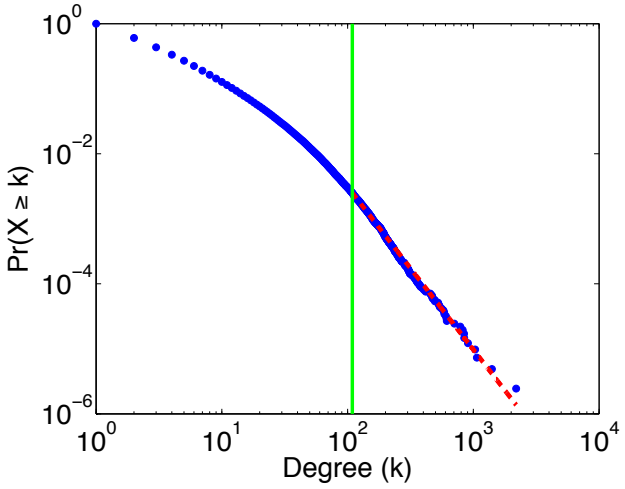


Figure A1: Log-log plot of the complementary cumulative distribution function (CCDF) of the degree distribution for a sample month (January 2009) network is shown (blue), along with the best fitting power law model ($\alpha = 3.50$ and $k_{\min} = 109$) using the procedure of Clauset, Shalizi, and Newman [41]. We test whether the empirical distribution is distinguishable from a power law using the Kolmogorov-Smirnov test and find no evidence against the null hypothesis ($D = 1.82\times10^{-2}$, $p = 0.35$, $n = 495881$) data. This distribution may be fit by a power law.

