A Mathematical Approach to Gauging Influence by Identifying Shifts in the Emotions of Social Media Users

Les Servi, *Senior Member, IEEE* and Sara Beth Elson

(Invited Paper)

***Abstract*—Although an extensive research literature on influence exists in fields like social psychology and communications, the advent of social media opens up new questions regarding how to define and measure influence online. In this paper, we present a new definition of influence that is tailored uniquely for online contexts and an associated methodology for gauging influence. According to our definition, influence entails the capacity to shift the patterns of emotion levels expressed by social media users. The source of influence may be the content of a user's message or the context of the relationship between exchanging users. Regardless of the source, measuring influence requires first identifying shifts in the patterns of emotion levels expressed by users and then studying the extent that these shifts can be associated with a user. This paper presents a new quantitative approach that combines the use of a text analysis program with a mathematical algorithm to derive trends in levels of emotions expressed in social media and, more importantly, detect breakpoints when those trends changed abruptly. First steps have also been taken to predict future trends in expressions (as well as quantify their accuracy). These methods constitute a new approach to quantifying influence in social media that focuses on detecting the impact of influence (e.g., shifts in levels of emotions expressed) as opposed to focusing on the dynamics of simple social media counts, e.g., retweets, followers, or likes.**

***Index Terms*—Breakpoint, emotion, influence, LIWC, social media, text analysis, Twitter.**

I. Introduction

SOCIAL MEDIA offers an important window into the emotions and influence of those who use the platform. In particular, media such as Facebook and Twitter can reveal what people are thinking and feeling. The rapidly growing number of social media users around the world creates a wealth of data that complements traditional public-opinion polls and other types of surveys. For example, according to one estimate, Twitter users generate 400 million tweets every day [58]. As such, much work has been done in recent years to develop ways to analyze this rich new dataset. Our area of focus has been on using social media to detect significant changes in emotional expressions, as revealed by the emotional dynamics of users of the media. A closely related goal has been to forecast future trends in these emotions. Being able to identify and forecast these trends with a high degree of accuracy is an essential first step toward characterizing the extent that individuals exert emotional influence on social media users.

Across the past several decades, research from a variety of fields, including social psychology, communications, political science, sociology, and marketing, has examined the issue of influence—both in offline and online contexts [4], [12], [28]. Studies on influence within the field of social psychology, in particular, have focused on quantification and capturing essential components like human choices so that they can be studied in a controlled laboratory setting.[1] For example, persuasion researchers have proposed dual-process models of persuasion, specifying that there are two primary ways in which people process information—either "centrally" (also called "systematic" processing) or "peripherally" (also called "heuristic" processing) [8], [49]. When engaging in central route persuasion, people focus on the content of the message and formulate their attitude on the topic based on factors such as the quality of the arguments. People are more likely to engage in central route persuasion if the topic is more important to them, if they have the cognitive resources available to process the message, if they know something about the topic, and if the arguments are written [8], [49].

On the other hand, people engaging in peripheral route persuasion are more likely to use decision cues or rules of thumb (also known as "heuristics") to formulate their attitude on a topic. For example, someone engaging in peripheral route persuasion may be more persuaded by the number of persuasive arguments rather than the quality. Such an individual will also more likely make a decision based on the perceived credibility of the influence agent rather than on the truthfulness of his/her statements. People are more likely to engage in peripheral-route processing when the topic is not important to them, they do not have the cognitive resources to process the message carefully, the mode of communication is one in which the influence agent is salient, and they know very little about the topic [8], [49].

Manuscript received March 25, 2014; revised December 15, 2014; accepted December 16, 2014. Date of publication January 23, 2015; date of current version January 26, 2015.

The authors are with The MITRE Corporation, Bedford, MA 01773 USA (e-mail: lservi@mitre.org; selson@mitre.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSS.2014.2384216

[1] We note the controversy surrounding research in controlled laboratory settings. Many argue that because such research is conducted in laboratory settings, there is little generalization to the real world. This is a valid concern that can be potentially mitigated with the advent of survey technologies enabling researchers to embed classic experimental designs into surveys [43]. Such technologies combine the strengths of controlled experimental designs with the generalizability to broader populations that surveys afford. In the future, therefore, it will be possible to test the generalizability of classic persuasion theories outside of laboratory settings.

It should be noted that individuals will not necessarily process every message using only central route or peripheral route processing. Rather, individuals may engage in either type, depending on the circumstances.

In parallel to the social psychological research described above, other studies have examined influence through social media. These studies have focused on items including directed links, indegree, retweets, and mentions [7]. A highly cited alternative approach to measuring influence looks at the dynamics of the spreading of a message through the network and seeks to identify a small fraction of people who, if exposed to a message are most likely to result in the message becoming known by a large fraction of the network [17], [35]. Other approaches which quantify diffusion through a network include [14], [27], [54], [59], and [60].

Our approach sits among these others and constitutes a new way of conceptualizing influence and measuring it quantitatively.

According to our definition, influence entails the capacity to shift the patterns of emotion levels expressed by social media users. Hence, it is the intensity of the emotional level impacted rather than the number of people impacted or the topology of the social network which is of interest to us.

We began our analysis with a large existing dataset of posts to Twitter. As we developed our new method, we used this existing dataset as the basis for a test case.[2] What made this dataset particularly useful was that the dynamics of Twitter users' emotions during the period examined were extraordinary, and the dataset was readily available. Our test case both demonstrates the method and provides concrete examples of the type of outputs it produces.

Our approach entails beginning with an automated text analysis program to analyze social media text. We used a program called "Linguistic Inquiry and Word Count 2007" (LIWC—discussed further in Section II) [46], [47]. Given the many effective content analysis systems that have been developed, we chose LIWC as a starting point because of the decades of research conducted to evaluate and apply it (for a review of this research, see [55]). However, we view LIWC (and the two specific LIWC categories examined in this study) as a starting point and emphasize that our technique can be applied using any content analysis program and corresponding set of emotion categories. In recent years, LIWC has been used successfully to probe the content of social media [29], [53], [57]. Starting with the results that LIWC can generate from sizeable datasets of social media posts, our approach employs a mathematical algorithm to detect changes in emotions of social media users. Points in time at which these emotions shift substantially are referred to as "breakpoints."[3] Our approach computes and mathematically updates estimates of when these breakpoints occur, as well as the trends in emotions in between breakpoints. It can do this either in real time or historically. The selection of breakpoints is unbiased by any person's subjective views of where the trends are going.

Our technique has also been designed to assist in forecasting the future direction of emotions. Forecasting, especially using social media, is challenging and is still far from an exact science, but we have made a start at applying a mathematical approach to forecasting LIWC indicators in a manner that brings desired rigor. In due course, our method is designed to apply these types of forecasts to changes in consumer behavior. A broad literature on forecasting shows that if one takes an average prediction of a group of forecasters that average will usually be more accurate than the majority of forecasters from whom the average was computed [3]. However, the best human forecasters could not outperform crude extrapolation algorithms [56]. Hence, we built on the idea of generating a collective forecast by developing a quantitative forecasting technique, applying it to a collective body of tweets, and measuring its accuracy. Although the Twitter users we examined were not trying to make forecasts *per se*, our technique entails aggregating levels of emotions they express on a topic, computing overall trends in those emotions, breakpoints in the trends, and then, based on those breakpoints, generating forecasts of where those emotions are headed.

Identifying emotional levels, detecting influence, and forecasting emotions are part of a larger process whose goal is to explain the meaning, in behavioral terms, of the raw trends and shifts that one is observing. Without first accurately identifying when such changes occur, reliable interpretations are not possible. But the ultimate objective is well-informed explanations and, possibly, recommendations stemming from those explanations. We have already made a foray into this work to date by developing a way to draw such behavioral implications from the output of our algorithm. That work is ongoing.

