University of Kentucky
UKnowledge

Information Science Faculty Publications | Information Science

4-2016

Personal Communication Networks and Their Positive Effects on Online Collaboration and Outcome Quality on Wikipedia

Michail Tsikerdekis
University of Kentucky, tsikerdekis@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/slis_facpub

 Part of the Communication Technology and New Media Commons, and the Library and Information Science Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Tsikerdekis, Michail, "Personal Communication Networks and Their Positive Effects on Online Collaboration and Outcome Quality on Wikipedia" (2016). *Information Science Faculty Publications*. 17.
https://uknowledge.uky.edu/slis_facpub/17

This Article is brought to you for free and open access by the Information Science at UKnowledge. It has been accepted for inclusion in Information Science Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Personal Communication Networks and Their Positive Effects on Online Collaboration and Outcome Quality on Wikipedia

Digital Object Identifier (DOI)
http://dx.doi.org/10.1002/asi.23429

Notes/Citation Information

Published in *Journal of the Association for Information Science and Technology*, v. 67, issue 4, p. 812-823.

© 2015 ASIS&T

This is the peer reviewed version of the following article: Tsikerdekis, M. (2016), Personal communication networks and their positive effects on online collaboration and outcome quality on Wikipedia. *Journal of the Association for Information Science and Technology*, 67: 812–823. doi: 10.1002/asi.23429, which has been published in final form at http://dx.doi.org/10.1002/asi.23429. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

This article is available at UKnowledge: https://uknowledge.uky.edu/slis_facpub/17

This is the peer reviewed version of the following article: Tsikerdekis, M. (2016), Personal communication networks and their positive effects on online collaboration and outcome quality on Wikipedia. *Journal of the Association for Information Science and Technology*, 67: 812–823. doi: 10.1002/asi.23429, which has been published in final form at http://dx.doi.org/10.1002/asi.23429. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Personal Communication Networks and their Positive Effects on Online Collaboration and Outcome Quality on Wikipedia

Michail Tsikerdekis

College of Communication and Information, University of Kentucky

316 Lucille Little Fine Arts Library, Lexington KY, USA

Phone: (859)-218-2298

Fax: (859)-257-4205

Email: tsikerdekis@uky.edu

Abstract

Online collaborative projects have been utilized in a variety of ways over the past decade such as bringing people together to build open source software or developing the world's largest free encyclopedia. Personal communication networks as a feature do not exist in all collaborative

projects. It is currently unclear if a designer's decision to include a personal communication network in a collaborative project's structure affects outcome quality. In this study, I investigated Wikipedia's personal communication network and analyzed which Wikipedia editors are utilizing it and how they are connected to outcome quality. Evidence suggests that people that utilize these networks are more experienced in editing high quality articles and are more integrated in the community. Additionally, these individuals utilize the personal communication network for coordinating and perhaps mentoring editors who edit lower quality articles. The value of these networks is demonstrated by the characteristics of the users who utilize them. These findings indicate that designers of online collaborative projects can help improve the quality of outcomes in these projects by deciding to implement a personal communication network in their communities.

Introduction

Personal communication is often seen as a necessary condition for collaboration and it is often considered a network phenomenon essential for online collaborative projects (Elliott, 2006). However, there is a knowledge gap regarding which users are making use of features that support personal communication in large online communities, what is the structure of personal communication networks and by how much their use affects collaborative outcome quality? Furthermore, stigmergic collaboration argued to take place in large collaborative groups eliminates the necessity for personal communication in favor of creative output. However, personal communication still exists in large collaborative groups and is ubiquitous but often not apparent. No prior work has examined the role of a collaborative platform supported by personal communication and investigated who utilizes this feature and how the individuals utilizing it affect collaborative outcome quality which can help determine the value of personal communication as a design component in online communities, a deeper understanding of its effects and a more meaningful contextualization of higher quality collaborative outcomes.

Personal communication is a component that can result in social networking. In turn, social networking is perceived to be an important element for supporting community aspects of online collaborative projects (Welser et al., 2011), however, its impact on the quality of collaborative outcomes remains unclear. In hindsight, one may intuitively assert that socializing should contribute to an improved outcome quality especially in task-oriented processes. However, it has been argued that impersonal communication and the lack of socializing is instead beneficial for task-oriented processes in collaborative projects (Walther, 1996) and especially in large groups (Elliot, 2006). This argument suggests that the usage of personal communication does not apply

to an online collaborative community as a whole. However, it is evident that users in collaborative project communities still make use of personal communication networks. Today, there are successful collaborative projects (e.g., Wikipedia) that do incorporate features that allow for social networking such as having a personal communication network. From a design standpoint, the creation of a space for personal communication in collaborative projects is not a requirement. Wikis, a technology used for collaborative projects, can be developed without spaces for personal communication. In their book on applications for wikis, J. A. West & West (2009) have mentioned cases of wikis aimed at promoting critical thinking that had discussion pages for tasks but not spaces for personal communication.

Wikipedia, the free encyclopedia collaborative project that ranks sixth most visited website worldwide ("wikipedia.org Site Overview," 2014), is an interesting case study because it allows users to voluntarily communicate with others through the so-called *user talk* pages, and, additionally provides evaluation of its collaborative outcomes (encyclopedic articles). User talk pages are public but owned by a specific individual and are used when Wikipedia editors desire to communicate with a particular editor. They constitute personal communication where personal is defined as something "carried on between individuals directly" ("personal", 2014) and under this definition privacy is not a necessary condition. These pages allowed me to analyze their data as a network based on messages exchanged between users. Given the task-oriented nature of Wikipedia, this network can be used by its users for social networking as much as coordinating. Therefore the term "*personal communication network*" in this paper is adopted as an inclusive term that views personal communication for analytical purposes as an action that occurs directly between individuals. In addition, Wikipedia provides an editorial review process through which

articles are rated based on a quality scale. However, little is known on what kind of users utilize Wikipedia's personal communication network and whether network utilization is connected to article quality. Understanding the effect of personal communication on article quality is important especially since the main expected outcome of Wikipedia is the production of encyclopedic articles ("Wikipedia:About," 2014). Additionally, establishing how utilization of the network occurs and by which users can be invaluable in understanding the value of personal communication networks. Collaborative projects are used in critical activities where quality is important. For example, they play a key role in the academic work of college students (Lim, 2009) and they offer unique capabilities for producing better information (Kane, 2011).

In this study, I provide findings that indicate that the success of Wikipedia could in part be attributed to the presence of its personal communication network, which is utilized more by users who tend to edit higher quality articles. The network, does not only support community aspects but also affects the performance of the overall community in achieving its collaborative outcome goal; high quality articles. Such finding invites a new perspective on the design of online collaborative projects and suggests that task-oriented processes could benefit from an independent space reserved for personal communication. In support of this, I analyzed Wikipedia's personal communication network in order to identify its structural properties and the likelihood of utilization by its users based on their characteristics. A visual analysis is provided for the network using k-core decomposition that helps examine its structural properties and highlight its hierarchical structure (Alvarez-Hamelin, Dall'Asta, Barrat, & Vespignani, n.d.). Additionally, I measured factors affecting participation in the personal communication network and I conducted a social network analysis aimed to obtain basic features such as reciprocity and transitivity that act

as indicators on how the network is being utilized. Finally, I used uniformed independent sampling (UIS) to develop a sampled version of the network and built an exponential random graph model (p^* models) to statistically test how user characteristics affect the probability of edge formation which is the likelihood of using the personal communication network.

