

# Direct Negative Opinions in Online Discussions

Claudiu-Cristian Musat, Boi Faltings  
École Polytechnique Fédérale de Lausanne  
Lausanne, Switzerland  
Firstname.Lastname@epfl.ch

Philippe Roussille  
ÍRIT laboratory  
Toulouse, France  
philippe.roussille@irit.fr

**Abstract**—In this paper we investigate the impact of antagonism in online discussions. We define antagonism as a new class of textual opinions - direct sentiment towards the authors of previous comments. We detect the negative sentiment using aspect-based opinion mining techniques.

We create a model of human behavior in online communities, based on the network topology and on the communication content. The model contains seven hypotheses, which validate two intuitions. The first intuition is that the content of the messages exchanged in an online community can separate good and insightful contributions from the rest. The second intuition is that there is a delay until the network stabilizes and until standard measures, such as betweenness centrality, can be used accurately. Taken together, these intuitions are a solid case for using the content of the communication along with network measures. We show that the sentiment within the messages, especially antagonism, can significantly alter the community perception. We use real world data, taken from the Slashdot<sup>1</sup> discussion forum to validate our model. All the findings are accompanied by extremely significant t-test p-values.

**Keywords**—opinion mining; direct opinions; text mining; centrality;

## I. INTRODUCTION

Information processing has become the backbone of modern society and the analysis of communication and its role in modeling interpersonal relations has a central role. So far, it has focused on objective content expressed in numbers and facts. However, human communication is much richer and many of its facets, including the expressed sentiment, have received comparatively little attention.

With the ever growing popularity of social networks, much research has been dedicated to understanding their inner mechanisms. Social network analysis (SNA) views social relations in terms of network theory with social actors as network nodes and social relations as links between them. The intuition is that a link between two individuals is a form of endorsement. By aggregating all the connections of a social actor one can thus obtain an indication of her value within the community.

Most of the research analyzes the implications of aggregating social links and not on what they represent. Various types of user centrality are computed with little regard to the underlying significance of how the network is formed. While undoubtedly a user's centrality is, to some extent, connected to her social importance, we believe complementary information extracted from the communication content can improve the quality of the overall picture.

Not all social interactions are positive, thus not all network links are a form of endorsement. Network measures sometimes provide good results, but in spite of this. Working with just the final value of a user's centrality is an approximation of her behavior. Potential explanations for this resilience abound. For instance, one might believe that, after an unacceptable behavior, the perpetrator is ostracized, thus ending up with a low centrality value. This might happen although, in the short term, people have noticed that low quality users thrive on negative feedback. This led to the saying *do not feed the trolls* - noting that a short term centrality boost is what those users are after. To the best of our knowledge, these intuitions have not been investigated.

In this paper, we propose a novel sentiment analysis method that identifies antagonism in a threaded discussion. Antagonism is defined as a negative opinion directed at the author of the parent comment. We show that its presence is correlated with the attention a comment or a user receive subsequently.

Using this method, we construct a model which connects the user behavior, extracted from their communication content, with the network topology. We name this approach a user level analysis. We also perform an analysis at the communication, or comment, level and hypothesize on what affects the community response to a message. We create seven hypothesis, pertaining to either a comment level or a user level analysis, which are grouped behind two intuitions. We test these hypothesis using real world data from the forum of the technology news website Slashdot, where users comment on news items.

The first intuition is that *the content of the communication can be used to separate useful messages and users from the rest*. We show that the sentiments expressed in messages exchanged within the network alter the community perception of those messages and their authors. This, in turn, leads to changes in the network topology.

The second intuition is that *centrality values have a significant delay in reaching their final values*. Within the studied dataset, certain behaviors trigger an initial centrality increase, while, in the long term, they are correlated with a centrality decrease. This delay makes the centrality value irrelevant for instantaneous, automatic forum moderating.

We start by reviewing the relevant previous work in Section II. We then describe the data structures in Section III, the experimental setting and methods used in Section IV, present the results in Section V and conclude our work in Section VI.

<sup>1</sup>www.slashdot.org

## II. RELATED WORK

Traditionally, social networks are studied as part of network theory. Degree related distributions, various types of node centralities and cluster analysis [1] are just a few of the methods and metrics used to study these networks. To determine social importance, algorithms like PageRank [2] are standard practice, and the use of homophily has been proven beneficial in detecting leaders [3].

However traditional approaches do not always work. [4] show that indegree is a poor indicator of popularity on Twitter. Additional information, available in text format, can be used in such cases. Text has been used to reveal the social structure [5], while [6] provide an influence measurement mechanism between the communication content and the social network structure.

