

Public Opinion Polling with Twitter

Emily M. Cody,^{1,*} Andrew J. Reagan,^{1,†} Peter Sheridan Dodds,^{1,‡} and Christopher M. Danforth^{1,§}

¹*Department of Mathematics & Statistics, Vermont Complex Systems Center,
Computational Story Lab, & the Vermont Advanced Computing Core,
The University of Vermont, Burlington, VT 05401.*

(Dated: August 9, 2016)

Solicited public opinion surveys reach a limited subpopulation of willing participants and are expensive to conduct, leading to poor time resolution and a restricted pool of expert-chosen survey topics. In this study, we demonstrate that unsolicited public opinion polling through sentiment analysis applied to Twitter correlates well with a range of traditional measures, and has predictive power for issues of global importance. We also examine Twitter's potential to canvas topics seldom surveyed, including ideas, personal feelings, and perceptions of commercial enterprises. Two of our major observations are that appropriately filtered Twitter sentiment (1) predicts President Obama's job approval three months in advance, and (2) correlates well with surveyed consumer sentiment. To make possible a full examination of our work and to enable others' research, we make public over 10,000 data sets, each a seven-year series of daily word counts for tweets containing a frequently used search term.

I. INTRODUCTION

Public opinion data can be used to determine public awareness, to predict outcomes of events, and to infer characteristics of human behaviors. Indeed, readily available public opinion data is valuable to researchers, policymakers, marketers, and many other groups, but is difficult to generate. Solicited polls can be expensive, prohibitively time consuming, and may only reach a limited number of people on a limited number of days. Polling accuracy evidently relies on accessing representative populations and high response rates. Poor temporal sampling will weaken any poll's value as individual opinions vary in time and in response to social influence[12, 43].

With the continued rise of social media as a communication platform, the ability to construct unsolicited public opinion polls has become a possibility for researchers through parsing of massive text-based datasets. Social media provides extraordinary access to public expressions in real time, and has been shown to play a role in human behavior [23].

With its open platform, Twitter has proved to be a boon for many research enterprises [37], having been used to explore a variety of social and linguistic phenomena [10, 30, 31]; harnessed as a data source to create an earthquake reporting system in Japan [42]; made possible detection of influenza outbreaks [7]; and used to analyze overall public health [38]. Predictions made using Twitter have focused on elections [22, 44], the spread of disease [41], crime [46], and the stock market [35]. These studies demonstrate a proof-of-concept, avoiding the more difficult task of building operational systems for continued forecasting.

We must be clear that for all its promise, prediction via social media is difficult. Indeed, we have seen a number of high profile failures such as Google Flu trends [29] and various attempts to predict election outcomes [21]. OpinionFinder was used in [9] to evaluate tweets containing direct expressions of emotion as in 'I feel', or 'I am feeling', and shown not to have predictive power for the stock market

Despite limitations which we address later in Sec. IV, Twitter data reveals an unprecedented view of human behavior and opinion related to major issues of global importance [32]. In a previous study [14], we analyzed the sentiment surrounding climate change conversation on Twitter. We discovered that sentiment varies in response to climate change news and events, and that the conversation is dominated by activists. Another study by Helmuth et al. analyzed tweets by United States Senators to determine which research oriented science organizations and which senators are best at getting science-related findings into the hands of the general public [24]. Twitter is also often used to analyze public opinion of political issues [8, 17, 45], and in several previous works as an opinion polling resource. In an application using neural networks called TrueHappiness, users enter one of 300,000 words to obtain a sentiment estimation based on this word's usage in a massive Wikipedia data set, and on previously collected sentiment scores for 10,222 words on Amazon's Mechanical Turk [16, 18], hereafter referred to as the labMT word set. In another application called RACCOON, users are invited to enter a query term and a rough sketch to obtain words or phrases on Twitter that correlate well with the inputs [6]. Google Correlate is a similar tool that discovers Google searches for terms or phrases that match well with real-world time series [3]. Financial term searches from Google Trends was shown by Preis et al. [39] to correlate with Dow Jones economic indices.

We argue that Twitter is a better source for opinion mining than Wikipedia, used in TrueHappiness, due

* emily.cody@uvm.edu

† andrew.reagan@uvm.edu

‡ peter.dodds@uvm.edu

§ chris.danforth@uvm.edu

to the personal nature of each post. RACCOON uses only user estimates for correlations and not actual survey data. Google Correlate uses only frequencies of Google searches to compare time series and may only be useful in specific situations.

In many studies that use text data from Twitter, the results are not compared to actual polling data, leaving the conclusions open to interpretation. One example work which does make a direct comparison is [36], where the authors use a Twitter data set from 2008 and 2009 to compare sentiments on Twitter, calculated with OpinionFinder, with daily and monthly public opinion polls. Our approach here is analogous to that of [36], but we use the sentiment analysis techniques developed in [14] to investigate public opinion regarding over 10,000 search terms.

Specifically, for each of 10,222 of the most frequently used English words, we calculate daily, weekly, monthly, and yearly sentiment time series of co-occurring words from September 2008 to November 2015. We compare many of these happiness time series to actual polling data, which is not typically available at such a high resolution in time. We investigate a wide range of topics, including politics, the economy, and several commercial organizations. Given the size of the dataset, we are unable to exhaustively compare all search terms to solicited opinion polls, and have released the data publicly along with this paper (see <http://compstorylab.org/share/papers/cody2016a/index.html>).

Overall, our aim is to determine the extent to which Twitter can be used to complement traditional public opinion surveys, ideally as a dashboard indicator accompanied by solicited feedback.

II. METHODS

We implement the “hedonometer”, an instrument designed to calculate a happiness score for a large-scale text, based on the happiness of individual words in the text. The hedonometer uses previously assessed happiness scores for the labMT word set, which contains the most frequently used English words in four disparate corpora [27]. We choose the hedonometer to obtain lexical coverage of Twitter text and produce meaningful word shift graphs [40].

The words were scored in isolation by human subjects in previous work on a scale from 1 (least happy) to 9 (most happy). We remove neutral and ambiguous words (scores between 4 and 6) from the analysis. For details regarding stop words, see Dodds et al. [18].

We use the hedonometer to calculate what we refer to as *ambient happiness* scores for each of the labMT words (first defined in [18]). We determine ambient happiness of a given word, w_j , by calculating the average happiness

of the words that appear in tweets with that word, i.e.,

$$h_{\text{amb}}(w_j) = \frac{\sum_{i=1, i \neq j}^N h_{\text{avg}}(w_i) f_i}{\sum_{i=1, i \neq j}^N f_i}. \quad (1)$$

Here, w_i is a word that appears in a tweet with word w_j , $h_{\text{avg}}(w_i)$ is the surveyed happiness score of word i , f_i is the frequency of word i , and N is the number of words in labMT (with stop words removed) appearing in tweets containing word j . Note that we do not include the frequencies or happiness scores of the given word (w_j) in the calculation of ambient happiness.

For example, $h_{\text{avg}} = 8.42$ for “love” and the ambient happiness for tweets containing “love”, averaged using Eqn 1 over the 7 year period, is 6.17. For “hate”, $h_{\text{avg}} = 2.34$ and we find the ambient happiness of “hate” is 5.75. As seen in the Appendix in Fig. B.1, we find that due to averaging, ambient happiness covers a smaller range of scores than labMT happiness for individual words.

We use the ambient happiness scores to create time series for each of the words in the labMT word set, and we correlate the happiness time series with polling data at various temporal resolutions.

A. Data

We collected tweets from Twitter’s gardenhose API from September 2008 to November 2015. During this time period, the volume of tweets grew by three orders of magnitude, but the random percentage of all public tweets fell from 50% to 10%. For each word in the labMT dictionary, e.g. “Obama”, we subsample the gardenhose for all tweets matching the word. We then tabulate the daily frequencies of labMT words in this term-defined collection of tweets, resulting in temporal counts of the words co-occurring with “Obama”. For example, the resulting collection of counts for “Obama” is a 2,628 (days) by 10,222 (words) matrix with entry (i, j) representing the frequency of labMT word j appearing in a tweet containing the term “Obama” on day i . This collection of counts is posted on the [online Appendix](#) for this paper.

Fig. 1 gives the average daily ambient happiness of “Obama”, along with the average daily happiness of all tweets during the same time period. Along with a general slow decline, we see spikes in happiness each year on August 4th, the President’s birthday, with the largest spike occurring on October 9, 2009 when President Obama was awarded the Nobel Peace Prize. We see a strong dip shortly after on October 26, 2009 when President Obama declares a state of emergency for the H1N1 virus. We see spikes in relative frequency of “Obama” on both election days in 2008 and in 2012.

