

ORIGINAL ARTICLE

Harvesting Social Signals to Inform Peace Processes Implementation and Monitoring

Aastha Nigam,^{1,2} Henry K. Dambanemuya,^{2,3} Madhav Joshi,³ and Nitesh V. Chawla^{1-3,*}

Abstract

Peace processes are complex, protracted, and contentious involving significant bargaining and compromising among various societal and political stakeholders. In civil war terminations, it is pertinent to measure the pulse of the nation to ensure that the peace process is responsive to citizens' concerns. Social media yields tremendous power as a tool for dialogue, debate, organization, and mobilization, thereby adding more complexity to the peace process. Using Colombia's final peace agreement and national referendum as a case study, we investigate the influence of two important indicators: intergroup polarization and public sentiment toward the peace process. We present a detailed linguistic analysis to detect intergroup polarization and a predictive model that leverages Tweet structure, content, and user-based features to predict public sentiment toward the Colombian peace process. We demonstrate that had proaccord stakeholders leveraged public opinion from social media, the outcome of the Colombian referendum could have been different.

Keywords: big data analytics; predictive analytics; peace process; social media; machine learning; Colombia; data mining

Introduction

Unpredictability of sociopolitical events in today's contemporary society is rendering their outcome susceptible to volatility and uncertainty, leading to the question of how do we maintain a pulse on the public opinion or sentiment on the event. There has been an influx of studies reflecting on the use of social media as an instrument to model and understand the public opinion in response to a political phenomenon.¹⁻²⁴ Studies on the 2016 U.S. presidential election and Brexit, for example, have shown that there was greater polarization of opinion in the society than as evidenced by polls, and that social media activity could have more accurately reflected the pulse of the society.^{15,21}

In this article, we ask the following fundamental question: can social media signals help explain, to some extent, the progression and execution of peace processes? Confounded by the challenge of a 40% chance of failure in the first 10 years of their implementation,²⁵ peace

processes face the challenge of accurately capturing the pulse of the society, which can be crucial for their successful implementation and monitoring. Although military victory used to be the predominant conflict termination outcome during the Cold War, since 1989 there has been a significant increase in ceasefires, peace agreements, and other civil war outcomes.^{26,27} In fact, between 1989 and 2015, 69% of 142 civil war terminations occurred through negotiated peace agreements.²⁸ However, the likelihood of negotiated peace agreement failure is higher than other types of civil war outcomes.²⁹ In general, negotiated peace settlements have a 23% chance of conflict reversion during the initial 5 years and 17% chance of reversion in the subsequent 5 years.²⁵ On average, negotiated peace agreements last 3.5 years before conflict resumes.³⁰

Negotiated peace agreements are driven by a framework informed by bureaucratic and institutional processes, which assume that polls accurately reflect

¹Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana.

²Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, Indiana.

³Kroc Institute for International Peace Studies, University of Notre Dame, Notre Dame, Indiana.

*Address correspondence to: Nitesh V. Chawla, E-mail: nchawla@nd.edu

citizens' opinions and perception toward the peace agreement. However, the low success rate of peace agreements also implies that there may be additional undercurrents about peace processes that may not be reflected by the polls. Thus, we ask, can social media, akin to its use in election and other referendums, provide additional insights to understand the pulse of the society about a peace process?^{31,32} In this article, we focus on the Colombian peace agreement. Although all the major polls indicated a predominant Yes outcome, many people were astounded by the prevailing No vote during the October 2016 national referendum.³³

The Colombian peace agreement ended the longest fought armed conflict in the Western Hemisphere (Fig. 1). The Colombian peace process demonstrated that peace processes are complex, protracted, and contentious—characterized by significant bargaining and compromising among various societal and political stakeholders.^{34,35} The most recent cycle of the Colombian peace process was initiated in August 2012 with the signing of a framework agreement that identified negotiation issues and provided a road map to the September 2016 final peace agreement. Despite the confidence demonstrated by all major Colombian polling organizations for a successful outcome, when the peace agreement was put to a yes and no referendum vote in October 2016, Colombians narrowly rejected the agreement. After renegotiations between the two parties who were in favor and against the peace process, Colombia's Congress finally approved the agreement in November 2016.

Unlike any other civil war termination in recent history, the Colombian peace agreement was widely discussed and campaigned through social media. We posit that the social media, despite being a sample of

the Colombian population, can potentially provide insights about public perception and sentiment in response to the peace process. To that end, we collected social media (Twitter) data around two main Colombian peace process events—the signing of the final peace agreement on September 26, 2016, and the national referendum on October 2, 2016, as shown in Figure 1.

Within the context of understanding the societal aspects during the peace process and the relatively surprising outcome of the Colombian peace referendum outcome, we study the following two questions: (1) Can social signals harvested from social media augment knowledge about the potential success or failure of a peace agreement—informing its structure, implementation, and monitoring? (2) Can insights drawn from social media analysis be indicative of the polarization and sentiment of the society toward the peace process, thereby predictive of its outcome?

To answer these questions, we model social signals from Twitter to develop two indicators: (1) intergroup polarization and (2) public sentiment. We find that the political environment before the referendum was polarized between proaccord (Yes) and antiaccord (No) groups. The two groups communicated differently and stood divided on key issues such as the Revolutionary Armed Forces of Colombia (FARC) and women's rights. Furthermore, we observe that the No signal was dominant with negatively charged sentiment and a well-organized campaign. In addition, by leveraging the Tweet structure, content, and user-based features, we were able to estimate public sentiment with 87.2% predictability. In comparison with contemporaneous polling data, our results better approximated the referendum outcome. We showcase that social media offers a compelling opportunity to enhance the understanding of public



FIG. 1. Colombian peace timeline from 1960s to present date. The Colombian peace process consisted of various peace processes at different times in history. The peace processes studied in our research have been highlighted (dotted line). The timeline was built based on literature review of the Colombian peace process.

perception and opinion around formation of peace agreements and the related outcomes.

This article contributes new insights into the dynamics of peace processes especially before public approval or rejection of the final peace agreement during a public referendum. We demonstrate how social media can help peacebuilding researchers and practitioners gather evidence of changes in human behavior and perception toward peace processes and empower them to take appropriate actions to promote favorable outcomes. Furthermore, this work highlights the importance of studying public opinion and sentiment for understanding events beyond peace processes such as monitoring military operations³⁶ and immigration.

Background

Colombia's polarization between Yes and No groups spans many decades in its formation and has divided Colombian politics on whether a negotiated peace settlement or a military onslaught would be an acceptable strategy to end the conflict. During President Alvaro Uribe's second tenure (2006–2010), military campaign was the preferred strategy to defeat the FARC. Defense Minister, Juan Manuel Santos' led military campaign severely weakened the FARC, convincing significant segments of the population in favor of militarily defeating the FARC. Despite this popular sentiment, newly elected President Juan Manuel Santos sought a negotiated peace settlement with the FARC. Therefore, when the peace agreement was signed on September 26, 2016, some groups still believed that the FARC could be defeated militarily. Other groups found the terms of the peace agreement lenient toward FARC combatants alleged of committing human rights violations. Noticeable opposition also came from fiscal conservatives on issues related to economic incentives for FARC combatants and rebuilding of war-torn communities, religious groups on gender-related issues, and business communities on land issues. For example, Catholic and Evangelical Christians interpreted gender provisions as government approval for gay marriage, sexual education in schools, and other liberal policies that they disapproved of.³⁷ Therefore, as soon as the final agreement was announced, individuals and groups opposed to specific issues in the peace agreement and the peace process coalesced under the No group loosely led by former President Alvaro Uribe.³⁸ In contrast, President Santos remained committed to the peace process with support from the Yes group.³⁹ These dynamics demonstrate that peace processes are particularly vulnerable to hardliners from polarized groups during the negotia-

Table 1. Keywords used for collection of Tweets

Date	Keywords
September 11	Colombia peace, farc, Colombia referendum, Colombia peace process, Colombia displaced, Havana peace process, Colombia final agreement, Colombia conflict victims, Colombia missing persons, Colombia United Nations
September 15	Colombia ceasefire, paz en Colombia, farc, Colombia referendum, Colombia acuerdo final, Colombia victimas del conflicto, fin del conflicto armado, proceso de paz de la Habana, proceso de paz Colombia, Colombia desplazados
September 26	#firmadelapaz, paz en Colombia, farc, Colombia referendum, Colombia acuerdo final, Colombia victimas del conflicto, fin del conflicto armado, proceso de paz de la Habana, proceso de paz Colombia, Colombia desplazados

The keywords suggested by the peace scientists were changed on the above listed dates.

tion and implementation phases.^{31,32} Notwithstanding these challenges, the number of armed conflicts resolved through negotiated peace settlements has increased significantly over the past 25 years.^{29,40} Previous studies also demonstrate that the level of peace agreement implementation is a significant predictor for durable peace between signatory and nonsignatory groups,^{41,42} thereby motivating our inquiry of the interplay between intergroup polarization, public sentiment, and peace process implementation.

Materials and Methods

In this section, we discuss the data collection and preprocessing steps, followed by general trends observed from the data, and lastly discuss the methods used in this research to study intergroup polarization and public sentiment.

Data collection and preprocessing

We studied the Colombian peace process by analyzing the political environment on Twitter 3 weeks before the referendum (October 2, 2016). The data were collected from September 11 to October 1, 2016, using a Python API, Tweepy,* to read Twitter data. We limited our data set to Spanish Tweets for two reasons: more than 99% of Colombians speak Spanish and we are mostly interested in capturing the opinions of Colombians about the peace process. To collect social media data for the Colombian peace process, we used a set of keywords related to substantive issues about the peace process as our tracking parameter as shown in Table 1. The tracking keywords were decided based on the

*www.tweepy.org/

feedback provided by peacebuilding professionals working in the field (Colombia) and peace scientists who have been studying and closely monitoring the peace process through newspaper and government reports.[†] The keywords were chosen such that they were reflective of the key issues of the process, including signing of the final peace agreement, public perceptions toward the FARC, victims of conflict, and the public referendum. The keywords as shown in Table 1 were kept pertinent to the main aspects of the peace process and were unbiased toward either the proaccord or antiaccord group.

During the process of data collection, the keywords were changed every couple of days to collect comprehensive and unbiased data. As shown in Table 1, the first set of keywords was in English. Using these keywords, we observed that most of the retrieved Tweets were generated from Colombia and were in Spanish. Therefore, we henceforth only used Spanish keywords. In addition, as mentioned earlier, for the analysis we only used the Tweets in Spanish (which was the majority). While using keywords, hashtags or following users is a standard process in sampling Twitter data feeds,^{4,15,16} they can limit the possible universe of Tweets, tilting the bias either in the favor of or against an issue. To overcome this challenge, we created an unbiased and comprehensive set of keywords (Table 1) focused on the Colombian peace referendum and informed by experts from the Kroc Institute for International Peace Studies at Notre Dame. We collected a sizable number of Tweets ($\approx 300K$) to have a representative data set. Future research will consider methods that allow us to go beyond keywords in collection of the data.

