

---

# Taking up the Mop: Identifying Future Wikipedia Administrators

**Moira Burke**

Human-Computer Interaction Institute  
Carnegie Mellon University  
Pittsburgh, PA 15217 USA  
moira@cmu.edu

**Robert Kraut**

Human-Computer Interaction Institute  
Carnegie Mellon University  
Pittsburgh, PA 15217 USA  
robert.kraut@cmu.edu

**Abstract**

As Wikipedia grows, so do the messy byproducts of collaboration. Backlogs of administrative work are increasing, suggesting the need for more users with privileged admin status. This paper presents a model of editors who have successfully passed the peer review process to become admins. The lightweight model is based on behavioral metadata and comments, and does not require any page text. It demonstrates that the Wikipedia community has shifted in the last two years to prioritizing policymaking and organization experience over simple article-level coordination, and mere edit count does not lead to adminship. The model can be applied as an AdminFinderBot to automatically search all editors' histories and pick out likely future admins, as a self-evaluation tool, or as a dashboard of relevant statistics for voters evaluating admin candidates.

**Keywords**

Wikipedia, administrators, management, collaboration

**ACM Classification Keywords**

H.5.3 [Information Interfaces]: Group and Organization Interfaces - Collaborative computing, Web-based interaction, Computer-supported cooperative work

**Introduction**

Wikipedia, the collaboratively edited online encyclopedia, reached 2.1 million English articles in December 2007. In the midst of exponential growth of both content and users [7], administrators keep things tidy: they delete copyright violations, protect frequently vandalized pages, block malicious users, move pages when there are name conflicts, exclude bulk vandalism from the recent changes list, and edit the front page. Approximately 1400 editors have successfully passed the rigorous peer review of the Request for Adminship (RfA) process and been given

---

Copyright is held by the author/owner(s).

CHI 2008, April 5 – April 10, 2008, Florence, Italy

ACM 978-1-60558-012-8/08/04.

administrator privileges, and they are considered trusted custodians of the successful encyclopedia and its community of contributors.

Kittur and colleagues found that while admins once accounted for nearly 60% of editing activity, their influence declined to approximately 10% due to an influx of new editors [7]. Yet admins are working harder than ever: while their edits per month have steadily increased, backlogs of work requiring admin privileges continue to grow<sup>1</sup>, suggesting a need for additional editors to become admins.

This paper presents a model of the admin promotion process. The model can be used to identify editors likely to be promoted, as a self-evaluation tool for potential admins, and to provide a dashboard of relevant behavior for RfA voters. The model is lightweight, based on edit counts and comment text available in the public Wikipedia database or the user's contribution page, and not requiring any full text from articles or talk pages. Thus, it can be run quickly to search the entire population of editors, or allow editors to calculate their own likelihood of success on the fly without taxing the server. Social science theory, such as Karau and William's collective effort model [6] suggests that people are more likely to contribute to a collective good such as Wikipedia when they know that they are uniquely qualified for the task, or that the likelihood of success is good. Thus, making strengths and weaknesses in their edit histories salient may prompt editors desiring adminship to behave in ways valued by the community of its admins, or may prompt editors who had not considered applying for adminship to step up. Wikipedia currently has a simple version of this idea; there is a hand-maintained page of users with high edit counts who are not administrators<sup>2</sup>. However, sheer number of edits does not make an RfA nominee likely to pass; "editcountitis" among the voters is frowned upon because some users have high counts from minor changes or not previewing work before publishing.

In addition to the practical benefits to the Wikipedia community, this paper also speaks to long-standing concerns of organizational scholars who have asked what

causes employees to get ahead in their jobs—for example, their experiences and skills, social networks, or job-irrelevant attributes like gender or looks (see [9] for review). Despite protestations that admins are lowly janitors mopping up, in many ways election to admin is a promotion, distinguishing an elite core group from the larger mass of editors. The research described uses policy capture [11] to compare nominally important attributes to those that actually lead to promotion. The behavioral data here is rarely available in conventional organizational settings. Although this paper examines only the influence of job-related experiences on promotion, understanding these influences is a base for any additional analyses.

