

Citizens' Use of Twitter in Political Information Sharing in South Korea

Jisue Lee
Florida State University
jl10n@my.fsu.edu

Hohyon Ryu
University of Texas, Austin
hohyon@utexas.edu

Lorri Mon
Florida State University
Lorraine.Mon@cci.fsu.edu

Sung Jae Park
Hansung University
spark@hansung.ac.kr

Abstract

This study examined citizens' use of social networking site Twitter in political information sharing in South Korea. Content analysis was used in classifying message types and sentiments from the most frequently re-tweeted (RT) messages including the names of three top political leaders running for general elections in 2012. Correlation analysis comparing citizens' use of Twitter in political information sharing online with results of public opinion polls offline indicated: 1) the volume and magnitude of re-tweeted messages are significantly correlated with results of public opinion polls; 2) types of messages are not correlated with the public opinion polling results; 3) positive and negative sentiment revealed in Twitter messages are highly correlated with the results of public opinion polls. Findings from this case study provide insights into citizens' use of Twitter in political communication.

Keywords: political communication, information sharing, re-tweeting (RT), message type classification, sentiment identification

Problem Statement

Social media tools such as Twitter, Facebook and YouTube are now considered as politically transformative communication technologies as radio and television. There are predictions that social networking sites (SNSs) such as Facebook and Twitter will transform democracy, allowing citizens and politicians to communicate, connect and interact in ways never before thought possible (Grant, Moon, & Busby Grant, 2010). In Barack Obama's Presidential campaigns in 2008 and 2012, over 100 staff members worked on Twitter outreach alone (@barackobama) (Sweet to Tweet, 2010; Campaigns Use, 2012). Current studies also show that the number of Japanese politicians using Twitter grew from three to 485 in under a year while in Germany, 577 politicians opened Twitter accounts (Hong & Nadler, 2011). Increasingly, politicians and elected officials are realizing the power of social media for communicating political information and interacting with citizens.

In considering the impact of social media in the political sphere, many researchers have explored how using SNSs such as Facebook and Twitter influences elections and public opinion poll results (Robertson, Vatrupu, & Medina, 2009; Hong & Nadler, 2011; Tumasjan, Sprenger, Sandner, & Welp, 2010; O'Conner, Balasubramanian, Rouledge, & Smith, 2010). These studies examine cases from the United States and European countries including Germany and the Netherlands. However, few studies from Asian countries were conducted. A case study on South Korea reviewed the evolution of hyperlinked networks, but did not explore use of Twitter during elections (Hsu & Park, 2011). Findings and insights from empirical studies for different countries are needed to understand how citizens and politicians worldwide share political information and opinion via social media. This study investigates in particular political information sharing in South Korea using Twitter.

Research Purpose

The purposes of conducting this study are twofold: to explore re-tweeting (RT) information behavior for political messages on Twitter by citizens in South Korea, and to compare the number and types of re-tweeted messages, and the sentiments captured from these messages with the results of public opinion polling about leading political figures. This research allows us to better understand the role of Twitter within citizens' political information sharing in South Korea, and offers insights into relationships between the message types and citizens' sentiments as expressed on Twitter and in public opinion polls.

Significance of Study

2012 was a significant year in South Korea. With a general election in April 2012 and the presidential election in December 2012, *change of regime* issues engaged citizens in South Korea. The National Election Commission in South Korea lifted a strict ban on using social networking sites (SNSs) such as Facebook and Twitter for election campaigns on the basis that using social media could broaden citizens' obtaining and sharing of information on candidates and elections at lower cost. With increased enthusiasm regarding this change in South Korea, the April 2012 election was considered as an important testbed for the role of social networking sites (SNSs) in political communication.

This paper examines how citizens in South Korea used Twitter in sharing political information and opinions on three candidates running in the April 2012 general election, and investigates relationships between citizens' use of Twitter and the results of public opinion polls. This study investigates South Korean citizens' re-tweeting of political messages about three candidates in the April 2012 election as community behavior indicating political sentiments, agreement, and consensus of political opinion. The study compares results of public opinion polling with citizens' re-tweeting of political messages in Twitter to provide insights into political information sharing in South Korea, and to add empirical findings in the growing body of research on social media use in political communication.

Backgrounds of Major Concepts and Literatures

Themes and research questions in this study concern the major concepts of Twitter as a social media tool, Diffusion of Innovation (DOI) theory and Re-tweeting (RT) in Twitter, and Use of Twitter and impact of on elections. Each will be examined for the further discussion.

Twitter as a Social Media Tool

Twitter, created in March 2006 and officially launched in July 2006, is a fast growing real-time social media tool allowing people to find and share information on what is happening worldwide (Chang, 2010). Twitter defines its service as "a real-time information network that connects you to the latest stories, idea, opinions and news." (Twitter, 2012). By January 2011, Twitter had over 200 million users, and by October 2011 was handling over 350 million tweets per day (Roosevelt, 2012; Twitter launched, 2011). Twitter's micro-blogging and messaging functionality has become a powerful tool for interpersonal, professional and academic communication (Java et al., 2007; Thomas, 2010; Dann, 2010).

Twitter messages allow a maximum length of 140 characters, and average 11 words per message (O'Connor et al., 2010). Messages, known as "tweets," can be made public or hidden, directed at another user by including the "@" symbol followed by another user's account name, i.e. @Friend_Username. Users can also share others' messages by "re-tweeting" (RT) them; which copies and disseminates the original message to the user's followers (Zhao & Rosson, 2009). Any message can be annotated with a topic or subject using hashtags, i.e. #Topic; clicking on or searching on a hashtag displays a choice of top tweets or all current tweets on Twitter that share the same hashtag. However, Twitter hashtags still suffer from their fragmentary and redundant nature (Chang, 2010). Therefore, this study excluded hashtag keywords (#Keywords) in the data collection process.

