Antagonism also Flows through Retweets: The Impact of Out-of-Context Quotes in Opinion Polarization Analysis

Pedro Calais Guerra, Roberto C.S.N.P. Souza, Renato M. Assunção, Wagner Meira Jr.
Dept. of Computer Science – Universidade Federal de Minas Gerais (UFMG)
{pcalais,nalon,assuncao,meira}@dcc.ufmg.br

Abstract

In this paper, we study the implications of the commonplace assumption that most social media studies make with respect to the nature of message shares (such as retweets) as a predominantly *positive* interaction. By analyzing two large longitudinal Brazilian Twitter datasets containing 5 years of conversations on two polarizing topics – Politics and Sports, we empirically demonstrate that groups holding antagonistic views can actually retweet each other *more often* than they retweet other groups. We show that assuming retweets as endorsement interactions can lead to misleading conclusions with respect to the level of antagonism among social communities, and that this apparent paradox is explained in part by the use of retweets to quote the original content creator out of the message's original temporal context, for humor and criticism purposes.

Introduction

In this paper, we study the implications of the commonplace assumption that most social media studies make with respect to the nature of message shares (such as retweets) as a predominantly *positive* interaction. Given that on general purpose social platforms such as Facebook and Twitter there are no explicit positive and negative signs encoded in the edges, it is commonly assumed (in general, implicitly) that a connection among users through message shares indicate increased homophily among them (Guerra et al. 2011; Conover et al. 2011). In general, studies of polarized online communities induced by topics such as Politics and public policies do not conduct any explicit analysis of antagonism at the edge granularity, and the degree of separation between communities as well as the controversial nature of the topic is accepted as sufficient evidence of polarization (Garimella et al. 2016). We provide a qualitative and quantitative analysis on the use of retweets as *negative* interactions. In particular, we analyze two large Brazilian Twitter datasets on polarizing topics – Politics and Soccer – which lead us to two main findings related to behavioral patterns on social-media based interactions:

1. Antagonistic communities tend to share each other's content *more often* than they share content from other less polarizing and conflicting groups. The immediate consequence of this observation is that a simplistic consideration of retweets as an endorsement interaction can lead to misleading conclusions with respect to the nature and polarity of group relationships, as a large number of retweets flowing from one community to another may be misinterpreted as a signal of support.

2. We observe retweets employed as a mechanism for *quoting out of context*, a known strategy of reproducing a passage or quote out of its original context with the intent of distorting its intended meaning (McGlone 2005). In particular, we found that Twitter users share old messages posted by someone from an opposing side with the goal of creating irony when putting the message out of its original temporal context. We observed that some messages are broadcasted even 6 years after they have been originally posted, with the intention of reinforcing an antagonistic and contrary position, rather than indicating support. In our datasets, a significant fraction of retweets crossing antagonistic communities are out of context retweets.

We believe the main reason these findings on the use of retweets to convey disagreement remain unnoticed in the social network analysis literature is due the focus on research on bipolarized social networks, characterized by the emergence of *exactly* two dominant conflicting groups, such as republicans versus democrats (Adamic and Glance 2005), pro and anti gun-control (Guerra et al. 2013), and pro-life versus pro-choice voices. In this setting, once you determine (automatically or by manual examination) the leaning of a group toward a controversial topic, their (negative) opinion w.r.t. the opposite viewpoint is implicitly determined, and no further analysis of edge polarities is usually performed.

Our work contributes to social media research in two distinct directions. Finding 1 adds to the recent trend on the pitfalls and drawbacks of making inferences based on social media data (Rost et al. 2013). Finding 2, on the other hand, explore how temporal information associated to retweets can be a rich signal to be incorporated into models focused on antagonism detection and real-time tracking of opinions in social media.

Related Work

On social networks whose edge signs are labeled, antagonistic relationships among communities are naturally reflected by the number of positive and negative edges flowing from the source community to a target community, and the communities themselves can be found by algorithms especially designed to deal with negative edges (Kunegis et al. 2010).

Many works qualitatively discuss and document the empirical observation that unlabeled social interactions on general purpose social platforms such as Twitter and Facebook can convey negative sentiment: replies and comments, as web hyperlinks, do not carry an explicit sentiment label and can be either positive or negative (Leskovec, Huttenlocher, and Kleinberg 2010). Message broadcasts, on the other hand, have been categorized by early works on behavioral analysis on Twitter as a strictly positive interaction (Boyd, Golder, and Lotan 2010). As users expertise evolved, they had begun finding uses of retweets that do not convey agreement. "Retweets are not endorsements" is a common disclaimer found in biographies of journalists and think tankers in Twitter, whereas some people share stuff that they vehemently disagree only to show the idiocy of the people they oppose. In summary, retweets and shares are often a "hate-linking" strategy – linking to disagree and criticize, often in an ironic and sarcastic manner, rather than to endorse (Tufekci 2014).