Our method provides a mathematically unbiased approach to detecting breakpoints and monitoring trends in emotions. The ingredients more classically applied to Twitter for measuring influence, such as the Klout measure of influence [2] (retweets, mentions, list memberships, followers, or replies) all are clearly indicative of influence. However, these ingredients do not directly measure the emotional impact the influence generates. This paper takes the point of view that it is useful to view (and quantify) influence according to the extent that users move the breakpoints or changed the trends from their forecasted values.

Each section of the paper can be read independently of other sections to suit the interests of various readers. Section II provides a background to LIWC and its application to Twitter data. Section III-A describes a new mathematical approach to best characterize abrupt shifts in the emotions of Twitter users—signaling influence in certain cases. We illustrate this approach first with a simple example and then with a Twitter data set. Section III-B then repeats the analysis of Section III-A in a mathematically precise manner. Expanding on the mathematical approach presented in Section III, Section IV-A presents a set of forecasting algorithms using the output of Section III and initial validation of those algorithms. Section IV-B then presents methodological considerations to address when using

[2]The data set presented in this paper is real and represents over 2.5 million Twitter messages from over 100 000 individual users However, we were requested not to identify any further details associated with the hashtag and search terms used to generate it.

[3]The working premise here is that, in general, emotional expression evolves quite smoothly, but on occasion, it changes very suddenly and dramatically. When this happens, it is as if a "reset button" were being pushed. In our work, we treat points in time before the "reset' as irrelevant to points in time after it.

the techniques presented in this paper. Finally, Section V presents future directions for this research.

II. Generating the Data Needed to Apply the Method: The LIWC Program

While we have used the computerized text analysis program, LIWC, there are alternative approaches which can be used to quantize the emotional content of words. For example, within the field of computer science, Kagan *et al.* [33] present the SentiMetrix SentiGrade scoring engine and discuss how they tuned it for the detection of PTSD-related signals using natural language processing techniques. Another example consists of the affective norms for English words (ANEWs) [6], [18], which associates a numerical value to valence and intensity of emotion words but has a more restrictive class of emotion categories than LIWC. AFINN was developed specifically with Twitter as the target source of words [22]. SentiWordNet [19], [20] is focused on sentiment scores (positive, negative, or neutral). An extension to (Profile of Mood States (POMS) [40], [41]), referred to as POMS-ex, has been developed and validated [48] and applied to the analysis of six individual dimensions of mood in Twitter messages [5]. In addition, the Sentimedir program [15], [13] identifies the sentiment polarity, the opinion holder (to whom the sentiment is attributed), and the target of the sentiment, and has been validated against manually annotated data (i.e., news, blogs, and tweets), but it also has a more restrictive class of emotion categories than LIWC.

Research on the development and evaluation of LIWC dates back to 1993 [24], [45]. Initially, this research entailed having successive panels of judges evaluate words in isolation for inclusion into categories. Over the ensuing decades, however, hundreds of studies have linked the language categories in LIWC to psychological processes operating in broader contexts [55]. Specifically, researchers have used LIWC to detect attentional focus [52], emotionality [25], cognitive styles [42], individual differences [44], and social relationships [31]. In addition, the use of certain function words (such as "the," "with," or "they")[4] has been linked with personality and social processes, psychological states including depression, biological activity, reactions to individual life stressors, reactions to socially shared stressors, deception, status, gender, age, and culture [10]. For example, people who are feeling physical or emotional pain tend to focus their attention on themselves and therefore use more first-person singular pronouns in their speech or writing while they are in that state [47].

As with function words, the use of emotion words (such as "happy," "terrified," or "jealous"), how people express emotion, and the valence of that emotion can tell us how people are experiencing the world [55]. Previous research using LIWC suggests that it identifies emotion accurately in language use. For example, people use positive emotion words (e.g., "love," "nice," or "sweet") when writing about a positive event and negative emotion words (e.g., "hurt," "ugly," or "nasty") when writing about a negative event [34]. In addition, LIWC ratings of positive and negative emotions words in written texts correspond with human ratings of the emotional content of those same written texts [1]. As such, the function and emotion words people use provide psychological markers of their thought processes, emotional states, intentions, and motivations [55].[5]

In the past, researchers have applied LIWC to a variety of different text genres, including college writing samples, science articles, blogs, novels, talking, and newspapers [25], [47]. Although previous research has laid a foundation for the use of Twitter and of LIWC to forecast the direction of Twitter users' emotions and events on the ground, any such forecasting would rely on manual inspection of trends in LIWC indicators. For this reason, we developed a set of mathematical algorithms to determine, objectively, when trends in emotion have shifted and to forecast where these trends are headed in the near future.

In essence, LIWC provided our means of converting the textual messages of tweets into quantitative data that we could analyze mathematically. When one uses LIWC to process texts, the program counts the total number of words contained in that text. It then counts the number of words falling into each of approximately 80 categories (such as emotion categories like "positive emotion," "sadness," social categories like "family," "friends," and other types of categories). For each of the roughly 80 categories, LIWC then computes a ratio of the number of words in the text falling into that category to the total number of words in the text. On both a daily and weekly basis, we combined all tweets posted in a day and in a week into a text file and then ran these files through LIWC. Doing so, enabled us to generate a series of ratios across time—one for each day or week—for each category. It was these ratios that we used as the raw material on which to run the algorithms presented in this paper.

Fig. 1 depicts raw tweet data. Each dot signifies an LIWC ratio for an indicator of anger (i.e., swearing) and an indicator of sadness, computed based on an aggregate of an entire week's worth of tweets. In looking at the pattern of these ratios, it is all too easy to make a judgment call as to when Twitter users' emotions shifted up or down and where to draw the line between various phases of their emotions. The danger in doing so lies in the fact that different observers may disagree as to where, exactly, to draw such lines. Moreover, if observers take into account newsworthy events, then doing so may lead to them to draw the lines differently depending on their views of the impact these events had on Twitter users' emotions.

Because of this potential problem, we developed a mathematical approach that determines, objectively, when Twitter users' emotions have shifted (or are shifting, as events unfold in real time). We introduce this algorithm in Section III.

III. An Automatic Breakpoint Analysis

There are a number of approaches one could use to automatically compute breakpoints. For example, Little and Jones

[4]Function words include the following categories: pronouns, propositions, articles, conjunctions, and auxiliary verbs [3]. These words have minimal lexical meaning yet they tie a sentence together. They frequently reveal psychological states which people are experiencing.

[5]Of course, like all other automated programs of its nature, LIWC does not perfectly capture the meanings of all words.

