2015

Request for Adminship (RFA) within Wikipedia: How Do User Contributions Instill Community Trust?

Christopher Kreider
Fulfillment.com, ckreider@fulfillment.com

Nima Kordzadeh
Idaho State University, kordnima@isu.edu

Follow this and additional works at: http://aisel.aisnet.org/sais2015

Request for Adminship (RFA) within Wikipedia: How Do User Contributions Instill Community Trust?

Christopher Kreider
Fulfillment.com
ckreider@fulfillment.com

Nima Kordzadeh
Idaho State University
kordnima@isu.edu

ABSTRACT

Research into user driven communities such as Wikipedia often focuses on community level characteristics, paying less attention to social processes performed within those communities. One important process is determining trust of community members. In Wikipedia, the access to specific tools requiring trust is determined through a community voting process known as request for adminship (RFA). In this study, we examine the impact of different forms of contribution made by adminship candidates on the community's overall decision as to whether to promote the candidate to administrator. To do so, we collected data on 754 RFA cases and used logistic regression to test four hypotheses. Our results supported the role of total contribution, and clarification of contribution in RFA success while the impacts of social contribution was partially supported and the role of content contribution was not supported. Also, both control variables (tenure and number of attempts) showed significant relationships with RFA success.

KEYWORDS

Wikipedia, community trust, request-for-adminship (RFA), editcountitis

INTRODUCTION

Wikipedia, self-billed as "The Free Encyclopedia", is an online collaborative project that provides a wide range of freely available encyclopedic articles. The widespread use of Wikipedia has also revolutionized the encyclopedia section of the publishing industry. For instance, Britannica stopped publishing their popular encyclopedia in 2010 after 244 years.[1]

Wikipedia is a large-scale, distributed collaborative project, made possible by a vast network of independent volunteers located around the world. Volunteers, also known as editors, or users, perform the variety of tasks necessary for success including: content contribution, social coordination, and content maintenance (Viégas, Wattenberg, McKeon, 2007; Welser, Cosley, Kossinets, Lin, Dokshin, Gay, and Smith 2011). While anybody can register for an account on Wikipedia.com, and contribute to the project, certain tasks are limited to those whom the community has deemed as trustworthy. Activities such as modifying the main page, deleting inappropriate content, blocking users and changing user login credentials are restricted to a class of volunteers known as administrators. At the core of this is the usually week-long request for adminship (RFA) process, during which a registered editor desiring to become an administrator, submits themselves to a community peer review process.

This process starts by creating a request page, and answering some community agreed upon pre-defined questions. After an RFA has been presented to the community, a community member provides a set of descriptive statistics on the users' history with the project providing an easily accessible high level summary of how the candidate has contributed to the project. These summary statistics provide information as to how long the user has been registered (tenure), the number of contributions the editor has made to various areas of the project, and how consistent the user is at following measurable desirable behavior, such as providing a descriptive summary of every change. The same statistics are provided for every user who starts the RFA process, and are readily available to all community members. To assess whether a candidate is suitable for promotion, community members are encouraged to assess the candidate's edit history and the cumulative contribution that the candidate has made to the project. During this process, all members of the community are encouraged to scrutinize the candidate's history with the project, and vote either "Support", or "Oppose". At the end of the process, a member of a very limited, highly trusted and well established group of volunteers analyzes the outcome to determine community consensus. Candidates who receive an appropriate level of support from the community (usually greater than 70% support) are promoted to administrators.

At the core of the RFA evaluation process is the permanent record of every change that a volunteer makes, called an "edit". An *edit* represents any change made to any editable page within the project. For example, the removal of an extra comma and

---

Proceedings of the Southern Association for Information Systems Conference, Hilton Head Island, SC, USA March 20th–21st, 2015

1

the writing of a new article could both be considered a single edit. Furthermore, contributions to coordinative and social aspects of the project also count as edits. The *total edit count* is the sum total of all edits a volunteer has made to the project, regardless of size, quality, and currency, and is usually the first summary statistic posted once a candidate has started their RFA. The summary statistic *total edit count* has many obvious problems, including the shortcoming of being unable to determine the quality of the contribution. To highlight this shortcoming, it would be trivial for a volunteer to perform 10,000 edits of adding a single space to the end of an article; while the magnitude of contribution may appear large according to the *total edit count*, even though they made no actual value adding contribution to the project. A remedy to this issue is considering contribution to namespaces as a supplemental measure of contributions.