Although documented in the literature as a known behavior, the impact of such "negative" retweets on community and network analysis has not been the focus of in-depth studies so far. Usually, social network analysis practitioners assume, implicitly or explicitly, that retweets (or more generally, shares) have a predominant endorsement nature. A recurrent pattern in community analysis works making sense of social media datasets is that they limit their analysis to social networks whose dominant topic induces a partition of the graph into exactly two conflicting sides: liberal versus conservative parties, pro-gun and anti-gun voices, pro-choice and pro-life (Conover et al. 2011; Adamic and Glance 2005). As we will show in the next sections, in bipolarized scenarios, it is harder to grasp the use of retweets to convey disagreement.

Data Collection and Preparation

We used Twitter's Streaming API to monitor two topics that motivate intense debate on offline and online media and thus are suitable for analysis of formation of antagonistic communities: Politics and Sports. Table 1 provides details.

In the political topic, our data collection was driven by the main candidates in the 2010 and 2014 Brazilian presidential elections, including Dilma Rousseff, elected for the presidency in both years. We monitored mentions to politician Twitter profiles and names, the hashtags used by each side participating in the political debate and the names of the presidents of the Brazilian Lower House and the Senate, which directly conducted Ms. Rousseff's impeachment process in the Congress. We also collected public tweets about the 2010 to 2016 editions of the Brazilian Soccer League. We monitored mentions to the 12 largest Brazilian soccer teams and match-related keywords ("goal", "penalty" and "yellow card", etc)

Community detection. Once collected we prepared the data for our various analysis as described next. The first step is to partition the social network induced by the messages and represented as a graph $G(V, E)$ into meaningful communities. In the case of the Twitter datasets we take into consideration, the official profiles of politicians, political parties and soccer clubs are natural seeds that can be fed to a semi-supervised clustering algorithm that expands the seeds to the communities formed around them (Guerra et al. 2011).

Table 1: General description of the two Twitter datasets we consider. Note the large variability on (native) retweet response times.

	Topic	
	Politics	*Soccer*
period	2010-16	2010-16
# groups	3	12
# tweets	20.5 M	103M
# users	3.1M	8.7M
manual RTs	46K	2K
quote RTs	67K	3K
native RTs	9.1M	30.9M
RT mean response time (hours)	29.5h	43.5h
RT median response time (hours)	0.24h	0.23h
RT response time std (hours)	255.4h	368.7h

Different graphs can be built based on the datasets described in Table 1; traditionally, a social network $G(V, E)$ represents a set of users V and a set of edges E that connect two users if they exceed a threshold of interaction activity. The limitation of this modeling is that it hides the individual user-message interactions; by representing interactions in a user-message bipartite retweet graph, as shown in Figure 1, we keep this more granular information.

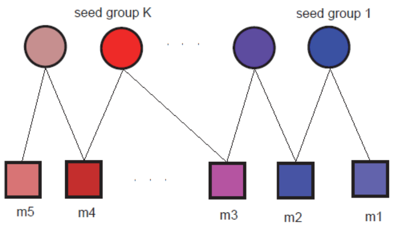

Figure 1: A bipartite user-message graph connecting users with messages they interact with. Node colors represent relative proximities to the the red/blue sides.

We assume the number of communities K formed around a topic T is known in advance and it is a parameter of our method. To estimate user and message leanings toward each of the K groups, we employ a label propagation-like strategy based on random walk with restarts (Tong, Faloutsos, and Pan 2008): a random walker departs from each seed and travels in the user-message retweet bipartite graph by randomly choosing an edge to decide which node it should go next. With a probability (1 - α) $= 0.85$, the random walker restarts the process from its original seed. As a consequence, the random walker tends to spend more time inside the cluster its seed belongs to. Each node is then assigned to its closest seed (i.e., community), as shown in the node colors in the toy example from Figure 1. For more details, please refer to (Guerra et al. 2011).

Finding 1: antagonistic groups retweet each other more than they retweet other groups

The intrinsic limitation of a bipolarized network is that only one separation metric value can be computed, since there is only one pair of communities. Since we are studying $K > 2$ cases, we now have $\binom{K}{2}$ pairwise community metrics to compare. For the sake of simplicity, for each pair of communities we compute the proportion of retweets triggered from users belonging to community i that flow toward messages posted by members of community j relative to all retweets that community i trigger to the other groups in the graph:

$$RT\_ratio(i,j) = \frac{RT_{i,j}}{\sum\limits_{k=1,k\neq i}^{K} RT_{i,k}} \quad (1)$$

We compare $RT\_ratio(i,j)$ considering the known local rivalries that exist in Brazilian Soccer among soccer clubs from the same Brazilian state, as listed in Table 2.

