

How to Become Famous in the Microblog World

Takeshi Sakaki

the University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
sakaki@biz-model.t.u-tokyo.ac.jp

Yutaka Matsuo

the University of Tokyo
Yayoi 2-11-16, Bunkyo-ku
Tokyo, Japan
matsuo@biz-model.t.u-tokyo.ac.jp

Abstract

Microblogging services are a social media platform in which users post their current states as a short text, similar to instant messaging (IM). Recently, microblogging services, including Twitter, have become extremely popular. Concomitantly with the growth of Twitter, several companies have begun to use Twitter as business tools. Some companies have succeeded in increasing sales. However, most companies are troubled at how to use Twitter and how to win popularity in the Twitter milieu. As described herein, we analyze how the number of followers changes in a certain period and determine which parameters dominate popularity in the Twitter world. We also classify Twitter users into groups based on social tagging and propose methods to become popular when using microblogging services.

Introduction

Social media have reached the masses and have become irreplaceable in daily life and business situations. "Social media" are media that general users generate and broadcast the contents of. Among all kinds of social network services, microblogging services, including Twitter, Jaiku, and Pownce, have become extremely popular in the past several years. Microblogging services are called "real-time blog services", in which users post a message about their current state in short text form, say within 140–200 characters, as in instant messaging (IM).

The number of unique Twitter users has increased rapidly. Figure 1 displays the number of Twitter unique visitors. This increasing tendency is expected to continue through the next several years.

Recently, the number of corporate accounts as well as those of individual users has been increasing. They want to use Twitter to increase their sales figures, to communicate with their customers, and to boost their reputation. Several companies achieved success in increasing sales figures using Twitter. However, most companies are fumbling at how to use Twitter for business. For companies or users who seek to broadcast information about themselves to many people, the most important thing is to become popular in the Twitter world.

Many studies of how to calculate the importance of Web pages have been done. The most famous method

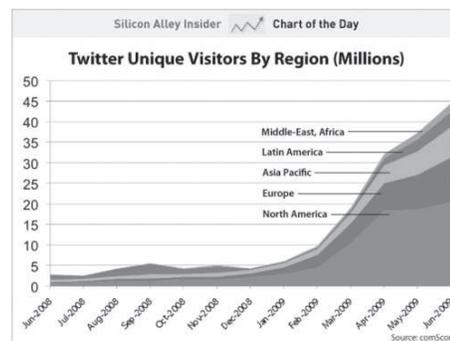


Figure 1: Twitter's unique visitor number (Alney 2009) is PageRank, invented by Larry Page and applied by google.com(Page et al. 1999). In addition, methods for ranking information and pages in search engines, so-called called "Search Engine Optimization," are well established these days.

In the Twitter world, we can know easily how important a user is, different from the Web. "follower counts" show how important accounts are in the Twitter world. If a user starts to Follow another user, then the user is inferred to be interested in the second user. For that reason, the simple count of followers can be treated as the popularity of a user. In fact, when we seek a useful user, we use follower counts as clues to discern how popular a user is.

It is important to know how to increase the number of followers. Nevertheless, such methods have never been established; no studies in the related literature describe such methods.

As described herein, we assess and report which feature quantities are more effective to become popular in the Twitter world.

- define the follower count as the popularity in the Twitter world.
- investigate users whose follower counts increase rapidly.
- calculate various amounts of them and which indices correlate strongly with the follower count.

In the section "feature quantities of Twitter", we explain functions of Twitter and list feature quantities adopted in this paper. In the "Experiments" section, we present our experiments and discuss the results. In the "Conclusion" section, we conclude this paper.

Feature quantities of Twitter

This paper identifies feature quantities that are effective for the popularity of Twitter users. Then, we calculate feature quantities of many kinds.

As described herein, amounts of two types exist.

The first type is a group of features related to user behavior. These features express action patterns of users when they tweet.

The second type is group of features based on Twitter functions. Twitter has numerous functions and feature quantities of many types, including friends, followers, retweets, etc. We analyze how these quantities affect follower counts of users.

Feature quantities related to user behavior

We defined feature quantities related to a user's behavior as follows:

- **Total Tweets** C
"Total Tweets" C are the number of tweets that a user has tweeted since registering with Twitter.
- **Average Tweet Length** \bar{L}
"Average Tweet Length" \bar{L} represents the average length of tweets. Here, \bar{L} is defined as

$$\bar{L} = \frac{\sum_i^{\text{recent}N} \text{Length}(tweet_i)}{N} \quad (1)$$

Ideally, all tweets a user has tweeted should be calculated. As described in this paper, we calculate the most recent 100 tweets ($N = 100$) for ease of analysis.

