

Evaluating Public Discussion of Occupying Wall Street on Twitter: Linking Twitter Streams with Search Queries, Opinion Polls, Media Coverage, and Stock Market Index

Cheng-Jun Wang¹ and Tai-Quan Peng^{1,2}

1. Web Mining Lab, Dept. of Media & Communication, City University of Hong Kong

2. Faculty of Humanities & Arts, Macau University of Science and Technology

Abstract

To evaluate the public opinion expression of occupying Wall Street on Twitter, this paper aims to investigate time series of tweet stream, search query, media coverage, stock market, and opinion polls. The findings reveal that: first, the dynamic change of total tweet stream, retweets, discussions, and hashtags are parallel; second, tweet stream and search query are strongly correlated, while tweet stream has only weak correlation with media coverage, opinion poll of president job disapproval rate, stock market index; third, tweet stream has significant influence on search query, however, we found no evidence of statistically significant impacts of digital traces on media coverage, stock market, and president job disapproval rate. The results shed light in understanding the role of digital traces in both reflecting the longitudinal opinion intensity and influencing real world indicators of public opinion.

Keywords:

Public Opinion, Digital Traces, News Sharing website, Opinion Intensity

Introduction

The public sphere have been extended through the public discussion of cyberspace (Dahlberg, 2001). Public opinion is shaped by public expression, information recommendation, and political discussion, all of which have been greatly facilitated by news

sharing networks (e.g. Twitter, Weibo, and Digg), which are defined as social media on which users can submit, share, disseminate, and discussing news.

Public opinion expression on news sharing website supplies real time information of public opinion (Gonzalez-Bailon, Banchs, & Kaltenbrunner, 2011), thus it's widely used to track the public opinions, stock market, elections, TV ratings (Chew & Eysenbach, 2010; Gonzalez-Bailon, et al., 2011; Livne, Simmons, Adar, & Adamic, 2011; O'Connor, Balasubramanian, Routledge, & Smith, 2010; Tumasjan, Sprenger, Sandner, & Welpe, 2010; Wakamiya, Lee, & Sumiya, 2011).

However, whether the online discussion of micro-blogs reflects the public opinion is still an open question asking for further research. First, online public opinion are criticized for its bias in representing the whole population rather than only Internet users (Gayo-Avello, 2011; O'Connor, et al., 2010). Second, although the tweeting behavior are both self-motivated and collectively influenced, and the vocal minority are very active, the majority is silent (Mustafaraj, Finn, Whitlock, & Metaxas, 2011). Based on these arguments, the validity of online discussion should be reconsidered.

To bridge the theoretical gap of evaluating the validity of online discussion on news sharing website, this paper aims in comparing tweet streams with, online search query, media coverage, traditionally offline opinion polls and stock price.

There are three dimensions of public opinion, that is: salience (e.g. intensity), valence (e.g. agree, disagree, neutral), and mood (e.g., positive or negative, opinion extremity, angry, happy, etc.) (Gonzalez-Bailon, et al., 2011). Scheufele (1999) distinguished cognitive and affective dimensions of public discussion. First,

Copyright is held by the authors.

The annual conference of the World Association for Public Opinion Research, Hong Kong, June 14-16, 2012.
Correspondence should be addressed
wangchj04@gmail.com.

citizens are rational in paying attentions to news events and issues (i.e. intensity of opinion), and deciding to support or oppose, however, second, they are also affective in making decisions of allocating their attentions and forming their attitudes. As an exploratory study, in this paper, we only focus on the first aspect of the cognitive dimension of public opinion, and aiming to find how the allocation of attentions on twitter mirrors the public opinion.

To compare across digital traces, news coverage, stock market, and opinion polls, this paper will focus only on the dimension of salience to detect and access the public opinion on twitter.

Literature Review

Public Opinion Expression

Public opinion expression is extensively studied in research of spiral of silence in terms of outspokenness (Glynn & Park, 1997; Noelle-Neumann, 1984; Scheufele, 1999). Scheufele (1999) argues that “wilingness of expression” could be replaced as a more general term “public discussion” which could be further distinguished as two dimensions: political talk and public opinion expression based on two modes of public opinion which were supplied by Nolle-Neumann (1984). Political talk, conceptually, is the cognitive dimension of public opinion, which is defined as the rational exchange of ideas and arguments towards specific issues. Public opinion expression, on the contrary, is the affective dimension of public opinion, which is linked with willingness to express under social pressures.

Opinion expression is an indicator of the intensity of public opinion. Weaver (1991) confirmed that issue salience is in accordance with public knowledge, opinion, and behavior. Glynn and Park (1997) argues that concern about issues is highly correlated with intensity of opinion, and they find that opinion intensity is a strong predictor of opinion expression.

Intensity of opinion measures to what extent respondents are concerned about the issues (Glynn & Park, 1997). Opinion expression can be measured in multiple ways. However, they primarily study opinion expression by gauging the willingness of expression, e.g., Noelle-Neumann measured opinion expression in multiple ways, including “train test”, a

battery of eleven ways to provide political support for a party, and filling in a sentence in a cartoon (Glynn & Park, 1997; Noelle-Neumann, 1984). Although traditional measures of intensity of opinion also measure the attitude (i.e. the intensity of agree or disagree), it’s necessary to distinguish the valence of public opinion and the salience of public opinion, and in this study, we merely focus on the salience of public opinion which is measured by the intensity of opinion expression.

