What is Twitter, a Social Network or a News Media?

Haewoon Kwak  Changhyun Lee  Hosung Park  Sue Moon
Department of Computer Science, KAIST, Korea

19th International World Wide Web Conference (WWW2010)
Twitter, a microblog service
Twitter, a microblog service

write a short message
Twitter, a microblog service

read neighbors’ tweets
In most OSN

“We are friends.”
In Twitter

“I follow you.”
Following on Twitter

“Unlike most social networks, following on Twitter is not mutual. Someone who thinks you're interesting can follow you, and you don't have to approve, or follow back.”

http://help.twitter.com/entries/14019-what-is-following
Following = subscribing tweets
http://twitpic.com/135xa – There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.
The goal of this work

We analyze how directed relations of following set Twitter apart from existing OSNs.

Then, we see if Twitter has any characteristics of news media.
media [mee-dee-uh]

1. a pl. of medium

2. the means of communication, as radio and television, newspapers, and magazines, that reach or influence people widely
The goal of this work

We analyze how directed relations of following set Twitter apart from existing OSNs.

Then, we see if Twitter has any characteristics of news media.
Summary of our findings

1. Following is mostly not reciprocated (not so “social”)

2. Users talk about timely topics

3. A few users reach large audience directly

4. Most users can reach large audience by WOM* quickly

*WOM: word-of-mouth
Data collection (09/6/1~9/24)

- 41.7M user profiles (near-complete at that time)
- 1.47B following relations *publicly available
- 4262 trending topics
- 106M tweets mentioning trending topics
  - Spam tweets removed by CleanTweets
How we crawled

- Twitter’s well-defined 3rd party API
- With 20+ ‘whitelisted’ IPs
  - Send 20,000 requests per IP / hour
Recent studies

• Ranking methodologies [WSDM’10]
• Predicting movie profits [HYPERTEXT’10]
• Recommending users [CHI’10 microblogging]
• Detecting real time events [WWW’10]

• The ‘entire’ Twittersphere unexplored
Part I.

1. Following is mostly not reciprocated (not so “social”)

2. Users talk about timely topics

3. A few users reach large audience directly

4. Most users can reach large audience by WOM* quickly
Why do people follow others?

- Reflection of offline social relationships

otherwise,

- Subscription to others’ messages
Sociologists’ answer

- “Reciprocal interactions pervade every relation of primitive life and in all social systems”
Is following reciprocal?

- Only **22.1%** of user pairs follow each other
- Much lower than
  - 68% on Flickr
  - 84% on Yahoo! 360
  - 77% on Cyworld guestbook messages
Low reciprocity of following

- Following is not similarly used as friend in OSNs
  - Not reflection of offline social relationships

- Active subscription of tweets!
Part II.

1. Following is mostly not reciprocated (not so “social”)

2. Users talk about timely topics

3. A few users reach large audience directly

4. Most users can reach large audience by WOM* quickly
Dynamically changing trends
User participation pattern can be a signature of a topic

(a) Topic ’apple’

(b) Topic ’#iranelection’
Majority of topics are headline news

31.5% “ephemeral”

54.3% “headline news”

6.9% “persistent news”

(a) Exogenous subcritical (topic ‘#backintheday’)

(b) Exogenous critical (topic ‘beyonce’)

(c) Endogenous subcritical (topic ‘lynn harris’)

(d) Endogenous critical (topic ‘#redsox’)

Friday, April 30, 2010
Part III.

1. Following is mostly not reciprocated (not so “social”)

2. Users talk about timely topics

3. A few users reach large audience directly

4. Most users can reach large audience by WOM* quickly
How many followers a user has?

![CCDF of number of followings and followers](image.png)

CCDF

# of followings/followers

Followings

Followers

10^0

10^{-1}

10^{-2}

10^{-3}

10^{-4}

10^{-5}

10^{-6}

10^{-7}

10^{-8}

10^0

10^1

10^2

10^3

10^4

10^5

10^6

10^7

27
CCDF

• Complementary Cumulative Density Function

• \( \text{CCDF}(x=k) = \int_{k}^{\infty} P(x) \, dx \)
Reading the graph

![Graph](image.png)

**CCDF**

- **Followings**
- **Followers**

**# of followings/followers**

- $10^0$
- $10^1$
- $10^2$
- $10^3$
- $10^4$
- $10^5$
- $10^6$
- $10^7$

- $10^0$
- $10^{-1}$
- $10^{-2}$
- $10^{-3}$
- $10^{-4}$
- $10^{-5}$
- $10^{-6}$
- $10^{-7}$
- $10^{-8}$
Plenty of super-hubs

Figure 1: Number of followings and followers
More super-hubs than projected by power-law

- Where do they get all the followers? Possibly from...
  - Search by ‘name’
  - Recommendation by Twitter

- They reach millions in one hop
Are those who have many followers active?

