Emotions and dialogue in a peer-production community: the case of Wikipedia

David Laniado[*, †]
david.laniado@barcelonamedia.org

Carlos Castillo[‡]
chato@acm.org

Andreas Kaltenbrunner[*]
kaltenbrunner@gmail.com

Mayo Fuster Morell[§]
mayofm@cyber.law.harvard.edu

[*] Barcelona Media Foundation, Spain
[†] DEI, Politecnico di Milano, Italy
[‡] Yahoo! Research Barcelona, Spain
[§] Berkman Center, Cambridge, MA, USA

ABSTRACT

This paper presents a large-scale analysis of emotions in conversations among Wikipedia editors. Our focus is on the emotions expressed by editors in talk pages, measured by using the Affective Norms for English Words (ANEW).

We find evidence that to a large extent women tend to participate in discussions with a more positive tone, and that administrators are more positive than non-administrators. Surprisingly, female non-administrators tend to behave like administrators in many aspects.

We observe that replies are on average more positive than the comments they reply to, preventing many discussions from spiralling down into conflict. We also find evidence of emotional homophily: editors having similar emotional styles are more likely to interact with each other.

Our findings offer novel insights into the emotional dimension of interactions in peer-production communities, and contribute to debates on issues such as the flattening of editor growth and the gender gap.

Categories and Subject Descriptors: H.5.3 [**Information Interfaces**]: Group and Organisation Interfaces – *Computer supported cooperative work, Web-based interaction*

Keywords: Wikipedia, discussion, emotions, gender gap

1. INTRODUCTION

As Goodwin et al. note, "like other aspects of culture, emotions can be seen as an aspect of all social action and social relations. They accompany rational acts as fully as irrational ones, positive experiences as much as negative ones" [19]. Online collaboration and peer-production [5], like any other social activity, evokes emotions – as well as it cultivates certain emotional environments or styles more than others.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
WikiSym'12, Aug 27–29, 2012, Linz, Austria.
Copyright 2012 ACM 978-1-4503-1605-7/12/08 ...$15.00.

We study one of the largest peer-production communities: the English Wikipedia. While there is quite a significant volume of data analytics on Wikipedia and the communications among their editors, not much is known about their emotional expression. Our focus is especially on the emotions expressed in discussion ("talk") pages. Talk pages play a fundamental role in the process of content creation at Wikipedia. They are one of the main places where dialogue and conflict between editors emerge, and where discrepancies are –eventually– solved.

The above-mentioned interactions evoke emotions in the editors. The way to deal with and express these emotions, for example through comments, can be an important descriptor of the editor base and of the interaction modes hosted at this community. The emotional expression of editors may depend on factors such as their level of experience and their gender. One would expect that the more experienced editors are, the larger their ability to channel their emotions towards constructive ends. One would also expect that gender plays a role, as common stereotypes assume men react with a different emotional mode than women; indeed gender has been found to explain in part diverse emotional expressiveness [8].

We expect to find in general a positive interaction environment in Wikipedia. According to survey data [41], editors describe each other as collaborative, intelligent, helpful and/or friendly (50% to 30%) more often than arrogant, unfriendly, rude and/or dumb (25% to 5%). However, Wikipedia is not devoid of conflict, as even a casual inspection of the discussion page of any controversial article shows. Additionally, serious problems including the gender gap (i.e. a strong inequality in the gender distributions of the participants) and the recent slow-down in the growth of the number of editors are a source of concern for the Wikipedia community [41].

Research questions and roadmap. As our review of previous works in Section 2 illustrates, there is a gap in our understanding of Wikipedia, and peer-production processes in general, in terms of the emotional expressions and responses of their contributors. Section 3 describes our approach, which tries to integrate data analytics and qualitative methods with a more socio-political, qualitative perspective [14, 17] to analyse Wikipedia.

The remainder of the paper is structured around a series of research questions (RQ1–RQ5) about emotional expressions in Wikipedia in general, about emotional styles of editors in collaboration with other editors in article talk pages, and about their relation with their own personal spaces:

- RQ1. How are the emotional styles of editors affected by their level of experience? (Section 4)
- RQ2. How are the emotional styles of editors affected by their gender and the topics they choose to work on? (Section 5)
- RQ3. How are the emotional expressions affected by interacting with others in comment threads (emotional congruence)? (Section 6)
- RQ4. How are the emotional styles of editors related to those of the editors they interact more frequently with (emotional homophily)? (Section 7)
- RQ5. How are the emotional styles of editors expressed in their personal spaces (user talk pages)? (Section 8)

After presenting our findings on each of these research questions, the last section summarises our conclusions, and provides some recommendations in terms of potential improvements on the socio-technical architecture of Wikipedia (and potentially, of other peer-production communities).

2. RELATED WORK

The literature on Wikipedia is vast: Okoli et al. [34] have identified at least 2,000 articles about Wikipedia. In this section, we outline previous works closely related to ours, but our coverage of these topics is by no means complete.

2.1 Wikipedia editors

Wikipedia's editor base has been analysed from several perspectives and profiled along many dimensions. Motivations to contribute and demographics are the aspects that have been studied in more detail through large surveys [41] as well as focused ones [13, 28]; personality traits have also been analysed [1].

Previous findings point to the diversity of user profiles: there are editors willing to contribute valuable content, and others that edit the Wikipedia for other reasons, for instance for self-promotion [23]. Furthermore, different editors prefer different types of tasks, such as contributing original content, or editing/reviewing content contributed by others [40].

Our findings also point to the diversity of editors' profiles, however we focus on emotional style, an aspect that has received little attention in the literature.

2.2 Wikipedia talk pages

Wikipedia provides a rich interaction mechanism between editors, including both *article talk* pages for discussing articles, and *user talk* pages that are used by most editors as a sort of public mail in-box. Both spaces are important to host users' discussion and interaction, but there are differences among them. Community well-being postings, such as courtesy, appear more frequently in user talk pages, while messages about the quality and accuracy of articles appear more frequently in article talk pages [22].

Viegas et al. [39] have described the conversation in article talk pages as "formalised and policy driven". According to them, about 60% of the edits in article talk pages can be considered requests/suggestions for editing coordination. The remainder are questions and references to Wikipedia policies. Networks of the interactions among users in talk pages have been studied extensively [12, 27].