To compare our findings with solicited opinions, we collected yearly and quarterly polling data from Gallup [2]. We focus on quarterly data, as it is the highest resolution

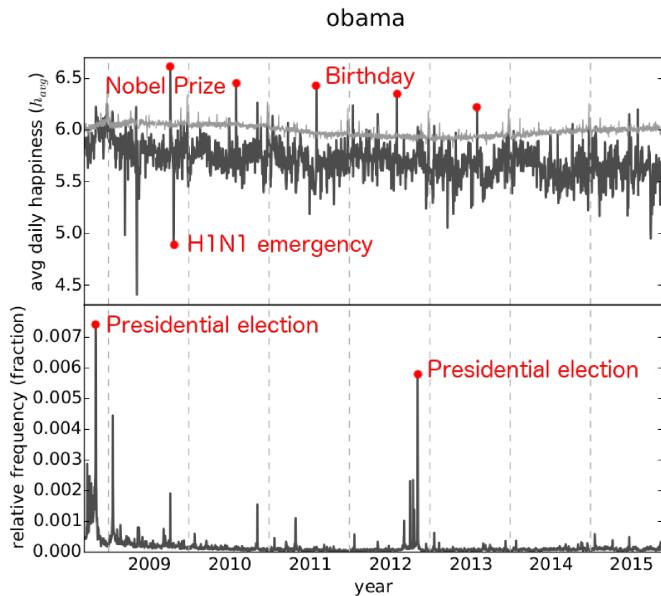


Figure 1. Average daily happiness of tweets containing “Obama” (top) with the relative frequency of “Obama” tweets (bottom). Spikes in happiness include President Obama’s birthday (August 4th) and his winning of the Nobel Prize (10/09/2009). Dips include a state of emergency for the H1N1 virus. Spikes in relative frequency occur on election days in 2008 and 2012.

we were able to obtain. The yearly analysis provides us with only 7 data points, and results are in the Appendix. We compare President Obama’s job approval rating on Gallup and on Pollster [4], which allows for daily data collection through their API. Finally, we use the University of Michigan’s Index of Consumer Sentiment data, which is collected monthly [5].

III. RESULTS

A. Unsolicited Public Opinions

Here we present happiness time series for several words for which we find interesting patterns. Figure 2 gives examples of ambient happiness and relative frequency time series for a few selected words. Happiness associated with certain religious words, e.g. “church” and “muslim” has decreased in recent years, with dips corresponding to several mass shootings including the Charleston shooting in June 2015 and the Chapel Hill shooting in February 2015. The relative frequency of “church” peaks each year on Easter Sunday. We see that ambient happiness of “snow” is seasonal, with the highest happiness during the northern hemisphere summer and lowest during the winter, while the relative frequency is highest during the winter and lowest during the summer. The saddest day for “snow” was June 14, 2015 when a popular Game of

Thrones character is presumed dead. The ambient happiness scores of “democrat” and “republican” are on a slow decline, with relative frequencies peaking during presidential and midterm elections. President Obama’s press conference after the Sandy Hook shooting is the saddest day for “democrat”, while the saddest day for “republican” coincides with an incident involving the Egyptian Republican Guard. Ambient happiness of “love” peaks around the holidays each year, and the relative frequency was increasing until recently. While ambient happiness of “love” peaks on Christmas each year, the relative frequency of “love” peaks on Valentine’s Day each year, which could be due to the difference in labMT scores for “christmas” and “valentines” (7.96 and 7.30 respectively).

In traditional polls, there may be large differences in public opinion from one time period to the next. In a yes/no or multiple choice survey question it is impossible to use that data to determine why differences occur. Here we use word shift graphs to determine the cause of a shift in ambient happiness.

A word shift graph ranks words by their contributing factor to the change in happiness between two pieces of text. For example, in Fig. 3 we investigate why the ambient happiness of “snow” is higher in the northern hemisphere summer (when its relative frequency is lowest) and lower in the winter (when its relative frequency is highest).

The word shift graph in Fig. 3 compares “snow” tweets during the winter months (December, January, February) to “snow” tweets in the summer months (June, July, August). English speaking countries like Australia and New Zealand will necessarily be included in the wrong season, however their contribution is small.

We find that Twitter users loathe the snow during the winter, and miss the snow during the summer, as indicated by the increase in the word “hate”, negatives, and profanity during the winter months and the decrease in the word “love”. The influence of the Disney classic “Snow White” is also visible, appearing to be referenced more often in summer months due to its motion picture release on June 1, 2012.

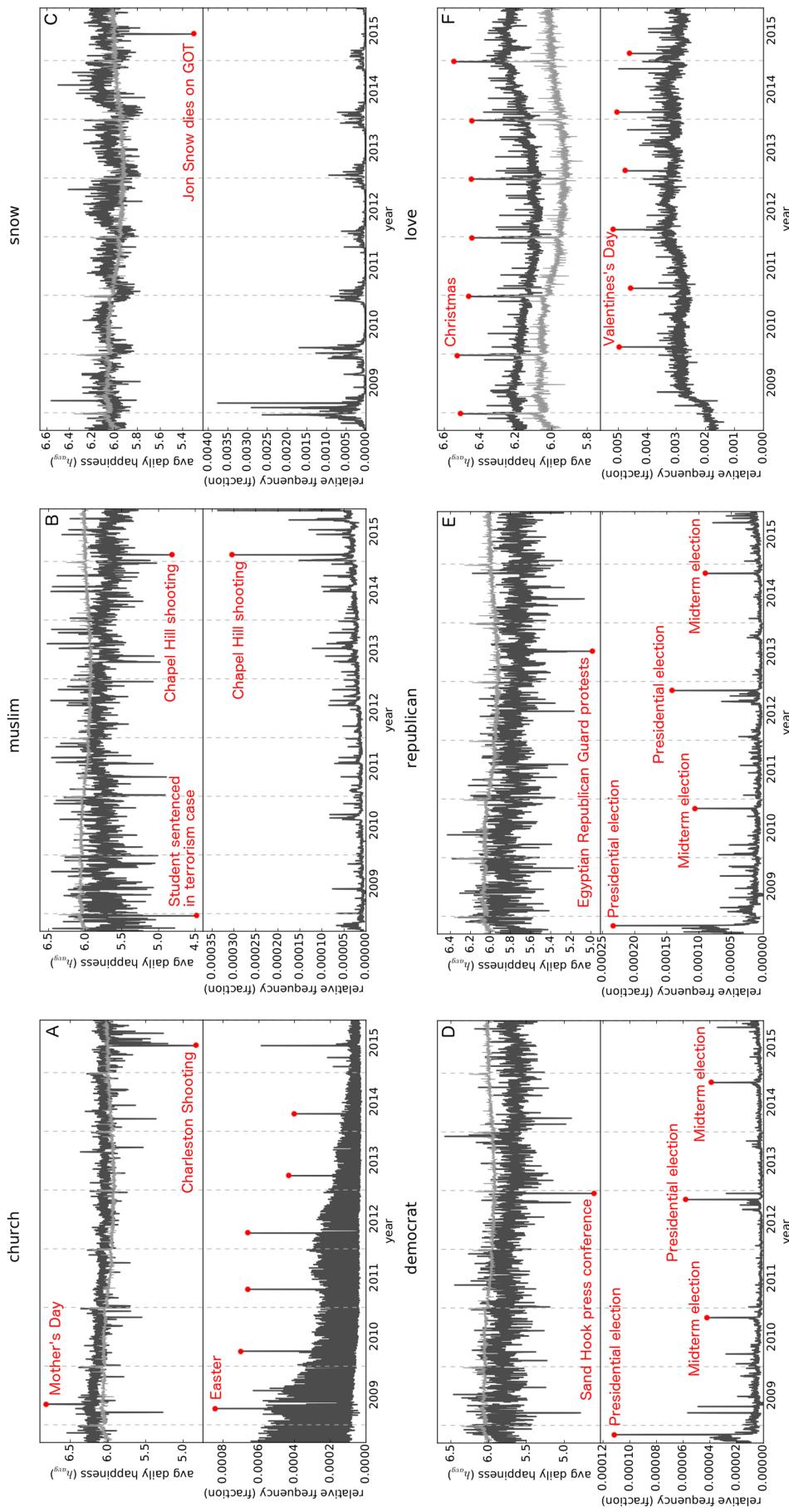


Figure 2. Six examples of ambient happiness time series (top, dark gray) along with relative frequency (bottom). Twitter's overall average happiness trend is in light gray for each plot. Relative frequency is approximated by dividing the total frequency of the word by the total frequency of all labMT words on a given day. (A) “church”: There is a large spike in happiness on Mother's day and a large dip following the Charleston church shooting in June 2015. There are spikes in relative frequency each Sunday, and yearly on Easter Sunday. (B) “muslim”: Two dips correspond to a sentencing in a terrorism case in late 2008, and the shooting at Chapel Hill in February 2015. (C) “snow”: Sentiment and relative frequency are seasonal, with a main dip when a character dies on the HBO show Game of Thrones. (D) “democrat”: Overall sentiment gradually decreases with a large dip after president Obama's press conference following the Sandy Hook shooting. There are spikes in relative frequency on election days. (E) “republican”: Overall sentiment gradually decreases with a large dip after protests of the Egyptian Republican Guard. (F) “love”: Sentiment peaks each year on Valentine's Day. Weekly and monthly ambient happiness time series for nearly 10,000 terms can be found in the online Appendix for the paper.

