Mixed-Initiative Social Media Analytics at the World Bank

Observations of Citizen Sentiment in Twitter Data to Explore “Trust” of Political Actors and State Institutions and its Relationship to Social Protest

Nadya A. Calderon and Brian Fisher, PhD
Simon Fraser University

Jeff Hemsley, PhD and Billy Ceskavich
Syracuse University

Greg Jansen and Richard Marciano, PhD
University of Maryland

Victoria L. Lemieux, PhD
The World Bank and
The University of British Columbia

***Abstract*—This paper discusses a project that studied the relationship between citizen trust and social protest using visual analysis of approximately 11 million sentiment classified Tweets from the period of the 2014 Brazilian World Cup. The results of the study reveal that the 2014 World Cup protests in Brazil sprang from a wide range of grievances coupled with a relative sense of deprivation compared with emergent comparative ‘standards’. This sense of grievance gave rise to sentiments that activated online protest that may have led to other forms of social protest, such as demonstrations. The paper describes an innovative approach to big data analytics – mixed initiative social media analytics – and discusses the potential of using big data in social science research of this kind, as well as some of the open methodological, technical and ethical issues still to be addressed.

Keywords—Visual analytics, interactive machine learning, political trust, governance, social protest, social media

I. Introduction

According to the OECD, trust in government is an important foundation for legitimacy and sustainability of political systems, and is essential for social cohesion and wellbeing as it affects governments’ ability to govern and enables them to act without having to resort to coercion [31]. Motivated by the significant decline of citizen’s trust in governments over the past decades and the events of the “Arab Spring”, this paper discusses a project that studied the relationship between citizen “trust”, specifically in connection to political actors and service delivery in state institutions, and social protest using visual analysis of approximately 11 million sentiment classified Tweets from the period of the 2014 Brazilian World Cup. The study sought answers to three key questions from the Tweets: 1) how Brazilian citizens felt about their state institutions at the time, 2) how these feelings connected to their sentiments about Brazilian Federal and State government and politicians and 3) how such sentiments translated into collective behaviors. Traditional methods of enquiry into these topics have often involved collection of data using household surveys to poll individuals for their opinions on politicians or public service delivery, or have employed field ethnography to explore, for example, the dynamics of social protest. Big social media data offers the promise of exploring these issues at scale and at a distance, and the possibility of uncovering new insights from innovative analytic techniques. At the same time, there remain many open methodological, technical and ethical questions to address.

II. Background

A. *Background on the context of the project*

The 2014 FIFA World Cup was the largest sporting event of its kind ever hosted in Brazil. In addition to this, it was also the most expensive, with an estimated cost of as much as US$14 billion according to some reports [48,49].

During and in the immediate aftermath of the 2014 World Cup there were public protests in several Brazilian cities, some resulting in arrests and injuries. Various reasons for the protests were cited in media reports at the time and in subsequent studies. These reports point to discontent with the level of spending on the Cup [49]. The lack of debate in Brazilian Society about the competition and its objectives has also been cited as a contributing factor [51]. Also a factor according to some sources was the threat of increased taxation and a higher cost of living due to World Cup spending [49]. News reports at the time also highlighted concerns about corruption [50]. Aside from the immediate causes of the 2014 World Cup protests, this research project sought to explore whether the protests related to declining levels of citizen trust in political actors and service delivery by state institutions.

B. Social media analytics in governance

The literature on social movements has devoted considerable attention to the relationship between social media usage and physical protests [1,6], the formation of networks of contention [10,21] and the increase in speed and scale of political networks [4,5]. The literature also has explored how social media has facilitated organization and horizontal logistical coordination [41] and informal learning about social movements, and social media platforms as positive settings for the construction of elective social affinities [32]. Bruns and Burgess find that hashtagged messages are associated with the formation of ad hoc publics, with Twitter tags being organized as "modern agoras" engineered to provide visibility and press coverage to events.

In other works, scholars have employed computational analysis on large datasets to quantify the impact of social media on social movements, though they have largely focused on the growth of online activity [39]. Others have examined the relationship between the on- and off-line spread of the Occupy Wall Street Movement [10]. These authors have generally focused on when Facebook pages for different cities (e.g. OccupyOakland or OccupyCincinnati) became active, or looked at the volumes of Tweets containing similar city-related hashtags or account names (e.g. #OccupyDenver or @OccupyDenver). Others [1,39] have examined the temporal dynamics of the volumes of tweets related to Occupy, finding that bursts of online activity can be correlated to off-line events that are reported in the media. Work by Nahon et al. suggests that when unexpected events happen on the ground, they tend to be reported on Twitter first, followed by the mainstream media, but with planned events the tendency is reversed [30]. In general, these authors have used computational methods on large datasets to look at temporal aspects of social media use during social movements. Often using simple text mining techniques to track term use over time, combined with social network analysis [39, 21]. The study of social movements using large scale Twitter data has thus far not employed sentiment analysis or mixed-initiative social media analytics methods

C. Trust and Social Protest

Concerns about the impact of declining public trust have given rise to an extensive body of literature aimed at explaining the phenomenon. Within this literature, the concept of trust is defined in many different, sometimes contradictory, ways. Indeed, Edwards notes that "The concept of trust is: vague, slippery, subjective, mutual, multidimensional, multidiscipinary and complex" [15]. Political trust generally equates to what is referred to in some of the literature as citizens' or public trust in government. The World Bank defines Political trust as "an assessment by citizens of a government's entitlement to enforce its decisions, laws and regulations as well as the probability that it will deliver on its obligations" [45].

Citizens' trust in governments has declined significantly over the past decades [28,42,24,15]. In this study, we were concerned with the relationship between political trust and social protest, e.g., whether rising distrust signals increasing likelihood of social protest.

Trust in government and its relationship with social protest has not received much recent attention in the political science literature informing the field of public sector governance [26], pointing to a lacuna in the literature and an area inviting greater research attention.