Within the field of text mining, our primary interest is sentiment analysis. Most work in sentiment analysis has focused on texts in isolation [7]. Social context has been used to improve the performance of opinion mining techniques [8]. Moreover, social networks permit the study of sentiment propagation and opinion cascade formation [9]. However, the way the opinions flowing through a network influence the network structure has been little investigated.

A direct way of using sentiment is to assign a polarity to social links. [10] and [11] study online social networks and allow edges to have both positive and negative weights. Another approach is to create links between social actors, based on the opinions between their messages. The authors of [12] use a sentiment dictionary to extract the overall polarity of replies in a forum. They then analyze the average polarity of the replies to a certain comment. While this can be considered a way to quantify the community opinion about the said communication instance, the sentiment extraction method is prone to significant errors. An overall negative opinion of a comment can be directed at the subject discussed or outside factors, rather than the previous comment. We distinguish between the various types of opinions and focus on opinions targeted towards other users.

[13], [14] use sentiment data to establish leadership status and assess user credibility. They use dictionary-based sentiment mining methods coupled with standard ranking algorithms [2] to obtain opinionated rankings. The network-based and opinion-based analyses are, however, separated.

For the Slashdot news site, [10] analyzed the network topology based on the Friend/Foe option. [15] also used the comment reply structure as the relevant links. The latter conclude that relations within the Slashdot message board are more generic than explicit friend links. The polarity of these social links has not been the object of previous studies.

## III. DATA STRUCTURE

The news site *Slashdot* is a collaborative platform sharing news items that can be commented on. It features a comment rating system based on human feedback. This can be viewed as a gold standard of the community perception of the ongoing interactions and a readily available labeling system. We gathered 934'612 comments written by 35'045 users on 4'922

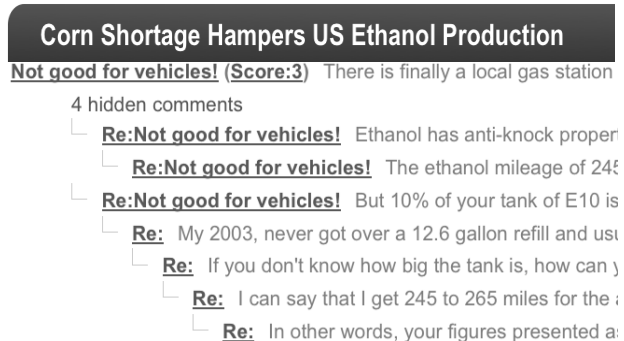


Fig. 1. A Slashdot story and sample comment chain

news stories between February 1, 2011 and February 1, 2012. The corpus is freely available for download<sup>2</sup>.

### A. Comment Level Structures

The comments are organized into discussion threads, as the one in Figure 1. Each new thread is a tree  $T = \{V, \rho\}$  that consists of a short story or news item  $s \in S$ , published by a member  $u \in U$ , which is followed by multiple comments  $c \in \mathcal{C}$ . Let  $\mathcal{P}(\mathcal{C})$  be the power set of  $\mathcal{C}$ . Similarly, let the power sets of all users, stories and text elements be  $\mathcal{P}(U)$ ,  $\mathcal{P}(S)$  and  $\mathcal{P}(\mathcal{C} \cup S)$  respectively. Each comment has an associated creation time, in POSIX format,  $t(c)$ ,  $t: \mathcal{C} \rightarrow \mathbb{N}$ .

Let  $\rho \subset \mathcal{C} \times \{\mathcal{C} \cup S\}$ , be the reply-to relation between a comment and another comment or news item and  $\rho^* \subset \mathcal{C} \times \{\mathcal{C} \cup S\}$  its transitive closure. The number of replies a comment  $c \in V_i$  receives is thus  $n(c) = \text{card}\{c' \in V_i | (c', c) \in \rho^*\}$ . We define a reply chain from  $c'$  to  $c$ , similar to the definition employed in [16], as  $\nu: \mathcal{C} \times \{\mathcal{C} \cup S\} \rightarrow \mathcal{P}(\mathcal{C})$ ,

$$\nu(c', c) = \{c'' \in \mathcal{C} | (c', c'') \in \rho^* | c' = c'', (c'', c) \in \rho^*\} \quad (1)$$

The length of the reply chain is  $n(c', c) = |\nu(c', c)|$ .

