Soc. Netw. Anal. Min. (2013) 3:787–801
DOI 10.1007/s13278-012-0092-6

ORIGINAL ARTICLE

Modeling Wikipedia admin elections using multidimensional behavioral social networks

Michal Jankowski-Lorek · Lukasz Ostrowski · Piotr Turek · Adam Wierzbicki

Received: 21 May 2012 / Revised: 7 December 2012 / Accepted: 12 December 2012 / Published online: 22 January 2013

Abstract Wikipedia admins are editors entrusted with special privileges and duties, responsible for the community management of Wikipedia. They are elected using a special procedure defined by the Wikipedia community, called Request for Adminship (RfA). Because of the growing amount of management work (quality control, coordination, maintenance) on the Wikipedia, the importance of admins is growing. At the same time, there exists evidence that the admin community is growing more slowly than expected. We present an analysis of the RfA procedure in the Polish-language Wikipedia, since the procedure's introduction in 2005. With the goal of discovering good candidates for new admins that could be accepted by the community, we model the admin elections using multidimensional behavioral social networks derived from the Wikipedia edit history. We find that we can classify the votes in the RfA procedures using this model with an accuracy level that should be sufficient to recommend candidates. We also propose and verify interpretations of the dimensions of the social network. We find that one of the dimensions, based on discussion on Wikipedia talk pages, can be validly interpreted as acquaintance among editors, and discuss the relevance of this dimension to the admin elections.

Keywords Wikipedia · Collaboration · Trust

M. Jankowski-Lorek · L. Ostrowski · P. Turek · A. Wierzbicki (✉)
Polish-Japanese Institute of Information Technology, Warsaw, Poland
e-mail: adamw@pjwstk.edu.pl

P. Turek
e-mail: turek@pjwstk.edu.pl

1 Introduction

Wikipedia is one of the most popular websites on the Internet. It is a collaborative effort to organize and present human knowledge, similarly to traditional encyclopedias. Its most distinctive feature is the fact that anyone may edit the content. Thanks to the Wiki technology, anyone may become the editor. This fact causes the sustained growth of Wikipedia (Spinellis and Louridas 2008), but also possible scalability problems in the future.

Nowadays in the Web 2.0 era, there are a lot of sites where user contributed content plays a major role. Many other public Wiki sites may face similar problems as Wikipedia. Due to Wikipedia's openess and lack of centralized supervision, authors need to overcome problems, that are not found in editing of traditional encyclopedias.

The most notable example is vandalism, which is mostly the deliberate deletion of content or putting false or irrelevant information. The effect of vandalizing Wikipedia may have serious consequences for real people, especially when a biographical article becomes vandalized. While the global impact of this kind of damage is rather low, it is rising (Priedhorsky et al. 2007). Even though the anti-vandalism bots created to automatically prevent the damage do a good job, there is always the need of human reviewers.

Another problem connected with lack of central supervision arises when editors have different points of view which may result in an edit war, when two or more contributors or groups try to enforce their version of the article. This violates one of the key Wikipedia rules, which mandates the contributors to keep a neutral point of view. Viégas et al. (2004) noted that edit wars are a threat not only for controversial articles.

The mentioned problems are caused mostly by human factors and at least some of their instances cannot be

resolved without another human intervention. This is the role of administrators to constantly monitor Wikipedia and make sure that rules established by the community are obeyed.

1.1 Problem statement

The growing amount of work for administrators caused by increased popularity and amount of content in Wikipedia (Kittur et al. 2007) causes a potential risk that administrators may become overwhelmed and their response time may become longer. There are also concerns that the number of newly elected Wikipedia administrators is decreasing, and that as a result of these two trends, the Wikipedia project itself will not be sustainable.

To avoid the possible degradation of Wikipedia quality, especially because there may be too little administrative workforce to accommodate Wikipedia growth, we have identified the need for new tools to evaluate potential new candidates for administrators. To get started we have taken a look at the current situation among administrators and performed quantitative studies on past Requests for Adminship (RfAs) (votes on new candidates for admins). This preliminary research has been summarized and published in (Turek et al. 2011). After examining the current situation, we have identified the main problem: the number of newly elected admins is indeed sharply decreasing. To investigate the possible causes of this phenomenon, we have formulated two hypotheses.

Hypothesis A states that new admins are elected on the basis of acquaintance. This hypothesis expresses the concern that the community of admins is forming a clique of acquaintances and it is more difficult to become part of this society as it grows. *Hypothesis B* states that new admins are elected on the basis of similarity of experience in editing of articles on various topis. According to this hypothesis, editors make voting decisions about candidates by comparing the candidate's experience to their own. A vote for a candidate will be cast if this candidate has a similar experience of editing articles in various topics as the voter. Over time the disproportion of experience among users of Wikipedia grows as new people are becoming active members while the "core" team of admins stays almost the same. The two hypotheses are not mutually exclusive, as editors could be using both criteria—acquaintance and experience similarity—in their nominations and voting decisions.

To verify the two hypotheses, we have constructed social networks from two sources: RfA votings and Wikipedia edit history, and then analyzed how the relationships from the edit history relate to the cast votes. The social network constructed from the edit history is a Multidimensional Behavioral Social Network (MBSN), based on our past research (Turek et al. 2010; Turek et al. 2011), that can be used as a general model of the Wikipedia knowledge community. The analysis presented in this paper focuses on using the MBSN to model RfA votings, but is also a demonstration of the relevance of the MBSN model to the Wikipedia knowledge community.

Based on the MBSN, we have created a data mining model to classify votes for and against admin candidates. In this way, we have tried to find which criteria are relevant for voters when making the decision about a candidate. The results of this analysis show that using our behavioral model, it is possible to recommend good candidates for admins. This recommendation could serve to increase the number of RfA votings and possibly also to increase the number of new admins. The data mining analysis also positively verifies hypothesis B.

The next step has been an attempt to validate the interpretation of the behavioral social networks. For each of the dimensions, we have formulated a hypothesis regarding the dimension's interpretation as a real social relationship. We have verified these hypotheses using a survey of over 100 active Polish Wikipedia editors. We present the results of our validation, focusing particularly on one dimension that has passed the validation successfully: the network based on Wikipedia talk pages that can be validly interpreted as an acquaintance relation among Wikipedia editors. We then refocus on the negative verification of hypothesis A.

This article has therefore four main contributions:

- An analysis of the MBSN of Wikipedia editors as a model of the Request for Adminship votes that shows how the multidimensional network can be used to recommend candidates for new admins
- A validation of a set of hypotheses concerning interpretation of MBSN dimensions as real social concepts, and a definition of a behavioral social network based on Wikipedia talk pages that can be validly interpreted as an acquaintance relation
- A negative verification of the hypothesis that new admins are elected on the basis of acquaintance
- A positive verification of the hypothesis that votes for admin candidates depend on the similarity of editing experience in various topics of the voter and the candidate

The rest of the paper is organized as follows: Next, we review past Wikipedia research, especially concerning adminship. In Sect. 3, we present a quantitative study on the current administrators and their RfA procedures. The study shows that the number of successful admin elections is declining; we formulate hypotheses A and B that can explain the reasons for this phenomenon. Section 4 focuses on the analysis of the votes using the multidimensional

behavioral social network based on edit history, using first a simple comparison of distributions, and then a data mining approach. This section describes a positive validation of hypothesis B. Section 5 presents the validation of the behavioral network of Wikipedia editors and discusses the negative validation of hypothesis A. Finally, in Sect. 6 we summarize the results and draw conclusions.