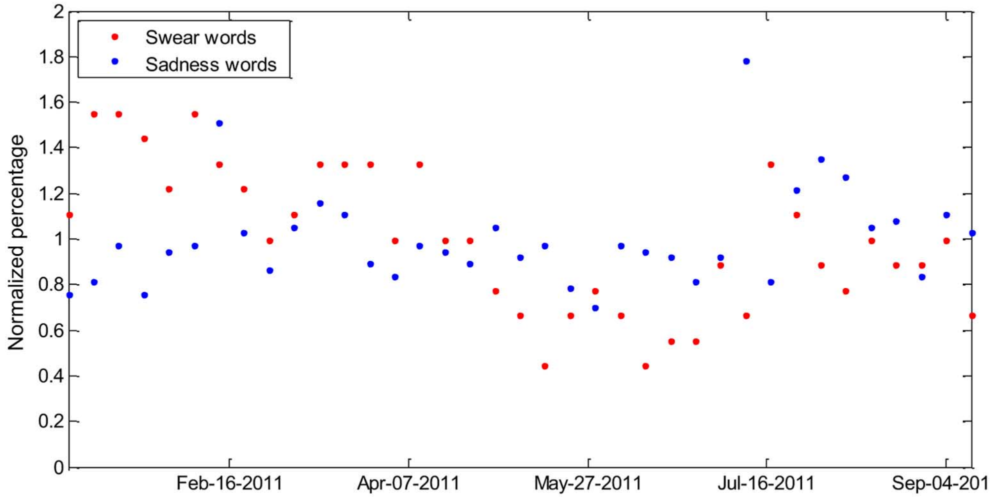
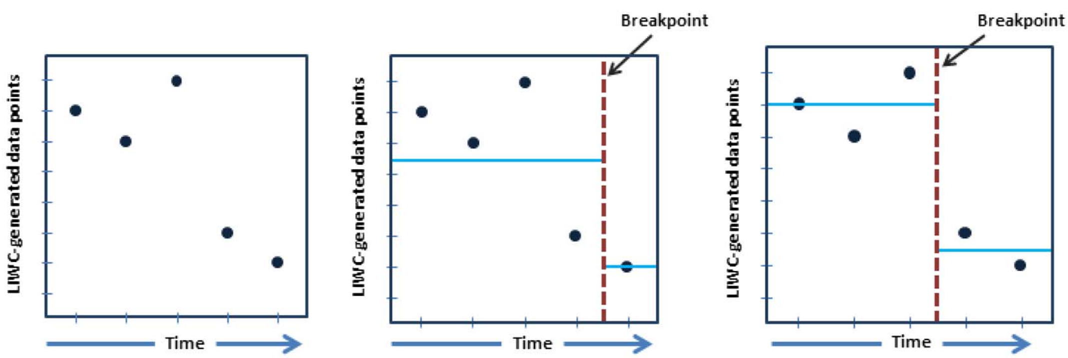


Fig. 1. Normalized percentages of LIWC swears and sadness words in weekly aggregates of Tweets.



(a) (b) (c)

Fig. 2. Five raw data points in a hypothetical set. (a) Five raw data points in a hypothetical set. (b) Alternative breakpoint #1. (c) Alternative breakpoint #2.

[38] provide an overview of methods for filtering piecewise constant signals, and Hawkins and Deng [30], Ross *et al.* [51], and Ross and Adams [50] describe recent algorithms that are all implemented in the **R** statistical language package **cpm**. From the computer science literature, Chung *et al.* [11] describe an approach to compute breakpoints while seeking to match the data to a predefined *set* of patterns. Currently, too little is known about the dynamics of social media to assume a more complicated pattern than a single simple linear dynamic. Motivated by computer graphics applications, Keogh *et al.* [36] examine greedy algorithms for determining discontinuous breaks which by design are very fast yet approximate. The social media data sets of interest in this paper are too small to require computational speed up. This paper's approach is based on [53] which is qualitatively described in Section III.A and the simpler nonrecursive algorithm described in Section III.B.

A. *Qualitative Description of the Approach*

Using a simple example with a single small hypothetical dataset, we will illustrate this mathematically unbiased approach to determining when shifts occurred in a dataset. Consider the set of five data points in Fig. 2(a), on the left:

Suppose the underlying dynamics of a system are such that it is reasonable to assume that these five data points represent noisy measurements of a system in one state which, some time while these data are being collected, discontinuously moves to another state (perhaps due to a newsworthy event). For example, if the discontinuity takes place after the fourth point in time, the horizontal blue lines (which correspond to the average value of the data during each time region) would correspond to Fig. 2(b), in the middle. Alternatively, if the breakpoint occurred between the third and fourth data point, the blue lines would similarly correspond to Fig. 2(c), on the right:

Which breakpoint is more reasonable and why?

An unbiased approach to determining which selection of breakpoints is better would be to drop a perpendicular orange arrow from each data point to the blue line (corresponding to the average in each region), [Fig. 3(a) and (b)], and summing the sizes, the distances, of these orange arrows.

Clearly, the division on the right has a smaller summation of the orange arrows. Hence, one arrives at the intuitive conclusion that the breakpoint on the right is the better choice.

For the case of finding a single division in the small hypothetical raw data in Fig. 2(a), the best division could have been found intuitively. However, for the raw Twitter data (Fig. 1),

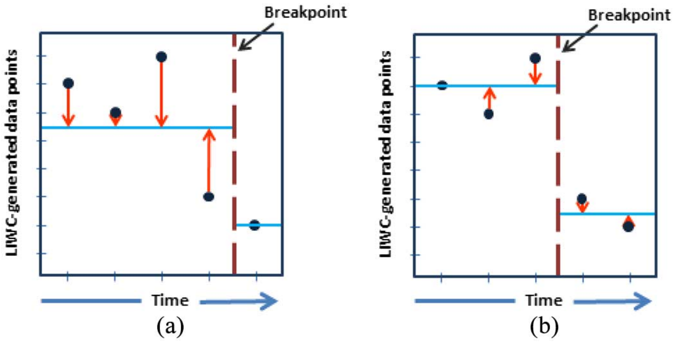


Fig. 3. Selecting the mathematically best breakpoint. (a) Mathematically sub-optimal breakpoint. (b) Mathematically best breakpoint.

intuition would not be sufficient. That is, for large data sets, we need the following generalizations. While the generalizations clearly require more computation to complete, it is conceptually no harder to find, and each step can be made mathematically quite rigorous.

1) Rather than examine only two possible sets of breakpoints, Fig. 2(b) and (c), examine *all* possibilities.
2) Rather than assume that the system has two regions, assume it may have more.
3) Rather than consider a single LIWC indicator consider multiple LIWC indicators.
4) Rather than assume emotions are constant in time, assume a larger class of dynamics, including but not limited to constant changes in time (i.e., straight lines).

Applying these first two assumptions to the raw weekly Twitter data of LIWC swears and sadness indicators,[6] one can algorithmically determine (in a manner described in Section III.B) the **best** set of breakpoints for the data. These breakpoints are shown in Fig. 4.

The above analysis is called a *constant model* because of the implicit assumption that the level of the LIWC emotions is a constant within a breakpoint region. This assumption can be relaxed to assume that the level of the LIWC emotions is a line within a breakpoint region. This will be referred to as the *linear model*. As indicated in the next section, the linear model can be analyzed in a manner analogous to the constant model. The result applied to the same data is shown in Fig. 5.

A similar analysis can be conducted using daily tweet data; this analysis is illustrated in Fig. 6.

Regarding influence, a researcher can use the breakpoint algorithm to detect situations in which a Twitterer posted a tweet creating so much influence that it caused a breakpoint in the levels of an emotion expressed—whether positive or negative. Upon seeing the breakpoint, the researcher would know which timeframe of tweets to examine for signs that a single user's tweet (or tweets) caused the breakpoint to occur—especially if that user's tweet(s) were retweeted. In fact, the breakpoint algorithm can be applied to retweets themselves—in order to determine whether a significant shift has occurred in the extent to which a tweet has been retweeted. This can be thought of as an additional type of influence—aside from causing a shift in the levels of emotions expressed.

While not used in this analysis, three additional generalizations that can be easily implemented would be

1) Rather than assume that each data point has the same importance, assume that the importance may be different for each (perhaps reflecting differing confidences in each point). This differing confidence might reflect the number of tweets sampled to generate each data point.
2) Rather than assume that each data set (derived, e.g., from LIWC indicators such as sadness) has the same importance, assume that the importance may be different for each (perhaps reflecting different experiences with each set for different applications).