Definitions of social protest are hard to find. Thus, this study borrows from a definition of social movements to define social protest as "collectivities [that] give voice to their grievances and concerns about rights, welfare and well-being of themselves and others by engaging in various types of collective action, such as protesting in the streets, that dramatize those grievances and concerns and demand that something be done about them" [37, 3]. Social protest is differentiated from social movements by the absence of continuity over time; in other words, social protest is episodic. Social protest also includes a variety of "repertoires of contention" [37]. These repertoires can range from marches, demonstrations and pamphleteering to riots and looting. In addition, new repertoires have emerged such as occupying [36], hactivism, and other digital forms of protest [14].

Various arguments have been advanced regarding why people protest. Classical social protest theories argue that people participate in protest to express their grievances stemming from relative deprivation, frustration, or perceived injustice [7,18]. Feelings of relative deprivation result from comparison of the protestors situation with a standard – be it their past, someone else's situation, or a cognitive standard such as equity or justice [16]. Foster and Matheson [17], showed that people who experience both personal deprivation and group deprivation are the most strongly motivated to take to the streets. Van Zomeren et al. [43] conclude that the cognitive component of relative deprivation (as reflected in the observation that one receives less than the standard of comparison) has less influence on action participation than the affective component (as expressed by such feelings as dissatisfaction, indignation and discontent about these outcomes)

III. METHODOLOGY

A. Background on the methodological approach

Brazilians are the second highest users of Twitter next to people in the US. There are approximately 41m users of this social media platform in the country [52]. The high Twitter usage rate in Brazil in combination with the 2014 World Cup protests make this country an ideal candidate for exploration of the relationship between trust in government and social protest using mixed initiative social media analytics. This approach relies upon big data generated from social media platforms (i.e., Twitter), and applies analytic techniques that draw from the emerging field of Data Science.

Data Science is defined as the study of generalizable extraction of knowledge from data with special emphasis in the power of prediction [11]. Predictive modelling is concerned with the use of algorithms, leveraging in statistical analysis, to train a model that has "learned" from the data and so, is able to make predictions or decisions about new, unknown observations, without following explicitly programmed instructions [33]. Typically, the event to predict is in the "future", based on learning from past data, but predictive modelling can be applied to any type of unknown event regardless of when it occurred (e.g. can we predict the sentiment valence of an opinion based on learning from other opinions?).

From the outset of the project the research team noted the potential benefits of a mixed-initiative approach where visualization, interaction and machine learning methods are combined. The combination of human and machine analysis offers several benefits. Provost and Fawcett [33], for example, argue that automated discovery and evaluation of patterns from data benefit from the analyst's business knowledge, creativity and common sense. Ribarsky et al. [35], emphasize the value of making "visually enabled tools", coupling visualizations and interactions with the human visual/understanding channel for maximum throughput integrated with human understanding and judgment. Interaction is the coupling mechanism between human reasoning and analytic processes, on the one hand, and computer-based processes on the other hand. Investigative analysis such as the one conducted to understand social media data in the context of political trust theories reflects a natural scenario for mixed-initiative systems where there is a need for human and computer to work together because the analysis involves constructing meaning from data and requires the human ability to create or extend hypotheses. At the same time, the data are too large, the dimensions too high, and the ramifications of a change or decision too many for a human to handle unaided. For this reason, the project team implemented an exploratory framework from data collection to data analysis using, among other tools, a set of tools developed by team members that integrate visualization, interactivity and machine learning.

Visual analytics, often defined as the "science of analytical reasoning facilitated by interactive visual interfaces" [55], provided us with knowledge of interactive methods and investigative methodological approaches while predictive modelling provided methods to operationalize difficult concepts, such as "trust", through a proxy of sentiment classification of opinion within the text of social media data. To highlight the power of a mixed-initiative social media analytics approach, we describe the exploratory analysis work carried out for analysis of collections of Tweets from Brazil. We also describe further work that develops data science methods to assist in the analysis of big social media data by supporting the definition of constructs of concepts of interest within the data, as well as assisting in the evaluation of evidence found in the data for hypotheses evaluation in an interactive-machine learning fashion.

B. Methodological steps

Our methodology comprised the following steps:

1. Operationalizing trust as a measure of sentiment in the content of twitter data
2. Instrumenting trust using a sentiment classifier
3. Initial "big picture" harvest of twitter data (the 11 million Tweets).
4. Visual iterative exploratory analysis and text analysis (Natural Language Processing) of "big picture" Twitter data to extract search terms representing the key concepts underlying our study i.e., state institutions, government, etc.
5. Use of search terms to harvest historical Twitter data for our period of interest i.e., 2014 World Cup in Brazil
6. Sentiment Classification of the harvested Tweets
7. Development and use of a Visual Analytics tool to explore sentiment-annotated historical Twitter collections.
8. Pair analysis for exploration of data.
9. Analysis of competing hypothesis aided by our VA tool to explore the historical Twitter collections and derive findings.

1) Operationalizing trust as a measure of sentiment

Our methodology involved a series of choices. One of the first, and most significant, of these was the choice about how to operationalize the construct of trust. Trust has been a notoriously difficult construct to measure, with much debate and many conflicting methodological approaches [15]. In this study, we chose to use sentiment as a proxy measure of trust, which we defined as affective valence, positive or negative. Sentiment analysis for the social web has become a useful tool for analysis of the role of emotion in online communication, public opinion and reaction to events via public texts [40]. The use of sentiment to measure trust is also common in the field of marketing (see, for example, Boertjes [9] on consumer trust in a cloud computing service). The approach has not been used, to the best of our knowledge, in studies of trust in government. Nevertheless, relying upon the use of feelings or emotions in measuring political trust is not without precedent. Feelings about politicians, institutions and groups has been used to measure political trust by Hetherington [22] as well as an American National Election Studies survey of trust in government (ANES n.d.). Newton and Norris also link the notion of public trust to feelings [29]. In light of these linkages, there is some justification for relying upon sentiment to operationalize the measure of trust.