Moreover, for a single comment  $c \in \mathcal{C}_i$ , we define its **depth** as the length of the longest chain  $\nu(c', c)$ : Its depth is also the height of the largest subtree with the root in  $c$ .

$$\delta(c) = \max_{c' \in V_i} (n(c', c)) \quad (2)$$

For any comment subset  $\mathcal{C}' \subseteq \mathcal{C}$  we define the average number of replies as  $\bar{n}(\mathcal{C}') = \text{avg}_{c \in \mathcal{C}'} (n(c))$  and average comment depth  $\bar{\delta}(\mathcal{C}') = \text{avg}_{c \in \mathcal{C}'} (\delta(c))$ .

### B. User Level Structures

Each comment is written by a single user. We define the belong-to relation  $\mu: \mathcal{C} \rightarrow U; c \mapsto \mu(c) = u \in U$  and denote the set of comments authored by a user with  $\mathcal{C}(u) = \{c \in \mathcal{C} | \mu(c) = u\}$ . For any user  $u_i \in U$  let  $t_i$  be the moment when the first comment from that user is written.

Based on the relations between individual comments within news threads, we can define relations between individual users. The link weight between two users  $u_1$  and  $u_2$  is the number of comments written by  $u_1$  as replies to comments posted by

<sup>2</sup>[http://lia.epfl.ch/data/Slashdot\\_2011\\_2012.zip](http://lia.epfl.ch/data/Slashdot_2011_2012.zip)

$u_2. \omega : U \times U \rightarrow \mathbb{N}; (u_1, u_2) \mapsto \omega(u_1, u_2) = \text{card}\{(c_i, c_j) \in \rho \mid \mu(c_i) = u_1, \mu(c_j) = u_2\}$

In the graph of user replies  $G = (U, E_u)$ , let  $\sigma_{st}$  be the number of shortest paths between users  $s$  and  $t$ , and  $\sigma_{st}(u)$  the number of shortest paths between users  $s$  and  $t$  that contain user  $u$ . Vertices that have a high probability to be on shortest paths between other nodes have a high betweenness centrality. The betweenness centrality of  $u$  is defined as

$$C_B(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}} \quad (3)$$

The centrality of a user at a given point in time  $\tau$ ,  $C_B^\tau(u)$ , is obtained by constructing the graph  $G$  only based on comments having  $t(c) < \tau$ . For a user subset  $U' \subset U$  the average betweenness centrality is:  $\bar{C}_B(U') = \text{avg}_{u \in U'}(C_B(u))$

### C. Comment and User Ratings

Each comment  $c \in \mathcal{C}$  is assigned a discrete rating  $r : \mathcal{C} \rightarrow R = \{-1..5\}$ ; A -1 rated comment is considered unnecessary and unpleasant, while a comment rated 5 is considered meaningful. Comment ratings represent the *human generated* ground truth. The rating are given by the users themselves as +1 or -1 increments or decrements, and the final rating is an aggregated value of these contributions. We define the comment subset that share a given rating  $i$  as  $\mathcal{C}_i \subset \mathcal{C}$ .

A user's average rating is computed as the mean of the rating values of the comments she wrote:  $r : U \rightarrow R_u = [-1, 5]$ ,  $u \mapsto r(u) = \text{avg}_{c \in \mathcal{C}(u)} r(c)$ . We use  $U_{i,j} \subset U$  to denote the users having an average rating between  $i$  and  $j$ ,  $\forall u \in U_{i,j}, i \leq r(u) \leq j$ . The average rating for the given user set is  $r(U_{i,j}) = \text{avg}_{u \in U_{i,j}} r(u)$ . Moreover, we define the maximum centrality of a user set with an average rating between  $i$  and  $j$  as  $C_{B_{max}}(U_{i,j}) = \text{max}_{u \in U_{i,j}} C_B(u)$ .

An important noise source is that some posts are not rated by users. Instead, they are assigned an automatic rating based on previous contributions. Since these ratings do not represent direct human feedback, we removed from the analysis all comments with a rating equal to the mode of all comments belonging to their author, rated in the 0-2 range.

## IV. PROPOSED EXPERIMENTS

We investigate the role of antagonism in online human interaction. We create a model of the Slashdot user behavior using topological and content measures. We formulate seven hypothesis, grouped around the two aforementioned intuitions. **Intuition 1** is that *the opinion content of the communication can be used to separate useful communications and users from the rest*. **Intuition 2** is that *centrality values have a significant delay in reaching their final values*. We separate the hypothesis into comment level and user level ones and we use the human generated ratings to validate them.

