

Does the Administrator Community of Polish Wikipedia Shut out New Candidates Because of the Acquaintance Relation?

Justyna Spychała

Faculty of Polish Studies
University of Warsaw
Warsaw, Poland

Email: jspychala@gmail.com

Mateusz Adamczyk

Faculty of Mathematics, Informatics and Mechanics
University of Warsaw
Warsaw, Poland

Email: adamczyk@mimuw.edu.pl

Piotr Turek

independent
Warsaw, Poland

Email: piotr@gmail.com

Abstract—Administrators of Wikipedia are its most dedicated users, which are granted special privileges and burdened with great responsibility for the Wikipedia. The administrators are usually nominated by the community and then elected by voting or by reaching a consensus. This paper examines period 2005 – 2011. In most recent years of examined period, decline in a number of newly appointed administrators can be observed. There are two main hypotheses of seen phenomena: with growth of Wikipedia it is harder to become an administrator or existing administrator community is shutting out new candidates due to acquaintance relation. This research is an attempt to find out whether the community is not becoming less open to new users and new potential administrators, because of their lack of chances to gain reputation. The key here is the understanding of social aspects driving the process of Request for Adminship (RfA) votings. Based on our previous work focused on social networks induced from collective activity of wikipedians, this paper extends it by the annual analysis of obtained statistics and examination of clustering coefficient as an approximation of social capital. We present the dynamics of relationships between voters and candidates across several years of Polish Wikipedia development. Obtained data allowed us to answer the question, whether administrator community of Polish Wikipedia shuts out new candidates because of the acquaintance relation.

Keywords—Multidimensional Behavioral Social Network; Wikipedia; Request for Adminship; Clustering Coefficient; Administrators' Lifespans; Administrator Community versus core; Wikipedia's users behavioural patterns.

I. INTRODUCTION

Spychała et al. [1] presented results of their research on The Third International Conference on Advanced Collaborative Networks, Systems and Applications (COLLA 2013 [2]) in Nice, France. Presented research concerned possibility of closing up the Administrator Community of Polish Wikipedia by shutting out new candidates. It was a refinement of work done by Turek et al. [3]. This paper is an extended version of paper written by Spychała et al. [1] and presented on COLLA 2013. Similarly to paper written by Spychała et al. [1], this paper covers the range of years 2005-2011 and its goal is to answer the question if the reason for decrease in successful RfA votings in Polish Wikipedia is choosing Administrators based on acquaintance. As the administrator community of Polish Wikipedia is rather small, especially compared to the English one, it could be valid claim. But, we argue that it is not

the case. Probably, it is caused by growing expectations about new candidates. As the Polish Wikipedia grows, the number of articles also grows. Articles are getting longer and more complex. Users gain more experience and have richer edits history. This paper contains some data and conclusions, which Spychała et al. [1] were unable to present due to limitations placed on length of the conference publication.

Administrators (or sysops) are very dedicated and trustworthy participants of the Wikipedia projects in all language versions. Thanks to community decision, they have received special privileges and use administrative tools to exercise preventive and policing functions. Administrators have the right to edit all the Wikipedia articles as well as many other privileges—understood rather as duties. These powers are not meant to give them editorial control over the project, but rather provide mentoring and technical assistance in other wikipedians' work. Administrators also serve by providing assistance, especially to newcomers, in editing of Wikipedia. All newly registered users get their guides—the administrators to whom they can always turn for help and be sure they will receive it as soon as possible.

Kittur et al. [4] claimed, that due to increasing amount of management work at Wikipedia, such as content quality control, coordination, maintenance, that are caused by the increasing popularity and amount of content in Wikipedia, the importance of administrators is increasing. This creates a potential risk that administrators may become overwhelmed by the amount of work and their response times become longer. Especially, Ortega et al. [5] showed that after peak of popularity of the Free Encyclopaedia, its user base growth slowed down. But in the same time, the amount of content is still increasing. If the Wikipedia is to keep its pace of growth, then some measures to increase its users base have to be taken.

However, the Polish community of Administrators is growing slower than expected; hence, the question whether and why this community shuts out candidates for new members. Currently, there are 149 administrators on Polish Wikipedia—for comparison, 1,147 administrators work currently on the English version. Of course, the English version is much more developed, but sheer number of people with administrative privileges is impressive. On the other hand, Ortega et al. [5] notes that the Polish Wikipedia has the highest rate of auto-

mated bots used to help administrators in their duties. This fact can explain why the number of administrators is not enough to prophesize doom of Polish Wikipedia.

Administrators are elected in a special procedure, the rules of which are clearly defined. This procedure is called Request for Adminship (RfA). As it was already mentioned, the privileges for administrators are granted by Wikipedia community, in a vote in which the right to vote is given to those Wikipedia users, who are well-known and respected members of the community and know and respect the established rules on the website. Wikipedians who are candidates for the administrator position must “have a minimum of 1,000 not deleted edits, first of which has to be made at least 3 months prior to the date of filing the candidacy”. Nominations for administrator candidates are adopted by a special form on the web page that also contains the regulations and the list of candidates. New administrators are elected during a voting that lasts a week (168 hours). Wikipedians, in order to be allowed to vote, must have been registered for at least one month and have a minimum of 500 not deleted edits.

Interestingly, in case of English version of Wikipedia, no formal conditions are required in order to declare a candidacy for an administrator. The only conditions are possession of an account and trust among other users. Despite this, the page with the declaration forms contains the information that in case of self-nominating, it is recommended to have at least 2,000 edits for a minimum period of 3 months. Another important difference is that in the case of English version of Wikipedia, new administrators are elected not by voting, but by discussion. Moreover, “the consensus in RFA is not achieved by exceeding a threshold, but by the strength of the justification of the candidacy”.

There is one aspect of this paper which distinguishes it from work presented by Spychała et al. [1], namely our analysis of lifespans of administrators. It allowed us to make sure that research, which is main topic of this paper, is well founded. What is more, analysis of lifespans of administrators allowed us to show analogy between our work and part of Ortega’s work [5]. Moreover, one can clearly see correspondence between behavioural patterns exhibited by administrators and those exhibited by the core of most active users. Analysing sub-communities of user community of Wikipedia is challenging and interesting research direction, which we plan to take in our future research.