2 Related work

Wikipedia has been a subject of several studies in the past few years. Most notable example of research topic is assessing content quality (Priedhorsky et al. 2007; Adler et al. 2008; Vuong et al. 2008; Zhang et al. 2010). The trustworthiness of Wikipedia is one of the key concerns related to its usefulness and generally, success.

The problem of recommending and evaluating candidates for administrators has not been extensively studied, but this topic is slowly growing in popularity. The most similar work that we have found is (Burke and Kraut 2008). The authors present an idea of recommending and evaluating candidates for administrators based on behavioral data and comments, not the page text. They counted each candidate's edits in various namespaces (article, article talk, Wikipedia, Wikipedia talk, Wiki projects etc.) to calculate total contribution as well as contribution diversity. They also measured user interaction, mainly activity on talk pages, but also participation on arbitration or meditation committee pages and a few others. There are also several other statistics, but the ones mentioned seemed to be the most relevant to the candidate's success. Especially successful were candidates with strong edit diversity, mere edits in Wikipedia articles did not add much more chance of success. In user interactions, article talk page edits were the best predictor of success, with other authors talk page edits being rather poor. Authors also confirmed Kittur et al.'s (2007) results that the percentage of indirect work (coordination, discussion, etc.) grows over time, the share of articles in all Wikipedia edits is decreasing.

The problem of evaluating voters and candidates has been also studied in the social context in (Leskovec et al. 2010). The authors found out that the probability of one person's vote to be positive is correlated with the basic *relative* figures such as: who—voter or candidate has more edits, who has more barnstars (awards given by other Wikipedia users), the extent of collaboration of the two, etc. Authors strongly noted that the vote value (positive or negative) is not just a function of candidate, but both voter and candidate. They also studied the relationship between past votes (which are public) and next votes given by other voters. The "response function" (function estimating vote value based on voter and previously cast votes) varied from one user to another. This suggests that each voter has a certain policy of looking or not looking at previous votes.

Multidimensional social networks have been studied in the work of Kazienko et al. (2011), Kennedy (2009) and Rodriguez and Shinavier (2009).

3 Polish Wikipedia adminship

As Wikipedia itself defines, an Administrator (sysop) is a committed and trustworthy participant of a project, who has received additional powers by a decision of the community. These powers do not suggest editorial control over the project. Administrators also provide help in editing Wikipedia, especially to newcomers. The basic administrative permissions are as follows:

- deleting pages and un-deleting them, so administrators have the access to content previously regarded as irrelevant or inappropriate for an encyclopedia,
- flagging and unflagging a page as editable only by administrators (mostly not encyclopedic pages, such as the main page) or only by registered users,
- blocking (and unblocking) users ability to edit pages, mostly used to disallow malicious individuals from damaging Wikipedia. Either user account or IP address (or a group of those) may be blocked.

As of November 1, 2010, Polish Wikipedia had 168 administrators. Since 2005 there have been held 281 votings for Requests for Adminship (hereafter—RfA). 171 were completed with granting an administrator's privileges, 110 were rejected the candidates, 39 were withdrawn before the end of the voting, and 34 were canceled (due to statutory requirements or no acceptance of the nomination by a candidate). Approximately 38 administrators were selected before introduction of the RfA procedure in March 2005.

Data on the RfA does not add up, inter alia, for the following reasons:

- "Verification" votings have been counted as ordinary (sometimes administrators want to confirm that they still have the support of the community and decide to verify their trustworthiness by standing for a re-voting).
- Some of the administrators gave up their powers. This happened both at the moments they stopped editing Wikipedia, and in the situations when they decided that after a break in editing they were not going to take it up again.
- some administrators resigned, and then applied for the adminship again, as has happened in the case of former administrators who returned to editing after previous conflicts within the community.

- A few administrators' permissions have been taken away by the Arbitration Committee.
- The first RfA procedure was performed on 3rd March 2005. Previously, administrators were elected on a mailing list.
- Some charts use only data from 86 cases, due to the lack of complete data in the logs of Wikipedia. This applies particularly to the initial contribution of administrators.

In the beginnings, when Wikipedia had only several active editors, the adminship was granted solely basing on technical needs, without social issues in mind. Soon after that, the mailing list was a place, where the emerging community discussed social aspects and particularly nominated candidates for administrators. The procedure implemented on a mailing list worked on a principle, that if nobody argued, whether a certain person should get the administrative permissions, they were granted. During the 4 years (up to 2005) of nominating candidates on a mailing list, 40 persons got the permissions, while only one candidate was rejected.

This way of granting adminship was questionable and did not leave a trace in the Wikipedia itself, who and when got the permissions. At the beginning of 2005 the new voting-based procedure was introduced. It caused a lot of problems, for example because of "free riders", who had very little and often disputable contribution to the project and applied for the position of administrator. There was also a problem on the other side—people who voted often had very little experience in editing Wikipedia. Additionally it was easy to rig the voting by using sock puppets (additional accounts owned by the same person). To remedy this situation the procedure was formalized and its final version were created almost a year later (December 2005).

The current version of the procedure mandates that a person standing for the voting must have the account for at least 3 months and with at least 1,000 edits. To be able to vote, user must have the account for at least 2 weeks and 500 edits in articles. The voting starts at the moment, when a candidate confirms that he or she is willing to become an administrator, as candidates may apply by themselves or be nominated by others. To get the administrative permissions, candidate must have at least 20 "for" votes and they must be at least 80 % of total "for" and "against" votes. After being rejected (due to not having enough support votes or not meeting the formal requirements) or resigning, candidate may re-apply in 60 days after voting ends.

3.1 Basic RfA statistics

All the statistics described in this chapter are based on the full population of Request for Adminship votes in the Polish-language Wikipedia, since the introduction of the RfA procedure.

3.1.1 During which voting a candidate was accepted

The first analysis we dealt with, was an attempt to determine during which voting a candidate is accepted. There is a noticeable and significant difference between the numbers of candidates who were admitted in the first and subsequent attempts. It is also evident that less than half the candidates who were rejected in the first approach were trying to get the administrator's rights for a second time. In total, 228 candidates have applied for the adminship at least once; 83 were rejected. For a second time the adminship was requested by 39 candidates; over a half of them, 21, were rejected; 14 candidates applied for the voting for a third time and 4 of them were accepted. Both in the case of a fourth vote (6 candidates) and a fifth vote (1 candidate), no one was accepted. Nobody applied for a sixth time.

3.1.2 Frequency of votings

Next, we proceeded to analyze the number of votings a year and the percentage of applications accepted yearly.

On the chart with the number of votings (Fig. 1) a peak can be observed in 2006 when the figure reaches the value 95, while a year before it was 34, and a year later it decreased to 60. Apart from the period 2006–2007 the number of votings has never exceeded 38. Only in 2010 the level was lower than 34. The number of RfAs between 2006 and 2010 got lowered over three times (from 95 to 26). However, this may be due to an incomplete testing period (data for the study were collected on November 1, 2010).

The percentage of accepted applications (see Fig. 2) can be divided into two periods, first, 2005–2008, when the percentage of accepted candidates ranged between 57 and 70 %. The second period, 2009–2010, are values below 50 % (respectively 47 and 42 %). Between 2008 and 2010 the percentage of successful RfAs fell almost by half (from 70 to 42%).