We found that hostile behavior impacts the dynamics of social interaction, both at a comment and a user level. We define the antagonism within a comment as a direct negative reference to the writer of the parent comment or story.

We perform a dictionary based subjectivity analysis. For instance, "nice" has a mildly positive polarity and "idiot"

carries a strongly negative one. We attached polarity values  $p(w)$  to the comment terms  $w \in W_c$ . The values are obtained from OpinionFinder [17] and online curse and foul language collections.

A **direct opinion** is an instance of a polarized word, or modifier, that refers to a second person pronoun or possessive determiner. A relation is defined as the presence of a syntactic dependency chain [18] between the two words. For example, the sentence "Your opinion is not right.", has a polarity of -1, because *right* has a dependency chain to *your*, has a polarity of +1, and is negated. For all valid constructs, let  $p(w)$  be the polarity of the modifier and  $\text{neg}(w) \in \{0, 1\}$  states whether it is negated. A comment is **antagonistic** if contains more negative than positive direct opinions. We define the antagonism property of a comment as  $p_\alpha : \mathcal{C} \rightarrow \{0, 1\}$

$$c \mapsto p_\alpha(c) = I_{(\sum_{w \in W_c} p(w) \times \text{neg}(w)) < 0} \quad (4)$$

Antagonism is also a property of reply to links:  $\alpha : \mathcal{C} \times \{\mathcal{C} \cup S\} \rightarrow \{0, 1\}$ ,  $(c, c') \mapsto \alpha(c, c') = p_\alpha(c)$ . In this relation,  $c$  is the **source** of antagonism and  $c'$  the **destination**.

Let  $\mathcal{C}_{\alpha s} \subset \mathcal{C}$  be the set of all antagonistic comment sources. The average depth of these comments is thus  $\bar{\delta}(\mathcal{C}_{\alpha s})$ . For the antagonistic source comments that have a rating  $i$ , the average depth is  $\bar{\delta}(\mathcal{C}_{\alpha s i})$ . We use similar notations for the malicious comment destinations:  $\mathcal{C}_{\alpha d} \subset \mathcal{C}$ , with an average depth  $\bar{\delta}(\mathcal{C}_{\alpha d})$  and  $\bar{\delta}(\mathcal{C}_{\alpha d i})$  for a rating  $i$ .

We create a Slashdot user behavior model. We believe that antagonism has a significant effect on the connectivity of its sources and destinations. In this respect we test the following hypothesis:

**Hypothesis 1.** A **high rated** comment is **less** likely to create controversy when it is the **destination** of antagonism.

**Hypothesis 2.** A **low rated** comment is **more** likely to create controversy when it is the **destination** of antagonism.

**Hypothesis 3.** A **high rated** comment is **more** likely than a low rated one to create controversy when it is the **source** of antagonism.

A user is defined as an antagonism source  $u \in U_{\alpha s} \subset U$  if he has written an antagonistic comment:  $\alpha_{su} : U \rightarrow \{0, 1\}$ ;

$$u \mapsto \alpha_{su}(u) = \text{max}(p_\alpha(c) \mid \forall c \in \mathcal{C}_u(u)) \quad (5)$$

Let  $U_{\alpha s i, j} \subset U_{\alpha s}$  contain all users having  $i \leq r(u) \leq j$ , with  $\bar{C}_B(U_{\alpha s i, j})$  their mean centrality. Similarly, we denote the set of all users who are the destination of antagonism with  $U_{\alpha d} \subset U$  and  $U_{\alpha d i, j} \subset U_{\alpha d}$  contains those with the rating bounded by  $i$  and  $j$ . We correlate the value of their centrality with their participation in antagonistic exchanges:

**Hypothesis 4.** Users that have an antagonistic behavior are less central to the community than the ones who do not.

**Hypothesis 5.** The centrality of users who are the destination of antagonism is lower than the centrality of users who are sources of antagonism.

**Hypothesis 6.** Low ranking users that are also antagonistic become pariah, with the smallest possible centrality.

But, earlier, Hypothesis 2 stated that low ranking comments receive greater community feedback if they are the antagonized. Hypothesis 7 explains the short term effects of antagonism on the centrality of those involved:

**Hypothesis 7.** The centrality of antagonistic users increases shortly after the negative interaction.

Taken together, Hypotheses 4 and 7 show the evolution of community interest in a user involved in negative interactions. While at first his centrality increases (possibly as a result of angry direct responses), the community then avoids further interactions, which leads to a final low centrality.