- **Average Tweet Interval** \bar{I}_t
The "Average Tweet Interval" is the average time span between tweets.

$$\bar{I}_t = \frac{\sum_i^{\text{recent}N} (T_{tweet_{i+1}} - T_{tweet_i})}{N} \quad (2)$$

$$(3)$$

(T_{tweet_i} : Time when $tweet_i$ was tweeted)

We calculate the most recent 100 tweets ($N = 100$) for ease of analysis here, too.

Feature quantities related to Twitter functions

The Twitter System has functions of many kinds. For that reason, we can extract relationships of many kinds from the Twitter world. Some feature quantities are related to a user's behavior in Twitter. We define feature quantities related to Twitter functions as follows.

- **Friends Count** F_r
In fact, F_r signifies the number of users a target user Follows.
- **Follower Count** F_{ol}
The number of users who Follow a target user is F_{ol} .

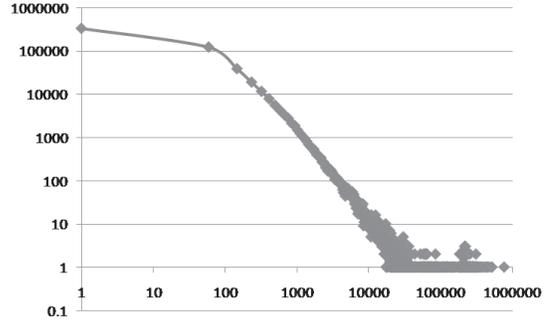


Figure 2: Frequency Distribution of Incremental Counts

- **Message Count (Active/Passive)** RP_a, RP_p
The number of messages a target user sends/receives for a certain period of time is RP_a / RP_p .
- **Retweet Count (Active/Passive)** RT_a, RT_p
The number of messages a target user retweets/is retweeted for a certain period of time is RT_a / RT_p .
- **List Count (Active)** L_a, L_p
The number of lists a target user makes/ include a target user is L_a / L_p .

Dataset

We collected information of many kinds from Twitter.com using Twitter API. These are various, so we explain them as follows: We crawled 10% of all tweet information from Twitter.com twice.

1. The first period was from 27 September 2009 to 3 October 2009.
2. The second period was from 27 December 2009 to 2 January 2010.

These two sets of data include the following relationship data of 200 million users at that time. Their size is 30 GB.

Analysis

We do two analyses in this paper. First, we analyze the transition of the number of followers. Second, we analyze how each feature quantity correlates with the increment of followers.

According to results of these two analyses, we classify users into some groups with the increasing number of followers(G_{Cn}) and with the increasing rate of followers(G_{Rn}). And we discuss what type of users could increase their followers.

Changes of follower counts

In this analysis, we prepare two sets of data crawled in September 2009 and December 2009. These datasets have follower counts of users at two points of time. We compare these two datasets and analyze how followers of Twitter users changed.

To compare these two sets of data, we calculate the increment counts I_c and increment rate I_r of each users. These two quantities are defined as follows:

Table 1: Correlation coefficient of all at Sep, 2009

Feature	Fr	Fol	RP_a	RP_p	RT_a	RT_p	L_a	L_p	C	\bar{L}	\bar{t}
I_c	0.01	1.00	0.52	0.98	0.17	0.97	0.07	0.99	0.01	-0.00	-0.02
I_r	0.22	-0.01	0.04	0.00	0.00	-0.01	-0.03	-0.01	-0.01	0.00	-0.06

