MIS Quarterly

RESEARCH ARTICLE

CONTENT SHARING IN A SOCIAL BROADCASTING ENVIRONMENT: EVIDENCE FROM TWITTER

Zhan Shi
Department of Information Systems, W. P. Carey School of Business, Arizona State University,
Tempe, AZ 85287 U.S.A. {zhan.m.shi@asu.edu}

Huaxia Rui
Simon Graduate School of Business, University of Rochester, Rochester, NY 14627 U.S.A. {huaxia.rui@simon.rochester.edu}

Andrew B. Whinston
McCombs School of Business, The University of Texas at Austin, Austin, TX 78712 U.S.A. {abw@uts.cc.utexas.edu}

Appendix A

Operationalization of Weak Ties

In this appendix, we discuss our operationalization of weak ties used in our empirical analyses. We define tie strength based on the following–follower relationships observed in the Twitter network, and, specifically, we claim that unidirectional relationships are *on average* weaker than bidirectional ones. We want to stress a few points regarding this assumption. First, we are not claiming that a bidirectional relationship in the Twitter world is a strong tie in the *absolute* sense. Twitter users, even if they are mutually connected online, often barely *know* each other in the real world, so to a certain extent, the claim that almost all ties on Twitter are weak is a fair one to make. The hypothesis only emphasizes the ordinal strength of the two types of ties, and the comparison is carried out in the sense of *probabilistic expectation*. The reason why reciprocity makes a difference is that frequent learning or regular interaction is more likely to happen when a reciprocal relationship exists. By reading each other’s posts, a pair of users can more easily develop mutual understanding about each other’s topics of interest and expertise, and sometimes even about detailed aspects of each one’s personal life. Over time, even though the pair are unknown to each other in the real world, they might become very familiar with each other’s activities and habits in the online community. Of course, reciprocal following does not guarantee such relationship development (which is why we emphasize the probabilistic nature of the hypothesis). However, without it, the relationship development is much less likely. Moreover, our operationalization is consistent with the previous sociological literature. Granovetter (1973) pointed out the importance of reciprocity by defining that “the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie” (p. 1361). In Friedkin (1982), asymmetrical contact between college professors was classified as a weak tie, and a reciprocal connection was classified as a strong tie. Marlow et al. (2009) also applied similar definitions in analyzing friendships on Facebook.

We perform an empirical test on the hypothesis, using the network graph data we collected. Note that we know not only the number of followings (followers) a user has, but also who the followings (followers) are (i.e., we observe the IDs of the user’s immediate social neighbors in our database). This information should give us more knowledge about, and in the meantime the ability to build important metrics of, a user’s network characteristics. In particular, knowing the IDs of two users’ social neighbors, we can compare how “similar” their social neighborhoods are. In deriving his theory, Granovetter, in his 1973 paper, claimed that the stronger the social tie between two persons, the larger the overlap of their friendship circles. Applying this statement in the Twittersphere, under our assumption, we would expect that two users who mutually follow each other, on average, have a larger overlap in their followings (followers) than two who don’t. Our test is based on this prediction. Operationally, we do so by empirically verifying whether $w_{ti} = 0$ positively correlates with a higher similarity between user and author ’s followings (followers). We measure similarity by computing two overlap indexes of followings (followers) of author *t* and user *ti*:

$$OI_{ti}^{V_1} = \frac{\overline{V}_{ti}}{\sqrt{V_t \sqrt{V_{ti}}}}, OI_{ti}^{V_2} = \frac{\overline{V}_{ti}}{\min\{V_t, V_{ti}\}} \tag{8}$$