Diffusion of Innovation (DOI) Theory and Re-Tweeting (RT) in Twitter

Rogers (1995) defines diffusion of innovation (DOI) as "the process by which an innovation is communicated through certain channels over time among the members of a social system" (p.5). For this

study, DOI provides a strong theoretical background to explain the phenomenon of adoption of innovation of political information seeking and sharing via re-tweeting (RT) in Twitter.

According to Rogers (1995), an innovation can be any "idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p. 11). The diffusion process includes four key elements: innovation, the social system that the innovation affects, the communication channels of that social system, and time (Rogers 1962, 1995). The notion of innovation in DOI has also been expanded to include new products, ideas, services, methods, and inventions (Chang, 2010). In this study, the innovation diffused among users particularly refers to the *idea of seeking and sharing political information and opinions in Twitter through re-tweeting (RT)*.

Communication is defined as "the process in which participants create and share information with one another in order to reach a mutual understanding" (Rogers, 1995, p. 17). Mass media and interpersonal communications are two major communication channels in the dissemination process of innovation (Rogers, 1962, 1995). Growing use of Twitter through creating and re-tweeting Twitter messages on computers and mobile devices can be seen as a major new pattern of mass-communication (Zhao & Rosson, 2009). Dann (2010) emphasizes that Twitter has rapidly "evolved through user innovations with the re-tweet (RT), reply (@) and hashtag (#) marks being introduced by consensus and community behavior" (p. 1). Roosevelt (2012) also mentions that re-tweeted messages should be analyzed with weights since re-tweeting a message can be interpreted as agreement with that particular message, and can also spread the message faster and influence more users. Therefore, it is important to investigate the *users' re-tweeting (RT) as community behavior of agreement and consensus* in this study.

Rogers (1995) defines a social system as "a set of interrelated units engaged in joint problem solving to accomplish a common goal" and further denoted characteristics of social systems as: social norms, opinion leaders, change agents, and types of innovation decisions, which can promote or hinder the diffusion of innovations (p. 23). The time aspect is essential for explaining the innovation-decision process, the impact of innovators on adopters and the growth rates of adoptions (Chang, 2010; Rogers, 1962, 1995). In this study, the time aspect offers findings on offline public opinion polling results in comparison with online users' concurrent re-tweeting of political messages for a comparative view of reflected political opinion over time.

Use of Twitter and Impact on Elections

Many researchers have studied how use of particular social networking sites (SNSs) by politicians and citizens relates to results of public opinion polls and elections (Hong & Nadler, 2011; Tumasjan et al., 2010; O'Conner et al., 2010; Vergeer et al., 2011).

Tumasjan et al. (2010) argue that Twitter message content reflects the offline political landscape, thus potentially predicting actual election results. In a German case study, numbers of tweeted messages were observed to closely match ranking by share of the vote in election results, and nearly approximated results of traditional election polling. O'Connor et al. (2010) observed sentiments in Twitter messages replicated 2008-2009 U.S. consumer confidence and presidential job approval polls. Hong and Nadler (2011) studied U.S. politicians' use of Twitter and its impact on public opinion, finding that the impact of the number of tweets was not significant for any of the tested opinion polls. Vergeer et al. (2011) explored the relationship between using Twitter and gaining votes in the Netherlands. Although the study showed a positive relationship between the number of Twitter messages and the number of votes, the size of the Twitter network was noted to be a limited indicator for voting outcomes.

Although there is some research on Twitter in political communication and elections in the U.S., Germany, and the Netherlands, as yet little research has explored political use of Twitter within Asian countries. While studies have examined Twitter messages created by politicians and citizens which include particular keywords, there has not been research particularly examining re-tweeted messages.

Therefore, this research studying re-tweeted messages in political communication in South Korea provides new insights into citizens' use of Twitter in the context of the national elections in South Korea.

Research Questions

The goal of this study is to investigate relationships between citizens' use of Twitter and public opinion polls, and to answer questions on how re-tweeting (RT) behavior in Twitter political communication may relate to public opinion polls.

- RQ 1. Does the number of citizens' re-tweeted messages correlate with results of concurrent public polls?
- RQ 2. Do the types of citizens' re-tweeted messages correlate with the results of public opinion polls?
- RQ 3. Do the sentiments of citizens' re-tweeted messages correlate with the results of public opinion polls?

Methodology

The procedure for collecting re-tweeted messages used the Python Twitter API to collect and store re-tweeted messages in a designated database. Collected Twitter messages were limited to those including three selected keywords for a specific time frame: the names of three leading political figures in South Korea from April 2 to May 11, 2012. Among those, the most frequently re-tweeted 200 messages for each class of keywords were selected as data sets of 600 messages. For this study, a set of 120 messages (20%) including the top 40 messages for each keyword class per one week were created, and total of 720 messages for three keyword classes for six weeks were used for the analysis.

Data Capture and Storage

Python Twitter API was used to collect re-tweeted messages including three specific keywords related to the general election of April 11, 2012 in South Korea - the names of three leading political figures: Bak GeunHye, Moon JaeIn, and Ahn ChulSoo (hereafter, Bak, Moon, and Ahn). Initially, re-tweeted messages including the hashtags #BakGeunHye, #MoonJaeIn, and #AhnChulSoo were collected. However, hashtags with the keywords showed very low usage among Korean citizens, and the fragmentary and redundant nature of hashtags as mentioned in Chang (2012) was also observed. Therefore, this study excluded hashtag keywords in data collection. The chosen keywords were appropriate in that these three leading political figures were expected to be the major candidates for the presidential election in December 2012.

The Python Twitter API collected tweets which included any one of the three presidential candidate names, automatically calculated frequency of messages being re-tweeted every 100 seconds, sorted the top 200 messages based on the re-tweeting frequency, and stored them in the designated database. As the Python Twitter API only included Twitter messages for the most recent six days, collected data needed to be stored in a different database.