Table 2: Correlation Coefficients with I_c at Sep,2009

GROUP	Fr	Fol	RP_a	RP_p	RT_a	RT_p	L_a	L_p	C	\bar{L}	\bar{t}
G_C1	-0.06	0.02	-0.08	-0.04	-0.05	0.02	-0.10	0.04	0.00	0.08	0.08
G_C2	0.02	0.00	-0.04	-0.02	0.09	0.09	-0.02	-0.05	0.00	0.08	0.08
G_C3	0.01	0.04	-0.04	0.04	0.01	-0.02	-0.01	0.14	-0.05	0.11	0.08
G_C4	-0.12	-0.06	0.01	0.03	0.04	0.02	-0.01	0.02	-0.08	-0.01	0.09
G_C5	0.03	-0.01	0.01	-0.03	-0.04	0.05	-0.08	-0.08	-0.09	0.01	-0.03
G_C6	0.00	-0.01	-0.06	-0.03	-0.11	-0.06	-0.11	0.00	-0.06	-0.04	-0.03
G_C7	0.02	0.05	0.00	-0.02	0.03	-0.01	-0.04	0.05	0.08	-0.01	-0.06
G_C8	0.08	0.11	0.04	0.00	0.00	0.06	0.01	0.03	0.09	-0.01	0.08
G_C9	0.11	0.33	0.14	0.17	0.11	0.13	0.03	0.17	0.02	0.19	-0.04
G_C10	0.12	0.47	0.03	0.16	0.05	0.10	0.01	0.48	-0.13	0.23	-0.11

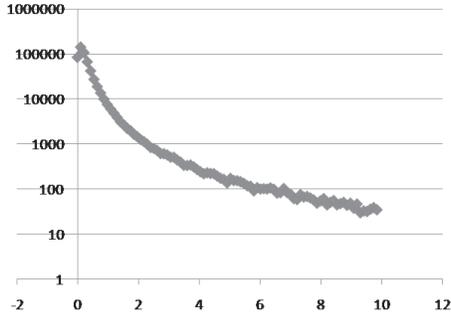


Figure 3: Frequency Distribution of Incremental Rate

$$I_c = Fol_{new} - Fol_{old} \quad (4)$$

$$I_r = \frac{I_c}{Fol_{old}} \quad (5)$$

Figure 2 shows the frequency distribution of I_c of all users. The x-axis of this frequency distribution graph shows I_c ; the y-axis shows the number of users. In Fig. 2, 80% of users are shown as having accumulated fewer than 60 new followers.

Figure 3 shows the frequency distribution of I_r of all users. The x-axis of this graph means I_r and the y-axis shows the number of users. In Fig. 3, 60% of users are shown as having increased followers by less than 30%; 80% of users increased followers by less than 50%, and 97% of users increased followers by less than 200%.

According to these two graphs, most users did not increase follower counts so much.

Correlation Coefficient of each quantity with the follower Incremental Rate

We evaluate feature quantities using a correlation coefficient. In this analysis, we calculate how feature quantities correlate with I_c and I_r .

The correlation coefficient is an index expressing how we correlate two probability variables. We performed correla-

tion analyses of two types. First, we calculated the correlation coefficient of a set of users who were sampled randomly. Second, we calculated the correlation coefficient of groups of users classified by I_c or I_r .

Correlation analysis of random sampling data Table 1 shows the correlation coefficient of each quantity with I_c and I_r of random sampling data.

Considering correlation with I_c , it is apparent that Fol_{old} correlate with I_c . It is extremely natural when one considers the fact that most users did not achieve a high incrementation rate I_r . That fact means that the order of follower counts was not changed so much. Therefore, changes of follower counts are affected by the follower count itself.

Additionally, RP_p , RT_p and L_p correlate with I_c . This fact means that those who are often retweeted, replied or listed have gotten many followers.

Considering correlation with I_r , it is apparent that each quantity does not correlate with I_r . However, only "friend counts" have a weak correlation with I_r .

The phenomenon is explainable by "Follow back", which is a cultural phenomenon in Twitter. When a user is followed by another user, the first user reciprocates and follows the second user. This action is called "Follow back".

Correlation Analysis of user groups Before analysis, we classify users into groups as I_c and I_r .

Table 2 shows ranges of groups by I_c and their correlation coefficient. Table 3 shows ranges of groups by I_r and their correlation coefficient.

According to Table 2, most features do not correlate with I_c . In " G_C9 " and " G_C10 ", both Fol and L_p have moderate correlation with I_c . The fact means that users who have more followers acquire more new followers, which is natural, considering that 80% of users have I_c less than 50%.

Additionally, both \bar{L} and RP_p show weak correlation with I_c . Users who post longer tweets and are replied more often accumulate more new followers.

