Mysterious Influential Users in Political Communication on Twitter: Users' Occupation Information and Its Impact on Retweetability

Jisue Lee, Florida State University
Jae-wook Ahn, IBM TJ Watson Research Center
Jung Sun Oh, University of Pittsburgh
Hohyon Ryu, Twitter Inc.

Abstract
This study attempts to examine the effect of user's self-disclosed identification to measure his influence and activity on Twitter. By looking at the most frequently shared top 1076 tweets written by 250 users during the 2012 presidential election campaign South Korea, we particularly examine the relation between user's occupation information and the measures of his 'influence': the number of followers and number of retweets by others. Influential users in South Korean political communication network on Twitter are classified as one group with self-disclosed occupation information and the other group without that information. User's occupation information clearly shows the impact on the number of followers for both groups. On the other hand, user group without self-disclosed occupation information has a higher level of producing influential political tweets and wide retweetability over the other group, regardless the small number of followers. It suggests that further study needs to identify other variables that may influence particular user or tweet's retweetability as an indicator of influence.
Keywords: Political Communication; Twitter; Retweet; Social Media; Influential User
Citation: Lee, J., Ahn, J., Oh, J.S., Ryu, H. (2015). Mysterious Influential Users in Political Communication on Twitter: Users' Occupation Information and Its Impact on Retweetability. In *iConference 2015 Proceedings*.

Research Data: In case you want to publish research data please contact the editor.
Contact: jl10n@my.fsu.edu, jaewook.ahn@gmail.com, jsoh@pitt.edu, hohyonryu@twitter.com

1 Introduction

Social media are increasingly considered as politically transformative communication technologies allowing citizens and politicians to easily connect, communicate, and interact with one another (Grant et al., 2010; Chadwick, 2006). Various social media platforms (e.g., Facebook, Twitter, YouTube) shaped citizens' political communication on political events and issues, and increased participation in social movements in the U.S. and around the world (Robertson et al., 2009; Hong & Nadler, 2011; Tumasjan et al., 2010; O'Connor et al., 2010; Younus et al., 2011). Among others, a microblogging service, Twitter, has become an influential medium for individual users to quickly engage in a variety of political communication activities.

The increased use of social media as a venue for political communication and participation casts a new light on a topic of long-standing interests in communications studies — opinion leaders. Opinion leaders are defined as "men who exert personal influence upon a certain number of other people in certain situations" (Merton, 1957, p. 410), especially by generating, filtering, or otherwise controlling the flow of information pertinent to others' decision-making. While the persistent interest in the idea of opinion leaders or influentials is evident in a large number of empirical studies investigating various issues related to this concept (e.g. the role of interpersonal influences in political decision making, traits of opinion leaders vs. followers, etc.), the problem of identifying and characterizing opinion leaders still remains elusive (Katz, 1957), partly due to the reflexive nature of evaluation of influence. Unlike offline settings, social media allows researchers to observe actual flows of political communication and different roles individual users play therein. Such observability may bring new insights into opinion leadership in general and traits of influential users in social media in particular.

In this study, we take up such opportunities and study 'influential' users on a particular social media. More specifically, we look at the usage of Twitter in a particular setting of political communication, the presidential election in South Korea in 2012, focusing on the problem of identifying and characterizing influential users (i.e., opinion leaders). As individuals' political information seeking and sharing activities particularly reached the peak during the election campaigns, elections are generally considered as important political opportunities for citizens' collective political deliberation (Huckfeldt et al, 2004), and thereby become an ideal setting for studying political communication at a large scale. In our view, the 2012 presidential election in South Korea holds additional merits for studying political communication on

Twitter. First, South Korea is one of the countries with the highest technology penetration rates —Internet (84%), mobile devices (110%), and social media (74%) (Global digital statistics, 2014). Twitter is reportedly used by 56% of Internet/mobile users in South Korea, making it the second most popular social media service following Facebook. Second, the South Korean government lifted its ban on social media use for campaigns early in 2012, the overall usage of social media was increased during General election in April and Presidential election in December, 2012 (National Election Commission, 2013). All in all, these facts point to the significant position Twitter had in the overall political communication taken place during the election.

For this study, we collected over 4 million tweet messages that contain certain keywords related to the presidential election and were retweeted at least once. We then selected the most frequently retweeted messages during a 25 day period before the election. Our analysis of influential users and their use of Twitter is based on this sample of highly retweeted messages.

1.1 Research Questions

Previous studies focus on the number of followers, retweets, and mentions to measure specific Twitter user' influence (Cha et al., 2010). While these measures based on Twitter network and communication features are certainly useful, they do not consider traits of the actors (users) themselves in the network. Some studies use certain characteristics or categories of users (e.g. celebrities) in tandem with these measures and see whether and to what extent these measures correlate with such user characteristics (Page, 2012). In this study, we are interested in the effect of self-disclosure of personal details, especially information on their occupation. More specifically, this study examines how influential users' self-disclosed occupation information correlate with the well known measures of influences, 1) number of followers and 2) number of retweets by others.

RQ1: Does a user's occupation information correlate with the number of followers?

RQ2: Does a user's occupation information correlate with the number of retweets by others?