As a final note, we wish to emphasize that as new data emerges over time, the algorithm recalculates the breakpoints to determine whether they still provide the best set of phases to the ever-changing data. Therefore, over time, breakpoints may appear, disappear, and reappear as the trends changes with newly acquired data. Alternatively, the breakpoints may remain quite stable with new data. One should not, therefore, consider the breakpoints to be permanent markers of where a phase-shift occurred in emotions. Thinking about this idea at a more conceptual/historical level, it often becomes clearer, with the passage of time, the points at which public opinion/mood entered a new phase regarding attempts to influence emotions. However, it may not be as clear at the very point when public mood is shifting that it is, in fact, shifting. This all creates a nuance to the quantification of influence which is yet to be explored.

B. Mathematical Description of the Approach

While Section III.A describes the key ideas behind the breakpoint analysis algorithm, it is useful to revisit it in a mathematically more precise manner for the interested reader.

Suppose we have a set of K time series x_n^k for $n = 1, \dots, N$ and $k = 1, \dots, K$ with time, n, separated into M partitions with breakpoints at $m_0(= 0)$, $m_1, \dots, m_{M-1}, m_M = N$. Suppose, within a breakpoint region, it is reasonable to assume the data arises either from a collection of piecewise constants, i.e., $x_n^k \approx y_j^k$, for $n = m_{j-1} + 1, \dots, m_j$, a collection of piecewise linear functions, i.e., $x_n^k \approx y_j^k(1) + n y_j^k(2)$, for $n = m_{j-1} + 1, \dots, m_j$, or more generally

$$\vec{x}_j^{\,k}(\vec{m}) \approx A_j(\vec{m})\vec{y}_j^{\,k}, \quad \text{for } j = 1, \dots, M \tag{1}$$

where

$$\vec{x}_j^{\,k}(\vec{m}) = (x_{m_{j-1}+1}^k, x_{m_j+2}^k, \dots, x_{m_j}^k)^T \tag{2}$$

$$A_j(\vec{m}) = \begin{pmatrix} 1 & m_j + 1 & (m_j + 1)^2 & \dots & (m_j + 1)^{D-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & m_{j+1} & (m_{j+1})^2 & \dots & (m_{j+1})^{D-1} \end{pmatrix} \tag{3}$$

[6]Combining the use of both the swear and sadness indicators yields more informative breakpoints than either indicator alone, so the rest of this paper uses both. Regardless of how many or which indicators one decides to use, one can algorithmically determine the *best* set of breakpoints, as illustrated in Fig. 4. In this figure, the vertical axis is a normalization of the percentages (making each have a mean of 1) for clarity.

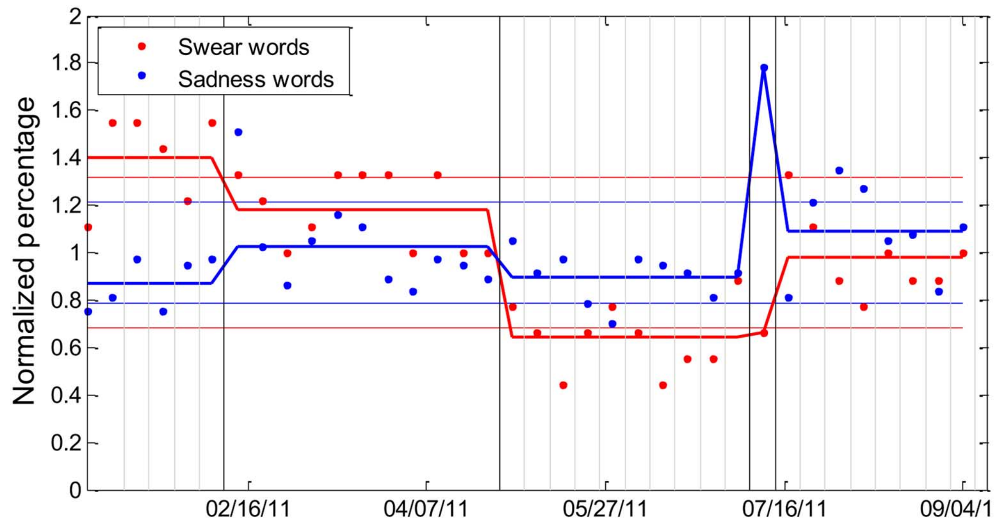


Fig. 4. Breakpoint analysis of normalized percentages of LIWC swears and sadness words in weekly aggregates of Tweets, using a constant model.

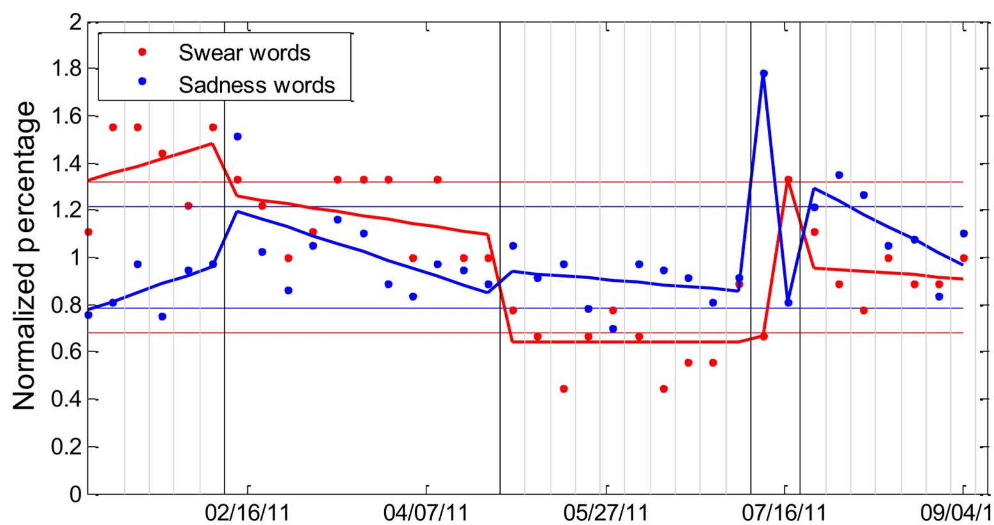


Fig. 5. Breakpoint analysis of normalized percentages of LIWC swears and sadness words in weekly aggregates of Tweets, using a linear model.

(where $D = 1$ and $D = 2$ are the cases of most interest) and $\vec{y}_j^k = (y_j^k(1), y_j^k(2), \ldots, y_j^k(D-1))^T$ is an unknown vector to be determined.

The *best* breakpoints are found from solving

$$m^*(\vec{m}) = \arg\min_{\vec{m}} \left[\sum_{j=1}^{M} \sum_{k} w_k \min_{\vec{y}_j^k} \|\vec{x}_j^k(\vec{m}) - A_j(\vec{m})\vec{y}_j^k\|_P^2 \right]. \tag{4}$$

Equation (4) can be interpretated as searching the space of all possible breakpoints $\vec{m}$ for the one which minimizes the distance between each data point of each data set and the fitted estimate from (1) (which might be a constant, a straight line, or another function) while the distances are weighted by the importance of the data point, (indicated by P), as well as the importance of the data set (indicated by $\vec{w}$). *P* might be set based on the number of tweets which generated the data point (as well as being diminished as one looks back in time. $\vec{w}$ might be set based on the seeming relevance of each data set on the issues of most interest.