Through exploratory analysis of sentiment our aim was to track trust in institutions and general public distress, hypothesizing that an increase in negative sentiment signalled declining public trust in government.

2) Instrumenting trust using a sentiment classifier

To instrument our approach to measuring trust, we chose to use a sentiment classifier, SentiStrength [40], as our sentiment analysis tool considering 1) successful evaluation results reported for analysis of short texts, 2) the capability to customize a domain-specific sentiment lexicon, and more importantly 3) the capability to customize language even without labelled training data. Research and open source initiatives have produced alternatives with good results but mainly for analysis of English language, thus, not fitting our main need.

Despite identifying sentiment polarities as sufficient for many applications, texts often contain a mix of positive and negative sentiment and for some applications it is necessary to detect both simultaneously and also to detect the strength of sentiment expressed. SentiStrenght accounts for this, classifying text with both a positive and a negative magnitude ranging between 1 and 5. SentiStrenght is a lexicon-based classifier that uses custom lists of words to identify sentiment. Besides the sentiment words, the algorithm includes booster words, negating words, question words, emoticons, and idioms if specified, meaning it has a rich set of language features. The customization of the Brazilian-Portuguese dictionaries was carried out by two native speakers (inter-raters of sentiment weight for the words) having as a base dictionary, the results of the translated and adapted Affective Norms of English Words (ANEW) for Brazilian Portuguese [27]. ANEW-BR is a set of 1,046 words with emotionality measures that were mapped to our scale of sentiment magnitude

3) Initial "Big Picture" harvest of Twitter data

For this project we collected publicly available Twitter data in two phases. In the first phase we collected data in real time using STACKS, an open source, social media data collection toolkit [20]. The first collection was bounded with a single location filter that fit roughly around the entirety of Brazil. It is important to note that only a small percentage of tweets (approximately 1 percent [3]) contain geolocation data, so this initial collection was intended to sample a large "universe" of Tweets in order to conduct an exploratory analysis of topics of discussion.

4) Visual iterative exploratory analysis and text analysis

We implemented an iterative inductive approach to narrow down the stream of data and build up collections defined by topical themes of interest. This "big picture" approach was used in order to capture ways in which citizens communicate or discuss themes, rather than constructing topics only from dictionary terms and definitions. Ultimately, this approach was used to determine the search phrases that would more naturalistically capture topics of interest - government service institutions and citizen opinion about government to track indicators of trust – within a body of Tweets, thereby avoiding imposing our own views on the topics of importance and the terms used to describe these topics.

Exploratory analysis of text can be done using a toolset of options ranging from manually reading hundreds of documents to selectively querying a database. However, the reading intensity and the time cost of these options are too high. Especially when working with thousands of short texts and trying to make sense of groupings of them. Visual Analytics is a robust approach for these types of tasks, being an analytical approach that accounts for the need of a human-in-the-loop in the analytic sense-making process.

For the analysis of the universe, or "big picture", of Tweets we used IN-SPIRE visual text analysis tool [46], and n-gram based text categorization using the NLTK Python package. IN-SPIRE is a visual analysis tool for the analysis of large sets of unformatted text documents by clustering similar documents together, determining key topics or themes in each document within the collection.

IN-SPIRE's galaxy view gave us a rough classification of easily identifiable popular culture themes such as TV shows, sports games, and mainstream actors and singers. By iteratively filtering out documents associated to these themes, which were not relevant to the focus of our study, political opinion commentary related to President Dilma Rousseff and Petrobras began to appear. In order to identify other documents relevant to our interests we searched (and collected a list of) general terms of services and service institutions such as: *educação, serviços, saúde, hospital, polícia, faculdade*. We extended the list of words people use for political commentary and to refer to institutions by filtering, searching and reading individual documents. We found commentary around the drought crisis in some states as a source of service delivery opinion. IN-SPIRE helped us to read the documents in a guided exploratory way, and to construct a list of terms to investigate further. However, due to constraints about the length of Tweets, language, and character encoding, we did not use IN-SPIRE to define the search phrases, as would be the usual method. Instead, we used the list of seed words to perform word co-occurrence analysis. In this way, we identified other words people use to talk about the water crisis, the current Petrobras events, references to service institutions, or general political opinion. We also performed bi-grams and tri-grams analysis that corroborated most of the phrases found by co-occurrence with a defined list of words

5) Use of search terms to harvest historical twitter data

Data collected and analysed in the first phase informed the selection of search terms employed in the second phase. For the second phase, we purchased historic data from DiscoverText [38] using a tool called Sifter, which is a reseller of data from Gnip, a subsidiary of Twitter. Table 1 summarizes the 3 different collections we compiled as a result, the number of Tweets collected, number of users contributing those Tweets, and examples of the search phrases used to collect the data.

TABLE I. SUMMARY OF HISTORICAL COLLECTIONS

Collection	Search Phrases (selection)	Date Range	Number of Tweets	Number of Users
Brazil "Big Picture"	*-geoLocation boundary*	Jan 29, 2015 - Feb 4, 2015	11,066,766	376,834
Political Opinion	*dilma lula pt, impeachment dilma, dilma precisa*	June 05, 2014 July 30, 2014	205,456	141,530
Service Delivery	educação saúde, educação dilma, falta água, crise hídrica	June 05, 2014 July 30, 2014	10,021	6,217
Petrobras	*petrobrás pt, petrobrás corrupto, corrupção petrobrás*	June 05, 2014 July 30, 2014	46	38

6) Sentiment classification of harvested Tweets

Once we customized the sentiment classifiers, the classification was built into the pipeline of data collection, data cleaning and data extraction. We implemented an integration of

STACK with SentiStrength to execute the classification of Tweets while they are being processed to insertion into the database. In other words, sentiment classification took place in near real-time.