Twitter messages were collected every 5 weekdays for 6 consecutive weeks, the same period as the collection of public opinion poll results. To study the relationships between Twitter messages and public opinion polling, correlation analysis was conducted. An example of Twitter data collected from April 2 to May 11, 2012 is shown below in Figure 1.

Rank	Time(KOR)	Text	Count
1	2012-04-03 05:56:28-2012-04-07 03:21:47	RT @ahnsarang: 유시민씨가 얼마 전 "한나라당은 신이 내린 정당입니다"라고 했었죠. 아무리 부패해도, 표절해도, 사찰해도 국민의 1/3은 간고히 새누리당을 지지합니다. 이 구도를 될 방법은 오직 투표율을 높이는 방법 뿐. 4월 11일 반드시 투표해야 하는 이유입니다.	691
2	2012-04-03 16:58:49-2012-04-05 23:22:20	RT @moonriver365: 김제동씨 등 연예인 사찰...연예인을 사찰하고 방송 출연 못하게 했던 것은 5공 때나 있었던 일이지. 청와대와 새누리당은 연예인 사찰도 참여정부 때부터 해왔던 일이라고 우릴까요?	579
3	2012-04-07 04:18:51-2012-04-07 10:15:12	RT @osoo: 저 분을 신분이 의심스럽습니다. RT @jinjumong: @osoo 김용민박말은 새누리당의원생경계 연극에 비하면 아무것도 아니다 새누리당(한나라당)의 저질 연극연극 '황생경계' http://tt.co/IZVHPa 우한RT를 요청합니다	539
4	2012-04-03 16:34:35-2012-04-06 22:01:12	RT @congjee: 문대성후보, 표절은 도둑질입니다. 도둑은 자기가 나쁘다고 생각이라도 하겠지만 이질로 박사 교수까지 따고 그 이력으로 출마한거 정말 안됩니다. 새누리당도 모르쇠하는 것도 어이없고, 이렇게 양심불량인사람들 국회에 가서 여짜자는건지, 사퇴해야 합니다.	478
5	2012-04-06 09:41:23-2012-04-07 09:45:55	RT @sangjungsim: 새누리당 박근혜 비대위원장이 기초노원연급 인상을 철회했군요. 박 비대위원장의 노원연급 인상 약속 파기는 정략적 노인멸하입니다. 박근혜 비대위원장은 어르신들에게 공약 파기를 죽자 사회해야 합니다.	474
6	2012-04-02 22:46:23-2012-04-07 09:27:44	RT @patriamea: 부산 문물 꼭 보셔야 할 사항! @kwzodo '한나라당 부산 열역 5명이 해수부 폐지 등을 골자로 한 정부조직법 개정안을 공동 발의 (2008년 1월 21일) 그러나 이번 4.11 총선 부산 새누리당 후보들의 핵심공약이 '해양수산부 부활'	448

Figure 1. Screen Capture of Collected Data

Analysis

Re-tweeted messages including the three keywords (political leader names) were harvested from April 2 to May 11, 2012. A total of 556,675 messages including the keywords were re-tweeted. The most frequently re-tweeted 200 messages from each keyword class for each week were listed as a data set of 600 messages, totaling 3,600 messages for the three classes for 6 weeks. From among these, sets of 120 messages including the top 40 re-tweeted messages for the three classes weekly were accumulated for 6 weeks. A total of 720 re-tweeted messages were analyzed for this study.

Messages were analyzed using two types of content analysis: 1) classifying message types and 2) identifying the general sentiment expressed in the messages. The results of content analysis for re-tweeted messages were compared with the results of public opinion polls through correlation analysis using SPSS 18.

Classification of Types of Message

Twitter messages are created and re-tweeted with multiple purposes and motivations. The objective in classifying types of messages was to understand various purposes of citizens in their message-creating and re-tweeting behavior. Therefore, coding was conducted with a focus of studying the purposes as to why citizens created and re-tweeted particular messages.

Coding Scheme. The coding scheme was based on criteria from the literature. Pear Analytics (2009) categorized tweets as primarily related to *pointless babble* (40%), *conversational* (38%), *pass along* (9%), *self-promotion* (6%), *spam* (4) and *news* (4%). Content categories from Java et al.'s study (2007) included *daily chatter* on the daily routine of individual users, *conversations* which included replies to other users, *information or URL sharing* which were classified according to the presence of full length or shortened URLs, and *news sharing* which include sports, weather and commentary on current affairs. However, these studies classified types of Twitter messages, not types of purposes in creating messages. Robertson et al. (2009) examined linkage patterns of politically-oriented community networking on Facebook, classifying five types of linkage motivation patterns shown on three presidential candidates' Facebook walls: evidence, rebuttal, action, joking and ridicule, and direct address. This study showed purposes for which posters created postings to share political information and opinions through social networking sites of Twitter.

In addition to adapting these five types of messaging from Robertson et al. (2009), three more types were created reflecting characteristics of the collected data set: Media Report, Human Report and Event Report. Detailed explanations on codes and definitions are shown in Table 1 (hereafter, E, R, A, JR, DA, MR, HR, ER).

Reliability Testing. Content analysis requires that researchers make "replicable and valid inferences from texts to the contexts of their use" (Krippendorff, 2004, p. 18). Methodological requirements of reliability and validity are critical demands in content analysis. This research undertook codebook development for content analysis through multiple intra-coder reliability tests and an interactive sequence of codebook revisions.

