Analyzing polarization in Twitter: The murder of Brazilian councilwoman and activist Marielle Franco

Abstract. Social media has allowed people to publicly express, at near zero cost, their opinions and emotions on a wide range of topics. This recent scenario allows the analysis of social media platforms to several purposes, such as predicting elections, exploiting influential users or understanding the polarization of public opinion on polemic topics. In this work, we analyze the Brazilian public perception related to the murder of a Rio councilwoman, Marielle Franco, member of a left-wing party and human-rights activist. We propose a polarity score to capture whether the tweet is positive or negative and then we analyze the score evolution over the time, after the murder. Finally, we evaluate our approach correlating the polarity score with human judgment over a randomly sampled set of tweets. Our preliminary results show how to measure polarity on public opinion using a weighted dictionary and how it changes over time.

Keywords: Social media, Sentiment analysis, Opinion Mining.

1   Introduction

Over the last years, an increasing number of people actively use the Internet to exchange information and convey emotions, allowing studies that examine how technology can influence people's feelings. Social media platforms have become an important source of data to capture those emotions, opinions and sentiments on several topics and debates – from the Mediterranean refugee's crisis (Coletto et al., 2016) to political leanings (Conover et al., 2011) (Tumasjan et al., 2010).

Twitter platform is especially powerful because of its very nature, which encourages people to have public conversations and debates, sharing their thoughts with others, creating solidary networks or politically engaged movements. However, such platforms also encourage public demonstrations of negative emotions, often with hate speeches, which discriminate against people's race, religion, ethnicity, gender, political views (Silva et al., 2016).

In this paper, we investigate the public perception, applying a polarity score extracted from tweets, of the recent murder of Rio councilor and activist Marielle Franco, on March 2018, in the city of Rio de Janeiro. Marielle was a black woman, bisexual, feminist, human rights activist and was born in the *Maré* favela, a low-income community in Rio de Janeiro. She was member of the left-wing party Socialism and Liberty party (PSOL) and an outspoken critic of the endemic police violence in Rio's favelas. The crime shocked not only Brazil but the whole world and remains unsolved.

We propose a simple yet effective method to classify tweets mentioning Marielle's case, based on a weighted dictionary of words and expressions that capture the polarization – *positive* or *negative*, to better understand the impact of events related to the crime on public opinion. Then we analyze how polarization changed in time, after the

murder. Finally, we evaluate the experiment on correlating a random sample of polarized tweets with human judgment.

2 Related work

Over the last years, the interest in Sentiment Analysis and Opinion Mining using social media such as Twitter has been increasing rapidly.

Some works surveys the techniques and approaches that addresses the new challenges raised by sentiment-aware applications, as compared with more traditional fact-based analysis (Pang et al., 2008). Combined methods mix lexicon-based and machine-learning based, in order to classify Twitter messages (Kolchyna et al., 2015). POS-specific prior polarity features are also exploited on sentiment analysis on Twitter data (Agarwal et al., 2011). Twitter is also used to understand the political sentiment on public opinion, for instance, on predicting the political alignment of Twitter users based on the political communication in the run-up to the 2010 U.S midterm elections (Conover el at., 2010). Other works exploit popular or influential users on Twitter, investigating the positive-negative influence measured between popular users and their audience (Bae et al., 2012) or comparing measures of influence, such as indegree, retweets, and mentions (Cha et al., 2010).

The contributions of this work are the following: (i) the creation of a weighted dictionary, in Portuguese, with *positive* and *negative* terms related to the topic, that can be profitably reused in other contexts; (ii) an analyze of the Marielle's murder impact on public opinion and how events changed the polarity over the period of time analyzed.

3 Data collection

We used the Twitter Standard Search API to gather tweets mentioning Marielle. We manually choose, as search input, some general keywords related to the topic *'#marielle', '#mariellefranco',* and also some hashtags frequently used to express support as *'#mariellevive'* (marielle lives)*, '#naofoiassalto'* (it was not a robbery)*, '#mariellepresente', '#mariellefrancopresente', '#todospormarielle'* (everyone for Marielle)*, '#justicapormarielle'* (justice for Marielle)*, '#pormarielleeanderson'* (for Marielle and Anderson, her car driver which was also murdered). We did not find explicitly negative hashtags related to the case. From the total tweets we collected, we filtered out only tweets written in Portuguese, most present language in the initial dataset (545,116 tweets, 73% of the total), as shown in Table 1.

Table 1. Summary of the dataset.