7) Development and use of a Visual Analytics tool to explore sentiment-annotated historical twitter collections

In this project, the exploration of social media data to understand trust, consisted of a set of exploratory and hypothesis testing tasks that required a structured way to assess the text-data. We developed a custom tool (see a snapshot in Figure 1) to facilitate the exploration with special emphasis on visualization of sentiment analysis patterns. Our tool allows the analyst to:

1. Explore aggregated overviews of temporal patterns of Tweets with the ability to overlay external events flags.
2. Explore the Sentiment trends using horizon charts [19]
3. Obtain an overview of the text content with a text cloud and line chart of individual words frequency
4. Read details of individual Tweets to understand digital opinion patterns that may not be obvious for an algorithm or from an aggregated overview of the data.
5. Filter Tweets list by search delimits (words or dates)
6. Sort data by sentiment, date and count of RTs if exploring the de-duplicated list of Tweets.

We designed representations of the data to study the temporal patterns of sentiment extracted by the classifier, to facilitate the text-content reading and to overlay time events relevant to our investigations. For the representation of sentiment over time we use horizon charts [19]. The value of using this uncommon type of time series chart is the ability to compare series with a wide scale of values. We represented nine sentiment categories (Positive from one to four, Neutral (0), and Negative from minus one to minus four). The number of Tweets for a category might range in tenths while others could range in thousands. The horizon charts allow a visual comparison of series with these different scales. The text content summaries are given by a bag-of-words representation using text clouds and line charts depicting distributions of the use of terms over the time frame of the collection. The exploration of individual Tweets was built into a simple sortable table, using number of ReTweets, date, and sentiment value as dimensions available to sort by.

8) Pair Analysis

Having developed the analytic tool and gathered relevant data, the next step involved conducting a visual analysis of the sentiment classified data. The analysis began with a simple pair analysis session aimed at making basic observations from the data about the study's three main research questions.

Pair analysis is a strategy inspired by pair programming practices, designed to conduct data analysis with visual interfaces [47]. It consists of two participants, one Subject Matter Expert (SME) and one Visual Analytics Expert (VAE) that are given one analytical task, one data set, and one computer with a visual analysis tool. While the VAE is in charge of "navigating" the dataset with the tool, the SME guides the exploration. The pairing of VAE and SME in the context of an analytic task is designed to generate a human-human dialog that enhances the expertise in visualization and cognitive processes that mediate analytic tasks, with the contextual knowledge required to conduct meaningful analysis of the data.

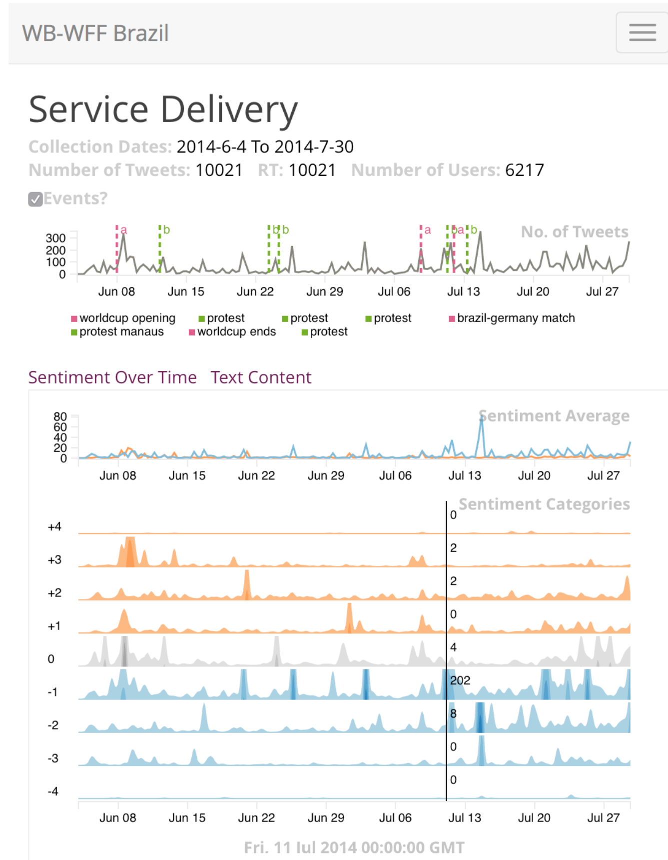

Fig. 1. Exploring a collection with: Summary data, Overlay of events and Sentiment Horizon Charts.

9) Analysis of Competing Hypotheses

Subsequent analysis was informed by Heuer's Analysis of Competing Hypotheses (ACH) methodology [23]. ACH "is an eight-step procedure grounded in basic insights from cognitive psychology, decision analysis, and the scientific method… [it is] a proven process that helps analysts avoid common analytic pitfalls.   Because of its thoroughness, it is particularly appropriate for controversial issues when analysts want to leave an audit trail to show what they considered and how they arrived at their judgment" [23]. ACH is typically conducted by a single analyst though the results may be corroborated by other analyses using the same or complementary methods. Thus, the analysis proceeded with:

1. Identifying the possible hypotheses to be considered.
2. Making a list (matrix) of the hypotheses, and identifying evidence for and against each one.
3. Analyzing the strength of each hypothesis.

4. Iterative refinement of the analysis, deleting evidence and arguments that have no diagnostic value.
5. Drawing tentative conclusions about the relative likelihood of each hypothesis by trying to disprove the hypotheses rather than prove them.

For the analysis of competing hypotheses (ACH), we drew upon the background literature to identify 68 hypotheses for consideration. The hypotheses can be divided into two groups: the first group derives from the literature on trust in government and the second group derives from the literature on social protest. Each hypothesis was considered in terms of the evidence in support of it, evidence against it, and the strength of the evidence in terms of diagnosticity and within sample representativeness of the evidence. The evaluation of the strength of the evidence for each hypothesis in the ACH analysis was based on a qualitative assessment

IV. Findings

A number of hypotheses could not be supported (or refuted) due to lack of evidence (see Table 2 for description of example hypotheses). Many of these hypotheses relate to research question three about how citizen sentiment translates into social protest, especially those "repertoires of contention" that involve public demonstration or collective action in the streets (see H4).