In contrast with previous works, we represent editor discussions by the emotions they express, and use them to quantify message-level congruence and user-level homophily in the interaction network.

2.3 Editor experience

The contributions and involvement of editors in the community vary over time, with some editors moving from peripheral participation to full involvement [9], and others abandoning Wikipedia [7, 21].

The contributions by registered users are in general of higher quality than those from anonymous users [3], but even changes by an long-time editor can be rejected or revised [20]. In general, the quality and quantity of work done by editors seems to be independent of their level of experience [35]. However, there are measurable differences in the way editors at different levels of experience interact with each other.

Panciera et al. [35] show that as editors spend more time in Wikipedia, they cite more often Wikipedia policies in their discussions. Yasseri et al. [42] indicate that less prolific editors express themselves in ways that tend to be more negative than more prolific ones. They also found that less active editors tend to address the most active ones rather than each other. From the inspection of some discussions, the authors suggest that "they do not consider low-activity editors worthy of commenting upon".

The present work explores editors at different levels of experience, validating some of the findings from previous works, and deepening them by focusing in the emotional content of the discussions by regular editors and administrators.

2.4 Gender bias

The April 2011 survey [41] indicated 9% of editors were female, down from 13% on the year before [18]. This is a concern for Wikipedians and the Wikimedia Foundation, and has prompted a significant amount of research on Wikipedia's *gender bias*.

Lam et al. [25] carried an in-depth study of several gender-driven differences among Wikipedia editors. They found that the gender gap depends strongly on the topic (e.g. geography and history have proportionally less female editors than people and arts). They also found that female editors participate more in the social and community areas of Wikipedia (talk pages and user pages), and face more adversity, i.e. their edits are reverted more, specially when they are newcomers, and overall they are more likely to be blocked indefinitely. Furthermore, as editors gain experience, the gender gap gets larger. On the other hand and perhaps more surprisingly, there is also a gender gap among Wikipedia *readers*. Lim and Kwon [30] found that among university students, males tend to use Wikipedia for academic work and for personal interests more than females.

Antin et al. [4] found that gender differences do not affect the type of works people do on the Wikipedia, but that among the most active Wikipedians, male editors tended to make more edits than female, while women tended to make more extensive revisions.

Collier and Bear [11] analysed a large survey of Wikipedia readers and editors; they concluded that women contribute

Table 1: Basic statistics of our dataset.

Articles	3 210 039	
Articles with talk page (ATP)	871 485	(27.1%)
Total comments in ATP	11 041 246	
Comments in ATP linking to policies	460 644	(4.1%)
Users who comment articles	350 958	
Users with ≥ 100 comments on ATP	12 231	(3.5%)
Registered users	12 651 636	
User talk pages (UTP)	1 662 818	(13.1%)
Comments in UTP	13 670 980	

less because they are afraid of being criticised or face conflict in Wikipedia, they are less confident in their expertise and value less their contribution, and prefer to collaborate and share rather than edit and delete other people's work. Worryingly, 22% of female editors have reported inappropriate messages or other unpleasant interactions [41].

In contrast with previous work, our research looks for quantifiable differences in the emotions expressed by male and female editors both in their discussions in article talk pages, as well as on their personal user talk pages.

2.5 Emotions and peer-production

A recent study by Kucuktunc et al. [24] focuses on emotions in a large-scale question-answering community. According to their findings, sentiments can be correlated with many factors including gender and experience in the system. Women express stronger and more positive sentiments than men; and community members tend to have a stronger tendency to give neutral answers as they gain experience.

There is also a strong correlation between the level of emotion expressed in questions and received answers: neutral questions tend to be answered in neutral terms, while emotionally-loaded questions tend to incite answers with higher emotional content. This "emotional congruence" or "emotional homophily" has been observed in other contexts such as online fora [10] and blogs [38].

Our work studies emotions in the context of Wikipedia, which has a more complex interaction mechanism (including article editions, article discussion, and user discussion among other elements) that goes beyond question answering or interactions in blogs.

3. FRAMEWORK OF ANALYSIS

In this section we describe how we acquired, sampled, and pre-processed Wikipedia discussions, and the framework for sentiment analysis used.

3.1 Data acquisition and pre-processing

We used a complete snapshot of the English Wikipedia,[1] processing discussion pages through a series of heuristics described in [27]. These heuristics reconstruct conversation "threads" from article talk pages. Table 1 reports some basic statistics of this dataset.

To get robust results when determining emotional styles, we focus on editors that are active in discussion pages. Specifically, we select all the $\approx 12,000$ editors that have written at least 100 comments.

We also detected comments in which editors invoke Wikipedia policies/norms. We identified all comments containing at least one link to a page in the `Wikipedia:` namespace, consistently with [35].[2]

To perform the per-topic analysis of Section 5 we use a topic classification described in [16], where each article is assigned to its closer macro-categories in the category hierarchy of Wikipedia. For the study on Section 7 we construct a social network of editors based on replies in article-talk pages, declaring as "connected" users who have exchanged at least one reply (similar results are obtained by adopting higher thresholds on the number of replies).

Table 2: Users with more than 100 comments by gender and administrator status. Category "unknown" includes 8,708 users that were not included in the crowd-sourced task, as well as 745 users whose gender could not be identified by evaluators.

	Non-admins	Admins	Total
Males	1 087	1 526	2 613
Females	68	97	165
Unknown	6 850	2 603	9 453
Total	8 005	4 226	12 231

3.2 Gender labelling

Editors can indicate their gender in their preferences, but few of them choose to do so (18.8% according to Antin et al. [4], 16.5% according to our observations). This information can be obtained through an API provided by Wikipedia.[3] In addition to this data, previous studies on Wikipedia and gender [25] have also searched for the presence of a user box – a kind of "badge" that users can easily incorporate in their user pages – reading "This user is a male [or female]."

We used a combination of previous methods. We first collected gender information provided through user preferences. Then, for a sample of 1,385 users who had this information missing, we collected assessments on their profile pages using crowd-sourcing through Crowdflower.[4] In this task we asked crowd-sourcing evaluators to look for any indication of gender, including user boxes but also the presence of a real name, of an implicitly-stated gender (e.g. "I am a father of two ..."), of a pronoun used when describing him/herself in third person (e.g. "User X lives in San Diego, she likes surfing, ..."), of a photo, etc.

