When the Wikipedians Talk: Network and Tree Structure of Wikipedia Discussion Pages

David Laniado[‡,*] **Riccardo Tasso**[*] **Yana Volkovich**[‡] **Andreas Kaltenbrunner**[‡]

[*] Politecnico di Milano
Dipartimento di Elettronica e Informazione
{david.laniado,tasso}@elet.polimi.it

[‡] Barcelona Media – Innovation Centre
Information, Technology and Society Group
{firstname.lastname}@barcelonamedia.org

Abstract

Talk pages play a fundamental role in Wikipedia as the place for discussion and communication. In this work we use the comments on these pages to extract and study three networks, corresponding to different kinds of interactions. We find evidence of a specific assortativity profile which differentiates article discussions from personal conversations. An analysis of the tree structure of the article talk pages allows to capture patterns of interaction, and reveals structural differences among the discussions about articles from different semantic areas.

Introduction

Wikipedia is the largest example of collaboration on the Web, accessed and edited each day by thousands of people. Behind the most visible part of Wikipedia, i.e. the articles, there are non-encyclopedic pages which are used for coordination, discussion and personal communication among the Wikipedians. While the growth of the encyclopedia in terms of numbers of articles, edits and active users has slowed down in the last years, activity on these pages has kept increasing at a higher rate (Suh et al. 2009). In this study we focus on this less visible side of Wikipedia, in order to shed light on communication patterns that accompany collaboration on the project.

Unlike other online discussions which often only satisfy the purpose of entertainment or of defending one's point of view, the discussion on Wikipedia article talk pages has a clear objective, i.e. to reach consensus and improve the content of the corresponding article. In many cases these pages can considerably outgrow the corresponding article in size. For example, the talk page associated to the article 'Barack Obama' contains more than 22 000 comments, which is more than the 17 500 edits done to the article itself. In Wikipedia there are also talk pages associated to registered users; these pages are somehow complementary to the article discussion pages, and are used for personal communication between the Wikipedians, as a sort of public in-box.

Communications in Wikipedia are part of a complex social system, where users are involved in the project to different extents and with different roles, either explicit or implicit. Several studies have focused on the analysis of the content of talk pages (Viégas et al. 2007; Stvilia et al. 2008; Schneider, Passant, and Breslin 2010), while some researchers have studied the correlation between the presence of discussion and article quality (Kittur et al. 2007; Kittur and Kraut 2008). Kittur et al. (2007) also identify the number of edits done to a discussion page as the best indicator of conflict on the corresponding article. Though, little attention has so far been devoted to the study of interaction patterns emerging from discussions on these pages. We claim that the study of these interactions on a large scale can reveal essential features of the Wikipedia community and its social structure.

In this paper we offer an extensive analysis of talk pages associated to articles and to users. We analyse the structural properties of the networks derived from interactions on these pages; in particular, the study of directed degree assortativity allows us to reveal specific patterns in the communications between the Wikipedians, which differ from the results obtained for the discussion board Slashdot.

To characterize the discussions on article talk pages, we analyse their structure according to different measures, such as depth and size of the discussion threads. We show how chains of direct replies between pairs of users can be an interesting indicator of particularly contentious topics, and we report a listing of the most discussed articles, according to different criteria. Finally, we investigate the relationship between structural properties of the discussions and the corresponding semantic areas.

Experimental Setup

From a technical point of view talk pages are simple wiki pages; however, their usage has evolved over the years to fit the community needs, and the wiki text syntax has been exploited to create a forum-like environment.

The freedom which is left to editors in the usage of discussion pages was a challenge for our analysis. There is no structure surrounding a single comment nor an always valid schema to detect its start and end. Moreover, signing comments is left to users, who can use a shortcut to add the signature containing a link to their personal page at the end of a post; for anonymous (not registered) users the signature reports their IP number. Though there are bots in charge of automatically adding missing signatures, many comments are

#articles	3 210 039	
#edits of article pages	402 851 686	
#articles with talk page (ATP)	871 485	(27.1%)
#total comments in ATP	11 041 246	
#signed comments in ATP	9 421 976	(85.3%)
#anonymous (ip signed) comments in ATP	1 000 824	(9.1%)
#users who comment articles	350 958	(2.8%)
#registered users	12 651 636	
#user talk pages (UTP)	1 662 818	(13.1%)
#comments in UTP	13 670 980	
#signed comments in UTP	13 493 254	(98.7%)
#anonymous (ip signed) comments in UTP	2 009 658	(14.7%)

Table 1: Basic quantities of the data analysed.

unsigned. To extract the thread structure with comment indentation, signatures and dates, we had to deal with many different explicit and implicit conventions, changing over years and not always attended by the users. Sometimes users reset indentation; we always consider these cases as the start of a new thread in the discussion.

For this study we relied on a complete dump of the English Wikipedia dated March 12th, 2010. In Table 1 we report some basic quantities of the data extracted.