2 Selected Related Literature

2.1 Political Communication on Twitter

Recently, increasing number of politicians and citizens have rapidly adopted social networking services and Twitter for their political communication during elections. In the United States—having the most Twitter users with 51% of the total—55% of the entire voting age population used the Internet and social media tools in order to seek and share political information and opinions with others since the 2008 presidential election (Smith, 2009). Rainie et al. (2012) also shows that 66% of social media users—or 39% of all American adults—engaged in multiple political activities during elections: liking or promoting politics or election related material (38%); encouraging people to vote (35%); posting their political thoughts or comments regarding political issues (34%); reposting politics and election related content created by others (33%); posting links to political stores or news articles (28%); and following elected officials and candidates (20%).

Previous studies of politicians' Twitter use show that they mainly use Twitter to publicize political events and articulate policy positions (Shogan, 2010), to disseminate information for self-promotion (Goldbeck et al., 2010), and to communicate with citizens on a more limited basis (Shogan, 2010). Lawless (2012) also claims that politician's major purposes of using Twitter include advertising (offering information on themselves or political events), position-taking (including explicit political positions), and credit-claiming (highlighting their accomplishments). Parmelee and Bichard (2012) show that citizens' primary motivations to follow political leaders in Twitter include social utility (to assist with social interactions with others), entertainment (to amuse and relax), self-expression (to express and communicate personal opinions), information seeking (to keep them knowledgeable and up-to-date on issues) and guidance (to help guide a decision), and convenience (to obtain information easy to access).

Regarding individuals' Twitter use during elections, research has been conducted to study the relations between individuals' Twitter use and actual election or public opinion polling results from many countries. Tumasjan et al. (2010) argues that Twitter is used as a platform for political deliberation and reflection of political sentiment as well as a predictor of the offline election results. While some studies demonstrate correlations between Twitter use and election/public opinion polling results (Tumasjan et al., 2010; O'Connor et al., 2010; Skoric et al., 2012; Soler et al., 2012; Lee et al., 2013), others claim that volume of tweets or sentiments from tweets do not have any predictive power for offline elections and public opinions (Bravo-Marquez et al., 2010; Hong & Nadler, 2010; Jungherr et al., 2012). It is notable that various contextual variables regarding data collection (e.g., data collection period, use of keywords for filtering, level of Twitter adoption, media freedom, competitiveness of elections, and other political conditions) may yield different estimation of tweet volume and prediction of elections or public opinion polls (Jungherr et al., 2012; Skoric et al, 2012).

As traditional political communication pattern does, online political communication exhibits political homogeneity and polarization within networks (Sunstein, 2007; Adamic & Glance, 2005; Yardi & boyd, 2010; Conover et al., 2011; Himelboim et el., 2013). Studies demonstrate that political blogs and websites commonly link to others sharing the same political ideology (Sunstein, 2007; Adamic & Glance, 2005). Studies of following, mention, and retweeting networks on Twitter show distinct clustering and information behavior patterns according to political orientation (Conover et al., 2011; Himelboim et al., 2013). They claim that citizen users from different types of communication networks on Twitter engage in dissimilar information behaviors according to their political orientations and selective exposure to certain information sources. Metaxas and Mustafajaj (2010) also find that ideologically segregated groups engage in exclusively retweeting information supporting their own political leanings.

2.2 Information Sharing on Twitter: Retweeting

Retweeting (RTing), the practice of forwarding original tweets to other users, became a unique and established many to many communication convention on Twitter. Users can retweet a tweet by copying the message, typically adding a text indicator RT followed by the original author in @username or click the RT button that Twitter provides. Users also can include additional content or slightly edit the original content when retweeting it. Page (2012) defines this collaborative practice of allowing users to involve themselves in the act of discourse creation as 'co-tellership.' The practice of retweeting can be considered as community behavior of public claim of agreement or consensus (Dann, 2010; boyd et al., 2010; Roosevelt, 2012).

Retweeting particularly well serves information diffusion in that a tweet is quickly propagated to new sets of audiences, the followers of the retweeters (Suh et al., 2010; Zhao & Rosson, 2009). This unique practice of retweeting creates new mass-communication patterns as user innovations (Zhao & Rossson, 2009; Dann, 2010; Suh et al., 2010). Particularly, Twitter's rapid many-to-many communication capacity has been dubbed as an electric form of 'word of mouth' (Jansen et al., 2009). Hashtag inclusion, URLs linkage, number of the followers, use of sentiment-related words within the tweets and others are found as the factors affecting retweetability (Suh et al., 2010; Stieglitz & Dang-Xuan, 2012; Hoang et al., 2011).

In the political sphere, the quickly shared retweets, as a many-to-many communication convention, are critical in terms of rapidly increasing citizens' political communication and deliberation with others (e.g., free and open discussion, information seeking, sharing, and exchanges). Information diffusion via retweeting may greatly influence when individual users form political opinions and decision-making for political changes (Papacharissi, 2002; Dahlgren, 2005; Shirky, 2011). This paper particularly examines the most frequently shared influential political retweets and the authors of those tweets during 2012 presidential election campaign in South Korea.

2.3 Influential Users on Twitter

Twitter, a network of *who follows and listens to whom*, is suited to studying influential users (Bakshy et al., 2011). Even though common measures of influence on Twitter include number of followers, number of retweets, and number of mentions, the most followed users do not necessarily have the highest number of retweets or mentions (Cha et al., 2011). Celebrities and other popular public figures (e.g., media representatives, subject matter experts, journalists) usually have large numbers of followers and small numbers of friends, and tend to post their own updates instead of retweeting others' tweets (Page, 2012).