The inner minimization is a simple linear regression so the best $\vec{y}_j^k$ in (4) is given by

$$\vec{y}_j^{k*} = \left(A_j^T(\vec{m})PA_j(\vec{m})\right)^{-1} A_j^T(\vec{m})P\vec{x}_j^k. \tag{5}$$

IV. Interpreting the Breakpoint Analysis and Forecasting

In Section IV, we introduce a mathematical algorithm to forecast, in the short term, the emotions of Twitter users. This algorithm can be used to understand the influence an individual might exert in the future, given what we know regarding her/his previous levels of influence. We designed this algorithm for use by anyone interested in generating forecasts of emotions for use in the analysis of influence or for its own right. However, Section IV also presents a more technical comparison (in Section IV-A) of the algorithm's predictive accuracy against that of two other algorithms—a baseline algorithm and an algorithm presented for comparison purposes. In the process, Section IV-A presents initial validation work for the

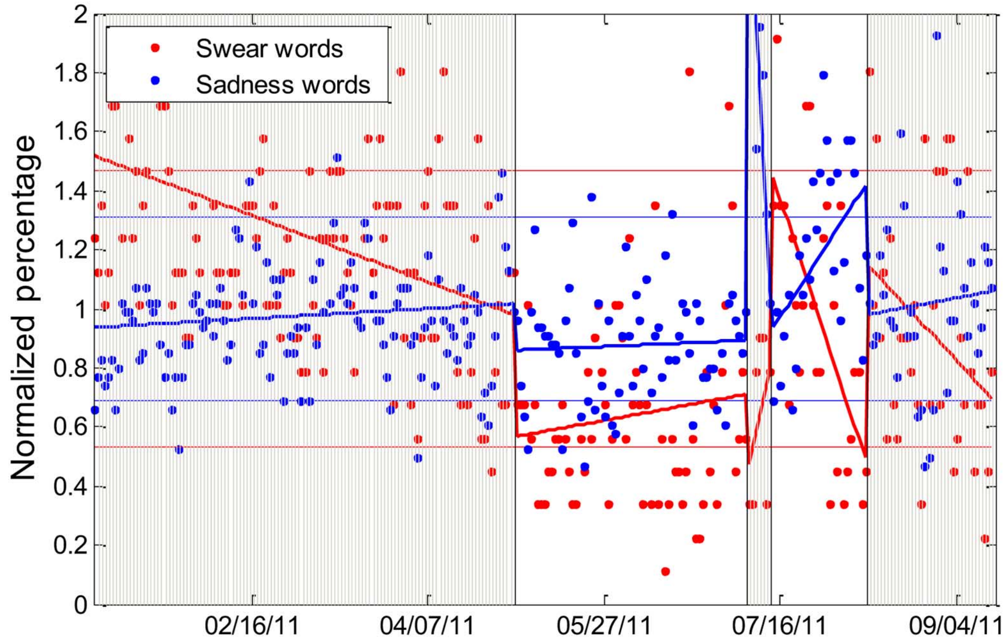


Fig. 6. Breakpoint analysis of normalized percentages of LIWC swears and sadness words in daily aggregates of Tweets, using a linear model.

most desired forecast algorithm for the data available. Readers wishing only to see the forecast algorithm itself may wish to skim Section IV-A and read Section IV-B in more depth, as Section IV-B presents an initial methodology for using the algorithm.

A. Forecasting: A Mathematical Forecasting Rule and Discussion of its Accuracy Compared with Other Rules

In Section III, we introduced the idea of mathematically calculating the points at which Twitter users' emotions shifted into a new phase. One can use this technique to understand how Twitter users' emotions changed in the present and recent past as events unfolded. This technique also opens the door to forecasting where those emotions are headed in the near future. Hence, this section introduces a mathematical extension of the breakpoint technique (in the form of a "forecasting rule") and presents a discussion of how that forecasting rule was selected over other potential forecasting rules. Specifically, this section presents a comparison of three potential forecasting rules and the accuracy with which they actually forecasted short-term trends in Twitter users' emotions.

An ultimate goal of those using this approach might be to use the output of the breakpoint analysis to forecast actions that will occur, such as whether consumers will buy more or less of a product in the near future. However, forecasting such actions using LIWC indicators would require, as a first step, making accurate forecasts regarding the future trends of the indicators themselves. Further research is needed to determine the extent to which LIWC indicators can be used to forecast actions—as opposed to just the emotions of Twitter users.

Forecasting, of course, is fraught with obstacles. Perhaps the most prominent is the possibility of self-negating prophesies, in which members of the public or leaders learn about a "prophesy" and then take steps to negate that prophesy.[7,8] Therefore, rigorous testing is crucial when developing forecasting techniques.

In Fig. 7, we illustrate and present the three mathematical approaches[9] ("Rules") to calculate a forecasted value (i.e., the ratio) of an LIWC indicator based on previous values of that indicator:

Rule 1: *Use the last dot*: Forecast the value of an LIWC indicator *without using the breakpoint analysis* by assuming that the last data point is the best estimate for the future. If this rule outperforms the others, it will mean that the breakpoint analysis is not necessary. As such, this might be considered the baseline forecast rule.

Rule 2: *Project ahead:* Perform a breakpoint analysis with the linear model and use this analysis to forecast by projecting ahead in time, starting from the most recent breakpoint. That is, extend the most recently derived trend line out longer (continuing along its current slope) and use that line to determine the forecasted value. This is the most natural approach. However, as indicated below, we expect it will demonstrate a comparative advantage only if there is an extensive amount of data in each individual breakpoint region and if the trend lines have large slopes.

[7] This phenomenon is closely related to the "paradox of warning" introduced and discussed in [37].

[8] For example, research on stock market prices has shown that the self-negating prophesy phenomenon hinders making forecasts of those prices according to the efficient market theory hypothesis [21]. Specifically, an anticipated future price change may spur traders to take immediate action, causing the change to take place in the present and, therefore, making the future unpredictable.

[9] The analysis in this paper builds on the assumption that the occurrence of a breakpoint results in the measurements before the breakpoint being statistically independent of measurements after the breakpoint. As such, there are limited options for prediction rules as the relevant points for prediction will be only those points after a breakpoint so a traditional ARMA prediction approach was not considered.

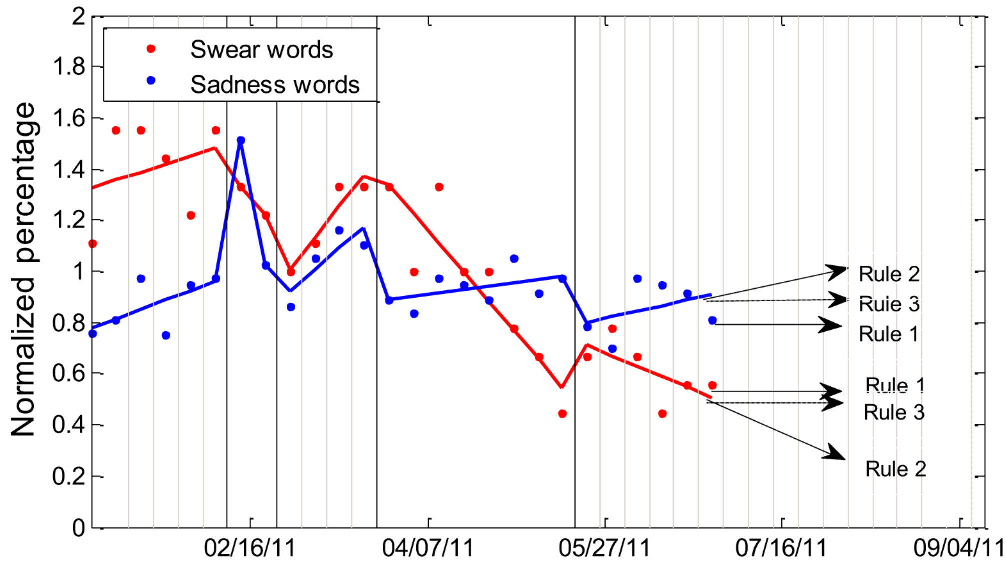


Fig. 7. Illustration of the three rules.