Namespaces are divisions of contributions to the Wikipedia project and are used for maintaining specific tasks and housing specific content (Beschastnikh, Kriplean, and McDonald, 2008). For instance, Main namespace "contains all encyclopedia articles, lists, disambiguation pages, and encyclopedia redirects" [2]; whereas, User namespace "contains user pages and other pages created by individual users for their own personal use" [2] and Wikipedia namespace "contains many types of pages connected with the Wikipedia project itself: information, policy, essays, processes, discussion, etc." [2] In fact, Main, User, and Wikipedia are the three major namespaces commonly used in the Wikipedia project.

Each namespace (e.g., *Main* and *User*) is associated with two pages: a subject namespace/page, which contains the article itself and Talk namespace/page, which is an area dedicated to coordination between asynchronous editors and discussions on the content of the associated subject page. For instance, the talk page for a discussion on improvements to the article *Australia* is named *Talk: Australia* and is used by editors for discussing how to improve that article. Thus, *Main*, *User*, and *Wikipedia* namespaces are associated with *Main Talk*, *User Talk*, and *Wikipedia Talk* pages, respectively. These six namespaces, that include subject namespaces and their associated talk namespaces, and the content within them reflect a major portion of editors' contributions to the Wikipedia project. When a candidate submits an RFA, a community member provides summary statistics such as those seen in Figure 1.

```
Mainspace:       3933
Talk:             107
User talk:       3203
User:             631
Wikipedia talk:    4
Wikipedia:        539
avg edits per article          1.81
earliest        17:38, 24 April 2012
number of unique articles       4686
total                           8499
2012/4            1
2012/7            0
2012/8          163
2012/9           67
2012/10        3456
2012/11        4812
```

Figure 1: Sample Summary Statistics From RFA

Those summary statistics can be used by the community to assess an RFA and make a decision on promoting a user in the project. In this study, we aim to understand to what extent the summary statistics on editors' contributions can explain the likelihood of successful promotions in the community through RFA process. In other words, we seek to understand what pieces of information are used by the community to decide on promotion requests. The results provide insights for community developers, providers, administrators, and managers as well as for researchers working in the relevant areas such as virtual communities and social processes within online environments and open source projects.

BACKGROUND

Wikipedia is a massive online encyclopedia whose content is added and managed completely by volunteers (Rahman, 2009). Research on Wikipedia has rarely looked at specific processes, and instead focuses on general characteristics of the site itself. Some research argues that Wikipedia projects can be considered communities of practices (CoPs), which have associated behaviors and norms (Hara and Shachaf, 2009). Others explore Wikipedia through a lens of social networks, examining the

---

[2] http://en.wikipedia.org/wiki/Wikipedia:Namespace, Retrieved December 2014.

concept of small-world theory (Ingawale, Dutta, Roy, and Seetharaman, 2009). Another area commonly studied in open community environments is why users contribute, and what differentiates those who contribute from those who don't (Olivera, Goodman, and Swee-Lin, 2008). A final area of study explores the above questions through a lens of economics and game theory, hypothesizing that Wikipedia should suffer from "under provision" (Rahman, 2009). Despite the growing body of research that explores the overall characteristics of this social website, research focusing on specific processes such as determining community trust and promoting users/editors in the community is less covered (Jin, Robey, and Boudreau 2007).

Among very few studies that focused on this subject, Burke and Kraut (2008a, 2008b) examined what characteristics are determinants of a successful promotion. Their model broke contribution characteristics down into areas such as strong edit history, varied experience, user interaction, helping with chores, and observing consensus and edit summary usage. We extend Burke and Kraut's (2008a) study and include other pieces of publicly available information about editors' contributions to the Wikipedia project and examine how those types of information may be considered by the community to assess RFAs.

RESEARCH FRAMEWORK AND HYPOTHESES

As discussed earlier, various statistics on user contribution are provided to the Wikipedia users who will evaluate one's request for promotion in the community. This information contains different values each of which belongs to a specific facet of the candidate's contribution that can be used by the voters as a basis for supporting or opposing a candidate. The first value is the total number of times a user has created or edited an article and indicates the user's total contribution. It can be a positive sign for the voters to support a candidate. We hypothesis:

H1: There is a positive relationship between total contribution and success in RFA process.