Wikipedia discussion networks

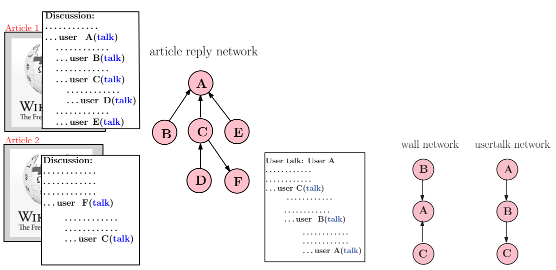
There are no explicit networks between users in Wikipedia. In order to study the patterns of communication and discussion, we extracted three implicit directed networks according to different types of interactions between users:

Article reply network (reply-NW) direct replies between users in article discussion pages.

User talk network (talk-NW) direct replies in user talk pages.

Wall network (wall-NW) personal messages posted on the talk page of another user.

In all networks we discard anonymous users, as IP numbers are not reliable identifiers. In Figure 1 we schematically explain the idea of how these networks are constructed. In the article reply network (Figure 1(a)) we establish a directed edge from a user B to a user A if B has written at least one comment indented under an entry by user A in any article discussion page. The user talk network (Figure 1(b)) is analogously defined, but based on the comments in user talk pages, while the wall network establishes a link from user B



(a) Article reply NW (b) User talk and wall NWs

Figure 1: Schema of the networks construction.

variable	reply-NW	talk-NW	wall-NW
#nodes with edges	204 017	114 258	1 861 702
w. in-degree ≥1	121 682	103 147	1 832 168
w. out-degree≥1	182 881	63 334	177 331
#edges M	1 489 734	852 065	4 412 212
size of giant comp.	88.5%	89.2%	96.3%
mean distance	4.10 (0.75)	3.86 (0.69)	4.06 (0.68)
maximal distance	15	11	12
Clustering coeff.	0.083 (0.19)	0.053 (0.16)	0.035 (0.14)
mean in-degree	7.30 (29.6)	7.46 (32.8)	2.37 (15.75)
mean out-degree	7.30 (35.2)	7.46 (41.5)	2.37 (103.79)
network density	$3.58 \cdot 10^{-5}$	$6.53 \cdot 10^{-5}$	$1.27 \cdot 10^{-6}$
reciprocity	0.44	0.45	0.15

Table 2: Global measures of the Wikipedia discussion and talk network. Values within parenthesis indicate stdv.

to user A if user B has written something on the talk page of user A.

Basic network parameters

In Table 2 we report some macroscopic features of the networks. Besides the dimension in terms of number of nodes, we report for each network the number of nodes having at least one outgoing or one in-going link, respectively. Interestingly, these quantities vary significantly from network to network. In the article discussions around 90% of users have replied to at least one user, while nearly 60% have received replies. On the contrary, in the other two networks almost all (wall: 98.4%, talk: 90.3%) users have at least one incoming link, while many do not have any outgoing link. In particular, only less than one over ten users in the wall network have written on another user's talk page. This result is due to the presence of *welcomers*, users and bots who write a welcome message on the wall of newly registered users; this also explains the larger size of the wall network, which contains many users who are not active. For this reason also reciprocity is lower in the wall network.

Network comparison

For the network comparison we modify the metric proposed by Szell, Lambiotte, and Thurner (2010), based on Jaccard coefficient of the link overlap, such that it takes values within $[0, 1]$ and it is independent of the network densities. It measures the co-occurrence of links between users in different social networks. More formally, let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be two networks with the same set of nodes V, and with the sets of edges E_1 and E_2, respectively. Then,

$$C_{jaccard} = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|} \cdot \frac{\max(|E_1|, |E_2|)}{\min(|E_1|, |E_2|)},$$

where we denote as $|\cdot|$ the number of elements in the set.

	reply-NW	talk-NW	wall-NW
reply-NW	1	0.11	0.09
talk-NW	0.11	1	0.35
wall-NW	0.09	0.35	1

Table 3: Jaccard coefficient between the networks.

	type	r	$\langle r_{rand} \rangle$	σ_{rand}	Z	ASP
Slashdot	(out, in)	-0.035	-0.046	0.00059	**17.677**	0.329
	(in, out)	-0.016	-0.033	0.00063	**26.613**	0.495
	(out, out)	-0.015	-0.038	0.00063	**35.843**	0.667
	(in, in)	-0.027	-0.040	0.00057	**24.143**	0.449
Reply	(out, in)	-0.025	-0.019	0.00063	**-8.629**	-0.485
	(in, out)	-0.018	-0.018	0.00061	0.062	0.003
	(out, out)	-0.027	-0.018	0.00062	**-14.179**	-0.797
	(in, in)	-0.015	-0.019	0.00063	**6.385**	0.359
Talk	(out, in)	-0.045	-0.030	0.00998	-1.526	-0.655
	(in, out)	-0.025	-0.026	0.00753	0.109	0.047
	(out, out)	-0.042	-0.028	0.00848	-1.753	-0.753
	(in, in)	-0.028	-0.029	0.00894	0.076	0.033
Wall	(out, in)	-0.126	-0.087	5.1e-5	**-769.81**	-0.936
	(in, out)	-0.039	-0.020	0.00020	**-93.51**	-0.114
	(out, out)	-0.063	-0.043	7.5e-5	**-26.04**	-0.317
	(in, in)	-0.061	-0.039	0.00026	**-84.21**	-0.102

Table 4: Directed assortativity results for the three networks of Wikipedians and for the Slashdot reply network. Values in bold are significant ($|Z > 2|$).