The rest of this work is divided as follows: in Section II, the related work is presented. Section III contains data motivating our research and base statistics, which show that growth of Polish-language Wikipedia Administrators group has slowed down. Data presented there is extended by data for year 2011, in comparison to data gathered by Jankowski-Lorek et al. [6]. In Section IV, Multidimensional Behavioural Social Network is used to analyse historical voting data. This analysis is the main contribution of this paper. Section IV also contains answer to question stated above. Section V presents conclusions and suggestions for future work.

II. RELATED WORK

The problem of evaluation and recommendation of users requesting for adminship in Wikipedia has been addressed in

several papers. In one of them, Burke et al. [7] try to indicate the features and qualities determinative for the user selection to the position of administrator. On the basis of publicly available tips for candidates [8], a set of attributes, that a future administrator should have has been developed. Behavioural data and comments, not page text, were used to evaluate candidates. Authors 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 mediation 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. Burke et al. also confirmed Kittur’s [9] results that the percentage of indirect work (coordination, discussion, etc.) grows over time, the share of articles in all Wikipedia edits is decreasing.

It is noteworthy that in Burke et al. [7] only the qualities of each user were evaluated. Leskovec et al. [10] have shown that the outcome of the voting depends on the candidate and his or her place in the community. They 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.

Kittur et al. [11], analysed the impact of the similarity of users on their mutual assessment. The examined data were collected from three websites: Wikipedia, Stack Overflow, and Epinions. The important feature of those websites is the possibility of mutual evaluation between their users. In case of Wikipedia it is the RfA voting. Two users were considered similar, when they have performed similar actions, which in case of Wikipedia were edits of articles. The authors concluded that, in case of Wikipedia, the possibility of casting a vote for a candidate increases with the increase of the similarity between the candidate and the voter. The voters, who are similar to the candidate, are less driven by the objective qualities (status in the community), such as experience in development of Wikipedia. Candidate’s status determines casting a vote for that candidate when the voter and candidate are only slightly similar.

An interesting observation is that during the voting, there are much more voters similar to the candidate in a group with higher status than in a group with relatively lower status. This may suggest that during RfA the voters do not constitute a representative sample of community. This allows for the outcome of the election to be predicted when profiles and similarity of a few first voters and the candidate are known. To effectively predict the result of the voting one does not even need to know the votes given by the first voters.

The quality of Wikipedia articles depends on the level of cooperation of the editors. Rad et al. [12] decided to examine

the history of article edits, and on that basis determine the mutual attitude of the editors and how controversial is the given article. Casting a vote during new administrator election was adopted as an indicator of relation between two users. If the voter has a positive attitude toward the candidate, the vote will be positive. In case of a negative attitude, the vote will be negative. The authors decided that the co-edit of the article is a pair of changes of the same section of the article, which were set apart in time by less than a fixed number of revisions. A social network with nodes labelled with users' profiles and directed edges labelled with users' co-edits, was also considered. This graph was used to induce a decision system and train a classifier, which was highly effective in predicting votes. What is important, is that this approach is complementary to the ones described earlier. It is based on the analysis of Wikipedia articles and their edit history and not on the aggregated statistics of the community. What is interesting, it turned out that it is relatively easy to predict positive votes. It seems that they are influenced by the most recent history of cooperation. On the other hand, the high quality of prediction of the negative votes required appropriately bigger and richer history of cooperation. The authors risked the statement that the users can remember disagreements for a long time and during a voting they can be guided by hidden qualities, like for example, the votes already cast in a given voting.

III. STATISTICS FOR REQUEST FOR ADMINSHIP PROCEDURE

During the period 2005 – 2011, the growth of Polish Wikipedia's content was steady. It can be seen on Figure 1, which presents an average number of new articles per day during examined period. Data for Figure 1 were obtained from freely available web-page presenting some statistics for the Polish Wikipedia [13]. On Figure 1, one can clearly see, that in the year 2006, there was a peak in pace of content growth, but in the following years the growth speed declined and steadied.

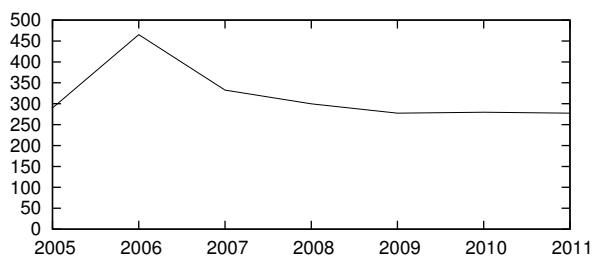


Fig. 1. The average number of new articles per day in each year

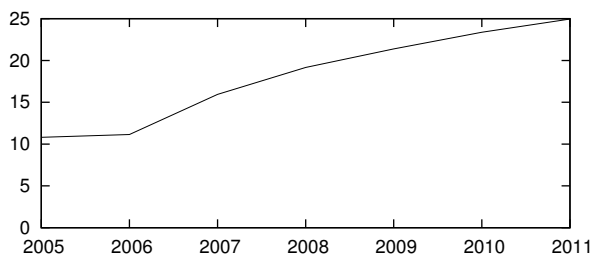


Fig. 2. The average number of edits per article in each year

In the same period, articles' content was getting more complicated. Because of that, there was increasing amount of management work, such as content quality control and maintenance. The increased articles' complexity can be approximated by an average number of edits per article. Values for averages for each year in examined period are presented on Figure 2. Data for Figure 2 were also obtained from above-mentioned web-page presenting some statistics for the Polish Wikipedia [13]. The above-mentioned figure shows, that the average number of edits per article was rapidly increasing during examined period. It indicates that amount of management work also increased, and leads to need for growth of the Administrators Community.

