Content Sharing in Social Broadcasting Environment: Evidence from Twitter

Huaxia Rui
(with Zhan Shi and Andrew Whinston)

Simon School of Business
University of Rochester

September 28, 2012

Outline

A Screenshot of Retweeting on Twitter

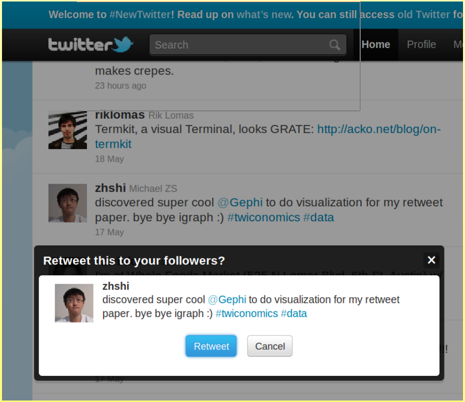
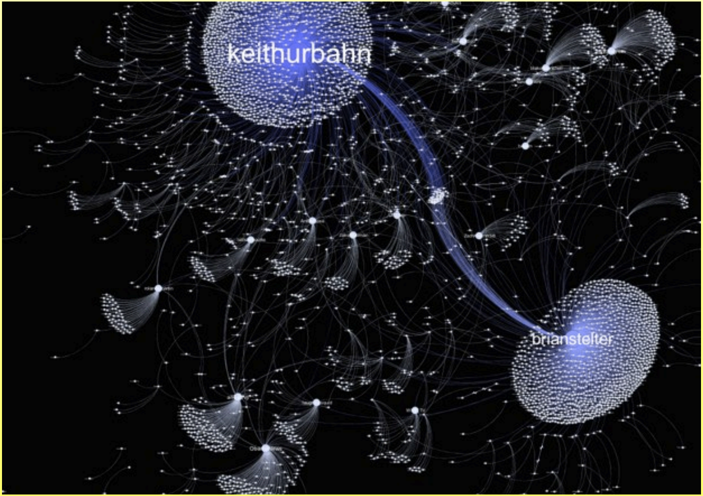


Figure 1: Retweeting is the action of a user rebroadcasting tweets received from the user's followings to the user's followers.

The Power of Retweeting

The Power of Retweeting[1]



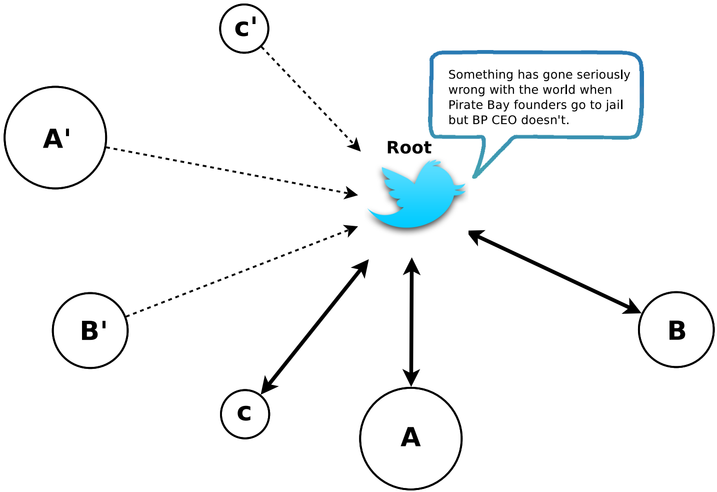
[1] http://blog.socialflow.com/post/5246404319

Replacement for the Follower Count?

- Twitter cofounder Evan Williams – who remains on the company's board – hinted on September 24, 2012, that follower counts may soon become the second most important number to users.
- According to Evan, "The thing I think would be more interesting than followers is... retweets," clarifying that a simple measure of followers "doesn't capture your distribution."

Motivation

Who will retweet?



Motivation

Research Question

How does the strength of the interpersonal tie moderate people's voluntary content sharing behavior in a social broadcasting network?

Significance

- Retweeting is the main instrument through which information diffuses on Twitter. Hence, a user's ability to generate more retweeting is considered an important aspect of influence. Therefore, the study of this research question could shed light on the measurement of influence.
- Studying people's retweeting behavior not only helps us understand how people relay information but also shed light on why people relay information.

Outline

How Did We Collect Data?