We compare the three networks and present the results in Table 3. In the wall network, we discarded all users who are not present in any of the two reply networks, to keep only active users. As it could be expected, the highest overlap is between the two networks extracted from user talk pages. Though, it is important to point out that these two networks capture different kinds of interaction, and none of the two is subsumed by the other. The overlap between edges in these networks of personal communications and the one extracted from articles is of about 10%, indicating substantially different networks.

Directed assortativity by degree

Assortativity by degree is a basic measure of diversity in networks, quantifying the tendency of nodes to link with other having similar number of edges (Newman 2002). This measure has been widely used to analyse various kinds of networks, and assortative mixing has been shown to be a characterising feature of social networks, with respect to technological and biological networks, that are mostly dissortative (Newman and Park 2003). On the contrary, recent studies have pointed out that many online social networks tend to dissortativity (Hu and Wang 2009).

To compute degree assortativity accounting for the direction of edges, we rely on the Assortativity Significance Profile (ASP) proposed in (Foster et al. 2010). Combining the degree types (in- and out-) of the source and target nodes we obtain a set of four assortativity measures; we denote as $r(out, in)$ the correlation between the out-degree of the source and the in-degree of the target node of each connection, and so on. For details about these metrics we refer to (Foster et al. 2010). To assess statistical significance of the results, we contrast each network with an ensemble of 100 randomly generated equivalents, having the same in- and out-degree sequences.

Results are reported in Table 4, together with the results for the reply network extracted from the Slashdot discussion board. We added these results to be able to compare discussions in Wikipedia with discussions from another large online community. The Slashdot reply network contains about 80 000 users and 1 million connections; for a detailed description of this dataset, see (Gómez, Kaltenbrunner, and López 2008).

None of the assortativity values computed for the talk network is statistically significant ($|Z > 2|$). The wall network exhibits dissortativity according to all four measures, which points a general tendency of socially active users to interact preferentially with users having few connections. In particular, the remarkably high value observed for the (out, in)-assortativity shall be imputed to the activity of users and bots who massively welcome new registered users writing on their personal talk page.

The reply network extracted from Wikipedia articles shows to be (out, out)- and (out, in)- dissortative, with significant Z-scores, pointing out a marked tendency of users having many outgoing links to interact preferentially with users having few connections, and vice versa. On the contrary, (in, in) assortativity is positive, revealing a tendency of users to reply more often to others having a similar in-degree. We do not observe this pattern in the Slashdot reply network, which is assortative according to all four measures[1]. The difference could be due to the peculiar nature of Wikipedia article talk pages, where discussions are usually aimed at taking decisions about content production according to the community policies. While a high out-degree is the result of an active behaviour, replying to many users, a high in-degree is achieved getting many replies from different users. These two measures seem to capture two distinct characteristics of Wikipedia influential users, resonating with a distinction between *hubs* and *authorities*. The Wikipedians who reply to many other users in article talk pages tend to interact mostly with users having few connections, i.e. newbies and inexperienced users, while the Wikipedians who receive replies from many users tend to interact preferentially with each other.

The discussion trees

In this section we focus on the shape and size of interactions in the discussion pages on Wikipedia. These interactions can be modelled in the form of discussion trees, where the root node corresponds to the article page on Wikipedia, and child nodes to comments or structural elements of the discussion pages. Unlike other online discussions, for example observed in blogs (Mishne and Glance 2006) or at Slashdot (Gómez, Kaltenbrunner, and López 2008), the Wikipedia discussion pages do not only consist of comments, which represent the actual interactions between the users, but may also contain many structural elements such as a separation of the total number of the comments into several sub-pages, or titles and subtitles to organize the content of the discussion. We model each of these different elements as a separate node in the discussion tree. The structure of the tree reflects

[1]Note that this is different from what has been reported previously in (Gómez, Kaltenbrunner, and López 2008), where no comparison with randomised networks was taken into account.

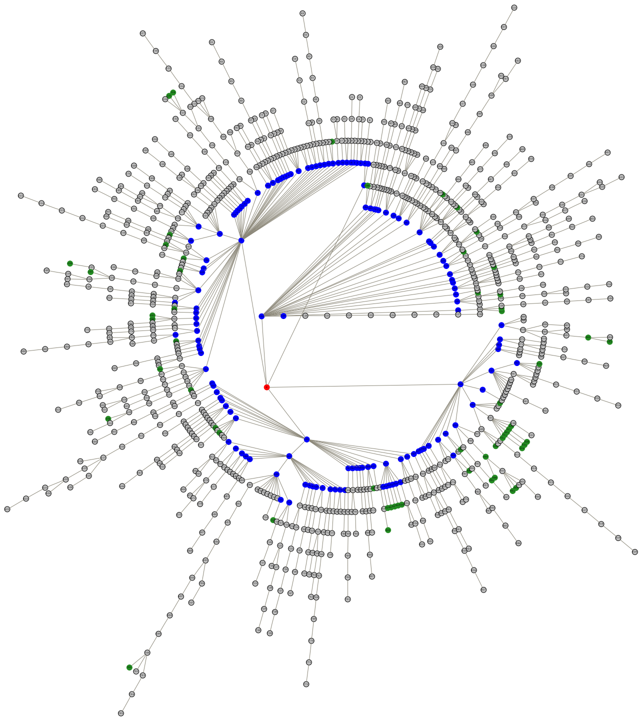


Figure 2: The structure of the discussion page of "Presidency of Barack Obama" (generated with Graphviz). Blue nodes are structural, green nodes are unsigned comments.