Several of these hypotheses were drawn from the literature on social protest. Another group of hypotheses that could not be supported (or refuted) due to lack of evidence relates to propositions about larger socio-political, technological or economic trends or values, morals and cognition (e.g. H17). The claims in these hypotheses were too distant from the evidence to be found in the Tweets. This suggests an inherent limitation in social media data analytics (i.e., it is difficult to map the data to very broad conceptual categories). Additional research and theoretical framing is necessary in order to find support for (or refute) these types of propositions. A further category of hypotheses for which there was no evidence pro or con were those that required additional data to properly investigate. These included the hypotheses on involvement in social protest mentioned above, but in addition included those that required geocoded data (H27).

Turning to those hypotheses for which we found support, the collection offers the strongest evidence in favor of the relative deprivation theory [16] of social protest coupled with notions of public trust that involve citizens' assessment of their government's ability to follow through on promises, deliver on policy objectives and act with integrity (see also H8, H12):

H48: Feelings of relative deprivation result from comparison of one's situation with a standard – be it one's past, someone else's situation, or a cognitive standard such as equity or justice.

In social theory, relative deprivation is the notion that an individual or group lacks something that another group has and to which they feel some entitlement. The important emphasis here is that they feel this deprivation in relation to some standard, whether static or emergent as in the case of our study (i.e., the 2014 World Cup) or other group. Evidence from this collection indicates that it is not performance or outcomes per se, but performance and outcomes for one group ("povo Brasilieros") relative to another (spending on campaign financing/World Cup). Many of the negative Tweets in the collection (and a good proportion of the positive Tweets) express strong sentiment (e.g., anger, disgust, discontent and disgruntlement) at the spending on the World Cup in relation to public services. Two public services, in particular, are the focus of Tweets in our political opinion collection: education and health, with water being mentioned most in our service delivery collection. We suspect, but cannot definitively prove, that emphasis on the two services in the political opinion collection is a reaction to a speech given by Dilma Rousseff just prior to the start of the World Cup in which the lack of investment in these two public services is mentioned in particular. In other words, resources were made available to spend on the Cup, depriving the "povo Brasilieros" or "Brazilian people" of investments in education, health care, and other public services and a better standard of living. In the public service delivery collection, the water shortage is mentioned more prominently, and there is less focus on relative deprivation in relation to spending on the World Cup and more on the behaviour of state level politicians (e.g., excessive spending on campaign financing) and SABESP, the water authority. It was this sense of relative deprivation that seems most evident during periods exhibiting spikes of heightened sentiment. This theory also explains the wide variety of issues that surfaced in the Tweets, from water, to education, health, police, transport, threat of increased taxes, and inflation – all contributed to the sense of relative deprivation. In short, growing distrust (negative sentiment) was driven by people's sense of a growing distance between the "haves" and the "have nots." This was also coupled with the perception that corrupt and dishonest politicians were personally benefitting, or at least not suffering like everyone else, and not delivering on promises to the people about spending (see H8, H12), which appears to have had an amplifying emotional effect.

TABLE II. Evaluation of competing hypotheses extract.

Hypothesis	Evidence Pro	Evidence Con	Strength[a]	Findings
H4. People who send negative tweets are involved in protest in the streets	Unable to find confirmatory evidence	Unable to find confirmatory evidence	D: Medium R: Low	No direct evidence that people who send negative tweets, or even those calling for protest in the streets, such as demonstrations, actually participate in protests.
H8. Political trust it is an assessment by citizens of a government's entitlement	Some of evidence that political trust (in the form of negative opinion expressed relates to	No evidence of questions of entitlement to enforce decisions (except for decision	D: Medium R: Low	Some weak support for this hypothesis. Regarding the representativenes there is not large volume of tweets that can be directly

to enforce its decisions, laws and regulations as well as the probability that it will deliver on its obligations (World Bank 2010, 52) H12. Political trust "happens when citizens appraise the government and its institutions, policy-making in general and/or the individual political leaders as promise-keeping, efficient, fair and honest".	probability that government will deliver on obligations. These obligations center on themes of public services (education, health, police, water, transport). Also tweets linked to government inefficiency, which is about the capacity to delivery, especially in reference to the water shortage.	to host the World Cup, laws and regulations). However, there are also other themes such as corruption, inflation, etc.		linked to this hypothesis in a systematic way; only tweets discovered while exploring the collections. H12: Very similar hypothesis to H8. Supported.
H17. Social trust is an important determinant of political trust.	No evidence	No evidence	No evidence	No evidence in collection.
H27: The political economy literature is on consensus that higher levels of trust are associated with wealthier areas. Reciprocally, lower degrees of trust go hand in hand with poorer areas.	Unable to test this hypothesis due to limitations of data (no granular level geocode metadata)	Unable to test	Unable to test	Unable to test
H48. Prominent among grievance theories was relative deprivation theory. Feelings of relative deprivation result from comparison of one's situation with a standard – be it one's past, someone else's situation, or a cognitive standard such as equity or justice [16]	There is evidence in the tweets to support this. Many of the negative tweets (and a good proportion of the positive tweets) express anger, disgust at spending on the World Cup, politicians access to fund their priorities (e.g., campaign fining) in comparison with funding on education, health, etc. In this sense, it is not so much the individual's situation but the "we" or collective situation i.e., the povo Brasilieros.	Most of the tweets do lend some form of support for this hypothesis	D: High R: High	Supported

[a.] Strength is determined by 1) Diagnosticity of the evidence, i.e., the degree to which the evidence can be used to support the relative likelihood of alternative hypotheses, indicated on a high, medium to low qualitative scale and 2) Representativeness, i.e., the degree to which the evidence within the collection (not in reference to the broader representativeness of the tweets) indicative of a wider pattern and can be used as the basis for further generalization, also indicated on a high, medium to low scale[23].