Intra-coder reliability. The initial coding scheme was based on Robertson et al.'s study (2009), adding additional codes emerging from the data to create a preliminary codebook. With the initial codebook, 3 rounds (6 iterations) of intra-coder reliability testing were conducted for three weeks. A sample set of 60 messages was coded to classify message types with a second coding was carried out of 60 messages one week later. The differences between the two trials of coding came from 19 out of 60 messages (31.6%). After revising the definitions of codes based on the analysis, a third coding with a different set of 60 messages was conducted. A week later, the fourth coding was conducted and differences between the third and fourth trial were discovered in 8 out of 60 messages (13.3%). With a re-revised codebook, a fifth coding with a different set of 60 messages was accomplished. In the following week, a sixth coding was completed with 100% agreement and no differences found between the coding results.

Table 1
Codebook of 8 Message Types

Code	Sign	Definition
Evidence	E	Tweets written in order to <i>provide evidence for a particular opinion or simple fact</i> . It may include actual links to evidence such as newspaper, blog post, images, video and etc. along with the texts.
Rebuttal	R	Tweets written in order to provide negative responses or comments to rebut others' ideas. It may sound like evidence, but it has additional components of <i>reaction or providing negative comments to specific persons or organizations</i> .
Action	A	Tweets written in order to <i>encourage others to take actions either on Internet</i> such as participating in a poll, joining a group, <i>or in the real world</i> , such as donating money or attending a rally.
Joking and Ridicule	JR	Tweets written in order to <i>ridicule something or someone</i> , reveal something fully about them or their behavior, or simply point people to <i>satirical content</i> .
Direct Address	DA	Tweets written in order to directly address <i>his/her own opinion or simply provide fact without any evidence</i> . It may include simple statements of his/her idea, fact, and opinion such as simple <i>messages of support or lack of support</i> as an expression.
Media Report	MR	Tweets written in order to <i>cite and report the contents from a Media Report</i> . It is usually followed by <i>the actual links of the materials it refers</i> , or can be written <i>in the direct quotation form</i> (" "). Even though the message itself includes neither links to the material it refers nor texts in direct quotation form, an indirect way of citation and reports of the contents from Media can be coded as MR. The main purpose of MR is to distribute and circulate the content of media.
Human Report	HR	Tweets written in order to <i>cite and report the contents from a Human's remarks</i> . It is usually followed by the actual link of the material it refers, or can be written in the direct quotation mark (" "). Even though the message itself includes neither links to the material it refers nor texts in direct quotation form, an indirect way of citation and reports of the contents from Human remarks can be coded as HR. The main purpose of HR is to distribute and circulate the content of Human remarks.
Event Report	ER	Tweets written in order to <i>report and distribute the real-time event news</i> , or his/her own experience of that event. The main purpose of this ER is to distribute the facts or news, which happen as real time events but are not likely to be broadcasted by mainstream media.

The main reason noted to cause inconsistencies was that multiple purposes could be employed in creating a message. For example, a tweet could encourage people to vote by citing a politically important figure's remarks. In this case, this message can be coded as Action (A) or Human Report (HR). Considering that ideas or opinions can be expressed using multiple methods including examples, rebutting, or making a joke, this ambiguity is intrinsic in interpreting human language itself. If one message includes characteristics of two types, it is possible that one main purpose could be understood better and stressed more over the other within the specific context so that each message can be coded as one specific type.

General Sentiment Identification

Tweeted messages include special features and conventions such as words, emoticons, and hashtags representing the author's emotions and feelings within 140 characters (Agarwal et al., 2011; Tumasjan et al., 2010). These Twitter-specific features allow researchers to identify sentiments and opinions in tweeted messages. Sentiment analysis studies in Twitter are mainly conducted at the word level (Agawal et al., 2011; Tumasjan et al., 2010; O'Connor et al., 2010). To measure the emotional

meaning or degree in words and texts, researchers used pre-defined dictionaries such as the Dictionary of Affect in Language (DAL) or other systems with embedded dictionaries such as Linguistic Inquiry and Word Count (LIWC), or OpinionFinder, classifying words according to three categories: positive, negative and neutral. This approach enables researchers to measure sentiments at the word level, and research has indicated that the sentiment word frequency in tweeted messages correlates with the public opinion polling results (O'Connor et al., 2010).

However, it is still difficult to analyze sentiments or opinions from tweeted messages at the message level. Branthwaite and Patterson (2011) argue that "opinions are divided on the ability to code the underlying sentiments as positive or negative" and it is certainly difficult for machines to understand irony or sarcasm (p. 432). Twitter features such as emoticons and hashtags add variations, and thus sentiment analysis requires human intervention and qualitative analysis to increase validity and accuracy at the message level (Branthwaite & Patterson, 2011).

This study attempts to identify general sentiments understood and interpreted within political contexts. General sentiments within re-tweeted messages were sorted into three classes: positive, negative and neutral. Sentiments expressed toward the three leading political figures can be interpreted as an indicator of how Twitter users perceived those leaders. Overall sentiments were identified considering the context in the messages, not by particular words or terms. Intra-coder reliability was tested through the two trials of coding and agreement without error was relatively easily reached.

Correlation with Public Opinion Polls

One of the most frequently referenced public opinion polls in South Korea is *Korean Gallup* (<http://www.gallup.co.kr>). Korean Gallup polls are administered through telephone and mobile phone interviews to 1550 citizens every five weekdays. Favorable impressions of three leading political figures were collected from Korean Gallup polls for six weeks from April 2 to May 11, 2012. Correlation analysis was conducted to study the relationship between citizens' expressed opinions online in Twitter and public opinion polling offline.

Findings

In answering the three research questions, general descriptive analyses including the magnitude of re-tweeted messages, classification of types of messages, and identification of sentiments were provided, with correlation analysis between results of content analyses of Twitter messages and the results of public opinion polls.