In political Twittersphere, celebrities may include both real world celebrities and Twitter celebrities who are laypeople but serve as influential users (e.g., opinion leaders, gate keepers, etc.) who provide quality political information and opinion regarding political issues to others (Parmelee & Bichard, 2012). Definition of influential users may vary depending the roles they take within the network. In this paper, we consider those who created messages that were widely shared via retweeting as influential users. We particularly study how influential users' occupation information (described in their profiles) correlates with their activities and reputation on Twitter.

Occupation has been considered a factor that measures social status of people (Hollingshead, 1975). Despite the popularity of demographic variables such as occupation to help understanding social behavior, few studies correlated user occupation and their online social network behavior, particularly online political communication. Hughes et al. (2012) emphasized that personality as well as demographic variables including user occupations were related online socializing and information seeking and exchange. Ajrouch et al. (2005) examined the main interactive effects of age and socioeconomic status (SES) on social networks. Ames et al. (2011) found parents' different socioeconomic classes had different values and practices around children' technology use including mobile phones. Even though personal characteristics such as sex, age, appearance, and occupation are often blurred, indiscernible, or faked in virtual settings (Rohde, 2004; Donath, 1998; Preece, 2000), occupation is still an important factor to better understand the nature of online social communication such as trust/community building mechanism among users as long as we can acquire the correct information.

2.4 Political Contexts of 2012 Presidential Election in South Korea

This study examines the data from the 2012 presidential election in South Korea. Three major candidates initially ran for the presidential election: Geun-hye Park, Jae-in Moon, and Cheol-soo Ahn (hereafter, Park, Moon, and Ahn). Park was the leader of the ruling conservative party, Saeuri Party. She is the eldest daughter of Dictator Jung-hee Park (who seized the power in a military coup in 1961, reigned over people for 19 years). Moon was the leader of the largest opposition liberal party, MinJoo Party. Moon was a prime minister during Moo-hyun Noh's presidency (the second liberal administration followed by Dae-jung Kim's administration). Ahn, a former physician and CEO of the biggest security solution provider (AhnLab), stood as a non-partisan independent candidate. After the two liberal administrations (1998-2007), Myung-bak Lee from conservative Saenuri party sat in presidency from 2008 to 2012. These three candidates (Park, Moon, and Ahn) initially ran for the presidency after Lee's, but Ahn withdrew his candidacy on the night of November 23 (twenty-six days prior to Election Day, December 19, 2012) expressing his support for Moon. The 2012 presidential election was basically a contest between Park and Moon.

3 Methods

3.1 Tweet Data Collection

In order to answer the research questions, we collected the most frequently retweeted political messages related to the presidential election from the official website of Twitter, using the names of three presidential candidates as keywords: Geun-hye Park, Jae-in Moon, and Cheol-soo Ahn (hereafter, Park, Moon, and Ahn). Even though Ahn gave up the candidacy in order to unite with Moon against Park, he was still an influential figure who received considerable advocacy from citizen supporters and likely to be mentioned in messages related to the election. Therefore, we decided to include his name as a keyword. Even though the candidates' names as keywords cannot include all the tweets discussing the presidential elections, they were considered as the most important keywords regarding the election. In our initial attempt, retweets including the hashtags #Geun-hyePark, #Jae-inMoon, and #Cheol-sooAhn were collected, but these hashtags with keywords showed very low usage among South Korean citizens. In addition, the fragmentary and redundant nature of using hashtags (Chang, 2012) was also observed from the patterns of South Korean Twitter users. We decided to exclude hashtags with keywords from isolating tweet data for this study.

Data collection spanned a period of 25 days from November 24 to December 18, 2012 (the day before the Election Day)—which included the official election campaign period of 22 days (November 27 to December 18). The Python Twitter API named Twython (https://github.com/ryanmcgrath/twython) was used to access Twitter REST Search API (https://dev.twitter.com/docs/api/1.1/get/search/tweets). In order to collect relevant Twitter messages as comprehensively as possible, the API search was performed every 10 seconds and the result was aggregated into a designated database. More than 4 million tweets were downloaded, with the rate of 117 tweets per minute.

For the analysis of this study, the 4 million messages were sorted by the frequency of retweets (RTs). The top 400 messages for each of three keywords were selected as a sample for analysis (the cutoff point for creating a sample set was at 10% of the number of summed frequencies of the entire set of retweets), resulting in a sample set of 1200 tweets. These 1200 messages (created by 277 unique individual users) are considered as highly influential political tweets: they account for the 10% of the entire retweets during the data collection period.

The additional information about users (users' profiles, number of followers, following/friends, number of tweets, dates of membership registration, and last activity on Twitter) was collected in May 2013 and April 2014. Due to the time lag between the main data collection (for Twitter messages) and the additional ones (for user profiles), specific numbers about individual users (e.g. the number of followers) might be changed from the time of message creation. However, a comparison of data collected in 2013 and 2014 has shown that there were few notable differences in relative sizes of those numbers. Rather, two additional data collections after the election campaign over the next two years allowed us to examine these influential users' post-election activities.

3.2 Content Analysis of User Profiles

To identify an individual user's occupation, we analyzed the user's self-disclosed description from user profile. The 277 unique users created those 1200 tweets (of a sample set) during the 25 days prior to the Election day. 27 user accounts and their tweets were excluded from the sample set, since they were unavailable at the point of data analysis (e.g., their accounts were neither available nor retrievable).