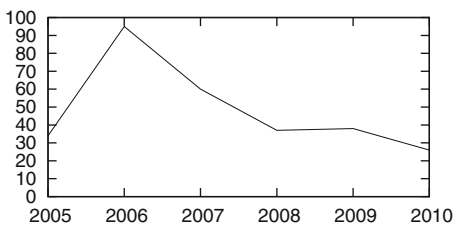

Fig. 1 Number of votings per year

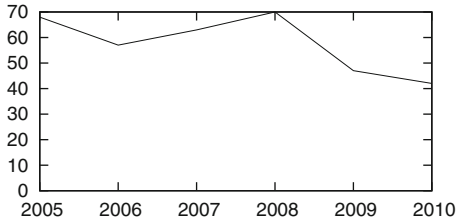

Fig. 2 Percent of accepted Requests for Adminship

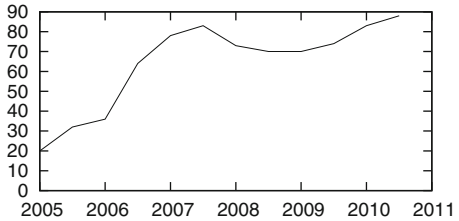

Fig. 3 Mean number of votes in single RfA per year

3.1.3 Number of votes in single voting

Another issue is the number of votes gave during a single voting. We decided to use the arithmetic mean value, not the median. The biggest difference between the arithmetic mean and the median did not exceed 4.71 of votes and that happened only when the analyzed values reached 80. As it can be seen on Fig. 3, the number of votes in the RfAs increased from a minimum of 20 votes in the first half of 2005 to a maximum of 88 votes in the second half of 2010. The chart shows two trends: one runs from the first half of 2005 to the first half of 2007, when the number of votes increased from 20 to 78; the second trend lasted from the first half of 2007 to the second half of 2010 when the number of votes remained at a similar level, ranging from 70 in the second half of 2008 and first half of 2009 to 88 in the second half of 2010.

3.1.4 Numbers of votes

Next, we present the statistical data on the minimum and maximum number of votes "for", "against" and "abstain." The lowest number of votes, when a candidate received adminship was 12/22/25 (for/against/abstain) in various polls. The highest number of votes when a candidate has not been granted the powers was 85/64/28 (for/against/abstain) in various polls. The highest number of votes was cast during a voting on the nomination for WarX—125 votes. In total, there were 14 votings in which the number of votes exceeded 100 (in 10 cases, the candidate was accepted, in 4 rejected).

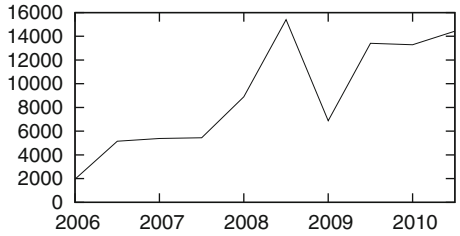

Fig. 4 Mean number of edits of successful candidates

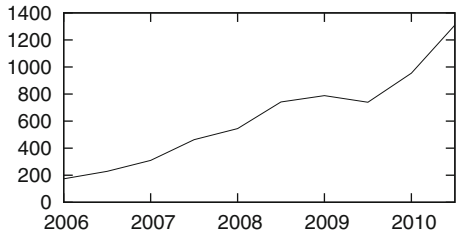

Fig. 5 Mean number of days since account registration of successful candidates per year

3.2 Candidates' experience on Wikipedia

Another study concerning candidates' experience prior to receiving administrator powers was conducted for the last 86 users who were elected as administrators. In the case of previously selected administrators, collecting complete data was not possible due to gaps in the logs of Polish Wikipedia (Figs. 4, 5).

3.2.1 Number of edits

One of the factors, that cause the most discussions during the votings is the number of edits made by a candidate. The RfA Rules contain a sentence that reads: *Candidates for administrators […] may be users who have at least 1,000 undeleted edits*.[1]. However, this value is often considered too low by the voters. On the basis of an analysis of the number of edits at the time of granting the privileges it can be observed that the minimum falls in the first half of 2006 and amounted on average to 2,037 edits. Then the values grow, achieving just over 14,000 edits in 2010. This shows that in the subsequent years the acceptance of candidates required growing experience and the difference between the level required by the Rules and the level a candidate is commonly accepted was constantly increasing. A similar phenomenon is observed on the German Wikipedia, where, according to the declaration of voters the candidates were accepted when they had over 10,000 edits in the second half of 2010.

[1] http://pl.wikipedia.org/wiki/Wikipedia:Przyznawanie_uprawnie%C5%84#Regulamin_przyznawania_uprawnie.C5.84

3.2.2 Time of Wikipedia practice

Another factor that triggers emotions during the votings is the time of practice. It is required by the rules of voting: *Candidates for administrators […] may be users who have at least 1,000 undeleted edits, the first of which took place at least 3 months before requesting the adminship*. We analyzed the time (in days) of the cadidates' practice between the date of registration and the date of being granted adminship, which is not exactly the same value as required in the regulations. The examined time of practice in the first half of 2006 was 182 days. The values gradually grew from 511 days in the second half of 2007, 870 days in the first half of 2009, with a small decline in the second half of 2009 (682 days). In the second half of 2010, it reached the value of 1,310 days, but this may be a slightly undependable result due to only two votings in this period. An overall analysis of the chart shows that in 2006, the candidates had less than one year practice, and since mid-2008 it is at least 2 years. The last two candidates with experience of less than 1 year were elected in February 2009 and November 2008.

3.2.3 Date of registration of recent successful candidates

The final factor we analyzed was the date of registration of the last 86 administrators (see Fig. 6). The analysis found that, as of November 2011, there were no administrators who created their accounts in 2009 and 2010. The latest was Magalia's account, created at the end of August 2008. Almost half of the administrators created their accounts in 2006 (41 of 86). The rest, according to the number of accounts, in 2005 (20), 2007 (16) and 2008 (9).

3.3 Causes of decreasing numbers of elected admins

Together, Figs. 1 and 2 demonstrate a decreasing trend in the overall number of elected admins from 2006 till 2010. This decrease gives rise to serious concern, as the amount of administrative work on the Wikipedia is increasing. Several possible explanations can be made for this phenomenon. The first explanation is that decreasing amounts of candidates accept nominations as admins (this would explain the decrease in the total number of RfA votings). The validation of this hypothesis is beyond the scope of this paper; related work has demonstrated that in recent years Wikipedia has noted some decline in users' contributions, showing a general decrease of motivation (Suh et al. 2009).

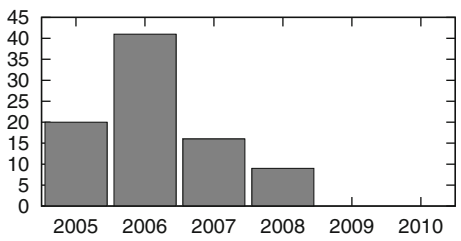

Fig. 6 Year of registration of last 86 administrators

The second explanation is that the number of successful elections decreases because of the changing criteria of candidate selection and acceptance. There can be many possible changes in the criteria, but our study suggests that the criteria are related to the experience of the candidate. This experience can be grossly estimated by the number of edited articles, but a more fine-grained measure (supported by previous work Burke and Kraut 2008) is the number of articles edited on specific topics.

A more detrimental possibility is that the community admins are elected on the basis of acquaintance of the current admins and the candidates for new admins. The next section describes an attempt to validate two hypotheses, formulated in the introduction:

1. Hypothesis A: new admins are elected on the basis of acquaintance
2. Hypothesis B: new admins are elected on the basis of similarity of experience in editing of articles on various topics

4 Analysis of RfA votes using the MBSN

The study of multidimensional behavioral Wikipedia social networks (Turek et al. 2010, 2011) is our ongoing research in effort to model the community of Wikipedia contributors with emphasis on the aspect of teamwork. The research tool that we have used is the social network analysis performed on the behavioral social network mined from the Wikipedia edit history. To create this dataset, we have analysed the entire edit history since the inception of Polish Wikipedia in 2001. The goal has been to find the real authors of content, not only those who copy or move the information around and to find the real social relationships between authors such as trust, criticism, acquaintance and common interests. This was accomplished by using various algorithms similar to those used in plagiarism detection.