RQ 1. Magnitude of Re-tweeted Messages and Public Opinion Polls

A total of 556,675 re-tweeted messages including the three keywords were collected for six weeks - April 2 to May 11, 2012. The total number of re-tweeted messages including Bak for six weeks was 354,284 (63.64%), with 102,000 (18.32%) for Moon and 100,391 (18.03%) for Ahn (Figure 2). The number of top 40 most frequently re-tweeted messages including Bak for six weeks was 56,281 (53.69%), 24,880 (23.73%) for Moon and 23,674 (22.58%) for Ahn (Figure 3).

The magnitude of the re-tweeted messages including the three keywords varied, and might depend on public awareness of the three leaders. Bak is a leader of the ruling party of Saenuri, and a daughter of the former president during the military dictatorship from 1960s through 1970s, JungHee Bak. Moon is the leader of the biggest opposition party of United MinJoo, and is well known as the successor of the former president, MooHyun Roh. Ahn is a non-partisan university professor without any official experience as a politician. The magnitudes of both total re-tweeted messages and top 40 re-tweeted messages showed one pattern: Bak was most frequently talked about by Twitter users, followed by Moon and Ahn.

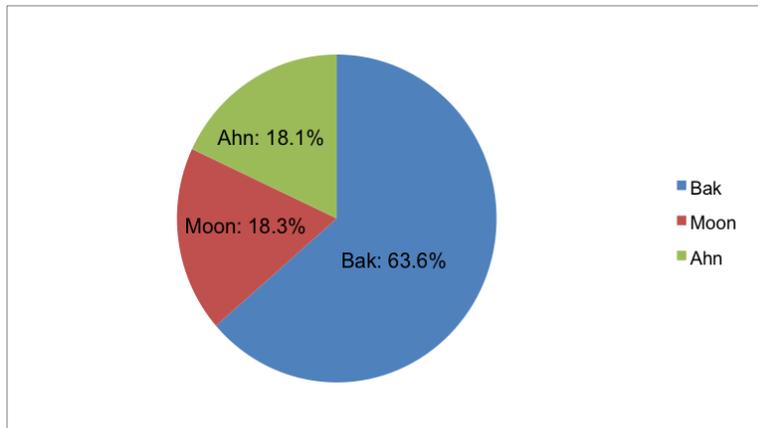


Figure 2. Magnitude of Entire Re-tweeted Messages

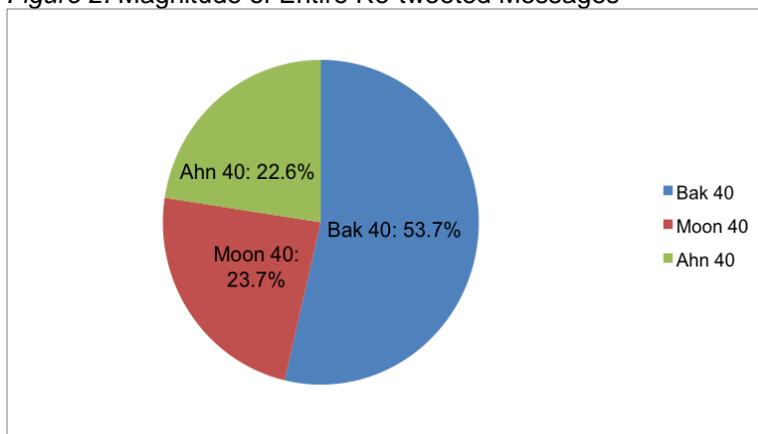


Figure 3. Magnitude of Top 40 Re-tweeted Messages

The extent to which the three leaders appeared in the Twitter timeline over the six weeks was not consistent with the percentages shown in public opinion polling results over the same period. Results from public opinion polls overall, as seen in Table 2, indicated that Bak obtained the most favorable impressions (averaging 36.83%) followed by Ahn (averaging 23%), and Moon (averaging 12.33%).

Table 2
Results of Public Opinion Polls for 6 Weeks

	Week1	Week2	Week3	Week4	Week5	Week6	Average
BAK	34	36	39	36	38	38	36.83 (%)
MOON	15	13	11	13	11	11	12.33 (%)
AHN	23	23	25	23	23	21	23 (%)

Notes. Public opinion polls results from nationwide data collection.

For correlation analysis, SPSS 18 was used. Total numbers of re-tweeted messages and of the top 40 most re-tweeted messages showed significant correlations with the percentages of public opinion polls (Table 3).

Table 3
Pearson Correlation between Re-tweeted Message and Public Opinion Poll

	RT_all	RT_40	Poll
RT_all		.985**	.676**
RT_40			.616**
Poll			

Notes. RT_all = number of all Re-Tweeted messages; RT_40 = number of top 40 Re-Tweeted messages; Poll = Public Opinion Polls percentage.

** $p < .01$

RQ 2. Classification of Messages Types and Public Opinion Polls

A total of 720 messages were classified into eight types according to the codebook. Direct Address (326, 45.3%), Human Report (142, 19.7%), Joking/Ridicule (113, 15.7%), and Evidence (54, 7.5%) were the most frequently assigned types (635/720, 88.2%) across all the re-tweeted messages. The results from classification of each type were shown in Figure 4 and 5.

- The main purposes of creating and re-tweeting were to directly address opinions and information without evidence (DA), distribute the cited contents from human remarks (HR), make jokes or ridicule others' ideas (JR) and provide evidence (E).
- Three categories, Media Report, Human Report, and Event Report (198/720, 27.5%), approximately one-fourth of messages, were created and re-tweeted to distribute and report cited contents from media, humans, and events. Media Report, Human Report, and Event Report represented the tendency of citizens to rely on other sources such as trustworthy media or leading figures' remarks. Evidence (54/720, 7.5%) indicated messages created based on evidence. In sum, political messages identified with types of Evidence, Media Report, Human Report, and Event Report (252/720, 35%) taken together demonstrated the tendency of citizens to circulate Twitter messages including supporting evidence.
- Rebuttal and Joking/Ridicule messages provided negative responses, comments, or jokes about others' ideas. Action messages sought to encourage people to take actions, and were usually created without including references or evidence. Direct Address messages directly address opinions and information without evidence. These four types of messages, Rebuttal, Action, Joking/Ridicule, and Direct Address (468/720, 65%), provided ideas or opinion without supporting evidence or references.
- Overall classification of message types found two major clusters with four types each: Supported (E, MR, HR, and ER; 35%) and Unsupported (R, A, JR, and DA; 65%). Supported message types included references to an information source such as a link, image, or remarks from particular media or figures, while unsupported message types did not.