To analyze users who have high I_c in detail, we collected names of list users. Table 4 presents summary data related to

Table 3: Correlation Coefficients with I_r at Sep,2009

GROUP	Fr	Fol	RP_a	RP_p	RT_a	RT_p	La	Lp	C	\bar{L}	\bar{It}
G_{R1}	0.07	0.05	-0.01	-0.02	0.02	-0.04	0.02	-0.03	-0.08	-0.05	0.01
G_{R2}	-0.04	-0.01	0.09	0.04	0.01	-0.06	0.04	-0.02	0.06	-0.02	0.01
G_{R3}	0.09	0.08	-0.05	-0.02	0.01	-0.04	0.06	0.04	-0.04	-0.05	-0.03
G_{R4}	0.30	0.39	0.19	0.27	0.27	0.11	0.10	0.47	0.25	-0.03	-0.09
G_{R5}	0.22	0.24	0.09	0.12	0.09	0.13	0.18	0.21	0.18	0.01	-0.06
G_{R6}	-0.05	-0.03	-0.04	-0.08	0.03	-0.03	0.08	-0.06	-0.06	0.04	0.02
G_{R7}	0.08	0.05	0.06	0.10	-0.07	-0.04	0.01	0.06	0.03	0.06	-0.01
G_{R8}	0.01	-0.09	-0.04	-0.06	-0.01	-0.01	-0.02	-0.07	0.03	-0.07	-0.05
G_{R9}	-0.06	0.03	0.01	-0.03	0.05	0.01	0.04	0.08	0.09	0.03	-0.05
G_{R10}	-0.03	-0.02	0.03	0.05	-0.02	-0.01	-0.03	0.14	-0.03	-0.01	-0.01

Table 4: Popular List Names

G_C10	G_R4	G_R10
celebs	friends	news
news	music	friends
celebrities	twibes-media	colirios
sports	sports	music
music	twibes-blogging	amigos
twibes-media	twibes-marketing	tech
tech	design	conversationlist
entertainment	twibes-writer	celebs
famous	gymnastics	sports

G_C10 . According to Table 4, G_C10 , which has high I_c , includes celebrities and some news sites. Combining this fact and results of correlation analysis, we can infer that celebrities who post longer tweets and are replied more acquire more new followers.

As presented in Table3, most features do not correlate with I_r . In " G_{R9} " and " G_{R10} ", both L_p have weak correlation with I_r . In " G_{R4} " and " G_{R5} ", Fr , Fol , RP_a , RP_p , RT_a , RT_p , L_p , C apparently have weak correlation with I_r .

To analyze this result in detail, we investigate names of lists. Table 4 presents summary data of "list name" of G_R4 and G_R10 . "Twib" is a web service to create groups of Twitter users interested in a common topic.¹ As shown in Table 4, G_R4 includes many names including "friends" and "twib". G_R10 has varieties of names.

Combining this fact with results of correlation analysis, in groups with moderate I_r , users who post many interesting tweets accumulate more followers. At the same time, users who often communicate with others also get more followers. According to the fact that C , RP_a , RP_p , RT_a , and RT_p all correlate with I_r , users with high I_r posted many tweets received many replies. Moreover, their tweets are retweeted many times. We can infer that they post many interesting tweets and that those tweets elicit the interest of others.

Discussion

Results of two analyses reveal two facts.

- users with high I_c : users who post longer tweets and are replied many times acquired more followers.

¹<http://www.twibes.com>

- users with moderate I_r : users who post many tweets or interesting tweets acquired more followers.
- users with high I_r : users who are listed by many users and who are tagged clearly acquired more followers.

We propose the following methods to increase follower counts.

- A user should Follow another user if Followed by a user. This practice of "Follow back" is effective for increasing follower counts.
- A user should reply diligently and post longer tweets if one is a celebrity.
- A user should specifically examine one or two areas when posting tweets. Thereby, the user becomes easy to tag.

Twitter users have relationships of many kinds. We can infer networks of many kinds. For that reason, we should have calculated feature quantities of network and compared other features.

In fact, we tried to sample users, compose networks, and calculate feature quantities of networks. However, their results differ entirely from one sampling set to another. Therefore, we attempt to calculate network quantities for future use.

Conclusion

As described herein, we showed which elements affect the increase in follower counts in Twitter. We calculate several feature quantities and test which are correlated with the increase of follower counts. In addition, we proposed four methods to increase followers.

Recently, the number of Twitter users is increasing rapidly. Many companies have started to use Twitter for business purposes. For that reason, the importance of searching for important or popular users is expected to become increasingly important. Therefore, our method is expected to become widely used and contribute to improvement in the use and analyses of microblogging services.

References

- Alney, D. 2009. Chart of the day: Twitter's boom around the world. Technical report, Firebelly Marketing.
- Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The pagerank citation ranking: Bringing order to the web.