We asked for 3 assessments of each user page and decided by simple majority. Overall, judges were able to identify about 50% of the users (47% male, 3% female). When it was possible to identify the editor's gender, it was in most cases because there was a clear indication of gender in the username or a real name (50% of those identified), or because there was an implicitly-stated gender (27% of identified males, 20% of identified females), or a pronoun (15% female, 10% male). The agreement among coders was 78%.

Table 2 summarises the overall numbers of users which we were able to identify as males or females, among administrators and regular users. As it can be noticed, administrators are much more likely to disclose their gender.

[1] `http://en.wikipedia.org/wiki/Wikipedia:Database_download`, downloaded on March, 2010.

[2] The namespace for information and discussion about Wikipedia, its policies and guidelines, as explained in `http://en.wikipedia.org/wiki/Wikipedia:Project_namespace`

[3] `http://www.mediawiki.org/wiki/API:Meta`

[4] `http://www.crowdflower.com/`

3.3 Sentiment analysis

We measure the emotional content of comments in article talk pages and user talk pages using the Affective Norms for English Words (ANEW), a list of words with emotional scores provided by human subjects [6]. The list contains about a thousand words that receive a score from 1 to 9 along three dimensions: valence, arousal, and dominance.

The **valence** dimension measures the extent to which words make subjects feel happiness, satisfaction and hope (higher values), or their opposites: sadness, dissatisfaction and despair (lower values). The **arousal** dimension measures the association of words with feelings of excitement, anger or frenzy (high) and their opposites (low); **dominance**, in turn, focuses on feelings of domination or being in control (high) versus feelings of submission or awe (low).

The algorithm to extract emotional scores from the discussions follows the method proposed in [15]. For every comment in the user and article talk pages, we identified and counted the number of occurrences of the different words in the ANEW list. We then matched every word with its scores in the three emotional dimensions, and calculated, in a third step, frequency-weighted averages over each score for each piece of text analysed.

Example messages. In Table 3 we present three example comments with their ANEW words highlighted, and the corresponding emotional scores. By design, ANEW words cover a large range of emotional variability. For instance, the 5 most positive words (highest valence) are *joy, love, loved, miracle, paradise* while the 5 most negative (lowest valence) are *cancer, funeral, rape, rejected, suicide.*

Normalisation by context. The last example in Table 3 suggests that emotions associated to a comment often depend on the topic discussed. Indeed, most words in the ANEW list do not directly describe feelings (such as *happiness* or *sadness*); instead, they describe concepts that evoke emotions. These concepts may match topics in the articles themselves, such as *war* or *game.* Hence, we also calculate a normalised average to address this topical bias. In the case of discussions associated to articles, emotional scores are normalised with respect to their context: the average score of words in the discussion page where they are written.

The first example has a remarkably positive tone, in the context of a quite neutral discussion about mathematics; the second contains evocative words in a technical discussion about astrology, while the third one stands out for pointing negative opinions on a college, subject of the article.

3.4 Statistical tests

Significance tests. When reporting statistical significance, it is the outcome of performing t-tests of the null hypothesis that two samples are drawn from normal distributions with equal means and equal but unknown variances, against the alternative of unequal means.

Assortativity test. When reporting assortativity we use a *shuffle test* [2]. This test measures the significance of correlations between nodes on a graph. It first measures the correlation coefficient r between some variable of a node (e.g. the average age of a person in a social network), and the same variable in users connected by a social link. Next, the social links are randomly "shuffled", i.e.: we generate a series of random graphs having the same degree sequence as the original one; the same correlation is averaged over the shuffled graphs (r_{rand}), and the standard deviation is computed (σ_{rand}). Finally, the Z-score is defined as $Z = (r - r_{rand})/\sigma_{rand}$. Large absolute values of Z evidence that the observed correlation is not accidental.

Table 3: Example messages with their corresponding Valence, Arousal, and Dominance scores.

	V	A	D
Sounds like a **good** challenge - to be proven or disproven. I'm **happy** if it can be shown to go further using closed cubic polynomial solutions. The **nice** thing about these are that they are **pretty easy** to test numerically ... –in "Exact trigonometric constants"	7.4	5.3	6.2
Seems you have not yet seen female lover after having **sex** who do not **wish** to have **sex** with the same lover any more :) Once you've seen it, you understand very well what **war** of Venus means compared to **war** of Mars. –in "House (astrology)"	5.5	7.0	5.2
What about the whirlie hazing, the alcohol **abuse**, the emotional **poverty**, the **suicide** in 1995/6, the biotech plans which were stopped by pitzer protests ... –in "Harvey Mudd College"	1.6	5.8	3.5

4. EMOTIONS AND EXPERIENCE

We first measure whether the level of experience of editors affects their emotional expression. We measure the experience of editors in terms of their number of edits and in terms of their status as Wikipedia administrators (measuring experience as the time passed since the editor's first contribution to Wikipedia, instead, gives little or no correlations).

4.1 Results

Administrators tend to be more positive than regular users, as depicted in Figure 1. The figure shows the probability distribution of the average ANEW scores of the administrators and normal users. The mean values of the two distributions ± the standard errors *in the estimations of their means* (SEM) are indicated by the two vertical bars, well separated from each other. The larger the distances between those bars the less likely the differences of the mean values are observed by chance. This difference is significant with $p < 0.001$. To illustrate the difference further we also plot the ratio between the two curves, indicating the likelihood of a user being an administrator given the average valence of his/her comments (and assuming equal sample sizes). This likelihood grows the more positive the comments become.

Figure 2 (left) depicts this difference showing also results aggregated by gender. The lines for "All" users correspond to the width of the vertical bars in Figure 1, markers indicate the mean and the surrounding lines are of the size of twice the standard error of the mean. On the right, a similar difference between normal users and administrators can be observed for the dominance expressed in the comments. The differences hold even after normalisation as described in Section 3.3 and are more marked for male than for female users; for the latter they are not statistically significant.