We define short term using a *user dependent* definition of time. For each user  $u_i \in U_{\alpha d} \cup U_{\alpha s}$ , we study the centrality variation after their negative interactions. Let  $t_{0,i}$  be the moment the first negative interaction involving  $u_i$  occurs. Also, let  $t_0 = \min_i(t_{0,i})$  be the moment of the first negative interaction that involves any user. We mark the end of the experiment with  $t_{100}$ , the same for all users.

We divide each time interval  $(t_{0,i}, t_{100})$  into 100 equal parts and denote the start of the  $k$ 'th interval with  $t_{k,i}$ . We compute the centrality of  $u_i$ , at the moments  $t_{k,i}$ ,  $C_B^{t_{k,i}}(u_i)$  with a focus on low  $k$  values,  $k \in D_k = \{1, 2, 5, 10\}$ , and compare it with the starting centrality for that user,  $C_B^{t_{0,i}}(u_i)$ . We thus inquire whether the difference, expressed relative to the starting value, is positive  $\delta_{C_B} : U \times D_k \rightarrow \mathbb{R}$ :

$$\delta_{C_B}^k(u_i) = \frac{C_B^{t_{k,i}}(u_i) - C_B^{t_{0,i}}(u_i)}{C_B^{t_{0,i}}(u_i)} \times 100 \quad (6)$$

We define the centrality variation of a user set given time increment  $k$  as  $\delta_{C_B} : \mathcal{P}(U) \times D_k \rightarrow \mathbb{R}$ ,  $\delta_{C_B}^k(U') = \text{avg}(\delta_{C_B}^k(u_i)), \forall u_i \in U'$ , with  $\delta_{C_B}(U')$  the average over all considered  $k$  values. For different values of  $k$ , we inquire whether  $\delta_{C_B}^k(U_a)$  are positive. A further question is if the aggregated centrality variation is larger for antagonistic users,  $\delta_{C_B}(U_a)$ , than for non antagonistic ones,  $\delta_{C_B}(U_{n\alpha})$ .

We focus on the variation of the centrality, relative to the initial value, rather than in absolute terms. Let user  $u_A$  have a very high centrality and user  $u_B$  a near zero one,  $C_B(u_B) = \epsilon \ll C_B(u_A)$ . A centrality variation of  $\epsilon$  for  $u_B$  is very significant, while for  $u_A$  it is hardly noticeable. It is thus important to consider the variation in the context of the user's past, rather than focusing on a narrow time window.

The computational cost of determining the centralities of users at various moments is high. We perform the analysis on a subset of antagonistic users,  $U_\alpha \subset (U_{\alpha d} \cup U_{\alpha s})$ , and non antagonistic users  $U_{n\alpha} \subset U \setminus \{U_{\alpha d} \cup U_{\alpha s}\}$ . We focus on users who had previous activity before the negative comment in  $t_{0,i}$  and whose centrality at that moment is not zero.

## V. RESULTS

### A. Comment Level Analysis

A comment's reply depth is a measure of the controversy it causes. Figure 2 depicts the comment set **depth**, for the possible rating values,  $\bar{\delta}(C_i)$ . The connection is again not linear and we observe that engagement rises for extreme rating values. We can observe that controversy reaches a maximum