Table 2: Local rivalries in Brazilian Soccer.

Brazilian state	local rivalries
M. Gerais	Cruzeiro, Atlético
S. Paulo	SPFC, Santos, Corint., Palmeiras
R. G. do Sul	Grêmio, Internacional
R. de Janeiro	Flamengo, Flumin., Vasco, Botafogo

In Figure 2 we plot $RT\_ratio(i,j)$ for all the $\binom{K}{2}$ pairs of communities formed around supporters of Brazilian soccer clubs, and we visually discriminate between pairs of rival communities (red triangles) and non-rival communities (green circles) according to the ground truth from Table 2. The graph shows a somewhat unexpected result: pairs of communities that are *more antagonistic* (i.e., the opposing sides belong to the same Brazilian state) tend to retweet each other's content *more often* than when there is less, or no antagonism between them. For example, Cruzeiro's community (id = 8) targets about 65% of its cross-group retweets to Atlético's community, their sole fierce rival in Brazilian state of Minas Gerais. As another example, community 1, which identifies supporters from Rio de Janeiro team Flamengo, prefers to retweet messages for their three local rivals. As a general rule, red triangles dominate green circles, i.e., retweets are targeted more often to antagonistic communities than to more neutral, less conflicting groups.

The fundamental insight to learn from Figure 2 is that retweets carrying a negative polarity directly impact the network structure and make antagonistic communities *closer* in the social graph. On traditional bipolarized domains in which current literature focuses, this apparent paradox is inherently unnoticeable, since there is only a single pair of antagonistic communities and thus only a single separation metric to be computed.

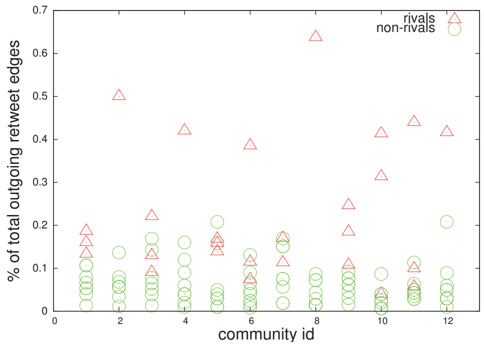

Figure 2: $RT\_ratio(i,j)$ for each pair of 12 communities discussing Brazilian soccer in Twitter.

Finding 2: out-of-context retweets are more prevalent on cross-group relationships

We now focus our analysis on the retweet response time – the time interval between the original message posting time and the retweet time. Earlier studies have related very short and very long retweet response times to fraudulent activity to boost user popularity (Giatsoglou et al. 2015); our goal is to analyze retweet response time under the perspective of the message polarity and the polarity that the user broadcasting the message is attempting to convey.

In Figure 3 we plot the cumulative distribution of retweet response times, measured in seconds. We plot this distribution for internal (intra-community) and cross-group (inter-community) retweets for both the Soccer and Politics dataset. Notice that cross-group retweets tend to occur later when compared to internal retweets. For instance, at least 30% of retweets connecting groups in both datasets occur after 16 hours of the original message posting time; on the other hand, in the case of internal retweets, only 10% of retweets occur temporally far from the original post.

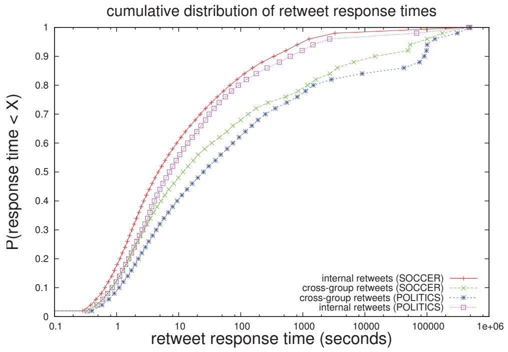

Figure 3: On average, retweets which cross antagonistic communities tend have larger response times than intercommunity retweets.

We now take a closer look at some messages. For instance, consider the following tweet posted by the official account of the Brazilian elected vice-president Michel Temer about a speech given on TV by his presidential candidate, Dilma Rousseff, during the 2010 Presidential Elections:

> 2010-08-05 11:11 PM: @MichelTemer: *Dilma is displaying confidence and knowledge.*

Six years after this post, President Rousseff has been suspended by the Brazilian Congress following an impeachment trial of misuse of public money. In response, she gave a speech on March 12th, 2015 accusing VP Temer's party (PMDB) to plan a coup against her. During her speech, many users contrary to Rousseff began retweeting Temer's 2010 message:

> 2016-05-12 12:23 AM: @randomRousseffOppositor: RT @MichelTemer: *Dilma is displaying confidence and knowledge.*

This is a clear attempt to retweet a message attaching to it a negative connotation; it does not support nor endorse its original content. On the contrary, the recent retweeters of this message attach to it a semantics which is exactly the opposite to the one stated in the direct interpretation of the message, what is precisely the definition of irony (Wallace 2013). While the "contextomy" practice usually refers to selecting specific words from their original linguistic context (McGlone 2005), we see that, in Twitter, such change of meaning is usually associated with some temporal evolution.