### **The Request for Adminship Process**

To become an administrator, an editor must undergo a week of scrutiny during which the community builds consensus about the candidate's experience and trustworthiness. Admin tools are not granted lightly; an inexperienced, biased, or ill-intentioned editor could cause significant damage, reducing the encyclopedia's credibility or demotivating other editors. Approximately 2700 editors have been nominated for adminship since 2001 with an overall success rate of 53%. However, the process has gradually grown more rigorous, dropping from a 75.5% success rate through 2005 to 42% in 2006 and 2007, and some early admins have expressed doubt that they would pass muster if their RfA were held today [4]. The process once called "no big deal" by the founder of Wikipedia has become a fairly big deal<sup>3</sup>. In the Guide to RfAs<sup>4</sup>, the community describes criteria many RfA contributors look for in nominees, including strong edit history, varied experience, and polite interaction with other users. This paper examines which of these criteria are most predictive of success, and which are only nominally used by the community in choosing its admins.

### **Modeling Successful RfA Candidates**

The data include all 1551 Requests for Adminship from January 2006 through October 2007, with 49 RfAs removed

<sup>1</sup> [http://en.wikipedia.org/wiki/Category:Administrative\\_backlog](http://en.wikipedia.org/wiki/Category:Administrative_backlog)

<sup>2</sup> <http://en.wikipedia.org/wiki/Wikipedia:NA>

<sup>3</sup> <http://en.wikipedia.org/wiki/WP:DEAL>

<sup>4</sup> <http://en.wikipedia.org/wiki/Wikipedia:GRFA>

for being multiple attempts by the same candidate in one month (all of which failed), bots, sockpuppets (multiple identities held by the same person), or because the nominee's edit history prior to the RfA was not available. An additional 1242 RfAs dating back to 2001 are described in the Conclusion section for changes in criteria over time, but the present model focuses on recent criteria in order to identify future admins. For each RfA, data from the user's contribution history page—up to the month before the RfA—was parsed and counted, and grouped according to the criteria described in the Guide to RfAs. Table 1 provides summary statistics for the features. Features applicable to multiple categories were placed in a single category, and two categories from the Guide to RfAs—trustworthiness and high quality of articles—were excluded from this analysis because they could not be captured from simple edit counts. These two categories are discussed further in the Conclusion.

*Strong edit history* includes the number of edits to articles and the number of months between the user's first edit and the RfA. Total edit count (Mean=5010.5, Std. Dev=5818.8) is included in the baseline analysis but is replaced by counts in the individual namespaces (e.g. articles, user talk pages) in the final model due to multicollinearity.

*Varied experience* divides edit counts into individual namespaces, including Wikipedia namespace edits (pages in a subsection focusing mainly on policy, with WikiProjects counted separately), and User pages. The *diversity* metric is a proxy for diverse experience consisting of a count of the number of different areas in which the user has participated, from the set {article, article talk, Wikipedia, Wikipedia talk, user, user talk, articles/categories/templates for deletion (XfD), (un)deletion review, other RfAs, village pump, admin intervention against vandalism (AIV), requests for protection (RfP), admin noticeboard, arbitration committee, mediation committee, and wikiprojects}. So, a user who has edited articles, her own user page, and posted once at the Village Pump would have a diversity score of 3. Actual number of edits in each of these sections is accounted for in the following categories to determine their relative importance.

*User interaction* includes edits to all talk pages, participation on arbitration or mediation committee pages or wikiquette alerts (an early stage in dispute resolution), posting "welcome" on user talk pages, and including common variants of "please" (including "pls" and "plz") or "thanks" (including "thx") in comment text.

*Helping with chores* includes reversion of vandalism (noted by "revert" or "rv" in the comment) requesting admin intervention for specific vandals (AIV), requesting protection for a frequently vandalized page, neutrality fixes (noted by "(n)pov" for "neutral point of view"), requesting admin attention (e.g. for inappropriate usernames), and participating in deletion discussions, including articles/categories/templates for deletion (XfD) and (un)deletion reviews. It also includes the percent of the user's total edits marked as minor, which is the designation for spelling or small formatting changes.

*Observing consensus* includes participation in other editors' RfAs, posting to the Village Pump (a forum for technical and policy discussions), or voting for articles to be deleted or rewritten. Though there are better ways to measure consensus building, they involve natural language processing and thus are prohibitively time-consuming at Wikipedia scale. These measures are lightweight proxies for consensus.

*Edit summaries* or *comments* include the percent of edits with a human-written comment (edits that only have automatically generated comments are not included), and the average length of comments. Comments are both summaries of changes and conversations between authors, often preempting objections or asking questions of each other [12].