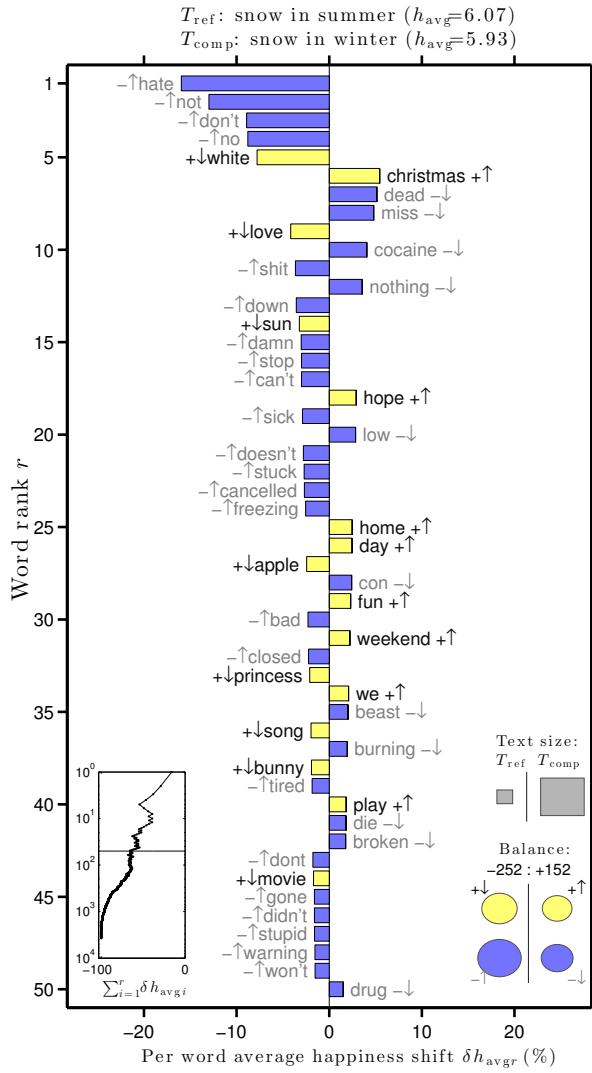


Figure 3. A word shift graph comparing tweets that contain the word “snow” during the summer months (reference text) and winter months (comparison text). A purple bar indicates a relatively negative word, a yellow bar indicates a relatively positive word, both with respect to the reference text’s average happiness. An up arrow indicates that word was used more in the comparison text. A down arrow indicates that word was used less in the comparison text. Words on the left contribute to a decrease in happiness in the comparison text. Words on the right contribute to an increase in happiness in the comparison text. The circles in the lower right corner indicate how many happy words were used more or less and how many sad words were used more or less in the comparison text.

B. President Obama’s Job Approval Rating

We next investigate the relationship between President Obama’s Job Approval Rating from two public opinion polling resources and the ambient happiness of “Obama” tweets.

President Obama’s quarterly job approval rating

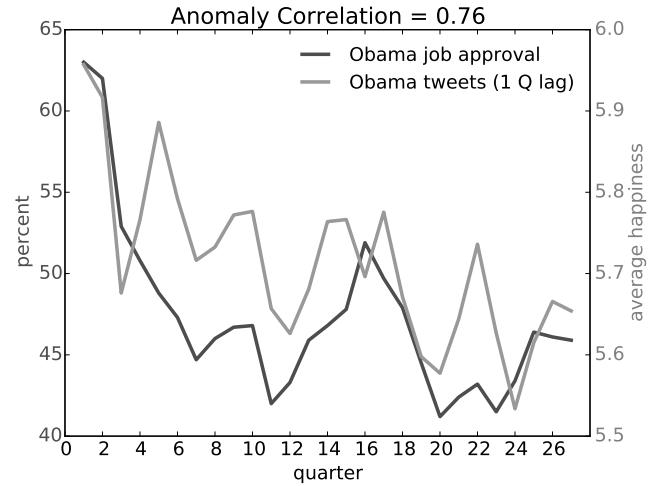


Figure 4. Average quarterly happiness of tweets containing “Obama” on a one quarter lag with Obama’s quarterly job approval rating. The high positive correlation indicates opinions on Twitter precede timely solicited surveys.

is freely available on [gallup.com](#) [2], and President Obama’s daily job approval rating is freely available on [pollster.com](#) [4].

We correlate the average quarterly happiness of tweets containing the word “Obama” with President Obama’s quarterly job approval rating and find a strong positive correlation (see Appendix Fig. B.3). However, we find the correlation is much stronger in Fig. 4, which gives the happiness time series at a one quarter lag. Similarly, we find a strong positive correlation between the daily approval rating available on Pollster and the daily ambient happiness of “Obama” (see Appendix Fig. B.4a) with an improvement in the correlation when the tweets are lagged by 30 days in Appendix Fig. B.4b. This indicates that real time Twitter data has the potential to predict solicited public opinion polls.

Figure 4 shows that President Obama’s highest approval rating in all three sources was during his first quarter (January–March, 2009). His lowest approval rating was during his 23rd quarter (July–September, 2014). Fig. 5 shows which words contributed most to this shift in ambient happiness. Tweets containing the word “Obama” discuss war and terrorism more often during his 23rd quarter than his first quarter.

C. Index of Consumer Sentiment

Next, we investigate a monthly poll on consumer sentiment designed by the University of Michigan [5]. This poll asks participants five questions about their current and future financial well being and calculates an Index of Consumer Sentiment (ICS) based on responses. In Fig. 6 we correlate this monthly time series with the ambient happiness of the word “job”. We find that the correlation

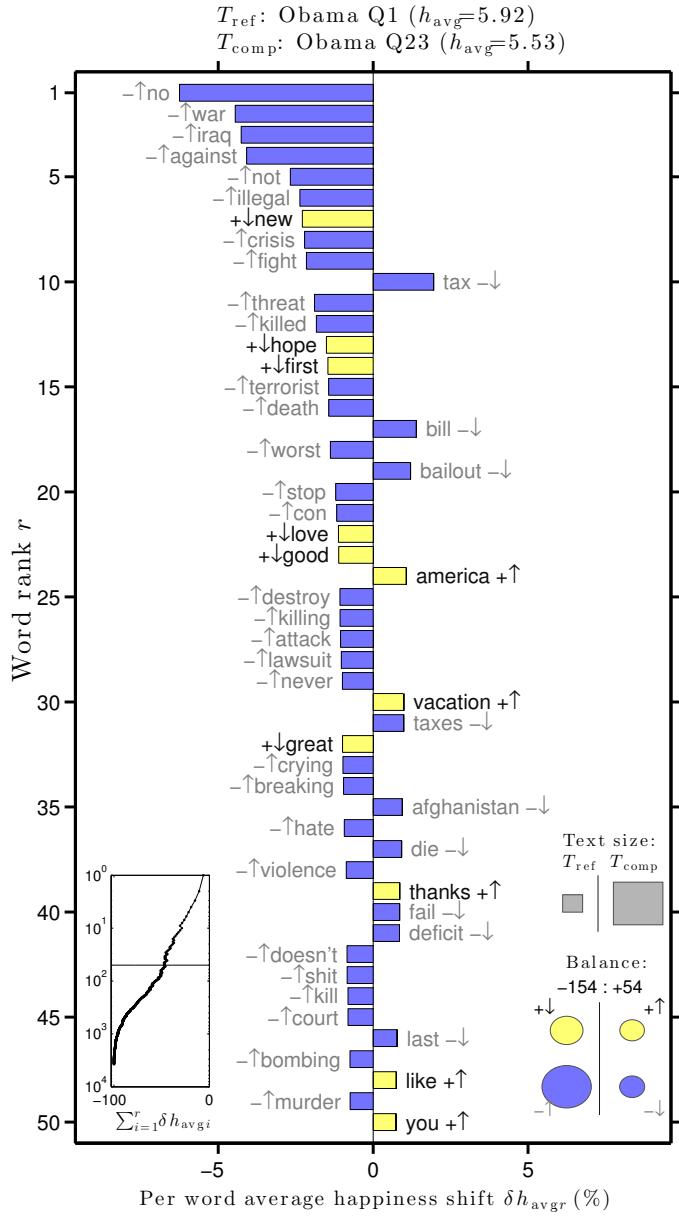


Figure 5. A word shift graph comparing tweets that contain the word “Obama” during the first quarter of his presidency, 2009/01–2009/03, (reference text) and 23rd quarter of his presidency, 2014/07–2015/09, (comparison text). Tweets referred to war and terrorism more often in quarter 1.

is much stronger starting in 2011 (Fig. 6b), and slightly stronger still when the ambient happiness is lagged one month (Fig. 6c). In Fig. 6d we correlate the ICS with the relative frequency of the word “job” on Twitter. We find a strong negative correlation, indicating that it is more likely that a user will tweet the word “job” when they are searching for one.