Literature review

In information science and technology, network data have been analyzed to study how social media spaces are utilized by their users. Additionally, within information science and technology, the study of computer-supported cooperative work (CSCW) contains works relating to collaborative outcome quality and studies on virtual space usage in collaborative communities. These literature fields serve as the necessary background information for the methods and analyses that follow in this paper. Since the data used in this study have been based on Wikipedia, most of the literature review is also focused on Wikipedia studies. However, the results could potentially be generalized to online collaborative projects beyond Wikipedia such as open source projects.

Social Network Analysis

Social network analysis has been suggested as a method for understanding group work in sociotechnical systems (Goggins, Valetto, Mascaro, & Blincoe, 2013) and has been used to study small online group formation of work practices and identity development (Goggins, Laffey, & Gallagher, 2011). In its simplest form, a social network is represented by a set of nodes that are connected with each other based on one or more types of relationships. The type of relationship is usually called an edge, it is exclusive between two nodes and can be one-directional or bi-

directional, which introduces the notion of reciprocity between nodes. Descriptive statistics can be used to measure how dense a network is (density) or how centralized it is (centralization) along with other measures. Nodes and edges can also be assigned properties based on different attributes. These can be used to determine more detailed descriptive statistics or used in social network statistical modeling (e.g., exponential random graph models) that can produce results such as the likelihood for edge formation based on a node attribute.

In social media studies, social network analysis has being utilized in order to understand online communities. Thewall (2009) has examined the presence of assortativity (also known as homophily) in relationships on MySpace, which is the preference of network nodes to form connections with other similar nodes based on a certain attribute (e.g., gender). The study discovered that assortativity is present for nodes based on ethnicity, country or age, but, for the case of gender a no assortativity, if not the opposite, a dissasortativity effect was found. Similarly, Panzarasa, Opsahl, & Carley (2009) have studied the evolution of an online social network of students and investigated the nature of hubs (e.g., users with a disproportionately high number of connections to others) and whether these individuals are responsible for holding the system together. Another study used social network analysis to determine the durability of relationships in a virtual gaming world and identified that many relationships tend to become stronger over time (Shen, Monge, & Williams, 2014). These examples demonstrate the potential for social network analysis to identify network characteristics such as who uses a network, the factors that affect participation, the hubs that represent more active users in a network, and, furthermore help determine the value of the network as a structure of a larger software design potentially contributing to the success and sustainability of an online community.

Quality of Collaborative Outcome

Work on collaborative outcome quality has been investigated in Group Decision Support Systems (GDSS) where the quality of produced decisions is critical. Task structure has been found to affect decision quality (Lam, 1997). Disjunctive and conjunctive tasks in particular were found to increase decision quality in GDSS. Others have pointed that familiarizing with GDSS and the technology involved is a requirement before achieving superior decision quality to face-to-face settings (Chun & Park, 1998). The environment and its design as well as the task involved are factors that can potentially influence quality.

Wikipedia is a collaborative project that contains encyclopedic articles. It is the sixth most visited website worldwide ("wikipedia.org Site Overview," 2014) and arguably a representative example within the domain of CSCW research (Stvilia, Twidale, Smith, & Gasser, 2008). Studies have found that Wikipedia articles contain overall accurate information (Chesney, 2006; K. West & Williamson, 2009) although inaccuracies in articles have also been found (Rector, 2008). Article quality assessment on Wikipedia is made by Wikipedia projects. These are groups of users interested in a particular topic and are tasked with the development and improvement of articles around that topic. During an article quality evaluation, these groups place their article quality grade on an article's discussion page. Each article has its own article discussion page. Multiple projects can assign their grade to an article discussion page that revolves around their topic of interest but all need to consistently adhere to the guidelines set forth by the Wikipedia community ("Wikipedia:Version 1.0 Editorial Team/Assessment," 2014). Further, Wikipedia's design allows

individuals to vote and comment anonymously, which has shown to reduce the chances of conformity for groups (Tsikerdekis, 2013). As such, expressions of article quality are also more likely to be objective. Bots have been created for the purpose of keeping track of grades that are being placed on Wikipedia articles. There are currently millions of articles in the English language of Wikipedia. Studies have investigated how people collaborate on such online collaborative projects and attempted to identify potential patterns that contribute to article quality and as a result, effective collaboration.

One of the earliest studies that attempted to identify conditions that could increase article quality was conducted by Kittur & Kraut (2008). In their study, they identified two types of coordination, implicit and explicit. Explicit coordination can be identified in a number of ways but the most apparent is through article discussion pages. Discussions can be held there in order to resolve issues about an article (Viegas, Wattenberg, Kriss, & Ham, 2007). Implicit coordination on the other hand is more subtle such as a user taking a leadership role and doing the majority of the work. Kittur & Kraut (2008) found signs of implicit coordination using editor concentration; a measure obtained using the gini coefficient, which is based on the distribution of edits made by users on an article. Editor concentration was found to have a larger impact on article quality than explicit coordination on article discussion pages. Their result indicated that there are few individuals that do most of the work and it was theorized that the positive effect on quality is due to a reduction in coordination costs between users. Other factors such as the age of the article were found to have a significant effect but to a lesser degree.

Another study attempted to investigate the effects of global inequality revised as a gini coefficient metric based on the total contributions on Wikipedia for a set of editors who edited a specific article (Arazy & Nov, 2010). Global inequality can be seen as the representation of involvement in Wikipedia. An individual with more contributions to articles and other administrative pages on Wikipedia can be seen as more involved than an individual with less contributions. The findings suggest that global inequality affects implicit coordination and explicit coordination. In turn, global inequality was linked directly with article quality. In other words, editors who are more involved in the Wikipedia community tend to affect coordinating processes that in turn affect article quality.

A similar work aimed to demonstrate that the experience in editing high quality articles may be a factor that affects an article's quality. Stein & Hess (2007) have shown that articles in German Wikipedia that are of the highest quality had a significantly higher percentage of individuals with experience in editing high quality articles. Their findings indicated that articles benefit when individuals experienced in editing high quality articles contribute in the early stages of an article's life. The authors of the study suggested that when this occurs, it increases the probability that early articles will eventually become high quality articles.

Space utilization

Collaborative projects offer a variety of options for designers and users. Wiki technology in particular presents a diverse and versatile tool that can be applied to a variety of collaborative projects. There are wiki projects aimed for collaborative learning, knowledge construction or

critical thinking that offer virtual spaces in which their users can collaborate. These wikis are composed of components called namespaces. Namespaces can vary from regular content pages to user profile pages (J. A. West & West, 2009). The way that one can change the contents of a namespace can be determined either by software, by using user accounts and rights, or by policy. A good example of the latter can be seen on Wikipedia and its user pages; a type of a user profile page. There is no software preventing users from editing other users' pages however users conform to policies set forth by Wikipedia mandating to avoid doing substantial editing on another user's page ("Wikipedia:User pages," 2014). Similarly, article pages are differentiated from other pages meant for administrative purposes by a set of policies and design rather than strict enforcement by software.