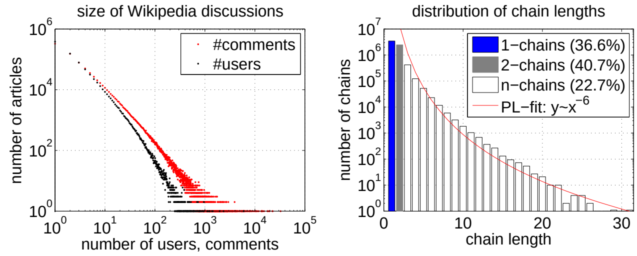


Figure 3: Distributions of the number of **(left)** comments and users per discussion page, and **(right)** discussion chains of different lengths in the dataset. Percentages indicate the proportion of comments in the different types of chains.

the hierarchy of the pages. A reply to a comment is a child node of this comment and comments which are placed below a title or a new page are child nodes of the corresponding structural node unless they reply to another comment. Note that there can be several nested levels of structural nodes as there can be several levels of titles and subtitles.

To help in the comprehension of the following analysis we show in Figure 2 one of these trees. It corresponds to the Wikipedia article "Presidency of Barack Obama" (represented by the red node in the centre of the radial tree) and contains 989 nodes of which 254 are structural nodes (in blue in Figure 2) and the rest comments. Note that this article is different from one just on "Barack Obama" cited earlier.

The size of the discussions

Out of the approx. 3.2 million articles in our dataset nearly 870 000 have an associated discussion page (about 27%), which contain more than 9.4 million signed comments, created by more than 350 000 users (See Table 1 for details). As one would expect the distribution of the number of comments and users among the different articles follows heavy tailed distributions as shown in Figure 3 (left).

Although more than 85% of all articles have discussions with only 10 or less comments, there is still a considerable number of articles (approx. 15 000) with more than 100 comments and 826 discussions even contain more than 1000 comments. The largest discussions reach more than 30 000 comments involving several thousand users.

What are the shapes of these discussions? The example of Figure 2 suggest that we can basically identify two patterns: comments that are placed directly after a structural node (a headline etc.) and do not receive any replies; and large chain-like subthreads of comments, containing a sequence of replies between several users. Only occasionally a comment receives more than just one reply in these subthreads, which contain about 65.4% of all comments. The remaining 36.6% of the comments correspond to isolated unanswered comments who themselves are also not replying to another comment.

To investigate those subthreads further we focus on the number of such chain-like subthreads that can be found per discussions and on their lengths. To formalise the concept of chains we consider only subthreads where exactly two users interact subsequently. We define as n-chains (or simply chains, if not stated otherwise) all sequences which include at least three comments. So the shortest n-chains are of the form $A \leftarrow B \leftarrow A$ where A and B are two different users and the arrows indicate a reply of B to A and a back-reply from A to B. These chains can grow considerably. The longest chain in our dataset is of length 31 in the discussion page of "Central Bosnia Canton"[2]. Figure 3 (right) shows the number of chains of different lengths. Given a chain of length k, we do not count any sub-part of it as a chain. We find that 22.7% of all comments form part of chains of length of at least 3 (n-chains), while 40.7% belong only to 2-chains (are either parent or reply but not part of an n-chain). The remaining comments are isolated (1-chains). The distribution of the number of n-chains with different lengths follows roughly a power-law with exponent 6 (red line in Figure 3).

In Figure 4 we show the distribution of the number of n-chains in the discussion pages. Again we find a heavy tailed distribution, with some discussions containing several thousand chains. The distribution can be fitted with both, a power law distribution (with cut-off) with exponent 2.23 and a truncated log-normal distribution. Both fits are not rejected by a Kolmogorov Smirnov test (see figure legend for the corresponding p-values).

The number of chains gives us an idea about how many times a controversy arises in the article discussion. In Table 5

[2] See http://en.wikipedia.org/wiki/Talk:Central_Bosnia_Canton

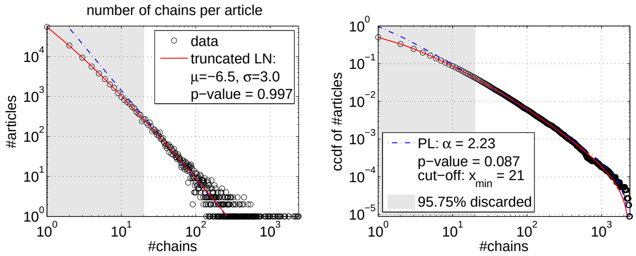


Figure 4: Number of discussion chains per discussion page.