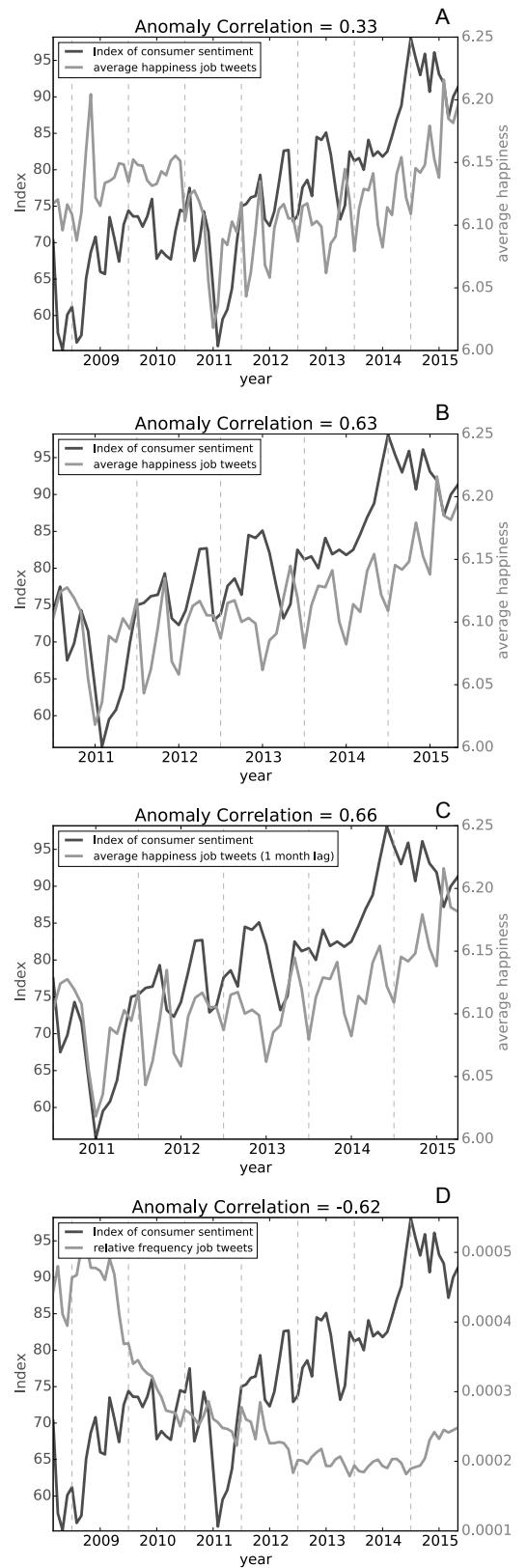


Figure 6. (A) Ambient happiness of “job” with the Index of Consumer Sentiment. We see a small positive correlation getting stronger after 2011. (B) Ambient happiness of “job” with ICS starting in 2011. (C) Ambient happiness of “job” is lagged by one month. (D) ICS with relative frequency of “job”.

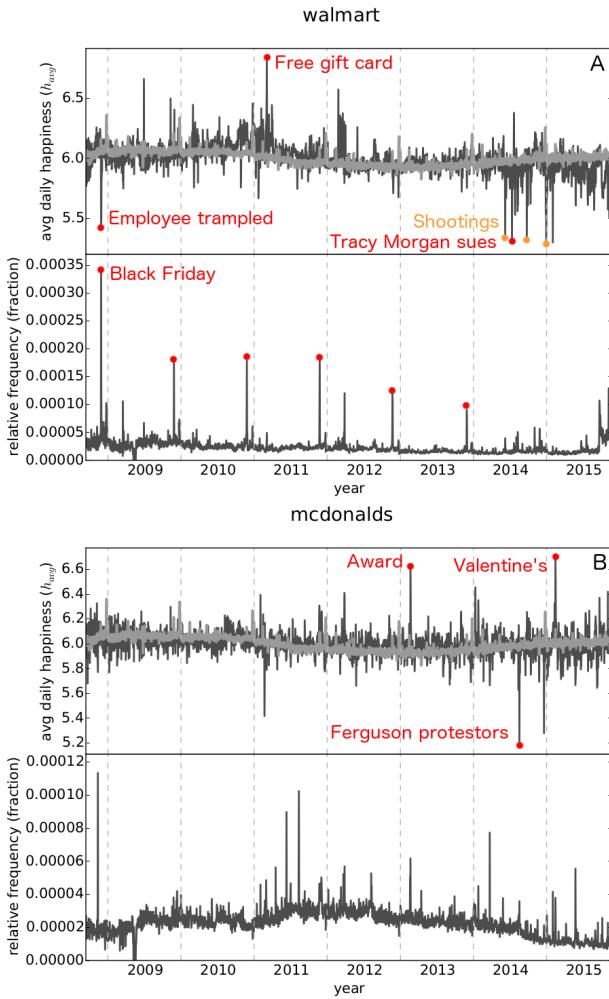


Figure 7. The ambient happiness and relative frequency time series for (A) “walmart” and (B) “mcdonalds”. Dips in sentiment correspond to deaths, lawsuits, and protests, while spikes in happiness correspond to awards, giveaways, and holidays. Spikes in the relative frequency of “walmart” appear largely on Black Friday. Time series for nearly 10,000 other terms can be found on the [online Appendix](#) for the paper.

D. Business Sentiment Shifts

In this section, we investigate the changes in Twitter sentiment surrounding two businesses, Walmart and McDonalds. We examine the ambient happiness time series to determine how sentiment changes in response to events that took place at specific stores. Fig. 7 gives the ambient happiness and relative frequency of the words “walmart” and “mcdonalds”.

Many of the spikes in the “walmart” ambient happiness time series correspond to free giveaways to which Twitter users are responding. A dip in November 2008 corresponds to the trampling to death of a Walmart employee on Black Friday (the day after Thanksgiving, notorious in the U.S. for shopping). Shootings that took place in

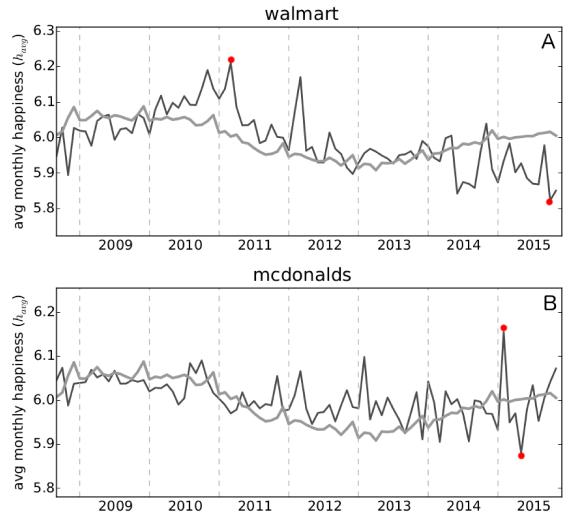


Figure 8. Monthly ambient happiness of (A) “walmart” and (B) “mcdonalds”.

Walmart stores in 2014 are shown with orange dots in Fig. 7a. In June 2014 the Jerad and Amanda Miller Las Vegas shootings ended with 5 deaths (including themselves) in a Nevada Walmart. In September 2014, the police officer who shot John Crawford in an Ohio Walmart was indicted. In December 2014, a 2 year old accidentally shot and killed his mother in an Idaho Walmart. We also see a dip in happiness on the day Tracy Morgan sues Walmart over a fatal crash with one of their tractor trailers in July 2014.

The happiest day in the “mcdonalds” ambient happiness time series is Valentine’s Day in 2015. Upon reading some example tweets from this day, we find that McDonalds was a popular ironic destination for Valentine’s Day dinner that year among Twitter users. A second spike corresponds to a prestigious award given to the McDonalds enterprise in February 2013. McDonalds was given the “Top Toilet Award” for the cleanliness of its restrooms. The saddest day for McDonalds on Twitter was August 18, 2014, the day that Ferguson protesters broke into a McDonalds to steal milk to relieve tear gas victims.

In Fig. 8 we explore the monthly ambient happiness of “walmart” and “mcdonalds”. We find that the ambient happiness of “walmart” reaches its maximum in March 2011, and its minimum in October 2015, and the ambient happiness of “mcdonalds” reaches its maximum in February 2015 and its minimum shortly after in May 2015. To investigate the texture behind these observations, we use word shift graphs to compare the happiest and saddest months for each business in Fig. 9.