- We fetch tweets from July 22$^{\mathrm{nd}}$ to Dec 2$^{\mathrm{nd}}$ 2010.
- Once a tweet enters our tweets database, the "fetch-retweeters" program began to track and fetch its retweeting data, and would be doing constantly during the subsequent 5 days.
- As retweeting data came in, another "fetch-graph" program worked on collecting relevant network graph information.
- By the end of the 140 day period, we had successfully completed data collection for 65 tweets. In order of the time when a tweet was posted, we indexed them by an integer, t, ranging from 1 to 65.

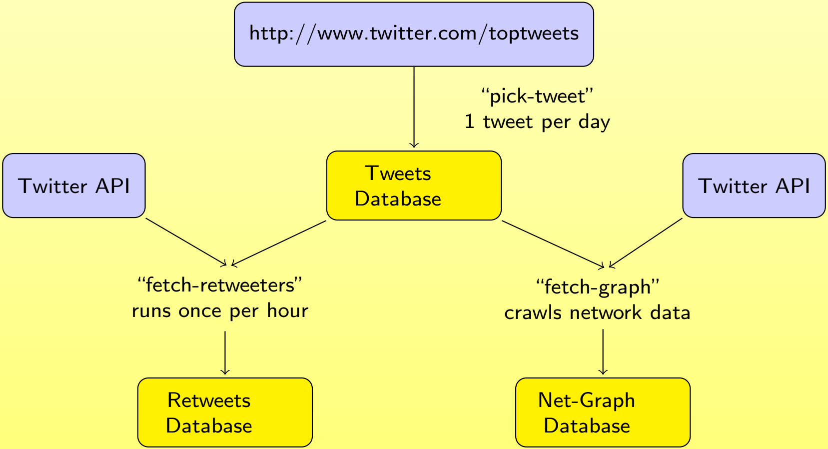


Figure 2: Data Collection Workflow

Outline

Key Variables

For a tweet t, we use $i \in \{1, 2, \ldots, n_t\}$ to index its observations (i.e. author t's followers). The definitions of the key variables are listed here

- y_{ti}: binary outcome, 1 if observation ti retweeted and 0 otherwise;
- w_{ti}: binary variable, 0 if author t and observation ti mutually follow each other, and 1 if only ti follows author t (weak tie);
- V_{ti}: outdegree, the number of ti's followings;
- W_{ti}: indegree, the number of ti's followers;
- m_{ti}, the number of times ti's followings retweeted tweet t.

Model

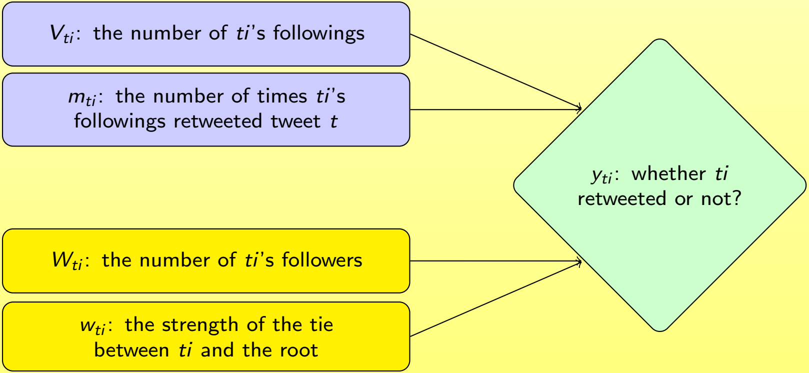


Figure 3: Conceptual Model

A Two-Stage Model

Given a tweet t by the root, and root's follower ti

$$\begin{aligned} \text{Prob}\,(i \text{ retweets } t) &= \text{Prob}(i \text{ reads tweet } t) \\ &\times \text{Prob}(i \text{ retweets } t \,|\, i \text{ reads the tweet } t) \end{aligned}$$

Notation

$$\begin{aligned} \text{P}_1 &= \text{Prob}(i \text{ reads tweet } t) \\ \text{P}_2 &= \text{Prob}(i \text{ retweets } t \,|\, i \text{ reads the tweet } t) \end{aligned}$$