Although total contribution can demonstrate the candidate's commitment and trustworthiness, a particular area of concern is the perception of editcountitis and entitlement (Collier, Burke, Kittur, and Kraut, 2008). That is, those who have exceptionally large numbers of edits, or have been members of the project for lengthy periods of time, feel entitled to be elected as administrators. As a result, editors desiring to become administrators may engage in editing behavior that inflates their edit counts and days since registration. These activities include making many menial edits (e.g. changing capitalization of a letter on thousands of pages), and making a few menial edits daily for lengthy periods of time (e.g. making a few minor changes each day for a year). Therefore, the voters may not rely merely on total contribution as the only indicator of the candidate's merit.

Therefore, voters can notice the distribution of the candidate's contribution in each type of collaborative effort. As explained earlier, contribution statistics are divided into namespaces. Wikipedia users can contribute to both subject and talk namespaces (pages). Thus, we divide the users' contributions into two forms: content contribution and social contribution. Content contribution pertains to the users' participation in content creation and maintenance on the basic namespaces (main, user, Wikipedia); whereas, social contribution refers to engaging in the discussions on a talk page that revolves around a specific page (talk, user talk, Wikipedia talk). Accordingly, we hypothesize:

H2: There is a positive relationship between content contribution and RFA success.

H3: There is a positive relationship between social contribution and RFA success.

The research by Burke and Kraut (2008a) identified *edit summary usage* as a factor that significantly influenced RFA success. Edit summary usage measures the information users provide about each contribution they make. Users with low edit summary usages generally provide no summary of their actions, decreasing transparency of the projects work and increasing difficulty of other users to summarize their work. In our study, accordingly, we conceptualize it as "clarification of contribution" and hypothesize:

H4: There is a positive relationship between clarification of contribution and RFA success.

Consistent with the extant literature, we include number of attempts and tenure (days since a user has registered on Wikipedia) as the control variables. Figure 2 presents the research framework and hypotheses.
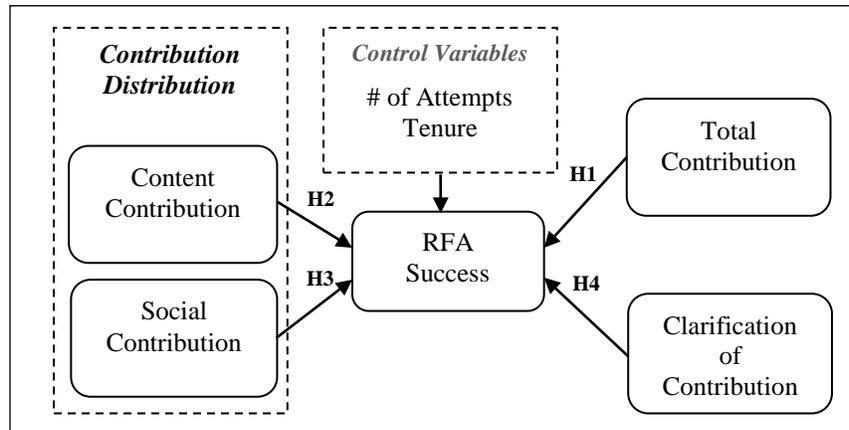
Figure 2. Research Framework

METHOD

We directly collected data from Wikipedia.com on successful and unsuccessful RFA's over a period of two years (N=954). From our sample, 200 abnormal attempts with no counting votes, or with no edits outside of the RFA process were eliminated from the sample. After removing these abnormal RFA's we were left with a final total number of N=754. Edit counts and number of support and oppose votes for our final sample were collected through a custom written edit counter program, developed in Java. Due to the nature of Wikipedia, past information is available for analysis with detailed date and time stamps indicating when the action occurred (Priedhorsky, Chen, Lam, Panciera, Terveen, and Riedl, 2007). A random selection of the RFAs was manually inspected to verify that the data was being properly collected; all manually gathered values matched those collected automatically.