Public Opinion and Public Attention

Public opinion and public attention are conceptually different. Public attentions are related to the scarcity of resources (e.g. time), due to which citizens have to allocate their attentions to publicly debated issues. The allocation of public attention is shaped by the issue salience (Zhu, 1992). While public opinion covers more than salience, it also includes the valence (e.g. attitude) and emotion towards public issues (Price & Neijens, 1997; Scheufele, 1999). Thus public attention is one component of public attention since it reflects the intensity of public opinion by measuring issue salience.

It’s necessary to note that although attitudes towards public issues are deeply rooted in the belief and value systems of the society, the public attentiveness for specific issues may not be so stable. Due to the dynamic development of social and political reality, public attentions will change dramatically (Newig, 2004).

Digital Traces and Issue Salience

As a virtual space where individuals leave their footprint there, Internet has become a perfect way of documenting human communication behaviors. First, in an unobtrusive way, digital traces document how users allocate their attentions to news events and important issues. Second, digital traces supply an alternative observational tool for tracing public agenda, giving traditional opinion polls’ heavy rely on the indirect measure of self-report. Undoubtedly, digital traces measure perceived issue salience and has increasing potential for public opinion study with the fast percolation of Internet into daily life. In this paper, we focus on two kinds of digital traces: tweet stream and search query, conceptually, both of which are salience-driven.

Tweet stream. As a new kind of unobtrusive digital traces, tweet streams integrate both monologues, dialogues, and diffusions by creating tweets, giving comments, speaking to other people (using @), and retweeting (using RT@). Collectively, people could express their opinions under the same topic (using the hashtag #).

Tweet stream is a good indicator of public opinion. First, the personal talk on Twitter is informal, and informal conversations are the loci of a great amount of political talk (Wyatt, Katz, & Kim, 2000). Second, the interactions on Twitter are based on interpersonal social networks. Interpersonal communication was found to enhance agenda-setting function of mass media (McLeod, Becker, & Byrnes, 1974), and it also outweighs news in predicting issue salience (Wanta & Wu, 1992). Third, twitter supplies a virtual but real public sphere on which users can speak to politicians and other elites. Traditional opinion polls are criticized for collecting 'pseudo opinions' developed outside of any meaningful public debate (Price & Neijens, 1997). As a real social setting, twitter enables individuals to be informed by the public opinion, express their own standpoints, and debate with other people.

Search query. In addition to tweet streams, there are many other digital trace data, and one of them is search query. A web search query is a query that a user enters into web search engine to satisfy his or her information needs. Documented by search engine, the search query generally reflects the Intensity of public attention of Internet users. Search query, as one kind of digital traces, is produced by an important kind of information seeking behavior which is the behavioral consequences of issue awareness and salience (Scharnow & Vogelgesang, 2011).

Real World Indicators of Public Opinion

Media Coverage. Media opinion influence public opinion by setting agenda. Although both digital traces and media coverage measure the mediated public attention and the intensity of the public opinion, it's necessary to distinguish them in terms of digital world (i.e., online environment) and the real world (i.e. offline environment). Since MacCombs and Shaw's study of agenda-setting, a lot of studies have devoted to this line of thought (MacCombs &

Shaw, 1972, 1993; Rogers & Dearing, 1988; Zhu, 1992).

Opinion Polls. The workhorse of tools for tracking public opinion is the opinion poll which has been widely used. However, individual responses measured by opinion polls formed outside the arena of public debate, thus deliberative polls in which individuals are informed of the public opinion are thought to be a better compliment (Fishkin, 1991, 1997; Price & Neijens, 1997). However, all the opinion polls have come under harsh criticism for manufacturing public opinion for its obtrusive way of data collection (Salmon & Glasser, 1995), and the unobtrusive digital traces may be a better choice in the future (if only it's representative). By combining traditional polls with Twitter analyses, a more accurate picture of public opinion could emerge.

Stock Market Index. Stock market index is an indicator of public opinion towards the economic situation. Stock Market will be widely influenced by both rationale estimations of economic state and irrational factors, e.g. weathers. The stock price was found to be significantly correlated with the weather of New York City in a long history (Hirshleifer & Shumway, 2003; Loughran & Schultz, 2004; Saunders, 1993; Trombley, 1997).

Prior studies of elections and the sentiment of tweets confirmed the influence of tweets on stock market, although there are different viewpoints. While in this study, we only focus on the intensity of opinion expression, rather than the sentiment of public opinion.

Research Question

Following the logic of the literature, I formulate the theoretical framework of evaluating how online discussion on microblogs as one kind of public opinion correlate with search query, opinion polls, media coverage, and stock market (see Figure 1).

Focusing on the relationship between digital traces and other indicators of public opinion, this study proposes to study two kinds of research questions: convergent validity between digital traces and real world indicators of public opinion; and second, the mutual influence between digital traces and real world indicators of public opinion.

Convergent Validity between Digital Traces and Real World Indicators

Both tweet stream and search query reflect the issue awareness and salience (Scharnow & Vogelgesang, 2011). Thus it's expected that there is correlation between search query and issue salience, and we formulate our research question:

RQ1: How digital traces are associated with news coverage, opinion polls, and stock market index?