Namespaces in Wikipedia have been the subject of studies that investigated how they are being utilized and attempted to establish their usefulness based on empirical evidence. A study that provided a bird's eye view of namespace usage was conducted by Kittur, Suh, Pendleton, & Chi (2007). Their findings point to a gradual increase in non-article namespace usage over the years since Wikipedia's launch. One of the namespaces that has seen a large increase in utilization is the user talk namespace. Substantial changes have been observed near the end of the year in 2003 as well as near the end of the year in 2006. The increase in user talk namespace usage is an indicator of an increasing one-on-one communication between Wikipedia editors. A potential reason for this increase may be due to the finding that the user talk namespace has been found to have a significant effect on the likelihood of an editor to be given administrative privileges (Okoli & Oh, 2007). Many editors may attempt to increase their community status through this namespace. On the other hand, this finding does not exclude the likelihood that the increase in user talk namespace

utilization could also prompt that this namespace is found to be useful for editors to coordinate with each other and increase effective collaboration.

Evidence of coordination has also been found by Viegas et al. (2007) that studied the article discussion namespace. Each article has its own article discussion namespace in which anyone can discuss issues related to an article (e.g., structural issues, missing points). The findings of the study indicated that Wikipedia pages focused on coordination and administration have increased since the website started in 2003. In particular, article discussion pages contain substantial planning about the article instead of disputes and resolutions on issues relating to an article that may arise. In their conclusion, the authors suggested a potential research direction to investigate the effect that technology has on forming conventional organization structures. Similar to other studies, it is evident that in Wikipedia there is a strong emphasis on group coordination and policy but it is also apparent that there is currently an unexplored potential on how the design of Wikipedia's technology may affect collaboration and as a result of this outcome quality.

Laniado, Tasso, Volkovich, & Kaltenbrunner (2011) studied various Wikipedia networks from article discussion networks to personal communication networks. In the article's discussion pages, editors had the tendency to reply to more inexperienced users. The same disassortativity effect was not found for the personal communication network. The authors concluded that the pattern observed in the article's discussion pages is due to their nature to serve as pages for articles that are meant to solve controversies as well as other Wikipedia-related policy issues. However, the metric that was used for assortativity was node degree as opposed to more rigorous user experience

metrics that were described in the previous section on quality metrics.

Research questions

Several questions arise when the aforementioned findings in the previous sections are taken into account. These focus on understanding which and how many of Wikipedia's users are making use of the personal communication network as well as how these users that make use of the network are influencing article quality in Wikipedia. The research perspective taken in this study views the personal communication network as a component that is utilized by some individuals based on their preference and attitudes. The individuals who utilize the network as well as how utilization takes place based on the structure of the network helps assess its value for the collaborative project and the potential effect it may have on outcome quality.

In previous studies (Laniado et al., 2011; Viegas et al., 2007), individuals who have never sent a message to another user through the user talk namespace were not taken into account since they were in no way connected in the personal communication network. However, if we are to understand the value of this network and who is using it, we need to examine what is the likelihood of being an isolate and why that may be the case. Furthermore, another unaddressed point is how experience affects the likelihood of being an isolate in the personal communication network. For example, if individuals who edit higher quality articles on average tend to make use of the network, that may imply that the network as a component is an important part of the interactions that lead to articles of higher quality.

Q1: Do isolates occur frequently in the personal communication network and what can the characteristics of non-isolates such as their attitude towards editing high quality articles tell us about the effects they may have on outcome quality?

The study conducted by Laniado et al. (2011) measured experience as the node degree, which is the amount of edges nodes have. However, other studies have used more rigorous metrics (Stein & Hess, 2008; Arazy & Nov, 2010) based on editor characteristics such as the proportion of editing high quality articles in order to determine high quality work experience in Wikipedia. As such, editors affecting article quality can be seen as more experienced and effective at what they do. Their characteristics can become metrics for evaluating how they utilize the personal communication network. Therefore, an aspect that needs to be investigated is how characteristics of high quality work determine the probability of having more edges and how likely it is to form edges between nodes that do similarly high quality work. The following research questions were formulated:

Q2: Does the network consist of editors who contribute more on articles of high quality demonstrating the value of the network towards outcome quality?

Q3: Do edges tend to form between nodes that edit on similar quality articles and what can this tell us of the value of the network towards outcome quality?

Methods

For this study, I chose the English-language branch of Wikipedia. At the time that this paper was written, Wikipedia had in total 33,414,234 pages on all namespaces of which 4,567,205 were articles belonging to the article namespace and a total of 21,888,378 registered users ("Special:Statistics," 2014). Out of the total registered users, some are solely readers (having made no contribution to the encyclopedia) while others are editors. There is a consensus among many researchers that just a small fraction of these editors are contributing most of the content in the online community (Kittur, Chi, Pendleton, Suh, & Mytkowicz, 2008; Kittur et al., 2007; Panciera, Halfaker, & Terveen, 2009; Priedhorsky et al., 2007).

In order to investigate my research questions, I obtained publicly available data through Wikipedia's database dumps. These datasets reach the orders of multiple terabytes ("Wikipedia:Database download," 2014) and include all pages and revisions that have ever been made to Wikipedia. I downloaded and processed the database snapshot up to January 2013 in order to retain article pages and other pages such as pages belonging to the user talk namespace as well as all revisions made to these pages. I retained all metadata but excluded text content and other data not relevant to the purposes of this study. Through the revision database, I also created a user database which included all users that made at least one revision to any namespace on Wikipedia. From the user database and further analyses, I excluded all the bots (active and inactive) on Wikipedia ("Wikipedia:Bots/Status," 2014) since this study is interested in human editing behavior. The study population is defined by these conditions.

Wikipedia contains a database of changes made to article quality grades maintained automatically

through bots that monitor changes in the article discussion pages where article quality grades are awarded. These bots retain a complete historical record of quality grade changes on an article and as such, one can inquire the article quality rating for a specific date. I obtained a complete copy of the database up until January 2013. Wikipedia also contains additional quality ratings for pages containing lists of articles that are maintained in the article namespace. These were excluded. The possible article quality grades that an article can obtain are the following: Featured Article, A, Good Article, B, C, Start and Stub. These were converted to corresponding numerical ratings ranging from seven to one. The distribution of article quality grades based on the maximum grade that they received by January 2013 are presented on Table 1.

Table 1. Distribution of article quality grades present in article quality database based on the maximum grade obtained by any Wikiproject evaluation per article.

Article quality grade	**Number of articles**
Featured Article	5,673
A	2,706
Good Article	20,686
B	107,111
C	140,382
Start	1,028,150
Stub	2,445,468

User behavioral patterns can be identified using database data. For example, every time a user edits an article he or she creates a revision in the revision database. The same applies for all pages in all

other namespaces on Wikipedia. Using the revision database, one can track all revisions ever made to articles as well as other namespaces on Wikipedia. By combining the articles where revisions have been made with the article quality database, one can also determine the article quality grade for each article at the time that a revision was made. Using the user database, revision database and article quality database, I calculated the following metrics for each user:

- *MQ*: mean quality of articles (namespace numeric identifier "0") that were edited. Articles with no quality rating were excluded. Article quality for each article was determined at the date and time where each revision was made on each article by a user.
- *MFA*: proportion of Featured Article (FA) quality articles. Articles with no quality rating were excluded. This metric is bounded between 0 and 1. Article quality for each article was determined at the date and time where each revision was made on each article by a user.
- *MA*: proportion of revisions on article namespace with numeric identifiers "0" and "6" in Wikipedia containing article pages as well as files.
- *MT*: proportion of revisions on article discussion namespace with numeric identifiers "1" and "7" in Wikipedia containing article discussion pages as well as file discussion page.
- *MU*: proportion of revisions on user namespaces with numeric identifiers "2" and "3" in Wikipedia containing user profile pages as well as user talk pages established for personal communication between members.
- *MW*: proportion of revisions on Wikipedia namespaces with numeric identifiers "4" and "5" in Wikipedia containing Wikipedia-related policy and regulation as well as discussion pages.
- *MO*: proportion of revisions on other namespaces with numeric identifiers "8" through "15", which contain other category pages such as meta pages.