where $\overline{V}_{ti}$, V_t, and V_{ti} are the number of mutual followings author *t* and user *ti* shared, the number of followings author *t* had, and the number of followings *i* had, respectively (a similar "neighborhood overlap" was defined by Onnela et al. 2007). Similarly, we can define and compute overlap indexes of followers $\left(OI_{ti}^{W_1}, OI_{ti}^{W_2}\right)$ by changing *V* to *W* in equation (8). Note that the two numerators in equation (8) are the same: $\overline{V}_{ti}$. The difference between $OI_{ti}^{V_1}$ and $OI_{ti}^{V_2}$ is in the denominators, or in the way by which we scale down $\overline{V}_{ti}$ based on the number of followings *ti* has. Both indexes are in the range [0, 1] because $\overline{V}_{ti} \leq \min\{V_t, V_{ti}\}$. The larger the indexes are, we say the more similar the two sets of followings are. When *t* and *ti* have no mutual followings shared, both indexes equal 0. When *t* and *ti* have exactly the same sets of followings, $OI_{ti}^{V_1}$ =1. When *ti*'s followings represent a subset/superset of *t*'s followings, $OI_{ti}^{V_2}$ =1.

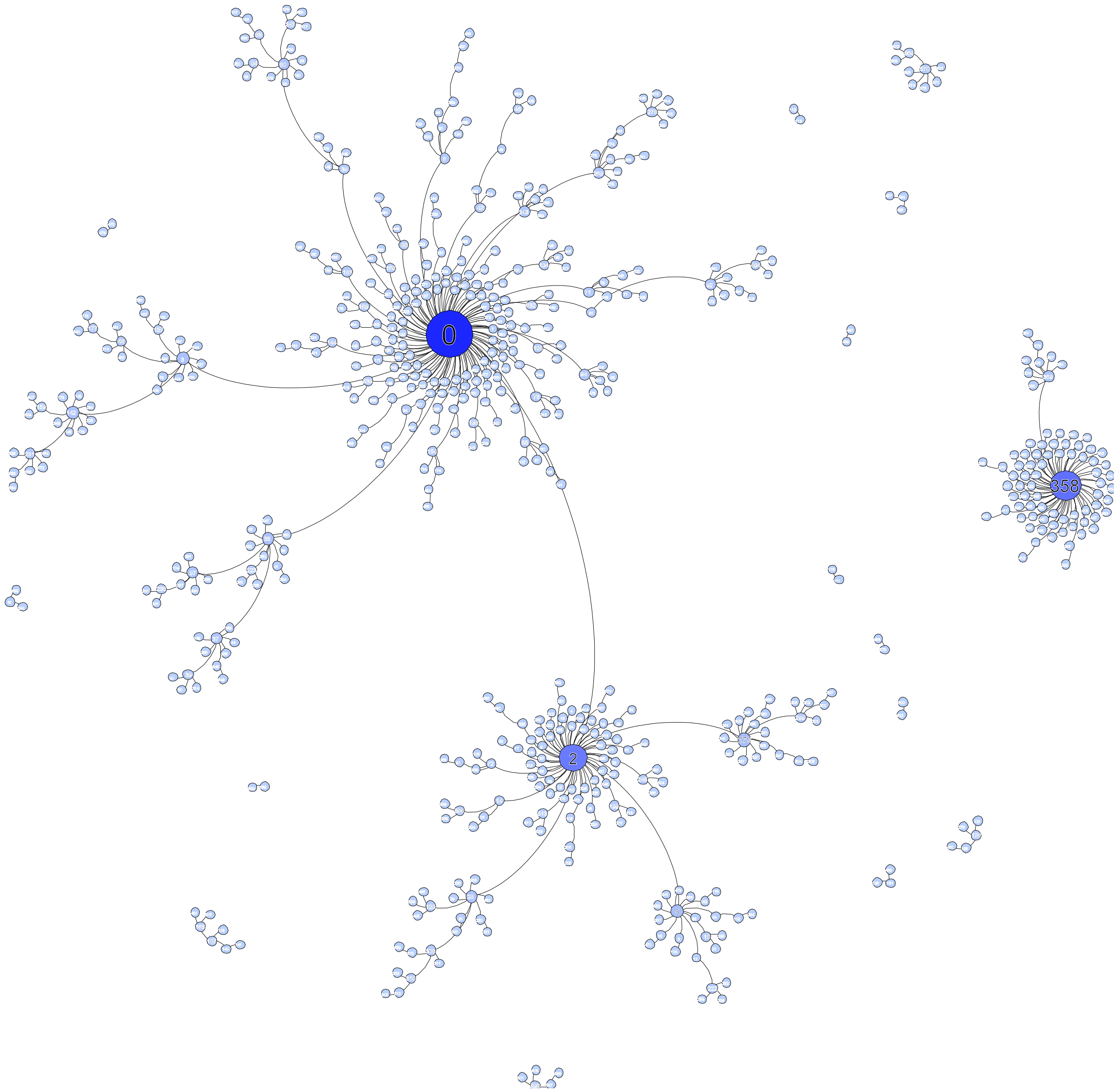
We investigate wether different w_{ti} values lead to significantly different overlap indexes by running a series of ANOVA tests, the results of which are given in Table A1. In all four tests, we control tweet-specific effects. As the regression coefficients in the first row show, we find that a unidirectional relationship w_{ti} = 1) is indeed associated with a smaller overlap in social neighborhoods. The *F* statistics and *p*-values indicate this difference is significant at the 0.1% level, no matter which index we use. Therefore, bidirectional relationships are associated with higher transitivity in social neighborhoods. The results thus support our hypothesis that unidirectional relationships are, on average, weaker than bidirectional ones.