V. Reflections on Results of the Study. Futute Work

A. Mixed-Initiative Analytics: Interacting with Machine Learning Models

Exploration and hypotheses testing using Twitter data with sentiment opinion trends provide an example of a visual exploratory analysis that benefits from integrating statistical learning modeling. The sentiment analysis classifier instruments a very relevant feature within our data, establishing a framework of reference to monitor trends and investigate the status of a specific problem. This type of problem sits in the middle of traditional "exploratory" vs. "predictive" analytics space because despite the fact that the tasks at hand are exploratory, handling the volume of data or computing additional features relies on computational modeling.

The ACH in our study demonstrated the need and opportunity for strengthening the qualitative assessment of evidence ideally, by capturing the data and the argumentation features relevant to a piece of information and using it to train a machine to find more.

In this project, exploration is not only the initial stage on a linear analysis pipeline, but rather a set of cognitive tasks that structure the access to, manipulation and transformation of the data in order to make sense of the "current state of events" and to systematically test hypotheses and search for evidence. We believe that traditional methods for analysis of social media data such as topic modelling, sentiment analysis, and social network analysis set a robust framework for investigation but going further, we aim at integrating interactive machine learning capabilities to exploit implicit and explicit human feedback, to evaluate labelled evidence in the space of the "unseen" data, and to reinforce transparency of the investigative process with visualization of the human-machine decisions. We elaborate on four features that could be integrated into our tool to take it from a visual exploration tool to an engine for argumentation in investigative analysis of social media.

B. Sentiment Analysis, Topic Modeling and Social Network Analysis

Text mining and natural language processing techniques allow us to extract structure from otherwise noisy data, yet due to the current state of automated text-processing, users conducting analysis of Twitter data would benefit from the opportunity to modify the algorithm's judgment. The ability to perform sentiment classification in Brazilian Portuguese represented the greatest advantage for our project, yet the classifier suffered from typical errors related to classification of irony and meaning in context (accurately positive-labeled text that expressed a negative feeling about the current government). Visualization and interactivity provide the means to correct for these errors and to allow the flow of analysis to continue. *Thus, in future, we would like to develop our interactive visualization of sentiment classifications to enable corrections and feedback of human judgment.* Additionally, the system should also keep track of such corrections. Whether those corrections take effect on the training of the classifier depends on the specific classifier being used, with this point we aim to highlight the ability to update values based on human evaluation, take those modifications to update the visualizations and continue with the exploration.

We did not explore the structural properties of the network of users in our data but we recognize the relevance for this type of analysis for the investigation of entities in relation to our interest in "political trust". Social network analysis and social contagion in online social protests is a well-documented area of research in the computational social sciences [13] and this is highlighted as part of the work we aim to integrate into our approach.

C. Interactive Machine Learning

Interactive Machine Learning is an area concerned with the development of systems that learn interactively from their end-users, particularly, systems based on Machine Learning (ML) algorithms [2]. Such approach enables the involvement of a human expert in a rapid, interactive and incremental ML process of analysis, for problems where human bias, experience and knowledge play an important role.

During the hypotheses evaluation analysis for our project, the researcher used an external instrument to keep track of findings of evidence to support and refute each of the hypotheses under evaluation (a simple Excel matrix). During this process, the investigator was executing tasks for 1) searching and refining search results (using the filtering mechanisms that the VA tool provided), and 2) classifying and annotating pieces of evidence (making a judgement and labelling findings in categories within the matrix). Performing these tasks in an interactive environment would not only facilitate the analyst's job, but would also create the opportunity for a machine to collect and learn from the human findings. For this reason, *we would like to enable the capture of summaries of arguments and evidence pro and against hypotheses under evaluation and provide the interface for defining and training a model with such argumentation evidence. Models can be derived by traditional ML algorithms, or relevance-feedback information retrieval algorithms*. We would like to develop an analysis tool with interaction mechanisms to allow people to define concepts that can be recognized by an intelligent system in a continuous iterative-interactive dialogue between the end-user and the system until a desired learner is achieved and used to investigate the larger body of "unseen" data.

It is within the space of research of this feature that most limitations come for our research. Initiatives like the ones we are describing: finding other pieces of data based on "features of the data" compared to a standard might indeed represent a practical aid for analytics, but in the end, what a truly mixed-initiative should aim at capturing is the argumentation, not just the data.

We rely on recent works in the Interactive Machine Learning literature, for example in *Power to the People* [2], authors describe cases of end-users of data-analysis tools involved in more rapid, focused and incremental learning cycles to create classifiers. These properties enabled users to interactively explore the model space through trial-and-error and drive the systems towards an intended behaviour effectively.

Propositions about interactive machine learning include: 1) rapid, focused and incremental learning cycles, which result in a tight-coupling between the user and the system, and with the resulting model being created from the influence that the user and the system have on each other; 2) expansion of the interaction strategies from what the skilled ML practitioners do (tuning parameters, defining new constraints, selecting features) to novel strategies (specifying error type preference, providing new labelled samples, not only acting as oracles); and 3) explicitly studying user interaction with machine learning models in this setting to challenge assumptions about traditional learning systems

D. Transparency, Collaboration and Governance

In our study, conclusions about the strength of the evidence for each hypothesis in the ACH analysis were based on a qualitative assessment, we have discussed how the analysis could be strengthened by the application of interactive machine learning techniques, but additionally, the analysis was conducted by a single analyst, and would benefit from an inter-coder reliability check (i.e., having another analyst follow the same research protocol, and comparing the results). To us, the integration of an analytic sandbox is not merely another interactive computational component. It is rather a space that captures and externalizes the evidence and reasoning for specific hypotheses, that also poses the possibility to create analysis by independent analysts and that stores the data about the analysis itself (documenting decisions, saving notes about computed values overwritten by human judgment, and documenting the features and evaluation of models developed). We believe these possibilities account for the need of transparent analysis, validity of qualitative inferences and analytical provenance to learn and iterate from every practice of analysis. *Thus, a future direction is to develop capabilities that allow for the recording of analytic reasoning and access* *to the record in order to more effectively employ the principles of visual analytics*.