- *MP*: proportion of revisions on portal namespace with numeric identifiers "100" and "101."
- *TR*: total revisions made by an editor to all namespaces throughout the history of the account.

The above data were merged with the user database in order to be used later as node properties once the social network is developed. The metrics *MQ* and *MFA* can be seen as representations of the experience and propensity of a user in conducting high quality work. *TR* is also meant to measure experience but in a more generalized way that does not incorporate quality. The sum of *MA, MT, MU, MW, MO,* and *MP* represents the work that a user does across all namespaces in Wikipedia. Additionally, I used the gini coefficient (Cowell, 2000) in order to develop a metric that establishes the distribution of revisions among these namespaces for each user. As described earlier in the literature review, the gini coefficient has been used as a measure before to study collaborative patterns on Wikipedia. *MU* was excluded from this metric since contributions to the user namespace translate as usage of the personal communication network. The metric (denoted as *GNR*) provides a value bounded between zero and one where zero is interpreted as an equal distribution between namespaces, while one represents unequal distribution calculated based on *MA, MT, MW, MO, MP*. The utilization of multiple namespaces is indicative of an individual integrating with all facets of the community rather than just focusing on one aspect (e.g., editing articles but not putting any effort in the development of policies, coordination and further administration in other namespaces).

Analysis and Results

Descriptive analyses

Even though previous studies (Laniado et al., 2011) have examined part of the structure of Wikipedia's personal communication network, there is still a need to understand what is the structure of the network today and what percentage of Wikipedia's population actually utilizes these pages. It is important to understand the extent of the network's utilization for the whole online community. Additionally, by determining if the network is central or decentralized as well as if it is composed of many local hubs, we can identify quantitatively how many individuals and to what degree are they using the network.

Using the revisions database that contains timestamps for each revision, I developed snapshots of the personal communication network for each year from 2004 until January 2013. The largest network built for January 2013 had 6,342,361 edges and 5,338,309 nodes. Given the size of the network, visualization and descriptive analysis is limited before having to apply uniform independent sampling (UIS) to reduce the size for further analyses. Traditional plotting algorithms are also impossible to render an image from which conclusions can be drawn. However, Large Networks Visualization Tool (LaNet-vi) can be used to determine the hierarchical structure of large networks and extract some conclusions from a k-core decomposition graph (Alvarez-Hamelin et al., n.d.). The graph is forged based on a series of concentric circular shells based on k-cores. Gradually increasing k-shells derived from k-cores are placed farther in the circle along with a gradual color change. The original degree of each vertex is shown based on a sizing scale that is adjusted using a logarithmic scale. Nodes that are found closer in the original graph are clustered together within each k-shell. More information on the features that can be extrapolated

from LaNet-vi can be found further online (Alvarez-Hamelin, Dall'Asta, Barrat, & Vespignani, 2006).

Figure 1 contains snapshots for the years 2004, 2005, 2008 and 2013. Numbers on the left side of the images signify the size of nodes based on their degree, while on the left legend, one can observe all the k-shells used in the graph and their corresponding color. In 2004, one can observe a centralized network with most of the large degree nodes located at the center of the graph. There is a minimal amount of clustering in the center of the graph but exterior k-shells contain substantial clustering between nodes. This is indicative of lower degree nodes being in close contact with one another, which is a positive indicator for a community under development. Within a year's time, in 2005, substantial differences in the social network can be observed. Coincidentally, this is the time when activity in user talk pages as well as in other namespaces on Wikipedia increased. One can observe nodes with a large degree, being farther out in the network (indicative of a decentralized network) as well as a further development and clustering of not only the outer but also the inner k-shells. In 2008, the network has increased dramatically and there is visible clustering that reveals many unified angular sectors. The clustering can be observed in inner k-shells as well as with just a few nodes that are isolated from their respected k-shell neighbors. In 2013, the network has retained most of the structure seen in 2008. Clustering even at this size has been retained at a high level even at such a large social network. There are also signs of decentralization with many high degree nodes appearing at outer shells of the graph. In summation, Wikipedia's personal communication network is structurally decentralized (with multiple contingency nodes) and is largely composed of a simple component. High quality images of the above described snapshots are also provided in the supplementary materials.