A major obstacle was the amount of data present in the edit history and the complexity of operations on this data. In case of the Polish Wikipedia the edit history is over 220 GB of text. Firstly, we needed a way to concisely represent the article text with authorship information. As a basic unit of content we considered a single word. We processed each revision of a particular article in order of the changes that were made and for each word we have assigned its author. So the first revision consisted of words written by the creator of the page and subsequent ones

contained the text at a particular time with their respective authors.

Between each two subsequent revisions we may have four kinds of actions: adding a word, deleting a word, moving a word from one place to another and changing a word. Adding is simply putting a new word in the text (whose author is the author of the revision, where it firstly appeared). Deleting is simply removing a word from the text. Moving is removing a certain portion of text in one place and putting exactly the same sequence in the other. Changing is an operation of replacing one word by the other (including for instance spelling corrections). We needed to separate moving from deleting followed by adding to preserve authorship information. There is a threshold to avoid regarding moving single words or common phrases as moving the text written by previous author. It works by identifying how many consecutive words were moved, if it was below the threshold, then the whole operation is considered a deletion followed by addition by the new author. The replacements of single words are considered also a deletion followed by addition.

4.1 Multidimensional behavioral social network of Wikipedia editors

The MBSN is a set of graphs consisting of nodes (Kazienko et al. 2011), each representing one Wikipedia contributor (some graphs may also contain other edges, such as Wikipedia categories) and edges, each representing one kind of relationship between them. Each edge has its specific weight represented by a numeric value. We have defined four dimensions (networks) of relationships between authors: co-edits, reverts, discussion and knowledge (interests). This network is completely behavioral, meaning that it does not contain any declared information about social relationships, but is completely based on edit history.

4.1.1 Co-edits

The main operation that influences edges strengths in the Co-edits network between contributors is adding text in the vicinity of text written by other author. We believe that when someone edits article text he or she has read the surrounding paragraphs (reviewed them). For this reason, we hypothesize that the Co-edits network may be interpreted as the social relation of trust (Zhang et al. 2010).

Coedits are defined as the amount of text (number of words) written by one author next to the text of other author. The exact measure is calculated as follows:

For each pair (w_1, w_2) of words in each revision, where w_1 is added in the current revision by author A_1 and w_2 has been previously written by A_2, we define D as the distance in words between them. We have:

$Coedits(A_1 -> A_2) = \sum (1/D)$ for each D, where $D < distance\_cutoff$

distance_cutoff is a user-defined parameter, typical values range from 10 to 100.

4.1.2 Reverts

Edge strength in the Reverts network is measured by the number of edits made by one author and reverted by another. This measure allows easy spotting of edit wars, where two or more authors or groups argue with each other. Revert operations have been frequently used in the literature to model conflict. We hypothesize that the Reverts network may be interpreted as the social relation of distrust or criticism (Vuong et al. 2008).

The strength of an edge in the Reverts network between authors A_1 and A_2 is counted as follows. For each revision R in the edit history we look if there was an identical revision R' before in the last *max_recent* revisions. For each such a pair (R, R') we have:

$Reverts(A_1 -> A_2) = count(author(R) = A_1$ and $author(R_i) = A_2)$ for each revision R_i between R and R' *max_recent* is a parameter describing how far we look back in the edit history trying to match the revert.

4.1.3 Discussion

To calculate edge strength in the Discussion network we looked at the articles' and users' talk pages. The measure is proportional to the amount of text added by one author next (that is in response) to the text written by the other author. Activity on talk pages has been used in the literature to evaluate the amount of collaboration between editors. We hypothesize that the Discussion network may be interpreted as the social relation of acquaintance.

The strength of a discussion edge between authors A_1 and A_2 is given by: $Discussion(A_1 -> A_2) = count(w)$ where word w is written by A_1 after the text by A_2 but no further than *discussion_distance* words away. *discussion_distance* is a parameter with a typical value of 20.

A typical case for the discussion on talk pages is that at least 20 words are written by each participant. However, increasing the *discussion_distance* parameter would result in ignoring shorter exchanges. The effect is that typically, after each exchange between authors A_1 and A_2, the value of the strength of the edge in the discussion network between them increases stepwise by $210 = \sum_{i=1}^{20} i$.

4.1.4 Topics

The Topic dimension is a bit different from the others, because to the set of nodes is extended by a subset of Wikipedia categories and edges form a bipartite graph

connecting authors to the categories of the articles that they have edited. The strength of the edges is proportional to the number of distinct articles in a particular category, in which the given editor has made at least one edit. Not all categories have been added to the set of nodes: we have attempted to filter out non-topical categories (for example, dates, or the "disambiguation" category). We hypothesize that the Topics dimension may be interpreted as a relation of interest or knowledge of an author in a topical category.

The edge strength in the Topics network from an author A to a category C is given by: $Edits - in - category(A -> C) = count(a)$ where article a was edited by A and belongs to category C.

4.2 Common links in behavioral social network and RfA networks

In our experiment, we have created networks where nodes represent editors who took part in a voting and edges represent the cast votes. The first network represents who voted for whom and the second, who voted against whom. Each vote has been converted to an edge in the graph connecting the person, who cast the vote with the person, for or against whom the vote was. We will take a look at each of those networks independently.

We have found the intersection of the voting networks and Wikipedia Behavioral Social Network dimensions: Co-edits, Reverts and Discussion.

This way we have social network measures for each pair voter-candidate. To find out how those measures are related to the voting, we have found the values in each dimension for each pair. Next, we present the cumulative relative frequency distribution graphs of those three measures separately for votes "for" and "against." Each graph is followed by a table with basic statistics: Minimum link strength value in given dimension (Min), first quartile of those values (1st Qu.), Median—the second quartile (Median), Average value (Mean), third quartile (3rd Qu.), maximum value (Max) and the percent of votes which actually has corresponding link in Wikipedia behavioral social network (coverage).

4.2.1 Co-edits and votes

As it can be seen in Fig. 7, the votes "for" suggest strong link between voter and candidate. The table summarizes the co-edits values for edges between voter and candidate in votes-for and votes-against networks. High coverage means that great majority of voter-candidate pairs have corresponding links in co-edits network.