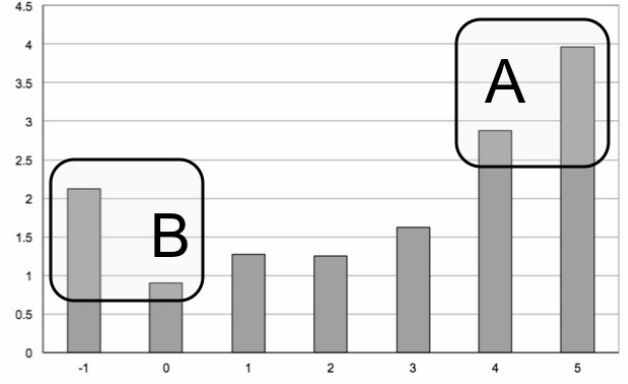


Fig. 2. Depth of a comment distributed by its rating

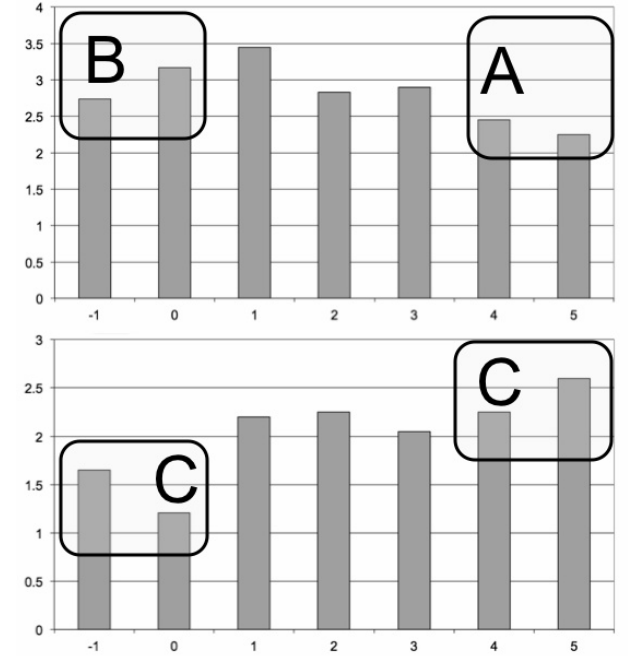


Fig. 3. Response depth distributed by comment rating - upper chart - antagonism destinations; lower one - sources

for the informative comments (rated 5), highly regarded by the community.

$$\bar{n}(C_5) \gg \bar{n}(C_{-1}) \gg \bar{n}(C_{0..2}) \quad (7)$$

$$\bar{\delta}(C_5) \gg \bar{\delta}(C_{-1}) \gg \bar{\delta}(C_{0..2}) \quad (8)$$

The bottom chart in Figure 3 shows the average depth for comments that are antagonism sources, and have a given rating  $i$ ,  $\bar{\delta}(C_{\alpha s i})$ , and the upper one for antagonism destination comments,  $\bar{\delta}(C_{\alpha d i})$ . We compare with the respective average depth for all comments,  $\bar{\delta}(C_i)$ , shown in Figure 2.

We find that for highly rated comments  $i \in \{4, 5\}$  the controversy they create decreases if they are antagonized:

$$\bar{\delta}(C_{\alpha d 4,5}) < \bar{\delta}(C_{4,5}) \quad (9)$$

This validates Hypothesis 1 and shows that overt criticism is destructive for pertinent arguments. In a constructive argument,

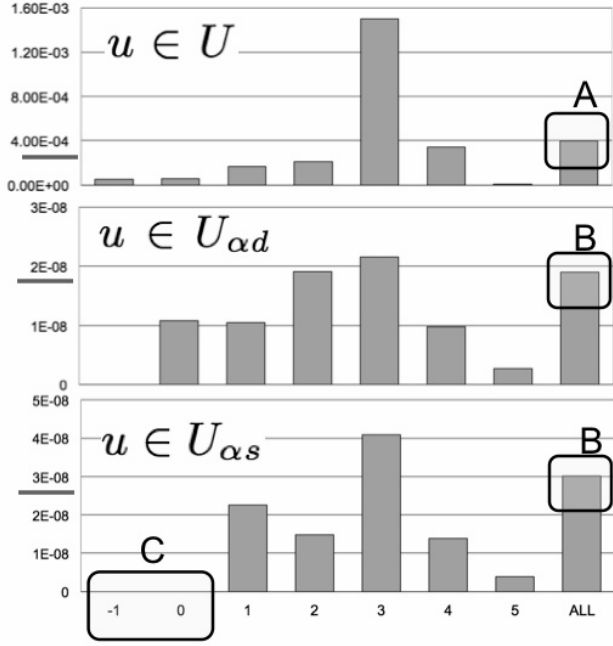


Fig. 4. Centrality distributed by user rating, for all users (top), antagonism targets (middle), and sources (bottom)

politeness is thus the key. The fact that the two pairs of means are different is backed by two significant p-values:  $1 \times 10^{-8}$  for  $i = 4$  and  $1 \times 10^{-5}$  for  $i = 5$ .

The boxes labeled B in Figure 3 show that disagreeing with impertinent users *significantly increases* the feedback they receive. The associated p-values are  $1 \times 1^{-8}$  for  $i = -1$  and  $1 \times 1^{-10}$  for  $i = 0$ . An interpretation of Hypothesis 2 is that we *shouldn't feed the trolls*.

$$\bar{\delta}(C_{\alpha d-1,0}) > \bar{\delta}(C_{-1,0}) \quad (10)$$

A third observation, that supports Hypothesis 3, is that constructive arguments create more controversy than less constructive ones. The boxes marked C in in Figure 3 show that comments rated 4 and 5 which criticize others have a higher depth than comments rated -1 and 0.

$$\bar{\delta}(C_{\alpha s4,5}) > \bar{\delta}(C_{\alpha s-1,0}) \quad (11)$$

The difference is shown to be significant by the p-value of  $4.81 \times 10^{-3}$ .