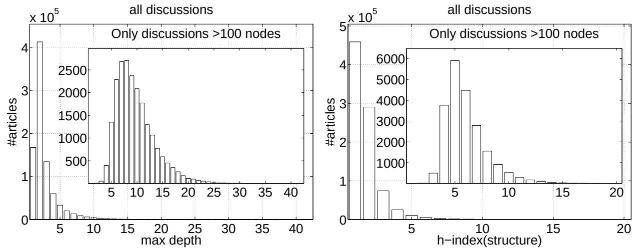


Figure 5: Distribution of the max. depth (left) and the h-index (right) of the article discussion pages.

we list the top 20 articles according to this measure and compare it with the total number of comments, registered users who comment and edits of the corresponding article page. The numbers in parenthesis indicate the rank number of the article when ordering by the corresponding variable. Note that nearly in all cases the number of comments is much larger than the number of actual edits on the corresponding article page. Most of the topics in the list represent also highly disputed subjects in real life, either due to political, ideological, religious or scientific disputes. They seem to be a good barometer of contentious discussion topics in the last few years.

We furthermore list as well the depth of these discussion trees which we will treat separately in the next subsection.

Depths of the discussions

We study the depth of the discussions using two measures: its maximal depth (i.e., the level of the deepest comment in the discussion tree) and its h-index, a balanced depth measure which was introduced in (Gómez, Kaltenbrunner, and López 2008) on the base of the h-index of (Hirsch 2005). To calculate the h-index of the tree structure we count the number of nodes per level, starting at level one (the root node) and descending the tree. The h-index of the tree is then the maximum level, for which the corresponding number of nodes is greater or equal to the level number (and all previous levels fulfil the same condition). Note that we also consider structural nodes in these calculations.

In Figure 5 we show the distributions of the maximal depths and the h-index for all discussion and only for the ones with more than 100 nodes (in the insets). We observe that the two distributions have a similar shape but slightly different modes.

The deepest discussion can be found about the article "Liberal democracy"[3]. It reaches a depth of 42, while its h-index is only 12. The maximal h-index is observed for "Anarchism" (h-index $= 20$).

In the Columns 6 and 7 of Table 5 we present the depth and h-index of the 20 most discussed articles. From the rank-values within parenthesis we can conclude that the rank by h-index is closer to the rank by number of chains than the rank of the maximal depth of these discussions. The maximal depth is very sensitive to the presence of isolated individual discussions between a small number of users, which can reach considerable depths while not being representative for the entire discussion. The h-index overcomes this limitation and we will use it therefore in the next section to account for the depth of the discussions.

Comparison with Categories

In this section we investigate whether the structure of the article discussions differs for the different topic categories of Wikipedia articles.

Assigning articles to macrocategories Assigning Wikipedia articles to a set of topics is not a trivial task, as each article is usually assigned only to low level categories, which can in turn be associated to many super-categories. Links to categories and super-categories are managed by users inside the wiki text, so the result is a rich but inconsistent hierarchical structure made of more than 500 000 categories, in which one can find several loops.

To deal with this disordered semantic information we relied on the approach proposed by Kittur, Chi, and Suh (2009) to classify articles according to a limited set of top level categories, or macrocategories. The algorithm starts by considering, for each article, all the categories to which it has been directly assigned. Each of these labels is in turn assigned to the closest macrocategory in the category graph. The extent to which an article belongs to each macro-category is computed as a weight, quantifying the proportion of directly assigned categories which belong to that macrocategory. In the case of equally short paths from a label to multiple macrocategories, the contribution of this category is split among the equidistant closest macrocategories.

Kittur, Chi, and Suh computed the article assignments in 2009 based on 11 macrocategories; we ran the same algorithm with 21 macrocategories shown in Figure 6, corresponding (with minor arrangements) to the current official Wikipedia top level categories. Though the category graph is based on directed relationships linking categories to super-categories, Kittur et al. considered it as an undirected graph to compute the shortest paths between each category and the macrocategories. This allows to assign all categories to some macrocategories, while considering the directed graph many categories would remain disconnected. However, our intuition is that link direction in the hierarchy matters. So we corrected the algorithm assigning a higher weight to the edges followed in the wrong direction. Penalising these edges by a factor of 3 brought to a significant

[3]See http://en.wikipedia.org/wiki/Talk:Liberal_democracy/Archive_2#The_prominence_of_the_liberalism_template