4.2.2 Reverts and votes

Figure 8 shows the relative frequency distribution of the Reverts measure. Table shows a summary of values in this measure for votes-for and votes-against. The Reverts network is based on reverting edits. It is clearly visible that for measure value around 20 and higher there are practically only "against" votes. Low values (around 1–3) slightly suggest that the vote will be "for", but there is not too big difference. In this dimension there is low coverage, which means that there are more not existing links (which may be

Measure	votes-for	votes-against
Minimum	0.05	0.05
1st quartile	13.92	9.35
Median	57.26	41.08
Mean	442.24	287.71
3rd quartile	217.83	164.24
Maximum	170,972.51	54,129.07
Coverage	87.44%	79.02%

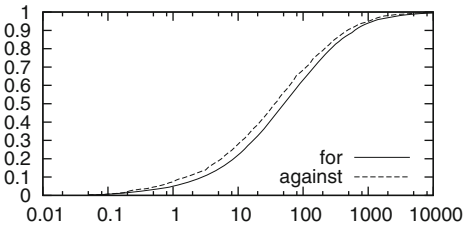

Fig. 7 Wikipedia Behavioral Social Network measures for *Co-edits* dimension (statistics and cumulative relative frequency distribution)

Measure	votes-for	votes-against
Minimum	1.000	1.000
1st quartile	1.000	1.000
Median	1.000	1.000
Mean	2.331	4.203
3rd quartile	2.000	3.000
Maximum	165.000	165.000
Coverage	11.17%	14.06%

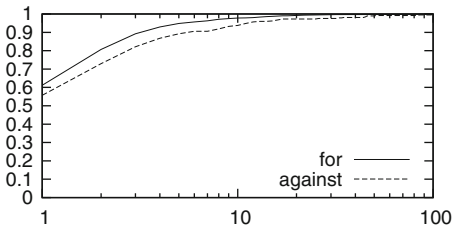

Fig. 8 Wikipedia Behavioral Social Network measures for *Reverts* dimension (statistics and cumulative relative frequency distribution)

Table 1 During which voting the candidate was accepted

Attempt	Accepted	Rejected	Percent successful
1st	145	83	63.60
2nd	18	21	46.15
3rd	4	10	28.57
4th	0	6	0.00
5th	0	1	0.00

Table 3 Statistics for networks created from votes

Measure	Votes "for"	Votes "against"
Co-edits median	57.26	41.08
Co-edits mean	442.24	287.71
Reverts mean	2.331	4.203
Reverts 3rd quartile	2.000	3.000
Discussion median	492	405
Discussion mean	1,035	751

regarded as links with zero value). More of them are in votes-for group, which suggests that not having any edit reverted supports getting a vote "for" (Table 1).

4.2.3 Discussion and votes

Similar to above, Fig. 9 describe results of analysis of the Discussion network. This dimension, however, is a bit more sparse than co-edits (lower coverage) allows better discrimination between "for" and "against" votes for values over around 200. For lower values of discussion it is difficult to tell the outcome of casting a vote.

4.3 Significance of difference between mean values

We have also performed the study of statistical significance of the results on the entire population of cast votes compared to the entire potential population of possible votes. One sample unit in the data sets corresponds to a pair of voter and candidate and their respective values in behavioral social network. The Table 2 shows the results of Welch two sample t test for the data from votes-for and votes-against. This test is an adaptation of Student's t test for use with samples with possibly not equal variances and shows the statistical significance of the difference between mean values of co-edits, reverts and discussion measures. The table shows the hypotheses verified, t statistic, number of degrees of freedom (df) and the p value which is a base to accept or reject the mentioned hypotheses. The significance level of a test is chosen to be 0.01.

The Table 3 summarizes the data from previous tables and presents only the relevant values, i.e. those which differ noticeably among "for" and "against." It is clearly seen that average co-edits value is almost two times higher for votes "for" even though median is not so distinctive. This is caused by some outliers in the "for" network, much higher than the average. They certainly predict vote value as positive.

The strengths in the Reverts network are very concentrated in lower values, it has a lot of links of value one, therefore min values, 1st quartiles and medians are equal to one. Very distinctive here is mean—this clearly suggest that even from very low values (below 4) we cannot reliably predict the vote, but if the value is much higher, the vote has very high probability of being negative.

Fig. 9 Wikipedia Behavioral Social Network measures for *Discussion* dimension (statistics and cumulative relative frequency distribution)

Measure	votes-for	votes-against
Minimum	1	1
1st quartile	210	210
Median	492	405
Mean	1,035	751
3rd quartile	1,186	811
Maximum	25,093	72,362
Coverage	64.49%	53.93%

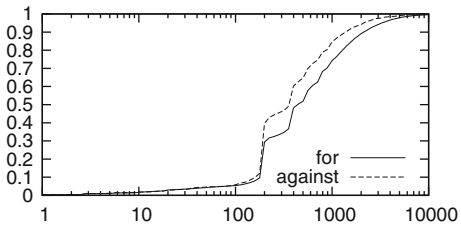

Table 2 The statistical significance of difference between mean values of network measures in votes-for and votes-against data sets

Measure	Hypothesis	t	df	p value	Result
Co-edits	Mean co-edits value is higher in votes-for than in votes-against	3.571	5537.2	0.00018	true
Reverts	Mean reverts value is higher in votes-against than in votes-for	−2.524	403.8	0.00600	true
Discussion	Mean discussion value is higher in votes-for than in votes-against	4.674	1674.0	0.000002	true

Table 4 Comparison of decision trees for predicting RfA votes

Tree name	Misclassification rate (%)
$Tree_1$	16.1
$Tree_2$	30.7
$Tree_3$	17.9
$Tree_4$	27

The third variable—discussion is better shown on a graph. According to the table the difference of values in median and mean are not great, but the distributions presented before suggest a threshold value, when probability of positive vote is significantly higher (Table 4).

4.4 Classifying RfA votes using the MBSN

In the previous section, we have compared the statistical properties of the relations in the MBSN that corresponded to votes for and votes against in the RfA procedure. The comparison has shown that the three dimensions of the MBSN that are relations between editors: co-edits, reverts and discussion, are related to the votes for or against in RfA votings. In this section, we will consider how to classify the RfA votes based on the MBSN.

We have used a standard data-mining approach: decision trees. The initial dataset of 15,556 votes (for or against) was modified so that the votes against (which were a minority of about 16 %) were proportionally repeated in order to create a balanced set. The balanced set was split up into two sets, the training set and the validation set (70 and 30 %, respectively), preserving the balance of votes for and against a candidate.

We have used the three dimensions (co-edits, reverts and discussion) to create variables that could directly be used to predict votes (since these variables are a function of the relation between the voter and the candidate). However, each of the dimensions was used twice, changing the direction of the relation. For example, the co-edits dimension was used to produce two variables: $coedits_{vc}$ (voter to candidate) and $coedits_{cv}$ (candidate to voter).

We have also used the fourth dimension of topics. The topics dimension was transformed into the following three variables: the number of topics in which a candidate for an admin was active ($topics_c$), the number of topics in which a voter was active ($topics_v$) and the number of common topics between the voter and the candidate (*topics_common*). The variables $topics_c$ and $topics_v$ were further modified by introducing thresholds on the number of edits in a category.

The variables created from the MBSM were complemented by the simple criterion of the total number of edits of a candidate.

We have used these variables to create multiple decision trees, with varying constraints. In our analysis, we have used SAS Enterprise Miner version 6.2. The trees were evaluated using a misclassification rate measure, which is the ratio of the sum of amounts of false positives and false negatives to the number of all cases.

The best decision tree used all variables and had a misclassification rate of about 16 %. We shall refer to this tree as $Tree_1$. Another tree was constrained so that it could only use relational variables (including $topics_common$), excluding the variables that counted the numbers of edits in categories for a voter or candidate ($Tree_2$). $Tree_3$ was constrained so that it could only use the total number of edits and numbers of edits in topical categories. Yet another tree was constrained by depth and width, resulting in a tree that had about 40 leaves ($Tree_4$).

A comparison of the classification correctness of the various trees shows that it is possible to classify the votes in RfA procedures with an almost 85 % accuracy. The best tree is quite large and uses all variables, although the variables derived from the Topics dimension are the most significant. Limiting the tree to just the non-relational topical variables decreases classification accuracy by about 2 %, which is significant if we want to recommend the best candidates. However, such a limitation has the advantage that the resulting tree excludes all relational variables, meaning that *the resulting classification is guaranteed to be impartial with regard to social relations among the voter and candidate*. $Tree_2$ also excludes the variables created from the Discussion dimension, which can be interpreted as acquaintance among editors (as we show in the next section). Using just the relational variables that depend on the relations between the voter and candidate (or vice-versa), it is possible to classify votes with almost 70 % accuracy, which shows that these variables are indeed significant for the RfA votings.