Some users explicitly stated their professional occupations (e.g., professor teaching sociology at A university, lawyer at B law firm, poet, doctor at C hospital, or news reporter/anchor at D newspaper company). Others did not offer specific occupational information; but they provided some information about their political orientations, hobbies, and or random personal interests. Based on this self-disclosed occupation information, we identified 250 influential users' occupations. Some users listed multiple occupations in their profiles. We used people search service in one of the most popular portals in South Korea (e.g., people.search.naver.com) and Google search (google.co.kr) to verify and assign the user's most current appropriate occupation.

Those who provide their occupation information are coded as self-disclosed user; those who did not are coded as un-disclosed user; and one government account at a macro level (Table 1). Of course, there may be a lot of users who decide not to reveal the occupational information on Twitter. However, in this study we classified influential user groups simply based on their self-disclosed profile texts. Table 2 shows three user groups' (self-disclosed, un-disclosed, and government account) activity profiles in terms of number of members, number of tweets of a sample set, and number of times their tweets were retweeted. Self-disclosed users include six sub groups with professional occupations; un-disclosed users—who did not provide their occupation information in their profiles—consist of three sub groups according to their post-election activity pattern (Table 3).

Micro	User's Occupation Information	N	Macro Type
AR	Artist (e.g., poet, writer, singer, composer, movie star, etc.)	12	Self-disclosed User
ED	Educator (e.g., university professors, researchers, etc.)	14	Self-disclosed User
LW	Lawyer (e.g., lawyer, judge, etc.)	2	Self-disclosed User
MD	Medical representative (e.g., surgeon, physicist, psychiatrist, etc.)	2	Self-disclosed User
MJ	Media representative (e.g., news reporter, anchor, etc.)	12	Self-disclosed User
PL	Politician (e.g., politician, election campaign staff, etc.)	19	Self-disclosed User
CT	Ordinary users who did not provide occupational information	188	Un-disclosed User
GA	Official Twitter account of South Korea	1	Government Account
Total		250	

Table 1. User's Social Types at Micro and Macro Levels

Macro Type	N of Users	N of Top Tweets	Avg. N of Top Tweets	N of Retweets	Avg. N of Retweets
Self-disclosed User	61 (24.4%)	244 (22.7%)	4	121,855	1,998.11
Un-disclosed User	188 (75.2%)	830 (77.1%)	4.41	373,211	1,985.16
Government Account (GA)	1 (0.4%)	2 (0.1%)	2	1,509	1,509
Total	250	1076	4.3	496,605	1,986.42

Table 2. Activity Profiles of Three User Groups

4 Findings

4.1 Descriptive Analysis

The sample set of the most frequently shared retweets (10% of the entire retweets during the data collection period) for 25 days prior to the Election day consists of 1200 messages (400 tweets from the 3 subsets including 3 keywords) written by 277 unique authors. We excluded the 27 user accounts and their tweets from the sample set, which were unavailable at the point of data analysis (e.g., their accounts were neither available nor retrievable) for this study. The sample data for this study includes a total of 1076 tweets, 250 unique authors, and at least 496,605 retweeters. The self-disclosed 61 users (24.4%) created the 244 most frequently shared tweets (22.7%) in the sample set, un-disclosed group of 188 users (75.2%) authored 830 tweets (77.1%), and only 2 retweets (0.1%) written by 1 government owned account (0.4%) (Table 2).

The main actors of South Korean political communication on Twitter consist of self-disclosed and un-disclosed individual users rather than official government or cooperation-related users. Un-disclosed user group created more than three quarters of most influential political information: self-disclosed users authored the one-quarter. It demonstrates that un-disclosed users engaged in authoring slightly more political tweets (average of 4.41) and their tweets were widely shared among others than self-disclosed user group did (average of 4). Regarding the number of retweets, this data set only captures the first round of retweeting by direct followers of the author. The top 244 tweets authored by self-disclosed user group were shared and distributed by at least 121,885 others; the top 830 tweets written by un-disclosed users were shared by at least 373,211 other users; and the top 2 tweets created by government user were shared by at least 1,509 other users (Table 2).

Except one government owned account (N=1), 249 users are classified as two groups (self-disclosed users who expressed their occupation information in their profiles and un-disclosed users who decided not to). Self-disclosed users (N=61) are grouped as 6 sub-groups according to their professional occupations. This self-disclosed group includes well-known public figures and celebrities—who are usually outliers having an extremely high number of followers (e.g., a popular writer with 1.7 million followers) compared to other users—as well as laypeople with professional occupations. Given having no occupational information in their profiles, un-disclosed users (N=188) are groups as three sub-groups according to the activity pattern at micro level (Table 3).