#	Title	chains	comments	users	h-index	max. depth	edits
1	Intelligent design	2413	22454 (3)	954 (13)	16 (20)	20 (358)	9179 (53)
2	Gaza War	2358	17961 (6)	607 (47)	19 (2)	27 (28)	11499 (29)
3	Barack Obama	2301	22756 (2)	2360 (2)	18 (6)	21 (245)	17453 (6)
4	Sarah Palin	2182	19634 (4)	1221 (9)	17 (10)	25 (56)	12093 (24)
5	Global warming	2178	19138 (5)	1382 (5)	17 (10)	20 (358)	14074 (15)
6	Main Page	2065	32664 (1)	5969 (1)	15 (34)	22 (169)	4003 (674)
7	Chiropractic	1772	13684 (13)	243 (389)	18 (6)	29 (17)	6190 (204)
8	Race and intelligence	1764	13790 (12)	410 (126)	17 (10)	24 (74)	7615 (100)
9	Anarchism	1589	14385 (9)	496 (76)	20 (1)	28 (22)	12589 (19)
10	British Isles	1556	12044 (16)	576 (56)	17 (10)	23 (113)	4047 (658)
11	Climatic Research Unit hacking incident	1551	11536 (17)	474 (88)	17 (10)	20 (358)	2346 (2364)
12	Jesus	1397	17916 (7)	1239 (7)	13 (119)	16 (1383)	17081 (7)
13	Circumcision	1356	10469 (21)	436 (113)	17 (10)	26 (42)	7354 (117)
14	Homeopathy	1323	13509 (14)	516 (68)	17 (10)	25 (56)	6902 (151)
15	George W. Bush	1281	15257 (8)	1969 (3)	14 (65)	18 (676)	32314 (1)
16	September 11 attacks	1250	13830 (11)	1244 (6)	16 (20)	26 (42)	11086 (30)
17	Evolution	1165	13404 (15)	942 (16)	13 (119)	23 (113)	9780 (44)
18	Catholic Church	1162	14104 (10)	620 (43)	15 (34)	18 (676)	14082 (14)
19	Cold fusion	1098	8354 (29)	359 (174)	15 (34)	20 (358)	4320 (557)
20	2008 South Ossetia war	1075	10596 (20)	853 (20)	17 (10)	23 (113)	9930 (43)

Table 5: Several structural measures of the top 20 Wikipedia discussions ordered by the number of n-chains (length ≥ 3).

improvement in the performance, evaluated over a random sample of 300 manually assigned articles (Farina, Tasso, and Laniado 2010).

Structural differences between the categories We investigate the proportion of pages with discussions among the different categories. We use the category weights for this calculation. So, if an article has a 60% weight in category A and a 40% weight in category B it contributes with the corresponding proportions to these two categories. The black bars in Figure 6 show these general proportions of articles with discussions (the corresponding %-value is written on the right y-axis). We observe a large heterogeneity among the different categories. "Geography and Places" and "History and events" are nearly of the same size and account together for more than 46% of all discussion pages. The next two categories "Culture" and "People" account for another 20%. Interestingly, if we restrict this analysis to only the top 1% or 0.1% of the article discussion pages (according to their number of comments) we observe rather different distributions (indicated by the grey and white bars in Figure 6). The "Geography and Places" proportion decays to a only half of its original value, while some other categories like "Belief", "Society", "Philosophy" or "Law" and "Politics" approximately double their share.

This change seems to indicate that these categories, although less frequent among the entire set of articles with discussions, attract more than an average number of comments. To investigate this further we calculated for every category (using the category weights of every article) the weighted average of several structural metrics presented in the previous subsections. The outcome of this analysis is presented in the form of two cross-plots in Figure 7. To verify the results we performed a bootstrap test ($N = 1000$) and depict the 95% interval of the observed average value with grey areas. In the cases where the area is absent the symbol size is larger

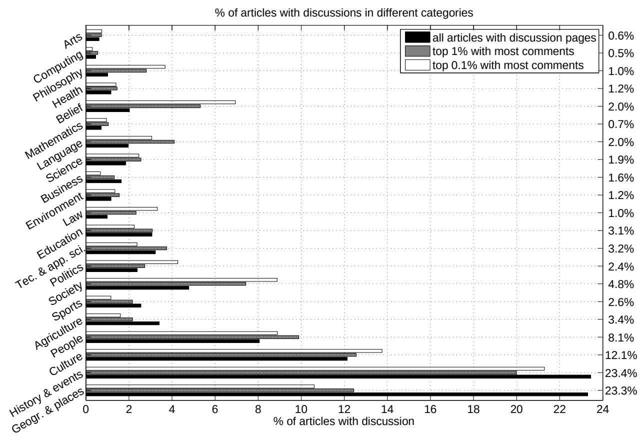


Figure 6: Proportion of articles within different categories for all discussion pages

than the corresponding confidence interval.

We can extract several interesting conclusions from the two cross-plots. From the right cross-plot we see a clear correlation between the average values of the depth of the discussion measured with the h-index and the number of users who left at least one comment in the discussion. As we discussed above the category "Geography and Places" is on average the one with the flattest discussions. On the other hand, the top categories according to these measures are "Philosophy", "Law", "Language" and "Belief". These categories trigger, on average, the deepest discussions involving the largest amount of users. They are also the top 3 categories if we use the number of discussion chains as a measure as can be seen from the left sub-figure of Figure 7 where we compare the number of edits with the number of discussion