In November 2015 (comparison text Fig. 9a), there were Black Friday altercations at many Walmarts throughout the country, often caught on camera, leading to an increase in negative words such as “caught”, “fail”, “ridiculous”, and “captured”. Twitter users were

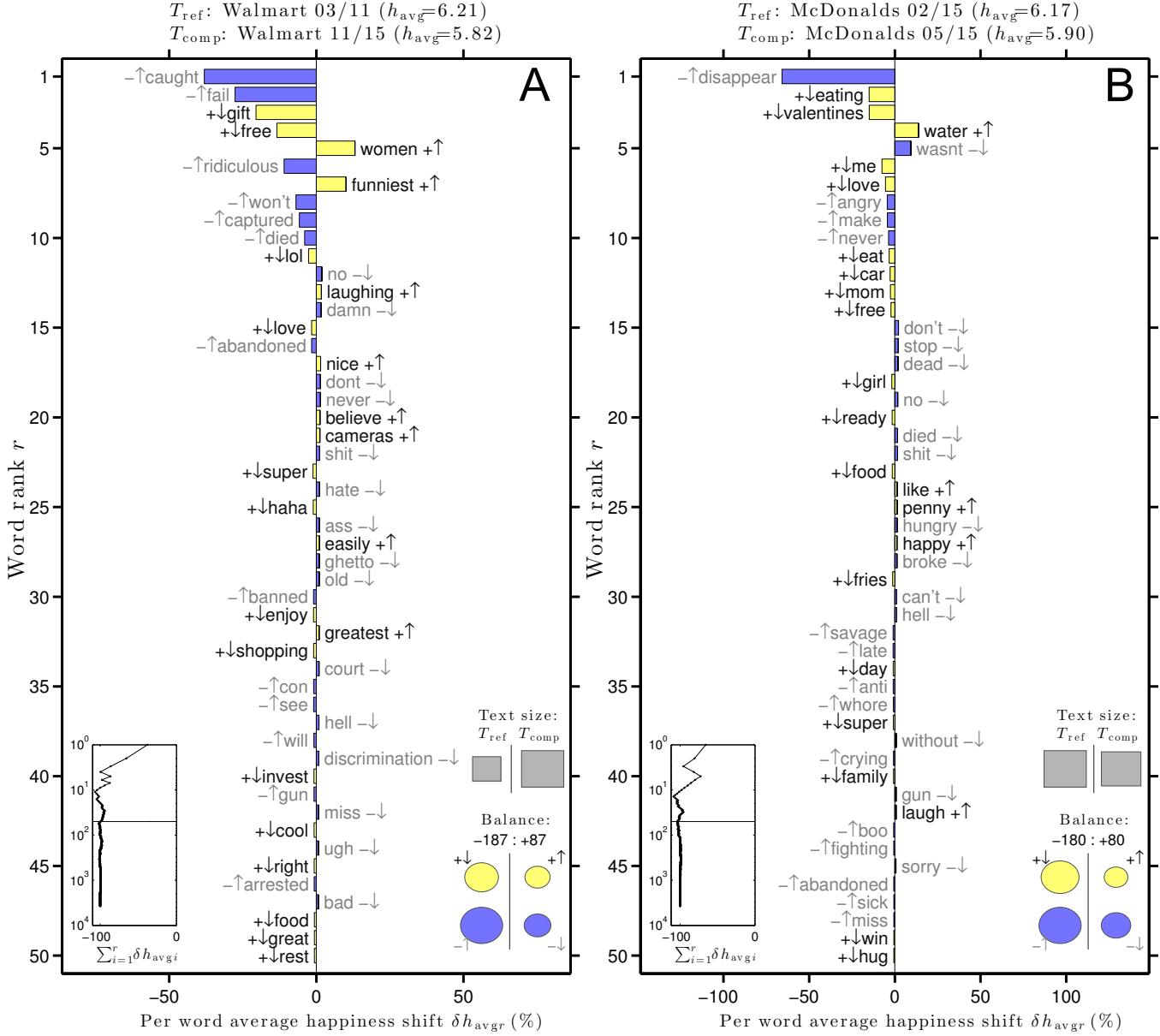


Figure 9. Word shift graphs comparing the happiest and saddest months for (A) “walmart” and (B) “mcdonalds”. The happiest month represents the reference text and the saddest month represents the comparison text.

happier about Walmart in March 2011 (reference text Fig. 9a) due in part to a free gift card giveaway. Happier tweets included the words “lol”, “love”, “haha”, and “super”. Surprisingly, we actually see more curse words during the happiest month than the saddest month.

The happiest month for McDonalds was February 2015 (reference text Fig. 9b) when a surprising number of Twitter users were spending Valentine’s Day there, hence the decrease in the words “valentines” and “love”. The decrease in happiness in May 2015 is in large part in an increase in the word “disappear”. During this time, a video of a Michigan McDonalds employee performing a practical joke, in which he claims he’s going to make a penny disappear in a bottle of water, went viral. Using word shifts, we are thus able to determine “disappear” was not a true indicator for negativity. The next step would be to add “disappear” to our stop word list and reevaluate the time series. In general, word shifts are very sensitive diagnostics that allow us to make sense of apparent sentiment patterns, and to adjust the hedonometer as needed.

IV. LIMITATIONS

Here we present positive correlations between ambient happiness on Twitter and solicited public opinion surveys. The methods used in this work present several limitations that we address here. First, we are unable to compare ambient happiness to public opinion surveys for the majority of subjects discussed on Twitter. Ambient happiness surrounding commercial businesses can be compared to stock prices, however the relationship is unlikely to be linear. Stock prices take into account much more than the success of one business, and thus reflect a different signal than public opinion.

We acknowledge that our dataset can be influenced by small events that are retweeted many times, and cause a transient spike or a dip in ambient happiness. Tweets reflect a broad spectrum of conversation and reactions to major events, including bots and sarcasm to name a few difficulties. The stochastic nature of Twitter timeseries data also makes it difficult to compare high resolution ambient happiness to low resolution opinion surveys. Appendix Fig. C.1 shows our attempt to compare ambient happiness to six important topics available on Gallup. Gallup surveys take place over several days, once a year, leading to a natural comparison with ambient happiness averaged for a full 365 days. We are therefore unable to report which method is a more accurate indicator of public opinion. Instead, we conclude that public opinion polling with Twitter has the potential to complement traditional public opinion surveys.

Finally, without knowing user demographics in detail, we are unable to adjust for the known sample bias of Twitter users relative to the general population [1]. Our conclusions pertain directly only to the Twitterverse and as suggestive results for the broader public. Twitter users

are a mixture of individuals, news outlets, corporations, and, perhaps most problematically, are not even always human. Previous work shows that Twitter contains many bots, which send tweets automatically, often to advertise a product or a political campaign [25]. We do not eliminate these tweets in this work, however many methods for uncovering them have been suggested [11, 13, 15, 19].

V. CONCLUSION

The objective of this research was to determine the extent to which ambient happiness on Twitter can be used as a reliable public opinion polling device. A central motivation is that solicited public opinion polling data is difficult to obtain at a high temporal resolution, if at all [33].

With data from Twitter we can investigate topics other than political or global issues, which are the focus of a large majority of solicited surveys. We can use ambient happiness to determine how people feel about seemingly neutral topics like “snow”, or how they are using the words “love”, and “feel”. We also have shown that Twitter users respond to various kinds of events taking place at commercial businesses, and thus ambient happiness could be used in market analysis to predict or improve a company’s sales.

Of the available polling data we were able to obtain, we find that ambient happiness of selected words correlates well with solicited public opinions. In some cases, the correlation increases when the tweets are lagged, indicating that real time Twitter data has the potential to predict solicited public opinion polls.

Not only can tweets anticipate survey responses, but we can use individual words within tweets to determine why one time period is happier than another, something that is not possible in traditional polls due to the multiple choice aspect of most surveys. Several other advantages of using tweets for public opinion polling include the ability to track movement [20] and make maps [28, 34] using geolocated tweets. Data collection itself is largely algorithmic, and does not rely on the responses of participants.

We find that for many topics, Twitter is a valuable resource for mining public opinions without solicited surveys. We encourage readers to explore the data online at <http://compstorylab.org/share/papers/cody2016a/data.html>. Social media may be the future of public opinion polling, revealing important signals complementary to traditional surveys.

ACKNOWLEDGMENTS

The authors are grateful for the computational resources provided by the Vermont Advanced Computing Core and the Vermont Complex Systems Center. PSD and CMD acknowledge support from NSF grant

#1447634. EMC acknowledges support from the Na-

tional Science Foundation under project DGE-1144388.