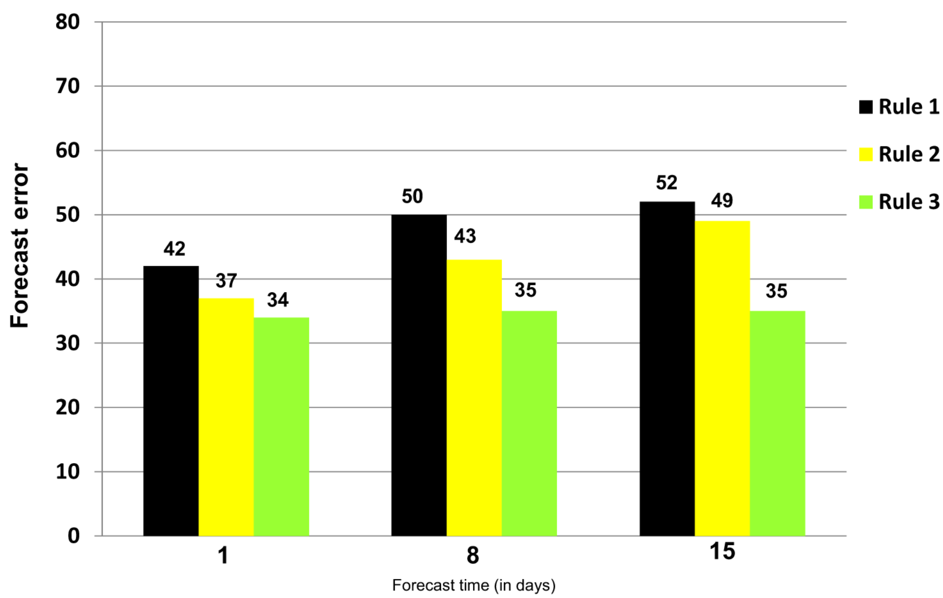


Fig. 8. Sum of square error in predicting normalized swear daily data.

Rule 3: Use last estimate: Perform a breakpoint analysis with the linear model and use this analysis to forecast that future values will be the same as the last *estimated* data point, i.e., extend, in time, the most recently derived trend line out from its end but with a *horizontal* slope—rather than continuing along the current slope of the line. While this may be less accurate and less intuitive than Rule 2 if the data are plentiful, this is a more conservative and robust rule which will perform well for a broader range of data availability and slopes of the trend lines.

In Figs. 8 and 9, we illustrate the accuracy of the three rules[10] using two LIWC indicators (swears and sadness) and examining the forecasted values generated at 1 day, 8 days, and 15 days into the future. We compared each set of predicted values[11] with the actual values that emerged and computed the amount by which our forecasts were off (i.e., the sum of squared errors).[12]

[10]More precisely, we performed the test by focusing on each day across the time period examined and comparing the accuracy of forecasts generated by each of the three rules for each given day. We began with the 101st day of the time period so that at least 100 prior daily data points would be processed for each day. The reason we chose to begin at this point was to ensure an adequate amount of data for the breakpoint analysis to be accurate.

[11]To facillitate comparisons between the swear and sadness predictions, all values were normalized so that the average value of both in the time region of interest was set to 1. In general, the more granular the data, the more accurate the forecasts will be. Our analysis was performed on a *daily* version of the data described in previous sections of this paper and not weekly data—as our preliminary studies indicate that daily data offer the minimal granularity needed for reasonably accurate forecasts.

[12]An alternative candidate measure for gauging the relative effectiveness of the different rules is to compute the Pearson correlation coefficient of the actual values and the predicted values. While this metric also has merit, this metric provides a relative measure of effectiveness in contrast to the sum of square approach which both quantifies the absolute error of each rule as well as the error relative to the baseline rule. This metric is also consistent with the metric in (4).

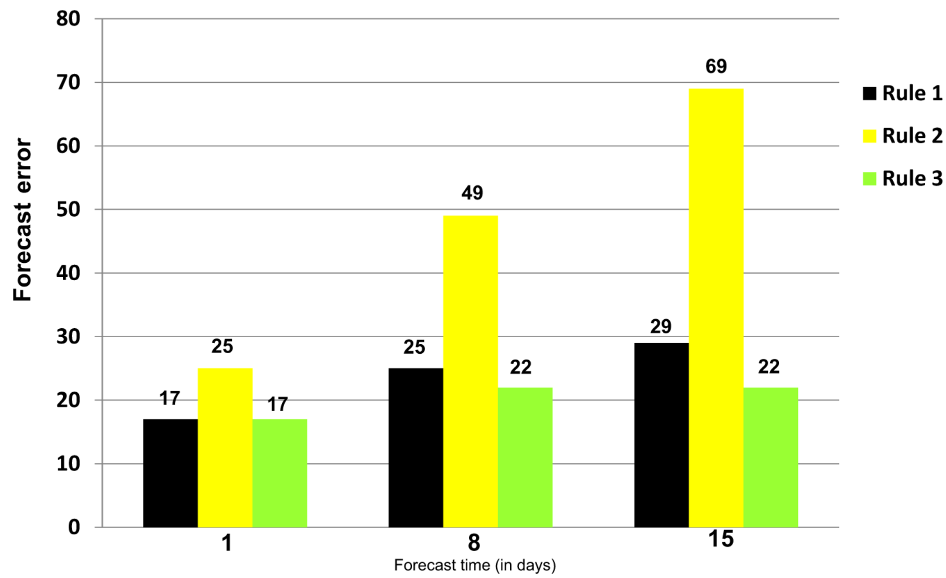


Fig. 9. Sum of square error in predicting normalized sadness daily data.

The key finding depicted on this figure is that for this data, Rule 3 consistently outperformed the baseline Rule 1 (which was generated without the breakpoint analysis) as well as Rule 2.[13] In addition, Rule 2 outperformed the baseline Rule 1 when applied against the normalized swear data but fell short when applied against the normalized sadness data.

We performed the above analysis with limited data which was available to us. As such, to make more definitive predictions and explore more subtle prediction rules would require a more exhaustive study with orders of magnitude more data.

B. Putting the Output of the Breakpoint Algorithm in Context

It should be clear that, by examining trend lines and breakpoints, one can gain an understanding of Twitter users' emotions that is more illuminating and subtle than examining raw data alone (such as in Fig. 1).

Yet, the trend lines and breakpoints must be put into a broader context in order to fully understand emotional patterns within Twitter. Certain questions are essential to address throughout the process of using the trend lines and breakpoints, such as: 1) are we applying them to an "extraordinary" period of emotional trends or to an "ordinary" one?; 2) how do we determine when a trend is really going "up" or "down"?; 3) what does it mean if a breakpoint disappears shortly after it appears?; and 4) how do we know that a trend line is really signifying the emotion we think it is?

[13]The estimated slope of the trend line generated by the linear model is an imperfect estimate. As such, in cases such as the one explored in this section, the precision of Rule 2 might be insufficient to outperform the more conservative Rule 3. However, one would expect Rule 2 to perform better as the number of data points in the breakpoint region increases and/or the slope of the trend line increases. Note, however, that there may be situations where Rule 3 outperforms Rule 2; yet Rule 2 (the linear model) can accurately predict the *sign* of the slope of the trend line (i.e., the direction which the emotions are moving).

When conducting an analysis using trend lines, breakpoints, and the forecast rule, we recommend, first, taking a step back to examine the larger picture of emotional trends. Examining this picture entails asking questions like "Have emotions truly entered a new phase compared with what they looked like before, during more typical times? For example, did emotions enter a new phase after beginning an ad campaign? Are the ups and downs I'm seeing in the LIWC indicators (e.g., sadness and swears) substantial, in the overall scheme of Twitter users' emotions, or are they relatively small?" One way to answer these questions consists of comparing the supposed new phase in LIWC indicators with "baseline" trends, collected before the new phase began. With twitter data, the "extraordinary" period may be as much as two orders of magnitude greater than the baseline period with regard to average LIWC ratios and /or the variation in these ratios (i.e., the standard deviation). Further empirical study will be needed to make recommendations regarding when LIWC values are elevated enough to warrant detailed breakpoint analysis. However, the higher the level of elevation, the more relevant the resulting interpretations are.