![Graph showing the relationship between number of followers and number of tweets](image.png)
How we plotted

<table>
<thead>
<tr>
<th>followers</th>
<th>tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>
How we plotted

<table>
<thead>
<tr>
<th>followers</th>
<th>tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

Avg. = 8
Med. = 9
More followers, more tweets

![Graph showing the relationship between number of followers and number of tweets.](image)
Many followers without activity

![Graph showing the relationship between the number of followers and the number of tweets per user.]
## Twitter user rankings by Followers, PageRank and RT

<table>
<thead>
<tr>
<th>Rank</th>
<th>Followers</th>
<th>PageRank</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>actor</td>
<td>actor</td>
<td>news</td>
</tr>
<tr>
<td>2</td>
<td>musician</td>
<td>president</td>
<td>news</td>
</tr>
<tr>
<td>3</td>
<td>show host</td>
<td>news</td>
<td>news</td>
</tr>
<tr>
<td>4</td>
<td>news</td>
<td>show host</td>
<td>journalist</td>
</tr>
<tr>
<td>5</td>
<td>show host</td>
<td>musician</td>
<td>news</td>
</tr>
<tr>
<td>6</td>
<td>twitter</td>
<td>show host</td>
<td>news</td>
</tr>
<tr>
<td>7</td>
<td>president</td>
<td>sports star</td>
<td>musician</td>
</tr>
<tr>
<td>8</td>
<td>show host</td>
<td>musician</td>
<td>news</td>
</tr>
<tr>
<td>9</td>
<td>sports star</td>
<td>twitter</td>
<td>comedian</td>
</tr>
<tr>
<td>10</td>
<td>model</td>
<td>show host</td>
<td>about iran</td>
</tr>
</tbody>
</table>
### Twitter user rankings by Followers, PageRank and RT

<table>
<thead>
<tr>
<th>Rank</th>
<th>Followers</th>
<th>PageRank</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>actor</td>
<td>actor</td>
<td>news</td>
</tr>
<tr>
<td>2</td>
<td>musician</td>
<td>president</td>
<td>news</td>
</tr>
<tr>
<td>3</td>
<td>show host</td>
<td>news</td>
<td>news</td>
</tr>
<tr>
<td>4</td>
<td>news</td>
<td>show host</td>
<td>journalist</td>
</tr>
<tr>
<td>5</td>
<td>show host</td>
<td>musician</td>
<td>news</td>
</tr>
<tr>
<td>6</td>
<td>twitter</td>
<td>show host</td>
<td>news</td>
</tr>
<tr>
<td>7</td>
<td>president</td>
<td>sports star</td>
<td>musician</td>
</tr>
<tr>
<td>8</td>
<td>show host</td>
<td>musician</td>
<td>news</td>
</tr>
<tr>
<td>9</td>
<td>sports star</td>
<td>twitter</td>
<td>comedian</td>
</tr>
<tr>
<td>10</td>
<td>model</td>
<td>show host</td>
<td>about iran</td>
</tr>
</tbody>
</table>
Great discrepancy among rankings

\[ K \]

\[ R_F - R_{PR} \quad + \quad R_F - R_{RT} \quad \triangle \quad R_{PR} - R_{RT} \]

top k ranking

Friday, April 30, 2010
Part IV.

1. Following is mostly not reciprocated (not so “social”)

2. Users talk about timely topics

3. A few users reach large audience directly

4. Most users can reach large audience by WOM* quickly

*WOM: word-of-mouth
Which is more efficient for WOM?
In Twitter

Information

Following
Average path length: 4.1
Retweet (RT)

- Relay tweets from a following to followers
Retweet (RT)

- Relay tweets from a following to followers
Retweet (RT)

• Relay tweets from a following to followers
Retweet (RT)

- Relay tweets from a following to followers
Retweet (RT)

- Relay tweets from a following to followers

RT @node0 Last day of WWW’10

Friday, April 30, 2010
Retweet (RT)

- Relay tweets from a following to followers
Retweet (RT)

- Not only 1 hop neighbors
4. WORD-OF-MOUTH

Retweet (RT)

- More goes further

2 hop neighbors

node0
node1
node2
node3
node4
node5
node6
We construct RT tree

- A tree with writer and retweeter(s)
Height of RT trees

1

1

2
Empirical RT trees
96% of RT trees = Height 1
Boosting audience by RT

Figure 14: Average and median numbers of additional recipients of the tweet via retweeting

People subscribe to mass media in various forms: radio, TV, and newspapers. They are immediate recipients and consumers of the news the established media produce. On Twitter people acquire information not always directly from those they follow but often via retweets. Assuming a tweet posted by a user is viewed and consumed by all of the user's followers, we count the number of additional recipients who are not immediate followers of the original tweet owner. Figure 14 displays its average and median per tweet against the number of followers of the original tweet user. The median lies almost always below the average, indicating that many tweets have a very large number of additional recipients. Up to about 1,000 followers, the average number of additional recipients is not affected by the number of followers of the tweet source. That is, no matter how many followers a user has, the tweet is likely to reach a certain number of audience, once the user's tweet starts spreading via retweets. This illustrates the power of retweeting. That is, the mechanism of retweet has given every user the power to spread information broadly. We recall that influentials by the number of retweets are dissimilar with those by the number of followers or PageRank. Individual users have the power to dictate which information is important and should spread by the form of retweet, which collectively determines the importance of the original tweet. In a way we are witnessing the emergence of collective intelligence.