This trend is further confirmed when looking at the number of edits of users. We find a weak but significant correlation ($\approx 0.12, p < 10^{-8}$) between the logarithm of the number

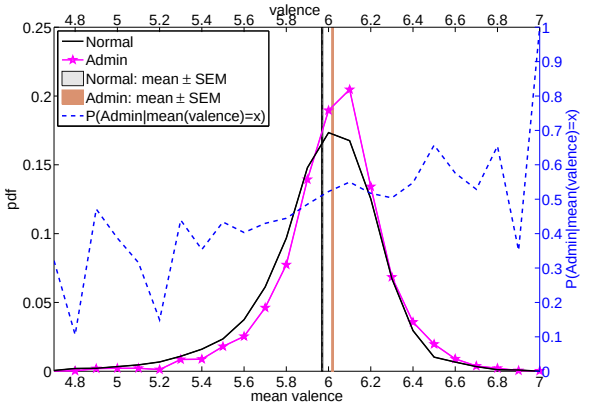

Figure 1: Distribution of the ANEW-valence scores of admins vs normal users, with the mean value and the interval of mean ± the standard error of the mean (SEM). Dashed lines indicate probability of being an administrator given the mean valence of a user (assuming equal sample sizes).

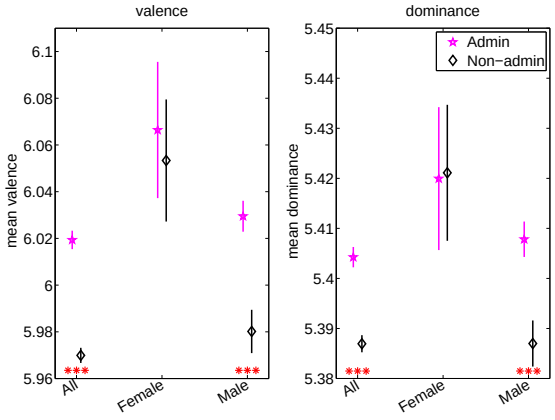

Figure 2: Mean(±SEM) of valence and dominance for regular users and administrators. Asterisks indicate statistically significant differences ($p < 0.001$).

of edits of a user and the (normalised) mean dominance and mean valence of the messages they write. Independently of their administrator status, as editors gain experience they tend to be more positive on their comments and express a higher dominance. The arousal variable is not affected.

Finally, we find that administrators tend to include nearly twice as often links to Wikipedia policies in their messages when compared to non-administrators (5.6% vs 3.0%; $p < 10^{-7}$), and tend to be significantly more brief, writing about 70 words per message on average, compared to 80 words per message on average for non-administrators.

4.2 Discussion

Our results are consistent with several observations from previous works. The relatively narrow range of emotional valence is expected considering that article discussion tends to be formal [39] and that most articles are not controversial [42]. The fact that administrators cite Wikipedia policy more often is consistent with [35]. Our quantitative findings agree with the qualitative observation that more prolific editors are more positive [42].

While in a question-answer system experienced contributors were found to give more neutral answers [24], in Wikipedia experienced contributors tend to have a more positive tone (valence) but not a stronger emotional expression (arousal).

The fact that there is no significant difference in the emotional variables between female administrators and non-administrators can be interpreted as indicating that the relatively few female editors that pass the 100-comments mark (the ones in our sample) tend to be more assertive –not necessarily aggressive, given the positive valence they express. This is in agreement with findings by Collier and Bear [11] on aversion to criticism and conflict as key variables driving female editors away from the Wikipedia; findings by Lam et al. [25] that female editors with a high level of experience are more likely to take an administrative role than their male counterparts at the same level of experience; and with gender differences in leadership styles in male-dominated organisations in general [37].

According to our results, the gender gap problem should be approached not only as a matter of having less female editors contributing to Wikipedia. It also implies that the female editors that decide to contribute are not a representative sample, and that as women gain experience in Wikipedia they tend to adopt the emotional tone of administrators.

5. TOPICS AND GENDER

Next we examine gender differences in terms of topics and emotions, which are suggested by the tag cloud of ANEW words presented in Figure 3.

5.1 Results

Topical gender bias. We find a negative correlation between the valence of the discussions within a topic category and the proportion of comments written by male editors ($r = -0.64$, $p < 0.01$). Women participate relatively more (15% or more of the comments –considering editors for which we were able to identify their gender) in discussions about *Arts*, *Health*, *Mathematics* and *Computing*, where more positive emotions are expressed, while less than 7% of comments are written by women in *History and events* and *Geography*

Figure 3: Tag cloud of ANEW words used more often by female (red) and male (blue) editors. Size accounts for difference in frequency.

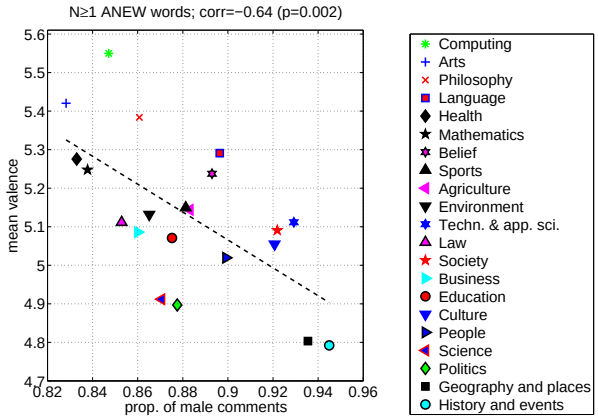

Figure 4: Mean valence for discussions of articles in different topic categories, vs the proportion of comments written by male editors (considering editors for which we were able to identify their gender).

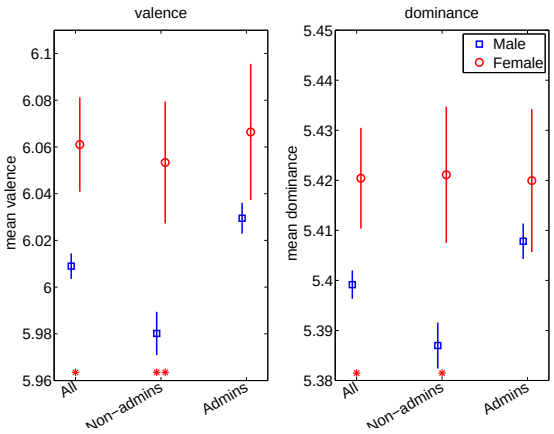

Figure 5: Mean valence and dominance for males and female users. Asterisks indicate statistically significant differences ($p < 0.01$, * $p < 0.05$). Significance disappears when normalising these values.**

and places, the two topics having the lowest scores of valence. These observations regarding gender bias in the article discussions are depicted in Figure 4, where markers represent a set of articles belonging to a topic; the y-axis indicates the average valence of the comments in these articles' talk pages and the x-axis indicates the proportion of these comments from male editors (vs female editors).