## Results

To examine the impact of these behavioral factors on the likelihood of an RfA's success, we first determined baseline performance to be 58.0% for a model that always predicted failure. Adding the nominee's total edit count improves performance to 66.2% accuracy. Then we compared the performance of several models including machine learning approaches using decision trees and support vector machines, and statistical models of regression, applied to all of the features mentioned above. All models performed

approximately the same (between 73% and 75% accurate), and identified similar behaviors as important so the results of the regression are reported here for ease of presentation.

We performed a probit regression on the binary dependent variable, RfA success. The coefficients in Table 1 represent the change in probability of success when all continuous variables are at their mean values. Results prior to 2006 are reported for comparison. Multiple attempts by the same candidate in a single month were excluded, leaving only one attempt per month, and the candidate's number of previous RfA attempts (in other months) is included as a control variable, showing that each subsequent attempt has a 14.8% lower chance of success than the previous one. All edit count variables with large means have been divided by 1000 in the regression and are noted in the table, but true means are reported. The model is 74.8% accurate in classifying RfA attempts as successful or not.

*Strong edit history and varied experience.* The results strongly support the notion that mere edit count ("editcountitis") is a poor predictor of Wikipedia adminship. Every 1000 article edits only increases the probability of success by 1.8%. Instead, variety of experience, particularly in the Wikipedia (policy) namespace, is a much better predictor. One Wikipedia policy edit or WikiProject edit is worth ten article edits; one thousand of those edits increases the probability of success by 19.6% and 17.1%, respectively. The diversity score also shows that making a single edit in any additional region of Wikipedia is correlated with a 2.8% increased likelihood of success. Length of experience helps slightly; each additional month between the user's first edit and her RfA helps by 0.4%. Not surprisingly, edits to user pages do not add information to the model, as the norm is to not edit other users' pages except rarely to add rewards known as barnstars. The majority of user page edits would be to these nominees' own user pages, which do not add much value to Wikipedia as a whole.

*User interaction.* Some forms of user interaction also help; every thousand article talk edits increases the likelihood of success by 6.3%. Article talk pages are mechanisms for coordination and dispute resolution [13], so it is not

surprising that future admins participate more heavily there than do unsuccessful nominees. Somewhat surprisingly, user talk page edits do not affect likelihood of adminship. This is perhaps because the norm is to hold disagreements over content on article talk pages, moving to user talk pages when the disagreement covers multiple articles or a user's overall behavior. Thus, user talk edits are likely to be mixed, and may have higher interpersonal conflict. This is supported by the finding that each edit to an Arbitration or Mediation Committee page, or to a Wikiquette notice, all of which are venues for dispute resolution, decreases the likelihood of success by 0.1%. Politeness helps modestly; though it was rare, every instance of saying "thanks" in a comment increases the likelihood of success by 0.3%. Saying "please" and welcoming newcomers had no effect.

*Helping with chores* does not have a large impact. Reverting vandalism is only marginally significant, and so it likely depends on the context of vandalism—such as whether the suspected vandal was an anonymous editor or a well-respected editor with a differing point of view. Pointing out neutrality (pov) issues helped adminship by 0.1%, but escalating problems to the administrators' noticeboard hurt by 0.1%. These results suggest that future administrators stay out of trouble, keeping disputes at the article talk level, rather than bringing them to committees or admin attention. Finally, making minor edits helps, although this may be because inexperienced editors making small changes inflate their importance or do not notice the checkbox for designating the changes as minor, and so it's simply that more responsible editors are promoted.

*Observing consensus.* Somewhat surprisingly, participating in consensus-building activities, such as other RfAs or the Village Pump, does not increase the likelihood of becoming an admin. However, these are simplistic measures of consensus-building and do not take full text from talk pages into account, which might reveal more successful language patterns of consensus.

*Edit summaries/comments.* Finally, leaving comments helps. While the length of the comment does not matter, diligently summarizing edits or having brief negotiations with other editors in the comments helps, consistent with

the statement in the Guide to RfAs that some voters expect edit summaries to approach 100%.