Macro	Micro	N of Users	N of Top Tweets	Ave. N of Top Tweets	N of Retweets	Ave. N of Retweets	Ave. N of Followers	Ave. N of Friends	Ave. N of Entire Tweets
SD	AR	12	37	3.08	19,221	1,601.75	366,031	4,861	17,900
	ED	14	76	5.43	34,431	2,459.36	111,610	9,613	20,774
	LW	2	2	1	699	349.5	48,322	29,408	5,332
	MD	2	3	1.5	1,428	714	78,166	1,198	29,222
	MJ	12	43	3.58	23,687	1,973.92	188,737	21,858	20,684
	PL	19	83	4.37	42,419	2,232.58	122,812	42,058	3,642
		61	244	4	121,855	1,998.11	177,150	21,566	14,625
UD	CT	143	579	4.05	262,887	1,838.37	19,166	15,786	20,706
	CT-1	37	141	3.81	60,750	1,641.89	1,032	1,366	2,395
	CT-2	8	110	13.75	49,574	6,196.75	2,344	2,426	1,675
		188	830	4.41	373,211	1,985.16	14,882	12,397	16,293
GA	GOV	1	2	2	1,509	1,509	743,000	282	568
		250	1,076	4.3	496,605	1,986.42	57,388	14,572	15,823

Table 3. Influential Twitter Users' Activity Profiles

The most distinct additional characteristic classifying sub-groups among un-disclosed users is whether they keep their memberships and continue using Twitter after the election. CT is a group of un-disclosed users who are active on Twitter even after the presidential election (on December 19 in 2012); CT-1 is a group of un-disclosed users who stopped using Twitter during the week of Election day (from December 17 to 21, 2012); CT-2 is a group of un-disclosed users whose accounts are available but all the tweets were deleted. Their memberships might have been changed after the election, or kept their accounts as protected ones so that their tweets are no longer available to the general public. The 76.1% of un-disclosed users (CT; N=143), who created 579 top tweets (69.7%), kept their memberships and activities after the election. 262,887 other users shared their tweets. The 23.9% of un-disclosed users (CT-1 and CT-2; N=45) created 251 top tweets (30.2%), and they were shard by 110,324 users. These users either stopped Twitter or changed their memberships into protected ones during the week of the Election day.

4.1.1 Number of Friends & Number of Entire Tweets

The number of friends (a.k.a. following) is an indicator of an individual user's willingness to listen to, and receive information from others. There is no significant difference between the self-disclosed and un-disclosed users' number of friends. While most of the subgroups in self-disclosed users (e.g., public figures and celebrities) tend to have a much smaller number of friends compared to their number of followers, un-disclosed users have similar numbers of friends and followers. This is in line with Page's study of Twitter celebrities (2012): following others is of no interest to celebrities. The number of entire tweets is a measure of that user's regular tweeting activities since the date of joining Twitter. With no statistical difference between groups, self-disclosed and un-disclosed users show the similar degree of creating tweets: un-disclosed users' number of entire tweets (an average of 16,293) is little higher than that of self-disclosed users' (an average of 14,625).

4.1.2 Post-Election In-Activity

Most of CT-1 users (N of users=37, N of tweets=141) joined Twitter during the five months prior to the Election day—August, September, October, November in 2012—and stopped using Twitter during the week of Election day (from December 17 to 21, 2012). CT-2 users' accounts (N of users=8, N of tweets=110) were either owned by someone else after the Election day or changed into protected accounts. In other words, the one fourth of un-disclosed users (CT-1 and CT-2) who had created one third of top 1076 tweets during the election campaign (N of users=45, 23.9% of all un-disclosed user group, N of tweets=251, 30.2%) stopped Twitter activities or made their tweets private right after the Election day.

Contrary to this inorganic activity pattern of un-disclosed users, the activities of users in the self-disclosed group appear persistent. Most of the accounts created well before the election, and stayed active after the election. Only two user accounts of self-disclosed user group (N of users=61, N of tweets=244) created their accounts during the five months prior to the Election day. One (@saenuritalk) related to the ruling party's election campaign accounts was created on September 27, 2012. The other one (@mooncamp1219) relevant to opposing party's election campaign was created on November 11, 2012. However, all the previous tweets written by @mooncamp1219 was deleted and is currently owned by someone else. The official election campaign account for the candidate from ruling party (@at_pgh) officially stopped Twitter activity after the candidate won the election. Except these cases, all of the self-disclosed users continued using Twitter after the Election day.

4.2 Users' Occupation Information and the Number of Followers

In order to answer the first research question, we compared the average number of followers between two groups: Self-disclosed and un-disclosed user groups. There is a significant difference between the two groups (one-way ANOVA, $F(2, 247) = 42.752$, $p = .000$) (Table 4). The number of followers is a well-known measure of user influence given the nature of directional information flow (A -> B) when user B follows user A. The self-disclosed group's average number of followers (177,150) is much higher than that of non-professional group (14,882). Especially, the artist group including famous writers, musicians, movie stars, and more are easily followed by massive groups of individual users (366,031). The average of 188,737 users follows media representatives; 122,812 users depend on politicians; and 111,610 users follow educators. Contrary to the professional group, non-professional users have much less number of followers.

		Sum of Squares	df	Mean Square	F	Sig.
Number of Retweets	Between Groups	230665.96	2	115332.98	0.008	0.992
	Within Groups	3.66E+09	247	14821672.43		
	Total	3.66E+09	249			
Number of Followers	Between Groups	1.69E+12	2	8.42E+11	42.752	0
	Within Groups	4.87E+12	247	1.97E+10		
	Total	6.55E+12	249			

Table 4. Results of ANOVA Test

4.3 Users' Occupation Information and the Number of Retweets

In political communication networks on Twitter, retweeting is a stronger indicator of public claim of agreement or group behavior of consensus (boyd et al., 2010). The average number of retweets shows the importance and influence of certain user or his tweets. We found no significant difference between the self-disclosed and un-disclosed users (one-way ANOVA, $F(2,247)=0.008$, $p=0.992$). Our data show that two groups have almost the same average numbers of retweets regardless of different size of followers and friends: self-disclosed group has 1,198.11 and un-disclosed group has 1,985.16. This means that even though the un-disclosed users have much less number of followers, their tweets per user are more widely retweeted and shared by others to almost the same extent. Particularly, the CT-2 users (whose tweets were deleted or hidden to the public) authored 110 top tweets and they were retweeted by 6,196.75 other users.