Description	# Total
Total tweets collected	794,890
Portuguese tweets	545,116
Distinct users	185,821

Tweets were collected comprising a period of 50 days, from 14th March to 9th May, with a 6 days gap (from 21th March to 24th March and from 24th April to 25th April) - due to infrastructure problems. Fig. 1 depicts the number of tweets collected along the days. The day with more mentions to the topic was the 15th March, the day just after the murder, with 74,418 tweets (the crime happened in the night of the 14th March).
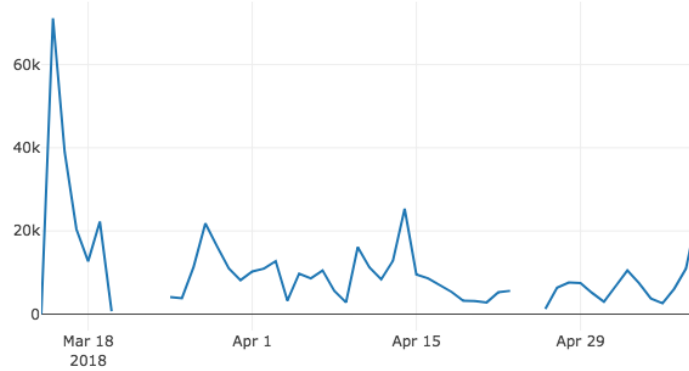
Fig. 1. Tweets collected per day from 14th March to 9th May 2018.

4 Polarizing tweets with a weighted dictionary

We are interested in finding whether a tweet express a *positive*, *negative* or *neutral* sentiment related to Marielle's case. In order to classify the tweets into these three classes, we take advantage of a dictionary with negative and positive terms, related to the topic.

The tweet polarity score p captures the tweet perception related to Marielle's case and rely on the tweet terms that match the dictionary terms as defined as following:

$$p = (f_t . w_t) \quad \text{for all } t \in T \text{ if } t \in D \tag{1}$$

where

- t is a term found in the tweet.
- T is the set of terms found in the tweet.
- D is the dictionary with positive and negative terms t and weights w.
- f_t is the frequency of each term t found in the tweet.
- w_t denotes the weight of t. If the term express negative feelings, w_t is negative, otherwise, w_t is positive

When p is zero, meaning that the tweet terms do not match any dictionary terms, the tweet is considered as *neutral*.

Initially, we manually created an initial dictionary (or *seeds*) with obvious terms, in Portuguese, representing general insults, outrages, and hate speeches. We ranked the tweets in ascendant order of polarity score p, in order to examine the tweets with more negative polarity. We found additional terms insulting her as bisexual woman (with misogynistic/homophobic language); swearing her party, offending her ideals as human

rights activist and even terms expressing a revenge feeling, implying that she deserved to die.

Later, we analyzed the most positive tweets, on ranking p in descendent order. We found very positive words supporting the councilwoman, her party, and also claiming the investigators for justice. Table 2 shows 6 out of the 102 terms of the dictionary, its weights and its frequency in the dataset, i.e, the number of occurrences.

Table 2. Weighted dictionary with terms and weights.

term	translation	# weight	# occurrences
semente	seed	+2	3492
lute como Marielle	fight as Marielle	+2	895
descanse em paz	rest in peace	+1	511
abortista	abortionist	-2	31
vereadora irrelevante	irrelevant councilwoman	-1	334

The term *semente* means *seed* (used to refer to Marielle's legacy, on inspiring other black and poor women on political career). The depreciative term *abortista* means *abortionist* (she was a pro-choice councilwoman). We considered the terms root (or *stems*), to capture all its variations, since we do not distinguish their categories – such as noun, adjective and verb. Table 3 shows some tweets, and their polarity scores. The symbol & refers to the logical AND (both terms have to be present and *vtnc* is a short for 'vai tomar no cu', means *fuck off*.

Table 3. Tweets with final polarity scores.

Tweet	polarity score p	terms and weights
pensei que fosse para mandar **canonizar** *a falecida* **vereadora** *comunista e* **protetora** *de* **bandidos**.	- 5	canoniz & vereadora -3 prote & bandid -2
#***VTNC*** *Essa vereadora era uma* **canalha** *que de-fendia bandidos*!	- 4	defen & bandid -2 vtnc -1 canalha -1
Dia triste *para quem tinha esperanças de uma pá-tria justa, igualitária e democrática. Profunda-mente* **arrasada** *com essa notícia. O racismo e a intolerância assumindo formas dantescas. preci-samos fazer a nossa voz ecoar.*	+ 2	dia & triste +1 arrasad +1
Tentaram calar *#MartinLutherKing e #Marielle-Franco, não sabiam que ambos são* **sementes**. *Es-tamos florescendo, mestres*	+ 3	semente +2 tentaram & calar +1

It is important to notice that we found several false negatives, classified as negative but actually positive. For instance, the tweet ('*as pernas chegam a tremer quando escuto um pessoal falando que Marielle defendia bandido', translated: my legs shake when I hear people saying that Marielle protected criminals*) got initially -2 as polarity score, because it contains the negative term *defend* & *bandid*. However, the user was actually criticizing someone else' speech, that, in turn, would be classified as negative. These sentences often include common expressions inferring that the opinion stated is not of that user, but from someone else (for instance, '*um pessoal falando que*'). As a turnaround to this challenge, we applied a simple heuristic: when we found these common sentences, we inverted the tweet final polarization (-2 turned 2 in this example, so the negative polarity turned positive).

