Skip all front matter: Jump to Page 1

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

Using Social Media to Gauge Iranian Public Opinion and Mood After the 2009 Election

Sara Beth Elson • Douglas Yeung • Parisa Roshan • S. R. Bohandy • Alireza Nader

RAND NATIONAL SECURITY RESEARCH DIVISION

This publication results from the RAND Corporation's Investment in People and Ideas program. Support for this program is provided, in part, by the generosity of RAND's donors and by the fees earned on client-funded research.

Library of Congress Cataloging-in-Publication Data

Using social media to gauge Iranian public opinion and mood after the 2009 election / Sara Beth Elson ... [et al.].
p. cm.
Includes bibliographical references.
ISBN 978-0-8330-5972-7 (pbk. : alk. paper)
1. Presidents—Iran—Election—2009—Public opinion. 2. Public opinion—Iran. 3. Social media—Political aspects. 4. Social media—Political aspects—Research. I. Elson, Sara Beth.

JQ1789.A5U75 2012
324.955'061—dc23

2011049687

Cover image courtesy of Jose Luis Pelaez Inc. (Blend Images/Getty Images)

Published 2012 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Summary

In countries where freedom of expression is limited, social media, such as Twitter, blogs, and Facebook, are playing an increasingly visible role in politics. For example, in the absence of democratic elections, an estimated 70 million bloggers in China have become the *de facto* voice of the people (Friedman, 2010). In another example, an increasing number of Pakistanis turned to YouTube, Flickr, Facebook, and Short Message Service (SMS) text messages as alternative media portals during the 2007–2008 "Pakistan Emergency," which began after Pakistan's president, General Pervez Musharraf, suspended the Chief Justice of the Supreme Court, and the government jammed cell phone networks and blocked news channels (Yusuf, 2009). And finally, in the months after the contested Iranian presidential election in June 2009, Iranians blogged, posted to Facebook, and, most visibly, coordinated large-scale protests on Twitter. Twitter users sent tweets—short text messages posted using Twitter—marked with the "IranElection" hashtag (i.e., labeled as being about the Iran election) at a rate of about 30 new tweets per minute in the days immediately following the election.

The use of social media in closed societies such as Iran's—where people's freedom of expression is curtailed—presents a novel opportunity for analysts to gain insight into public opinion and mood where previously they faced significant barriers. Analysts studying public opinion in open societies have multiple tools at their disposal: face-to-face surveys, on-site observation, media analysis, and telephone poll results. But the inherent nature of a closed society makes it difficult, and sometimes impossible, for analysts to apply these methods, which require people to cooperate openly and media to be uncensored. In contrast, social media, because they can be used in anonymity, offer an alternative to which numerous people in closed societies are turning in lieu of making public or traceable statements. As such, they offer analysts an alternative source of data on public opinion and mood in places where people may fear retribution for expressing themselves freely.

Computer-Based Analysis: A Promising Counterpart to the More Established Approach to Analyzing Social Media

At present, analysts typically use a "manual" approach to study the content of social media. For example, they might read a few select blogs or follow a specific set of people on Twitter. Then, they interpret the postings they read and, typically, present their analysis in a report. The manual approach can provide analysts with an in-depth understanding of posted content, allowing them to interpret the nuance, subtlety, and implicit meanings inherent in what people write. However, it also has significant drawbacks. Analysts may read the wrong material. They are constrained by the amount of text they can read in a given workday. It will be impossible for them to gain an understanding of public opinion or mood on a mass scale. And analysts' biases may affect their interpretations of what they read.

Given the shortcomings of the manual approach, using a computerized method to study the content of social media can serve as a useful complement, compensating for some of these limitations. Such a tool exists: an automated content analysis program called "Linguistic Inquiry and Word Count 2007" (LIWC, pronounced "Luke") (Pennebaker, Booth, and Francis, 2007; Pennebaker et al., 2007). Able to analyze thousands of social media posts in only a few seconds, LIWC offers a number of advantages. For instance, analysts can gain a bird's-eye view of what people are saying and feeling across many social media sites over time. Analysts can also quantify their data, in this way reducing the chance that their biases affect their interpretations of social media texts. Ultimately, combining an automated approach with the more traditional, manual approach is most likely to yield optimal results.