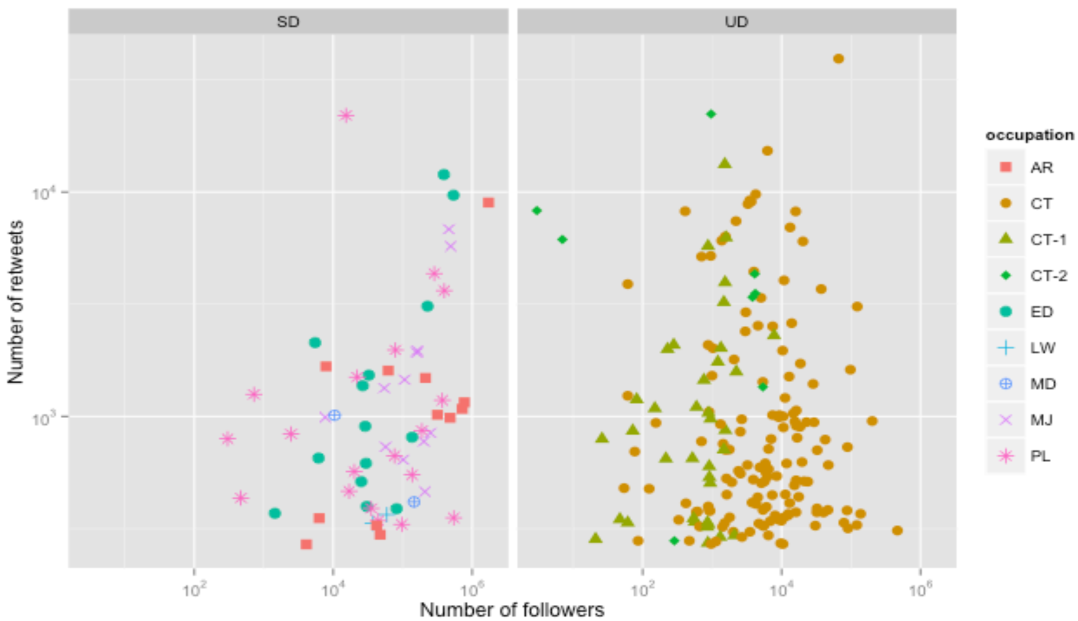

Figure 1. Comparison of number of followers (X-axis) and the number of retweets being shared by others (Y-axis). UD shows the log-log scatterplot of un-disclosed users (CT, CT-1, CT-2). SD shows self-disclosed users with professional occupations of AR, ED, LW, MD, MJ, and PL. We adopted the log-log scale for visual clarity.

Figure 1 visually compares the number of followers (X-axis) and the number of retweets (Y-axis) of subgroups with occupation information. As previously described, the two groups of self-disclosed (SD) and un-disclosed (UD) groups show the same number of retweets (R2), whereas the self-disclosed users have bigger numbers of followers than the un-disclosed users do (177,150 versus 14,882 in Table 3). Within the un-disclosed user group, CT (orange disk) and CT-1 (green triangle) are noticeably different. They have the same level of retweets by others (1,838 vs. 1,641 in Table 3) but CT-1 is clearly located on the left side of CT (less followers). CT-2 is located top-left position than the other two. They have higher numbers of retweets with much lower number of followers, even though we could not test their statistical significance.

From the subgroups of self-disclosed users, we see that the educator group's tweets are most widely shared and distributed, even though the artist group has the biggest followers. Regarding the political information and opinion, Twitter users more share more tweets written by educators, politicians, and media representatives rather than artists.

5 Discussion

This study attempts to examine the effect of user's self-disclosed occupation information to measure his/her influence on Twitter. By looking at the most frequently shared top 1076 retweets written by 250 unique users during the South Korean presidential election in 2012, we wanted to characterize these influential users, specifically what social types they consider themselves, and how the social factor of occupation impact their Twitter activities and reputation. More specifically, by using self-disclosed occupation information from the profiles we tested the well-known two measures of influence on Twitter: number of followers and number of retweets by others.

5.1 Effect of Self-disclosed and Un-disclosed Information

Using the descriptions in user profiles, we divided the influential users (those who created tweets that are frequently retweeted during the data collection period) into two groups – self-disclosed user group (24.4%) and un-disclosed group (75.2%). Self-disclosed user group was then grouped with 6 different professional occupations based on the occupation information from their profiles, while un-disclosed group was classified based on their activity patterns. During the election campaign, un-disclosed group authored 77.1% of most widely shared political 1076 retweets while the self-disclosed group authored only 22.7%. Given the rapid diffusion of political information and opinion among users through retweeting practice on Twitter during the election campaign, closer probes on who those un-disclosed users are and what they tweet about must be continued.