### B. User Level Analysis

We prove a user's centrality depends heavily on whether he emits or is targeted by antagonism. Figure 4 presents the betweenness centrality in three different settings. The top chart portrays all users,  $\overline{C}_B(U_{i,j})$ , the middle one antagonistic comment destinations,  $\overline{C}_B(U_{\alpha d i,j})$ , and the bottom one the users that are sources of antagonism,  $\overline{C}_B(U_{\alpha s i,j})$ . The rightmost bars show the centrality values for all the users, regardless of their rating: top -  $\overline{C}_B(U)$ , middle -  $\overline{C}_B(U_{\alpha d})$ , bottom -  $\overline{C}_B(U_{\alpha s})$ .

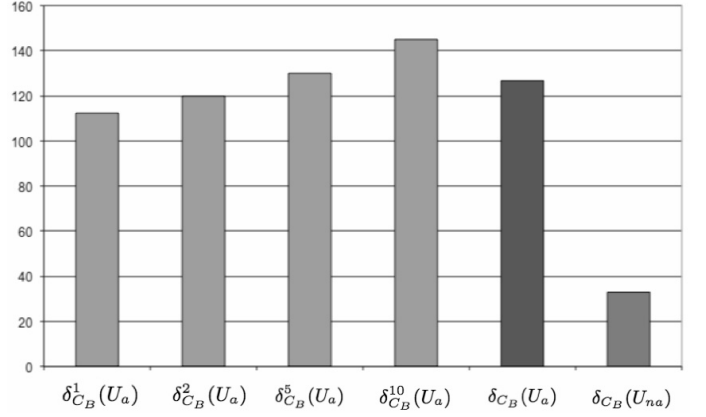


Fig. 5. Short term centrality variation of antagonistic and regular users

A first observation, shown in boxes labeled B versus the box labeled A, supports Hypothesis 4. We prove that antagonistic users have a much lower centrality than the average:

$$\overline{C}_B(U) \approx \overline{C}_B(U_{\alpha d}) \times 10^4; \overline{C}_B(U) \approx \overline{C}_B(U_{\alpha s}) \times 10^4 \quad (12)$$

The differences are backed by highly significant Welch test p-values of  $6.9 \times 10^{-7}$  for antagonism sources and  $6.54 \times 10^{-4}$  for destinations.

A second user level observation is that user centrality is significantly higher if you emit a negative opinion than if you are being criticized, as in Hypothesis 5. The values highlighted in the boxes labeled B on Figure 3 show that

$$\overline{C}_B(U_{\alpha s}) = \overline{C}_B(U_{\alpha d}) \times 1.59 \quad (13)$$

This finding is however less significant statistically, as the p-value obtained is only  $9.57 \times 10^{-2}$ .

A third observation, that supports Hypothesis 6, is that giving negative explicit feedback when having a low average score leads to marginalization. The null subsequent engagement values are highlighted in box C in Figure 4.

1) *Centrality Variation:* A final experiment regards the short term variation of antagonistic user centrality. As previously mentioned, the experiment was performed on two sets of users: antagonistic  $|U_{\alpha}| = \{112\}$  and non antagonistic ones  $|U_{na}| = \{133\}$ . The first four columns of Figure 5 portray the values of  $\delta_{C_B}^k(U_{\alpha})$ , with  $k$  varying from 1 to 10. All four values verify the claim in Hypothesis 7, that, in the short term, antagonistic users increase their centrality. We observe that, on average, they more than double their centrality for a short time period after the negative interaction.

We compare the aggregate variation value for antagonistic users,  $\delta_{C_B}(U_{\alpha})$  with the corresponding one for regular users,  $\delta_{C_B}(U_{na})$ . Results show that the short term average centrality variation for antagonistic users is almost four times larger than for regular users (126.7 versus 32.9 per cent). This result is backed by a p-value of  $1 \times 10^{-5}$ .

## VI. CONCLUSIONS

The main contribution of this work is the study of antagonism in online discussions. We defined antagonistic behavior as the presence of negative opinions regarding the author of

the previous comment. Results showed that its presence has a profound impact on the community perception and on the importance within the network of both the source and the destination.