We then computed the total polarity score for each day analyzed in the period, summing the polarity scores of all tweets of the 50 days analyzed, after Marielle's murder. Figure 2 shows (i) the total positive scores on the given days (blue bars), (ii) the total negative scores on the given day (red bars), and (iii) the final polarity scores on each day (black line). For most of the analyzed days, the positive p scores were overwhelmingly superior to the total negative p scores, i.e., tweets were much more positive than negative. However, there were few days when the negative scores overpassed the positive ones.
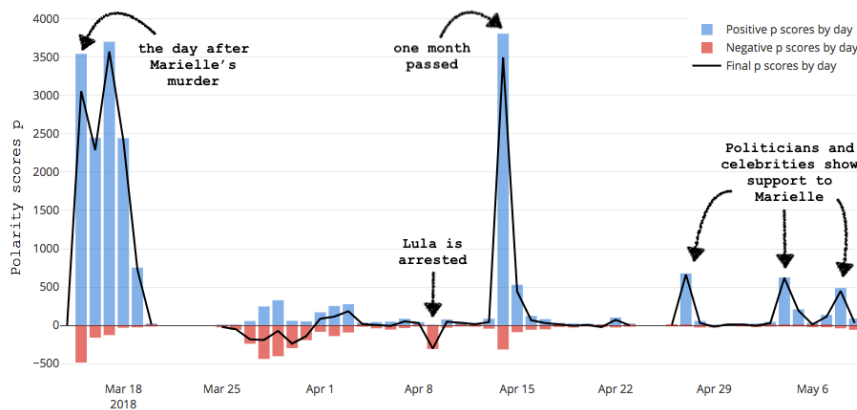

Fig. 2. Polarity scores over the period of time analyzed.

The five days following Marielle's murder were marked by a majority of supportive tweets, albeit that several fake news stating that Marielle was married with a traffic dealer were being spread from the day after the murder, which very likely increased the negative scores on that period, as can be seen in Figure 2. We identified also several tweets mentioning the Brazilian ex-president Lula and his arrest (that happened on 7th April 2018). A possible reason for that is both are left-wing politicians, therefore the associations made in those tweets attacked Marielle's reputation by associating her with a politician that was on the verge of being arrested. One month after the murder (on 14th April), a spike in the total positive p score reflected tweets demanding justice and answers from investigators (see Figure 2). Other positive peaks reflect some tweets from celebrities and politicians, supporting Marielle and claiming for justice, that got viral.

5   Evaluation

We evaluate our polarity score precision with a simple experiment, inspired by (SILVA et al., 2016). We randomly sampled 100 tweets from the dataset classified as negatives, then we manually verified if the tweet was really negative. We sampled only negative tweets because they represent a small part of the entire classified dataset. We observed that 90% of the sampled tweets were correctly classified as negative, while 10% were false negatives.

6   Conclusion and future work

In this work, we presented an ongoing work on analyzing the polarity on public opinion related to the murder of Brazilian councilwoman and activist Marielle Franco on classifying tweets using a dictionary with positive and negative terms. As next steps, we aim to (i) enrich the dictionary to improve the precision on polarizing the tweets, (ii) combine our approach with machine learning approaches, and (iii) identify in the negative polarized tweets hate speeches categories, such as race, religion and gender.

References

AGARWAL, Apoorv et al. *Sentiment analysis of twitter data*. In: Proceedings of the workshop on languages in social media. Association for Computational Linguistics, 2011. p. 30-38.

BAE, Younggue; LEE, Hongchul. *Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers*. Journal of the Association for Information Science and Technology, v. 63, n. 12, p. 2521-2535, 2012.

CHA, Meeyoung et al. Measuring user influence in twitter: *The million follower fallacy*. Icwsm, v. 10, n. 10-17, p. 30, 2010.

COLETTO, Mauro et al. *Sentiment-enhanced multidimensional analysis of online social networks: Perception of the Mediterranean Refugees crisis*. In: Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on. IEEE, 2016. p. 1270-1277.

CONOVER, Michael D. et al. *Predicting the political alignment of twitter users*. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011. p. 192-199

KOLCHYNA, Olga et al. *Twitter sentiment analysis*: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955, 2015.

PANG, Bo et al. *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, v. 2, n. 1–2, p. 1-135, 2008.

SILVA, Leandro Araújo et al. *Analyzing the Targets of Hate in Online Social Media*. In: ICWSM. 2016. p. 687-690.