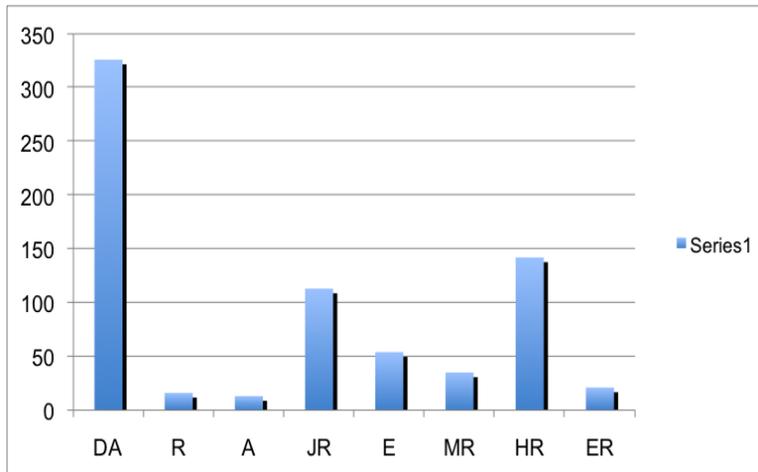


Figure 4. Results from Classification of Types of Messages

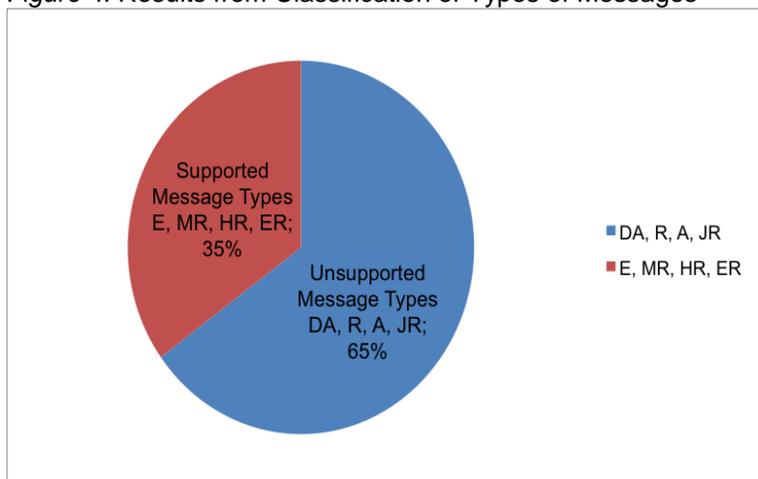


Figure 5. Clustering of Supported and Unsupported Message Types

Results from message type classification were shown in Table 4 and Figure 6. 240 tweeted messages for each political figure were analyzed.

- For Bak, Direct Address (114/240, 47.5%), Joking/Ridicule (50/240, 20.8%) and Evidence (27/240, 11.3%) were the most frequently classified message types. Supported message types (E, MR, HR, ER) were 65 (65/240, 27.1%) cases, while unsupported message types (DA, R, A, JR) were 175 (175/240, 72.9%) cases.
- For Moon, Direct Address (118/240, 49.2%), Joking/Ridicule (36/240, 15%) and Human Report (36/240, 15%) were the most frequently assigned types. As the senior advisor of the biggest opposition party against Saenuri Party, he is regarded as a progressive politician running against Bak. Supported message types (E, MR, HR, ER) were 81 (81/240, 33.7%) cases, while unsupported message types (DA, R, A, JR) were 159 (159/240, 66.3%) cases.
- For Ahn, Direct Address (94/240, 39.2%), Joking/Ridicule (27/240, 11.3%) and Human Report (86/240, 35.8%) were the most frequently assigned types. Unlike Bak and Moon, HR (86/240, 35.8%) was a dominant type for Ahn, and all of the HRs cite Ahn's own remarks. This suggests that Ahn's remarks were regarded as meaningful and influential among citizens, deserving repeated re-tweeting (RT). Even though unsupported types of DA, R, A, and JR (134/240, 55.8%) were observed more than supported types of E, MR, HR, and ER (106/240, 44.2%) cases, the contrast between two categories was more balanced than observed in the tweeted messages regarding Bak and Moon.

Table 4
Results from Classification of Message Types for Three Keywords

		DA	R	A	JR	E	MR	HR	ER	Sum
Bak	W1	20	0	4	5	8	1	2	0	40
	W2	21	1	4	6	6	0	2	0	40
	W3	20	2	0	7	7	2	1	1	40
	W4	18	0	0	7	4	5	4	2	40
	W5	15	0	0	11	2	5	6	1	40
	W6	20	0	0	14	0	1	5	0	40
Moon	W1	18	2	1	3	4	2	4	6	40
	W2	27	0	2	2	6	0	0	3	40
	W3	19	0	0	5	2	5	8	1	40
	W4	30	0	0	3	0	2	5	0	40
	W5	11	0	0	6	2	0	14	7	40
	W6	13	0	0	17	5	0	5	0	40
Ahn	W1	9	0	0	1	2	3	25	0	40
	W2	13	0	1	1	3	1	21	0	40
	W3	16	4	0	5	1	4	10	0	40
	W4	15	2	1	10	2	4	6	0	40
	W5	15	5	0	9	0	0	11	0	40
	W6	26	0	0	1	0	0	13	0	40
	Sum	326	16	13	113	54	35	142	21	720
	%	45.3	2.2	1.8	15.7	7.5	4.9	19.7	2.9	100.0

Notes. D = Direct Address; R = Rebuttal; A = Action; JR = Joking and Ridicule; E = Evidence; MR = Media Report; HR = Human Report; ER = Event Report. W1 means week 1 of 5 weekdays for which data collection is conducted.