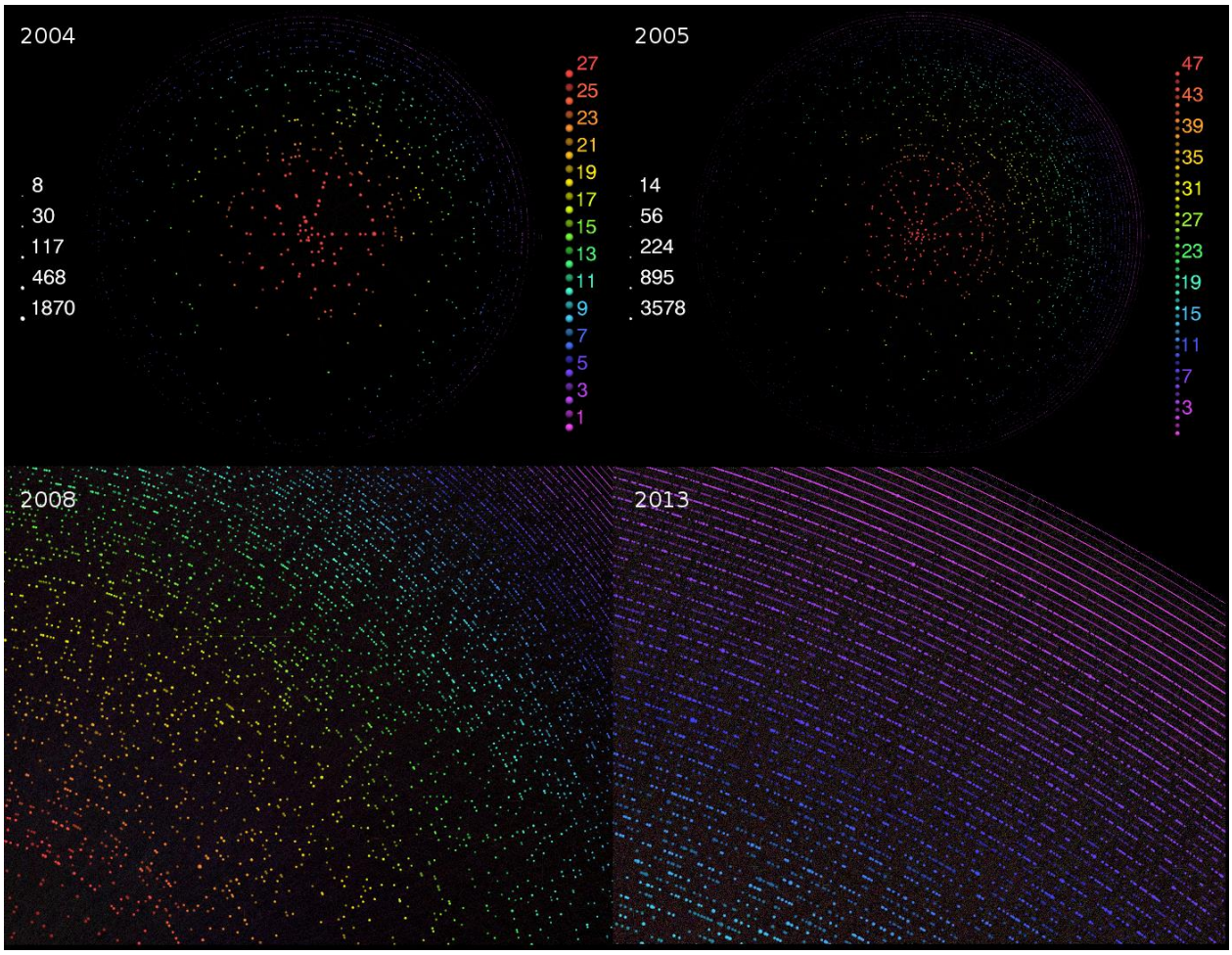


Figure 1: LaNet-vi (k-core decomposition) visualization of Wikipedia's personal communication network snapshots for the years 2004, 2005, 2008 and 2013. Numbers on the left side of the images signify the size of nodes based on their degree, while on the left legend, one can observe all the k-shells used in the graph and their corresponding color. Snapshots for years 2008 and 2013 are magnified since the number of individuals nodes are too many to appear clean in a full view. Notable high clustering is gradually more pronounced over the years.

Besides using network visualization in order to understand the network's high level structure, I have also used descriptive metrics in order to identify other characteristics of the network. As indicated visually, the network is highly decentralized. In fact, when measuring for its degree centralization (Wasserman & Faust, 1994) the result was 0.0037. The degree can also be expressed based on how many incoming and how many outgoing edges there are in the network. Without counting isolates, the network's in-degree centralization was 0.0015 and out-degree centralization was 0.0082. The mean total-degree for connected nodes is 6.16 with in-degree being 3.16 and out-degree being 19.4. The network's transitivity is 0.0081. The overall density of the network is extremely low, 2.23e-07 with isolates and 3e-07 without isolates included. The reciprocity of the network is 0.242 which translates to about a quarter of the communication on the network being two-way. These metrics are indicative of a network that is probably not used for socializing (or initiating relationships) but task specific usage. Furthermore, in the k-core decomposition produced by LaNet-vi no isolates were included. However, there is a substantial number of isolates in the network. Approximately 61.44 percent of users that have made at least one revision up until January 2013 are actually non-participating users in Wikipedia's personal communication network. This is a finding that relates to Q1 and it is further explored in the statistical analyses of isolate editors that follow.

Comparing isolates to participating users

While Wikipedia users do not have demographic information available, node characteristics can still be identified based on user behavior metrics that were calculated earlier on. I developed a binomial logistic regression model using a random sample of 100,000 nodes in order to determine

if any of the user patterns can act as predictors in a user's likelihood to participate in the personal communication network. To answer Q1, it is important to understand if the work conducted by users utilizing the network is different than those that do not. Users who participate in the network define the value of the network based on the work they do. Due to the fact that one of the variables appeared to have a perfect separation on the binary outcome, a binomial GLM using bias-reduction method developed by Firth (1993) was used (Kosmidis, n.d.). The regression model with all variables is displayed on Table 2.

Table 2: Binomial logistic regression model with variables predicting the likelihood of an editor participating in the personal communication network.

		95% CI for Odds Ratio		
	B (SE)	**Lower**	**Odds Ratio**	**Upper**
Non-isolates vs isolates				
Intercept	-3.679 (0.011)***	0.021	0.025	0.0311
MQ	-0.077 (0.005)***	0.917	0.926	0.935
MFA	-0.382 (0.084)***	0.579	0.682	0.804
GNR	5.846 (0.135)***	265.57	345.95	450.67
TR	-0.046 (0.001)***	0.954	0.955	0.957
Note: R^2 = .147 (Cox & Snell), .198 (Nagelkerke). Model χ^2 (11) = 15864, p <. .001. * p < .05, ** p < *.01*, *** p < .001.				

There is a strong indicator that individuals who tend to edit high quality articles utilize the personal communication network more. The model indicates that *MQ* and *MFA*, which represent editing patterns on quality articles, have an effect on the probability of someone being an isolate. The odds for *MQ* are low (too close to one) to draw any definitive conclusions but *MFA* odds clearly indicate that a unit increase in *MFA* results in a change of odds of being an isolate of 0.682. Another powerful indicator is provided by the odds of *GNR* where one unit increases result in a change of 345 of the odds of someone being in isolate. Values close to one for *GNR* indicate a skewed distribution of revisions among namespaces. Individuals who contribute to multiple namespaces and therefore are seen as being more integrated with the community will have a *GNR* closer to 0 and therefore will tend to be more likely in utilizing the personal communication network.

Network Statistical analyses

The size of the network along with the time complexity of metrics that can be applied to social networks was restrictive to proceed for further analyses in order to answer Q2 and Q3. For additional analyses, it was necessary to sample the network in order to produce a smaller and more computationally feasible version of the network for the analyses to follow. Uniform sampling has been used as a "ground truth" to be compared with other sampling methods (e.g., random walk) (Gjoka, Kurant, Butts, & Markopoulou, 2010). Given that there was a total overview of the complete network a uniform sampling method was an unbiased way of including representative nodes across the network.

I obtained a random sample of 5,000 nodes and their subsequent edges. The final graph consisted of 10,998 edges and 9,714 nodes. The sampled network's properties were similar to the full

network's although slightly more centralized (in-degree, 0.094; out-degree, 0.149; total-degree, 0.121) and denser (with isolates, 8.032e-05; without isolates, 1.053e-04). The statistics are expected to differ because the size of the network is different. On the other hand, local statistics such as reciprocity 0.246 and other node and local network characteristics aimed to answer our research questions are likely to remain similar to those of the larger network. Isolates were removed beyond this point to avoid having them affect the final statistical estimation. The network consisted of 7,921 nodes and 9,714 edges. A LaNet-vi visualization of the sampled network is shown in figure 2. The trail of nodes at the lower bottom of the rendered network, represent nodes that are connected with others (mostly dyadic connections) and are disconnected from the main component.