The number of friends, an indicator or a user's willingness to listen to and receive information from does not show any difference between two groups. The number of entire tweets also shows that both group regularly engaged in posting tweets. However, un-disclosed group shows inorganic information behaviors (eliminating Twitter memberships or hiding their tweets from the public right after the election) contrary to the self-disclosed user group's activities. CT-1 and CT-2 groups (one-quarter of un-disclosed users) joined Twitter during the five months prior to the Election day (December, 19, 2012) and stopped using Twitter during the week of the Election day or changed their tweet private after the election. This study could not explain the reasons behind the inorganic activities by un-disclosed user group. However, it will be an important further study to examine why this certain suspicious activity happens particularly to un-disclosed user group.

5.2 Users' Reputation and Influence

We found there is a significant difference between the two different groups' number of followers. Through the following activity on Twitter, users selectively choose their information sources (other users) and information quickly flows among users within networks. Therefore, the number of followers is commonly used to measure the user's influence to others. This study also shows that self-disclosed users (e.g., public figures and celebrities) have much more followers than un-disclosed users do. However, the un-disclosed user group in this study demonstrates their equivalent or even stronger influences based on the number of top tweets and number of retweets among others. Un-disclosed user group shows almost the same average number of retweets despite their distinctly small size of followers and friends. It means that there may be impacts or influences derived from other communication features in Twitter such as retrieving tweets in timeline and retweeting them or so. Further study needs to explain more about this phenomenon of information sharing and or influence exchanges.

6 Limitation

This research has a few limitations as well. First, this study relies on the data sample, which is restricted to the tweets including the names of three presidential candidates. The names of candidates as keywords are highly common and straightforward keywords relevant to the presidential elections. Many previous studies select the names of candidates or their political affiliated political party as keywords for collecting Twitter data. However, these keywords cannot capture all tweets discussing presidential elections that do not contain the candidates' names within them. Second, this study exclusively examined the most frequently shared retweets. Even though this data set has a merit of showing the most commonly and widely shared political information and retweeting activities, other types of communication such as mentions and replies could not be comprehensively covered. Further study needs to incorporate various methods collecting tweets with hashtags, keyword variations, and other types of tweets and use a bigger size of sample for the analysis.

7 Conclusion

This study explored the South Korean political communication network during the presidential election in 2012. Particularly, it was useful to study the influence of main user groups in the communication network from the perspective of occupation information reflecting their social types. It demonstrates that self-disclosed occupation information clearly shows the impact on the number of followers for both groups. On the other hand, un-disclosed users show higher levels of influence in terms of producing influential political tweets and its wide retweetability even without the large numbers of social contacts (followers). Further study needs to identify other variables that may influence particular user or tweet's retweetability as an indicator of influence.

After the 2012 presidential election in South Korea, it turned out that National Intelligence Service (NIS) of South Korea was accused of swaying the public opinions by posting 1.2 million tweets in favor of Geun-hye Park, leader of conservative ruling party. Due to this attempt, the credibility and trustworthiness of influential political tweets and retweets in this data set cannot be fully guaranteed. Therefore, further investigation identifying these mysterious influential user groups who did not disclose self-identification information and show inorganic activity pattern after the election must be conducted. The finings from this study are important and useful in that understanding the vulnerability of social media data and virtual public sphere—such as spamming attempts during election campaigns (Metaxas & Mustafaraj, 2010), which may be a critical threat to the increasing users engaging in virtual political communication and interactions in both politically stable and less stable countries.

References

Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery.*

Ajrouch, K. J., Blandon, A. Y., and Antonucci, T. C. (2005). Social Networks Among Men and Women: The Effects of Age and Socioeconomic Status. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(6):S311– S317.

Ames, M. G., Go, J., Kaye, J. J., and Spasojevic, M. (2011). Understanding technology choices and values through social class. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, page 55, New York, New York, USA. ACM Press.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Bravo-Marquez, F. B., Gayo-Avello, D. G., Mendoza, M., & Poblete, B. (2012). Opinion dynamics of elections in Twitter. *Proceedings of 2012 Eighth Latin American Web Congress.* doi:10.1109/LA-WEN.2012.11

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM*, *10*, 10-17.

Chadwick, A. (2006). *Internet politics: States, citizens, and new communication technologies*. New York: Oxford University Press.

Chang, H. C. (2010). A new perspective on Twitter hashtag use: Diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology 2010, 47*(1), 1-4.

Conover, M. D., Franciso, R. M., Goncalves, B., Flamminin, A., & Menczer, F. (2011). Political polarization on Twitter. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media.* Retrieved from http://truthy.indiana.edu/site_media/pdfs/conover_icwsm2011_polarization.pdf

Dahlgren, P. (2005). The Internet, public spheres and political communication: Dispersion and deliberation. *Political Communication, 22*(2), 147-162.

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

Donath, J.S. Being Real. In: Goldberg, K. eds. (2001) *The Robot in the Garden: Telerobotics and Telepistemology in the Age of the Internet*. MIT Press, Cambridge, MA

Global Digital Statistics 2014 (2014). We are social website. Retrieved from http://wearesocial.sg/blog/2014/01/social-digital-mobile-2014/

Golbeck, J., Grimes, J., & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information and Technology, 61*(8), 1612-1621.

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.

Himelboim, I., McCreery, S., & Smith, M., (2013). Birds of a feather Tweet together: Integrating network and content analysis to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication, 18*, 154-174.