ANALYSIS AND RESULTS

Each RFA has a dichotomous outcome, promotion or no promotion; hence, binomial logistic regression was chosen for our analysis (Hosmer and Lemeshow, 2004). The regression model has ten potential predictors. The first three variables were logarithm of days since registration (a.k.a. tenure), logarithm of total edit counts since registration, and number of attempts (e.g. an editor that failed their first attempt may try again at a later date). Due to the varying nature of days and total edits, we chose to take the log transformation of these measures to make their values more linear. The model also initially included six variables representing edit count within each of the six namespaces, divided by total edit count. These variables show what percentage of one's contribution is devoted to each of the six namespaces. In addition, the model included edit summary usage, which measures the information users provide about each contribution they make. As discussed earlier, edit summary usage is a measure of clarification of contribution. Among those ten potential predictors, log(tenure) and number of attempts are considered control variables since they are not directly related to the proposed hypotheses in this study.

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
TENURE	1.470	.352	17.430	1	.000	4.348	2.181	8.670
NUM_ATTEMPTS	-.681	.137	24.654	1	.000	.506	.387	.662
TOTAL_EDITS	1.170	.247	22.474	1	.000	3.223	1.987	5.229
EDIT_SUM_USAGE	8.262	1.384	35.628	1	.000	3875.646	257.080	58427.962
USER_TALK	-2.315	.777	8.879	1	.003	.099	.022	.453

We

WIKIPEDIA_TALK	9.926	3.813	6.777	1	.009	20455.409	11.624	3.600E7	used
CONSTANT	-15.610	1.702	84.154	1	.000	.000			
Cox & Snell R-Square = 0.31									

Table 1. Model Coefficient Estimates

forward selection approach based on which the statistical package (SPSS 18.0) decides what predictor variables may have significant relationship with the outcome; hence, should be included in the final model. In result of this process, six predictor variables showed significant impact on the outcome and were included in the model (Table 1). Those six variables are total edits, edit summary usage, contribution to User Talk, contribution to Wikipedia Talk, days since registration, and number of attempts.

We further analyzed the predictive power of the final model using a classification table measuring correct classification percentages. This procedure applied the empirically collected voting data to the model to determine if the model would predict whether the decision would be to promote, or not to promote. The model predicted outcome is then compared to the empirically collected outcome to determine how accurate the model is at predicting the true outcome. Overall, the outcome of 76.6% of the cases were predicted correctly using the regression model (Table 2).

Observed			Success/Fail		Percentage Correct
			Fail	Pass	
Final	Success/Fail	Fail	444	79	84.9
		Pass	97	133	57.8
	Overall Percentage				76.6

Table 2. Classification Comparisons

Hypothesis Testing

According to the results, total edits has a significant relationship with RFA success (*p*-value < 0.001) indicating that H1 is supported. None of the subject namespaces (Main, User, and Wikipedia) showed significant relationship with RFA success. Therefore, H2 is not supported. The results of the relationship between social contribution and RFA success, however, is mixed. Among Main Talk, User Talk, and Wikipedia Talk namespaces, only User Talk and Wikipedia Talk showed significant relationships with RFA success (*p*-value < 0.01); whereas, the relationship between Main Talk and RFA success is not supported. The relationship between User Talk and RFA success, however, is demonstrated to be in the opposite direction of what we expected it to be. In other words, the results show that those who contribute more to Talk User pages, are less likely to be successful in their RFA. Thus, H3 is partially supported. According to the results, there is also a significant relationship between edit summary usage and RFA success (*p*-value < 0.001). Thus, H4 is supported.

Regarding the control variables, both tenure (number of days since registration) and number of attempts showed significant relationships with RFA success. The positive relationship between tenure and RFA success implies the overall community's trust in senior members. The negative relationship between the number of attempts and RFA success suggests that those who have tried to become administrators several times are less likely to be trusted by the community.

In summary, two of the four hypotheses in our study were supported (H1 and H4); while, one hypothesis was partially supported (H3) and the other hypothesis was not supported (H2).

DISCUSSION AND CONCLUSION

The results of this study are an indication that the Wikipedia community as a whole tends to utilize the readily available information it deems important in making promotion decisions. Those information include total contribution to the community as well as specific form of contribution such as those related to Wikipedia Talk pages. Moreover, the results demonstrated that more senior and those who have tried less number of times to become administrators are more likely to be successful in their RFA process. Additionally, the negative coefficient for the percentage of contributions to the User Talk namespace (B = -2.315) may represent the communities awareness of wasted resources in that User talk contributions are supposed to be used for coordination among users, not for general socialization. It could also be the fact that users who are in trouble more often than not, may have a higher number of edits to User Talk pages as they are required to explain their actions. Lastly, our results imply that edit summary percentage is valued because it shows the community that the user is not lazy (i.e entering edit summaries requires effort), and the user is in support of transparency.