As of December 31, 2011 the Polish-language Wikipedia had 171 administrators. Since 2005, there were 307 votings on RfA. 177 of those ended with granting the administrator privileges to a candidate, in 110 of those, the candidates were rejected and in about 40 votings, the candidates resigned before the end of the voting and about 30 votings were cancelled (due to statutory requirements or lack of acceptance of the nomination by the candidate). About 38 administrators were chosen before the introduction of RfA procedures in March 2005. The data on RfAs do not sum up for several reasons. Among them are: verification votings and losing privileges by administrator either by giving them up or being revoked by the Arbitration Committee.

In the current version, the procedure states that a candidate for an administrator must have an account for at least three months and at least one thousand not deleted edits. In order to participate in the voting, user must have an account for at least 2 weeks and at least 500 article edits. Voting begins at the moment, when the candidates confirm, that they are willing to take the administrator position, as users can apply for the position themselves or be nominated by other users. In order to receive administrator privileges, the candidate need to receive at least 20 votes "for" and it must constitute of at least 80% of the sum of the votes "for" and "against". If the candidate does not receive the required number of positive votes or do not meet the formal requirements, he or she can apply again after at least 60 days since the end of last voting. A similar rule applies to the administrators who resigned from their position but would like to receive the privileges again.

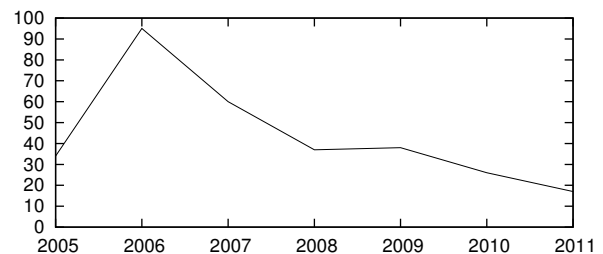


Fig. 3. Number of votings in each year

Figure 3 is presenting the number of votings in each year; the peak can be observed in the year 2006, when that number reaches 95, while a year before it reached only 34. One year after the peak, the number of votings dropped to 60. With the exception of the years 2006-2007, the number of votings never exceeded 38. In the years 2010-2011 that number declined

below 34. The number of RfAs between year 2006 and 2011 decreased by nearly three quarters (from 95 to 26).

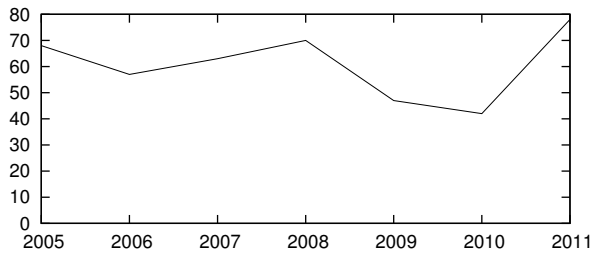


Fig. 4. The percentage of accepted RfA in each year

The percentage of the accepted nominations in each year (see Figure 4) can be divided into three periods. The first one consists of the years 2005-2008, when the percentage of the accepted candidates ranged from 57 to 70. The second period are the years 2009-2010, with the percentage below 50 (47% and 42% respectively). Between the years 2008 and 2010 the percentage of the positive RfAs fell by almost a half (from 70% to 42%). The third period, which accounts for the year 2011, is characterized by the relatively high number of positive RfAs. However, it should be noted that the number of the votings performed at that time was significantly lower than in the previous years.

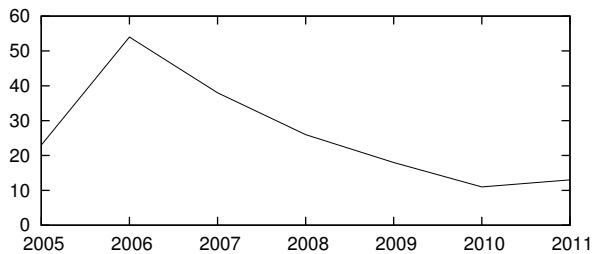


Fig. 5. The number of elected administrators in each year

The number of elected administrators is presented on Figure 5. The peak, as with number of votings, can be observed in the year 2006, when that number reaches 54, while a year before it reached only 23. One year after the peak, the number of elected administrators dropped to 38. In the following years, number of elected administrators steadily went down to 11 in the year 2010 and 13 in the year 2011. With the exception of the year 2006 and the year 2007, the number of elected administrators never exceeded 30. In the most recent years, that number declined below 20.

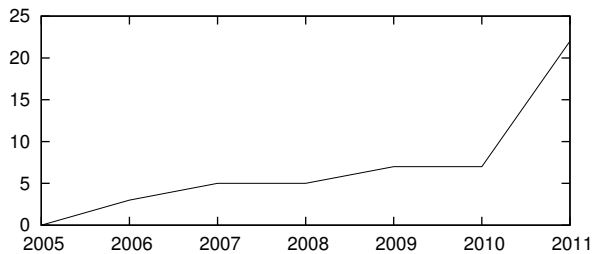


Fig. 6. The number of administrators who left the project in each year

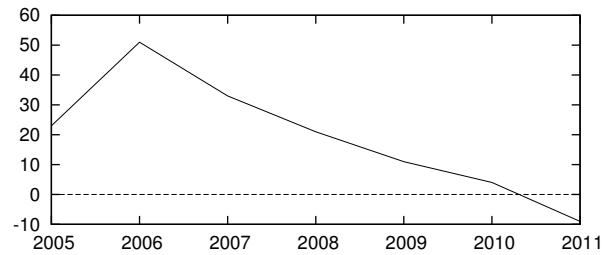


Fig. 7. The growth of the Administrator community in each year

The number of administrators who were relieved from duty is presented on Figure 6. There is steady upward trend in number of former administrators. The value in the year 2011 is equal to 22, which is probably a little too big. The errors are possible because our parsers for the Wikipedia's pages [14], [15] were not accurate enough. But the key fact is illustrated on Figure 7, which presents growth of number of Administrators of the Polish Wikipedia. The graph is an approximation of the real growth, because some face were not considered in data used to build it. One issue was with administrators who gave up their duties and were later re-elected. Since an administrator can leave the project and then come back again, in our analysis we used only the date of his first election to this function and the date of his last departure from the project. If some administrator has departed from the project and then was later re-elected, then his case lowered our approximation of growth of the Administrator Community of Polish Wikipedia by one. Fortunately, set of re-elected administrators has low cardinality. Because of that, our approximation although not accurate, can show overall trend. And it can be clearly seen, that after the peak of popularity of Polish Wikipedia in the year 2006, the above-mentioned trend in the growth of the Administrator Community of Polish Wikipedia is downward.