Mutual Influence between Digital Traces and Real World Indicators

The causal relationship between digital traces and other measures of public opinions are too complicated and asking for more attentions. Scholars are interested in whether the online salience-driven behavior has impact on the real world public opinion. Prior work has been done by exploring the mechanism of agenda-setting, inter-media agenda-setting, and agenda-building.

First, according to the study of agenda-setting, media coverage attracts public attentions and sets agenda for the public. News produced by mass media were observed to have great impact on online discussion, e.g. Wu et al's study found that on Twitter, 50% of URLs consumed are generated by just 20K elite users, where the media produces the most information (Wu, Hofman, Mason, & Watts, 2011).

Second, as one kind of social media, twitter is one important step of inter-media agenda setting. According to Kwak et al's finding, the majority (over 85%) of topics on twitter are headline news or persistent news (Kwak, Lee, Park, & Moon, 2010). Thus Twitter is more widely recognized as news media. And according to the study of inter-media agenda-setting (Wanta & Foote, 1994), online discussion on twitter hereafter may have impact on news production.

Third, in addition to that, scholars are getting more interested in how media agenda are shaped in terms of agenda-building (McCombs, 1992). Users collectively participated in filtering information to share news, adding comments, and producing new information. Twitter users intentionally participate in mobilizing the elites (e.g. mass media).

Forth, important news benefits more from

interpersonal effect on social media. The relationship between tweets and news is influenced by the mechanisms of news diffusion. According to the study of classic news diffusion (Adams, Mullen, & Wilson, 1969; Banta, 1964; Budd, MacLean Jr, & Barnes, 1966; Deutschmann & Danielson, 1960; Funkhous.Gr & McCombs, 1971; Funkhouser, 1970; Greenberg, 1964; Hill & Bonjean, 1964; Sheatsley & Feldman, 1964), especially J-curve model, news of great importance will be diffused by both mass media and interpersonal networks. Compared with news of less importance, interpersonal effect plays a more important role. Twitter as one kind of social media is more closed connected with interpersonal communications. Occupying movement is a global movement of great importance, thus interpersonal networks are more sensitive in discussing the development of the movement.

Online public discussion of occupying Wall Street is supposed to influence the stock market and the satisfaction to the government, especially president's job disapproval rate. First, practically, Hill's poll of occupying movement found that 33% of the people blame Wall Street, while 56% of the people blame Washington (see table 3). Stock market is sensitive to the public opinion. It's possible that online political talk and relevant information seeking behavior will have impact on the stock market and Obama's job disapproval rate. Second, empirically, online opinion and opinion polls are found to be positively correlated. By examining about 1 billion tweets post during 2008 and 2009, and measuring consumer confidence and President Obama's approval rating, O'Connor et al (O'Connor, et al., 2010) find Twitter posts are in line with opinion polls.

Based on the arguments above, we formulate the second research question:

RQ2: Whether digital traces (e.g., tweet streams and search query) have significant influence on news coverage, opinion polls, and stock market index?

Figure 1 about here

Method

Traditional studies employs surveys (e.g. opinion polls) to measure opinion expression (Glynn & Park,

1997; Noelle-Neumann, 1984), e.g. the “train test” is used by Nolle-Neumann (1984). The intensity of opinion is measured using scales (Glynn & Park, 1997). While in our study, we measure the salience of public opinion by the frequency of tweeting or searching “occupying Wall Street”, news coverage, stock market index, and also offline opinion polls. Multiple time series analysis is used to gauge the research question.

Data and Measure

Using the tweets data collected from Twitter stream, search engine, opinion polls, news coverage, and stock market, this study compares the public opinion of occupying Wall Street on the aggregate level with survey, online search query, media coverage, and stock price.

Tweet Stream. The tweet stream data is retrieved from tweets of Occupying Wall Street which is collected by R-shief¹ through Twitter’s API, and it contains 6317976 tweets within 120 days (between Oct 6th 2011 and Feb 11th 2012), among which, there are 4285404 retweets, 6513780 discussions (which are identified by the symbol of “@”), and 15558606 hashtags.

Figure 2 about here

Search Query. The search query data are retrieved from Google Trend by searching for the key words of “occupy Wall Street”.

News Coverage. The data of news coverage about occupy Wall Street on Wall Street Daily is retrieved from the online archive of the official website of Wall Street Daily. Only 6 articles are found. Media coverage of occupying Wall Street on Washington Post is retrieved from the archive of Proquest, and the news coverage of New York Times and The Guardians are collected through their API. The growth curves of three newspapers share the same patterns. To make the analysis convenient, the news coverage of the three newspapers is combined together (see Figure 2).

Opinion polls. There are three polls relevant to Occupying Wall Street conducted, including: Zuccotti

Park Poll², The Hill Poll³, and USA Today/Gallup Poll⁴. The findings give a static picture for the public opinion for both participants of OWS and national voters.

Table 1 about here

Since the specific surveys of occupying Wall Street are not time series data (e.g. the measures are not consistent, the time span are different, and people being surveyed are different), thus it’s impossible to use them as the baseline. Instead, we use the daily opinion poll of president approve/disapprove rate as the offline public opinion. Although the president approval/disapproval rate is just a general survey of the American society, it could be used as control variable in time series analysis. The daily president approve/disapprove rate is retrieved from Gallop’s daily opinion polls.