What Affect P_1

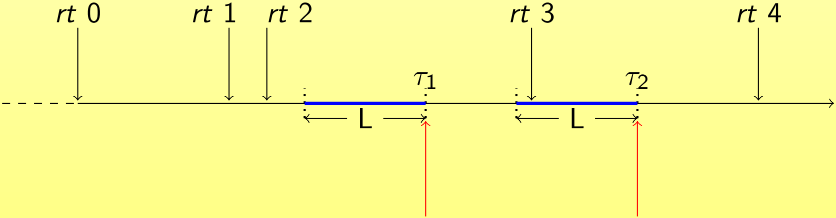


Figure 4: (Re)Tweets Entering A Twitter User's Timeline

Stage 1

Assumptions

$$\mathrm{P}_1 = \mathrm{Prob}\left(\frac{m_{ti}}{V_{ti}} > \epsilon_{ti}\right), \text{ where } \log(\epsilon_{ti})|t \sim \log(\epsilon_{ti}) \sim N(\bar{\epsilon}, \sigma_\epsilon^2)$$

$$\begin{aligned}
\log(m_{ti}) - \log(V_{ti}) &> \log(\epsilon_{ti}) \\
-\frac{\bar{\epsilon}}{\sigma_\epsilon} + \frac{1}{\sigma_\epsilon}\log(m_{ti}) - \frac{1}{\sigma_\epsilon}\log(V_{ti}) &> \frac{\log(\epsilon_{ti}) - \bar{\epsilon}}{\sigma_\epsilon} \\
\mathrm{P}_1 = \Phi(-\frac{\epsilon}{\sigma_\epsilon} + \frac{1}{\sigma_\epsilon}\log(m_{ti}) &- \frac{1}{\sigma_\epsilon}\log(V_{ti}))
\end{aligned}$$

$$\mathrm{P}_1 = \Phi\left(\alpha_0 + \alpha_1 \log(m_{ti}) + \alpha_2 \log(V_{ti})\right)$$

Stage 2

Assumptions

$$\mathrm{P}_2 = \mathrm{Prob}\left(\hat{c}_t + \hat{\beta}_1 w_{ti} + \hat{\beta}_2 W_{ti} > \eta_{ti}\right),$$

where

$$\eta_{ti} \perp \epsilon_{ti}, \quad \eta_{ti}|t \sim \eta_{ti} \sim N(\bar{\eta}, \sigma_\eta^2)$$

$$-\frac{\bar{\eta}}{\sigma_\eta} + \frac{\hat{c}_t}{\sigma_\eta} + \frac{\hat{\beta}_1}{\sigma_\eta} w_{ti} + \frac{\hat{\beta}_2}{\sigma_\eta} W_{ti} > \frac{\eta_{ti} - \bar{\eta}}{\sigma_\eta}$$

$$\mathrm{P}_2 = \Phi(-\frac{\bar{\eta}}{\sigma_\eta} + \frac{\hat{c}_t}{\sigma_\eta} + \frac{\hat{\beta}_1}{\sigma_\eta} w_{ti} + \frac{\hat{\beta}_2}{\sigma_\eta} W_{ti})$$

$$\mathrm{P}_2 = \Phi\left(c_t + \beta_1 w_{ti} + \beta_2 W_{ti}\right)$$

Full Model

$$
\begin{aligned}
\mathrm{P}(y_{ti} = 1) =& \mathrm{P}_1 \times \mathrm{P}_2 \\
=& \Phi(\alpha_0 + \alpha_1 \log(m_{ti}) + \alpha_2 \log(V_{ti})) \\
& \times \Phi(c_t + \beta_1 w_{ti} + \beta_2 W_{ti})
\end{aligned}
$$

$$\theta = \{\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2, c_1, c_2, \ldots, c_T\}$$

Outline

1. Motivation
2. Data
3. Model
4. Results

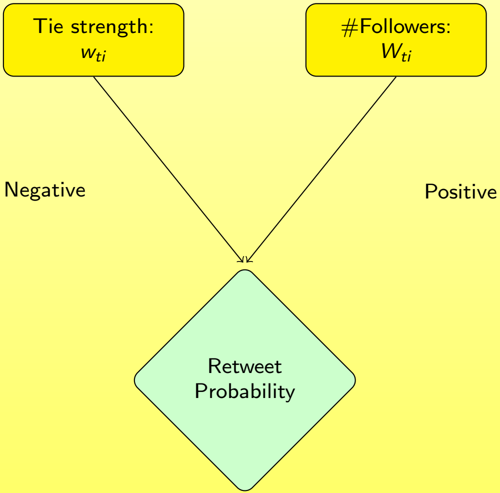
Result of MLE with Probit Specification

dependent variable: y_{ti}, # of tweets: 65, # of obs: 24403					
probit					
	$\log(m_{ti})$	$\log(V_{ti})$	w_{ti}	$\log(W_{ti})$	$w_{ti}\log(W_{ti})$
	α_1	α_2	β_1	β_2	
eq 1: probability of consumption upon receipt					
1)	0.494**	-0.639**			
	(4.81)	(-10.38)			
2)	0.485**	-0.641**			
	(4.69)	(-10.33)			
eq 2: probability of retweeting upon consumption					
1)			0.284**	0.115**	
			(5.57)	(6.32)	
2)			0.364*	0.127**	-0.014
			(2.33)	(4.15)	(-0.56)