We extracted the Tweets from the JSON object returned by the Twitter API. We then performed basic data preprocessing steps to extract the text and hashtags from each Tweet. After data preprocessing, we split the Tweet text into word-based tokens and used regular expressions to strip off punctuation marks. In addition, we cleaned the data using Spanish stop words provided by nltk module⁴³ in Python and domain-specific words such as *despues* (after) and *poder* (power). The word *Colombia* (name of the country) was also considered as a stopword for our experiments since all Tweets were relevant to Colombia and did not provide any new information. We also restricted our data to Tweets from users whose gender

could be retrieved (more details are in the Intergroup polarization section).

For our study period, we retrieved a collection of 280,936 Spanish Tweets (and Retweets generated by 34,190 users) related to the Colombian peace process. The entire data set was used for the polarization analysis. However, for the sentiment analysis we studied a subset of the Tweets since manual annotation of the sentiment was required. The Tweets were manually labeled by a native Spanish speaker[‡] based on the sentiment polarity of the adjectives and adverbs contained in the message. We assumed a binary opposition (positive or negative) in the polarity of the sentiment. The annotator described rules for sentiment labeling to ensure repeatability and generalizability of the annotation process. Given the onerous task of manual labeling, we downsampled to 3000 unique Tweets.[§] The data set had about 1071 labeled as a positive sentiment class. To retain the balance between the negative and positive Tweets, we sampled the same number of Tweets from the negative sentiment class, creating a balanced data set of 1071 Tweets in each of the positive and negative sentiment classes—generating balanced samples using oversampling or undersampling is a fairly standard approach when dealing with imbalanced data. Our goal here was to decipher the effects of different features in their propensity to label sentiment (results discussed in the Public sentiment section). However, we realize a limitation of this analysis arises from both the sample size and labels generated by a single annotator. As part of future work, we aim to expand the analysis to include a larger set of Tweets and multiple human annotators.

General trends

Figure 2a depicts the volume of Tweets obtained from September 11, 2016, to October 2, 2016, using the keywords. As shown, we observed a sharp increase in number of Tweets on September 26, which marked the signing of the final peace agreement in Cartagena. As previously discussed, the referendum was conducted on October 2, 2016; therefore, we only utilize the data collected until October 1 in our study. Figure 2b depicts the number of Tweets generated by each user. It can be observed that it is a heavy tailed distribution with most users sharing only a small number

[†]The Kroc Institute for International Peace Studies is responsible for monitoring the Colombian peace process as stipulated in the peace agreement. The peace scientists at Kroc Institute are considered domain experts who have been studying the peace process for many years and have field knowledge (also coauthors on this study).

[‡]The annotator was fluent in Spanish and well versed with the Colombian peace process.

[§]Since we are interested in the sentiment of a Tweet, Retweets were not considered for this analysis. Only original Tweets were annotated.

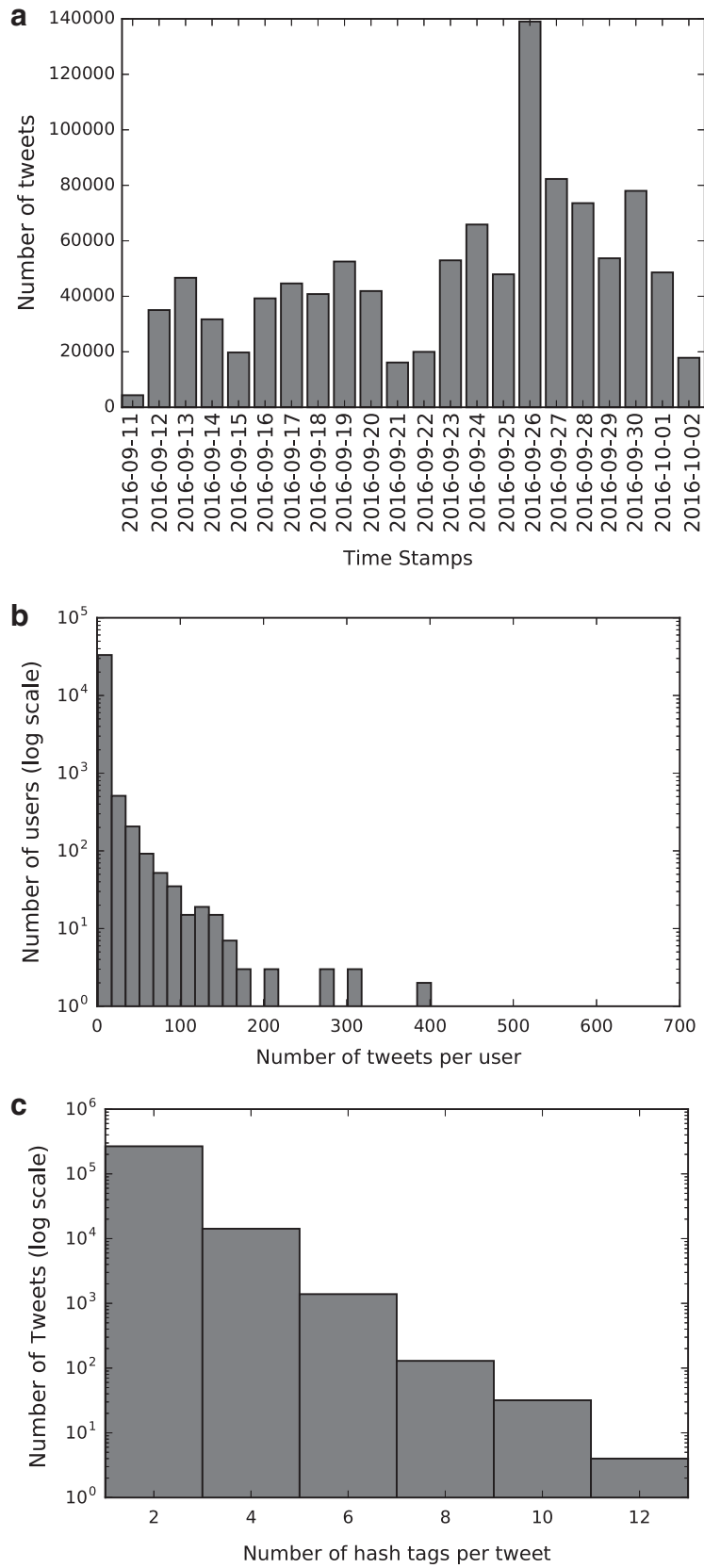


FIG. 2. Data statistics. General trends observed in the Twitter data for the Colombian peace process: **(a)** Daily Tweet volume across the study period (before the referendum), **(b)** histogram distribution of number of Tweets by users, **(c)** histogram distribution of number of hashtags observed in a Tweet. Authors' calculations based on the collected Twitter data set.

of Tweets. Moreover, since the polarity of a Tweet is inferred based on hashtags used (explained later), we present a distribution of number of hashtags used in a Tweet as observed in our data, as shown in Figure 2c. We observed that most Tweets contain between one and eight hashtags.

The Twitter data set consisted of 27,862 and 10,368 unique words and hashtags, respectively. Tables 2 and 3 list the top 15 words and hashtags with their English translations and proportions, respectively. The most frequent word and hashtag recorded was farc (the rebel group). Moreover, we observed words such as no, paz (peace), acuerdo (agreement), and Santos (president of the country) appearing as the most prominent words. Similarly, we observed No polarized hashtags such as #VoteNo (vote no), #ColombiaVotaNo (Colombia vote no) and #EncartagenaDecimosNo (in Cartagena we say no) emerging as the most popular hashtags.

Methods

To understand the complex phenomena of peace processes, we leverage the data collected from social media to harvest two indicators: intergroup polarization and public sentiment. Intergroup polarization is characterized by extreme and divergent opinions by different political groups on key issues,⁴⁴ whereas public sentiment captures people's emotions toward the peace agreement as either negative or positive. The interplay of these two indicators can help us better understand peace processes. We discuss the two indicators in the subsequent subsections.

Intergroup polarization. Our analysis for intergroup polarization covers four main components: (1) hashtag

Table 2. Top 15 words obtained from Tweets on Colombian peace agreement

Word	Translation	Proportion (%)
FARC	FARC	8.62
No	No	4.13
Paz	Peace	4.08
Acuerdo	Agreement	1.70
Santos	Juan Manuel Santos	1.29
Acuerdos	Agreements	0.95
Gobierno	Government	0.94
Firma	Signing	0.81
Si	Yes	0.73
Victimas	Victims	0.58
Guerra	War	0.49
Plebiscito	Referendum	0.48
Colombianos	Colombians	0.46
Armas	Weapons	0.43
Pol	In	0.43

Authors' calculations based on the collected Twitter data set.

Table 3. Top 15 hashtags obtained from Tweets on Colombian peace agreement

Hashtag	Translation	Proportion (%)
#FARC	FARC	6.36
#AcuerdoDePaz	Peace agreement	2.59
#FirmaDeLaPaz	Sign the peace	2.55
#VotoNo	Vote no	2.06
#ColombiaVotaNo	Colombia vote no	1.67
#EncartagenaDecimosNo	In Cartagena we say no	1.59
#Paz	Peace	1.56
#NoAlasFarc	No to FARC	1.54
#SiAlaPaz	Yes to peace	1.41
#CartagenaPiTano	Cartagena honks no	1.07
#VoteNo	Vote no	1.06
#Alaire	To the air	1.04
#VotoNoAlPlebiscito	Vote no to plebiscite	0.98
#HagaHistoriaVoteNo	Make history vote no	0.98
#ColombiaconelNo	Colombia with no	0.93

Authors' calculations based on the collected Twitter data set.

spread and evolution analysis, (2) word association analysis, (3) emergent topic analysis, and (4) polarized users analysis. In this section, we explain the method used for each of those analyses.

Hashtag spread and evolution analysis. As discussed previously, we identified a set of keywords (Table 1) for data collection based on the input from peace scientists. Once we collected the Tweets from those keywords, we tabulated all the hashtags in the Tweet data set (top 15 hashtags are listed in Table 3). The set of hashtags retrieved from our Twitter data set was leveraged to infer polarity. For all the retrieved hashtags, based on the presence of the word no or si, we classified each as No or Yes hashtag, respectively. Hashtags containing the keyphrase no or its variation were classified as No hashtags. Similarly, hashtags containing the keyphrase si or its variation were marked as Yes hashtags. Consequently, the hashtag labels were used to infer the polarity of the Tweets. Tweets with only No hashtags were marked as belonging to the No group, and conversely Tweets with only Yes hashtags were classified as belonging to the Yes group.** We studied the evolution patterns of Tweets and characterized prominent hashtags across the following five key attributes:

- **Daily volume:** We computed mean volume of Tweets observed with a hashtag on each day of our study period.