Stock Market. Many financial websites have been writing articles about the Occupy Wall Street protests. At least one article about the movement seems to be present on Google Finance at all times and it is starting to become a hot topic at the water cooler in offices across the country. The stock market Index is collected from website of Yahoo Finance.

Result

Parallel relationship within tweets and stock indexes

To simplify the time series analysis by avoiding involving too many variables of the same category, OLS regression is used to test the time series of different tweets. Interestingly, we find the time series of total tweets, retweets (i.e., RT@), discussions (i.e. @), and hashtag (i.e., #) are parallel (see figure 7).

Similar to the parallel relationship among retweets, retweets, discussion, and hashtags of tweet stream, we observe similar parallel relationship among different stock indexes. Based on these parallel

²

http://www.douglasschoen.com/pdf/Occupy_Wall_Street_Poll_Douglas_Schoen.pdf

³

http://thehill.com/images/stories/news/2011/10_october/crossstabs_20111013_thehill.pdf

⁴

http://www.gallup.com/file/poll/150167/Occupy_Wall_Street_1_111018.pdf

¹ <http://www.r-shief.org/>

relationships, I don't need to involve all the variables of tweets and stock market indexes. Instead, I choose the total tweets and NYSE stock market index as the measures of tweet streams and stock market index, respectively.

Figure 3 about here

Figure 4 about here

Convergent Validity

Correlation Matrix. To address the first research question about convergent validity, the face validity was detected first. With regard to face validity, the degree to which the measure looks like it is designed to measure should be checked. To inspect the face validity, normalized measures (i.e., mean equals 0, and SD equals 1) are plotted together in Figure 3. Intuitively, the measure of digital traces has modest face validity. Tweet stream and search query change simultaneously. The president job disapproval rate synchronously grow with digital traces at first, however, after one and a half month, it fluctuates dramatically. Daily based news coverage seems to overlap with digital trace. However its fluctuation is too big. Stock market index has reversed relationship with digital traces, and there is a time lag.

Figure 5 about here

To further test the convergent validity of different measures, the correlation matrix of measures is demonstrated in Table 2. The time series of tweet stream has strong correlation with search query ($r = .78, p < .001$), while weak correlation with news coverage ($r = .4, p < .001$), Obama job disapproval rate ($r = .31, p < .01$), Obama approve rate ($r = -.37, p < .001$), and stock index ($r = -.34, p < .01$).

Compared with tweet stream, search query are more strongly correlated with news coverage ($r = .55, p < .01$), Obama job disapproval rate ($r = .57, p < .01$), Obama approve rate ($r = -.60, p < .01$), and stock index ($r = -.61, p < .01$).

Table 2 about here

Mutual Influence

Unit Root and Cointegration Test. A requirement of the VAR technique is the stationarity of the series. The Augmented Dickey Fuller (ADF) test is one of the most frequently used tests for unit roots. The null hypothesis is that there exists a unit root (i.e., that the series is not stationary) for search query, stock market index, and president job disapproval rate. However, all the time series turn to be stationary after first differencing (see table 3). Thus only the first difference of the indicators will be included in further analysis.

Table 3 about here

Phillips-Ouliaris Cointegration Test supports to reject the null hypothesis (Phillips-Ouliaris demeaned = -135.8872, Truncation lag parameter = 1, p-value = 0.01), and thus reveals that the time series are cointegrated (i.e. there exists a long run relationship among the five variables) which implies that the usual vector auto regression may be spurious. To test the cointegrated influence between digital traces and real world indicators, vector error correction model should be employed.

Vector Error Correction Model. The result of VECM is given in table 4. Tweet stream has significant influence on search query. However, no significant relationship was found between digital traces and real world indicators.

Table 4 about here

Discussion and Conclusion

Two concerns were explored in this study: first, the convergent validity of different measures; and second, the mutual influence of different measures. To address these two research questions, a case study of public discussion about occupying Wall Street is used.

Digital traces are a somewhat different but still related measure of the public opinion intensity compared with traditional survey measures. Tweet stream is weakly correlated with opinion poll of Obama's job disapproval rate, while search query has modest size of influence on opinion polls. Similar findings have been found in Scharkow et al's study (2011). They find the correlation coefficient is .49. In our study, the correlation between search query

(about occupying Wall Street) and Obama's job disapproval rate is .57.

Compared with tweet stream correlation coefficient between tweet stream and real world indicators, the higher correlation coefficient between search query and real world indicators, may indicate digital divide between Internet users and general people. The composition of information searchers has more similarity with general people compared with Microblog users. Thus to make twitter streams as an representative measure of public opinion, micro-blog still has a long way to go to representatively measure the overall public opinion.

However, our finding of search query's correlation with media coverage is relatively low (.55) compared with a similar study which is based on weekly aggregated search query data (Ripberger, 2011). This may be caused by the large fluctuation of the daily based media coverage, which is a shortcoming a monitoring public opinion by media coverage based on daily data.

Although we found weak correlation between tweet stream and real world indicators of public opinion, however, no evidence of statistically significant impacts of digital traces on media coverage, stock market, and president job disapproval rate was found. This finding implies that although Internet, and especially microblogs, has been considered as an important tool to influence the real world, so far this proposition has not been sufficiently supported, at least in this study.

