A Comparative Study of Social Media and Traditional Polling in the Egyptian Uprising of 2011

Lora Weiss[1], Erica Briscoe[1], Heather Hayes[1], Olga Kemenova[1], Sim Harbert[1], Fuxin Li[1], Guy Lebanon[1], Chris Stewart[2], Darby Miller Steiger[2], and Dan Foy[2]

[1] The Georgia Institute of Technology, Atlanta, Georgia
{lora.weiss,erica.briscoe,heather.hayes,olga.kemenova,
sim.harbert}@gtri.gatech.edu,
{fli,lebanon}@cc.gatech.edu
[2] The Gallup Organization, Washington D.C.
{chris_stewart,darby_miller_steiger,dan_foy}@gallup.com

Abstract. Because social network sites such as Twitter are increasingly being used to express opinions and attitudes, the utility of using these sites as legitimate and immediate information sources is of growing interest. This research examines how well information derived from social media aligns with that from more traditional polling methods. Specifically, this research examines tweets from over 40,000 Egyptian users from both before and after the Egyptian uprising on January 25, 2011 and compares that information with polling data collected by The Gallup Organization during the same time period. This analysis ascertains trends in sentiment and identifies the extent to which these methodologies align over time. The results show that trends across the two sources are not consistent. Focusing solely on Twitter data, individuals expressed increasingly negative opinions after the uprising, whereas survey results indicated that individuals were increasingly positive post-uprising. We discuss the implications of these differences for the use of social media as a real-time information source.

Keywords: Social Media, Twitter, Sentiment analysis, Polling, Arab Spring.

1 Introduction

Increasingly, social networking sites have become an important tool for social interaction and information sharing. Use of social networks has increased exponentially in the past few years. Facebook went from 100 million users in 2008 to over 1 billion users in 2012 [13]. Twitter has also seen an exponential increase in users since its inception, where there are more than 500 million current users, up from 100 million in 2011 and 5.5 million in 2005 [7]. These users generate over 230 million tweets per day, which is an increase of 110% since 2011 [7].

Not only are social network sites used for self-expression, seeking and sharing of information, and social interaction, but these outlets also contribute to revolutionary movements [17]. Mainstream news and communication media are often restricted or

A.M. Greenberg, W.G. Kennedy, and N.D. Bos (Eds.): SBP 2013, LNCS 7812, pp. 303–310, 2013.

blocked by governments, so that social network sites have become a prominent means of quickly disseminating information. Twitter is often associated with opinion, attitude, and rapid information sharing as compared to the more socially-interactive, group-oriented Facebook [12, 14]. Twitter is also used more often to track events that change quickly, such as disasters, accidents, and riots [8]. Twitter became increasingly popular in Egypt and other Arab countries during the Arab Spring. There was a significant increase in internet use for Egypt; namely, a 39.61% increase from 2010 to 2011 [2]. The greatest increase in users was during the first three months of 2011, where Egypt was in the top five Arab countries in number of Twitter users and tweets [3], with an increase of over one million tweets between the week of January 16, 2011 (122,319 tweets) to the week of January 24, 2011 (1.3 million tweets) [18], where the date associated with the Egyptian uprising is canonically January 25, 2011. Thus, Twitter effectively becomes a new paradigm for journalism, as evidenced by the increase of its use during the government-controlled blackout in Egypt after the 2011 uprising [3], [17].

The goal of the research reported in this paper is to ascertain how individuals' sentiment towards various topics in Egypt varied during the revolution and how this variance manifests in both Twitter and Gallup polling data. Polling data was obtained by The Gallup Organization by conducting major surveys in Egypt in 2010 and 2011 involving face-to-face interviews with approximately 1000 adults per survey [19, 20]. This paper describes research that examined a subset of many tweets from over 40,000 Egyptian users both before and after the Egyptian uprising on January 25, 2011 and compares that information with polling data collected by The Gallup Organization in Egypt during the same time period.