**Table 1.** Descriptive statistics and probit regressions on the likelihood of success in a Request for Adminship in 2006-7 and pre-2006.

|                                   | 2006-2007 |        |            | Pre-2006 |       |          |
|-----------------------------------|-----------|--------|------------|----------|-------|----------|
|                                   | Mean      | StDev  | Δ Prob.    | Mean     | StDev | Δ Prob.  |
| Attempt number                    | 1.2       | 0.6    | -14.8% *** | 1.1      | 0.3   | -11.1%*  |
| <b>Strong edit history</b>        |           |        |            |          |       |          |
| Article edits ‡                   | 2611.1    | 3804.3 | 1.8% ***   | 2082.6   | 2.8   | 1.1%     |
| Months since first edit           | 9.6       | 8.0    | 0.4% *     | 9.4      | 7.1   | 0.2%     |
| <b>Varied experience</b>          |           |        |            |          |       |          |
| Wikipedia (policy) edits ‡        | 433.8     | 913.8  | 19.6% ***  | 265.9    | 0.5   | 0.4%     |
| WikiProject edits ‡               | 144.0     | 569.1  | 17.1% **   | 53.0     | 0.2   | 7.2%     |
| Diversity score                   | 11.0      | 3.7    | 2.8% ***   | 9.9      | 2.6   | 3.7%***  |
| User page edits ‡                 | 219.0     | 296.7  | -8.6%      | 101.8    | 0.2   | 11.5%    |
| <b>User interaction</b>           |           |        |            |          |       |          |
| Article talk edits ‡              | 415.2     | 775.4  | 6.3% *     | 211.6    | 0.3   | 15.4%*   |
| User talk edits ‡                 | 786.6     | 1169.9 | -1.8%      | 187.3    | 0.3   | -4.8%    |
| Wikipedia talk edits              | 87.6      | 179.5  | 0%         | 34.7     | 64.2  | 0%       |
| Arb/mediation/wikiquote edits     | 9.8       | 47.1   | -0.1% ***  | 3.7      | 19.5  | -0.2%**  |
| Newcomer welcomes                 | 76.9      | 321.1  | 0%         | 20.2     | 90.9  | 0%       |
| "Please" in comments              | 31.7      | 83.8   | 0%         | 20.9     | 112.5 | 0%       |
| "Thanks" in comments              | 21.8      | 39.3   | 0.3% ***   | 7.2      | 13.2  | 0%       |
| <b>Helping with chores</b>        |           |        |            |          |       |          |
| "Revert" in comments ‡            | 257.6     | 563.2  | 7.0% †     | 98.5     | 0.2   | -6.8%    |
| Vandal-fighting (AIV) edits       | 26.5      | 108.7  | 0%         | 0.7      | 6.5   | 0.1%     |
| Requests for protection           | 3.7       | 12.2   | -0.2%      | 0.7      | 2.8   | -2.4%*** |
| "POV" in comments                 | 26.7      | 46.9   | 0.1% **    | 18.0     | 28.8  | 0.0%     |
| Admin attention/noticeboard edits | 18.7      | 57.9   | -0.1% *    | 2.9      | 14.4  | 0.2%     |
| X for deletion/review edits ‡     | 504.5     | 1027.3 | -2.9%      | 280.7    | 0.7   | 2.5%     |
| Minor edits (%)                   | 25.5      | 23.0   | 0.2% ***   | 35.3     | 24.3  | 0.2%***  |
| <b>Observing consensus</b>        |           |        |            |          |       |          |
| Other RfAs                        | 27.4      | 58.9   | 0%         | 11.1     | 30.4  | 0.0%     |
| Village pump                      | 9.5       | 34.6   | 0%         | 12.8     | 56.1  | 0.0%     |
| Votes                             | 3.4       | 15.1   | 0.1%       | 1.5      | 4.0   | -0.5%    |
| <b>Edit summaries / comments</b>  |           |        |            |          |       |          |
| Commented (%)                     | 63.2      | 24.0   | 0.5% ***   | 75.1     | 24.8  | 0.4%***  |
| Avg. comment length (log2 chars)  | 5.0       | 0.8    | 1.6%       | 4.8      | 0.7   | 3.1%     |

N=1502 \*\*\*p<.001 \*\*p<.01 \*p<.05 †p<0.1 ‡Edit count divided by 1000 for ΔProb

### Applications

This model can be applied in three ways: to automatically search all editors' histories and pick out likely future

admins, as a self-evaluation tool, or as a dashboard of relevant statistics for RfA voters.

Following the lessons learned by Cosley and colleagues' SuggestBot [1], which matched pages needing work with editors who had similar interests, a kind of "AdminFinderBot"—a user account for a computer program that runs the model automatically—would need to follow Wikipedia norms and work with the bot approval committee to be most effective and accepted. As a very lightweight process, it already meets one of the main criteria for Wikipedia bots: it would not be a server hog. As the discussion on the non-admin high edit count page shows, some highly contributing editors do not want to become admins, so their privacy will need to be considered in the implementation.