Tweet stream has great internal validity as one kind of digital traces (e.g., search query), since it has a strong correlation with search query, and it has significant influence on search query. The political discussion of occupying Wall Street on twitter will promote general Internet users to search for the movement using search engine. In addition to that, the strong parallel relationship among total tweets, retweets, hastags, and discussions implies that it's possible to focus only one part of these indicators to study the opinion intensity reflected by twitter in general and the issue salience in specific, e.g., only retrieving the relevant retweets rather than collecting all the tweets.

This study suffers from two other aspects: first,

comparing specific events with general opinion polls. The measure of tweet stream, search query, and media coverage centers on the occupying movement, while the opinion poll is about the president job approval rate, and this may account for the insignificant finding of agenda-setting effect between media coverage and the president job approval rate; and second, this paper merely focuses on the salience of public opinion, and valence and emotion are not included. Thus, it should be cautious to conclude that tweet stream can or can't influence (not predict) opinion polls only based on the intensity of opinion expression without evaluating the valence and emotion of opinions. The emotion extremity may lead to polarization of public opinion (Conover et al., 2011), which is consistent with Scheufele's study which reveals that Schumpeter's assumption is right, and an informed and rational citizen does not exist (Scheufele, 1999). It's also worthy to note that this paper aims in assessing the public expression on twitter in terms of digital traces, rather than predicting.

Overall, then, this paper gauges the convergent validity between digital traces and real world indicators, and reinforces the argument that there is a gap between digital traces and real world public opinion, although there is a strong convergent validity within digital traces. The findings again remind us to be cautious of using digital traces or real world indicators of public opinion or arguing for the mutual influences, without being sufficiently informed of their disadvantages.

References

- Adams, J. B., Mullen, J. J., & Wilson, H. M. (1969). Diffusion of a "minor" foreign affairs news event. *Journalism Quarterly*, 46, 545-551.
- Banta, T. J. (1964). The Kennedy assassination: Early thoughts and emotions. *Public opinion quarterly*, 28, 216-224.
- Budd, R. W., MacLean Jr, M. S., & Barnes, A. M. (1966). Regularities in the diffusion of two major news events. *Journalism Quarterly*, 43(22), 1-230.
- Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, 5(11), e14118.
- Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., & Menczer, F. (2011).

Political polarization on twitter.

- Dahlberg, L. (2001). The Internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4), 615-633.
- Deutschmann, P. J., & Danielson, W. A. (1960). Diffusion of knowledge of the major news story. *Journalism Quarterly*, 37(3), 345-355.
- Fishkin, J. S. (1991). *Democracy and deliberation: New directions for democratic reform* (Vol. 217): Cambridge Univ Press.
- Fishkin, J. S. (1997). *The voice of the people: Public opinion and democracy*: Yale Univ Pr.
- Funkhous.Gr, & Mccombs, M. E. (1971). Rise and Fall of News Diffusion. *Public opinion quarterly*, 35(1), 107-&.
- Funkhouser, G. R. (1970). A probabilistic model for predicting news diffusion. *Journalism Quarterly*, 47, 41-45.
- Gayo-Avello, D. (2011). Don't turn social media into another 'Literary Digest' poll. *Communications of the ACM*, 54(10), 121-128.
- Glynn, C. J., & Park, E. (1997). Reference groups, opinion intensity, and public opinion expression. *International Journal of Public Opinion Research*, 9(3), 213-232.
- Gonzalez-Bailon, S., Banchs, R., & Kaltenbrunner, A. (2011). Emotions, Public Opinion and US Presidential Approval Rates: A 5 Year Analysis of Online Political Discussions.
- Greenberg, B. S. (1964). Person to person communication in the diffusion of a news event. *Journalism Quarterly*, 41, 489-494.
- Hill, R. J., & Bonjean, C. M. (1964). News Diffusion: A Test of the Regularity Hypothesis. *Journalism Quarterly*(41), 336-342.
- Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3), 1009-1032.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April 26-30). *What is Twitter, a social network or a news media?* Paper presented at the WWW 2010, Raleigh, North Carolina, USA.
- Livne, A., Simmons, M. P., Adar, E., & Adamic, L. A. (2011). *The Party is Over Here: Structure and Content in the 2010 Election*. Paper presented at the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain.
- Loughran, T., & Schultz, P. (2004). Weather, stock returns, and the impact of localized trading behavior. *Journal of Financial and Quantitative Analysis*, 39(2).
- McCombs, M. E. (1992). Explorers and surveyors: Expanding strategies for agenda-setting research. *Journalism & Mass Communication Quarterly*, 69(4), 813-824.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, 36(2), 176-187.
- McCombs, M. E., & Shaw, D. L. (1993). The evolution of agenda- setting research: Twenty- five years in the marketplace of ideas. *Journal of Communication*, 43(2), 58-67.
- McLeod, J. M., Becker, L. B., & Byrnes, J. E. (1974). Another look at the agenda-setting function of the press. *Communication Research*, 1(2), 131-166.
- Mustafaraj, E., Finn, S., Whitlock, C., & Metaxas, P. (2011). Vocal minority versus silent majority: Discovering the opinions of the long tail. *Proc. of IEEE SocialCom*.
- Newig, J. (2004). Public attention, political action: The example of environmental regulation. *Rationality and Society*, 16(2), 149-190.
- Noelle-Neumann, E. (1984). *The Spiral of Silence: Public opinion-Our Social Skin*. Chicago.
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). *From tweets to polls: Linking text sentiment to public opinion time series*. Paper presented at the ICWSM-2010.
- Price, V., & Neijens, P. (1997). Opinion quality in public opinion research. *International Journal of Public Opinion Research*, 9(4), 336-360.
- Ripberger, J. T. (2011). Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness. *Policy Studies Journal*, 39(2), 239-259.
- Rogers, E. M., & Dearing, J. W. (1988). Agenda-setting research. *Communication yearbook*, 11.
- Salmon, C. T., & Glasser, T. L. (1995). The politics of polling and the limits of consent *Public opinion and the communication of consent* (pp. 437-458): Guilford Press.
- Saunders, E. M. (1993). Stock prices and Wall Street weather. *The American Economic Review*, 83(5), 1337-1345.
- Scharkow, M., & Vogelgesang, J. (2011). Measuring the Public Agenda using Search Engine Queries.