**We do not explicitly follow yes or no hashtags; as previously discussed, we followed keywords based on expert guidance and then retrieved hashtags from the collected data to infer polarity.

- **Variability:** We observed that some hashtags were used consistently in the campaign compared with few that were popularized for a short period of time. To distinguish between such hashtags, we introduced the concept of variability (to understand the fluctuations in the evolution of a hashtag over time). It was computed as shown in Equation (1).

$$variability = mean(|v_i - v_j| \text{ for } j > i), \quad (1)$$

where v_i and v_j represent the volume of Tweets on days i and j , respectively.

- **Influence:** To capture the influence of a hashtag across a diverse audience, we computed this metric by leveraging the number of unique users who had used the given hashtag in a Tweet.
- **Popularity:** To estimate the reach of a given hashtag, we computed the average Retweet count of Tweets that had used the given hashtag.
- **Prominence:** To calculate the prominence of the hashtag, we computed the average number of followers for users who had used the given hashtag in their Tweets.

Word association analysis. To characterize the two groups, we learned the word embeddings using a word2vec model^{45,46} that captured word associations and context. The word2vec model uses neural networks to produce word embeddings based on context. Since we used hashtags to detect intergroup polarization, we excluded them from the word2vec model. Furthermore, we studied the word associations, as obtained by word2vec model, for key entities identified by peace scientists. In addition, we leveraged pointwise mutual information (PMI)^{47,48} to estimate the importance of the entity in each group. PMI is given in Equation (2).

$$pmi(x, y) = \log \frac{p(x, y)}{p(x)p(y)}, \quad (2)$$

where x and y are the word and class, respectively. $p(x)$ and $p(y)$ are the probabilities of observing x and y in the data set. $p(x, y)$ captures the joint probability that is computed by using number of Tweets that belong to class y and contain the word x . In our case, the two classes would be the proaccord (Yes) and antiaccord (No) groups.

Emergent topic analysis. To further understand the public opinion and infer topics of discussion, we used a popular topic modeling algorithm called Latent Dirichlet Allocation (LDA)⁴⁹ to analyze the emergent topics. LDA

is a generative model wherein each document (Tweet) is described as a mixture of topics distributed on the vocabulary of the corpus. It allows us to study emerging topics using contributing words from the training vocabulary and the spread of each topic in a document. To decide the number of topics to be learned, we used topic coherence (C_v) as the metric.⁵⁰ C_v leverages the indirect cosine measure with the normalized PMI.

Polarized users analysis. After performing analyses on the Tweets and studying their content, we identified the No, Yes, and Undecided users based on the polarity of their Tweets. If the users exclusively had Yes Tweets, they were classified as Yes users. Similarly, users who only had No Tweets were marked as No users. Users with both Yes and No Tweets were categorized as Undecided users. Using this method, we were able to identify the alignment of the users in an extremely polarized environment. However, we do understand the biases implied by this assumption. A user expressing opposing views might not always be an undecided user. Therefore, to build a comprehensive understanding of the user, for next steps, we would like to leverage the number of Yes and No Tweets spread across a longer period of time and estimate the probability of a user belonging to either group instead of using frequency.

We first studied the content similarity between users by computing the Jaccard similarity⁵¹ over the Tweets shared by them. Jaccard similarity is given by Equation (3).

$$jaccard_similarity = \frac{|A \cap B|}{|A \cup B|}, \quad (3)$$

where A and B are set of words used by the users in their Tweets.

In addition, we extracted user gender and location features from the Tweets. We use the location feature to compare the base population's location demographic with the population composition from polling and Twitter data. Furthermore, as previously discussed in the Background section, gender was one of the most polarizing issues in the Colombian peace agreement; therefore, we study the distribution of polarized users across gender. The users' location was inferred based on their self-reported location in their profile, and to infer users' gender we used a method similar to Mislove et al.⁵² We queried users' first names against a dictionary of 400 common English and Spanish names, comprising 200 names for each gender equally distributed between each language. We were able to retrieve the

location and gender for 63.78% and 100% of the total users, respectively.

Public sentiment. To build a model to predict the sentiment of each Tweet, we extracted three categories of features: (1) Tweet-based, (2) content-based, and (3) user-based. We computed Tweet-based features by extracting basic attributes such as number of hashtags and number of words. Furthermore, we computed the distribution of a Tweet over different parts of speech (POS) using the Stanford Spanish POS tagger.^{53,54} We extended the Tweet-based features, by augmenting information about number of positive and negative words based on a publicly available Spanish sentiment lexicon.^{††} For the content-based features, we used two methods (1) term frequency-inverse document frequency (TF-IDF)⁵⁶ representation and (2) topic representation using LDA⁴⁹ model. For the first content representation using TF-IDF, we utilized TFIDFVectorizer implementation by the sklearn module⁵⁷ in Python. The TF-IDF score is given by Equation (4)

$$tf_idf(t, d) = tf(t, d) \times idf(t), \quad (4)$$

where $tf(t, d)$ is given by number of times term t appears in document d and $idf(t)$ is given by Equation (5)

$$idf(t) = \log \frac{1 + n_d}{1 + df(d, t)} + 1, \quad (5)$$

where n_d is the total number of documents (or Tweets) and $df(d, t)$ is the number of documents (or Tweets) that contain term t . For the second content representation, as previously discussed, LDA computes the distribution of topics across each document (Tweet). From the learned model, we can project a new test document on the trained topic model. Lastly, for user's network-usage features, we used features such as number of friends and followers as observed from their profiles. An exhaustive list of all the features used for the sentiment prediction model is provided in Table 4.

As mentioned before, the Tweets were manually labeled by a native Spanish speaker based on the sentiment polarity of the adjectives and adverbs contained in the Tweet. Furthermore, we assume a binary opposition (positive or negative) in the polarity of the sentiment. Leveraging three categories of features and the manually annotated sentiment label, we trained various machine learning models such as Support Vector

Table 4. Prediction features used for learning public sentiment

Category	Features
Tweet-based	Retweet (yes/no), no. of mentions, no. of hashtags, no. of yes group hashtags, no. of no group hashtags, no. of urls, no. of positive (+ve) emoticons, no. of negative (-ve) emoticons, no. of words, no. of positive (+ve) words, no. of negative (-ve) words, Retweet count, favorite count, parts of speech information: [no. of nouns, pronouns, adjectives, verbs, adverbs, prepositions, interjections, determiners, conjunctions, and punctuations.]
Content-based	TF-IDF representation of the Tweet vs. Tweet represented across topics computed by LDA
User-based	Follower count, friends count

The features are categorized into Tweet-based, content-based, and user-based. The metrics were derived based on literature review of studies using Twitter-based data.

LDA, latent Dirichlet allocation; TF-IDF, term frequency-inverse document frequency.

Machines (SVMs),⁵⁸ Random Forests (RFs),⁵⁹ and Logistic Regression (LR)⁶⁰ to estimate the sentiment of a Tweet. We employed precision, recall, F1 score, area under the receiver operating curve, and accuracy to evaluate the performance of the prediction tasks in this study.⁶¹

Results

We evaluated the outcome of the Colombian peace process through two keys indicators: intergroup polarization and public sentiment.

Intergroup polarization

Colombian politics has been polarized for many decades on whether a negotiated peace settlement or a military onslaught would be an acceptable strategy to end the conflict. In this section, we present our results on intergroup polarization and discuss the major issues that drove the polarization as measured through Twitter.

Hashtag spread and evolution analysis. We studied polarization in Colombian society by analyzing the political environment on Twitter 3 weeks before the referendum (September 11 to October 1, 2016). As previously discussed in the Intergroup polarization section in Materials and Methods, the hashtags were classified as a No/Yes hashtag. In total, we retrieved 1327 No and 798 Yes hashtags. Using the set of hashtags present in a Tweet, the political alignment of each Tweet was inferred. A Tweet with only No hashtags was classified as a No Tweet. Similarly, a Tweet with only Yes hashtags was labeled as a Yes Tweet. Based on this criterion, we obtained 17,783 Yes group Tweets, 84,742 No group

^{††}We used the open source Spanish sentiment lexicon from Stony Brook University's Data Science Lab⁵⁵ to estimate number of positive and negative words.

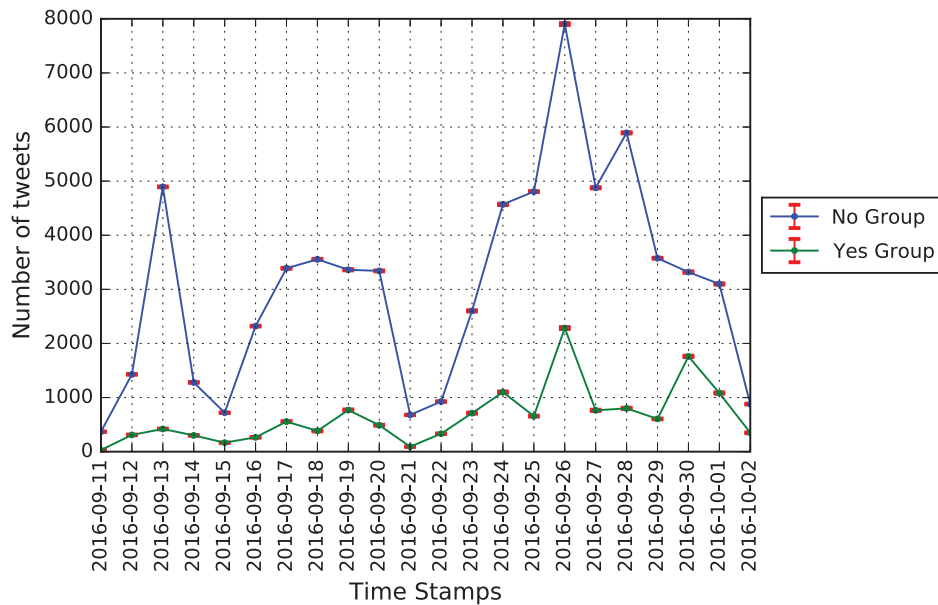


FIG. 3. Evolution of Tweets for the polarized groups. Evolution of Tweets from No and Yes groups over time. The plot shows error bars by taking random samples from each day and plotting mean results for 100 iterations. Authors’ calculations based on the collected Twitter data set.

Tweets, and 1731 as neutral Tweets (presence of both yes and no hashtags). Since the number of neutral Tweets was not significant in comparison with Yes and No Tweets, we excluded them from further analysis. We observed that Colombians were divided between Yes and No camps, as shown in Figure 3, with the No Tweets dominating political conversations leading up to the referendum on October 2, 2016.