The reduced $Tree_4$ has a manageable size of about 40 leaves and achieves an accuracy of about 73 %, which may be considered still good enough to recommend candidates for RfA procedures. The tree uses the following variables: the number of categories that the candidate edited in, the total number of edits of the candidate, the coedits of the candidate and the voter, the number of common categories that the candidate and voter edited in, and the discussion of the candidate and the voter.

The SAS Enterprise miner package allows for a comparison of the variable importance in a decision tree. A variable's importance depends on the strength of the influence and the number of cases influenced. Variable importance is expressed as an average percentage of the variable's importance in predicting the class of each individual (for details, see Neville 1999). Table 5 shows the importance of selected variables in the best decision tree ($Tree_1$).

Table 5 Variable importance in $Tree_1$

Variable name	Variable importance (%)
total_edits	99
$topics_c$	100
topics_common	73
$coedits_{vc}$	31
$discussion_{vc}$	20
$strong\_discussion_{vc}$	18

As expected, the number of edits is a strong criterion of candidate evaluation. However, the topic diversity of a candidate's edits is even stronger. These two simple measures could be used to formulate criteria for nominating new admin candidates by the admin community.

The number of topics in common between a voter and a candidate is a very significant variable. This finding positively validates hypothesis B of the paper: new admins are indeed elected on the basis of similarity of experience in editing of articles on various topics.

On the other hand, the importance of other variables derived from the relations of the MBSN is low when compared to the other criteria. The last variable in the table, *strong_discussion*, has been created by selecting the edges in the Discussion dimension that had a strength of at least 630. As will be shown in the next section, this threshold is significant for an interpretation of this variable as acquaintance among the voter and candidate. If this interpretation holds, we can conclude that hypothesis A does not hold: acquaintance does not play an important role in the election of new admins. In the simplified tree ($Tree_4$) the relative importance of *strong_discussion* is even weaker, indicating that in $Tree_1$ the variable is used to enhance the classification correctness above the level of 73 %.

5 Validation of the MBSN

The MBSN can be considered as just a set of behavioral social networks that can be used for various purposes, like recommendation. In the previous section, we have shown that four dimensions of this network are tied to the votes cast in the RfA procedures. The MBSN may be a new, valuable tool used for recommending candidates for admins.

However, we can only consider the MBSN as a valid social model of the community of Wikipedia editors if we can validly interpret the dimensions as meaningful social relationships. In the literature, this interpretation has usually been assumed, and validated partially through indirect evidence. For example, if the Reverts dimension can be found to be related to edit wars, it can be argued that this indirectly validates that the Reverts network can be interpreted as a social relation of conflict or distrust. In the previous section, we have shown that the strength of edges in the Reverts dimension is higher for votes against than for the votes for a candidate in the RfA procedure. This finding can also be used as indirect evidence for interpreting the Reverts dimension as conflict or distrust.

In our research, we have attempted to directly validate the hypotheses concerning the interpretation of dimensions of the behavioral social network. We have used a survey of over 100 Wikipedia editors (survey results included several thousands of declared relations) to gain declarative data that can be used to validate our behavioral social networks. Respondents have been invited to the survey through an announcement at a Wikipedia event (WikiMania) and through a Polish-language Wikipedia mailing list, personal contact and using snowball sampling. Therefore the choice of respondents is not representative. However, for every respondent we have randomly chosen a subset of his/her relations using stratified sampling scheme to provide both weak and strong relations. Subsequently, relations have been weighted to adjust sample structure to population structure and so to improve representativeness of the sample.

We have validated five hypotheses for the current dimensions of the behavioral social network:

1. Hypothesis 1: The Co-Edits network may be interpreted as trust in the ability of an editor to produce content of good quality
2. Hypothesis 2: The Reverts network may be interpreted as conflict between editors
3. Hypothesis 3: The Discussion network may be interpreted as acquaintance among editors
4. Hypothesis 4: The Topics network may be interpreted as interest of an editor in a topic
5. Hypothesis 5: The Topics network may be interpreted as expert knowledge of an editor about a topic

The survey included eight questions for validating particular relations and eight questions regarding respondents' demography and Wikipedia usage, out of which we list the most important ones here:

1. "We would like to know how many Wikipedians do you know. Please name every one that you can remember (use nicknames)."
2. "Look at a list of Wikipedians with whom you have edited the same articles. Mark nicknames that you recognize (you remember that you have seen them before)"
3. "Please select the nicks of editors that have, in your opinion, edits of a good quality."
4. "Please select the nicks of editors with whom you have at any time disagreed with or argued with."

5. "Please look at the following list of Wikipedia categories. Please select the categories in which you have expert knowledge."
6. "Now select the categories that you are interested in."

Hypothesis 3 has been found to be supported by our data, while the other hypotheses were not. The negative validation of the other hypotheses points to an important difficulty in the use of behavioral social networks. A common-sense interpretation of a behavioral network, even if supported by indirect evidence, may turn out not to be valid when confronted with declarative data which have a direct social interpretation. A particularly interesting case is the Reverts dimension that has often been interpreted as conflict in the literature, yet turned out to be very weakly connected to declared conflict in our data. We consider this result to be one of the more important contributions of the paper, since it points out the need for further research that could create new behavioral networks (perhaps using more complex operational definitions) that would be better suited to be interpreted as valid social relations.

We are currently investigating new dimensions of the MBSN that could be used to operationalize the remaining social relations. The initial results of using a data mining approach to create new dimensions that would fit the declared social data are promising. However, in this article we shall focus on the successful validation of the Discussion network as acquaintance among editors. We will now describe our validation approach and its results in more detail.

5.1 Validation of the discussion network

The common meaning of acquaintance is quite obvious and intuitive one but for the application in network analysis more precise definition and operationalization is needed—one that allows for measurement and empirical research.

How do we know if two people are acquaintances or not? How do we know that two people know each other? What are indices of acquaintance? One way to know if people are acquaintances is to ask them. If they declare "yes, we know each other" we can assume they are acquaintances. One could say that in that case we learn that through "declaration-based indice" of acquaintance.

Another way is to observe how people behave. For example if they shake hands we can be quite sure they know each other. In case of Wikipedia we cannot watch people shaking hands but we can observe how they communicate via talk pages. If their posts are next to each other we can be quite sure that some kind of conscious communicative interaction between them has taken place. Therefore we may assume that they are acquaintances at least at some basic level. In that case we base our knowledge on "behavior-based indice" of acquaintance.

Both types of indices, declarative and behavior-based, give legitimate and common sense ways to measure acquaintance. We use them in everyday life to recognize social relations around us. Both indices have their strengths and weaknesses. Declaration-based indice is very straightforward to understand, but depends heavily on a person's memory—usually we have more acquaintances that we can name. But even if one cannot name her acquaintance some time after last interaction they are not strangers any more—some kind of acquaintance still exists which makes much easier to refresh the relation.

On the other hand, the behavior-based indice is totally memory independent—in Wikipedia dump data one can detect a trace of acquaintance even years after the last interaction. But it can be misleading, too. A small talk may be not enough to constitute a relation. One could give not enough attention, the topic could be insignificant, one could barely noticed the fact of interaction or even have not noticed it at all. Moreover, some scholars question the validity of comparing "virtual" relations with "real" ones, claiming that there is a qualitative difference between the two. In the case of virtual acquaintance, another important concern arises: the number of virtual acquaintances may grow even larger than "real" ones, increasing the likelihood that virtual interactions may not be an indicator of "real" relations.