The results of this study has contributions for research and practice. From a theoretical standpoint, our results extended prior studies' results, such as Burke and Kraut's (2008a) findings, on what factors may increase the likelihood of success in RFA processes. From a practical perspective, our findings may be used by managers and providers of Wikipedia and other open source projects. For example, our results indicated that users' contributions to User Talk and Wikipedia Talk have negative and positive impacts on community's decisions on RFAs, respectively. Thus, community providers can make users aware of this and accordingly, encourage them to contribute more to Wikipedia Talk pages and less to User Talk pages to increase the likelihood of their success in their requests for promotion within the community. Moreover, our results showed that clarification of contribution may significantly impact users' success in RFAs. Therefore, users should be informed about this and be encouraged by the community managers to provide sufficient and clear summary of the edits they make to the Wikipedia pages.

Limitations of this study include incorporating the narrow view of readily available information as opposed to a more in depth analysis of edit history. In fact, contributions to each of the namespaces could be further broken down into the type of activity performed for a better analysis of the factors leading to a successful promotion. Furthermore, this study may not be used for inferences of causality by the fact that the measures are not experimental. It is possible and even likely that successful promotion is not due to the number of edits made, but the characteristics of users that may be evident in these measures. Future research may perform a factor analysis to determine what underlying factors are important to the community, and correlate those factors with objective measures.

REFERENCES

1. Beschastnikh, I., Kriplean, T., and McDonald, D. (2008) Wikipedian self-governance in action: Motivating the policy lens, *ICWSM,* 2008.

2. Burke, M., and Kraut, R. (2008a) Mopping up: modeling wikipedia promotion decisions, *Proceedings of the ACM conference on Computer supported cooperative work (CSCW 08),* November 8 – 12, San Diego, CA, USA, ACM Press, 27-36.

3. Burke, M., and Kraut, R. (2008b) Taking up the mop: identifying future Wikipedia administrators, *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 08),* April 5 - 10, Florence, Italy, ACM Press, 3441-3446.

4. Collier, B., Burke, M., Kittur, A., and Kraut, R. (2008) Retrospective Versus Prospective Evidence For Promotion: The Case Of Wikipedia, *Proceedings of the Academy of Management*, April 8 – 13, Anaheim, CA, USA.

5. Hara, N., and Shachaf, P. (2009) Cross-Cultural Analysis of the Wikipedia Community, *Proceedings of the Americas Conference on Information Systems (AMCIS 2009)*, August 6 – 9, San Francisco, CA, USA, 770.

6. Hosmer, D. W. and Lemeshow, S. (2004) Applied logistic regression. John Wiley & Sons.

7. Ingawale, M., Dutta, A., Roy, R., and Seetharaman, P. (2009) The Small Worlds of Wikipedia: Implications for Growth, Quality and Sustainability of Collaborative Knowledge Networks, *Proceedings of the Americas Conference on Information Systems (AMCIS 2009),* August 6 – 9, San Francisco, CA, USA, 439.

8. Jin, L., Robey, D., and Boudreau, M. (2007) Beyond development: A research agenda for investigating open source software user communities, *Information Resources Management Journal,* 20,1, 68-80.

9. Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005) Applied Linear Statistical Models, (Fifth ed.) McGraw-Hill Irwin, New York, NY.

10. Olivera, F., Goodman, P.S., and Swee-Lin Tan, S. (2008) Contribution Behaviors In Distributed Environments, *MIS Quarterly* 32,1, 23-42.

11. Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., and Riedl, J. (2007) Creating, destroying, and restoring value in Wikipedia, *Proceedings of the International ACM Conference on Supporting Group Work (GROUP 07)*, November 4 – 7, Sanibel Island, FL, USA, 259 – 268.

12. Rahman, M. (2009) An Analysis of Wikipedia, *Journal of Information Technology Theory and Application (JITTA)* 9,3, 6.

13. Viégas, F., Wattenberg, M., and McKeon, M. (2007) The Hidden Order of Wikipedia, *Online Communities and Social Computing*, 445-454.

14. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. (2011) Finding social roles in Wikipedia, *Proceedings of the iConference (iConference 2011)*, Februaru 8 – 11, Seattle, WA, 122 – 129.