International Journal of Public Opinion Research, 23(1), 104-113.

Scheufele, D. A. (1999). Deliberation or dispute? An exploratory study examining dimensions of public opinion expression. *International Journal of Public Opinion Research*, 11(1), 25-58.

Sheatsley, P. B., & Feldman, J. J. (1964). The Assassination of President Kennedy: A Preliminary Report on Public Reactions and Behavior. *Public opinion quarterly*, 28(2), 189-215

Trombley, M. A. (1997). Stock prices and wall street weather: Additional evidence. *Quarterly Journal of Business and Economics*, 11-21.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*.

Wakamiya, S., Lee, R., & Sumiya, K. (2011). Crowd-powered TV viewing rates: measuring relevancy between tweets and TV programs. *Database Systems for Adanced Applications*, 390-401.

Wanta, W., & Foote, J. (1994). The president- news media relationship: A time series analysis of agenda- setting. *Journal of Broadcasting & Electronic Media*, 38(4), 437-448.

Wanta, W., & Wu, Y. C. (1992). Interpersonal communication and the agenda-setting process. *Journalism & Mass Communication Quarterly*, 69(4), 847-855.

Weaver, D. (1991). Issue salience and public opinion: are there consequences of agenda-setting? *International Journal of Public Opinion Research*, 3(1), 53-68.

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, March 28-April 1). *Who says what to whom on twitter*. Paper presented at the WWW 2011, Hyderabad, India.

Wyatt, R. O., Katz, E., & Kim, J. (2000). Bridging the spheres: Political and personal conversation in public and private spaces. *Journal of Communication*, 50(1), 71-92.

Zhu, J. H. (1992). Issue competition and attention distraction: A zero-sum theory of agenda-setting. *Journalism & Mass Communication Quarterly*, 69(4), 825-836.

Figure 1 Digital Traces and Real World Indicators of Public Opinion

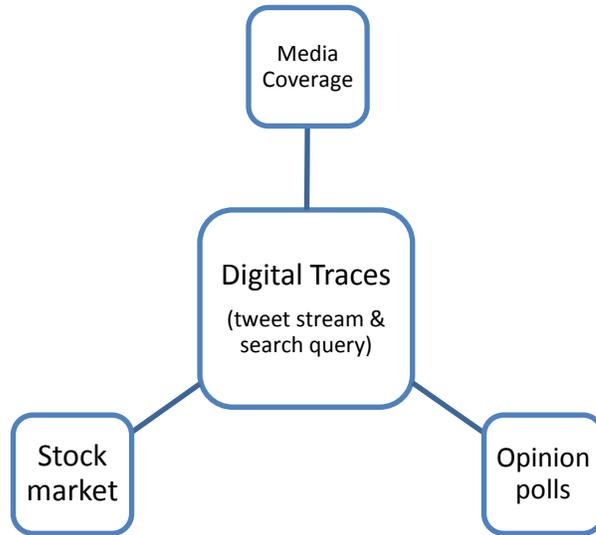


Figure 2 Time series of Tweet stream, search query, News Coverage about occupying Wall Street, and Obama job approval/disapproval rate, Stock Market Index