Being conscious of strengths and weaknesses of presented measures, we consider both of them good indices of social relation of acquaintance in terms of "face validity" (Babbie 2007)—in our opinion they adequately depict meaning of the notion. But can we prove their usefulness by showing some evidence that they are measuring the same concept? In social science this is a question of so-called "criterion-related" or "predictive" validity (Babbie 2007) which is tested by studying indexes correlation and ability to predict value of one indice knowing value of another.

To answer the questions of "predictive validity" we conducted a survey study of $n = 111$ polish Wikipedia editors. For each editor we inquired about his relations with other Wikipedians that we detected using our "behavior-based indice" of acquaintance so we could compare values of "behavior-based" and "declaration-based" indices.

To test the predictive validity of the "behavior-based indice" for every relation assumed to be acquaintance we have asked the respondent two questions:

1. "Look at a list of Wikipedians with whom you have edited the same articles. Mark nicknames that you recognize (you remember that you have seen them before)".

2. "We would like to know how many Wikipedians do you know. Please name every one that you can remember (use nicknames)."

Both questions are "declaration-based" indices. We call the first one a "recognition indicator" and the second one a "recall indicator". Recall depends more heavily on memory so respondents reported much fewer acquaintance than recognition. On the other hand recall allows for more spontaneous answer which is not aided with some kind of pre-made list. Total number $n = 874$ relations were evaluated with this procedure.

On that basis we could assess value of our "behavior-based indice" for predicting acquaintance relation operationalized with our "declaration-based indice". We found out that 45 % of editors identified as acquaintance with "behavior-based indice" are recognized by the respondent and 6 % are recalled.

Since this result is far from satisfying our need of a valid measure, we have decided to improve it by increasing the cut-off point of behavior indice strength. Using this approach, edges with low strength were dropped out and no longer treated as an indicator of acquaintance relation. We found out that the prediction validity varies with strength of "behavior-based indice" cut-off point. For recognition, it ranges from about 45 % of recognized editors for very low values of "acquaintance behavior-based indice" to 96 % of recognition for very high values of indice (Fig. 10). For recall, the prediction validity increases from 6 to 53 %.

The stepwise shape of the plot on Fig. 10 is explained by the typical increase in strength of the edges in the Discussion network by 210 (see Sect. 4.1.3). The shape of the figure also explains the threshold of $630 = 3 \times 210$ used to select the strongest relations in the Discussion network in the data mining analysis described in the previous section. Over 80 % of the relations in the Discussion network that exceed the strength threshold of 630 are recognized as real acquaintances in our survey.

A high quality of prediction of declarative acquaintance with behavior-based index is not enough to acknowledge validity of these indicators. For example, if all editors were declared acquaintances the prediction value of behavior-based index would be 100 %, even if only a few declared acquaintances were discovered with that index. Therefore we had to test the ability to predict "behavior-based indice" with respondents declarations. This was possible only for recall declaration—due to methodological reasons it was not possible to aid recognition with some pre-made list of random editors.

We have gathered data for $n = 270$ declared relations. Next we have studied how many of the listed editors were identified with "behavior-based indice". Again, we found out that the result depends on the "behavior-based indice" cut-off point. It starts from 90 % of identified relations for very low values of "acquaintance behavior-based indice" and sharply falls with increasing values of that indice (Fig. 11).

According to the data presented above, increasing the behavior-based indice cut-off point improves prediction value of behavior-based indice but at the same time worsens the prediction value of declarative-based indice. In order to choose an optimal cut-off point we have decided to use Pearson's R correlation between indices as an overall measure of their "predicitive validity". We have estimated Pearson's R between "behavior-based indice" (binary variable based on strength of Discussion relation) and recall for various cut-off points of discussion relation's strength.

Figure 12 illustrates the relationship between "behavior-based indice" cut-off point and "predictive validity" measured with estimated Pearson's correlation. It turns out that indices are most valid when cut-off point belong to interval [220;800], with a maximum in the interval [630,770]. So it is worth recommending to use cut-off point equal to 630 which maximizes both "predicitive validity" and number of identified acquaintance relations.

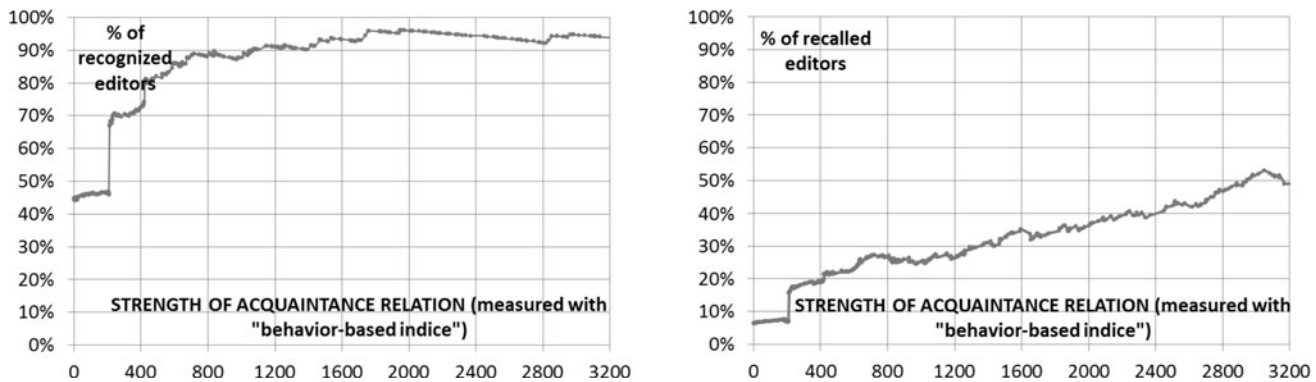

Fig. 10 Predicting recognition and recall of acquaintance with Discussion network

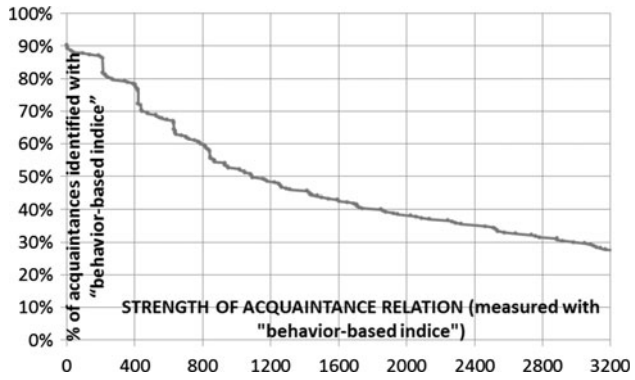

Fig. 11 Percentage of recalled acquaintances as a function of increasing edge strength in Discussion network

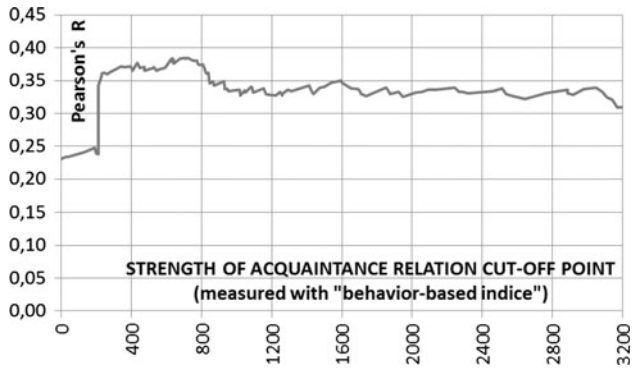

Fig. 12 Estimated correlation between declared acquaintance and edge strength in Discussion network

At the end of this section, we would like to briefly summarize the negative evaluation of the hypotheses concerning the other dimensions of our behavioral social network. We have used a similar approach as described for the Discussion dimension. The predictive validity of the Co-edits dimension for the declared trust in the ability of editors to produce good content has been about 50 %. The predictive validity of the Reverts dimension for declared conflict has been about 40 %. The predictive ability of the Topics dimension for declared expert knowledge or topical interest has been below 50 %. In our opinion, these results do not exclude the possibility of using behavioral social networks to operationalize valid social relations (since the validation of the Discussion network's interpretation is positive); however, more work is needed in order to propose better dimensions for the social concepts described above. These results also, our results do not exclude the possibility of using the Co-edits, Reverts and Topics dimensions in practice for recommendation purposes. However, an application that relies on the interpretation of these dimensions as trust, conflict, knowledge or interest should be regarded with caution and tested independently.