Another important consideration when examining trendlines is the fact that LIWC, like many (but not all) automated content analysis programs, ignores context, irony, sarcasm,[14] and the use of metaphors [10]. In that sense, although studies are yielding evidence that function words indicate emotional and biological states, status, honesty, and several individual differences, the imprecise measurement of word meaning and psychological states implies that researchers will not be able to detect those states with 100% accuracy solely by examining words automatically.

Another consideration in applying the breakpoint analysis is that breakpoints may shift as new data emerge, as discussed

[14]There has been some early research on identifying sarcasm [9], [26], and irony [23].

in Section III. This possibility highlights the need to gain confidence that a breakpoint (signifying a shift in Twitter users' emotions) is robust—meaning that their emotions have truly entered a new phase—before trying to forecast where their emotions are going. By the same logic, if a breakpoint disappears quickly with more data, there may be flux in Twitter users' emotions, making it hard to forecast where those emotions are headed. One could gain confidence that a breakpoint is robust by monitoring the breakpoint over time, and as the days continue, one's confidence increases. At this time, we advise initial caution when a breakpoint first appears before concluding that Twitter users' emotions have shifted into a new phase.

At this point, it is worth noting how valuable it can be to observe flux in the breakpoints, which signifies flux in the underlying Twitter users' emotions. The value of observing this flux is that one is implicitly learning that the emotions have not settled yet. One also is learning to temper one's certainty about the future.

While monitoring breakpoints to determine when Twitter users' emotions have shifted, one can also learn the directional change in emotions by examining whether trend lines are rising/falling and how sharply those rises/falls are (i.e., the slopes). Further research will need to determine, mathematically, how sharp a rise/fall must be for one to consider it significant. Naturally, as a trend line angles more sharply upward/downward, one gains confidence that emotions are changing quickly.

V. Conclusion

By combining an automated text analysis program with a new mathematical approach to detect shifts in Twitter users' emotional expressions, the work presented in this paper offers an opportunity to use the data obtained from social media to gain objectively derived insights into the dynamics of Twitter users. In real time, this method can provide insight into Twitter users' expressions in two ways:

1) identifies abrupt shifts in Twitter users' emotional expressions that occurred in the recent past, and pinpoint when they took place;
2) draws trend lines that indicate the direction in which present Twitter users' emotional expressions are changing in real time.

The work presented here has also laid the groundwork for forecasting, mathematically, what Twitter users' emotions will look like in the near future, and, ultimately, interpreting the meaning of changes with a high degree of reliability.

If, indeed, influence in social media entails the capacity to change the dynamics of the emotions of large groups of social media users then the new methods presented in this paper begin to create a foundation necessary to quantify that influence.

Acknowledgment

The authors would like to thank F. Chang, P. Garvey, and P. Lehner for their thoughtful comments regarding this paper. They would also like to thank S. Lunsford for his many long hours spent developing tools to facilliate this work.

References

[1] G. W. Alpers *et al.*, "Evaluation of computerized text analysis in an Internet breast cancer support group," *Comput. Human Behav.*, vol. 21, pp. 361–376, 2005.

[2] J. J. Anger and C. Kittl, "Measuring influence on Twitter," in *Proc. 11th Int. Conf. Knowl. Manage. Knowl. Technol.*, New York, NY, USA: ACM, 2011, vol. 31, no. 4, pp. 1–31.

[3] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA, USA: Kluwer, 2001.

[4] J. Barden and R. E. Petty, "Persuasion," in *Encyclopedia of Human Behavior*, vol. 3, V. S. Ramachandran, Ed., 2nd ed. New York, NY, USA: Academic, 2012, pp. 96–102.

[5] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. ICWSM*, 2011, p. 452.

[6] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective rating," Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Tech. Rep. C-1, 1999.

[7] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy" in *Proc. 4th Int. AAAI Conf. Weblogs Social Media*, 2010, pp. 10–17 [Online]. Available: www.aaai.org

[8] S. Chaiken, W. Wood, and A. H. Eagly, "Principles of persuasion," in *Social Psychology: Handbook of Basic Principles*, E. T. Higgins and A. W. Kruglanski, Eds. New York, NY, USA: Guilford, 1996, pp. 702–744.

[9] C. Chew and G. Eysenbach, "Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak," *Plos One*, vol. 5, no. 11, pp. e14118, 2010, accessed on Dec. 14, 2014 [Online]. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0014118#abstract0.

[10] C. K. Chung and J. W. Pennebaker, "The psychological function of function words," in *Social Communication: Frontiers of Social Psychology*, K. Fiedler, Ed. New York, NY, USA: Psychology Press, 2007, pp. 343–359.

[11] F. L. Chung, T. C. Fu, V. Ng, and R. W. Luk, "An evolutionary approach to pattern-based time series segmentation," *IEEE Trans. Evol. Comput.*, vol. 8, no. 5, pp. 471–489, Oct. 2004.

[12] R. Cialdini, *Influence: The Psychology of Persuasion*. New York, NY, USA: Morrow, 1993.

[13] B. Costa and J. A. Boiney, "Social radar," in *Proc. NATO Human Factors Med. Panel HFM-201—Spec. Meeting Social Media Risks Oppor. Mil. Appl.*, 2012, pp. 3.1–3.11.

[14] H. Cruz, R. C. Coelho, and C. Codeco, "Epigrass: A tool to study disease spread in complex networks," *Source Code Biol. Med.*, vol. 3, no. 3, pp. 1–9, 2008.

[15] D. Day, J. Boiney, T. Brown, and M. Ubaldino, "Multi-channel sentiment analysis," in *Proc. Int. Cross-Cultural Decis. Making Appl. Human Factors Eng. Conf.*, San Francisco, CA, USA, 2012, pp. 4333–4342.

[16] P. S. Dodds and C. M. Danforth, "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents," *J. Happiness Stud.*, vol. 11, pp. 441–456, 2010.

[17] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. 7th Int. Conf. Knowl. Discov. Data Min.*, 2001, pp. 57–66.

[18] Z. Estes and J. S. Adelman, "Automatic vigilance for negative words in lexical decision and name: Comment on Larsen, Mercer, and Balota (2006)," *Emotion*, vol. 8, pp. 441–444, 2008.

[19] A. Esuli and F. Sebastiani, "Determining term subjectivity and term orientation for opinion mining," in *Proc. Eur. Chapter Assoc. Comput. Linguistics (EACL)*, 2006, pp. 617–624.

[20] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proc. Lang. Resource Eval. (LREC)*, 2006, pp. 417–422.

[21] E. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[22] F. A. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," in *Proc. ESWC Workshop 'Making Sense Microposts*,' May 2011, pp. 93–98 [Online]. Available: http://arxiv.org/abs/1103.2903

[23] E. Filatova, "Irony and Sarcasm: Corpus generation and analysis using crowdsourcing," in *Proc. Lang. Resour. Eval. (LREC)*, 2012, pp. 392–398.