Hoang, T.-A., Lim, E.-P., Achananuparp, P., Jiang, J., and Zhu, F. (2011). On Modeling Virality of Twitter Content. In Xing, C., Crestani, F., and Rauber, A., editors, *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, volume 7008, chapter 27, pages 212– 221. Springer Berlin Heidelberg.

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. *Proceedings of the 12th Annual International Digital Government Research Conference,* 182-186.

Huckfeldt, R. R., Johnson, P. E., & Sprague, J. (2004). *Political disagreement: The survival of diverse opinions within communication networks*. New York: Cambridge University Press.

Hughes, D. J., Rowe, M., Batey, M., and Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169-2188.

Jungherr, A., Jurgens, P., & Schoen, H. (2012). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting elections with Twitter: What 140 characters reveal about political sentiment." *Social Science Computer Review, 30*(2), 229-234.

Katz, E. (1957). The Two-Step Flow of Communication: An Up-To-Date Report on an Hy- pothesis. *Public Opinion Quarterly*, 21(1, Anniversary Issue Devoted to Twenty Years of Public Opinion Research):61.

Lawless, J. L. (2012). Twitter and Facebook: New ways for members of Congress to send the same old messages? In R. L. Fox and J. M. Ramos (Eds.), *iPolitics: citizens, elections, and governing in the new media era.* New York: Cambridge University Press.

Lee, J., Ryu, H., Mon, L. & Park, S. J. (2013). Citizens' use of Twitter in political information sharing in South Korea. *Proceedings of iConference 2013*, 351-365. doi:10.9776/13210

Merton R. K. (1957). *Social theory and social structure.* Glencoe, Ill.: Free Press.

Metaxas, P. T., & Mustafajaj, E. (2010). From obscurity to prominence in minutes: Political speech and real-time search. *Proceedings of Web Science Conference*, April 26-27, Raleigh, NC, USA.

National Election Commission. (2013). The comprehensive survey of the 18th presidential election (No. 34-9761030-130011-14). Retrieved from http://www.nec.go.kr/BBS_201311240443435171.pdf

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

Page, R. E. (2011). *Stories and Social Media: Identities and Interaction*. NY: Routledge.

Pennacchiotti, M. and Popescu, A. (2011). A Machine Learning Ap- proach to Twitter User Classification. *ICWSM*, pages 281–288.Parmelee and Bichard 2012

Preece, J. (2000) *Online Communities. Designing Usability, Supporting Sociability.* Wiley, Chichester u.a.

Prosecutor detail attempt to sway South Korean election (2013, November 21). New York Times. Retrieved from http://www.nytimes.com/2013/11/22/world/asia/prosecutors-detail-bid-to-sway-south-korean-election.html?_r=0

Rainie, L, Smith, A., Scholozman, K. L., Brady, H., & Verba, S. (2012). *Social media and political engagement*. Retrieved from Pew Research Center: http:// http://pewinternet.org/~/media//Files/Reports/2012/PIP_SocialMediaAndPoliticalEngagement_PDF.pdf

Robertson, S., Vatrapu, 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.

Rohde, M., Reinecke, L., Pape, B., and Janneck, M. (2004). Community-Building with Web-Based Systems ? Investigating a Hybrid Community of Students. *Computer Supported Cooperative Work (CSCW)*, 13(5-6):471–499.

Roosevelt, C. M. (2012). Social Media Analytics: Data mining applied to insurance Twitter posts. *Casualty Actuarial Society E-Forum, 2*, 1-36.

Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs, 28(January/February).* Retrieved from http://www.bendevane.com/FRDC2011/wp-content/uploads/2011/08/The-Political-Power-of-Social-Media-Clay-Sirky.pdf

Shogan, C. J. (2010). Blackberries, tweets, and YouTube: Technology and the future of communicating with Congress. *PS: Political Science & Politics, 43*(2), 231-233.

Skoric, M., & Poor, N. (2012). Tweets and votes: A study of the 2011 Singapore general election. *Proceedings of 45th Hawaii International Conference on System Sciences.* doi:10.1109/HICSS.2012.607

Smith, A. (2009). *The Internet's Role in Campaign 2008*. Retrieved from Pew Research Center: http://pewinternet.org/~/media//Files/Reports/2009/The_Internets_Role_in_Campaign_2008.pdf

Soler, J. M., Cuartero, F., Roblizo, M. (2012). Twitter as a tool for predicting elections results. *Proceedings of 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Retrieved from https://ruidera.uclm.es/xmlui/bitstream/handle/10578/2972/fi_1361186031-twitter%20ieee.pdf?sequence=1

Stieglitz, S., & Dang-Xuan, L. (2012). Political Communication and Influence through Microblogging – An empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior. *Proceedings of 2012 45th Hawaii International Conference on System Sciences.* doi:10.1109/HICSS.2012.476

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweets in Twitter network. *Proceedings of IEEE International Conference on Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust*. doi:10.1109/SocialCom.2010.33

Sunstein, C. R. (2007). *Republic.com 2.0.* Princeton: Princeton University Press.

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

Yardi, S. & boyd, d. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society, 30*, 316-327.

Younus, A., Qureshi, M. A., Asar, F. F., Azam, M., Saeed, M., and Touheed, N. (2011). What Do the Average Twitterers Say: A Twitter Model for Public Opinion Analysis in the Face of Major Political Events.

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