As a self-evaluation tool or voter dashboard, this model would be useful, allowing editors or voters to size up an editor compared to previous RfA nominees, indicating areas where the editor needs improvement, or highlighting the editor's varied experience. Decision-making research shows that people are poor at making these kinds of decisions, and that statistical models often perform better [3][5], so at least providing statistics should improve the RfA process.

### Conclusion

Merely performing a lot of production work is insufficient for "promotion" in Wikipedia. Candidates' article edits were weak predictors of success. They also have to demonstrate more managerial behavior. Diverse experience and contributions to the development of policies and WikiProjects were stronger predictors of RfA success. This is consistent with findings that Wikipedia is a bureaucracy [1] and that coordination work has increased substantially [8][13]. However, future work is needed to examine more closely what the admins are doing. Future admins also use article talk pages and comments for coordination and negotiation more often than unsuccessful nominees, and tend to escalate disputes less often.

Although this research has shown that judges pay attention to candidates' job-relevant behavior and especially behavior that suggests the candidate will be a good manager and not just a good worker, it is silent about whether other factors

identified in the organizational literature [9]—social networks, irrelevant attributes, or strategic self-presentation. Indeed, recent evidence that Wikipedia admins use a secret mailing list to coordinate their actions toward others<sup>5</sup> suggest that sponsorship may also play a role in promotion. Future research in Wikipedia using techniques like those in the current paper can be used to test theories in organizational behavior about criteria for promotion.

An important limitation of the current model is that it does not take the quality of contribution into account. We plan to improve the model by examining measures of length, persistence, and pageviews of edits, which are already being used in more processor intensive models of existing admin behavior [7] and impact of edits [10].

Criteria for admins have changed modestly over time. Success rates were much higher (75.5%) prior to 2006, and collaboration via article talk pages helped more in the past (+15% for every 1000 article talk edits, compared to +6.3% today). The diversity score performs similarly prior to 2006 (+3.7% then, +2.8% now). However, participation in Wikipedia policy and WikiProjects was not predictive of adminship prior to 2006, suggesting the community as a whole is beginning to prioritize policymaking and organization experience over simple article-level coordination.

### Acknowledgments

The authors thank Niki Kittur, Tom Murphy, and Ben Collier for feedback. This work was supported by NSF IIS-0325049 and an NSF Graduate Research Fellowship.

### References

- [1] Butler, B., Joyce, E., and Pike, J. Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. *Proc. CHI 2008*, ACM Press (2008).
- [2] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: Using intelligent task routing to help people find work in Wikipedia. *Proc. IUI 2007*, ACM Press (2007), 32-41.
- [3] Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582.
- [4] Forte, A., and Bruckman, A. Scaling consensus: Increasing decentralization in Wikipedia governance. *Proc. HICSS 2008*.
- [5] Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage Publications.
- [6] Karau, S., and Williams, K. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65 (1993), 681-706.
- [7] Kittur, A., Chi, E., Pendleton, B., Suh, B., and Mytkowicz, T. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Proc CHI 2007*, ACM Press (2007).
- [8] Kittur, A., Suh, B., Pendleton, B. A., Chi., E. (2007). He says, she says: Conflict and coordination in Wikipedia. *Proc CHI 2007*, ACM Press (2007), 453-462.
- [9] Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58(2), 367-409.
- [10] Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L., and Riedl, J. 2007. Creating, destroying, and restoring value in Wikipedia. *Proc GROUP 2007*, ACM Press (2007), 259-268.
- [11] Stumpf, S. A., & London, M. (1981). Capturing rater policies in evaluating candidates for promotion. *The Academy of Management Journal*, 24(4), 752-766.
- [12] Viegas, F., Wattenberg, M., and Dave., K. Studying cooperation and conflict between authors with history flow visualizations. *Proc CHI 2004*, ACM Press (2004), 575-582.
- [13] Viegas, F., Wattenberg, M., Kriss, J., and van Ham, F. Talk before your type: Coordination in Wikipedia. *Proc HICSS 2007*, 575-582.

<sup>5</sup> [http://www.theregister.co.uk/2007/12/04/wikipedia\\_secret\\_mailing](http://www.theregister.co.uk/2007/12/04/wikipedia_secret_mailing)