[24] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. (2007). "The development and psychometric properties of LIWC2007," Austin, TX, USA: LIWC. Net [Online]. Available: www.liwc.net

[25] E. M. Gortner and J. W. Pennebaker, "The archival anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M bonfire tragedy," *J. Social Clin. Psychol.*, vol. 22, pp. 580–603, 2003.
[26] R. González-Ibáñez, S. Muresan, and N. Wacholder, "Identifying sarcasm in Twitter: A closer look," in *Proc 49th Annu. Meeting Assoc. Comput. Linguistics: Human Lang. Technol.*, 2011, vol. 2, pp. 581–586.
[27] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Sociol.*, vol. 83, no. 6, pp. 1420–1443, 1978.
[28] R. Guadagno and R. Cialdini, "Online persuasion and compliance: Social influence on the Internet and beyond," in *The Social Net: Understanding Human Behavior in Cyberspace*. New York, NY, USA: Oxford Univ. Press, 2005, pp. 91–113.
[29] J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," *Present Tense*, vol. 27, no. 16, pp. 42–48, 2012.
[30] D. M. Hawkins and Q. Deng, "A nonparametric change-point control chart," *J. Qual. Technol.*, vol. 42, no. 2, pp. 165–173, 2010.
[31] M. E. Ireland *et al.*, "Language style matching predicts relationship initiation and stability," *Psychol. Sci.*, vol. 22, no. 1, pp. 39–44, 2011.
[32] M. O. Jackson and L. Yariv, "Diffusion of behavior and equilibrium properties in network games," *Amer. Econ. Rev.*, vol. 97, no. 2, pp. 92–98, 2007.
[33] V. Kagan, E. Rossini, and D. Sapounas, *Sentiment Analysis for PTSD Signals*. New York, NY, USA: Springer, 2013.
[34] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson, "Measuring emotional expression with linguistic inquiry and word count," *Amer. J. Psychol.*, vol. 120, pp. 263–286, 2007.
[35] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2003, pp. 137–146.
[36] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," in *Data Mining in Time Series Databases*. New York, NY, USA: World Scientific, 2004, vol. 57, pp. 1–22.
[37] P. Lehner and A. Michelson, "Measuring the forecast accuracy of intelligence products," *Stud. Intell.*, 2014, submitted for publication.
[38] M. A. Little and N. S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals," *Proc. Roy. Soc.*, vol. 467, no. 2135, pp. 3088–3144, 2011.
[39] J. Mathieu, M. Fulk, M. Lorber, G. Klein, and B. Costa, "Social radar workflows, dashboards, and environments," in *Proc. NATO Human Factors Med. Panel HFM-201—Spec. Meeting Social Media: Risks Oppor. Mil. Appl.*, 2012, pp. 25.1–25.6.
[40] D. McNair, J. P. Heuchert, and E. Shilony, *Profile of Mood States. Bibliography 1964–2002*. San Diego, CA, USA: Multi-Health Systems, 2003.
[41] D. McNair, M. Loor, and L. Droppleman, *Profile of Mood States*. San Diego, CA, USA: Educational and Industrial Testing Service, 1971.
[42] M. R. Mehl and J. W. Pennebaker, "The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations," *J. Pers. Social Psychol.*, vol. 84, pp. 857–870, 2003.
[43] D. C. Mutz, *Population-Based Survey Experiments*. Princeton, NJ, USA: Princeton Univ. Press, 2011.
[44] M. L. Newman, C. J. Groom, L. D. Handelman, and J. W. Pennebaker, "Gender differences in language use: An analysis of 14 000 text samples," *Discourse Processes*, vol. 45, no. 3, pp. 211–236, 2008.
[45] J. W. Pennebaker, "Putting stress into words: Health, linguistic, and therapeutic implications," *Behav. Res. Ther.*, vol. 31, pp. 539–548, 1993.
[46] J. W. Pennebaker, R. E. Booth, and M. E. Francis, *Linguistic Inquiry and Word Count: LIWC2007—Operator's Manual*. Austin, TX, USA: LIWC.net, 2007.
[47] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, *The Development and Psychometric Properties of LIWC2007*. Austin, TX, USA: LIWC.net, 2007.
[48] A. Pepe and J. Bollen, "Between conjecture and memento: Shaping a collective emotional perception of the future," in *Proc. AAAI Spring Symp. Emotion Pers. Social Behav.*, 2008, pp. 111–116.
[49] R. E. Petty and J. T. Cacioppo, "The effects of involvement on responses to argument quantity and quality: Central and peripheral approaches to persuasion," *J. Pers. Social Psychol.*, vol. 46, pp. 69–81, 1984.
[50] G. J. Ross and N. M. Adams, "Two nonparametric control charts for detecting arbitrary distribution changes," *J. Qual. Technol.*, vol. 44, no. 2, pp. 102–116, Apr. 2012.
[51] G. J. Ross, D. K. Tasoulis, and N. M. Adams, "Nonparametric monitoring of data streams for changes in location and scale," *Technometrics*, vol. 53, no. 43, pp. 159–178, 2011.
[52] S. S. Rude, E. M. Gortner, and J. W. Pennebaker, "Language use of depressed and depression-vulnerable college students," *Cognit. Emotion*, vol. 18, pp. 1121–1133, 2004.
[53] L. D. Servi, "Analyzing social media data having discontinuous underlying dynamics," *Oper. Res. Lett.*, vol. 41, no. 6, pp. 581–585, 2013.
[54] P. Shakarian, V. S. Subrahmanian, and M. L. Sapino, "Using generalized annotated programs to solve social network optimization problems," in *LIPIcs-Leibniz Int. Proc. Informat.*, 2010, vol. 7, pp. 182–191.
[55] Y. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Social Psychol.*, vol. 29, pp. 24–54, 2010.
[56] P. E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*. Princeton, NJ, USA: Princeton Univ. Press, 2005.
[57] A. Tumasjan *et al.*, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proc. 4th Int. Conf. Weblogs Social Media (ICWSM'10)*, 2010, pp. 178–185.
[58] H. Tsukayama. (2013, Mar. 21). "Twitter turns 7: Users send over 400 million tweets per day," *The Washington Post* [Online]. Available: http://articles.washingtonpost.com/2013-03-21/business/37889387_1_tweets-jack-dorsey-twitter
[59] M. O. Jackson, "A survey of network formation models: Stability and efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, G. Demange and M. Wooders, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2005, pp. 11–49.
[60] T. C. Schelling, *Micromotives and Macrobehavior*. New York, NY, USA: Norton, 1978.

Les Servi (S'79–M'84–SM'88) received the M.S. and Ph.D. degrees in engineering from Harvard University, Cambridge, MA, USA, and the Sc.B. and Sc.M. degrees in applied mathematics from Brown University, Providence, RI, USA.

He runs the Decision Analytics Group at The MITRE Corporation, Bedford, MA, USA. Prior to joining MITRE, he worked with Bell Laboratories, Murray Hill, NJ, USA, Verizon, MIT Lincoln Laboratories, Lexington, MA, USA, and was a Visiting Research Scientist with Harvard University and the Massachusetts Institute of Technology, Cambridge, MA, USA for a year. While at MITRE, he served on a Defense Science Board task force. He is a former Editor of *Operations Research*, *Management Science*, and the *ORSA Journal on Computing*. He has an extensive publication record and 10 patents derived from his operations research and most recently social media analysis.

Dr. Servi is an INFORMS Fellow, a former member of the Board of Directors of INFORMS, and a former Chair of the INFORMS Applied Probability subdivision and the INFORMS Telecommunication subdivision. He is the current Chair of the INFORMS Boston Chapter and the founding Chair of the INFORMS Social Media Analytics subdivision.

Sara Beth Elson received the Ph.D. degree in social psychology from the Ohio State University, Columbus, OH, USA.

She is a Behavioral Scientist with The MITRE Corporation, Bedford, MA, USA. She leads a research program aimed at developing new techniques for analyzing social media, with a focus on psycholinguistic indicators of emotion, social processes, and cognitive dispositions and how change in these indicators can be analyzed mathematically. She also provided analytical support to the Office of the Secretary of Defense Human Social Culture Behavior (HSCB) Modeling Program. Prior to joining MITRE, she worked with RAND Corporation, Santa Monica, CA, USA, where she initiated the current program of research on social media and worked on a variety of other projects, such as improving U.S. information operations in Afghanistan, Army strategic communication, communication of assessments to policymakers, and manpower planning.