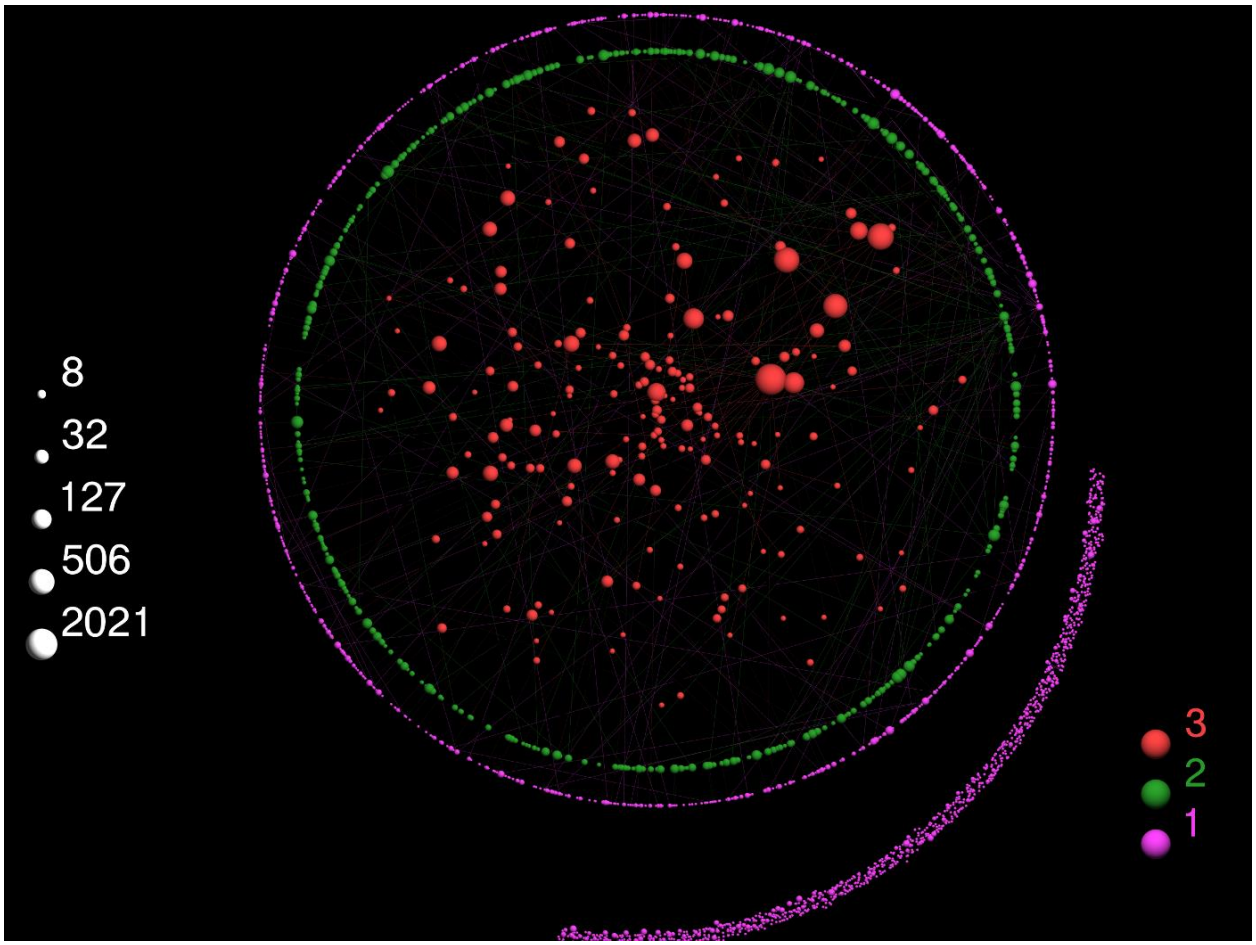


Figure 2: LaNet-vi (k-core decomposition) visualization of the sampled network. The left legend represents the size of nodes based on their degree and the right legend represents each k-shell along with its corresponding color. A small percentage of edges are also visible. The network presents a high degree of clustering within k-cores. The lower cluster of nodes represent nodes connected with each other but otherwise disconnected from the main component. The network appears to be centralized although nodes with relatively high degree can be found in outer k-cores as well.

In order to provide more concrete answers for Q2 and Q3, I developed an exponential random graph model (*p** model) to understand if there is assortativity in the network and identify what

leads people to communicate with others or being the recipients of communication. The use of such models can also assist in understanding further how edges are formed in the personal communication network. While descriptive statistics are limited to only the observed social network, *p** models can form the basis of statistical inference since their results incorporate the observed network as well as all other possible instances of alternative networks (Robins, Pattison, Kalish, & Lusher, 2007). To achieve this, the models utilize the Markov chain Monte Carlo (MCMC) algorithm which helps to estimate how likely is the observed network over all other probable networks that could have potentially existed. The outcome variable in *p** models is essentially the network itself while predictor variables become all terms that describe conditions within a network. These may include simple terms such as mutual ties but also more complex such as a term that determines the likelihood of edge formation between two nodes based on the numeric difference of one of their attributes (e.g., age). The ergm package in R (Hunter, Handcock, Butts, Goodreau, & Morris, 2008) provides several network terms that describe edges and nodes along with their attributes. The following terms were included for the analysis and a short description is included. Additional details for all the terms are provided in the library reference manual of the statistical language *R* for the *ergm* package ("CRAN - package ergm," 2014). I formed a model using the following terms:

- *Edges*: the overall probability of an edge being formed
- *Asymmetry*: the propensity of the network to have non-reciprocal edges

In order to answer Q2, I used the metrics that determine the average quality of articles that editors have previously edited as a node property. These were the metrics *MQ* and *MFA* and they were used with the following terms on my *p** model:

- *NodeCov.MQ*: whether there is an association between *MQ* and the probability of forming an edge
- *NodeCov.MFA:* whether there is an association between *MFA* and the probability of forming an edge

In order to answer Q3, I utilized the same metrics that represent quality of work but with terms that aim to determine assortativity between nodes as well as in-degree and out-degree assortativity:

- *AbsDiff.MQ*: whether there is assortativity between nodes in terms of their *MQ* scores
- *AbsDiff.MFA*: whether there is assortativity between nodes in terms of their *MFA* scores$Idegree_x.MQB_{val}$: whether there is an association between in-degree (receiving personal communication) and assortativity based on *MQB* for *x* being the degree of one and two. *MQB* is a binary metric ("low", "high") derived from *MQ* based on the middle value 3.5 from the range of possible values that *MQ* can obtain (one through seven). *val* represents one of the two values of *MQB* used to determine assortativity for nodes with the same value.
- $Odegree_x.MQB_{val}$: whether there is an association between out-degree (sending personal communication) and assortativity based on *MQB* for *x* being with degree of one and two.*MQB* and its parameters have been derived similar to $Idegree_x.MQB_{val}$.

All of the above mentioned terms were added to a model which was computed using Markov chain Monte Carlo (MCMC) estimation. For evaluation of the performance of MCMC, which can affect the statistical outcome, I used diagnostic tools provided by the package ergm (mcmc.diagnostics function). Plots and statistics indicated that the model converged, which is a good indication that

results are accurate. Model results are presented on table 3.

Table 3: Exponential random graph model for sampled Wikipedia personal communication network. Terms are presented along with their coefficient (B) and standard error (SE) as well as the significance (*p*) for each term. Odds ratio is also as derived from the coefficients.

	B (SE)	**Odds Ratio**
Edges	-4.833 (0.032)***	0.008
Asymmetry	-3.943 (0.019)***	0.019
AbsDiff.MQ	-0.161 (0.01)***	0.851
AbsDiff.MFA	- 0.505 (0.367)	0.604
NodeCov.MQ	0.039 (0.006)***	1.04
NodeCov.MFA	0.136 (0.34)	1.145
$Idegree_{1}.MQB_{high}$	-0.510 (0.081)***	0.601
$Idegree_{2}.MQB_{high}$	-1.749 (0.163)***	0.174
$Idegree_{1}.MQB_{low}$	0.039 (0.025)	1.040
$Idegree_{2}.MQB_{low}$	-1.248 (0.046)***	0.287
$Odegree_{1}.MQB_{high}$	0.959 (0.084)***	2.609
$Odegree_{2}.MQB_{high}$	-0.453 (0.141)**	0.636
$Odegree_{1}.MQB_{low}$	0.452 (0.025)***	1.572

$Odegree_2.MQB_{low}$	-1.263 (0.050)***	0.283
Note: * $p < .05$, ** $p < .01$, *** $p < .001$.		