The next study, related to the experience of candidates prior to granting them administrator privileges, has been conducted on 97 users, who recently received them. In case of those elected before, the gathering of data was impossible because of gaps in the logs of Polish Wikipedia.

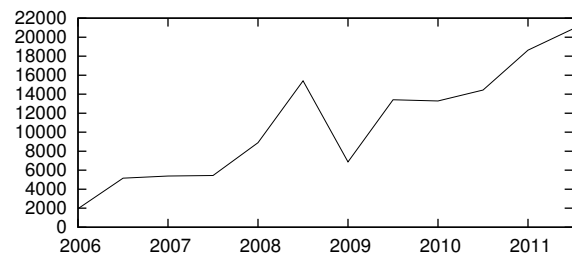


Fig. 8. The average number of edits made by user at the moment of receiving administrative privileges

One of the factors causing the most discussion during the voting is the number of edits performed by the candidate. RfA rules contain the following sentence: "Users who want to candidate for adminship (...) must have at least 1000 not deleted edits". Often, however, this number is considered by the voters to be too small. Basing on the analysis of the number of edits, it can be seen (Figure 8) that the minimum falls on

the first half of 2006 with an average of 1,957 edits. This value then grows up to 2011 when it slightly exceeds 20,000 edits. This indicates that year by year, candidates needed to have greater experience in order to be accepted as administrators. The difference between the level of experience required by the regulations and the level widely accepted has been increasing as well. A similar phenomenon can be observed in the German Wikipedia, where—according to the voters—in the second half of 2010, candidates were accepted only if they had over 10,000 edits.

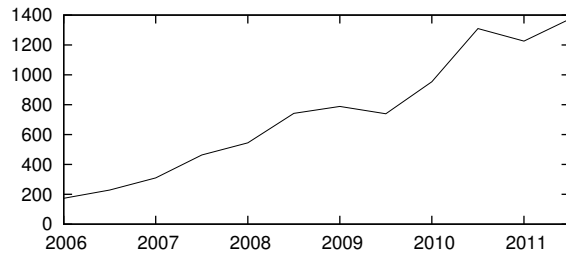


Fig. 9. The average number of days since creation of the account at the moment of receiving administrative privileges

Another factor that stirs up emotions at the time of voting is the seniority (understood as time since the first not deleted edit) on Wikipedia. The terms of voting set the following requirements: “Users who want to candidate for adminship (...) must have at least 1000 not deleted edits, the first of which took place at least 3 months before the date of candidacy proposition”. The seniority (in days) of candidates, before the date of registration and acquiring the administrator rights, had been analysed (see Figure 9). This, however, is not exactly the same value as the required by the regulations. The measured seniority in the first half of the year 2006 was 173 days. This value has been gradually increasing: from 463 days in the second half of 2007, to 788 days in the first half of 2009, with a slight decline in the second half of 2009 (739 days). In the second half of 2010 the value reached 1310 days. This result, however, may be unreliable due to the fact that during that period only two votings took place. In the second half of 2011 the measured value reached 1374 days. The overall analysis of the chart shows that in the year 2006 candidates had less than a year of seniority, however, since mid-2008 the seniority is at least two years. The last two candidates who had less than one year of experience were selected in February 2009 and November 2008.

All three figures (Figure 3, Figure 4, and Figure 5) show the downward trend in the total number of newly appointed administrators between 2006 and 2011. This decrease gives reason for serious concern as the amount of required administrative work on Wikipedia is constantly growing. This phenomenon may have several possible explanations. The first explanation is the declining number of candidates who accept their nominations for administrators (that would explain the decreasing number of RfA votings), but the confirmation of this hypothesis is beyond the scope of this paper. Nevertheless, related works have shown that in recent years Wikipedia has experienced a downtrend in the amount of user contributions, which reflects the general decline in motivation as shown by Suh et al. [16].

The second explanation states that the number of positive nominations decreases due to the changing criteria for selecting and accepting candidates. Those criteria can vary in many ways; however, our research shows that they are connected to the candidate’s experience. This experience can be initially estimated on the basis of the edits performed, but the more accurate measurement (presented by Burke et al. [7]) represents the number of article edits in a specific category.

The more damaging prospect is the fact that the administrator community is chosen on the basis acquaintance between current administrators and candidates. The next section discusses, if that is the case.

IV. ACQUAINTANCE IN THE ADMINISTRATORS SOCIETY

Acquaintance is a social phenomenon, which can lead to shutting out new members of some society. To approximate it, we used implicit social connections contained in history of talk pages in Wikipedia. We claim, that connections obtained in above-mentioned way, are indeed good approximation of acquaintance relation. Jankowski-Lorek et al. [6] proven mentioned claim to be true.

A. Data description

Data and multidimensional behavioural social network used for this paper were gathered, aggregated and made available by the team led by dr. Adam Wierzbicki. Methodology, data and networks are described in greater detail by Jankowski-Lorek et al. [6]. Examined period encompasses the years 2005-2011.

Basically, the network consists of four dimensions:

- Co-edits,
- Reverts,
- Discussion,
- Topics.

Weights in co-edits dimension are based on number of words written by one author next to the text written by some other one in the text of articles. The authorship information for a particular fragment of text was obtained by analysing its first occurrence in the whole edit history of examined page.

Edge strength in reverts dimension is based on the number of edits made by one author and reverted by other. It was obtained by searching identical revisions before the examined one. If it was found, each pair of examined revision and revisions after the other identical one was used to calculate number of reverts.

Similar to co-edits, edge strengths in discussion network were stated as number of words written by one author next to text created by other one. But in this dimension, the talk pages were considered.

The last dimension, topics, was a little different to other ones. It was a bipartite graph connecting authors with categories in which they have edited at least one article. The edge weight was exactly the number of article edits made by given author in the particular category.