6 Conclusions

In this article, we have studied the Request for Adminship votes on the Polish-language Wikipedia. We have noticed the decreasing amount of successful admin elections and have formulated two hypotheses that could explain this phenomenon. Hypothesis A stated that new admins are elected on the basis of acquaintance of the voter and candidate. If this would be a valid explanation, we could conclude that the community of admins is becoming increasingly closed, which would be detrimental to the sustainable development of the Wikipedia.

Hypothesis B stated that new admins are elected on the basis of similarity of experience in editing various topics of the voter and candidate. Since voters are other active admins whose experience increases with time, their thresholds of accepting a candidate are likely to increase (as has been observed from the simple statistics of RfA votings).

It should be possible to improve the likelihood of electing new, good admins by changing the criteria of nomination for the RfA procedure. Currently, these criteria are based just on the number of edits and are much lower than the real thresholds of candidate acceptance. Our research suggests that there could be two criteria: one based on the number of edits, the other based on edit diversity (measured by the number of topical categories in which a candidate has edited). It would also be possible to use automatic classification based on the MBSN to recommend candidates with a high likelihood of acceptance (based on past votes). The recommendation could be made in an impartial manner (not depending on the identity of the voter) by using only variables based on the Topics dimension of the MBSN.

In order to validate the two hypotheses, we have used the Multidimensional Behavioral Social Network created from edit history of the Polish-language Wikipedia. The MBSN network can be used as a general model of the Wikipedia knowledge community and is versatile enough to model various social phenomena, such as teamwork (Turek et al. 2010, 2011). This article shows that the MBSN can also model admin elections. A data mining model for classifying RfA votes for and against a candidate has been based mostly on variables derived from the MBSN. The model has an accuracy which of about 84 %, which should be sufficient to recommend good candidates for elections.

A variable which expresses the number of common topics in which both the voter and the candidate have edited has been found to be highly important in decision trees for classifying votes. This observation validates the hypothesis B: similarity of experience in editing various topics among the voter and the candidate significantly increases the likelihood of a vote for that candidate. On the

other hand, the same does not hold for a variable derived from the Discussion dimension of the MBSN (based on close edits on Wikipedia talk pages) by selecting the strongest relations. We have used a survey of Wikipedia editors that supports the interpretation of such strong relations in the Discussion dimension as real acquaintance among editors. Because of this fact, we claim that hypothesis A does not hold: acquaintance of the voter and candidate does not play a significant role in the RfA votings. This is fortunate and shows that the admin elections are open to new candidates outside the acquaintances of current admins. The increase in acceptance criteria is not a sign of a closing community of admins.

The sustainable development of the Wikipedia is of importance to all Internet users worldwide. Our study of the Polish-language Wikipedia shows that it is possible to understand and model the process of electing new admins, who play a critical role in maintaining and increasing the quality of the Wikipedia. It should also be able to support that process by recommending new candidates based on edit history. We hope that these contributions can play a small role in supporting Wikipedia development.

Acknowledgments This research has been supported by the Swiss grant Reconcile: Robust Online Credibility Evaluation of Web Content through the Swiss contribution to the enlarged EU. The authors would like to thank Maciej "Nux" Jaros for extracting data from Wikipedia database for statistics in Sect. 3

References

Adler BT, Chatterjee K, de Alfaro L, Faella M, Pye I, Raman V (2008) Assigning trust to Wikipedia content. In: WikiSym 2008: international symposium on Wikis and open collaboration

Burke M, Kraut R (2008) Taking up the mop: identifying future Wikipedia administrators. In: CHI '08: CHI '08 extended abstracts on human factors in computing systems. ACM. New York, pp 3441–3446

Babbie ER (2007) The practice of social research. Wadsworth Publishing, Cengage Learning, Belmont

Kazienko P, Musial K, Kukla E, Kajdanowicz T, Brdka P (2011) Multidimensional social network: model and analysis. In: ICCCI (1), pp 378–387

Kennedy KT (2009) Synthesis, interdiction, and protection of layered networks. Department of The Air Force. Air Force Institute of Technology, USA, PhD Dissertation

Kittur A, Chi E, Pendleton B, Suh B, Mytkowicz T (2007) Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In: Proceedings of CHI 2007. ACM Press

Kittur A, Suh B, Pendleton BA, Chi EH (2007) He says, she says: conflict and coordination in Wikipedia. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI 2007. ACM. New York, pp 453–462

Leskovec J, Huttenlocher D, Kleinberg J (2010) Governance in social media: a case study of the Wikipedia promotion process. In: Proceedings of the 4th international AAAI conference on weblogs and social media, ICWSM 2010. AAAI Press. Menlo Park, CA

Neville P (1999) Decision trees for predictive modeling. SAS Institute Inc., USA

Priedhorsky R, Chen J, Lam SK, Panciera KA, Terveen LG, Riedl J (2007) Creating, destroying, and restoring value in Wikipedia. In: Proceedings of the 2007 international ACM conference on supporting group work, GROUP 2007. ACM. New York, pp 259–268

Rodriguez MA, Shinavier J (2009) Exposing multi-relational networks to single-relational network analysis algorithms. J Informetr 4(1):29–42

Spinellis D, Louridas P (2008) The collaborative organization of knowledge. In: Communications of the ACM—designing games with a purpose, vol 51(8), August 2008. ACM. New York

Suh C et al (2009) The singularity is not near: slowing growth of Wikipedia. In :Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym 2009). Orlando, Florida, pp 1–10

Turek P, Wierzbicki A, Nielek R, Hupa A, Datta A (2010) Learning about the quality of teamwork from Wikiteams. In: Proceedings of the 2010 IEEE second international conference on social computing, SocialCom/IEEE international conference on privacy, security, risk and trust, PASSAT 2010. Minneapolis, pp 17–24

Turek P, Wierzbicki A, Nielek R, Hupa A, Datta A (2011) WikiTeams: How do they achieve success? IEEE Potentials 30(5):2–7

Turek P, Spychala J, Wierzbicki A, Gackowski P (2011) Social mechanism of granting trust basing on polish Wikipedia requests for adminship. In: Proceedings of the international conference on social informatics (SocInfo 2011). Singapore, pp 212–225

Viégas F, Wattenberg M, Kushal D (2004) Studying cooperation and conflict between authors with history flow visualization. In: Proceedings of the 2004 conference on human factors in computing systems. ACM, New York

Vuong B, Lim E, Sun A, Le M, Lauw H (2008) On ranking controversies in wikipedia: models and evaluation. In: WSDM 2008, pp 171–182

Zhang Y, Sun A, Datta A, Chang K, Lim E (2010) Do Wikipedians follow domain experts? In: A domain-specific study on Wikipedia knowledge building. JCDL 2010, pp 119–128