2 Methodology

A traditional method of assessing sentiment, opinions, and attitudes of a study population is through the use of administered surveys, which is one of the most common approaches to measuring a wide variety of constructs. Such methodologies are maximized through reliability and validity testing [15]. They can also serve as useful comparisons with newer, more open-ended, and less constricted analyses of data derived from open-source mediums such as social network sites. It would be ideal if the two methodologies yielded consistent results, thereby making them parallel measures of the same opinions or attitudes. Congruence would suggest that open-source data may serve as an augmentation or, when traditional methods are infeasible, a replacement of survey-collected information. In the current study, we consider the consistency that arises between the dynamic reflection of attitudes toward various topics, such as the economy or the government, as measured through Twitter data and Gallup polling.

Data was collected from two separate sample populations with possible, though unlikely, overlap. The first was Gallup poll data that surveyed Egyptians; the second was Egyptian Twitter users. Both populations are described below. Analysis of these populations included pre- and post-uprising attitudes, with the differentiation based on the date of the Egyptian uprising of January 25, 2011.

2.1 Survey Poll Data

Gallup conducted several nationally representative surveys of the Egyptian population during the target timeframe. Sampling involved multiple layers of stratification based on population size, geography, and randomized household and respondent selection procedures. This data was collected through face-to-face interviews in Arabic by native speakers and was weighted by key demographic variables including: household size, age, gender, education, and socioeconomic status to correct for disproportionalities across these variables. Surveys were conducted in October 2010 with 1,011 respondents ages 15+ and in April 2011 with 1,005 respondents. These dates correspond to periods of pre- and post-Egyptian uprising, centered about January 25, 2011.

From the results of a larger, comprehensive set of questions, respondents' opinions about four primary areas were obtained (see Table 1).

Table 1. Gallup poll questions categorized into indicator areas

Statement	Response Options
Positive Emotive Indicators	
You will invent or discover something that will change the world.	Agree, Disagree, Don't know
You never give up until you reach your goals, no matter what.	Agree, Disagree, Don't know
Even when things go wrong, you feel very optimistic.	Agree, Disagree, Don't know
Do you have relatives or friends who are living in another country whom you can count on to help you when you need them, or not?	Agree, Disagree, Don't know
Now, please think about yesterday, from the morning until the end of the day. Think about where you were, what you were doing, who you were with, and how you felt. Were you treated with respect all day yesterday?	Agree, Disagree, Don't know
Negative Emotive Indicators	
Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Enjoyment?	Yes, No, Don't know
Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Happiness?	Yes, No, Don't know
Economic Indicators	
Right now, do you feel your standard of living is getting better or getting worse?	Getting Better, The Same, Getting Worse, Don't know
Right now, do you think the economic conditions in Egypt, as a whole, are getting better or getting worse?	Getting Better, The Same, Getting Worse, Don't know
Institutional Indicators	
In Egypt, do you have confidence in each of the following, or not? How about Financial Institutions or banks?	Yes, No, Don't know
In Egypt, do you have confidence in each of the following, or not? How about Honesty of elections?	Yes, No, Don't know

2.2 Twitter Data

Twitter collection dates were September 9, 2010 to October 12, 2010 for the pre-uprising stage and March 8, 2011 to April 9, 2011 for the post-uprising stage. Initial analysis included tweets translated from Arabic, but validation testing of the machine translation of Arabic tweets by native Arabic speakers found these methods unreliable; therefore, we concentrated only on English tweets. Analysis of Twitter data required several steps to ensure samples were representative of Egyptian users. First, we obtained 6.5 million tweets based on relevant hash tags and keywords from the Twitter message archiving site TwapperKeeper.com. Tweets that included the hash tags "#jan25" or "#egypt" and any that included the word "Egypt" were kept. Then we filtered out those in Arabic, leaving 2.14 million English tweets. Next we filtered on those users who reported their location as somewhere in Egypt, which resulted in a pool of approximately 820,000 tweets from 48,077 Egyptian users. Next, categories of key words relevant to the focus areas of the Gallup surveys were created. For example, for job satisfaction, relevant key words included: chore, employment, occupation, put to work, busy, subcontract, farm out, job, etc.