-
- [1] The demographics of social media users. <http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/>. Accessed: 2016-07-18.
- [2] Gallup trends. <http://www.gallup.com/poll/trends.aspx>. Accessed: 2016-03-08.
- [3] Google correlate. <https://www.google.com/trends/correlate>.
- [4] Pollster API. <http://elections.huffingtonpost.com/pollster/api>. Accessed: 2016-03-08.
- [5] University of Michigan Index of Consumer Sentiment. <http://www.sca.isr.umich.edu/tables.html>. Accessed: 2016-03-08.
- [6] D. Antenucci, M. R. Andwerson, P. Zhao, and M. Cafaella. A query system for social media signals. 2015.
- [7] E. Aramaki, S. Maskawa, and M. Morita. Twitter catches the flu: Detecting influenza epidemics using Twitter. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1568–1576. Association for Computational Linguistics, 2011.
- [8] P. Barberá. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data.
- [9] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
- [10] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2649–2658, 2012.
- [11] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Detecting automation of Twitter accounts: Are you a human, bot, or cyborg? *Dependable and Secure Computing, IEEE Transactions on*, 9(6):811–824, 2012.
- [12] R. B. Cialdini and N. Garde. *Influence*, volume 3. A. Michel, 1987.
- [13] E. M. Clark, J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds. Sifting robotic from organic text: A natural language approach for detecting automation on Twitter. *Journal of Computational Science*, 2015.
- [14] E. M. Cody, A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth. Climate change sentiment on Twitter: An unsolicited public opinion poll. *PLoS ONE*, 10(8):e0136092, 2015.
- [15] J. P. Dickerson, V. Kagan, and V. Subrahmanian. Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*, pages 620–627. IEEE, 2014.
- [16] P. U. Diehl, B. U. Pedroni, A. Cassidy, P. Merolla, E. Neftci, and G. Zarrella. TrueHappiness: Neuromorphic emotion recognition on TrueNorth. *arXiv preprint arXiv:1601.04183*, 2016.
- [17] J. DiGrazia, K. McKelvey, J. Bollen, and F. Rojas. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS ONE*, 8(11):e79449, 2013.
- [18] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one*, 6(12):e26752, 2011.
- [19] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *arXiv preprint arXiv:1407.5225*, 2014.
- [20] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth. Happiness and the patterns of life: A study of geolocated tweets. *Scientific reports*, 3, 2013.
- [21] D. Gayo-Avello. “i wanted to predict elections with twitter and all i got was this lousy paper”—a balanced survey on election prediction using twitter data. *arXiv preprint arXiv:1204.6441*, 2012.
- [22] D. Gayo-Avello. A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, page 0894439313493979, 2013.
- [23] M. Glenski and T. Weninger. Rating effects on social news posts and comments. *arXiv preprint arXiv:1606.06140*, 2016.
- [24] B. Helmuth, T. C. Gouhier, S. Scyphers, and J. Mocarski. Trust, tribalism and tweets: Has political polarization made science a “wedge issue”? *Climate Change Responses*, 3(1):1, 2016.
- [25] P. N. Howard and B. Kollanyi. Bots, # Strongerin, and # Brexit: Computational propaganda during the UK-EU referendum. *Available at SSRN 2798311*, 2016.
- [26] S. Kamvar and J. Harris. *We feel fine: An almanac of human emotion*. Simon and Schuster, 2009.
- [27] I. M. Kloumann, C. M. Danforth, K. D. Harris, C. A. Bliss, and P. S. Dodds. Positivity of the English language. *PLoS ONE*, 7(1):e29484, 2012.
- [28] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3):e1500779, 2016.
- [29] D. Lazer, R. Kennedy, G. King, and A. Veselnitskaya. The parable of Google Flu: Traps in Big Data analysis. *Science Magazine*, 343:1203–1205, 2014.
- [30] Y.-R. Lin, B. Keegan, D. Margolin, and D. Lazer. Rising tides or rising stars?: Dynamics of shared attention on Twitter during media events. *PloS one*, 9(5):e94093, 2014.
- [31] Y.-R. Lin, D. Margolin, B. Keegan, and D. Lazer. Voices of victory: A computational focus group framework for tracking opinion shift in real time. In *Proceedings of the 22nd international conference on World Wide Web*, pages 737–748. International World Wide Web Conferences Steering Committee, 2013.
- [32] Y. Mejova, I. Weber, and M. W. Macy. *Twitter: A digital socioscope*. Cambridge University Press, 2015.
- [33] G. Miller. Social scientists wade into the tweet stream. *Science*, 333(6051):1814–1815, 2011.
- [34] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth. The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE*,

- 8(5):e64417, 2013.
- [35] A. Mittal and A. Goel. Stock prediction using Twitter sentiment analysis. *Stanford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, 2012.
- [36] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2, 2010.
- [37] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 1320–1326, 2010.
- [38] M. J. Paul and M. Dredze. You are what you tweet: Analyzing Twitter for public health. *ICWSM*, 20:265–272, 2011.
- [39] T. Preis, H. S. Moat, and H. E. Stanley. Quantifying trading behavior in financial markets using Google Trends. *Scientific reports*, 3, 2013.
- [40] A. Reagan, B. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds. Benchmarking sentiment analysis methods for large-scale texts: A case for using continuum-scored words and word shift graphs, 2015.
- [41] J. Ritterman, M. Osborne, and E. Klein. Using prediction markets and Twitter to predict a swine flu pandemic. In *1st international workshop on mining social media*, volume 9, pages 9–17. ac. uk/miles/papers/swine09. pdf (accessed 26 August 2015), 2009.
- [42] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
- [43] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.
- [44] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185, 2010.
- [45] C. Vaccari, A. Valeriani, P. Barberá, R. Bonneau, J. T. Jost, J. Nagler, and J. Tucker. Social media and political communication: A survey of Twitter users during the 2013 Italian general election. *Rivista italiana di scienza politica*, 43(3):381–410, 2013.
- [46] X. Wang, M. S. Gerber, and D. E. Brown. Automatic crime prediction using events extracted from Twitter posts. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 231–238. Springer, 2012.

Appendix A: Anomaly Correlation

We use anomaly correlation (Pearson Correlation) to determine the relationship between the Twitter happiness time series and the polling data. When doing so, we subtract the mean of the series, m , from each data point, h_i , to determine anomalies, and then calculate the cosine of the angle between the series of anomalies, i.e.,

$$H_{an} = \{h_i - m\}_{i=1}^L \quad (\text{A1})$$

$$\text{Corr}_{an}(H, P) = \frac{H_{an} \cdot P_{an}}{\|H_{an}\| \cdot \|P_{an}\|} \quad (\text{A2})$$

The variables H and P represent happiness time series and polling time series respectively, and L is the length of the time series.

Appendix B: Additional Figures and Tables

Each word in our data set was previously assigned a happiness score through Amazon's Mechanical Turk (labMT scores). We investigate the relationship between surveyed scores and ambient happiness scores in Fig. B.1. We find a positive slope, indicating that ambient happiness rises with surveyed happiness, however we see a much smaller range of scores, which can be attributed to averaging a large amount of words. We give the top 10 and bottom 10 words sorted by ambient happiness in Table B.1. Top words included birthday wishes and prize giveaways, and bottom words suggest legal news stories.

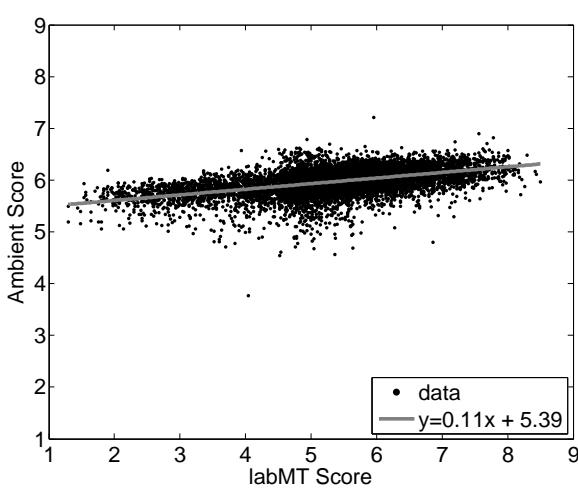


Figure B.1. Surveyed happiness versus ambient happiness for all words in the labMT dataset. The small positive slope indicates that ambient happiness increases with surveyed happiness, however ambient happiness covers a smaller range of values. An interactive version is available in the online Appendix.

Rank	Top 10				Bottom 10			
	Word	Ambient	labMT	Rank	Word	Ambient	labMT	
1.	collected	7.21	5.96	9780.	defendants	4.87	4.26	
2.	merry	6.90	7.56	9781.	prosecutors	4.87	4.20	
3.	birthday	6.82	7.78	9782.	suspects	4.86	3.60	
4.	iya	6.79	4.94	9783.	suspected	4.81	3.52	
5.	prizes	6.73	7.20	9784.	indicted	4.81	3.60	
6.	b-day	6.71	7.68	9785.	seas	4.80	6.84	
7.	2-bath	6.69	5.28	9786.	pleaded	4.78	3.84	
8.	entered	6.65	5.82	9787.	sentenced	4.71	3.70	
9.	giveaway	6.62	6.38	9788.	civilians	4.68	5.84	
10.	shipping	6.61	5.46	9789.	welt	3.77	4.04	

Table B.1. The top 10 and bottom 10 words sorted by ambient happiness. Ambient happiness is calculated using word frequencies from September 2008 through November 2015. Non-English words and words with frequencies under 1000 are removed, leaving 9789 remaining in our ambient dataset.

Rank	Top 10				Bottom 10			
	Word	Ambient	labMT	Rank	Word	Ambient	labMT	
1.	birthday	6.82	7.78	9780.	seas	4.80	6.84	
2.	b-day	6.71	7.68	9781.	civilians	4.86	5.84	
3.	merry	6.90	7.56	9782.	defendants	4.87	4.26	
4.	prizes	6.73	7.20	9783.	prosecutors	4.87	4.20	
5.	giveaway	6.62	6.38	9784.	welt	3.77	4.04	
6.	collected	7.21	5.96	9785.	pleaded	4.78	3.84	
7.	entered	6.65	5.82	9786.	sentenced	4.71	3.70	
8.	shipping	6.61	5.46	9787.	indicted	4.81	3.60	
9.	2-bath	6.69	5.28	9788.	suspects	4.86	3.60	
10.	iya	6.79	4.94	9789.	suspected	4.81	3.52	

Table B.2. The top 10 and bottom 10 words according to ambient happiness, sorted by labMT score.