One of the most important observations made by Jankowski-Lorek et al. [6], is that discussion network can be

interpreted as social relation of acquaintance. Jankowski-Lorek et al. [6] conducted another research, a survey among Polish Wikipedia users. However, interpretations of other dimensions have not been confirmed.

The data contained two more graphs: positive votes network and negative ones. If, during RfA procedure, user has cast positive vote for candidate, then an edge in the positive votes network has been created. Its weight was equal to number of positive votes cast by the user for the candidate. Weights of more than one were possible only if the user was a candidate more than once. Network of negative votes has been created in an analogous manner, but taking the negative votes instead.

Turek et al. [3] and Jankowski-Lorek et al. [6], intersected each dimension with positive and negative votes networks, in order to examine correlation between social network dimensions and RfA votings. Both graphs were analysed separately and features distinguishing them have been found.

Research presented in this paper studies only on the discussion dimension. The reason for such decision is that discussion network can be interpreted as a real relation—acquaintance. For each year, graph of discussion network has been intersected with positive and negative votes networks. Maniu et al. [17] suggest using one, signed network, especially, when there is a strong correlation between both networks as shown by Leskovec et al. [18]. Two separate graphs were used for two reasons:

- To maintain consistency with analysis presented by Turek et al. [3] and Jankowski-Lorek et al. [6],
- To separately check positive and negative impact on RfA procedure of acquaintance relation.

Ortega et al. [5], defined the core of the Wikipedia user community as a group of most active editors in analysed month. Most of the Wikipedia content is created by the group of core users. The core existed in all examined language versions of the Wikipedia, in particular in Polish and English versions, but the behavioural patterns of users were different in both language versions. The first indication of that fact are the differences in the structures of core user groups of both Wikipedias. To name just one, in the English version new Administrator is elected by reaching consensus among the users fulfilling a set of requirements, while in the Polish version – by "formal" voting within similar group. Ortega notes, that structure of the core user group of Polish Wikipedia is different than structures of similar groups of other top ten Wikipedias.

The other difference between social behaviour in case of users of the Polish Wikipedia and the English one is what we could call the survival patterns of the core users. Although in all top ten Wikipedias users entered to the core relatively quick – after less than a half year, the stability of a core group's composition was quite different. In case of the English Wikipedia, users forming the core often left it after a month. In case of the Polish version of the free encyclopaedia, users stayed in the core for three months. But in both cases there exists part of the core formed by users who stay there for years.

There is also a difference between commitment level of authors after leaving the core group. Users of the English

TABLE I. STATISTICS FOR LIFESPANS IN MONTHS FOR ENGLISH WIKIPEDIA

	active	former	combined
min	28	2	2
max	120	40	120
average	75.68	12.73	72.36
median	73	13	73

version of the Wikipedia often leave the project, after leaving the core group. In case of the Polish version, this is not the case. Polish users show strong commitment to the project. After leaving the core, for long period they often revise articles and sometimes come back to the core.

One can argue, that the rotation of Administrators is high enough to prevent from forming closed groups based on acquaintance. But to decide if such claim is well founded, it is necessary to describe the commitment part of behavioural patterns in case of the Administrators of the Wikipedia. The intuition is that, the Administrators are a group of some of the most active users of the Wikipedia and will show behavioural patterns similar to members of the core users group. But the question is whether it will be more similar to the English core group or to the Polish.

In order to answer above-mentioned question, we analysed list of Administrators of the English Wikipedia [19] and list of Administrators of the Polish Wikipedia [14], [15] and gathered statistics for lifespans of Administrators for mentioned Wikipedias. In our research, *lifespan* of administrator is the length of period, in which user has been granted administrator privileges. If user resigned and was later appointed for being administrator one more time, then his lifespan is period from first election to the last resignation.

For each Wikipedia, we selected three sets of Administrators: all active Administrators, all former Administrators and a combined set of all Administrators (active and former ones). Within each set we calculated lifespans of its members. For obtained lifespans we calculated: minimum, maximum, average and median. Lifespans and statistics relate to events up to the end of the year 2011. Lifespans of Administrators, who were active after the year 2011, were calculated to the 1st of January 2012. As the unit of lifespan we have chosen one month. In case if there were some remaining days, lifespan was rounded to the nearest month. It is worth noting, that in all cases average lifespan and median are almost equal to each other. Because of that, we will analyse average lifespans keeping in mind, that our conclusions are also backed up by medians.

Statistics of lifespans of Administrators of the English Wikipedia are presented in Table I. It is worth noting, that there is large difference between lifespans of active and former Administrators. Active Administrators are active for at least 2 years. They remain active for about six years on average. The maximum length of practice in the case of an active Administrator is ten years. On the other hand, most former Administrators are active for about a year on average. Former Administrator with the longest service was on duty for more than three years. Because of such short average lifespan of former Administrators, one could not infer about possibility of forming strong groups based on acquaintance. The other fact, which is noteworthy, is that the group of Administrators

TABLE II. STATISTICS FOR TIME OF SENIORITY IN MONTHS AFTER BEING FREED FROM ADMINISTRATORS DUTY FOR ENGLISH WIKIPEDIA

	time of seniority
min	0
max	120
average	50.36
median	48

TABLE III. STATISTICS FOR LIFESPANS IN MONTHS FOR POLISH WIKIPEDIA

	active	former	combined
min	7	1	1
max	79	75	79
average	46.57	38.96	44.08
median	45	40	44

of the English Wikipedia yields similar behavioural patterns to the its core.

In Table II statistics for time of seniority in months after being freed from administrators duty for English Wikipedia are presented. Data for Table II was obtained from [20]. There are some cases when administrator lost his privileges and left the project. But on average, after quitting administrators group, users tend to support the free encyclopaedia. And it is substantial support as a median of time spent in project after quitting administrator role is about four years. It is noteworthy difference from the core users described by Ortega et al. [5] and should receive more attention in future studies.

