BIG DATA
CASE Study

WORLD BANK GROUP

Where citizens lack means of voicing discontent, they can embrace forms of protest that can quickly escalate

Observing People's Feelings About State Institutions

Victoria L. Lemieux

SUMMARY

Where citizens lack means of voicing discontent with institutions and governments, they can embrace extreme forms of protest that quickly escalate. To explore the relationship between citizens' feelings about governance institutions, their trust in government, and civil unrest, this project conducted sentiment analysis of tweets made during protests surrounding the 2014 Soccer World Cup in Brazil. The team harvested 11 million real-time tweets and used visual analytics techniques to extract search terms for the collection of relevant historic tweets. They used a visual analytics tool to carry out sentiment and text analysis, then followed a structured approach to explore how Brazilian citizens felt about their state institutions, how these feelings connected to their sentiments about government and politicians, and how such sentiments translated into collective behaviors.

The analysis showed citizens expressing negative sentiment about the national government's low investment in services such as education, health and water, relative to lavish spending on the World Cup. The negative tweets were forms of social

protest that may have led to other forms of protest, including demonstrations. These findings supported the relative deprivation theory of what causes social protest (the theory that people feel grievance leading to protest when they feel deprived relative to some external standard). They suggest domestic policy priorities for Brazil's government and resonate with wider policy discussions on wealth inequality.

CHALLENGE

Civil protest has been described as moving from discontent of the populace, to politicization of that discontent, to actualization of frustration as aggression against the state. Such unrest can lead to fragile states, corruption, terrorism and economic impediments. It can be causally linked to poverty, but dissatisfaction with institutions can also be relevant. In countries where citizens lack means of voicing discontent with institutions and governments, they can embrace extreme forms of protest that quickly escalate – as demonstrated by the so-called 'Arab Spring'.

Typical methods for studying internal conflict include qualitative case studies and econometric analyses. However, case studies can be dangerous and costly, and econometric approaches can overlook localized conditions and the evolving dynamics of protest. There is a need for granular information about local conditions and the causal chain of events, without the dangers and costs of case study approaches. Analysis of microblog data has the potential to deliver this information. This project sought to discover whether such analysis could uncover how citizens feel about their institutions and government, and how these sentiments translate into collective behavior.

Evidence from online media data offers an innovative approach to investigating social and political issues

INNOVATION

The team chose to analyze Twitter postings in Brazil during the 2014 Soccer World Cup. The country has relatively high levels of inequality and high social media use. It is the world's second-biggest Twitter user (with roughly 41.2 million tweeters). Costing up to an estimated US$14 billion, the World Cup was the country's then-largest and most expensive sporting event. During and afterwards, it sparked public protests in several Brazilian cities.

To explore whether the protests related to citizens' trust in state institutions, the project produced an innovative visual analytics tool and a novel analytic methodology. Visual analytics combines human reasoning with machine reasoning through an interactive visual interface. This 'mixed initiative' approach overcomes the limitations of each type of reasoning. Human analysts can detect subtleties of humor or satire that a computer might miss, and computers can rapidly process and manipulate volumes of data that humans cannot.

How do you feel?

Through sentiment analysis and text analysis (machine reasoning) of a large-scale collection of tweets, the project tracked general public distress and trust in institutions, hypothesizing that increased negative sentiment signaled declining public trust in government. The team collected publicly available Twitter data, in two phases. The first was an initial 'big picture' harvest of approximately 11 million tweets. The sample was then analyzed using Natural Language Processing and a visual analysis tool that clusters documents

together by determining key themes in each. Collections of texts are displayed like a galaxy of stars, in which each star is a single document, and clusters represent document similarity. Through this, the team identified tweets on political opinion and analyzed them to define 'naturalistic' search terms representing key concepts (such as government and institutions) underlying the study. These search terms were then used to harvest historical tweets for the 2014 World Cup period.

The team performed sentiment classification of the harvested tweets using SentiStrength, a tool which classifies text in terms of positive and negative sentiment. It can also compute a single evaluation based on both positive and negative classifications. The project used this single-sentiment approach, resulting in nine sentiment categories: Positive or negative with a magnitude of 1 to 4, and neutral. The results of the sentiment classification were visually presented as horizon charts, allowing analysts to see how sentiment varies in polarity and intensity over time. From a basic text analysis of historical tweets, they depicted the top 500 word-frequency terms in a text-cloud, as well as the single use of terms over time.