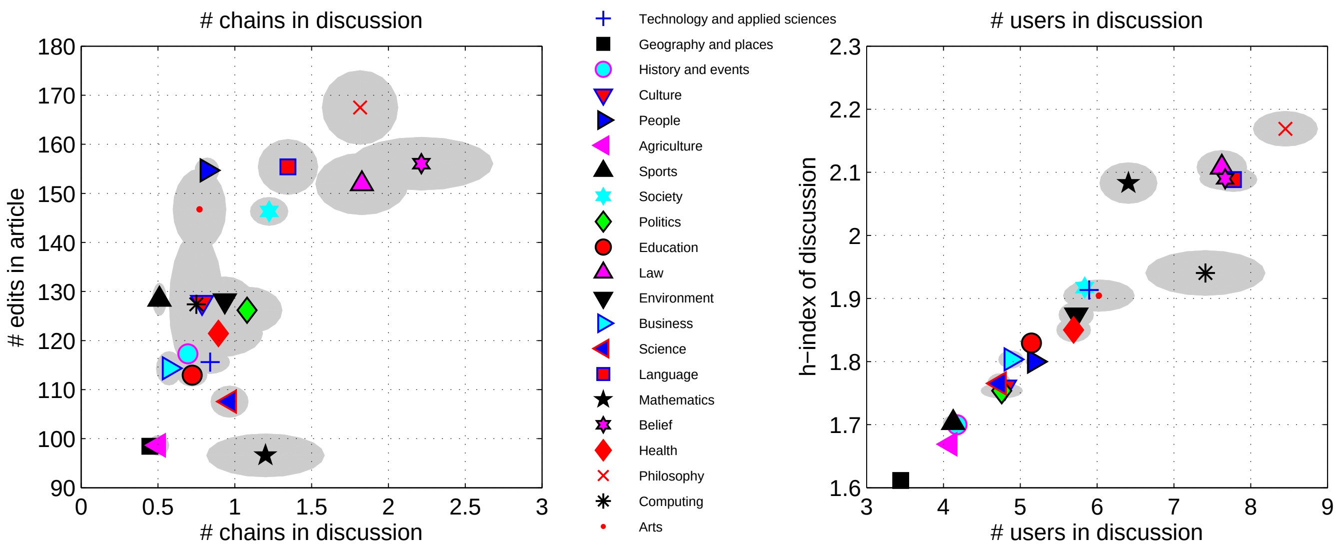


Figure 7: Differences between Categories: (left) the average number of chains in the article discussion pages vs. the average number of article edits, (right) the average number of users versus the h-index of the tree structure. Gray areas indicate (when larger than the corresponding symbol size) the 95% confidence interval.

chains. This seems to agree with (Kittur, Chi, and Suh 2009), where the amount of conflict in different macrocategories is estimated according to a page-level heuristics, and "Belief" and "Philosophy" are identified as the most contentious categories. While the authors find the lowest level of conflict in category "Mathematics", from Figure 7 we can observe that, although this category has the lowest average number of edits per article, it still reaches a considerable amount of debate in the form of discussion chains. Some article categories like "People", "Arts" or even "Sports" are on average less discussed than their number of edits would suggest.

In the right subplot we also observe two outlayers to the otherwise quite correlated averages. The categories "Computing" and "Mathematics" obey a different behaviour. Discussions on "Computing" articles involve more users, than their average h-index would suggest, while articles of the "Mathematics" category have the opposite behaviour. They are deeper but involve less users than expected.

To summarise, we observe quite different relations between the discussion structures and the number of edits among the different categories. We have also found that the size and shape of the discussion varies significantly among the different categories. A more detailed analysis involving the study of individual user activities in the different categories might shed further light on whether these differences are community or content based.

Related Studies

Kittur et al. (2007) describe the growth of the *hidden side* of Wikipedia, comprehending talk pages and all Wikipedia-specific pages deputed to conflict and coordination. A longitudinal study to investigate the role of coordination in the improvement of Wikipedia articles' quality is described in (Kittur and Kraut 2008); positive improvements are observed as effect of discussion only on small and "young" pages. Presence of discussion is just measured in terms of size of the article talk pages, and patterns of communications are not considered.

Few researchers focused on Wikipedia talk pages to study social relationships between users. Crandall et al. (2008) investigate the interplay between social ties, modelled as interactions in discussion pages, and similarity, modelled as editing activity on the same articles. They find evidence of a feedback effect between the two phenomenons. Only a 15% overlap is found between the graph of social interactions and the one of similarity; interestingly, properties of the social network reveal to be better predictors of future behaviour than properties of the similarity network. A qualitative description of different patterns corresponding to different social roles in Wikipedia is offered in (Gleave et al. 2009), based on the local network of personal communications around single users.

Tang, Biuk-Aghai, and Fong (2008) propose a model for the representation of weighted co-authorship relationships, while in (Laniado and Tasso 2011) the main editors of each article are selected as authors in order to build a collaboration network over the whole Wikipedia. More detailed models have been proposed to represent different kinds of interactions in editing activity (Brandes et al. 2009). Both the approaches of extracting co-authorship networks and edit networks are complementary to ours and could be integrated to contrast direct replies in discussion pages with relationships emerging from editing activity.

To the best of our knowledge, in this paper we propose the first extensive study of Wikipedia as a discussion space. Similar analysis have been performed for blogs (Mishne and Glance 2006) and online discussion boards (Gómez, Kaltenbrunner, and López 2008). A generative model for the structure of the discussion threads analysed here has been presented in (Gómez, Kappen, and Kaltenbrunner 2011). The model parameters show important structural differences between the discussions in Wikipedia and those of other social media platforms.