Table III presents statistics for lifespans Administrators of Polish Wikipedia. The first thing to note is, that in case of the Polish Wikipedia maximum and average lifespans are quite similar for active and former Administrators. The longest service of active Administrator is more than six years. And it is also length of the longest service of former one. Both, active and former Administrators will be on duty for three to four years on average. Is is a period which length is comparable to length of the whole examined period (years 2005–2006). This, in contrast to the English Wikipedia, leads to concern if there is possibility that some kind of "elite" in the Administrator community of the Polish Wikipedia is forming. Another fact is that, the Administrator community of the Polish Wikipedia shows similar behavioural patterns to the core of Polish Wikipedia and is quite different from the Administrators community of the English Wikipedia. Unfortunately, due to the time constraints, we were unable to ensemble table with statistics for time of seniority in months after being freed from administrators duty for Polish Wikipedia. It is our priority to gather data and compare it to corresponding data for the English version.

B. Base statistics

In order to compare graphs resulting from intersecting the discussion dimension graph and votes nets for each year, base statistics were obtained for edges' weights. The used measures were: minimum, maximum, arithmetic mean, median, first and third quartile.

For each graph, i.e., discussion network, discussion intersected with positive and discussion intersected with negative votes in each year, empirical distribution functions were calculated. Distribution graphs for examined years are presented in Figure 10 and Figure 11. Values of x-axis are logarithms

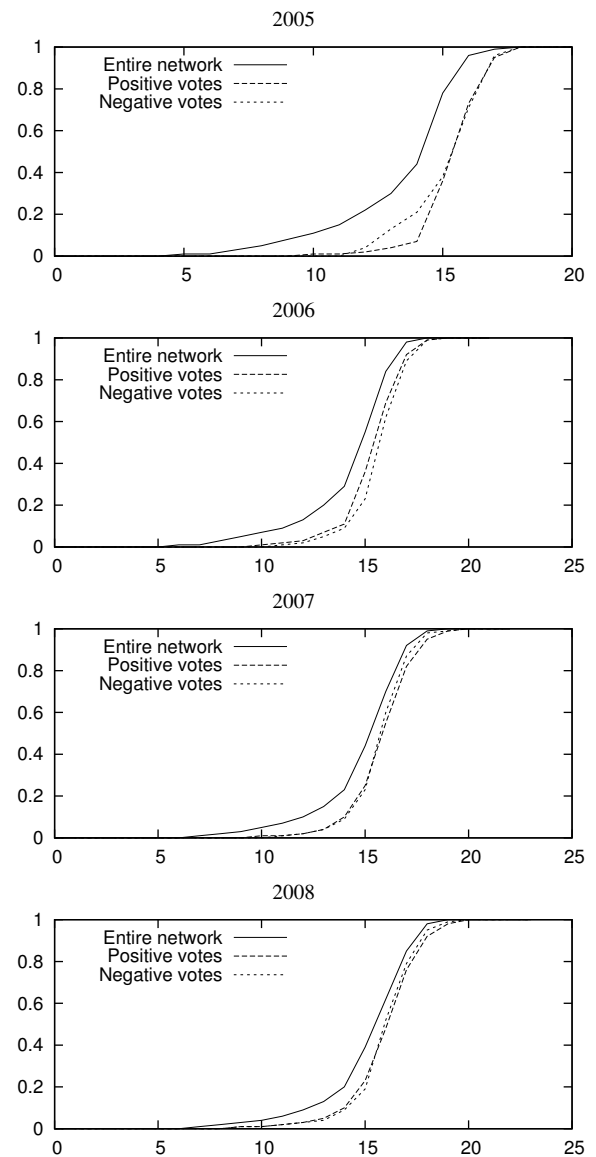


Fig. 10. Distributions of discussion network edge strengths

of edge strengths. Since 2007, the distribution of data is analogous to that described in the article. Both arithmetic mean and median are significantly higher for positive votes. As a result, it can be concluded that the Wikipedia user community is developing steadily.

The stability of the development of the Wikipedia administrator community is also reflected by the empirical cumulative distribution charts. The shapes of the curves are similar, so it can be concluded that the probability distributions describing different parts of the Wikipedia user community originate in the same distribution family. This means that the behaviour of the voters is not subject to sudden changes, but at most it undergoes a calm evolution.

C. Clustering coefficients

Clustering coefficient is a measure of degree to which nodes in a graph tend to be clustered together. The global version, which is used in this article, was designed to give an