However, LIWC is largely untested in political contexts. In the past, researchers have used LIWC in a variety of ways—for example, to study language patterns after traumatic events (Gortner and Pennebaker, 2003; Stone and Pennebaker, 2002), to investigate how men and women communicate differently (Newman et al., 2008), and to detect deception (Hancock et al., 2008; Newman et al., 2003). But LIWC has not been widely applied to understanding a non-Western political context.

Recognizing the potential of this computer-based tool to shed light on public opinion in closed societies, we tested LIWC by applying it to the case of Iran. Focusing on Twitter, we used LIWC as a means of tapping into Iranian public opinion and mood during the tumultuous months following the highly controversial 2009 presidential election. The post-election period may be viewed as extending from the election itself in 2009 until major protests ended nine months later, in February 2010. With this period as our focal point, we examined

- how Twitter users tweeting about the election felt, in general, over these nine months
- whether their sentiments tracked with the outbreak of protests
- how they felt about political leaders, including Supreme Leader Ali Khamenei, President Mahmoud Ahmadinejad, opposition presidential contenders Mir Hussein Mousavi and Mehdi Karroubi, and U.S. President Barack Obama
- how they wrote about certain groups affiliated with either the Iranian government or the opposition, including the Islamic Revolutionary Guards Corps (IRGC), the Basij paramilitary force, and the Green Movement
- how they wrote about certain countries, including the United States, Israel, and Iran.

Given that LIWC is largely untried in non-Western political contexts, we used Iran during this period as a test case. On the one hand, we sought to shed light on how public opinion and mood evolved after the 2009 election. But at the same time, we intended to examine the validity of a new methodology—one incorporating the LIWC tool—for analyzing foreign public sentiment on political topics, as expressed through the social media platform, Twitter.

How Does LIWC Work?

LIWC was developed to analyze the characteristics and patterns of written text, allowing one to draw conclusions about people's psychological states (e.g., emotions, desire for social interaction) on the basis of their usage of specific categories of words. LIWC contains approximately 80 such categories: first-person singular pronouns, positive emotion words, and swear words are three examples. For a given text, LIWC first counts the total number of words in that text. Then it searches for all words contained within each of the 80 categories, keeping tally of the number of instances in each category. Each time LIWC encounters a word in a given category, it increments that category by one. Finally, LIWC calculates a ratio of the number of words within each category divided by the total number of words in the text.

Such LIWC-generated ratios (the quantitative data) can then be interpreted qualitatively, by using established research linking the use of specific words to psychological states and behaviors. On this basis, by examining, for example, how large or small the ratios are or how they change over time, one can gain insight into how widely felt certain implicit sentiments are among a group of people or examine how their opinions and sentiments have shifted.

Ample Precedent for Using Word-Usage Analysis to Assess Public Opinion

A strong precedent exists for using LIWC in this way, in the form of previous research that has connected people's use of certain word categories with their emotions, attitudes, and behaviors. For example, greater use of first-person singular pronouns—talking and thinking about oneself—has been shown to suggest feelings of depression (Rude, Gortner, and Pennebaker, 2004) and poor coping with traumatic events. Second-person or plural pronouns indicate reaching out to others (Chung and Pennebaker, 2007) and a sense of community or group identity. (These words are typically used more frequently around large-scale, shared traumas [Cohn, Mehl, and Pennebaker, 2004; Gortner and Pennebaker, 2003; Stone and Pennebaker, 2002].) Validations of LIWC suggest that it accurately characterizes emotions in written language (Kahn et al., 2007) and that the results it generates are comparable to those produced with other content-analysis methods (Alpers et al., 2005). In addition, LIWC has been successfully applied more recently to various forms of social media (e.g., blogs) (Pennebaker et al., 2007).

Our Approach to the Analysis

Our computer-based methodology centered on examining the rates at which people posting on Twitter about the Iran election used certain categories of words. To interpret the attitude or emotion conveyed by these categories, we used established precedent from the LIWC literature. We then (1) tracked the rates of use over the nine-month period following the 2009 Iranian presidential election and (2) examined whether patterns in these rates coincided with political events, to gain insight into how people may have felt before, during, and after those events.