Translating sentiment into behavior

Results of the sentiment classification and text analyses were then represented visually, so analysts could detect patterns in the data (human reasoning). They used a structured analytic methodology, aided by the tool's interactive features, such as the ability to search, sort and see data from different views. The sentiment analysis enabled the team to observe patterns in the data such as high negative or positive sentiment toward a particular institution, increasing or decreasing positive or negative sentiment over a period, or correlation of negative sentiment about an institution with

rising negative sentiment about government as a whole. They explored correlations between patterns of sentiment, government policies and observed citizen behavior (such as protests).

Using the visual analytics tool, the team carried out a pair analysis to make basic observations from the data. This analysis pairs a subject-matter expert and a visual analytics expert, combining contextual knowledge with technical expertise. The tool was also used for an analysis of competing hypotheses, a methodology which analyzes the degree to which evidence supports the relative likelihood of alternative hypotheses. From background literature, the team identified 68 hypotheses on the relationship between citizen trust and social protest. The analysis drew tentative conclusions about the relative likelihood of each by trying to disprove rather than prove each hypothesis.

RESULTS

By connecting the visual analysis to hypotheses on the relationship between citizen trust and social protest, the study found support for the relative deprivation theory of social protest. This suggests that an individual or group lacks something that another group has and to which they feel entitled. Deprivation is felt in relation to some external standard, not in absolute terms.

The methodology showed that around the 2014 World Cup, Brazilians expressed negative sentiment about low investment in services such as education, health and water, relative to spending on the tournament. At state level, water was a key issue, with tweeters criticizing investment relative to politicians' spending on priorities such as campaign financing. The negative tweets themselves constituted social protest, as well as forming part of larger politicized groups using protest hashtags and calling for other forms of protest (such as demonstrations).

While media reports and subsequent studies of the 2014 World Cup protests generally focused on single immediate causes, the analysis showed that the protests sprang from a range of long-standing grievances, coupled with relative deprivation triggered by spending on the World Cup and campaign financing. This sense of deprivation fueled sentiments that activated protest.

The project offers an innovative approach to investigating social and political issues. The methodology has already been used in a World Bank evaluation of higher education in Brazil, enabling analysts to link how citizens feel about higher education to existing hypotheses about education in developing countries. Evidence from online media data suggests further avenues for research using complementary methodologies, such as surveys or comparative analysis with other data sources. The approach could also support the development of an observatory of citizen sentiment to inform investment decisions around strengthening institutions. It could ultimately lead to predictive models for negative sentiment towards particular institutions.

LESSONS LEARNED

The project demonstrated the potential of big data analytics to contribute to knowledge about development, as well as for evaluating development outcomes.

- **Combine human and computational power for optimum results**
  Visual analytics are useful for unpicking complex socio-political issues, such as those surrounding citizen trust. Alongside computational methods, analysis of online social data requires human input, supported by interactive visual interfaces.

- **Build strong partnerships**
  Development of this approach required close collaboration with experts in fields such as Brazilian society, history and government. Using 'design thinking' techniques, the project team partnered with experts to test assumptions and obtain feedback on approaches and results. This improved the project outcome

- **Use big data analytics alongside other research approaches**
  Big data analytics makes observations possible from a distance, both in terms of space and time, and the project provided insight into Brazilians' thoughts, as expressed naturalistically. In contrast, surveys require researchers to spend time on the ground, and may prime citizens with questions that do not reveal their own thoughts. However, this type of big data analytics does not offer a representative sample, and is best used to complement other approaches to understanding development issues.

- **Protect individuals' privacy**
  Twitter users do not consider use of their data for research when they tweet. Many tweets sent impulsively – especially during social protest – contain candid or critical remarks. Take care to protect citizens from potential harm, such as retaliation for expression of opinion or lawsuits for defamation.

INNOVATIONS IN
BIG DATA
ANALYTICS

#BIGDATAINNOVATE

WORLD BANK GROUP