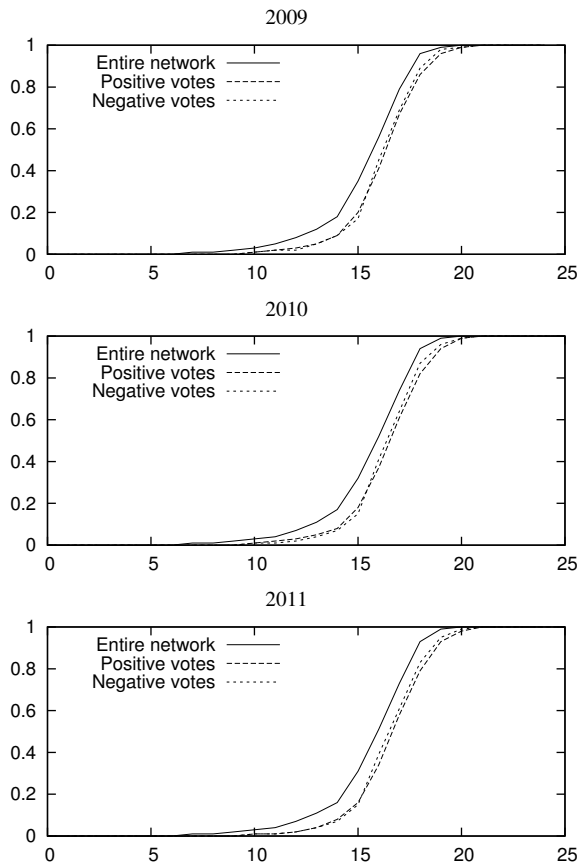


Fig. 11. Distributions of discussion network edge strengths

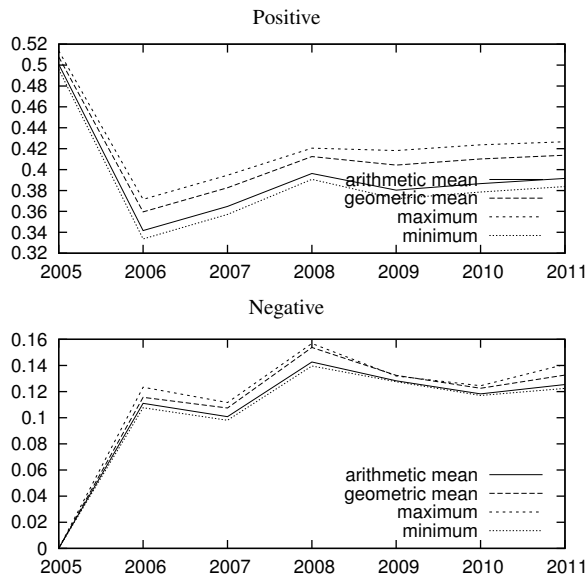


Fig. 12. Clustering coefficients

overall indication of the clustering in the network. Basically, for undirected graphs, it is a ratio between number of closed triplets (three nodes connected by two links) to number of all triplets (three nodes connected by either two or three links).

For directed, weighted graphs a generalization was pro-

posed, it is described in detail Opsahl et al. [21]. Opsahl et al. proposed four measures to calculate triplet value:

- Minimum of edges' weights (mi),
- Maximum of edges' weights (ma),
- Arithmetic mean of edges' weights (am), and
- Geometric mean of edges' weights (gm).

The intuition is as follows: the minimum version is used to find the weakest group in graph, the maximum to find the strongest. Both means give an indication of the strength of ordinary clusters. Opsahl et al. [21], [22] also created tnet library [23] for R software [24].

For each year from 2005 to 2011, the clustering coefficients were obtained for intersections of acquaintance networks with graphs of positive and negative votes. Those coefficients are presented in Figure 12. There are four values (calculated for each of the measures mentioned before) for both graphs.

A few facts can be observed. The first is that there are no very weak or strong groups in Polish Wikipedia society. There is no "elite", which governs RfA procedure or has taken over the administrator society and has power to rule Polish Wikipedia.

The second fact is that clustering coefficients are relatively low and their growth rate is low and negligible. We argue, that decrease in successful administrator elections is not a result of a building up acquaintance relation. Voters do not cast positive votes for their acquaintances or cast negative votes for strangers. The anomaly in year 2005, that clustering coefficients have abnormal values, is most likely caused by the fact, that data for year 2005 were not complete.

V. CONCLUSIONS AND FUTURE WORK

This paper presented the analysis of the development dynamics of the community of administrators of Polish Wikipedia. We have used multidimensional behavioural social networks as a tool to model relationships between wikipedians. The aforementioned analysis included examination of the community in each year from 2005 to 2011 as well as the analysis of the social network corresponding to the final state of the community. The analysis was based on the data from public Wikipedia data dumps.

The fundamental question which we sought the answer to was: "Is the administrator community of the Polish Wikipedia closing up?" It turns out that the answer is not straightforward and it depends on what aspects of the problem one put the greater emphasis, or how to define the "closing up" society.

The conducted analysis of the social network allows us to draw conclusions about the impact of the social system on the nominations of the new administrators. The results of this analysis clearly show that this phenomenon does not exist in the Polish Wikipedia. This is one of the arguments for the statement that the community of the administrators is not shutting out candidates. The administrator community is open to new members in the same way as it was in the beginning of the Polish-language Wikipedia.

However, the pace of growth of the administrator community is lower than it could be expected in case of a young and dynamically growing society. In the early years of development the number of votings was much higher than in the recent years and the number of new administrator appointments strongly declined. That could indicate, however, that the community is closing up after all.

Slower pace of growth and acceptance of new members can be caused by various factors. One such factor may be higher entrance requirements for candidates. Both administrators and regular editors of Wikipedia continue to develop and gain experience in new areas. At the same time, the history of their activity is freely available. For that reason, new users may have trouble with showing equally high achievements and contribution to Wikipedia development. This can be interpreted as closing up of the community by making prohibitive requirements for the new candidates, or as a kind of professionalization aiming to increase the substantive level of the Polish Wikipedia.

It is also possible, that potential candidates do not request for administrator's privileges or withdraw their own candidature from voting because they assess their achievements and contribution to Wikipedia development as not high enough to succeed as an administrator. That would lead to decreasing number of new candidates and appointed new administrators. In our opinion, such case is extremely hard to distinguish from other ones. The best thing to do for the Wikimedia foundation is to do social campaign among users of Wikipedia promoting administrator's duties and boosting confidence in users. This could lead to weakening self-censure among candidates.

Our conclusion is that it cannot be claimed with certainty that the Polish Wikipedia community is shutting off new candidates for administrators. We believe that the increase in the requirements of the current administrator community and users entitled to speak during RfA process toward administrator candidates stems from the community's desire to raise the quality and ensure maximum involvement of all the administrators in the development of Wikipedia.

Our research leads to another interesting conclusion. In Section IV, we have shown the statistics concerning lifespans of the Administrators of the Polish Wikipedia and the English one. Conclusions about the Administrator communities, which can be drawn from above-mentioned statistics, are coherent with ones presented by Ortega et al. [5] about core users of Wikipedias. Namely, users of the Polish Wikipedia will support the free encyclopaedia longer than their English counterparts. It is promising research direction to check if sub-communities of Wikipedias' communities are in some sense fractal. To check, if all sub-groups of users share similar behavioural patterns determined by language version of the project.

The results presented in this study describe the community of Polish Wikipedia administrators only partially. Further research should focus on the detection of new relations between the users and social networks associated with them. It is important to find methods that will allow the development of community to be automatically analysed on the basis of widely available data. The multidimensional behavioural social networks seem to be an ideal tool for this purpose. Richer description of the community could help predict the direction

of its development, which may result in the early identification of threats. This will give the opportunity to counteract those threats and ensure the correct development of Wikipedia.