Collecting and Preparing the Sample

We began by selecting a set of relevant Iranian political topics to analyze using LIWC. Next, we constructed an automated software program to parse and clean the social media texts in our sample, and we stored the texts in a database. We analyzed 2,675,670 tweets marked with the "IranElection" hashtag,[1] posted by 124,563 distinct individuals and dated from June 17, 2009, to February 28, 2010. As this hashtag was 2009's second most popular topic across all of Twitter (Twitter Blog, 2009b), this set of data should provide broad coverage of the election discussion among Twitter users. This sample necessarily included observers throughout the world, as well as people in Iran, who were all communicating with each other. In addition, all of the tweets in this dataset were written in English. But a review of many of them (shown below as they appeared) suggested that their authors were Iranians living inside Iran:

> Just went to Vanak square site of previous protests. All calm so far
>
> oh my dear god please help us. no one helping us but we still fighting the Basij.
>
> We been beaten tortured and killed for 30yrs! Nothing the SUPREME LIAR says can break our will![2]

Processing and Interpreting the Data

We then processed the data with LIWC and conducted qualitative interpretations of the quantitative LIWC output. Table S.1 lists the word categories we studied to gain insight into the Iran election and describes how we interpreted the attitude or mood expressed.

Table S.1
Word Categories Indicating Public Opinion and Mood and How We Interpreted Them

Word Categories	Attitude or Mood Expressed
First-person singular pronouns	Feelings of depression within the population[a]
Second-person pronouns	An intent and desire to interact with others[b]
Plural pronouns	A sense of (1) group or collective identity and (2) coping with shared trauma[c]
Positive emotions	Feeling generally good or happy (based on LIWC's accuracy in capturing emotions)[d]
Negative emotions	The degree to which people have been affected by a given trauma; also feeling angry, anxious, or sad[e]
Swear/curse words	Frustration or anger

[a] Pennebaker and Chung (2005); Gortner and Pennebaker (2003).
[b] Slatcher, Vazire, and Pennebaker (2008).
[c] Cohn, Mehl, and Pennebaker (2004); Pennebaker, Mehl, and Niederhoffer (2003); Chung and Pennebaker, (2007).
[d] E.g., Alpers et al. (2005); Kahn et al. (2007).
[e] Cohn, Mehl, and Pennebaker (2004).

---

[1] Hashtags are words or phrases preceded with the # symbol. Twitter users use hashtags to label their tweets, so tweets with the "IranElection" hashtag are labeled as being about the election.

[2] The reason for tweeting in English may have been a strong sense of needing to communicate across national boundaries what was happening in the country.

Placing the LIWC Output into a Politico-Historical Context

In the final phase of our work, we looked for patterns in the data and interpreted those patterns in relation to specific events on the ground and to public figures and topics important during the post-election period. For example, we sought spikes and dips in word usage—i.e., sudden reactions to a specific event or action (e.g., a protest or holiday) that might offer insight into public opinion about that event.[3] We also looked for patterns in comparisons of certain word categories against others.

A Demonstration of LIWC Using Twitter: Insights into Public Mood and Opinion in Iran in the Aftermath of the 2009 Presidential Election

Our work with LIWC enabled us to draw a number of conclusions about the opinion and mood of Twitter users on a wide range of topics during the post-election period. We present a selection here.[4]

As of the end of February 2010, the opposition movement did not appear likely to protest against the government in the foreseeable future. We examined whether increases or decreases in the levels of anger people expressed about the election coincided with—or even forecasted—the outbreak of protests. One word category conveying strong emotion stood out the most clearly: swearing. With each large-scale protest,[5] rates of swearing spiked on Twitter, or rose in the weeks leading up to the event. This suggests that swearing levels could predict the outbreak of protests (see Figure S.1). For example, levels of swearing were elevated in the weeks before the Quds Day protest (September 18, 2009), one of the major protests of the post-election period, and there was a large spike at the end of December, when a protest on Ashura Day occurred. The overall trend in swearing—a gradual decline over time to a relatively stable level lower than the initial level observed in June and July 2009—suggests that the opposition movement had probably resigned itself to the political situation nine months after the election.