We further investigated the usage and evolution of 10 most commonly used hashtags by both groups. The dominance of the No group is demonstrated through their consistent and well-organized campaign as shown in Figure 3. We observed greater activity around September 26 when President Juan Manuel Santos and FARC rebel leader Rodrigo Timochenko Londono signed the final peace agreement in Cartagena, Colombia. Over the 21-day period, the No campaign was characterized by strong and negatively charged hashtags such as #EncartagenaDecimosNo (In Cartagena we say no), #HagaHistoriaVoteNo (Make history vote no) and #Colombiaconelno (Colombia with no) as shown in Figure 4a.^{††} By contrast, we observed less activity from the Yes group (Fig. 4b), which gained momentum few

days before the referendum. Although most of the hashtags were popularized 2 to 3 days before the referendum, #sialapaz (Yes to peace) is the only hashtag that persisted throughout our study period.

We further studied the usage of each hashtag across five key attributes⁶² as discussed in the Intergroup polarization section. Daily volume captures the number of Tweets shared daily containing the given hashtag; Variability measures the variation in the daily volumes of a hashtag to capture its peak usage and volume fluctuations; influence computes the number of distinct users Tweeting with the given hashtag; popularity calculates the reach of a hashtag measured by average Retweet count of Tweets containing the hashtag, and prominence captures the average number of followers each Tweeter who used the given hashtag had.

As shown in Table 5, we observed the No group dominating social media conversations related to the peace process based on their higher daily volumes, wider No-related hashtag use by diverse users, and greater popularity and prominence of No group Tweets compared with their Yes counterparts. No group hashtags such as #noalafarc (No to FARC) and #encartagenadecimosno (In Cartagena we say no) influenced (as defined in the Intergroup polarization section) a diverse set of users attaining higher popularity and prominence scores even

^{††}We would like to note that the scales for plots in Figure 4 are different because we want to highlight the difference in the activity. To capture the difference in the magnitude, we present the actual volumes instead of normalized values.

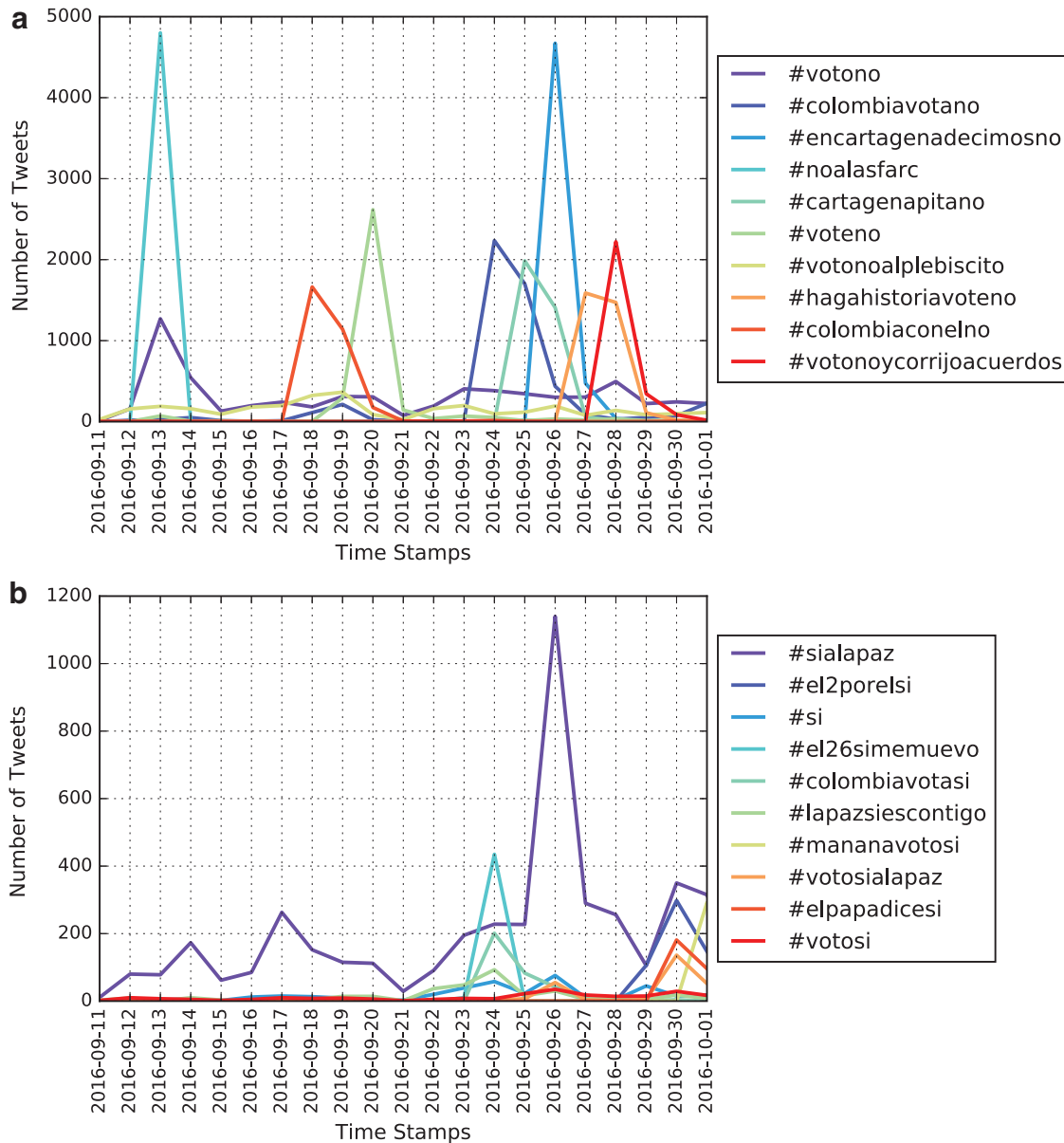


FIG. 4. Hashtag trends for No and Yes groups. Evolution based on the volume of key hashtags for the No (a) and Yes groups (b) as recorded on Twitter from September 11 to October 1, 2016. Please note that there is a difference in scale for both plots. Authors’ calculations based on the collected Twitter data set.

though they persisted over short periods of time and were associated with high variability. Yes group hashtags such as #elpapadicesi (The Pope says yes), although not frequently used by diverse users, resulted in high Retweet volumes as indicated by their popularity scores.

Word association analysis. Although polarized, we observed that Tweets from both groups frequently com-

prised words commonly associated with the peace agreement such as farc (rebel group) and paz (peace) (as shown in Table 6), making it difficult to differentiate and characterize the political beliefs of each group. We argue that even though both groups used the same words, understanding the usage and context of these words is crucial to deciphering public opinion. Therefore, we extended our analysis to characterize the

Table 5. Lists frequently used hashtags by both groups and quantifies them across five key attributes

Group	Hashtag	Daily volume	Variability	Influence	Popularity	Prominence
No group ^a	#votono	5.63 (1.09)	3.83 (1.21)	10.15 (1.20)	28.87 (5.22)	11.79 (3.16)
	#colombiavotano	4.85 (2.46)	5.12 (2.38)	8.65 (3.874)	27.92 (6.65)	16.17 (4.25)
	#encartagenadecimosno	4.69 (4.19)	8.80 (5.73)	17.03 (14.04)	47.43 (15.31)	8.25 (5.20)
	#noalafarc	4.59 (4.54)	9.59 (6.55)	1.49 (1.13)	49.01 (20.97)	6.56 (1.81)
	#cartagenapitano	3.16 (2.15)	3.75 (2.25)	14.49 (8.99)	55.23 (14.33)	8.70 (6.26)
	#voteno	3.08 (2.38)	5.28 (3.15)	4.97 (3.44)	21.70 (6.19)	6.99 (1.76)
	#votonoalplebiscito	2.24 (0.33)	1.21 (0.26)	5.01 (0.75)	35.45 (6.18)	11.16 (2.59)
	#hagahistoriavoteno	2.88 (1.90)	2.91 (1.92)	16.54 (9.64)	53.47 (17.42)	3.22 (1.25)
	#colombiaconelno	2.65 (1.72)	3.02 (1.73)	10.06 (5.94)	33.33 (11.22)	8.57 (4.39)
Yes group ^a	#votonoycorrijoacuerdos	2.33 (1.95)	4.07 (2.58)	13.31 (9.77)	49.33 (18.92)	8.39 (3.89)
	#sialapaz	2.89 (0.73)	2.48 (0.76)	6.48 (1.68)	17.70 (5.09)	11.97 (4.37)
	#el2porelsi	0.44 (0.27)	0.50 (0.26)	5.53 (2.70)	46.17 (23.03)	13.98 (5.42)
	#si	0.36 (0.09)	0.42 (0.09)	0.86 (0.24)	23.75 (6.43)	4.23 (0.89)
	#colombiavotasi	0.21 (0.12)	0.23 (0.12)	1.69 (0.83)	34.22 (14.06)	7.09 (1.84)
	#lapazsiescontigo	0.27 (0.09)	0.25 (0.08)	0.58 (0.20)	28.43 (6.96)	11.49 (3.11)
	#votosialapaz	0.21 (0.13)	0.30 (0.16)	1.802 (1.01)	40.84 (18.03)	5.53 (1.42)
	#elpapadicesi	0.27 (0.18)	0.33 (0.19)	4.68 (2.24)	55.79 (29.44)	15.85 (9.71)
	#votosi	0.11 (0.03)	0.09 (0.02)	0.25 (0.08)	16.56 (7.06)	2.81 (0.77)
#votosiel2deoctubre	0.15 (0.11)	0.27 (0.14)	1.56 (1.04)	26.93 (18.99)	4.26 (1.33)	
	#yovotosi	0.07 (0.02)	0.09 (0.02)	0.18 (0.06)	15.67 (8.17)	4.08 (1.50)

The values are presented as mean (scale: 0–100) over our study period with standard error (in parentheses). Authors' calculations based on the collected Twitter data set.

^aWe study the frequent hashtags for the No (against the accord) and Yes (in favor of the accord) groups separately.

two groups based on their Tweets by building content profiles using word-level associations. Using Tweets from each group (7825 and 11,675 unique words for Yes and No group Tweets, respectively), we built independent word2vec models^{45,46} to understand word associations and the context of each word as used in their respective group's Tweets. We used cosine similarity score to interpret the (dis)similarity between word contexts of common words. We observed that although there exists a significant word-usage overlap between the content generated by both groups, the same words are used in dissimilar contexts marked by relatively lower cosine similarity scores as shown in Figure 5.