Emotional gender bias. Figure 5 shows that female editors write on average more positive comments, and with higher dominance. However, this difference is no longer statistically significant when we normalise by topics. In other words: the emotional content of discussions of male and female editors in article talk pages is to a large extent determined by the different topics that male and female editors choose to participate on.

Links to Wikipedia policies. Female editors include links to Wikipedia policies in about 9.8% of their messages in article discussions, which is more than twice the percentage observed for males (3.9%). These differences are statistically significant both within administrators (12.4% vs 4.9%) and non-administrators (6.2% vs 2.5%), with $p < 0.01$.

Additionally, we find that comments containing links to Wikipedia policies are characterised on average by higher valence (+1.03), arousal (+0.38) and dominance (+1.04). All results are significant with $p \ll 0.001$. This indicates that, when editors invoke community norms, they tend to do it with a remarkably positive and dominant tone, and with stronger emotional load than in the rest of the discussion.

Message length. We also observed a small, but statistically significant ($p < 0.05$) difference in the average length of messages written by females and males (83 vs. 71 words). Interestingly, this difference is accentuated for administrators (85 vs. 68).

5.2 Discussion

Our results on the dependency of emotions on topics agree with observations on a question-answering portal [24]. The results on this section help to interpret the finding by Lam et al. [25] that female editors prefer certain topics: such topics tend to have discussions in which words expressing a more positive valence are found.

The finding that female editors double the fraction of comments with links to Wikipedia policies is significant, considering the strong correlation with the level of experience of Wikipedia editors [35]. This reinforces our observation from the previous section that female editors are special and behave more as administrators than their male counterparts.

6. EMOTIONAL CONGRUENCE

In this section we focus on emotions in responses to comments, and we study whether editors receive replies that are emotionally consistent with the messages they write. This has been dubbed homophily or emotional congruence in previous works [24, 38]. In the present work we denote by *congruence* the correspondence at the level of messages, and by *homophily* the correspondence at the level of users.

6.1 Results

We calculate the differences of the average ANEW scores of every comment in an article discussion page, and its replies; we then average these differences for all comment-reply pairs. In this specific case we consider all comments, not just those of editors having more than 100 comments.

We observe that on average editors tend to reply with significantly ($p < 0.05$) higher valence (+0.05 per comment) and dominance (+0.04), while differences are not statistically significant for arousal. This may imply that users tend to be more positive and use words which underscore their arguments (have a larger dominance) when replying, but without recurring to words evoking stronger sentiments.

6.2 Discussion

The above-mentioned results help explain why conflictive articles are a minority. Having positive responses that do not evoke stronger sentiments can help reduce the number of discussions that spiral down into conflict. This keeps most article discussions in a state of consensus or at least temporary consensus [42].

The Wikipedia dispute resolution policy explicitly recommends refutation and counter-argument over contradiction, responses to the tone of the other person's message, ad-hominem argumentation and name calling; in a sense, this expresses a preference in the community for less emotionally-loaded messages (lower arousal).

7. EMOTIONAL HOMOPHILY

Our next step is to determine if editors tend to interact with other editors having similar emotional styles and participation patterns, aggregating comments at the level of the users that author them. As explained in Section 3, the network we study here connects users who have exchanged at least one reply to a comment in an article talk page.

7.1 Results

Figure 6 depicts a sub-graph of the discussion network (for clarity, depicting only editors who have exchanged at least ten messages). Colours highlight users that overall exhibit a particularly high emotional valence in their comments (red) or a particularly low one (blue). Visually, we can see groups of users with similar colours connected together.

To quantify this intuition, we use the concept of *assortativity* [32]. In simple terms, a social network is assortative if people tend to relate to others similar to them; and disassortative if people tend to relate to others who are different from them. In general, social networks tend to be assortative with respect to the number of connections of people [33]: those who are well-connected tend to be connected among them. We measure assortativity using the *shuffle test* described in Section 3.4.

Assortativity according to emotions. Table 4 shows that users send and receive messages to users exhibiting a similar emotional style across the three emotional dimensions we measure. This result holds (with less intensity) also after normalisation with respect to each article's discussion.

Assortativity according to gender. We also studied mixing by gender. As shown in the first row of Table 5, the network is assortative with respect to gender. This result is especially due to the preference of female editors to communicate with other female editors, as the number of messages exchanged among women is much higher than expected.

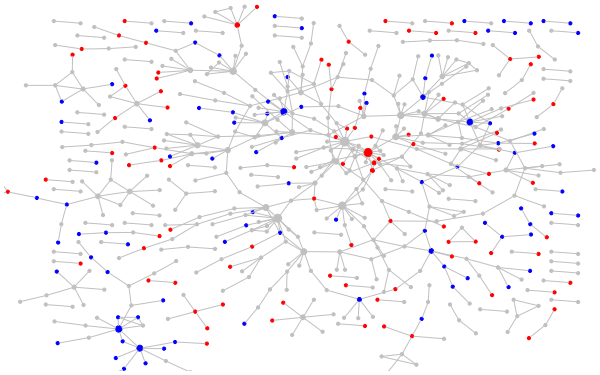

Figure 6: Discussion network of editors that have exchanged at least 10 messages. Colours indicate average valence: red for the 15% most positive, blue for the 15% most negative. Sizes are proportional to degree.

Table 4: Large Z-scores indicate emotional homophily in the directed "reply" network, according to messages sent and received by each user. Normalised values are emotional scores divided by the average score of the pages in which they are found.