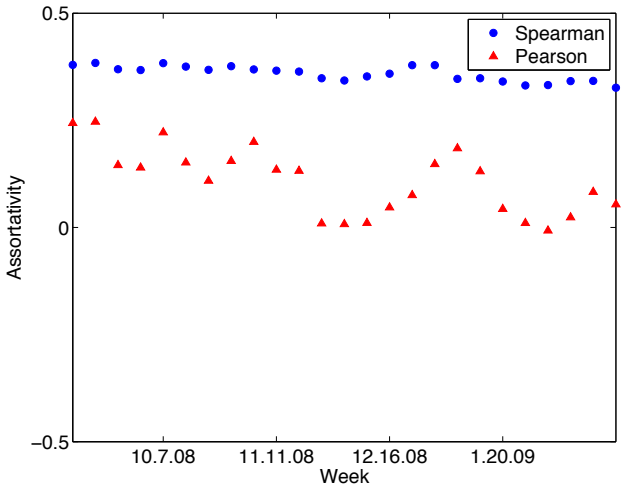
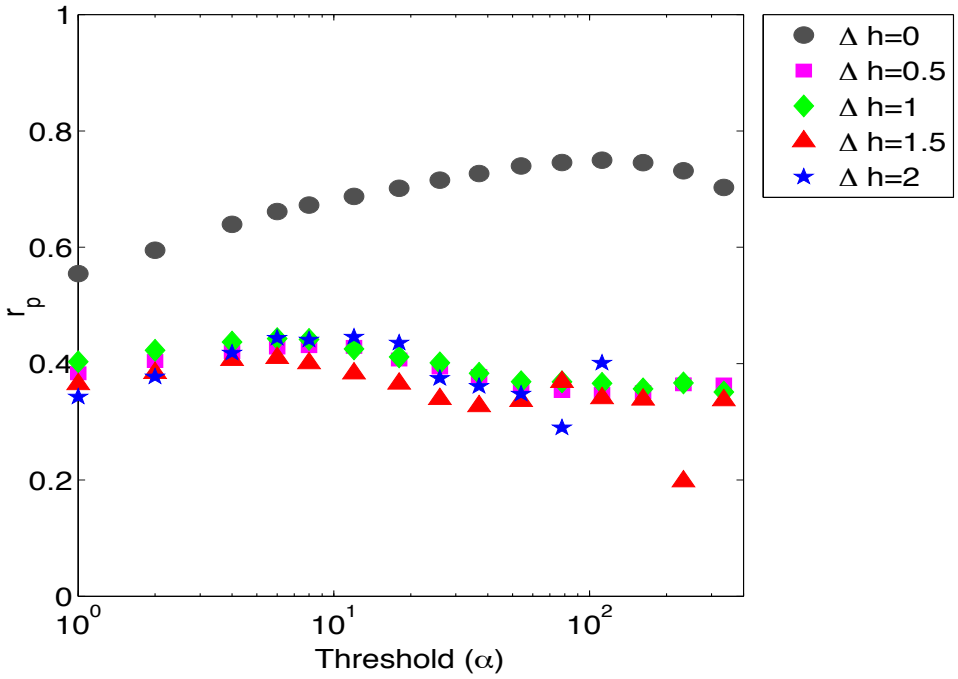
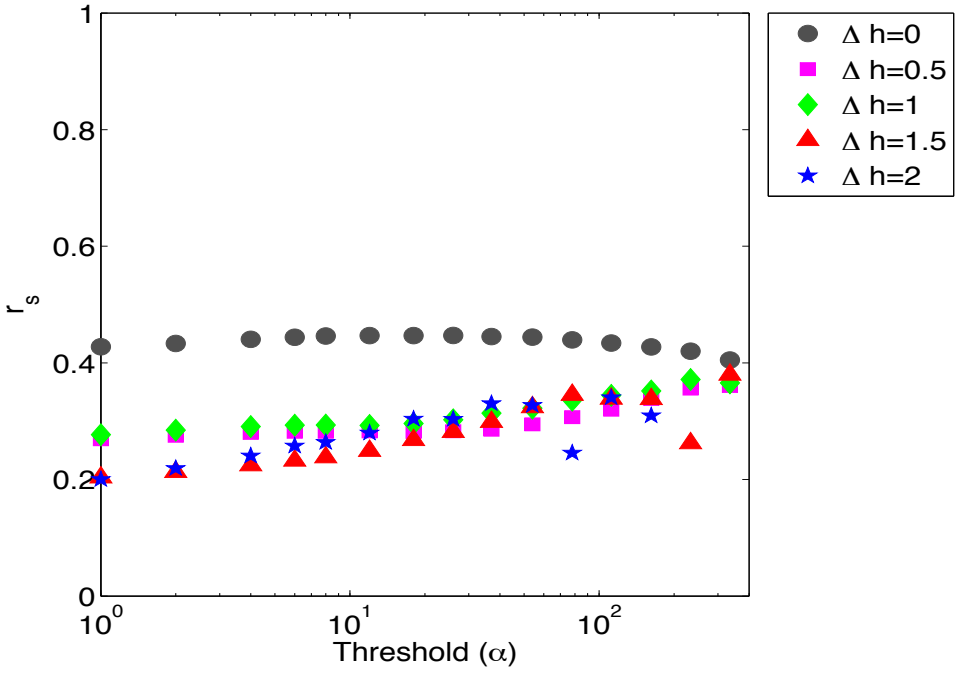


Figure A2: Spearman and Pearson correlation coefficients are used to measure degree assortativity. The Pearson correlation coefficient is more sensitive to extreme values. As a result, the Pearson correlation coefficient obscures the trend that the network is assortative with respect to the rank of node degrees. Given the nature of the degree distribution and the questions that we are asking, we use the Spearman correlation coefficient for our study.



(a) Assortativity of happiness, Pearson's r



(b) Assortativity of happiness, Spearman's r

Figure A3: Measured happiness assortativity as threshold for labMT word usage increases for a single week network. The Spearman correlation coefficient (right) exhibits less variability as compared to the Pearson correlation coefficient (left). Notice that when $\Delta h = 0$, there is less variation due to the numerous words centered around the mean happiness score, regardless of the threshold, α.