We further contrasted the content profile of each group across seven key entities (or keywords) identified

Table 6. Top 10 words occurring in polarized yes and no group Tweets

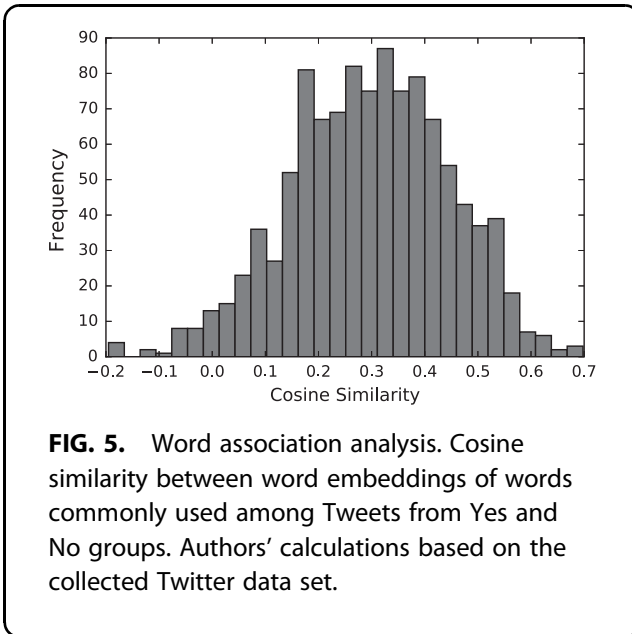
Yes group			No group		
Word	Translation	Proportion	Word	Translation	Proportion
Farc	FARC	6.67	Farc	FARC	10.35
Paz	Peace	5.52	No	No	6.66
No	No	3.54	Paz	Peace	2.12
Si	Yes	1.64	Santos	President Santos	1.85
Acuerdo	Agreement	1.18	Acuerdos	Agreements	1.59
Conflict	Conflict	1.08	Acuerdo	Agreement	1.21
Guerra	War	0.92	Octubre	October	0.91
Fin	End	0.89	Colombianos	Colombians	0.83
Armado	Armed	0.67	Si	Yes	0.75
Firma	Signing	0.59	Farc-santos	FARC-Santos	0.73

Authors' calculations based on the collected Twitter data set.

by domain experts—representing the main concepts associated with the peace process and capturing the main players involved in the peace process.^{§§} The seven entities include (1) Colombian President Juan Manuel Santos, (2) former President Alvaro Uribe, who loosely led the No campaign, (3) FARC rebel leader Rodrigo Timochenko Londono, (4) Plebiscito, which translates to referendum, (5) FARC rebel group, (6) Acuerdo, which translates to agreement, and (7) Paz, which translates to peace. We used PMI to understand the relative importance of each word for a given class (Yes/No group)^{47,48} as explained in Equation (2).

Table 7 shows the associations and class-level importance for each entity. The results show that both groups are less likely to mention their leaders in their Tweets. Rather, members of each group are more likely to mention the other group's leader than their own leader. For example, members of the Yes group are less likely to Tweet about their own leader, Santos, and more likely to Tweet about the No group leader, Uribe, than their No counterparts who are also more likely to Tweet about Santos than their own leader, Uribe. Although the finding that the other group's leaders were mentioned more frequently than that group's leaders may at first seem surprising, it resonates with the nature of the referendum given that it is the actions of the op-

^{§§}The entities presented were suggested by the already mentioned domain experts as they capture the most relevant events during our study period, related to the Colombian peace process.



position leaders and their followers on each camp that swayed the outcome of the referendum. For example, although each group's leader may be able to galvanize their followers toward voting for or rejecting the peace accords in the referendum, each group must try as much as possible to influence and persuade the other group's leaders and members to join their group. This can be achieved by criticizing the opposition's leaders' policies to try to win their followers, hence resulting in each group Tweeting more about their opposition leader than their own leader.

We also observed that the No group is more likely to mention the plebiscite and peace agreement in their Tweets than the Yes group. This observation mirrors events on the ground because the No group was largely responsible for opposing the peace agreement and voting it down during the October 2nd national

referendum. Although less likely to mention the peace agreement or the plebiscite in their Tweets, the Yes group was more likely to mention peace in their Tweets than the No group. This observation supports existing theories of peace and justice because the proaccord (Yes) group is mostly interested in achieving peace even at the expense of justice compared with the antiaccord (No) group that would rather see justice prevail and FARC combatants held accountable and punished for their crimes at the expense of the peace agreement. We conclude that even though both groups used similar words associated with the peace process, their content profiles are strikingly different and they stand divided on various key points.

Emergent topics analysis. We used LDA⁴⁹ to compute topics for the Yes and No groups based on each group's Tweets. Based on a grid search for number of iterations and topics (10, 20, 30, 50, 100) using topic coherence (C_v) as the metric, we computed 30 yes topics and 100 no topics. Topic coherence scores for the Yes and No groups were 0.414 and 0.413, respectively. Tables 8 and 9 illustrate key topics discovered by the topic models for the Yes and No groups, respectively.

We observed that the majority of the topics are related to the FARC. The topics are also related to Santos and Uribe as well as the national referendum and various aspects/provisions of the peace agreement such as victims of conflict, women, disarmament, prisoners of war, reparations, institutions, and justice. Although both groups' topics frequently contain FARC, we observed different word associations in topics that include the word FARC (rebel group) between the two groups. The words observed along with FARC (rebel group) in No topics appear to be more polarized and negatively charged than the Yes group word associations. Although both groups' topics contain words such as terrorists,

Table 7. Word associations for key entities associated with Colombian peace process and corresponding class importance computed with pointwise mutual information

Word	Yes group		No group	
	Associations	PMI	Associations	PMI
Santos	gusta, desarmar, paras, derrotadas	-1.37	santos-farc, comandante, aprobar, narcoasesinos	0.15
Uribe	seguridad, despiden, desmovilizadas, mentiras	0.28	viudas, odio, acepta, firmando	-0.07
Timochenko	confirma, atenci, familias, secreta	-0.09	asustaba, susto, ratificado, ofrezco	0.02
Plebiscito	votar, quiero, aprobar, puedo	-0.52	propaganda, papa, narcoasesinos, ganarnos	0.08
Farc	Timochenko, invito, promover, acabe	-0.42	impuesto, libres, constitucional, rechaza	0.07
Acuerdo	gobierno, final, firman, apoyan	-0.05	final, gobierno, comunicado, acuerdos	0.01
Paz	acuerdos, firmar, retos, apoyamos	0.67	apoya, insisten, alcanzar, entender	-0.22

Authors' calculations based on the collected Twitter data set. PMI, pointwise mutual information.

Table 8. Topics obtained from Latent Dirichlet Allocation models on Tweets from yes group

<i>Yes topic</i>	<i>Context</i>
entrevista, exguerrillero, vivo, encargado, rutas, exclusiva, paz, farc, cuenta, septiembre, coca, asesinos, nacional, universidad, agenda, alapaz, balas, dejamos, martes, dinero	Rodrigo Timochenko Londono's exclusive interview with The Observer
paz, acuerdo, firma, nueva, mundo, guerra, farc, final, cartagena, acaba, apoyan, nico, mas, buenos, artistas, alcanzar, asegura, objetivo, sociales, plena	President Santos and Timochenko sign peace agreement in Cartagena
paz, papa, francisco, une, habla, justicia, proceso, farc, eln, destruccion, desbocada, detengamos, naturales, nacionales, parques, no, auc, generacion, impunidad, m19	President Santos announces Pope Francis will visit Colombia in 2017
no, farc, paz, victimas, armas, uribe, gusta, santos, desplazados, inocentes, desarmen, gustan, si, dejen, bendici, paras, van, gente	Victims of conflict, displaced people, and disarmament of paramilitaries
farc, paz, oportunidad, acuerdo, venes, sociedad, volver, doy, ahora, guerra, puede, terminaci, no, conflicto, final, cerca, gobierno, firman	Final peace agreement and termination of war

Each topic is explained by appropriate context as interpreted by domain experts. Authors' calculations based on the collected Twitter data set.

guerrilla, death, war, rich, and victims, the Yes group topics include more positive and neutral words such as forgiveness, support, favor, and peace, whereas the No group topics comprise more negatively charged and accusatory words such as extortion, abduction, drug traffickers, bloodthirsty, killers, and psychopaths. Based on this observation as well as the contents of their Tweets, we can conclude that the Yes group seems to be more sympathetic toward the FARC rebels than the No group. This observation is consistent with the word association results wherein the No group is more vocal about justice for the FARC and their leader, Timochenko, as well as the peace agreement and plebiscite than the Yes group that seems to be only vocal about prospects for peace even at the expense of justice.

Polarized users' analysis. In the previous sections, we have discussed the extent of polarization as seen on social media through hashtags and studied emergent topics and words that characterize the content of the No and Yes groups. An important aspect of the polarized environment is the Twitter users. In this analysis, we categorize the users into three classes: Yes users (only Yes Tweets), No users (only No Tweets), and Undecided users (mixture of Yes and No Tweets showing no distinct preference toward either group).*** In total, we recorded 199 Yes, 841 No, and 452 Undecided users. We investigate the user groups and their alignments through the following dimensions:

- **Content similarity:** We studied the users' inter-class similarity based on the content they shared through their Tweets using Jaccard similarity⁵¹

[computed using Equation (3)]. As shown in Figure 6a, we observed that users belonging to the No and Undecided groups share overlapping content. However, the Yes group's content is remarkably different.

- **User demographic:** We also studied the user groups' demographics such as gender and location based on political regions. As previously described, we inferred user's gender from their self-reported first name based on combined English and Spanish name lexicons.⁵² Similarly, we leveraged users' self-reported locations to infer their political regions. From Figure 6b, we observe that irrespective of political alignment, men are more vocal about their opinions on Twitter. The ratio for men to women is comparable among all three groups with the most users belonging to the No group. Furthermore, as shown in Figure 8, we observed No and Undecided users dominating Twitter conversations related to the peace process across all regions. We also observed comparable proportions of No and Undecided users in urban regions such as Bogota, Caribe, and Centro.
- **User network and activity:** Finally, we studied the user groups across two social network metrics: (1) number of followers and friends to understand their reach and (2) number of Tweets and Retweets to interpret their usage activity. As shown in Figure 7, we observed distinct patterns for all three political alignment groups. Furthermore, we observed that Yes supporters have fewer friends and followers (Fig. 7a), affecting their social media usage and reach as shown in Figure 7b and c, respectively. Broadly, we observed that a Twitter user has more followers than friends and that their number of Tweets and Retweets increases proportionately. After comparing the No and Yes group users, we

***Like most Twitter-based studies, we consider a user to be a unique account. Therefore, we do realize that an account could belong to an individual or an organization. However, we do not make any distinctions on the types of accounts for any of the user classes; therefore, the grouping is uniformly applicable across the classes.