Normalised		r	r_{rand}	σ_{rand}	Z
valence	(sent)	0.0269	-0.0003	0.0011	**23.8**
	(received)	0.0109	-0.0004	0.0010	**10.8**
arousal	(sent)	0.0253	-0.0004	0.0009	**28.2**
	(received)	0.0187	0.0013	0.0012	**14.8**
dominance	(sent)	0.0380	-0.0001	0.0015	**26.2**
	(received)	0.0121	9.8e-08	0.0011	**10.8**

Table 5: Large absolute Z-scores (in bold) indicate assortative (positive) or disassortative (negative) mixing between editors.

	r	r_{rand}	σ_{rand}	Z
gender	0.0443	-0.0008	0.0059	**7.63**
#comments written	-0.0177	-0.0014	0.0017	**-9.51**
#replies received	-0.0060	-0.0013	0.0014	**-3.50**
#replied users	-0.0340	-0.0023	0.0020	**-16.23**
#replying users	-0.0237	-0.0014	0.0015	**-14.35**
#discussed articles	-0.0009	-0.0011	0.0014	0.12
avg #words (comments)	0.0802	0.0003	0.0013	**61.08**
avg #words (replies)	0.1875	-5.9e-05	0.0012	**156.56**
first comments %	0.0663	0.0006	0.0012	**52.61**
received/written	0.1906	0.0004	0.0023	**80.97**

Disassortativity according to involvement. Next, we studied the assortativity of this network according to measures of the volume of activity in article talk pages: the number of comments written and replies received, the number of users they have interacted with writing (*#replied users*) or receiving (*#replying users*) comments, and the number of article discussions in which they have participated.

As shown in Table 5, all measures depending on the users' volume of activity are disassortative, with the exception of the number of articles discussed, where no significant pattern is found. This result is in line with previous studies [26, 27], and indicates that even within our sample of editors having more than 100 comments in discussions, active editors tend to interact especially with less active ones, and vice-versa.

Assortativity according to style. Finally, we studied the assortativity of the social network of editors according to stylistic characteristics of their messages: the average number of words in comments written [*avg #words (comments)*] or received [*avg #words (replies)*], the proportion of messages written without replying to any comment (*first comments %*), and the ratio between received replies and written comments (*received/written*). All these measures indicate assortative mixing, i.e. more intense communication between similar users.

7.2 Discussion

Emotional homophily seems to be a prevalent phenomenon in community sites [10, 24, 38] and Wikipedia does not seem to be an exception, as is confirmed across the three emotional variables we consider, even after controlling for topics.

The mixing pattern of experienced and less experienced editors deepens previous findings [42] through a larger set of metrics, and supports the idea of "mentoring", where experienced editors help less experienced ones to soften their

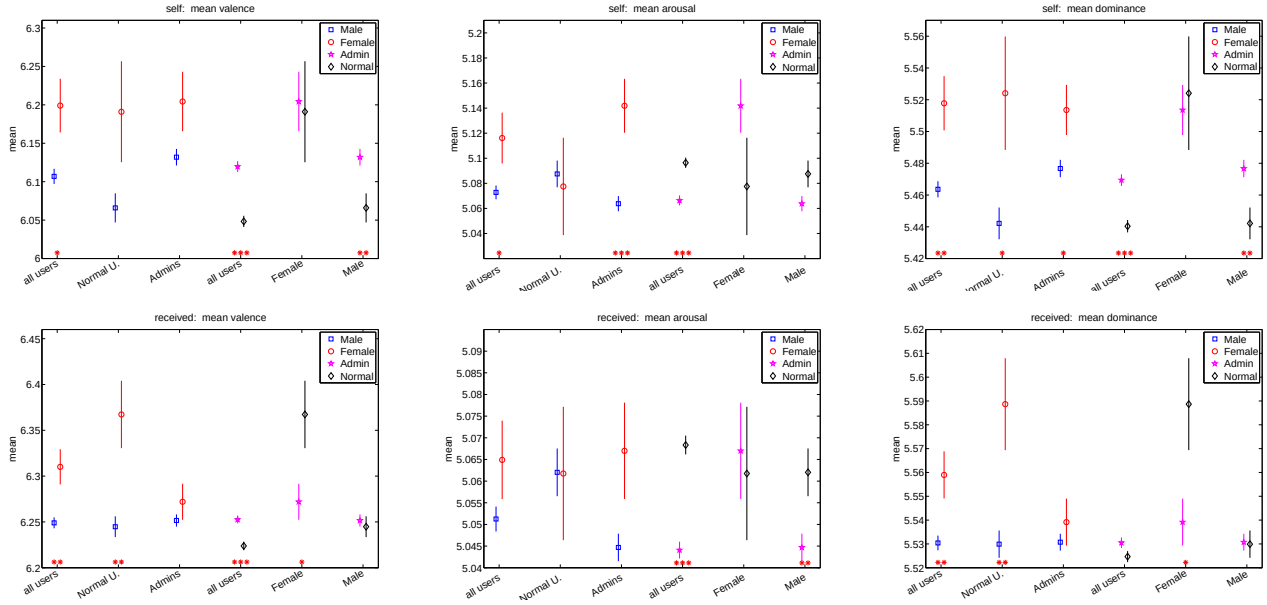

Figure 7: Average ANEW-scores of different users classes in the user-talk pages. Top: comments written by the users on their own talk page. Bottom: comments received from other users. Asterisks indicate statistically significant differences ($***: p < 0.001$, $**: p < 0.01$, $*: p < 0.05$).

learning curve [31]. It also provides a possible explanation to the observation that editors tend not to vote (e.g. in the process to grant administrator privileges) for other editors that have a similar edit count to them [29].

Interestingly, female editors tend to interact with each other much more frequently than what would be expected.

8. PERSONAL ("USER TALK") PAGES

So far we have focused on article discussions ("article talk") pages, instead in this section we analyse the emotions expressed and received by editors in their personal spaces ("user talk" pages).

8.1 Results

Own messages in user talk pages. In the top row of Figure 7 we measure the emotional variables of comments written by the users *on their own user talk pages*. We find that administrators express more positive emotions, use less emotional content in general (lower arousal) together with a higher dominance on their personal pages. This is consistent with the valence and dominance they use in article discussion pages, which is also higher than for regular users as shown in Section 4. The lower arousal in their user pages may indicate that even in that personal space, administrators are less willing to express emotions and try to keep a more neutral tone. However, this effect is only visible for male administrators. Female administrators express (with high significance) more emotions (i.e. higher arousal) on their personal talk pages than their male colleagues.

Messages left by other users. Next, we analyse the average sentiment of comments *written by other users* on a user's talk page. We notice that female editors receive comments that are significantly more positive and express more dominance than those received by males.

However, this trend is not observed for female administrators which receive on average the same emotional content as male administrators, while receiving significantly less positive comments with less dominance than normal female users. No such difference exists for males. A supposed gender difference seems here to be clearly out-weighed by the type of interactions the administrator tasks bring with them.