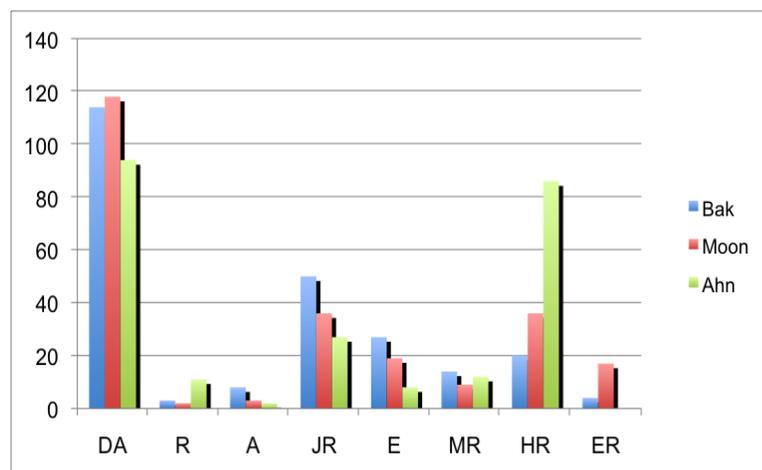


Figure 6. Classification of Messages for Three Classes

The eight types of Twitter messages did not show any correlation with results from public opinion polls (Table 5). Thus, re-tweeted (RT) messages as classified by the purpose of the message did not relate to the public opinion poll results.

Table 5
Pearson Correlations Among Types of Message and Public Opinion

	DA	R	A	JR	E	MR	HR	ER	Poll
DA		-.165	.230	-.274	.050	-.187	-.569*	-.122	-.023
R			-.104	.090	-.181	.018	-.041	-.085	.074
A				-.193	.666**	-.331	-.365	-.067	.232
JR					-.015	.029	-.373	-.177	.237
E						-.168	-.487*	.131	.233
MR							.004	-.098	.214
HR								-.084	-.233
ER									-.370
Poll									

Notes. DA = Direct Address; R = Rebuttal; A = Action; JR = Joking and Ridicule; E = Evidence; MR = Media Report; HR = Human Report; ER = Event Report.

* $p < .05$; ** $p < .01$.

RQ 3. General Sentiment Identification and Public Opinion Polls

Sentiments in tweeted messages were sorted into three classes: positive, negative and neutral. Sentiments were identified from the overall sentiment observed in messages, not by coding for particular terms, and are shown in Table 6 and Figure 7.

- For a total of 720 messages, 44.9% (323/720) negative sentiment, 46.1% (332/720) positive sentiment, and 9.0% (65/720) neutral sentiment were observed.
- For Bak, 98.3% (236/240) negative sentiment, and 1.67% (4/240) positive sentiment were identified. Tweets including Bak, an icon of conservatism, primarily showed negative sentiments.
- For Moon, 18.8% (45/240) negative sentiment, 75% (180/240) positive sentiment and 6.5% (15/240) neutral sentiment were observed. As a counterpart figure against Bak, tweeted messages showed mostly positive sentiments.
- For Ahn, 17.5% (42/240) negative sentiment, 61.7% (148/240) positive sentiment and 20.8% (50/240) neutral sentiment were found. Although overall sentiment toward Ahn was positive, a high portion of neutral sentiment was captured in tweeted messages.

Table 6
Results from Sentiments Identification

		NG	PS	NT	Sum
Bak	W1	39	1	0	40
	W2	40	0	0	40
	W3	40	0	0	40
	W4	40	0	0	40
	W5	38	2	0	40
	W6	39	1	0	40
Moon	W1	8	31	1	40
	W2	2	38	0	40
	W3	8	30	2	40

	W4	19	18	3	40
	W5	7	31	2	40
	W6	1	32	7	40
Ahn	W1	0	37	3	40
	W2	4	32	4	40
	W3	1	25	14	40
	W4	15	18	7	40
	W5	16	17	7	40
	W6	6	19	15	40
	Sum	332	332	65	720
	%	44.9	46.1	9.0	100.0

Notes. G = Negative Sentiment; PS = Positive Sentiment; NT = Neutral Sentiment.

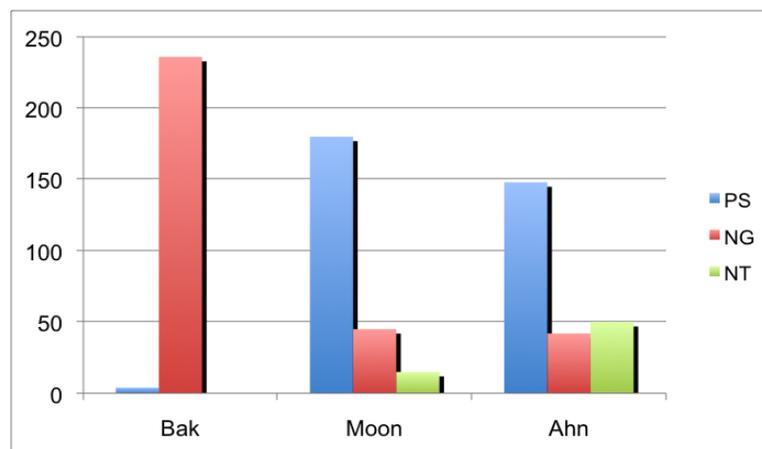


Figure 7. Results from Sentiment Identification

Positive and negative sentiment observed in re-tweeted messages showed significant correlation with percentages in public opinion polling; however, neutral sentiment did not show any correlation (Table 7).