Based on table 3, the probability of forming an edge is 0.79 percent (which is also the fraction of realized edges over all possible edges). This coincides with previous findings that personal communication usage is sparse. The odds ratio of adding a one-way edge is 0.019 which indicates that asymmetric relations reduce the probability for edge formation. There is an additional number of local terms (e.g., triads) that could have been included in the model but go beyond the scope of the research questions addressed by this study and some local properties are the subject of another future study. These representative basic terms were used as a control to verify that basic network properties of the sampled network approximate those of the full network. Other variants of terms and parameters were attempted (such as having results for more in-degree and out-degree), however, the models did not converge.

Q2 and Q3 were addressed using further terms in the model. The estimate for node assortativity (*AbsDiff.MQ)* based on the mean rating of articles that users have edited in the past (*MQ*) is significant and negative, which indicates that there is assortativity between nodes. Users that edit high quality articles tend to contact users that also edit high quality articles. However, the statistic for node assortativity using the percentage of featured quality articles edited by users (*MFA*) is not significant although still negative. On the other hand, there were statistically significant results for edge formation (*NodeCov.MQ)* based on a user's *MQ*. The higher the *MQ* is, the higher the

likelihood for edge formation.

Further, behavioral patterns for users are apparent based on the in-degree and out-degree statistics based on the *MQB* and the tendency for assortativity. Users that belong to the "high" *MQB* category are less likely to receive communication (in-degree) from other users that belong to the "high" *MQB* category. The same disassortativity is also seen for users that are in the "low" *MQB* category ($Idegree_1.MQB_{low}$ is not significant but $Idegree_2.MQB_{low}$ is significant and negative). On the other hand, out-degree of one displays an assortativity effect for both users in the "high" and "low" *MQB* categories. However, that effect is diminished and switched to a disassortativity for out-degree 2. That means that users that were in contact with two other users tend to edit on average different quality articles than their connections. Given the findings from the rest of the statistics, a person likely to have a high out-degree is one that edits higher quality articles and is therefore likely to communicate with editors that on average edit lower quality articles. On the other hand, users who receive many messages from others tend to edit lower quality articles than the people that are contacting them. Simply put, there is an opposite attraction between users in terms of *MQB* value.

In order to visually observe the effect discovered for Q2, I developed a time-lapse visualization of the sampled network with appropriate colors, where lighter colors indicate higher *MQ* and darker colors for nodes indicate lower *MQ*. This visualization can be found in the supplementary materials, where the tendency of users with higher *MQ* contacting others with lower *MQ* can be observed.

Discussion

The results provided by the various analyses conducted in this study have provided a better picture on the presence of a personal communication network in collaborative projects. These results indicate certain conditions that confirm there is only a portion of the community that utilizes the network. The personal communication network on Wikipedia is a large, non-centralized, and, sparse network with a higher tendency for having nodes with multiple outgoing communications rather than incoming communications. Transitive relations are also uncommon. All these suggest that the network is not primarily used for socializing. In fact, the network contains only a small percentage of reciprocal edges.

A more interesting finding with a relatively medium level of accounted variance was the model identifying the probability for non-participating users based on user patterns in order to address Q1. The findings suggest that users who do not utilize the personal communication network tend to contribute less to high quality articles and deliver content to fewer namespaces than participating users to the network. To the best of my knowledge, this is the first finding in literature to connect users who are associated with high quality work and the use of a personal communication network. The network as a component of Wikipedia functions to serve a particular group of users. The result indicates that people that utilize the network tend to edit high quality articles and are well integrated with the community through editing on average more in various namespaces on Wikipedia. Even more importantly, it seems that the total number of revisions that users have throughout their account's lifetime do not seem to affect the likelihood of participating in the personal communication network. In other words, even senior users who have made many

revisions on Wikipedia are not guaranteed to participate on the personal communication network. This is another indication that only users who do quality work end up utilizing the network.

Furthermore, based on the findings of the *p** model, there seems to be a clear connection between users that edit high quality articles and those that edit on average low quality articles. Models suggest that editors that are experienced and have on average edited higher quality articles tend to communicate more with editors that edit on average lower quality articles. A likely explanation of this phenomenon could be that more experienced users tend to communicate (or perhaps implicitly mentor) more inexperienced users in terms of editing high quality articles.

In summary, the personal communication network is linked with users that edit high quality articles and are more active in Wikipedia, which suggests that Wikipedia's success may be in part attributed to its personal communication network. The value of the network is determined by the users who are using it. Since these users conduct work on high quality articles, it implies that the personal communication network allows them to achieve their tasks and as such the network itself as a component of the online community contributes to collaborative outcome quality.

Finally, the broader implications of this study also invite a reevaluation of the role that communication plays in large collaborative projects. Based on the results, communication channels that allow for coordination and perhaps even socializing are necessary to improve the quality of work being performed even for large groups such as the ones present in Wikipedia. This

is not to suggest that collaboration cannot exist without these channels but rather that the quality would not be as high as with projects that do have such channels for personal communication. The effect may be enhanced for open source collaborative projects where communication may be necessary to articulate the reasons behind the implementation or changes to certain parts of a code. Comments for revisions may not be sufficient for such tasks and therefore a channel for personal communication may be necessary. Personal communication networks can be considered as inexpensive solutions to implement for most websites. They have been around for a long time in social media and that is perhaps indicative of users preferring to have such a feature. It is therefore recommended that website developers and designers aim to implement features that allow for personal communication networks.

Limitations

This study has investigated the effects of Wikipedia's personal communication network and therefore caution is advised in generalizing the results to other online collaborative projects with different designs. For example, if Wikipedia's user talk namespace were to be private, results may have been different. In general, under this design and any similar design these effects are likely to appear. Additionally, while *p** models have been used in other studies, the use of MCMC always introduces a certain amount of noise that may affect the results. Finally, while a disassortativity effect has been found for nodes based on *MQB* it is unclear on whether these results remain for nodes that have higher degrees than 2.

Future work

While statistical analyses provided by this study present a bird's eye view of Wikipedia's personal communication network, future studies should also focus on qualitative findings that demonstrate with cases, what is indicated here using statistics. Furthermore, we do not know the motivations behind the use of the personal communication network. Surveys should also explore if editors with high out-degree identify themselves in a leadership or mentoring position within the community.

Conclusion

In this study, I have demonstrated a link between collaborative outcome quality and the current design of Wikipedia allowing for the existence of a personal communication network. The personal communication network was found to be a valuable tool in this context by examining the user population that utilizes it, which is a subset of Wikipedia's total population and has specific characteristics relating to high quality work. The network is utilized by a fraction of Wikipedia's users who tend to edit high quality articles and are more integrated with the community. While causality cannot be identified, the fact that users that make an impact are utilizing the personal communication network is an indication of the benefits of having a personal communication network in online collaborative projects. According to the findings, it is a tool for coordination and perhaps even mentoring. Wikipedia's design is one of many that can exist in online collaborative projects. Given that the success of these projects is dependent on their communities, it is essential that we understand how social interactions are affected by the platform design. Evidently, these designed features may play a large role in the success of online collaborative communities.