President Ahmadinejad initially spurred more anger than Mir Hussein Mousavi, but the opposite had become true by the end of the post-election period. Trends in swearing show that in the initial weeks after the election, people used more profanity when tweeting about Ahmadinejad than about leading opposition candidate Mousavi. However, this pattern had reversed itself by the end of the post-election period, suggesting that when the opposition movement flagged in February 2010, Twitterers felt angrier at Mousavi than Ahmadinejad.

After some early frustration, Twitter users generally viewed President Obama without ill feeling in the aftermath of the election. Our analysis indicates more negative sentiment toward Supreme Leader Khamenei and President Ahmadinejad than toward President Obama throughout the entire post-election period. Indeed, Twitterers displayed little anger toward Obama. One of the few exceptions was early after the election, when they expressed a strong desire for him to take a public stand in support of the Iranian protesters. Linguistic patterns in the tweets also indicated that Twitterers felt a greater sense of community and shared ties

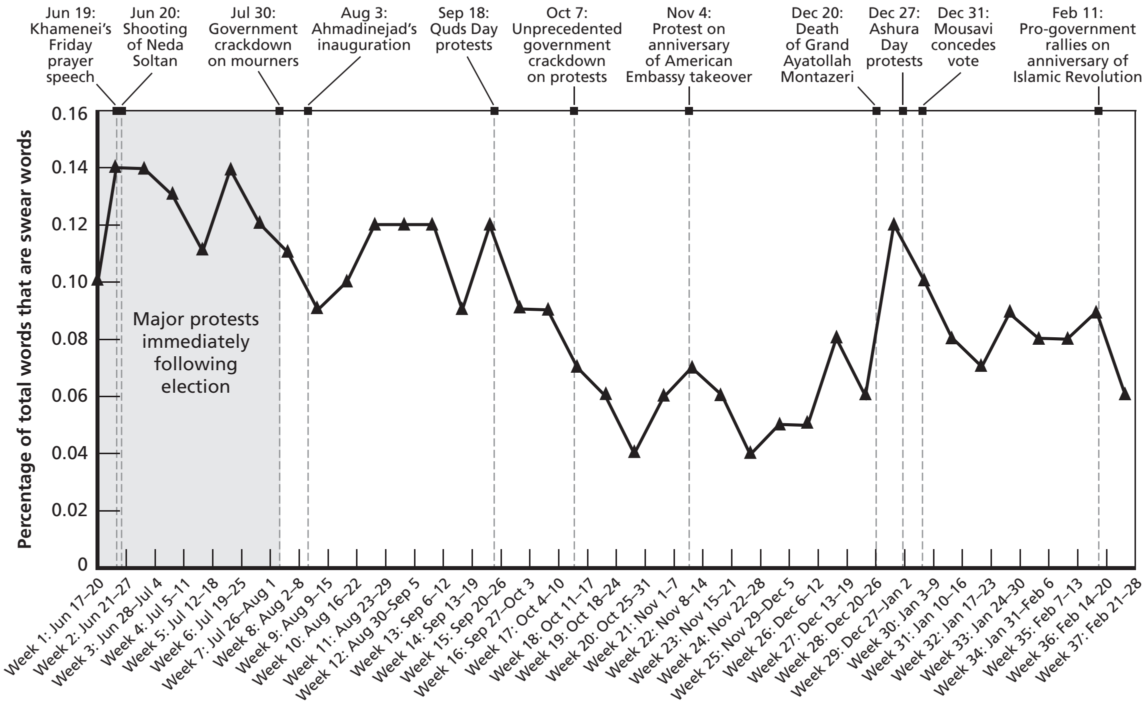
---

3 Such correlations do not necessarily indicate causal relationships.

4 An interactive web tool presenting Twitter data from this study is available on RAND's website (RAND National Security Research Division, 2011). Users may select and view combinations of word categories and political topics. The trends displayed show how public opinion and mood changed across time about the topic selected.

5 By *large-scale*, we mean protests involving hundreds of thousands or millions of people.

Figure S.1
Rates of Swearing on Twitter After the 2009 Presidential Election



NOTES: The data points signify a ratio of the number of swear words to total words across an entire week's worth of tweets. Significant events in the aftermath of the election are noted at the top of the figure.
RAND *TR1161-S.1*

with each other when discussing Obama than when they discussed Ahmadinejad, Khamenei, Mousavi, or Karroubi.