Rank	Top 10				Bottom 10			
	Word	Ambient	labMT	Rank	Word	Ambient	labMT	
1.	laughter	5.96	8.50	9780.	died	5.76	1.56	
2.	happiness	6.11	8.44	9781.	kill	5.71	1.56	
3.	love	6.17	8.42	9782.	killed	5.56	1.56	
4.	happy	6.48	8.30	9783.	cancer	5.93	1.54	
5.	laughed	5.87	8.26	9784.	death	5.66	1.54	
6.	laugh	6.01	8.22	9785.	murder	5.39	1.48	
7.	laughing	5.71	8.20	9786.	terrorism	5.16	1.48	
8.	excellent	6.31	8.18	9787.	rape	5.46	1.44	
9.	laughs	6.06	8.18	9788.	suicide	5.49	1.30	
10.	joy	6.19	8.16	9789.	terrorist	5.19	1.30	

Table B.3. The top 10 and bottom 10 words according to labMT score.

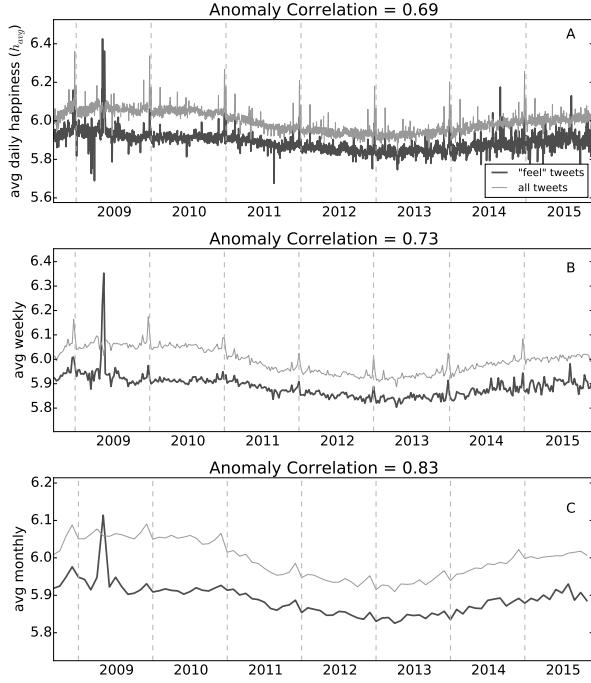


Figure B.2. Ambient happiness of “feel” compared to overall happiness by (A) day, (B) week, and (C) month (motivated by the work of Kamvar and Harris [26]). The ambient happiness of the word “feel” correlates strongly with the average happiness of tweets that do not contain “feel”, and the correlation grows stronger as we decrease the temporal resolution. This indicates that the shape of overall happiness remains the same whether a user is directly or indirectly expressing an emotion on Twitter. An interactive version of the overall signal can be found at hedonometer.org.

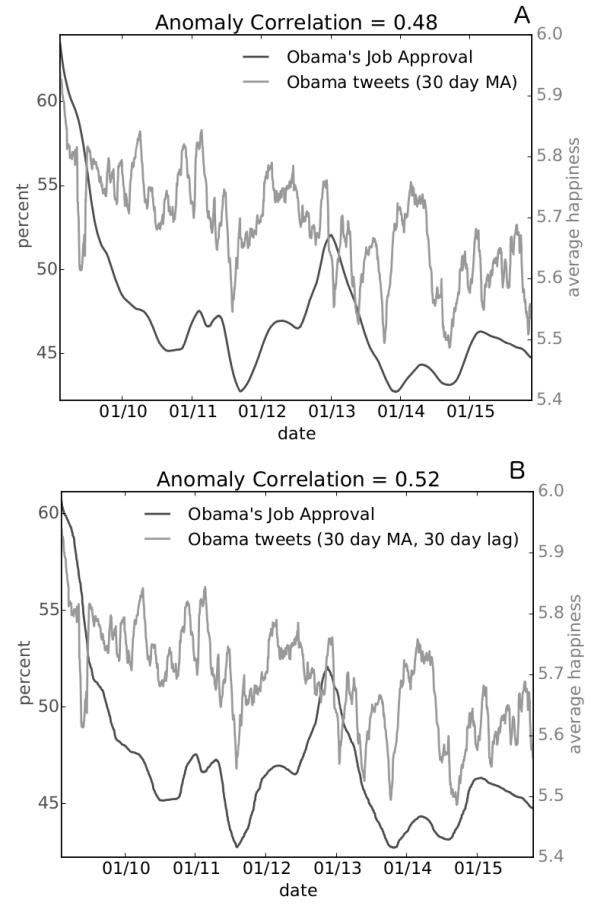


Figure B.4. (A) Average daily happiness of tweets containing “Obama” with Obama’s daily job approval rating from Pollster. (B) 30 day lag. We find a relatively high correlation with solicited polling data.

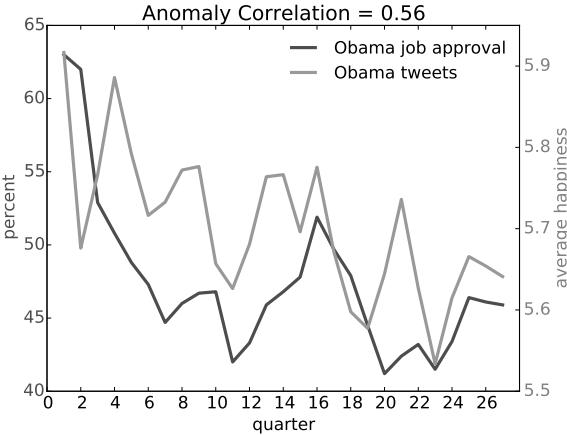


Figure B.3. Average quarterly happiness of tweets containing “Obama” with Obama’s quarterly job approval rating from Gallup. We find a relatively high correlation with solicited polling data.

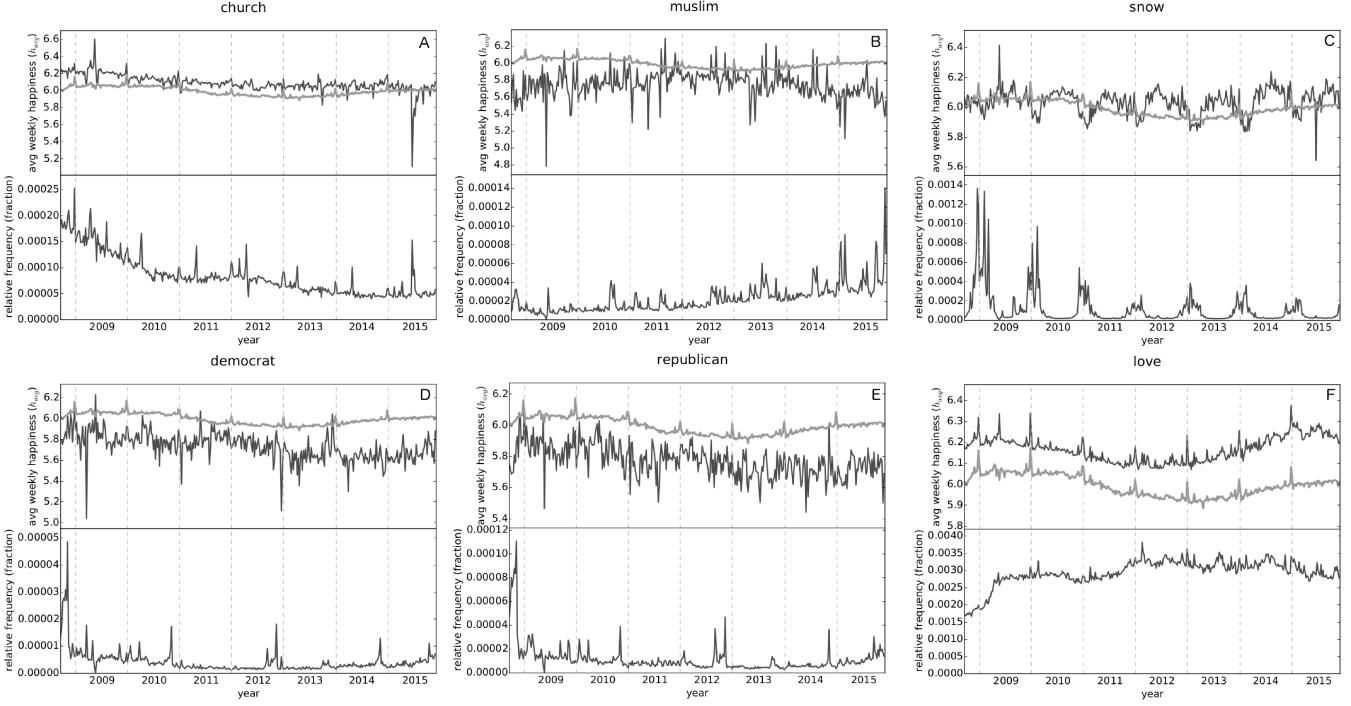


Figure B.5. Six examples of weekly ambient happiness time series (top) with the weekly relative frequency for the word (bottom). Relative frequency is calculated by dividing the total frequency of the word by the total frequency of all words on a given week. (A) “church” (B) “mulsim” (C) “snow” (D) “democrat” (E) “republican” (F) “love”