6.2 Retweet Trees

Knowing that retweet actually delivers information to far more people than a source’s immediate followers, we are now interested in how far and deep retweets travel in Twitter. In order to answer the question we build an information diffusion tree of every tweet that is retweeted and call it a retweet tree. All retweet trees are subgraphs of the Twitter network.

We illustrate all the retweet trees of the topic 'air france flight' in Figure 15. In every connected component different colors represent different tweets. The forest of retweet trees has a large number of one or two-hop chains. We find interesting retweet patterns such as repetitive retweet and cross-retweet; the former is repeatedly retweeting the same tweet, and cross-retweet is retweeting each other.

In Figure 16 we plot the CCDFs of the retweet tree heights and the number of users in a retweet tree. The height of 1 is the most
Additional readers

2 additional readers by retweeter

3 followers
A retweet brings a few hundred additional readers

![Graph showing the relationship between the number of followers of the source and the number of additional recipients via retweeting.](image)

- **X-axis:** Number of followers of the source
- **Y-axis:** Number of additional recipients of the tweet via retweeting

**Legend:**
- **Avg.**
- **Med.**

**Graph Description:**
- The graph illustrates the distribution of additional recipients based on the number of followers of the source tweet.
- The data points represent the actual retweeted tweets, with the median line indicating the central tendency.
- The average line shows the expected value based on the followers count.

---

### 4. Word-of-Mouth

A retweet brings a few hundred additional readers.
Time lag between hops in RT tree

Graph showing CDF of elapsed time for each hop in a retweet tree.
Fast relaying tweets by RT: 35% of RT < 10 min.
Fast relaying tweets by RT: 55% of RT < 1hr.
Summary

1. We study the entire Twittersphere

2. Low reciprocity distinguishes Twitter from OSNs

3. Twitter has characteristics of news media:
   - Tweets mentioning timely topics
   - Plenty of hubs reaching a large public directly
   - Fast and wide spread of word-of-mouth
Resources

Supplementary info.
About Twitter

“Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now.”

Twitter asks “what’s happening” and makes the answer spread across the globe to millions, immediately.
About Twitter

“Twitter is a real-time information network powered by people all around the world that lets you share and discover what’s happening now.”

Twitter asks “what’s happening” and makes the answer spread across the globe to millions, immediately.

makes the answer spread across the globe to millions, immediately.
A few numbers

- 105M registered accounts
- 55M tweets a day
- 180M unique visitors a month
- 19B searches a month
Homophily in terms of followers
Assortative mixing
Homophily in terms of location

![Graph showing the relationship between the average time difference (ΔT) and the number of reciprocal friends.]
Favoritism in RTs?

- A few informative users?
Disparity in weighted network

\[ Y(k, i) = \sum_{j=1}^{k} \left\{ \frac{|r_{ij}|}{\sum_{i=1}^{k} |r_{il}|} \right\}^2 \]
Favoritism in RTs

\( \gamma = 0.801 \)

\( \gamma = 0.892 \)

(a) \( k_{out} Y(k_{out}) \sim k_{out} \)

(b) \( k_{in} Y(k_{in}) \sim k_{in} \)
Fast WOM by retweet

![Graph showing elapsed time from (N-1) hop to N hop](image)

- Elapsed time:
  - 1 d
  - 12 h
  - 6 h
  - 1 h
  - 30 m
  - 10 m

- From (N-1) hop to N hop:
  - 0
  - 2
  - 4
  - 6
  - 8
  - 10
  - 12
  - 14

- Graph lines:
  - 25%
  - med
  - 75%

Discussion:
- The graph illustrates the fast WOM (Word of Mouth) diffusion through retweeting in social networks.
- Each point on the graph represents the time it takes for the retweet to reach a certain number of hops away from the source.
- The graph shows that most retweets occur within a short time frame, with a peak at around 1 hour.
- The 25% and 75% lines indicate the interquartile range of retweet times, showing a wide distribution.
- The med line represents the median retweet time, which is closer to the 25% line in this case.

Conclusion:
- The rapid spread of information through retweeting is a key feature of social networks, allowing for fast WOM diffusion.

References:

Further reading:
- Cyworld, the biggest social networks in Korea.
- Internet chain letter forwarding and meme propagation in blogspace.
- Linking patterns in blog graph and technology.
- The success of online social networks opens extensively from sociology, marketing, and epidemiology.