Finally, we conducted a qualitative assessment of the pages of the 30 users having the highest and lowest valence scores. The highest valence scores were associated in general with encouraging, positive messages. The lowest valence scores were associated with a more neutral tone, never expressing a strong negativity. Women are seen expressing more welcoming, affectionate feelings in their user talk pages, while men tend to maintain a more neutral tone. In some cases, we see women addressed in a tone that can be read as paternalistic or condescending.

8.2 Discussion

Our findings confirm that user talk pages have generally a positive tone, agreeing with the observation by Hara et al. [22] regarding courtesy in these spaces. However, there are significant differences between administrators and non-administrators and between male and female editors.

9. CONCLUSIONS

We have provided data and insights about the emotional dimension of Wikipedia, and how emotions are related to the profiles of editors and to their interactions. Below we summarise our main conclusions and recommendations.

C1. Wikipedia editors express themselves in general with a positive emotional tone. According to our analysis, most Wikipedia discussions occur within a range of emotional expressions that can be interpreted as going from neutral to positive tone. Furthermore, people tend to reply

to comments in discussions with a more positive tone than the original messages, and to write in other editors' user talk pages with a more positive tone than on their own pages. This corresponds to a positive attitude of Wikipedians towards each other, in line with the "assume good faith" [36] and other principles that Wikipedia promotes.[5]

C2. Female editors are different in their emotional expression and relationships. A gender-based comparison of emotional styles reveals remarkable differences in the behaviour of male and female editors.

Female editors tend to work in topics in which the discussion has a more positive tone. They also receive more positive comments on their user talk pages. Both can be interpreted as signals of positive emotions brought by women, although we have also seen that sometimes messages addressed to them have a paternalistic tone that may not be constructive.

Our results raise interesting questions on the role of women with respect to power in a prevalently masculine community. Women are heavily under-represented in the general editor population, are even less represented at higher levels of experience [25]. However, active female editors express themselves similarly to male administrators irrespectively of their own administrator status, and female administrators and non-administrators behave similarly; they also tend to cite more often Wikipedia policies, a trait shared by experienced editors and administrators. This is in agreement with [37] and references therein: in male-dominated organisations, women tend to conform to leadership styles that mimic masculine values (e.g. being more task-oriented than interpersonal-oriented).

C3. Administrators and experienced users play a pivotal role. The study of assortativity in the editor network confirms the tendency of active users to interact with less active ones. Experienced editors and administrators express themselves with a more positive tone. These facts suggest that they (together with non-administrator female editors) are fundamental to promote a positive working environment.

C4. Wikipedians of a feather flock together. Editors communicate more with others having a similar style, both in terms of emotions in messages written and received, and also according to more superficial indicators such as the average length of their comments. Female editors tend to interact with other female editors with a higher frequency than what would be expected.

9.1 Recommendations

Based on our conclusions, we provide three main recommendations for the Wikipedia community.

R1. Design for positive emotional expression. At a high level, our results show that emotions do play an important role in a peer-production community. Contributors will experience emotions associated to both the content they generate and the actions of their peers. The challenge lies in taking this into account in the technological platforms and in the development of community policies and the promotion of social norms.

While Wikipedia (rightly) promotes a "neutral point of view" in the encyclopedic content, the needs and practices of the community of editors when interacting with each other requires an expressive environment. Providing spaces for emotional expression, favouring positive exchanges, creating ways to channel negative feelings in a non-destructive manner and in general providing a way of incorporating emotions towards productive contributions are ways of promoting a more diverse and gender-balanced community.

For instance, our results suggest that the community should build upon the positive tone of user talk pages and keep them as spaces in which users experience positive emotions. According to our qualitative assessment, it seems a safer option to do so in a gender-neutral way.

R2. Encourage experienced editors to contribute to positive emotions. To contribute to recruit and retain new editors over time, experienced editors and administrators in particular should be encouraged to maintain and foster a positive climate in Wikipedia, particularly considering that they tend to interact with new editors.

The fact that emotional styles are to some extent contagious could be used to "cool down" conflictive discussions. Some Wikipedia editors have been observed to specialise in mediation roles [23] and this should be encouraged. The emotional tone of messages with links to Wikipedia policies, favoured by administrators and female editors, is definitively not neutral. Policies and norms about linking to such messages should take this into account, for instance, by encouraging editors to explain clearly why a certain policy is invoked and to avoid the perception of an emotionally dominant tone as an arrogant one. Additionally, appropriate wording, graphics and/or visual cues should be used.

R3. Gender-aware recruiting. Given the clear preference of women for interaction with other women, it may be advisable that new female editors are invited and welcomed by other female editors. Invitations to women should also consider that certain topics will be more attractive to them. In that sense, it might be advisable to emphasise in such messages the topics in which the presence of female editors is already observed.

Conversely, editors on the topics for which the gender gap is more acute should actively seek the participation of more female editors, through direct messages or through postings in the community areas of the Wikipedia.

9.2 Future work

The study of the emotional dimension of collaboration is a relatively new subject, and therefore much remains to be done. Our study deals only with one peer-production community using one set of emotional metrics, and it would be valuable to conduct similar studies within other communities, over a long time period, and with more metrics. Indeed, we are currently applying other sentiment-analysis metrics to the same data and our preliminary results seem to be largely consistent with those reported here.

In Wikipedia, particularly in user talk pages, emotions are also expressed through non-textual forms. These include emoticons "`:-)`", "barn stars" (an informal award used extensively by Wikipedia editors), and virtual gifts such as "cookies" and "kittens". We have also focused on relatively experienced Wikipedia editors when trying to determine personal emotional styles. A study that includes non-textual emotional aspects and covers less experienced editors is likely to have to rely more on human annotations, and less on automatic methods.

[5] `http://en.wikipedia.org/wiki/Wikipedia:Assume_good_faith`

A deepening of the understanding of communicational and leadership styles in Wikipedia is also necessary. For instance, the fact that female editors, specially administrators, write longer messages may be a clue pointing to a more interpersonal-oriented leadership style as opposed to task-oriented, but this cannot be concluded from our data. There are other research questions that would benefit from more human annotation effort, directed towards goals such as measuring the extent of "condescending" or "paternalistic" language in comments addressed at female editors, or the use of sarcasm. This goes beyond what can be achieved presently by automatic analysis, and it is just one aspect of open research questions for future work.