Table 9. Topics obtained from Latent Dirichlet Allocation models on Tweets from no group

No topic	Context
pagar, impuestos, privilegios, financiar, farc, tener, colombianos, golpe, presidente, fiscal, consejo, campanazo, ciudadanos, victimas, petici, reunidos, justicia	Fiscal conservatives on issues related to economic incentives for FARC combatants
gobierno, paz, farc, acuerdo, reparaci, santos, no, estable, ello, duradera, acuerdos, mentiras, lograr, comunicado, verdadera, vox, instituciones, victimas, contempla	Peace agreement expected to achieve stable and lasting peace
victimtas, no, farc, fortuna, reparar, entreguen, injusticia, inmensa, centavo, nobel, paz, riqueza, EEUU, logro, refiere, malas, reitera, solicitud, invitados, conservar	President Santo's Nobel Peace achievement and reparations for FARC victims
octubre, no, acuerdos, farc-santos, guerrilla, masivamente, votando, redireccionar, presentes, representantes, devuelto, latina, nicocampo, capo, descubierto, renegociemos	Calls for President Santos and FARC to renegotiate peace agreement
dinero, farc, narcotr, fico, no, secuestros, reparar, monos, victimas, violadores, asesinos, colombianos, mosles, risita, psicopatas, real, exigent	Depicting FARC as killers, terrorists, narcotraffickers, kidnapers, and corrupt

Each topic is explained by appropriate context as interpreted by domain experts. Authors' calculations based on the collected Twitter data set.

observed greater reach and more activity among No users. These results confirm the prominence of the No group before the referendum. However, we observed that the Undecided users were the most active group as shown in Figure 7b.

Relevance to the outcome of referendum. Before the national referendum, polls were conducted across six political regions in Colombia where they predicted an overwhelming Yes vote for the peace agreement. However, the referendum was rejected with a narrow margin—witnessing a low voter turnout with fewer than 38% of the voters casting their votes.⁶³ A comparison between the poll estimates and actual referendum votes^{†††} (Fig. 8) shows that the polls were unable to estimate the dominant no signal leading to a surprising referendum outcome. In the figure, we only showcase the percentage of the No voters as estimated by polls and the No voter proportion in the referendum. However, since the numbers are normalized on a scale from 0 to 1, the Yes voter proportion for both can be computed by taking the complement.

Although multiple factors contributed to the unfavorable outcome, according to the referendum data, the country was divided regionally. Outlying provinces that suffered the most during the conflict voted in favor of the peace agreement, whereas most inland provinces voted against the peace deal. Furthermore, it was also seen that regions with more supporters for Alvaro Uribe rejected the agreement, whereas regions with greater support for Juan Manuel Santos voted in favor of the agreement.⁶³ Given the controversial nature of the referendum, we were interested in understanding whether the no signals harvested from social media (before the referendum) were able to follow

the trends of the No voter turnout in the referendum better than the polling estimates.

In Figure 8 we showcase the distribution of Twitter users in the Undecided and the No groups, for each political region, based on our analysis. For simplicity and clear understanding, we present the Undecided and No proportions, and omit the Yes Tweeter proportion. However, since we use normalized scores, the Yes proportion can be estimated by subtracting the sum of Undecided and No volume from 1.

Although our results indicate the dominance of the No opinion through Figures 3, 4, and 6b, we also observe that the No Tweeters proportion is more representative of the referendum No voter turnout for most regions such as Bogota, Caribe, Cafetero, and Pacifica in comparison with the polling estimates. Particularly, for urban regions such as Bogota, Caribe, and Pacifica that witnessed a voter turnout of 62.8% during the referendum, we are able to capture the trends more closely. Being urban centers of the country, we were able to obtain a more representative Twitter data sample from these regions, constituting 89.8% of the total Tweeters—leading to better approximations.^{†††} Furthermore, we observe that in regions where the No Tweeter volume was not able to model the referendum result, the Undecided Tweeter volume was able to compensate for the difference, thereby highlighting the level of doubt and uncertainty among Colombia's political milieu. We posit that the uncertainty around a referendum outcome is introduced by the Undecided users. Although we have previously showcased that the Undecided Tweeters are closer to the No Twitter

^{†††}<http://plebiscito.registraduria.gov.co/>

^{†††}Twitter representation is influenced by various demographic, social, and economic factors. It is observed that as a city's population increases, its Twitter representation rate also increases.⁵² In addition, economic factors such as the availability of technology infrastructures, like high-speed Internet and connectivity, affect the number of Twitter (or social media) users. Subsequently, the Twitter data set recorded 89.8% of the data emerging from urban centers Bogota, Caribe, and Pacifica.

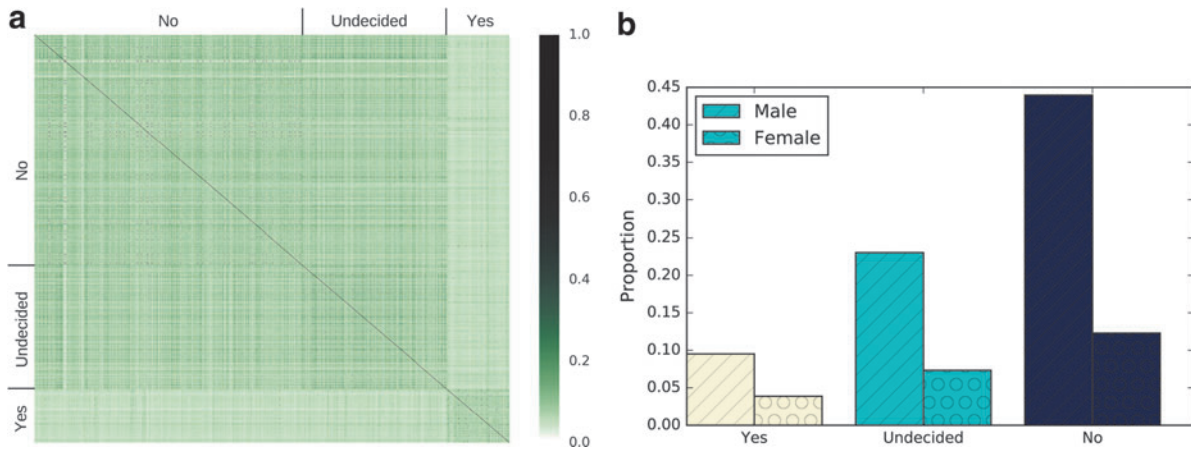


FIG. 6. User differences based on content and gender. Analysis to characterize the differences in the users of polarized groups based on the Tweet content **(a)** and demographic attributes **(b)**. **(a)** User based content overlap, as computed by jaccard similarity between three classes of polarized users. **(b)** Distribution of users across Yes, Undecided, and No classes based on gender. Authors' calculations based on the collected Twitter data set.

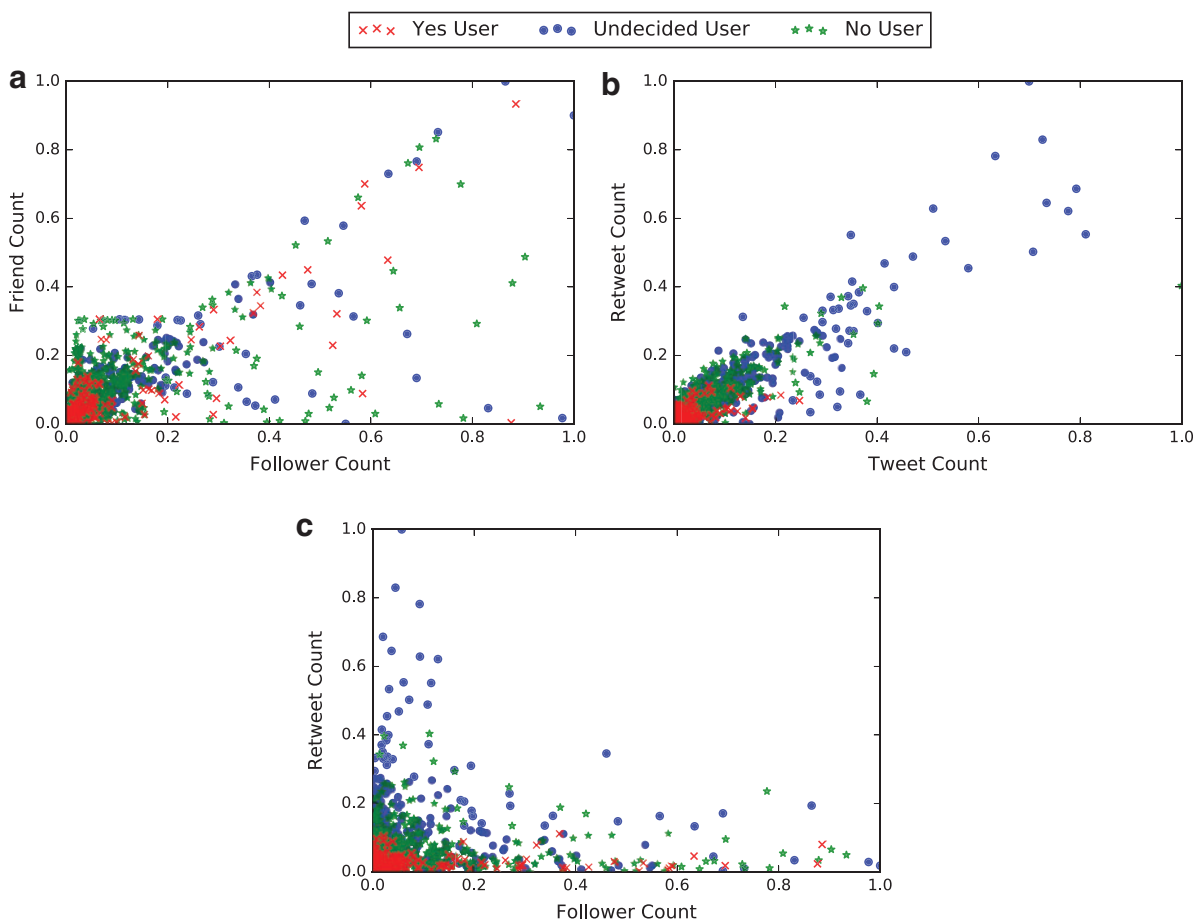


FIG. 7. User analysis based on network and activity. Polarized user activity distribution across key social network and activity features. Analysis includes **(a)** friend and follower count **(b)** Retweet and Tweet count, and **(c)** Retweet and follower count. Authors' calculations based on the collected Twitter data set.