The response to a comment is influenced by being antagonized, which only helps low rated posts. Regarding the users, we found that being rude and having an antagonistic stance leads in the end to social marginalization. This effect is more visible for users who are targeted by negative opinions than for users who generate them. Another observation is that user centrality only decreases for antagonistic users after a long period of time, while the initial effect is reversed.

We used these observations to create a behavior model of people posting online comments, that was validated using real world data from Slashdot. We studied two intuitions about the usefulness of traditional network oriented methods and the proposed content based one. We have shown that, for the considered network, including opinion related information into the analysis of social networks improves the understanding of the analyzed societies.

We believe the impact of our work is threefold. We introduce a new method of analyzing attitudes in social network based on aspect-based sentiment analysis. The method can help differentiate positive links from negative ones, leading to a finer grained analysis of the network.

The presented results are a case against indiscriminately using network measures, like the betweenness centrality. We hope this will increase the efforts put into finding other ways to define connections between social actors. We have shown that text based connections are a reliable alternative.

A direct application of this work is the automatic moderating of discussion boards. Some systems employ rating self-moderating schemes, but many rely solely on costly administrator moderating. Others, like Slashdot, use a hybrid approach. An automatic moderating tool can greatly reduce costs and improve the user experience. We have shown that using user centrality values for moderating is only helpful after a significant period of time. Alternative, instantaneous, methods are needed and we proposed two text based ones - the topic relevance of posts and their antagonism content.

## REFERENCES

- [1] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 29–42.
- [2] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, November 1999. [Online]. Available: <http://ilpubs.stanford.edu:8090/422/>
- [3] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 7–15.
- [4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *ICWSM 10: Proceedings of international AAAI Conference on Weblogs and Social Media*, 2010.
- [5] J. Diesner and K. Carley, *Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis.*, ser. Causal Mapping for Information Systems and Te. Harrisburg, PA: Idea Group, 2005, ch. 4.
- [6] S. Wang and P. Groth, *Measuring the dynamic bi-directional influence between content and social networks*. Springer, 2010, vol. 6496, pp. 814–829. [Online]. Available: <http://www.springerlink.com/index/WV88V4666P3Q2Q61.pdf>
- [7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Empirical methods in natural language processing-Volume 10*. ACL, 2002, pp. 79–86.
- [8] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 1397–1405. [Online]. Available: <http://doi.acm.org/10.1145/2020408.2020614>
- [9] M. Miller, C. Sathi, D. Wiesenhal, J. Leskovec, and C. Potts, "Sentiment flow through hyperlink networks," in *ICWSM 11, Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 2011.
- [10] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 741–750.
- [11] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 641–650.
- [12] A. Stavrianou, J. Velcin, and J.-H. Chauchat, "A complementary model to the social networks for mining forums," in *From Sociology to Computing in Social Networks*, 2010, pp. 59–79.
- [13] K. S. Cho, J.-S. Ryu, J.-H. Jeong, Y.-H. Kim, and U.-M. Kim, "Credibility evaluation and results with leader-weight in opinion mining," *2010 International Conference on Distributed Computing and Knowledge Discovery*, pp. 5–8, 2010. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5615503>
- [14] B. He, C. Macdonald, and I. Ounis, *Ranking opinionated blog posts using OpinionFinder*. ACM Press, 2008, vol. 31, pp. 727–728. [Online]. Available: <http://dx.doi.org/10.1145/1390334.1390473>
- [15] V. Gómez, A. Kaltenbrunner, and V. López, "Statistical analysis of the social network and discussion threads in slashdot," in *Proceedings of the 17th international conference on World Wide Web*. New York, USA: ACM, 2008, pp. 645–654. [Online]. Available: <http://doi.acm.org/10.1145/1367497.1367585>
- [16] A. Stavrianou, J. Velcin, and J.-H. Chauchat, "Definition and measures of an opinion model for mining forums," in *ASONAM*, 2009, pp. 188–193.
- [17] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "Opinionfinder: a system for subjectivity analysis," in *Proceedings of HLT/EMNLP on Interactive Demonstrations*. Stroudsburg, PA, USA: ACL, 2005, pp. 34–35. [Online]. Available: <http://dx.doi.org/10.3115/1225733.1225751>
- [18] M.-C. de Marneffe, B. MacCartney, and C. D. Manning, "Generating typed dependency parses from phrase structure trees," in *LREC*, 2006. [Online]. Available: [http://nlp.stanford.edu/pubs/LREC06\\_dependencies.pdf](http://nlp.stanford.edu/pubs/LREC06_dependencies.pdf)