This process of filtering and targeting relevant tweets resulted in a reduced number of tweets per topic: 509 tweets for job satisfaction, 661 tweets for confidence in military, 80 tweets for confidence in judicial system, 596 tweets for confidence in government, and 148 tweets for confidence in the honesty of elections. We then ran a sentiment analysis using a regressor trained on 146,291 LiveJournal blogs labeled with 32 different emoticons, where multi-dimensional scaling was performed on the mean document conditioned on each emoticon, in order to automatically retrieve a latent emotion manifold that accurately reflected human perceptions of the emotion space (see [21] for a more detailed description of this procedure). For this work, we concentrate on the mapping of the tweets to the first dimension of the manifold, which depicts the sentiment, ranging from depressed/sad to happy/excited. This analysis assigned a sentiment score on a scale from -1 (least positive) to 1 (most positive) to each tweet. A sample of the sentiment ratings, concentrating on those at the more extreme ends of the scale, was validated using human coders.

3 Analysis and Results

Before the analysis, poll question responses were selected for their appropriateness in representing each indicator area of sentiment (see Table 1). This step involved dichotomizing item responses as 1 for negative responses and 2 for positive responses (e.g., saying "yes" to a question in the "positive emotive indicator" section). Negative emotion scores were reverse scored such that higher scores represented lower negativity. The following two indices were investigated (i) the average (tetra-) correlations among items in each sentiment area (i.e., coefficient alpha) and (ii) each item's correlation with the total score for each dimension, as described above. In both cases, correlations were in the range of 0.60 - 0.80, which is acceptable for assuming that composite scores taken across individual item scores for each area of sentiment are

appropriate for use in further analysis. A general sentiment score was also computed based on the sum of the specific composite scores.

For the Gallup analysis, there was inconsistency in the number of respondents retained in pre- versus post-test. Therefore, the data was "paired" by randomly sampling from the larger sample size (pre versus post). As a result, the sample sizes are lower than that originally attained. The mean of each composite score and the overall composite score were compared pre- and post-uprising, where the higher score meant a more positive sentiment. This analysis consisted of paired sample t-tests, in which the p-value was corrected for family-wise error ($p < 0.05/5 = 0.01$). Paired sample t-tests were used because the same set of individuals were used at two time points, resulting in the need to compare pre- and post- means via repeated measure analysis. Table 2 provides descriptive statistics. Table 3 provides t-test results. Table 3 shows an increase in positive sentiment for all areas as well as an overall increase. Based on the paired t-tests, all of these changes were statistically significant except for negativity (Table 3).

For the Twitter analysis, a similar event occurred wherein data was paired by random selection of larger sample size for pre versus post, resulting in a lower overall sample size. Sentiment for each area as well as overall sentiment was compared pre- and post-uprising similarly to that performed on the Gallup data. Based on the descriptive statistics provided in Table 4, it is apparent that respondents were generally more negative toward all topics except for job satisfaction. Moreover, sentimentality toward all areas, as well as overall sentiment, decreased from pre- to post-uprising. Based on the paired sample t-tests (Table 5), only one of these changes was significant: overall sentimentality. Thus, participants expressed more negatively in general after the uprising compared to before the uprising. Four main reasons are likely attributable to a lack in statistical significance in other changes: (i) the sample sizes were lower for other areas of sentiment and (ii) p-values were corrected for family-wise error such that statistical significance required a p of $0.05/7 = 0.007$; (iii) dimensions were flawed due to inadequate or inappropriate items used for a given dimension (e.g., too broad, doesn't necessarily mention Egypt or refer to current events); (iv) alternate sentiment and topic analysis methods could produce different results. Nevertheless, these findings demonstrate that survey and Twitter data did not provide consistent results and were therefore not well aligned.