Acknowledgements We wish to thank Riccardo Tasso for his fundamental contribution to the parsing of Wikipedia talk pages, Jacopo Farina for computing article topics, and Erik Zachte and the Cooperation group of the Berkman centre for Internet and Society (particularly Benjamin M. Hill) for their feedback. This work was partially funded by the CDTI and the Cenit Social Media project; by the Institute of Govern and Public Policies (AUB) and by the Beatriu de Pinos program of the Regional Govern of Catalonia.

Key references: [24, 25, 27]

10. REFERENCES

[1] Y. Amichai-Hamburger, N. Lamdan, R. Madiel, and T. Hayat. Personality characteristics of wikipedia members. *CyberPsychology & Behavior*, 11(6):679–681, 2008.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proc. of KDD*, 2008.

[3] D. Anthony, S. Smith, and T. Williamson. The quality of open source production. *Hanover: Dartmouth College*, 2007.

[4] J. Antin, R. Yee, C. Cheshire, and O. Nov. Gender differences in Wikipedia editing. In *WikiSym*, 2011.

[5] Y. Benkler. *The wealth of networks: how social production transforms markets and freedom*. Yale Univ. Press, 2006.

[6] M. Bradley and P. Lang. Affective norms for English words (ANEW). Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[7] U. Brandes and J. Lerner. Is editing more rewarding than discussion? a statistical framework to estimate causes of dropout from Wikipedia. In *Proc. of WWW*, 2009.

[8] J. Brebner. Gender and emotions. *Personality and Individual Differences*, 34:387–394, 2003.

[9] S. L. Bryant, A. Forte, and A. Bruckman. Becoming wikipedian. In *Proc. of GROUP*, 2005.

[10] B. Chee. Sickness and health: Homophily in online health forums. In *iConference*, 2010.

[11] B. Collier and J. Bear. Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. In *Proc. of CSCW*, 2012.

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

[13] U. Cress and J. Kimmerle. A systemic and cognitive view on collaborative knowledge building with wikis. *I. J. Comp. Supp. Collaborative Learning*, 3(2):105–122, 2008.

[14] D. della Porta and M. Keating, editors. *Approaches and Methodologies in the Social Sciences*. Cambridge University Press, New York, NY, USA, 2008.

[15] P. Dodds and C. Danforth. Measuring the happiness of large-scale written expression. *Journal of Happiness Studies*, 11(4):441–456, 2010.

[16] J. Farina, R. Tasso, and D. Laniado. Automatically assigning Wikipedia articles to macrocategories. In *Proc. of Hypertext*, 2011.

[17] M. Fuster Morell and B. M. Hill. Socio-political approaches to open collaboration. In *Proc. of WikiSym*, 2010.

[18] R. Glott, P. Schmidt, and R. Ghosh. Wikipedia Survey. Technical report, United Nations University, Mar. 2010.

[19] J. Goodwin, J. Jasper, and F. Polletta. *Passionate Politics. Emotions and Social Movements*. The University of Chicago Press, Chicago, 2001.

[20] A. Halfaker, A. Kittur, R. Kraut, and J. Riedl. A jury of your peers. In *Proc. of WikiSym*, 2009.

[21] A. Halfaker, A. Kittur, and J. Riedl. Don't bite the newbies: how reverts affect the quantity and quality of wikipedia work. In *Proc. of WikiSym*, 2011.

[22] N. Hara, P. Shachaf, and K. F. Hew. Cross-cultural analysis of the Wikipedia community. *JASIST*, 61(10):2097–2108, 2010.

[23] T. Iba, K. Nemoto, B. Peters, and P. Gloor. Analyzing the creative editing behavior of wikipedia editors. *Procedia - Social and Behavioral Sciences*, 2(4):6441–6456, 2010.

[24] O. Kucuktunc, B. B. Cambazoglu, I. Weber, and H. Ferhatosmanoglu. A large-scale sentiment analysis for Yahoo! answers. In *Proc. of WSDM*, 2012.

[25] S. K. Lam, A. Uduwage, Z. Dong, S. Sen, D. R. Musicant, L. Terveen, and J. Riedl. WP:clubhouse? an exploration of Wikipedia's gender imbalance. In *Proc. of WikiSym*, 2011.

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

[27] D. Laniado, R. Tasso, Y. Volkovich, and A. Kaltenbrunner. When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. In *Proc. of ICWSM*, 2011.

[28] K. K. Lee and G. G. Karuga. The role of cognitive conflict in open-content collaboration. In *Proc. of AMCIS*, 2010.

[29] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Governance in social media. In *Proc. of ICWSM*, 2010.

[30] S. Lim and N. Kwon. Gender differences in information behavior concerning Wikipedia, an unorthodox information source? *Library and Information Science Research*, 32(3):212 – 220, 2010.

[31] D. R. Musicant, Y. Ren, J. A. Johnson, and J. Riedl. Mentoring in Wikipedia. In *Proc. of WikiSym '11*, New York, New York, USA, 2011.

[32] M. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.

[33] M. Newman and J. Park. Why social networks are different from other types of networks. *Arxiv preprint cond-mat/0305612*, 2003.

[34] C. Okoli, M. Mehdi, M. Mesgari, F. r. Nielsen, and A. Lanamäki. The people's encyclopedia under the gaze of the sages. *SSRN eLibrary*, 2012.

[35] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made. In *Proc. of GROUP*, 2009.

[36] J. M. Reagle. *Good Faith Collaboration*. MIT Press, 2010.

[37] N. Z. Stelter. Gender differences in leadership. *Journal of Leadership & Organizational Studies*, 8(4):88–99, 2002.

[38] M. Thelwall. Emotion homophily in social network site messages. *First Monday*, 15(4), 2010.

[39] F. B. Viegas, M. Wattenberg, J. Kriss, and F. van Ham. Talk before you type: Coordination in wikipedia. In *Proc. of HICSS*, 2007.

[40] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith. Finding social roles in Wikipedia. In *Proc. of iConference*, 2011.

[41] Wikimedia Foundation. Editor survey, April 2011. [Online; accessed 13-September-2011].

[42] T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertész. Dynamics of Conflicts in Wikipedia. *PLoS ONE*, 7(6):e38869, 2012.