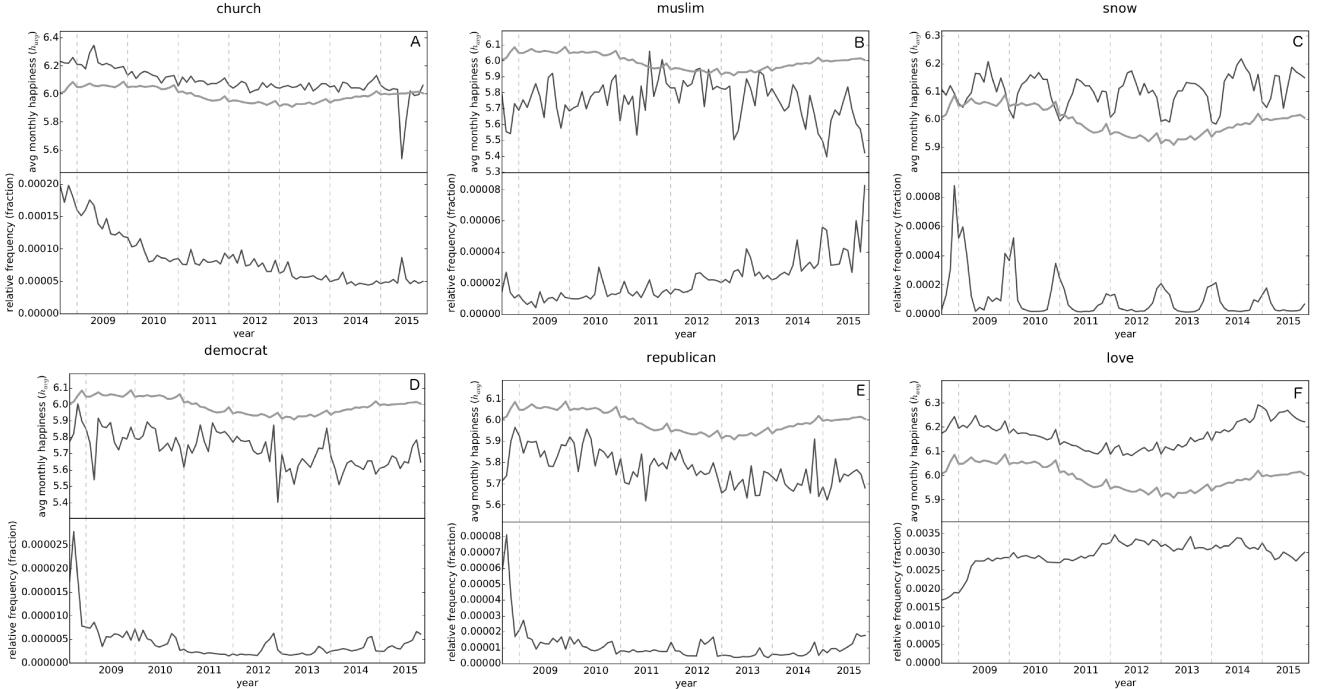


Figure B.6. Six examples of monthly ambient happiness time series (top) with the monthly relative frequency for the word (bottom). Relative frequency is calculated by dividing the total frequency of the word by the total frequency of all words on a given month. (A) “church” (B) “mulsim” (C) “snow” (D) “democrat” (E) “republican” (F) “love”

Appendix C: Gallup Yearly Polling

Gallup trends provide yearly polling data on many topics without a subscription. The Gallup survey questions can be found in Table C.1. These polls, however, take place only once a year in the same month over several days. This presents a challenge as to the amount of Twitter data we should include in our correlations, as opinions may change. For each Gallup datapoint, we use the current year's worth of tweets from 2009 through 2015 for various subjects of national or global interest. Fig. C.1 shows several topics that correlate quite well with ambient happiness on Twitter. We find that the favorability of two major countries, Iran and Iraq, has a positive correlation with the ambient happiness of "Iran" and "Iraq". We also find that the United States opinion on religion has a strong positive correlation with yearly ambient happiness of "religion". Other topics, including the United States opinion on Afghanistan and immigration show no significant correlation to Twitter data. There is a strong negative correlation between the satisfaction of the position of the United States in the world, indicating there may be some sarcasm associated with "usa" tweets.

Topic	Survey Question	Frequency	Source
Iraq	What is your overall opinion of Iraq? Is it very favorable, mostly unfavorable, mostly unfavorable, or very unfavorable?	Yearly	Gallup
Iran	What is your overall opinion of Iran? Is it very favorable, mostly unfavorable, mostly unfavorable, or very unfavorable?	Yearly	Gallup
Afghanistan	What is your overall opinion of Afghanistan? Is it very favorable, mostly unfavorable, mostly unfavorable, or very unfavorable?	Yearly	Gallup
USA	On the whole, would you say you are satisfied or dissatisfied with the position of the United States in the world today?	Yearly	Gallup
Religion	Please tell me how much confidence you , yourself, have in the church or organized religion – a great deal, quite a lot, some, or very little?	Yearly	Gallup
Immigration	On the whole, do you think immigration is a good thing or a bad thing for this country today?	Yearly	Gallup
Obama	Do you approve of disapprove of the way Barak Obama is handling his job as president?	Quarterly	Gallup
Obama	Average of latest opinion polls on Obama's job approval	Daily	Pollster

Table C.1. Survey questions for polling data from various resources used in our analysis.

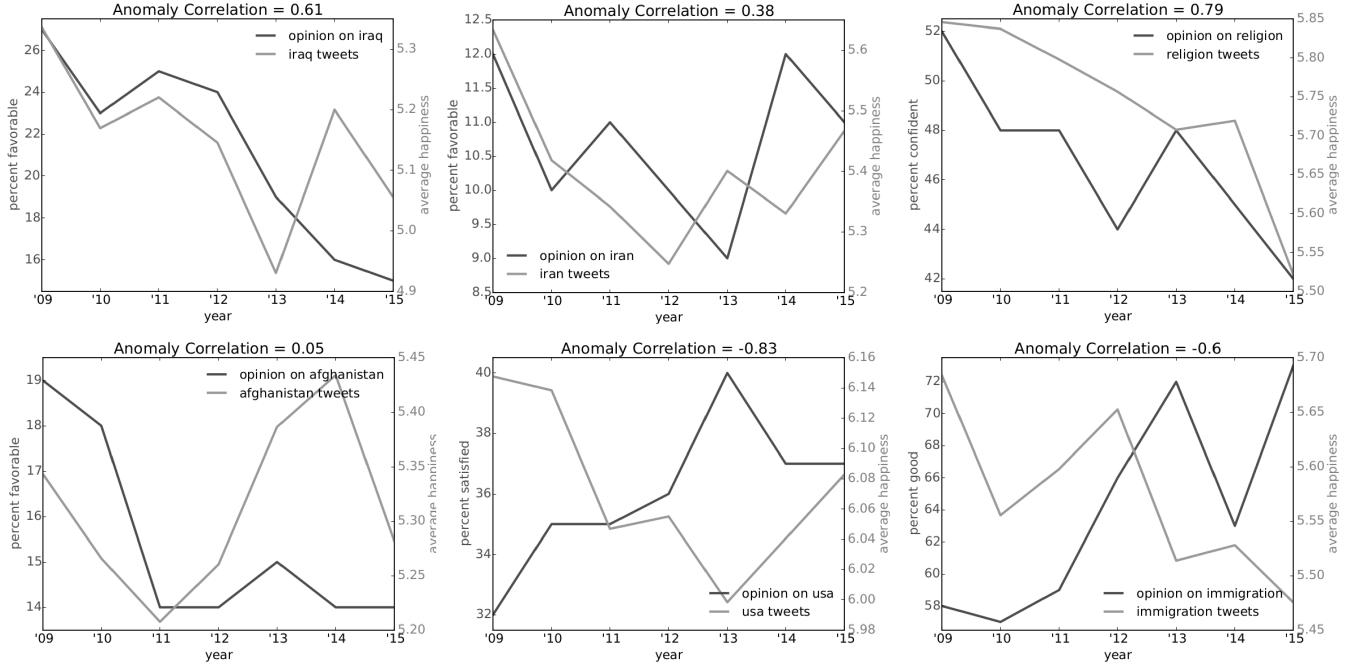


Figure C.1. Correlations between average ambient happiness and opinion polls on various global subjects. We obtain varying levels of correlation between the topics due the limited availability of traditional polling data. For example, Twitter sentiment tracks public opinion surrounding Iraq and religion quite well, but performs poorly on Afghanistan. The specific questions can be found in Table C.1.