Table 2. Descriptive Statistics for Gallup Data

Dimension	Sample Size	Mean	Mean	Standard Deviation	Standard Deviation
		Oct-10	Apr-11	Oct-10	Apr 2011
Positive	1009	3.71	9.05	0.58	1.54
Economic Optimism	439	1.33	1.89	0.74	0.99
Institutional Confidence	678	1.61	2.6	0.49	0.99
Negative	1011	2.94	2.9	0.91	0.97
Overall Positivity	1011	8.89	15.3	1.67	2.49

Table 3. Paired Sample *t*-test Results for Gallup Data (* = p <0.01)

Dimension	Mean Difference	*t*- statistic	Degrees of Freedom	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Positive	-5.34	-102.27*	1008	-5.44	-5.24
Economic Optimism	-0.56	-9.09*	438	0.44	0.68
Institutional Confidence	-0.99	-23.71*	677	-1.07	-0.09
Negative	0.03	0.85	1010	-0.05	0.11
Overall Positivity	-6.41	-65.79*	1010	-6.61	-6.22

Table 4. Descriptive Statistics for Twitter Data

Dimension	Sample Size	Mean Sept-Oct 2010	Mean Mar-Apr 2011	Standard Deviation Sept-Oct 2010	Standard Deviation Mar-Apr 2011
Work Satisfaction	156	0.09	0.046	0.176	0.167
Confidence in Army	43	-0.059	-0.063	0.115	0.144
Confidence in Judicial System	29	-0.018	-0.07	0.123	0.094
Confidence in Government	236	-0.014	-0.046	0.139	0.169
Confidence in Elections (honesty)	106	-0.03	-0.069	0.129	0.13
Overall Sentiment	1231	0.029	-0.04	0.169	0.139

Table 5. Paired Sample *t*-test Results for Twitter Data (* = $p < 0.007$)

Dimension	Sample Size	Mean Difference	*t*-statistic	Degrees of Freedom	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Work Satisfaction	156	0.044	2.23	155	0.005	0.084
Confidence in Army	43	0.005	0.17	42	-0.053	0.062
Confidence in Judicial System	29	0.052	1.59	28	-0.015	0.12
Confidence in Government	236	0.031	2.24	235	0.004	0.059
Confidence in Elections (honesty)	106	0.039	2.18	105	0.004	0.075
Overall Sentiment	1231	0.068	12.47*	1230	0.058	0.079

4 Discussion

Analysis of sentiment in the Twitter data identified an increase of negativity experienced by individuals between pre- and post-uprising, whereas for the Gallup data, respondents demonstrated greater positivity post-uprising as compared to pre-uprising. Several possible explanations exist for this effect, such as the possibility that the unstructured nature of social media reflects a bias towards the expressing only the chaos and confusion that the population was experiencing following the uprising. Those who use Twitter are likely to be unrepresentative of the general population, where, for example, they may tend to be people who are inherently outwardly emotional. It may also be the case that, given the interview nature of the survey, respondents provided answers that were not as emotional as they would be in open-ended platforms, especially because polling often involves in-person interviews. Impression management and experimenter demands may explain an increase in positivity in the survey data.

The results of this analysis serve as a cautionary warning toward the use of social media as a 'pulse' of a population. While traditional survey methods may have limited reach, the nature of social media presents several different sources of error. First, there exists an obvious socio-economic bias in the sample of users of both the Internet and social media, where members of more disadvantaged populations are grossly under-represented (Facebook penetration in Egypt is only 5.49% as of December of 2010 [5]). Second, though increasingly intelligent methods of natural language processing of social media text are in development, the limitations afforded by processing data at such a scale necessitates the introduction of additional sources of error, such as the use of keywords that could be interpreted in multiple contexts (e.g., 'work' as in 'I hate going to work on Mondays' versus 'This election process is never going to work'). Third, in addition to the complexity of analyzing new language constructs constantly being created (e.g., the use of emoticons and abbreviations such as 'lol'), language translation proves difficult in informal and often ungrammatical usage such as in social media. Follow-up work to this study will include performing a comparison between sentiment analysis techniques, including human coding, and the traditional polling data. We also intend to perform an analysis of the Arabic tweets so as to determine if English tweets are significantly different in the attitudes conveyed.