Result of MLE with Logit Specification

logit					
	$\log(m_{ti})$	$\log(V_{ti})$	w_{ti}	$\log(W_{ti})$	$w_{ti} \log(W_{ti})$
	α_1	α_2	β_1	β_2	
eq 1: probability of consumption upon receipt					
3)	0.812**	-1.131**			
	(4.80)	(-10.19)			
4)	0.800**	-1.135**			
	(4.69)	(-10.13)			
eq 2: probability of retweeting upon consumption					
3)			0.556**	0.243**	
			(5.60)	(7.06)	
4)			0.697*	0.263**	-0.025
			(2.25)	(4.54)	(-0.49)

Discussion


Tie strength:
w_{ti}
#Followers:
W_{ti}
Negative
Positive
Retweet
Probability

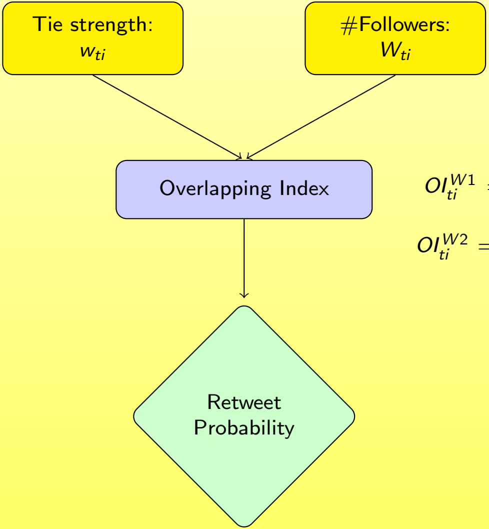
Discussion

Alex Moderator 05/08/2010 07:47 PM

One thing I would point out is that I don't re-tweet or mention certain people because I assume my friends are either already following them or simply not interested.[a]

[a] http://blogs.hbr.org/research/2010/05/influence-and-twitter.html

Discussion



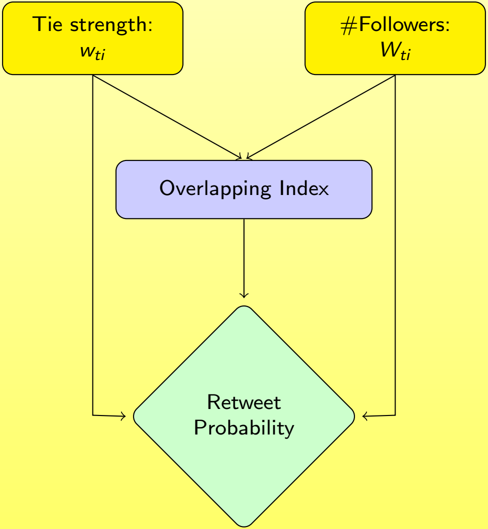
$$OI_{ti}^{W1} = \frac{\bar{W}_{ti}}{\sqrt{W_t}\sqrt{W_{ti}}}$$

$$OI_{ti}^{W2} = \frac{\bar{W}_{ti}}{\min\{W_t, W_{ti}\}}$$

Result of MLE with Probit Specification and *OI* included

dependent variable: y_{ti}, # of tweets: 65, # of obs: 24403					
probit					
	$\log(m_{ti})$	$\log(V_{ti})$	w_{ti}	$\log(W_{ti})$	OI_{ti}
	α_1	α_2	β_1	β_2	β_3
eq 1: probability of consumption upon receipt					
1)	0.445**	-0.586**			
	(6.04)	(-7.72)			
eq 2: probability of retweeting upon consumption					
1)			0.179**	0.134**	-2.947**
			(4.15)	(6.87)	-4.29

Discussion


Tie strength:
w_{ti}
#Followers:
W_{ti}
Overlapping Index
Retweet
Probability

Conclusion

- Through delicately designed data collection and analyzing procedure and carefully constructed econometric model, we studied how social network characteristics are associated with information diffusion on Twitter.

- We found retweeting is more likely to occur from weak ties .

Managerial Implications

Our results have direct implications on how influence should be measured in a social broadcasting network, which is of great importance to managers who wish to harness the power of social broadcasting networks.

Influence

Simply counting the number of a user's followers is a coarse measurement of this user's influence. For a user, we should also take into account of:

- the tie strength between the user and his/her followers;
- the number of followers/followings of each of the user's follower;
- the overlapping index between the user and each of his/her followers.