References

Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2014, June 4). k-core decomposition : a tool for the visualization of large scale networks. Retrieved from http://arxiv.org/abs/cs/0504107

Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2006). *LaNet-vi in a Nutshell*. Technical Report. Indiana University School of Informatics. Retrieved from http://lanet-vi.soic.indiana.edu/lanetvi-2_2.pdf

Arazy, O., & Nov, O. (2010). Determinants of wikipedia quality. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10* (pp. 233–236). New York, New York, USA: ACM Press. doi:10.1145/1718918.1718963

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, *11*(11). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/1413

Chun, K. J., & Park, H. K. (1998). Examining the conflicting results of GDSS research. *Information & Management*, *33*(6), 313–325. doi:http://dx.doi.org/10.1016/S0378-7206(98)00038-X

Cowell, F. A. (2000). Measurement of inequality. In F. Bourguignon & A. B. Atkinson (Eds.), *Handbook of Income Distribution* (pp. 87–166). Amsterdam, Netherlands: North-Holland.

CRAN - package ergm. (2014, January 31). *cran.r-project.org*. Retrieved from http://cran.r-project.org/web/packages/ergm/index.html

Elliott, M. (2006). Stigmergic Collaboration: The Evolution of Group Work. *M/C Journal*, *9*(2). Retrieved from http://journal.media-culture.org.au/0605/03-elliott.php

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. doi:10.1093/biomet/80.1.27

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2010). Walking in Facebook: A Case Study of Unbiased Sampling of OSNs. In *2010 Proceedings IEEE INFOCOM* (pp. 1–9). IEEE. doi:10.1109/INFCOM.2010.5462078

Goggins, S. P., Laffey, J., & Gallagher, M. (2011). Completely online group formation and development: small groups as socio-technical systems. *Information Technology & People*, *24*(2), 104–133. doi:10.1108/09593841111137322

Goggins, S. P., Valetto, G., Mascaro, C., & Blincoe, K. (2013). Creating a model of the dynamics of socio-technical groups. *User Modeling and User-Adapted Interaction*, *23*(4), 345–379. doi:10.1007/s11257-012-9122-3

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, *24*(3).

Kane, G. C. (2011). A multimethod study of information quality in wiki collaboration. *ACM Transactions on Management Information Systems*, *2*(1), 1–16. doi:10.1145/1929916.1929920

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2008). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proceedings of Alt.CHI at CHI 2007*. San Jose, California, USA: ACM.

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work - CSCW '08* (pp. 37–46). New York, New York, USA: ACM. doi:10.1145/1460563.1460572

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 453–462). New York, NY: ACM. doi:10.1145/1240624.1240698

Kosmidis, I. (2014, June 4). brglm: Bias reduction in binary-response Generalized Linear Models. *cran-r-project.org*. Retrieved from http://cran.r-project.org/web/packages/brglm/index.html

Lam, S. S. K. (1997). The Effects of Group Decision Support Systems and Task Structures on Group Communication and Decision Quality. *Journal of Management Information Systems*, *13*(4), 193–215. Retrieved from http://www.jmis-web.org/articles/v13_n4_p193/index.html

Laniado, D., Tasso, R., Volkovich, Y., & Kaltenbrunner, A. (2011). When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages. In *Proceedings of International AAAI Conference on Weblogs and Social Media*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2764

Lim, S. (2009). How and why do college students use Wikipedia? *Journal of the American Society for Information Science and Technology*, *60*(11), 2189–2202. doi:10.1002/asi.21142

Okoli, C., & Oh, W. (2007). Investigating recognition-based performance in an open content community: A social capital perspective. *Information & Management*, *44*(3), 240–252. doi:10.1016/j.im.2006.12.007

Panciera, K., Halfaker, A., & Terveen, L. (2009). Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 51–60). New York, New York, USA: ACM. doi:10.1145/1531674.1531682

Panzarasa, P., Opsahl, T., & Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, *60*(5), 911–932. doi:10.1002/asi.21015

personal. 2014. In Merriam-Webster.com. Retrieved July 25, 2014, from http://www.merriam-webster.com/dictionary/personal

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Conference on supporting group work - GROUP '07* (pp. 259–268). New York, New York, USA: ACM. doi:10.1145/1316624.1316663

Rector, L. H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*(1), 7–22. doi:10.1108/00907320810851998

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191. doi:10.1016/j.socnet.2006.08.002

Shen, C., Monge, P., & Williams, D. (2014). The evolution of social ties online: A longitudinal study in a massively multiplayer online game. *Journal of the Association for Information Science and Technology*, n/a–n/a. doi:10.1002/asi.23129

Special:Statistics. (n.d.). In *Wikipedia*. Retrieved July 27, 2014, from http://en.wikipedia.org/wiki/Special:Statistics

Stein, K., & Hess, C. (2007). Does it matter who contributes: a study on featured articles in the german wikipedia. In *Proceedings of the 18th conference on Hypertext and hypermedia HT 07* (pp. 171–174). New York, New York, USA: ACM. doi:10.1145/1286240.1286290

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001. doi:10.1002/asi.20813

Thelwall, M. (2009). Homophily in MySpace. Journal of the American Society for Information Science and Technology, 60(2), 219–231.

Tsikerdekis, M. (2013). The effects of perceived anonymity and anonymity states on conformity and groupthink in online communities: A Wikipedia study. *Journal of the American Society for Information Science and Technology*, *64*(5), 1001–1015. doi:10.1002/asi.22795

Viegas, F., Wattenberg, M., Kriss, J., & Ham, F. (2007). Talk Before You Type: Coordination in Wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (pp. 78–78). Washington, DC, USA: IEEE. doi:10.1109/HICSS.2007.511

Walther, J. B. (1996). Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, *23*(1), 3–43. doi:10.1177/009365096023001001

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference on - iConference '11* (pp. 122–129). New York, New York, USA: ACM. doi:10.1145/1940761.1940778

West, J. A., & West, M. L. (2009). *Using wikis for online collaboration: The power of the read–write web*. San Francisco, California, USA: John Wiley & Sons.

West, K., & Williamson, J. (2009). Wikipedia: friend or foe? *Reference Services Review*, *37*(3), 260–271. doi:10.1108/00907320910982758

wikipedia.org Site Overview. (2014, May 1). *Alexa.com*. Retrieved from http://www.alexa.com/siteinfo/wikipedia.org

Wikipedia:About. (n.d.). In *Wikipedia*. Retrieved June 4, 2014, from https://en.wikipedia.org/wiki/Wikipedia:About

Wikipedia:Bots/Status. (n.d.). In *Wikipedia*. Retrieved June 4, 2014, from https://en.wikipedia.org/wiki/Wikipedia:Bots/Status

Wikipedia:Database download. (n.d.). In *Wikipedia*. Retrieved June 4, 2014, from https://en.wikipedia.org/wiki/Wikipedia:Database_download

Wikipedia:User pages. (n.d.). In *Wikipedia*. Retrieved June 4, 2014, from https://en.wikipedia.org/wiki/Wikipedia:User_pages

Wikipedia:Version 1.0 Editorial Team/Assessment. (n.d.). In *Wikipedia*. Retrieved June 4, 2014, from http://en.wikipedia.org/?title=Wikipedia:Version_1.0_Editorial_Team/Assessment