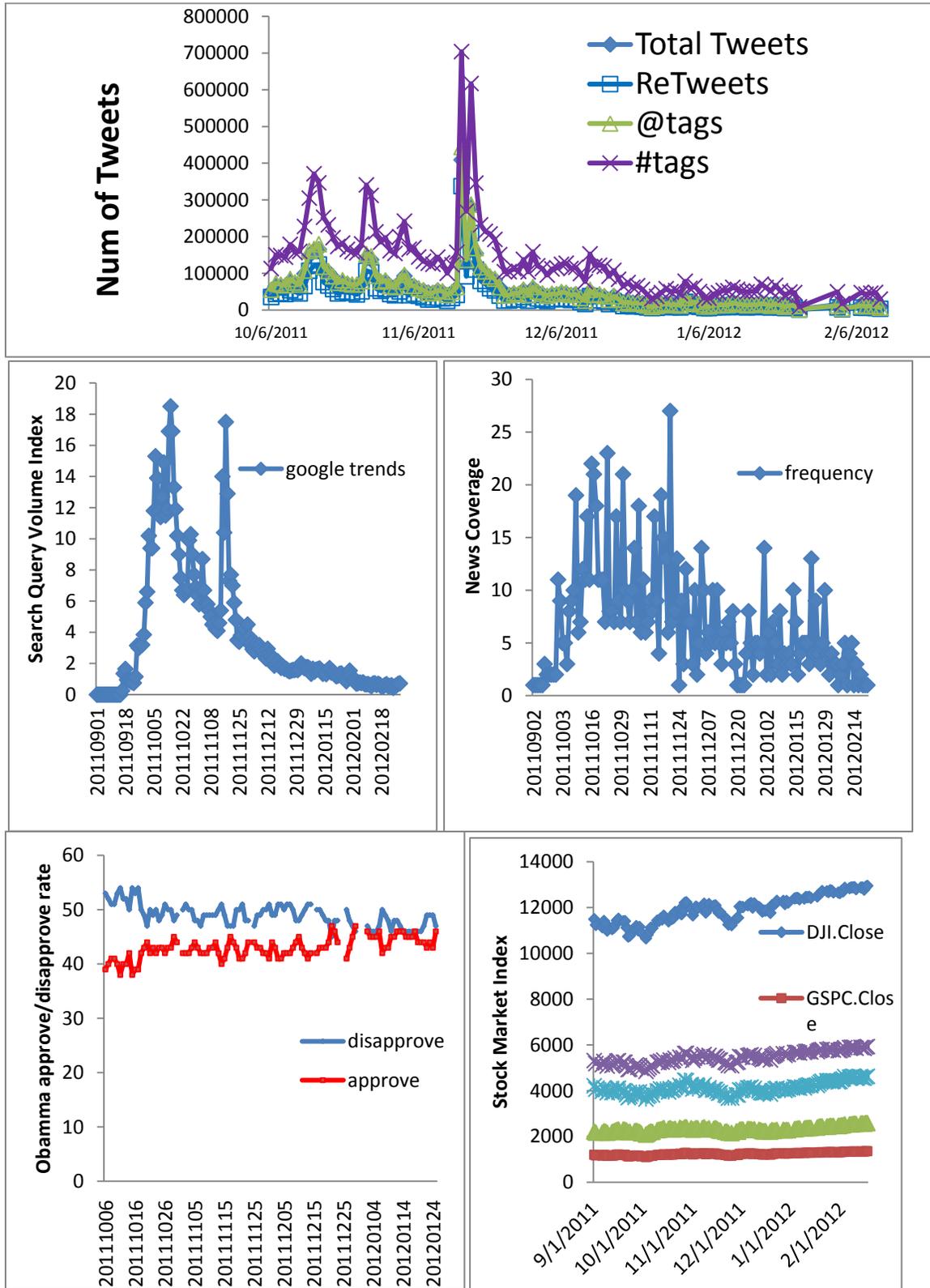


Figure 3 The parallel growth of Total Tweets, retweets, discussions, and hashtags

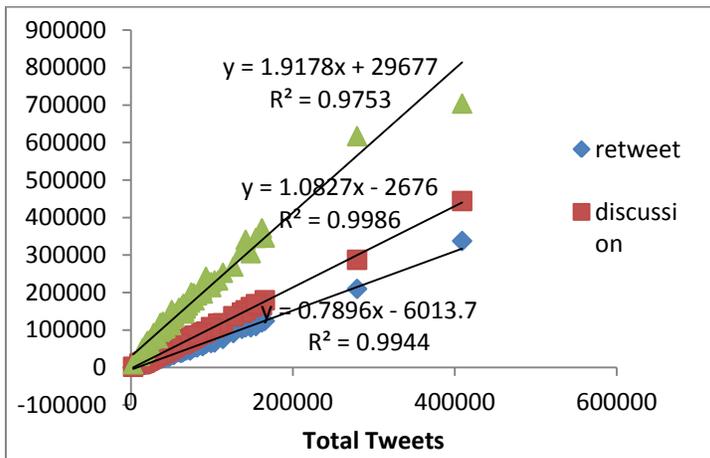


Figure 4 Parallel growth of stock indexes

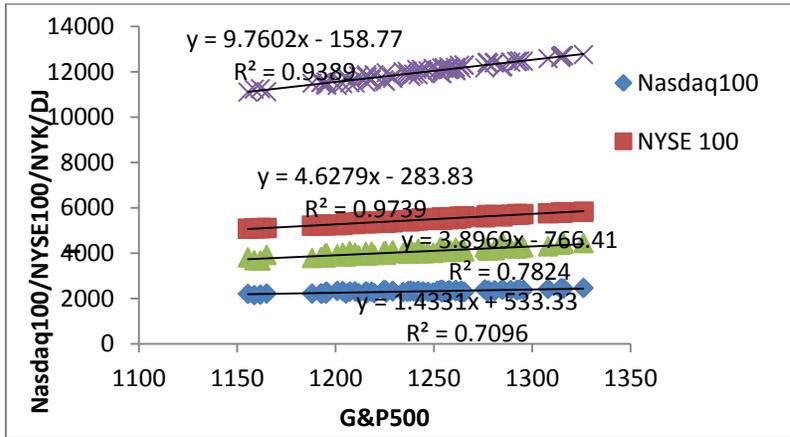


Figure 5 Normalized Time series (sd = 1, Mean = 0)