5 Conclusion

This analysis sought to ascertain trends in sentiment relative to various topics for the Egyptian population using both traditional (polling) and new (social media) information sources and to identify the extent to which these two sources aligned during the 'Arab Spring'. The results showed that trends across the two sources are not exactly consistent. Focusing solely on Twitter data, individuals expressed increasingly negative opinions after the uprising, whereas polling indicated that individuals were increasingly positive post-uprising. These results suggest a greater need for caution when drawing conclusions based on the use of social media as a sole descriptor of a population or event.

References

1. Al Ahram Weekly, http://mohamed-salah.com/2010/11/17/Twitter-stats-in-egypt
2. ArabCrunch, http://arabcrunch.com/2011/04/egypts-mcit-egypt-has-23-51-million-internet-users-71-46-million-mobile-subscribers-3972-ict-companies.html
3. Dubai School of Government: Arab Social Media Report Civil Movements: The Impact of Facebook and Twitter, 1 (2011)
4. Bailar, B., Bailey, L., Stevens, J.: Measures of interviewer bias and variance. Journal of Marketing Research 14, 337–343 (1977)
5. Mourtada, R., Salem, F.: Arab Social Media Report: Facebook Usage 1 (2011)
6. Crocker, L., Algina, J.: Introduction to Classical and Modern Test Theory. Harcourt Brace, New York (1986)
7. Dugan, L.: Unofficial Reports Suggest Twitter Surpassed 500M Registered Users in June, http://www.mediabistro.com/allTwitter/Twitter-500-million-registered-users_b26104
8. Farhi, P.: The Twitter explosion. American Journalism Review 31 (2009)
9. Flores-Macias, F., Lawson, C.: Effects of interviewer gender on survey responses: Findings from a household survey in Mexico. International Journal of Public Opinion Research 20, 100–110 (2008)
10. Howard, P.: The digital origins of dictatorship and democracy: Information technology and political Islam. Oxford University Press, London (2011)
11. Huberman, B.A., Romero, D.M., Wu, P.: Social networks that matter: Twitter under the micro-scope. First Monday 14, 1 (2009)
12. Hughes, D.J., Rowe, M., Batey, M., Lee, A.: A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior 28(2), 561–569 (2012)
13. Number of active users at Facebook over the years, http://finance.yahoo.com/news/number-active-users-facebook-over-years-214600186-finance.html
14. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter: A social network or news media. In: Proceedings of the 19th International Conference on the World Wide Web (2010)
15. Nunnally, J.C.: Psychometric Theory. McGraw Hill, New York (1967)
16. Papacharissi, Z.: A private sphere: Democracy in a digital age. Polity Press, Cambridge (2010)
17. Papacharissi, Z., de Fatima Oliveira, M.: Affective news and networked publics: The rhythm of news storytelling on Egypt. Journal of Communication 62(2), 266–282 (2012)
18. Sysomos, http://blog.sysomos.com/2011/01/31/egyptian-crisis-twitter
19. Gallup: Worldwide Research – Country Data Set Details (2010)
20. Abu Dhabi Gallup Center: Egypt: The Arithmetic of Revolution, An empirical analysis of social and economic conditions in the months before the January 25 uprising (2011)
21. Kim, S., Li, F., Lebanon, G., Essa, I.: Beyond Sentiment: The Manifold of Human Emotions, arXiv:1202.1568 (2011)