Table A1. Results of Anova Tests

	OI^{V_1}	OI^{V_2}	OI^{W_1}	OI^{W_2}
w_{ti}	-0.042***	-0.069***	-0.034***	-0.064***
F	(2322.21)	(1476.43)	(3837.34)	(2158.65)
p-value	0.00	0.00	0.00	0.00

Appendix B

The Spead of a Single Tweet (t = 1) in Our Sample



Appendix C

Immediate-Follower Retweeters and Other Retweeters

In Table C1, we provide a breakdown of different types of retweeters for each tweet in our sample, including the bidirectional followers of the original author, the unidirectional followers of the original author, those second or higher order retweeters, and other retweeters who are either non-connected or protected. The average ratio of other retweeters is about 38 percent.[1] These other retweeters are most likely users who became exposed to the tweets in our sample after they searched certain keywords because tweets classified as Top Tweets often appear in the top part of the first page of search results if they match the keywords.[2]

Table C1. Number of Immediate-Follower Retweeters and Other Retweeters

t	Immediate Followers		Second and Higher Order Retweeters	Other Retweeters
	Bidirectional	Unidirectional		
1	78	10	545	42
2	7	3	681	787
3	0	11	927	493
4	5	6	423	351
5	2	4	452	1
6	0	8	413	70
7	1	6	399	265
8	2	31	775	341
9	2	6	615	229
10	1	1	292	230
11	1	14	584	46
12	3	15	399	52
13	2	3	264	310
14	7	46	246	117
15	1	9	471	17
16	4	25	435	16
17	3	15	437	110
18	2	13	437	213
19	0	6	363	433
20	2	61	627	58
21	0	27	263	290
22	0	16	275	296
23	1	9	336	319
24	2	10	175	122
25	2	9	256	96
26	1	5	134	280
27	0	23	226	9
28	2	13	350	306
29	70	64	631	87
30	6	1	309	418
31	3	10	215	371
32	1	15	361	281

[1]Note that this ratio is significantly larger than the ratio for randomly selected tweets (Appendix D). For example, among 52 retweets from 200 randomly selected tweets, 32 retweeters are immediate followers of the original authors, 14 are higher order followers, and 6 are from other retweeters.

[2]Indeed, one phenomenon that is consistent with our conjecture is that the proportion of "other retweeters" is significantly higher for tweets with a hashtag (45%) than for tweets without (34%). This is because when people click on a hashtag, Twitter automatically shows the search results containing that particular hashtag.