User community of Polish Wikipedia—in contrast to other language versions—is relatively little known and researched, although, it is an ideal subject for researchers dealing with social informatics. It can be an interesting subject for two types of research: new research, previously not conducted on such a social group, and repeated research, taken from a different version of Wikipedia and performed on the Polish version in order to compare the results and draw conclusions on the development of the latter in comparison to other versions.

Tools used to create multidimensional behavioural social networks for Polish Wikipedia were unable to create such graph for larger instances, e.g., English one. In order to conduct comparative research, scalability problems should be addressed. There is also possibility, that more scalable algorithms can be made on-line. This can allow development of on-line recommendation algorithm for RfA votings.

ACKNOWLEDGEMENT

Authors of this paper thank the team led by dr. Adam Wierzbicki from Polish-Japanese Institute of Information Technology in Warsaw for sharing processed Polish Wikipedia data and multidimensional behavioural social network of Wikipedia users.

The second author was supported by the Polish National Science Centre grant DEC-2011/01/B/ST6/03867.

REFERENCES

- [1] J. Sychała, P. Turek, and M. Adameczyk, "Does the acquaintance relation close up the administrator community of Polish Wikipedia? analysing Polish Wikipedia administrator community with use of multidimensional behavioural social network," in *COLLA 2013, The Third International Conference on Advanced Collaborative Networks, Systems and Applications*, 2013, pp. 28–33.
- [2] (2014, 05) The third international conference on advanced collaborative networks, systems and applications (colla 2013). [Online]. Available: <http://www.ariajournals.org/conferences2013/COLLA13.html> [Accessed: 20.05.2014]
- [3] P. Turek, J. Sychała, A. Wierzbicki, and P. Gackowski, "Social mechanism of granting trust basing on Polish Wikipedia requests for adminship," in *SocInfo*, ser. Lecture Notes in Computer Science, A. Datta, S. W. Shulman, B. Zheng, S.-D. Lin, A. Sun, and E.-P. Lim, Eds., vol. 6984. Springer, 2011, pp. 212–225.
- [4] A. Kittur, B. A. Pendleton, B. Suh, and T. Mytkowicz, "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie," *World Wide Web*, vol. 1, no. 2, p. 19, 2007.
- [5] J. F. Ortega Soto, "Wikipedia: A quantitative analysis," 2012.
- [6] M. Jankowski-Lorek, L. Ostrowski, P. Turek, and A. Wierzbicki, "Modeling Wikipedia admin elections using multidimensional behavioral social networks," *Social Network Analysis and Mining*, pp. 1–15, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s13278-012-0092-6> [Accessed: 15.05.2013]
- [7] M. Burke and R. Kraut, "Taking up the mop: identifying future Wikipedia administrators," in *CHI Extended Abstracts*, M. Czerwinski, A. M. Lund, and D. S. Tan, Eds. ACM, 2008, pp. 3441–3446.
- [8] (2014, 05) Wikipedia: Guide to requests for adminship. [Online]. Available: http://en.wikipedia.org/wiki/Wikipedia:Guide_to_requests_for_adminship [Accessed: 20.05.2014]
- [9] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi, "He says, she says: conflict and coordination in Wikipedia," in *CHI*, M. B. Rosson and D. J. Gilmore, Eds. ACM, 2007, pp. 453–462.

- [10] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Governance in social media: A case study of the wikipedia promotion process," in *ICWSM*, W. W. Cohen and S. Gosling, Eds. The AAAI Press, 2010.
- [11] A. Anderson, D. P. Huttenlocher, J. M. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," in *WSDM*, E. Adar, J. Teevan, E. Agichtein, and Y. Maarek, Eds. ACM, 2012, pp. 703–712.
- [12] H. S. Rad, A. Makazhanov, D. Rafiei, and D. Barbosa, "Leveraging editor collaboration patterns in wikipedia," in *HT*, E. V. Munson and M. Strohmaier, Eds. ACM, 2012, pp. 13–22.
- [13] (2014, 05) Wikipedia statistics polish. [Online]. Available: <http://stats.wikimedia.org/EN/TablesWikipediaPL.htm> [Accessed: 20.05.2014]
- [14] (2014, 05) Wikipedia:administratorzy. [Online]. Available: <http://pl.wikipedia.org/wiki/Wikipedia:Administratorzy> [Accessed: 20.05.2014]
- [15] (2014, 05) Wikipedia:pryzyznawanie uprawnień/przyjete. [Online]. Available: http://pl.wikipedia.org/wiki/Wikipedia:Przyznawanie_uprawnien%C5%84/Przyj%C4%99te [Accessed: 20.05.2014]
- [16] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli, "The singularity is not near: slowing growth of wikipedia," in *Int. Sym. Wikis*, D. Riehle and A. Bruckman, Eds. ACM, 2009.
- [17] S. Maniu, B. Cautis, and T. Abdessalem, "Building a signed network from interactions in wikipedia," in *DBSocial*, D. Barbosa, G. Miklau, and C. Yu, Eds. ACM, 2011, pp. 19–24.
- [18] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Signed networks in social media," in *CHI*, E. D. Mynatt, D. Schoner, G. Fitzpatrick, S. E. Hudson, W. K. Edwards, and T. Rodden, Eds. ACM, 2010, pp. 1361–1370.
- [19] (2014, 05) User:noseptember/list of administrators. [Online]. Available: http://en.wikipedia.org/wiki/User:NoSeptember/List_of_Administrators [Accessed: 20.05.2014]
- [20] (2014, 05) Wikipedia:missing wikipedians. [Online]. Available: http://en.wikipedia.org/wiki/Wikipedia:Missing_Wikipedians [Accessed: 20.05.2014]
- [21] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155–163, May 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.socnet.2009.02.002> [Accessed: 15.05.2013]
- [22] T. Opsahl, *Structure and Evolution of Weighted Networks*. University of London (Queen Mary College), London, UK, 2009. [Online]. Available: <http://toreopsahl.com/publications/thesis/> [Accessed: 15.05.2013]
- [23] (2013, 05) tnet: Analysis of weighted, two mode, and longitudinal networks. [Online]. Available: <http://toreopsahl.com/tnet/> [Accessed: 15.05.2013]
- [24] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org> [Accessed: 15.05.2013]