Conclusions

In this paper we explore the observation that, in the vast majority of social media studies, especially those based on Facebook and Twitter data, there is no explicit positive and negative signs encoded in the edges. Since inferring individual edge polarities in a unsigned graph is not a trivial task, most social studies assume that retweets and shares are endorsement interactions. No specific analysis on the polarity of the links crossing the communities is usually conducted and antagonism is assumed due to the modular division of the social graphs into two communities historically known to be antagonistic, such as democrats and republicans.

Although very recent papers on retweeting activity still qualify retweets as a strictly positive interaction (Garimella et al. 2017; Metaxas et al. 2015), we show that retweets can actually carry a negative polarity, conveying a sentiment which is opposite to the one explicited in the tweet's text. We believe the neglected impact of negative retweets explain, in part, the low accuracy levels obtained in some user polarity classification experiments (Cohen and Ruths 2013).

We found that one of the reasons that motivate Twitter users to broadcast tweets they disagree with is to create irony by broadcasting a message in a different temporal context, especially when a real-world event that disproves the original message argument happens. Such behavior finds similarity on *quoting out of context*, a practice already described in the Communications literature (Boller and George 1989).

We believe the better understanding of retweets as multifaceted social interactions which can be (1) possibly negative and (2) have a temporal component may support the design of algorithms that exploit the network structure in conjunction with opinionated content to better perform tasks typically offered by social media platforms, such as content recommendation, event detection, sentiment analysis and news curation.

Acknowledgments

This work was supported by CNPQ, Fapemig, InWeb, MASWeb, BIGSEA and INCT-MCS.

References

Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 u.s. election: divided they blog. In *LinkKDD*, 36–43.

Boller, P. F., and George, J. H. 1989. *They never said it : a book of fake quotes, misquotes, and misleading attributions*. Oxford University Press New York.

Boyd, D.; Golder, S.; and Lotan, G. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *43rd HICSS*.

Cohen, R., and Ruths, D. 2013. Classifying political orientation on Twitter: Its not easy! In *ICWSM*.

Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Flammini, A.; and Menczer, F. 2011. Political polarization on Twitter. In *5th ICWSM*.

Garimella, K.; De Francisci Morales, G.; Gionis, A.; and Mathioudakis, M. 2016. Quantifying controversy in social media. In *9th WSDM*.

Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2017. Balancing opposing views to reduce controversy. In *10th WSDM*.

Giatsoglou, M.; Chatzakou, D.; Shah, N.; Faloutsos, C.; and Vakali, A. 2015. Retweeting activity on twitter: Signs of deception. In *19th PAKDD*.

Guerra, P. H. C.; Veloso, A.; Meira, Jr, W.; and Almeida, V. 2011. From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In *17th SIGKDD*.

Guerra, P. H. C.; Jr., W. M.; Cardie, C.; and Kleiberg, R. 2013. A measure of polarization on social media networks based on community boundaries. In *7th ICWSM*.

Kunegis, J.; Schmidt, S.; Lommatzsch, A.; Lerner, J.; Luca, E. W. D.; and Albayrak, S. 2010. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proc. SIAM Int. Conf. on Data Mining*, 559–570. SIAM.

Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010. Predicting positive and negative links in online social networks. In *19th WWW*.

McGlone, M. S. 2005. Contextomy: the art of quoting out of context. *Media, Culture & Society* 27(4):511–522.

Metaxas, P. T.; Mustafaraj, E.; Wong, K.; Zeng, L.; O'Keefe, M.; and Finn, S. 2015. What do retweets indicate? results from user survey and meta-review of research. In *9th ICWSM*.

Rost, M.; Barkhuus, L.; Cramer, H.; and Brown, B. 2013. Representation and communication: Challenges in interpreting large social media datasets. In *CSCW*.

Tong, H.; Faloutsos, C.; and Pan, J.-Y. 2008. Random walk with restart: fast solutions and applications. *Knowl. Inf. Syst.* 14(3):327–346.

Tufekci, Z. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *8th ICWSM*.

Wallace, B. 2013. Computational irony: A survey and new perspectives. *Artificial Intelligence Review* 1–17.