VI. Conclusion

In this paper, we outlined an ongoing research work to create mixed-initiative systems to investigate factors that may have motivated the protests during and in the immediate aftermath of the 2014 World Cup. We have connected this analysis to various explanatory hypotheses on the relationship between citizen trust and social protest, which can be used to conduct further exploratory and confirmatory studies related to the 2014 World Cup protests, as well as to reflect back to the literature on citizens' trust in government and social protest in order to contribute to scholarly discourse on these topics.

The work described in this paper has been developed in the context of the Governance and Inclusive Institutions practice at the World Bank and is aimed at creating proactive tools to support them (as well as other practice areas in the bank interested in the use of social media data) in research and operations. We have designed an initial exploratory tool and outlined a methodological framework that includes the harvesting of the data, definition of concepts and constructs, and conducting analysis, using methods from the visual analytics literature.

Although the work developed and presented in this paper is at the stage of exploratory analysis for a case in Brazil, we are also describing the larger scope of our research hoping to better support human judgment in studying these complex issues. We described our desired features to support the analytical discourse by allowing the human analyst 1) to correct machine classifications and to reflect updates in visualizations in real time, 2) to collect evidence that supports argumentation to confirm or reject hypotheses under study 3) to interactively train classifiers that can assist in finding evidence in the larger space of collected data, and 4) to use visualization of analytic data itself throughout the investigative process.

For now, our goal is that different operations in the World Bank can benefit from a framework that recognizes the value of visualization and interaction in the study of the kind of research we conducted i.e., research that aims to unpick complex theoretical and methodological relationships, such as those surrounding research on citizen trust and social protest.

More than a "dashboard" of sentiment analysis data, the framework we describe is a set of tools and methods to investigate social and political issues, an analytical observatory of social media. With this in mind, it is also necessary to consider that the evidence gathered from tweets and online media data must be seen at this point only as exploratory, and as pointing the way to interesting avenues of further exploration and confirmatory studies using complementary research methodologies, such as surveys, ethnographic studies or comparative analysis with other sources of data. Mixed-initiatives data analysis of what citizens are expressing in Tweets offers an innovative new way for researchers to explore socio-political dynamics. We believe that, as institutions like the World Bank see examples of cases relevant to their operations, there will be a motivation to train and use more data-driven approaches facilitated by visualization, interactivity and data science methods to innovate policy and the way governments engage in dialogue with citizens.

References

[1] Agarwal, S., L. Bennett, C. Johnson, S. Walker, (2014). A Model of Crowd Enabled Organization: Theory and Methods for Understanding the Role of Twitter in the Occupy Protests. International Journal of Communication 8(0): 27.

[2] Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105-120.

[3] Bamburic, M. (2012). Twitter: 500 million accounts, billions of tweets, and less than one percent use their location. Retrieved from http://betanews.com/2012/07/31/twitter-500-million-accounts-billions-of-tweets-and-less-than-one-percent-use-their-location/

[4] Bennett, W.L., Breunig, C., & Givens, T. (2008). Communication and political mobilization: Digital media and the organization of anti-Iraq war demonstrations in the US. Political Communication, 25(3), 269-289.

[5] Bennett, W. L., & Segerberg, A. (2009). Collective action dilemmas with individual mobilization through digital networks. Center for Communication and Civic Engagement Working Paper2009, 2

[6] Bennett, W. L. (2012). The personalization of politics political identity, social media, and changing patterns of participation. The ANNALS of the American Academy of Political and Social Science, 644(1), 20-39.

[7] Berkowitz, L. (1972). Frustrations, comparisons, and other sources of emotion arousal as contributors to social unrest. Journal of Social issues, 28(1), 77-91.

[8] Butler, B. N., & Aicher, T. J. (2015). Demonstrations and displacement: social impact and the 2014 FIFA World Cup. Journal of Policy Research in Tourism, Leisure and Events, (ahead-of-print), 1-15.

[9] Boertjes, E., Gerrits, B., Kooij, R., van Maanen, P. P., Raaijmakers, S., & de Wit, J. (2012). Towards a social media-based model of trust and its application. In ICT Critical Infrastructures and Society (pp. 250-263). Springer Berlin Heidelberg.

[10] Conover MD, Ferrara E, Menczer F, Flammini A (2013) The Digital Evolution of Occupy Wall Street. PLoS ONE 8(5): e64679. doi:10.1371/journal.pone.0064679.

[11] Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64–73. doi:10.1145/2500499

[12] Gleason, B. (2013). # Occupy Wall Street: Exploring informal learning about a social movement on Twitter. American Behavioral Scientist, 57: 966-982.

[13] Golder, S. A., & Macy, M. W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. Annual Review of Sociolog y, 40(1), 129–152.

[14] Earl, J., & Kimport, K. (2009). Movement Societies and Digital Protest: Fan Activism and Other Nonpolitical Protest Online*. Sociological Theory, 27(3), 220-243.

[15] Edwards, M. (2015). The trust deficit - concepts and causes of low public trust in governments. Committee of Experts on Public Administration, 14th Session. E/CN.16/2015/CRP.3/Rev.1.

[16] Folger R. (1986). Rethinking equity theory: A referent cognitions model. In Bierhoff HW, Cohen RL and Greenberg J (eds) Justice in Social Relations. Plenum, New York, 145–162.

[17] Foster M.D. & Matheson K. (1999). Perceiving and responding to the personal/group discrimination discrepancy. Personality and Social Psychology Bulletin 25, 1319–1329.

[18] Gurr, T. R., & Leggewie, C. (1970). Why men rebel.