Conclusions

In this paper we have focused on Wikipedia talk pages to detect structural patterns of interaction which accompany collaboration on the project. The study of directed assortativity reveals the existence of a characterizing pattern in the reply network extracted from article discussion pages. Users who reply to many other users tend to reply preferentially to inexperienced users, while the Wikipedians who receive comments by many users are more likely to interact with each other. This pattern is not observed in the Slashdot reply network neither in personal conversations in Wikipedia. We suggest that it derives from the nature of discussion on article talk pages, focused on solving issues and controversies according to codified community policies, and reflects the existence of different social roles among the more influential users.

The study of shape and size of the discussions at the article level reveals interesting patterns and suggests some metrics to characterize different talk pages. The number of chains of direct replies between pairs of users seems to be a good indicator of contentious discussion topics, while h-index of the tree is a compact measure to capture the actual depth of a discussion. We found evidence of significant differences in discussions from different semantic areas. For example, discussions about Mathematics tend to reach a much higher depth than the number of users involved and of edits in the corresponding articles would suggest.

This work proposes a first insight into Wikipedia as a space of discussion and offers many directions for improvement and for future investigation. The comparison of users' behaviour in the different networks (and maybe also in networks derived from interactions in article editing) could help in the identification of social roles. A more fine grained analysis involving the time-stamps of the comments may allow for a better understanding of social dynamics on a temporal dimension, and to detect contentious topics during a certain interval of time.

Acknowledgements

Yana Volkovich acknowledges support from the Torres Quevedo Program from the Spanish Ministry of Science and Innovation, co-funded by the European Social Fund.

References

Brandes, U.; Kenis, P.; Lerner, J.; and van Raaij, D. 2009. Network analysis of collaboration structure in Wikipedia. In *Proc. of WWW 2009*.

Crandall, D. J.; Cosley, D.; Huttenlocher, D. P.; Kleinberg, J. M.; and Suri, S. 2008. Feedback effects between similarity and social influence in online communities. In *Proc. of KDD 2008*.

Farina, J.; Tasso, R.; and Laniado, D. 2010. Automatically assigning Wikipedia articles to macrocategories. http://airwiki.ws.dei.polimi.it/index.php/WCG.

Foster, J. G.; Foster, D. V.; Grassberger, P.; and Paczuski, M. 2010. Edge direction and the structure of networks. *PNAS* 107(24):10815–10820.

Gleave, E.; Welser, H. T.; Lento, T. M.; and Smith, M. A. 2009. A conceptual and operational definition of 'social role' in online community. In *Proc. of HICSS 2009*.

Gómez, V.; Kaltenbrunner, A.; and López, V. 2008. Statistical analysis of the social network and discussion threads in Slashdot. In *Proc. of WWW 2008*.

Gómez, V.; Kappen, H. J.; and Kaltenbrunner, A. 2011. Modeling the structure and evolution of discussion cascades. In *Proc. of Hypertext 2011*.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *PNAS* 102(46):16569–16572.

Hu, H.-B., and Wang, X.-F. 2009. Disassortative mixing in online social networks. *Europhysics Letters* 86(1):18003.

Kittur, A., and Kraut, R. E. 2008. Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In *Proc. of CSCW 2008*.

Kittur, A.; Suh, B.; Pendleton, B.; and Chi, E. 2007. He says, she says: Conflict and coordination in Wikipedia. In *Proc. of SIGCHI 2007*.

Kittur, A.; Chi, E. H.; and Suh, B. 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proc. of CHI 2009*.

Laniado, D., and Tasso, R. 2011. Co-authorship 2.0: Patterns of collaboration in Wikipedia. In *Proc. of Hypertext 2011*.

Mishne, G., and Glance, N. 2006. Leave a reply: An analysis of weblog comments. In *Proc. of WWE 2006*.

Newman, M., and Park, J. 2003. Why social networks are different from other types of networks. *Physical Review E* 68(3):36122.

Newman, M. E. J. 2002. Assortative mixing in networks. *Physical Review Letters* 89(20):208701.

Schneider, J.; Passant, A.; and Breslin, J. 2010. A qualitative and quantitative analysis of how Wikipedia talk pages are used. *Proc. of the WebSci 2010*.

Stvilia, B.; Twidale, M. B.; Smith, L. C.; and Gasser, L. 2008. Information quality work organization in Wikipedia. *JASIST* 59(6):983–1001.

Suh, B.; Convertino, G.; Chi, E. H.; and Pirolli, P. 2009. The singularity is not near: slowing growth of Wikipedia. In *Proc. of WikiSym 2009*.

Szell, M.; Lambiotte, R.; and Thurner, S. 2010. Multirelational organization of large-scale social networks in an online world. *PNAS* 107(31):13636–13641.

Tang, L. V.-S.; Biuk-Aghai, R. P.; and Fong, S. 2008. A method for measuring co-authorship relationships in mediawiki. *Proc. of WikiSym 2008*.

Viégas, F.; Wattenberg, M.; Kriss, J.; and van Ham, F. 2007. Talk Before You Type: Coordination in Wikipedia. In *Proc. of HICSS 2007*.