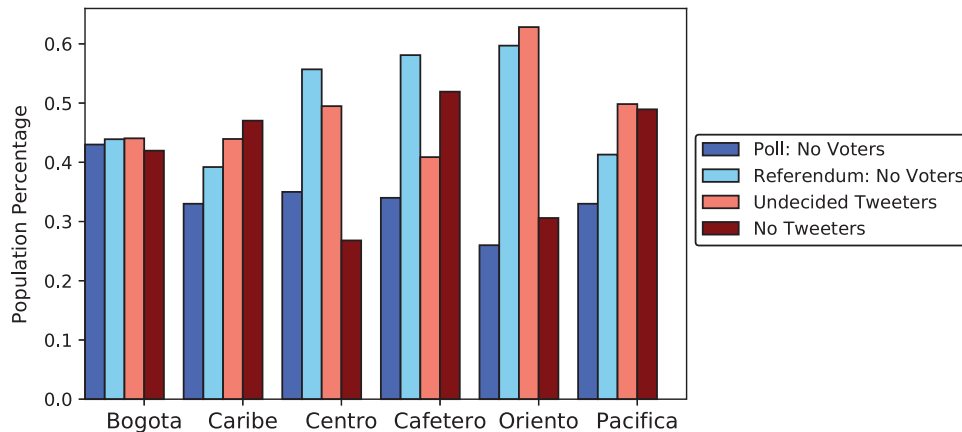


FIG. 8. Relevance of Twitter results to the referendum outcome. Distribution of voters across political regions based on polls and referendum results. Distribution of Tweeters (polarized users) depicted across same regions. Authors' calculations based on the collected Twitter data set.

users in terms of content, demographics, and activity, the unpredictability of their alignment at the time of voting makes the referendum uncertain.

Based on these results, we believe that social media platforms can be leveraged to provide novel insights about public opinion and sentiment toward election and referendum outcomes. Especially, when polling mechanism might not be able to estimate the pulse of the nation, social media can be used in conjunction to harvest signals toward better understanding of major processes and events of great social, economic, and political impact.

Public sentiment

The analyses presented in the previous sections demonstrate the utility of studying the influence of intergroup polarization toward the outcome of a peace process. In this section, we study the predictive power of social media to understand another important peace process outcome indicator—public sentiment toward the peace process. Sentiment analysis can play a pivotal role in discerning public perception about a peace process and can be instrumental in predicting the outcome of a referendum. From a subset (2142 Tweets manually annotated by a Spanish speaker as discussed in the Data collection and preprocessing section) of Tweets collected in the 21 day period of our study, we observed large variations in public sentiment toward the peace process. To achieve this, we built a predictive framework to distinguish negative

from positive Tweets. We extracted three categories of features from users' Tweets (as shown in Table 4): (1) Tweet-based features such as POS distribution and number of negative words, (2) content-based features such as TF-IDF representation of the text, and (3) user-based features such as follower count.

We split the data set into stratified train (80%) and test (20%) samples (1713 train and 429 test Tweets). For the LDA content representation, we learned 100 topics from the train data and the test data were projected on the learned topics. We chose the number of topics by performing a grid search using topic coherence as the metric.⁵⁰ For the TF-IDF representation, we chose 500 most frequent words. We used Tweet, content, and user-based features to train various models such as SVM,⁵⁸ RF,⁵⁹ LR,⁶⁰ Bagged Decision Tree,⁶⁴ Adaboost,⁶⁵ Naive Bayes,⁶⁶ and a random classifier. For all experiments, parametrization was done by fivefold cross validation. The results are shown in Table 10.

Through exhaustive experimentation of various features and feature combinations, we were able to estimate the sentiment of a Tweet with a predictive power of 87.2%, using all the features and TF-IDF representation of the text (as shown in Table 10). We further studied how different categories of features compare for the prediction of sentiment. Using the best performing model from our previous experiment (TF-IDF representation and RF), we compared different permutations of the features. From Figure 9,

Table 10. Public sentiment prediction results

Model	Precision	Recall	F_1	AUC	Accuracy
LDA^a					
RF	0.758	0.808	0.782	0.833	0.776
SVM	0.700	0.546	0.614	0.691	0.657
LR	0.698	0.714	0.706	0.781	0.703
BDT	0.747	0.761	0.754	0.805	0.752
ADB	0.726	0.742	0.734	0.800	0.731
NB	0.507	0.990	0.667	0.708	0.508
Random	0.509	0.518	0.513	0.515	0.510
TF-IDF^a					
RF	0.775	0.855	0.813	0.872	0.804
SVM	0.498	1.000	0.665	0.311	0.498
LR	0.763	0.752	0.757	0.831	0.759
BDT	0.733	0.808	0.768	0.819	0.757
ADB	0.769	0.766	0.768	0.834	0.769
NB	0.503	0.990	0.667	0.709	0.508
Random	0.540	0.542	0.542	0.517	0.543

Metrics used precision, recall, F1 are reported for the negative sentiment label, AUC and accuracy. Authors' calculations based on the collected Twitter data set.

^aLDA and TF-IDF are the two different content-based representations used for the text of the Tweet. With LDA, we achieve AUC=0.833 and accuracy=0.776. However, using TF-IDF representation we achieve the best performance of AUC=0.872 and accuracy=0.804.

ADB, Adaboost; AUC, area under the ROC; BDT, Bagged Decision Tree; LR, Logistic Regression; NB, Naive Bayes; RFs, Random Forests; SVMs, Support Vector Machines.

Table 11. Feature importance: Ranks 30 most predictive features for public sentiment

Category	Feature	Rank	Importance	
Tweet-based	No. of urls	3	0.0339	
	No. of words	5	0.0302	
	No. of adverbs	7	0.0265	
	Retweet Count	9	0.0254	
	No. of -ve words	10	0.0249	
	No. of verbs	11	0.0239	
	No. of nouns	12	0.0238	
	No. of prepositions	14	0.0194	
	No. of determiners	15	0.0191	
	No. of pronouns	16	0.0185	
	No. of conjunctions	17	0.0173	
	No. of adjectives	18	0.0163	
	Retweet (Yes/No)	19	0.0161	
	No. of mentions	20	0.0153	
	No. of +ve words	21	0.0151	
	No. of hashtags	23	0.0129	
	Favorite count	27	0.0089	
	Content-based	Farc	1	0.0468
		No	2	0.0354
		Paz	8	0.0257
		Plebiscito	13	0.0219
		Acuerdo	22	0.0141
		Si	24	0.0125
		Conflicto	25	0.0118
		Santos	26	0.0093
		Fin	28	0.0087
		Proceso	29	0.0086
User-based	Armado	30	0.0072	
	Follower count	4	0.0309	
	Friend count	6	0.0288	

Authors' calculations based on the collected Twitter data set.

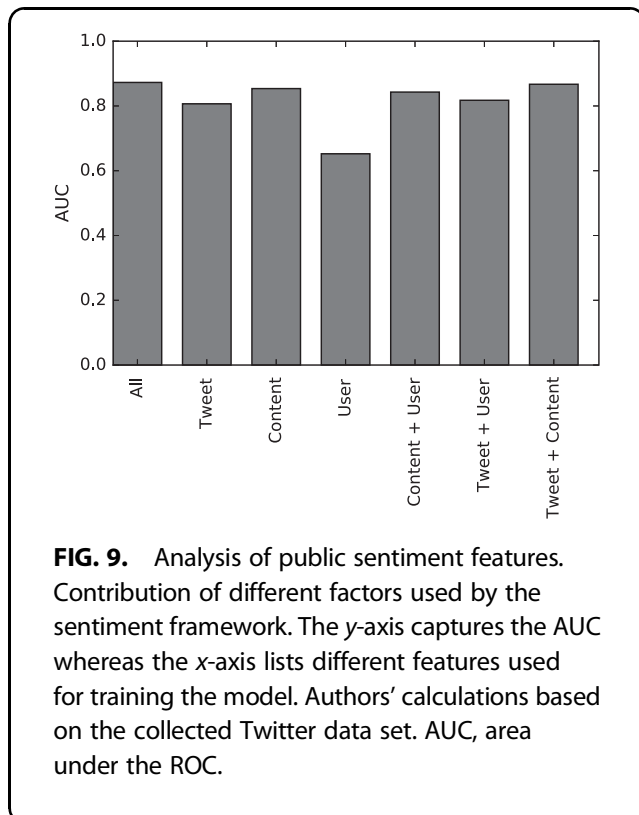


FIG. 9. Analysis of public sentiment features. Contribution of different factors used by the sentiment framework. The y-axis captures the AUC whereas the x-axis lists different features used for training the model. Authors' calculations based on the collected Twitter data set. AUC, area under the ROC.

we observed that model with all three categories of features performed the best.

Furthermore, we investigated the relative importance of features for prediction. Table 11 shows the 30 most predictive features as ranked by RFs. According to our results, the five most important features were (1) TF-IDF score of word farc, (2) TF-IDF score of word no, (3) number of urls present in a Tweet, (4) follower count of the user, and (5) number of words in the Tweet. The two most important features highlight the importance of words and their context (meaning and usage) in understanding sentiment. We also observed that the number of negative and positive words and the distribution over POS are contributing factors, demonstrating their importance and relevance to sentiment prediction.

Conclusion

Our study demonstrates how social media can be used to understand complex and emergent sociopolitical phenomena in the context of civil war termination through negotiated peace settlements. We believe

that, before the October 2, 2016, national referendum, had proaccord stakeholders used social media as a tool or as a platform to listen and understand public opinion and sentiment toward the Colombian peace agreement, the outcome of the referendum could have been different and in the longer term it could have resulted in a more successful implementation of the peace process.

Through this study, we show the importance of two social signals, intergroup polarization and public sentiment, toward peace process outcomes. We posit that social media can further our knowledge and understanding of critical world issues such as civil war termination and postconflict peacebuilding. Considering the pervasiveness of social media platforms such as Twitter and the wide acceptance of technology (or wider penetration of technology in everyday life), ordinary people are empowered to share their opinions about important issues that matter most to them and influence major societal-scale events that can lead to peace or conflict. Therefore, despite the demographic, social, and economic biases seen on social media sites such as Twitter, we observe that the signals can still be indicative of public opinions and perceptions toward peace processes. Therefore, the ability to harvest social signals from such mediums to predict huge sociopolitical outcomes can provide early warning signals to help negotiators and policymakers adjust their approaches to strategic peacebuilding, and ultimately avoid negative cascading effects in a timely manner.

Although we restrict our study to the Colombian peace process, it provides a framework to evaluate social media's impact on peace processes. Our work provides a foundation for future that must be devoted toward developing a concurrent social media-based model for monitoring the implementation of peace agreements to understand public (dis)satisfaction with the delivery or lack of delivery of specific reforms and stipulations negotiated in a peace agreement. It is particularly important to harvest social signals from such mediums because, to an extent, social media platforms shape our political landscape by determining the way we engage in political discourse and with each other. The digital landscape, therefore, influences our decision-making environment, particularly when deciding whether and how to participate in politics, which, in turn, determines how mobilization, support, and disapproval of peace processes build up. Moreover, beyond peace processes, our work highlights the impor-

tance and potential use of social media platforms to better understand public opinion and sentiment on other major public policy issues such as healthcare reform, immigration, climate change, refugees, or women's rights to name a few.