Table C1. Number of Immediate-Follower Retweeters and Other Retweeters (Continued)				
t	Immediate Followers		Second and Higher Order Retweeters	Other Retweeters
	Bidirectional	Unidirectional		
33	0	14	798	133
34	0	6	113	576
35	2	8	487	14
36	1	10	248	548
37	0	9	230	537
38	8	17	378	85
39	0	30	111	502
40	0	43	261	219
41	0	5	263	487
42	1	5	438	31
43	1	17	293	531
44	3	1	104	12
45	0	36	148	71
46	1	41	113	21
47	0	8	297	748
48	1	8	312	405
49	1	25	669	347
50	2	4	91	19
51	1	11	254	512
52	0	25	373	377
53	0	3	257	419
54	1	12	234	291
55	1	19	136	114
56	0	9	459	703
57	1	3	575	358
58	0	2	382	512
59	6	0	515	726
60	1	3	625	116
61	2	17	227	348
62	12	25	607	16
63	3	13	453	289
64	1	7	922	31
65	0	21	170	593

Appendix D

Random Sample

To investigate the generalizability of our results further, we also collected a random sample of (relatively recent) tweets from the entire Twittersphere (as opposed to TopTweets alone). In this appendix, we show key statistics from this sample. Interested readers can compare them with the ones shown in the main text.

This random sample contains 200 tweets, which were selected from Twitter's (official) public time line in September 2012.[3] We tracked the retweeting activity over a period of two weeks. In terms of the network structure, we collected the IDs of both the followings and the followers of the 200 authors, so that we could identify the unidirectional versus the bidirectional followers.

The 200 authors have 110,672 followers in total. Across the authors, the mean number of followers is 558, the median is 207, the minimum is 0, and the maximum is 13,894 (compared with the first row in Table 2); the mean percentage of proportion of unidirectional followers is 41.9%, the median is 36.4%, the minimum is 0.0%, and the maximum is 100.0% (compared with Figure 7). So, on average, the authors in the TopTweets sample (used in the main text) have a larger proportion of unidirectional followers.

Of the 200 tweets, 20 generated at least 1 retweet for a total of 52 retweets, distributed as follows: 1 tweet generated 16, 1 tweet generated 10, 3 tweets generated 3 each, 2 tweets generated 2 each, and 13 tweets generated 1 each. The authors' immediate followers account for 32 of the 52 retweets. In other words, the retweeting rate among the immediate followers is about 0.03 percent (32/110,672), much lower than that of the TopTweets sample (4.27%), which is not a surprise. *The 32 immediate followers are all unidirectional followers*. This finding is consistent with our key result in the main text.

Several key statistics that depict the distribution of the number of followings and followers are given in Table D1, which can be compared with the third and fourth rows of Table 3.[4] The statistics of the random sample are all larger than those of the TopTweets sample. This finding may be due to the growth of the Twitter network over the more-than-two-year period between July 2010 and September 2012.

Table D1. Descriptive Statistics of the Random Sample

		Mean	*std*	5%	15%	50%	85%	95%
V_{ti}	followings	3,223	16,475	45	120	552	1,984	9,274
W_{ti}	followers	4,527	73,676	11	42	289	1,594	12,300

References

Friedkin, N. 1982. "Information Flow Through Strong and Weak Ties in Intraorganizational Social Networks," *Social Networks* (3), pp. 273-285.

Granovetter, M. 1973. "The Strength of Weak Ties," *The American Journal of Sociology* (78:6), pp. 1360-1380.

Marlow, C., Byron, L., Lento, T., and Rosenn, I. 2009. "Maintained Relationships on Facebook," Overstated (http://overstated.net/2009/03/09/maintained-relationships-on-facebook).

Onnela, J., Saramäki, J., Hyvõnen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. 2007. "Structure and Tie Strengths in Mobile Communication Networks," *Proceedings of the National Academy of Sciences of the U.S.A.* (104:18), pp. 7332-7336.

[3]The public time line is an aggregated stream of all public tweets. We wrote a program to visit the public time line and pick the most recently published tweets. We ran the program every hour to ensure that the tweets were uniformly distributed across the clock hours.

[4]We know the *number* of followings and followers of the 110,672 users, but not their *IDs*.