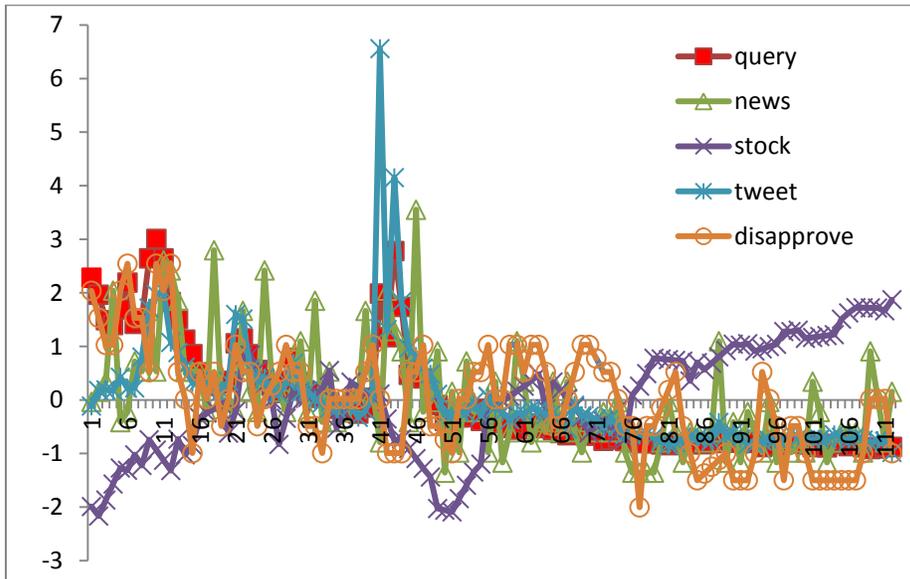


Table 1 Summary of Opinion Polls of OWS

	Valence of Opinion	Time & People
Zuccotti Park Poll	52% participated in social movement before; 73% plan to vote in 2012; 77% support raise taxes on the wealthiest Americans;	Oct. 10 and 11 200 protesters
The Hill Poll	33% blame Wall Street; 56% blame Washington	Oct. 13 1,000 likely national voters
USA Today/Gallup Poll	26% support OWS; 19% anti OWS	October 15-16 1,026 national voters

Table 2 Correlation matrix of measures

	tweet	query	News	stock	disapprove	approve
tweet	1	0.78	0.4	-0.34	0.31	-0.37
query	0.78	1	0.55	-0.61	0.57	-0.6
news	0.4	0.55	1	-0.42	0.33	-0.35
stock	-0.34	-0.61	-0.42	1	-0.55	0.61
disapprove	0.31	0.57	0.33	-0.55	1	-0.9
approve	-0.37	-0.6	-0.35	0.61	-0.9	1

Note: all of the correlation coefficients are significant

Table 3 Results of the Augmented Dickey-Fuller test for Unit Roots

Variable	ADF	First Difference.
Tweet stream	-4.3056 (4) **	-6.3194(4) **
Search query	-3.353(4)	-5.3314(4) **
News coverage	-3.6276(4) *	-6.1671(4) **
Stock market index	-2.5538(4)	-4.8387(4) **
President job disapproval	-3.3567(4)	-6.5687(4) **

Note: * P < 0.05. ** P < 0.01. *** P < 0.001.

Table 4 OLS regression of Vector Error Correction Model (VECM)

	Tweet	Query	News	Disapprove	Stock
tweetd.d11	-1.7822*** (.164)	-1.24E-05* (5.73E-06)	3.06E-05 (2.03E-05)	-2.05E-05 (2.43E-04)	-9.11E-06† (5.48E-06)
queryd.d11	8134.028† (4224.122)	-.654*** (.148)	-4.64E-01 (5.23E-01)	1.24 (6.26)	.176 (.141)
newsd.d11	-526.34 (764.7242)	-1.20E-02 (2.67E-02)	-1.77*** (.095)	-.202 (1.13)	1.13E-02 (2.55E-02)
stockd.d11	-37.5143 (68.322)	-2.54E-03 (2.39E-03)	9.47E-04 (8.46E-03)	-1.02*** (.101)	-1.63E-04 (2.28E-03)
disapproved.d11	4411.066 (3020.238)	.159 (.106)	.492 (.374)	4.52 (4.47)	-1.15*** (.101)
tweetd.l2	-1.6322 (.262)	1.28E-05 (9.14E-06)	4.39E-05 (3.24E-05)	5.29E-05 (3.88E-04)	-1.40E-05 (8.74E-06)
queryd.l2	1660.714 (5327.091)	-1.29*** (.186)	-5.17E-01 (6.59E-01)	-4.14E+00 (7.89E+00)	.236 (.178)
newsd.l2	-304.787 (1344.976)	1.33E-03 (4.70E-02)	-2.14E+00*** (.167)	-.422 (1.99)	-1.54E-02 (4.49E-02)
stockd.l2	-50.557*** (97.319)	-2.07E-03 (3.40E-03)	4.75E-03 (1.21E-02)	-1.07*** (.144)	-4.49E-04 (3.25E-03)
disapproved.l2	6154.219 (4548.565)	.099 (.159)	.884 (.563)	.530 (6.74)	-1.10*** (.152)
const	-308.634 (4055.106)	-.121 (.142)	-.087 (.502)	6.25 (6.01)	-.0133 (.135)
R square	.796	.691	.825	.533	.602

Note: * P < 0.05. ** P < 0.01. *** P < 0.001. † P < 0.1.