The Green Movement was viewed more positively than the Islamic Revolutionary Guards Corps and the Basij. The Green Movement is the broad-based opposition movement that developed in the weeks following the presidential election. The Revolutionary Guards are an elite military force and currently Iran's most powerful economic, social, and political institution. The Basij comprises a set of pro-government paramilitary organizations serving under the leadership of the Revolutionary Guards.

In the initial weeks after the election, Twitterers expressed high levels of positive sentiment toward the Green Movement. However, this sentiment dropped sharply after the initial post-election period and remained low until the end of February 2010. With regard to both the Revolutionary Guards and the Basij, Twitterers expressed consistently low levels of positive sentiment throughout the entire period. Between the two, Twitterers expressed more anger at the Basij forces than at the Revolutionary Guards, on the whole.

Countries with traditionally tense relations with Iran were not the target of anger, unlike the "Islamic Republic." Our results suggest that Twitterers did not focus their anger on Iran's traditional enemies—Israel and the United States—in the aftermath of the election. We saw similarly low levels of negative sentiment toward "Iran." But the opposite was true of the "Islamic Republic." This indicates that Twitter users in our sample were angry not at the country of Iran itself but, rather, at the government that rules Iran.

A Methodological Note for Potential Users of the LIWC Software Tool

Our exploratory research raised important considerations for those wishing to use this method. One was especially important. In the current project, the automated analysis provided initial results—suggesting, for example, that tweets containing swear words expressed anger at a certain target. In most cases, the tweets truly did express the emotions signified by the LIWC words they contained. However, in certain cases, tweets that contained words thought to denote sadness, positive emotion, and anxiety did not express the emotion we would have expected, given the words they contained. In that sense, word categories thought to denote a particular set of emotions did not always perform as expected. To manage this problem, analysts could potentially construct alternate, tailored word categories that contain words that seem most useful in context and then validate those categories. Performing this automated analysis therefore alerts the analyst to check specific portions of the data for such anomalies, regardless of whether noteworthy patterns are found.

The Policy Uses of Computer-Based Analysis of Social Media

Our test case of Iran suggests that using the LIWC-based method as a means of analyzing social media holds much promise, particularly in countries where freedom of expression is limited. The potential policy uses are multiple. With this approach, analysts can

- **Make informed assessments of public opinion in countries of interest, in retrospect as well as in real time.** An example is the multiple trends observed among Twitter users that suggested that the opposition movement in Iran had flagged by the end of the nine-month post-election period and would not protest further.
- **Forecast events, such as large-scale protests.** The increase in Twitterers' use of swear words just before large protests formed a suggestive pattern that spikes in swearing could signal an imminent action.
- **Assess the impact of political events or actions on public opinion.** The fact that Twitterers swore at such high levels about Ahmadinejad in the immediate aftermath of the election substantiates the general impression that the election negatively affected public opinion of the Iranian president on a large scale.
- **Inform outreach efforts to foreign populations.** The cursing we saw of President Obama in the earliest weeks after the election, indicating that Twitterers wanted him to take a stand in favor of the opposition, is the type of information that could help the U.S. government, in particular, better understand and act on public opinion in foreign countries.
- **Pinpoint intelligence gaps.** For example, the reversal in public sentiment about President Ahmadinejad and leading opposition candidate Mousavi after the Quds protest raises the question of what drove that change at that time.

Ideas for a Second Phase of This Comprehensive Program of Research

We envision this exploratory work with LIWC as the first phase of a larger program of research. Having now established that LIWC can generate informative output about public mood and opinion in a closed society, we can confidently incorporate it into a multimethod research design to answer timely policy questions, both to validate the method and to explore additional research questions. The upcoming presidential elections in Iran would present one such opportunity. The tightly controlled political climate makes it difficult to obtain direct data on how people inside Iran feel about the candidates and relevant political issues. As such, one could turn to social media to gain needed insights, using LIWC alongside manual analyses of social media posts, interviews of people traveling to and from Iran, and content analyses of articles in the Iranian news media. Similar studies of questions about the Iranian economy or Iran's nuclear weapons program would also be possibilities. Ultimately, we view LIWC and other automated content analysis as an important part of research designs for studies of countries in conflict generally (such as Pakistan or Egypt, as well as Iran)—both to examine them on their own terms and to make comparisons between them.