Table 7
Pearson Correlations Among Sentiments and Public Opinion

	NG	PS	NT	Poll
NG		-.962**	-.555*	.842**
PS			.306	-.869**
NT				-.288
Poll				

Notes. G = Negative Sentiment; PS = Positive Sentiment; NT = Neutral Sentiment
* $p < .05$; * $p < .01$.

Conclusions

This study investigated political information sharing in social networking site of Twitter in South Korea. *Re-tweeting as community behavior of agreement and consensus in Twitter* was examined by both quantitative and qualitative approaches: counting the number (magnitude) of re-tweeted messages, using content analyses to classify types of messages and sentiments, and conducting correlation analysis between expressed opinions on Twitter and results of public opinion polls.

Although the magnitude of re-tweeted messages showed significant correlations with public opinion polling results over time, rankings of message magnitude differed from the rankings of public opinion polls. The magnitudes also appeared vulnerable to changes in political issues and events in real life. The 8 most frequently re-tweeted message types did not correlate with public opinion polling results. Contents in tweeted messages were highly subjective, complicated and contextual as a representation of political communication. However, sentiments in tweeted messages, while subjective and contextual as well, did show a correlation with public opinion polling results. This implies that capturing sentiments from tweeted messages dealing with broader political issues can be useful in gauging public opinion.

These findings require researchers to closely look at Twitter message content to understand purposes and underlying sentiments in context. In this perspective, the qualitative approach to classifying the purposes and sentiments of re-tweeted messages was appropriate in attempting to study this particular political communication phenomenon via social networking sites (SNSs). The significant correlations between the use of Twitter and public opinion polls indicate possible potential for utilizing current social networking sites (SNSs) to explore public opinion in political information sharing.

This study had several limitations. The sample used for this study may not be representative of South Korean electorate, as not everyone uses Twitter. Tweets collected for this study were also limited only to those including names of three top political leaders. Other specifically topic-based keywords such as titles of specific policy or political issues could be employed to collect targeted political data for future studies. This study also examined data for an abbreviated timeframe of six weeks. Therefore, further studies need to be conducted for a longer period of time in order to add more and stronger empirical findings.

References

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 30-38.
- Branthwaite, A., & Patterson, S. (2011). The power of qualitative research in the era of social media. *Qualitative Market Research: An International Journal*, 14(4), 430-440.
- Chang, H. C. (2010). A New Perspective on Twitter Hashtag Use: Diffusion of Innovation Theory. Paper presented at the Proceedings of the American Society for Information Science & Technology 2010, Pittsburgh, Pennsylvania.
- Campaigns Use Social Media to Lure younger Voters. (2012, October 8). *The New York Times*. Retrieved from <http://www.nytimes.com/2012/10/08/technology/campaigns-use-social-media-to-lure-younger-voters.html>
- Dann, S. (2010). Twitter content classification. *First Monday*, 15(12). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2745/2681>
- Grant, W. J., Moon, B. & Busby Grant, J. (2010). Digital Dialogue? Australian Politicians' Use of the Social Network Tool Twitter. *Australian Journal of Political Science*, 45(4), 579-604.
- Hong, S., & Nadler, D. (2011). *Does the early bird move the polls?: The use of the social media tool 'Twitter' by U.S. politicians and its impact on public opinion*. Paper presented at the Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, Maryland.
- Hsu, C., & Park, H. W. (2011). Sociology of Hyperlink Networks of Web 1.0, Web 2.0 and Twitter: A Case Study of South Korea. *Social Science Computer Review*, 29(3), 354-368.
- Huberman, B., Romero, D., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday*, 14(1). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/2317/2063>

-
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). *Why we twitter: understanding microblogging usage and communities*. Paper presented at the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis, San Jose, California.
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 122-129. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536/1842>
- Pear Analytics (2009). Twitter study. Retrieved from <http://www.pearanalytics.com/wp-content/uploads/2009/08/Twitter-Study-August-2009.pdf>
- Robertson, S., Vatraru, R. K., & Medina, R. (2009). The social life of social networks: Facebook linkage patterns in the 2008 U.S. Presidential election. *Proceedings of the 10th Annual International Conference on Digital Government Research, Peubla, Mexico*, 6-15.
- Roosevelt, C. M. (2012). Social Media Analytics: Data Mining Applied to Insurance Twitter Posts. *Casualty Actuarial Society E-Forum*, 2, 1-36.
- Rogers, E. M. (1962). *Diffusion of Innovations* (1st ed.). New York: Free Press.
- Rogers, E. M. (1995). *Diffusion of Innovations* (4th ed.). New York: Free Press.
- Sweet to tweet. (2010, May 6th). *The Economist*. Retrieved from <http://www.economist.com/node/16056612>
- Thomas, L. (2010). Twitter at the office. Social eyes. *Journal of Web Librarianship*, 4(1), 79-82.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>
- Twitter. (2012). *About Twitter*. Retrieved from <http://twitter.com/about>
- Twitter, launched five years ago, delivers 350 billion tweets a day. (2011, July 15). *L.A. Times*. Retrieved from <http://latimesblogs.latimes.com/technology/2011/07/twitter-delivers-350-billion-tweets-a-day.html>
- Vergeer, M., Hermans, L., & Sams, S. (2011). Is the voter only a tweet away? Micro-blogging during the 2009 European Parliament election campaign in the Netherlands. *First Monday*, 16(8). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3540/3026>
- Zhao, D., & Rosson, M. B. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. *Proceedings of the ACM 2009 International Conference on Supporting Group Work* (pp. 243-252), Sanibel Island, Florida, USA, 243-252. doi:10.1145/1531674.1531710