[19] Heer, J., Kong, N., & Agrawala, M. (2009). Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1303-1312). ACM

[20] Hemsley, J., Ceskavich, B., Tanupabrungsun, S. (2014). STACK (Version 1.0). Syracuse University, School of Information Studies. Retrieved from https://github.com/bitslabsyr/stack DOI: 10.5281/zenodo.12388

[21] Hemsley, J., & Eckert, J. (2014, January). Examining the Role of" Place" in Twitter Networks through the Lens of Contentious Politics. In System Sciences (HICSS), 2014 47th Hawaii International Conference on (pp. 1844-1853). IEEE.

[22] Hetherington, M. J. (2005). Why trust matters: Declining political trust and the demise of American liberalism. Princeton University Press.

[23] Heuer, R. J. (1999). Psychology of intelligence analysis. Center for the Study of Intelligence, Central Intelligence Agency.

[24] Keele, L. (2007). Social capital and the dynamics of trust in government. American Journal of Political Science, 51(2), 241-254.

[25] Keim, D. A., Mansmann, F., & Thomas, J. (2010). Visual Analytics: How Much Visualization and How Much Analytics? *SIGKDD Explor. Newsl.*, *11*(2), 5–8.

[26] Klandermans, B., & Roggeband, C. (Eds.). (2007). Handbook of social movements across disciplines. Springer Science & Business Media.

[27] Kristensen, C.H., Gomes, C.F.A., Justo, A.R., Vieira, K. (2011). Brazilian norms for the Affective Norms for English Words. Trends Psychiatry Psychother. 33(3) – 135-146.

[28] Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. Annual review of political science, 3(1), 475-507.

[29] Newton, K., & Norris, P. (2000). Confidence in public institutions. Disaffected democracies. What's troubling the trilateral countries.

[30] Nahon, K., Hemsley, J., Mason, R. M., Walker, S., & Eckert, J. (2013). Information Flows in Events of Political Unrest. In iConference 2013 Proceedings. Fort Worth, TX.

[31] OECD (2013) Government at a Glance 2013, OECD, Paris

[32] Papacharissi, Z., & de Fatima Oliveira, M. (2011, September). The rhythms of news storytelling on Twitter: Coverage of the January 25th Egyptian uprising on Twitter. In World Association for Public Opinion Research Conference (Vol. 312, p. 3188).

[33] Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.

[34] Reuters. (2013). Brazil Beats Japan, Protests Spoil Confederations Cup Opening Day. Retrieved from http://www.voanews.com/content/brazil-japan-confederations-cup/1682679.html

[35] Ribarsky, W., Fisher, B., & Pottenger, W. M. (2009). Science of Analytical Reasoning. Information Visualization, 8(4), 254–262. doi:10.1057/ivs.2009.28

[36] Sassen, S. (2011). The global street comes to Wall Street. Possible Futures. Retrieved from http://www.possible-futures.org/2011/11/22/the-global-street-comes-to-wall-street/

[37] Snow, D. A., Soule, S. A., & Kriesi, H. (Eds.). (2008). The Blackwell companion to social movements. John Wiley & Sons.

[38] Shulman, S. (2011). DiscoverText: Software Training to Unlock the Power of Text. UMass Amherst, Political Science Department.

[39] Tan, L., Ponnam, S., Gillham, P., Edwards, B., & Johnson, E. (2013, August). Analyzing the impact of social media on social movements: a computational study on Twitter and the Occupy Wall Street movement. In Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on (pp. 1259-1266). IEEE.

[40] Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558

[41] Theocharis, Y. (2013). The wealth of (occupation) networks? Communication patterns and information distribution in a Twitter protest network. Journal of Information Technology & Politics, 10(1), 35-56.

[42] Tolbert, C. J., & Mossberger, K. (2006). The effects of e‐government on trust and confidence in government. Public Administration Review, 66(3), 354-369.

[43] Van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy. Journal of personality and social psychology, 87(5), 649.

[44] Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). The Sandbox for Analysis: Concepts and Methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 801–810). New York, NY, USA.

[45] World Bank. (2010). Results, Performance Budgeting and Trust in Government. World Bank, Washington, DC.

[46] Wong PC, EG Hetzler, C Posse, MA Whiting, SL Havre, NO Cramer, A Shah, M Singhal, AE Turner, and J Thomas. (2004). "IN-SPIRE InfoVis 2004 Contest Entry." In IEEE Symposium on Information Visualization. InfoViz 2004.

[47] Arias-Hernandez, R., Kaastra, L. T., Green, T. M., & Fisher, B. (2011). Pair Analytics: Capturing Reasoning Processes in Collaborative Visual Analytics. In *2011 44th Hawaii International Conference on System Sciences (HICSS)* (pp. 1–10).

[48] R. Downie (2014). Soccer-Brazil World Cup stadiums on track, but costs soar. Available: http://www.reuters.com/article/2012/04/03/soccer-world-brazil-idUSL2E8F2GG820120403

[49] Associated Press (2014), World Cup set to be most lucrative ever. Available: http://www.espnfc.com/fifa-world-cup/story/1830732/2014-world-cup-set-to-be-most-lucrative-ever

[50] T. Panja, T. (2014), Corruption to Blame for Some Brazil World Cup Cost Rises. Availbale: http://www.bloomberg.com/news/articles/2014-05-22/corruption-to-blame-for-some-brazil-world-cup-cost-rises

[51] B.N. Butler & T.J. Aicher, T. J. (2015), Demonstrations and displacement: social impact and the 2014 FIFA World Cup, *Journal of Policy Research in Tourism, Leisure and Events*, (ahead-of-print), 1-15

[52] Associated Press (2014), World Cup set to be most lucrative ever. Available: http://www.espnfc.com/fifa-world-cup/story/1830732/2014-world-cup-set-to-be-most-lucrative-ever

[53] The Global Web Index (2015), Data from the GlobalWebIndex using Q1 2014 numbers. Available at: http://wearesocial.net/blog/2015/01/digital-social-mobile-worldwide-2015/

[54] Thomas, J.J. and Cook, K.A. (eds.). 2005. Illuminating the Path: the Research and Development Agenda for Visual Analytics. Los Alamitos: IEEE.