Acknowledgments

We would like to thank Heather Roy and Jonathan Bakdash from U.S. Army Research Lab for providing feedback on the article. This work is supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 and by the National Science Foundation (NSF) Grant IIS-1447795.

Author Disclosure Statement

No competing financial interests exist.

References

- Bond RM, Fariss CH, Jones JJ, et al. A 61-million-person experiment in social influence and political mobilization. *Nature*. 2012;489:295–298.
- Hong S, Nadler D. Does the early bird move the polls? The use of the social media tool 'Twitter' by U.S. politicians and its impact on public opinion. In: *dg.o: Digital Government Innovation in Challenging Times*, 2011. College Park, MD, pp. 182–186.
- Wang D, Abdelzaher T, Kaplan L. *Social sensing: Building reliable systems on unreliable data*, 1st edition. San Francisco, CA: Morgan Kaufmann Publishers, Inc. 2015.
- Borge-Holthoefer J, Magdy W, Darwish K, et al. Content and network dynamics behind Egyptian political polarization on Twitter. In: *CSCW*, Vancouver, Canada, 2015. pp. 700–711.
- Lietz H, Wagner C, Bleier A, et al. When politicians talk: Assessing online conversational practices of political parties on Twitter. *CoRR*. 2014, abs/1405.6824.
- Vera Liao Q, Fu W, Strohmaier M. #snowden: Understanding biases introduced by behavioral differences of opinion groups on social media. In: *CHI*, ACM, San Jose, CA, 2016. pp. 3352–3363.
- Kavanaugh A, Sheetz SD, Skandrani H, et al. The use and impact of social media during the 2011 Tunisian revolution. In: *dg.o., ACM*, Shanghai, China, 2016. pp. 20–30.
- Stieglitz S, Dang-Xuan L. Social media and political communication: A social media analytics framework. *Soc Netw Anal Mining*. 2013;3:1277–1291.
- Conover M, Ratkiewicz J, Francisco JR, et al. Political polarization on Twitter. *ICWSM*. 2011;133:89–96.
- Conover MD, Gonçalves B, Ratkiewicz J, et al. Predicting the political alignment of Twitter users. In: *PASSAT and SocialCom*, IEEE, Boston, MA, 2011. pp. 192–199.
- Golbeck J, Hansen D. Computing political preference among Twitter followers. In: *SIGCHI*, ACM, Vancouver, Canada, 2011. pp. 1105–1108.
- Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. *Bus Horiz*. 2010;53:59–68.
- Margetts H. Political behaviour and the acoustics of social media. *Nat Hum Behav*. 2017;1:0086.
- Asur S, Huberman BA. Predicting the future with social media. In: *Proceedings of the 2010 IEEE/WIC/ACM WI-AIT*, IEEE Computer Society, Toronto, Canada, 2010. pp. 492–499.
- Llewellyn C, Cram L. Brexit? Analyzing opinion on the uk-eu referendum within Twitter. In: *ICWSM*, Cologne, Germany, 2016. pp. 760–761.
- Khatua A, Khatua A. Leave or remain? deciphering brexit deliberations on Twitter. In: *ICDMW, DMIP Workshop*, Barcelona, Spain, 2016, 2016.
- Lazer D, Tsour O, Eliassi-Rad T. Understanding offline political systems by mining online political data. In: *WSDM*, ACM, San Francisco, CA, 2016. pp. 687–688.
- Wang Y, Luo J, Niemi R, Li Y, Hu T. Catching fire via "likes": Inferring topic preferences of trump followers on Twitter. *arXiv*. 2016;arXiv:1603.03099.

19. Wang Y, Li Y, Luo J. Deciphering the 2016 us presidential campaign in the Twitter sphere: A comparison of the trumpists and clintonists. arXiv. 2016;arXiv:1603.03097.
20. Wang Y, Luo J, Niemi R, Li Y. To follow or not to follow: Analyzing the growth patterns of the trumpists on Twitter. arXiv. 2016;arXiv:1603.08174.
21. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. From Tweets to polls: Linking text sentiment to public opinion time series. ICWSM. 2010;11:1–2.
22. Birmingham A, Smeaton AF. On using Twitter to monitor political sentiment and predict election results. In: SAAIP, Chiang Mai, Thailand, 2011. pp. 2–10.
23. Sunstein CR. Republic.Com 2.0. Princeton, NJ: Princeton University Press, 2007.
24. Sunstein CR. Republic.Com. Princeton, NJ: Princeton University Press, 2001.
25. Call CT, Cousens EM. Ending wars and building peace: International responses to war-torn societies1. *Int Stud Perspect*. 2008;9:1–21.
26. Kreutz J. How and when armed conflicts end: Introducing the UCDP conflict termination dataset. *J Peace Res*. 2010;47:243–250.
27. Licklider R. Peace time: Cease-fire agreements and the durability of peace by Virginia Page Fortna. *Pol Sci Q*. 2005;120:149–151.
28. Joshi M, Melander E, Quinn J. Systemic peace, multiple terminations, and a trend towards long-term civil war reduction. In: ISA Annual Conference, Atlanta, GA, 2016. pp. 16–19.
29. Joshi M, Quinn JM. Is the sum greater than the parts? The terms of civil war peace agreements and the commitment problem revisited. *Negotiation J*. 2015;31:7–30.
30. Hartzell C, Hoddie M, Rothchild D. Stabilizing the peace after civil war: An investigation of some key variables. *Int Organ*. 2001;55:183–208.
31. Stedman SJ. Spoiler problems in peace processes. *Int Secur*. 1997;22:5–53.
32. Darby J, MacGinty R. Contemporary peacemaking. Springer. 2002.
33. Moffett M. "The spiral of silence": How pollsters got the Colombia-FARC peace deal vote so wrong. *Vox*. 2016.
34. Cederman L, Weidmann NB. Predicting armed conflict: Time to adjust our expectations? *Science*. 2017;355:474–476.
35. Collier P, Elliott VL. Breaking the conflict trap: Civil war and development policy. World Bank Publications, 2003.
36. US Army. Fm 3-07 stability operations. Department of the Army (US), 2008. Available online at <http://usacac.army.mil/cac2/repository/FM307/FM3-07.pdf>
37. Campoy A. Colombia and the FARC have another peace deal, and this one's not being left up to a referendum. *Quartz*. 2016.
38. Colombo S. The leadership clash that led Colombia to vote against peace. *Harvard Business Review* 2016.
39. Alpert M. There is no plan B for the FARC deal. *The Atlantic*. Atlantic Media Company, 2016.
40. Bell C. Peace agreements: Their nature and legal status. *Am J Int Law*. 2006;100:373–412.
41. Joshi M, Quinn JM. Implementing the peace: The aggregate implementation of comprehensive peace agreements and peace duration after intrastate armed conflict. *Br J Political Sci*. 2017;47:869–892.
42. Joshi M, Quinn JM. Watch and learn: Spillover effects of peace accord implementation on non-signatory armed groups. *Res Polit*. 2016;3:2053168016640558.
43. Loper E, Bird S. Nltk: The natural language toolkit. In: ACL ETMTNLP, Philadelphia, PA, 2002. pp. 63–70.
44. Abrams D, Wetherell M, Cochrane S, et al. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization*. *Br J Soc Psychol*. 1990;29:97–119.
45. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In: NIPS, Curran Associates, Inc., Lake Tahoe, CA, 2013. pp. 3111–3119.
46. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Jebara T, Xing EP (Eds.): ICML-JMLR Workshop and Conference Proceedings, Beijing, China, 2014. pp. 1188–1196.
47. Bouma G. Normalized (pointwise) mutual information in collocation extraction. In: Biennial GSCL Conference, Postdam, Germany, 2009. p. 156.
48. Church KW, Hanks P. Word association norms, mutual information, and lexicography. *Comput Linguist*. 1990;16:22–29.
49. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Machine Learn Res*. 2003;3:993–1022.
50. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: WSDM, ACM, Shanghai, China, 2015. pp. 399–408.
51. Niwattanakul S, Singthongchai J, Naenudorn E, et al. Using of Jaccard coefficient for keywords similarity. In: IMECS, Hong Kong, volume 1, 2013. p. 6.
52. Mislove A, Lehmann S, Ahn Y, et al. Understanding the demographics of Twitter users. In: ICWSM, Barcelona, Spain, 2011.
53. Toutanova K, Manning CD. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: EMNLP SIGDAT ACL, Hong Kong, 2000. pp. 63–70.
54. Toutanova K, Klein D, Manning CD, Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In: NAACL, 2003, pp. 173–180.
55. Chen Y, Skiena S. Building sentiment lexicons for all major languages. In: ACL, Baltimore, MD, 2014. pp. 383–389.
56. Salton G, Wong A, Yang CS. A vector space model for automatic indexing. *Commun ACM*. 1975;18:613–620.
57. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*. 2011;12:2825–2830.
58. Cortes C, Vapnik V. Support vector machine. *Machine Learn*. 1995;20:273–297.
59. Liaw A, Wiener M, et al. Classification and regression by randomforest. *R News*. 2002;2:18–22.
60. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression, volume 398. John Wiley & Sons, 2013.
61. Davis J, Goadrich M. The relationship between precision-recall and roc curves. In: ICML, ACM, Pittsburgh, PA, 2006. pp. 233–240.
62. Lin Y, Margolin D, Keegan B, et al. # Bigbirds never die: Understanding social dynamics of emergent hashtag. arXiv. 2013;arXiv:1303.7144.
63. BBC. Colombia referendum: Voters reject FARC peace deal. BBC 2016.
64. Freund Y, Mason L. The alternating decision tree learning algorithm. In: ICML, volume 99, Bled, Slovenia, 1999. pp. 124–133.
65. Rätsch G, Onoda T, Müller K-R. Soft margins for adaboost. *Machine Learn*. 2001;42:287–320.
66. McCallum A, Nigam K. A comparison of event models for naive bayes text classification. In: AAI-98 Workshop on Learning for Text Categorization, volume 752, Madison, WI, 1998. pp. 41–48.

Cite this article as: Nigam A, Dambanemuya HK, Joshi M, Chawla NV (2017) Harvesting social signals to inform peace processes implementation and monitoring. *Big Data* 5:4, 337–355, DOI: 10.1089/big.2017.0055.

Abbreviations Used

LDA = Latent Dirichlet Allocation
 LR = Logistic Regression
 POS = parts of speech
 PMI = pointwise mutual information
 RF = Random Forest
 SVM = Support Vector Machine
 TF-IDF = term frequency-inverse document frequency