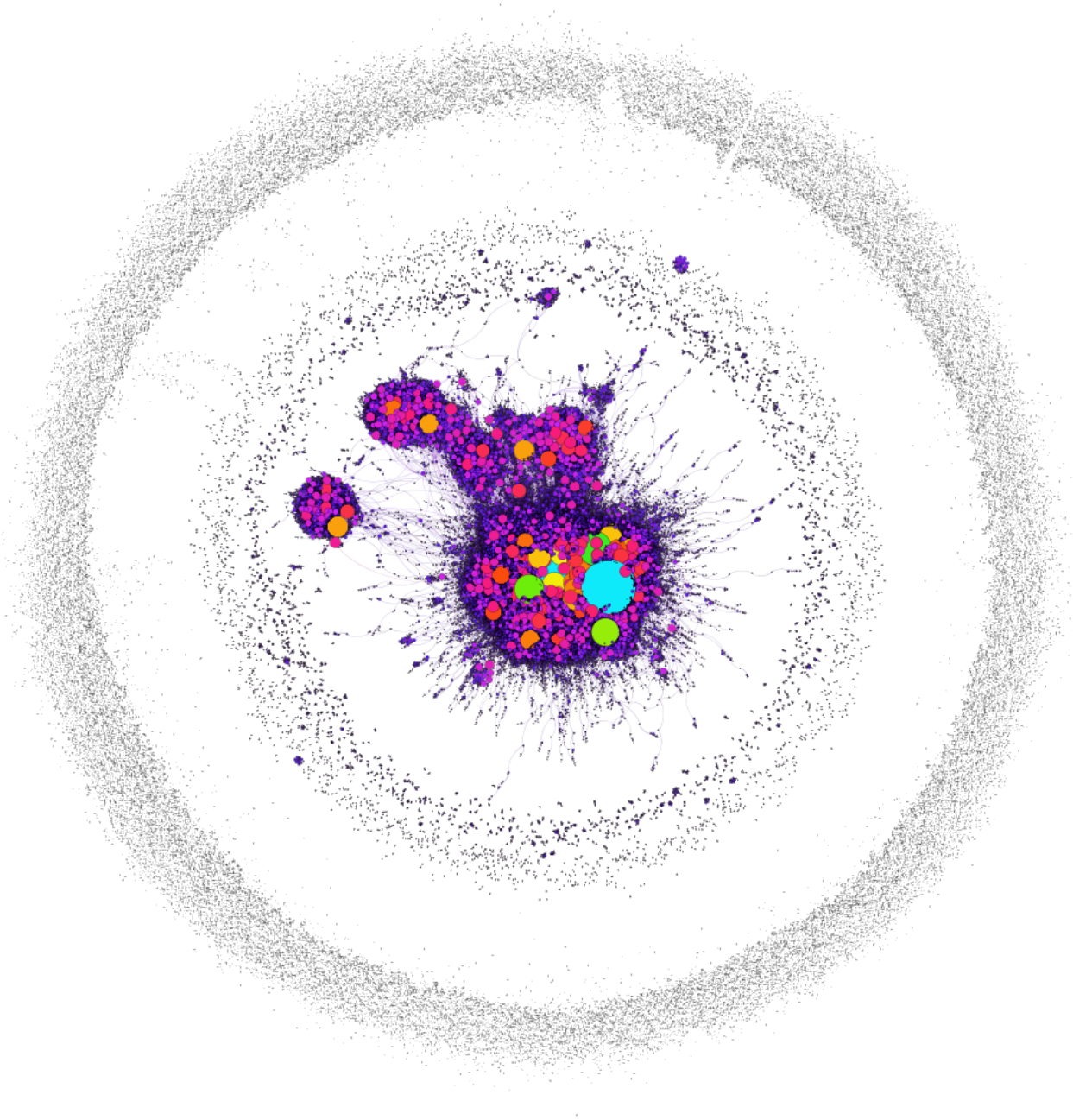


Figure A4: A visualization of the reciprocal reply network for the week beginning September 9, 2008 (Week 1) is depicted. The size of a node is proportional to the degree, and colors further emphasize the degree detected by Gephis implementation of the algorithm suggested by Blondel et al. [40].

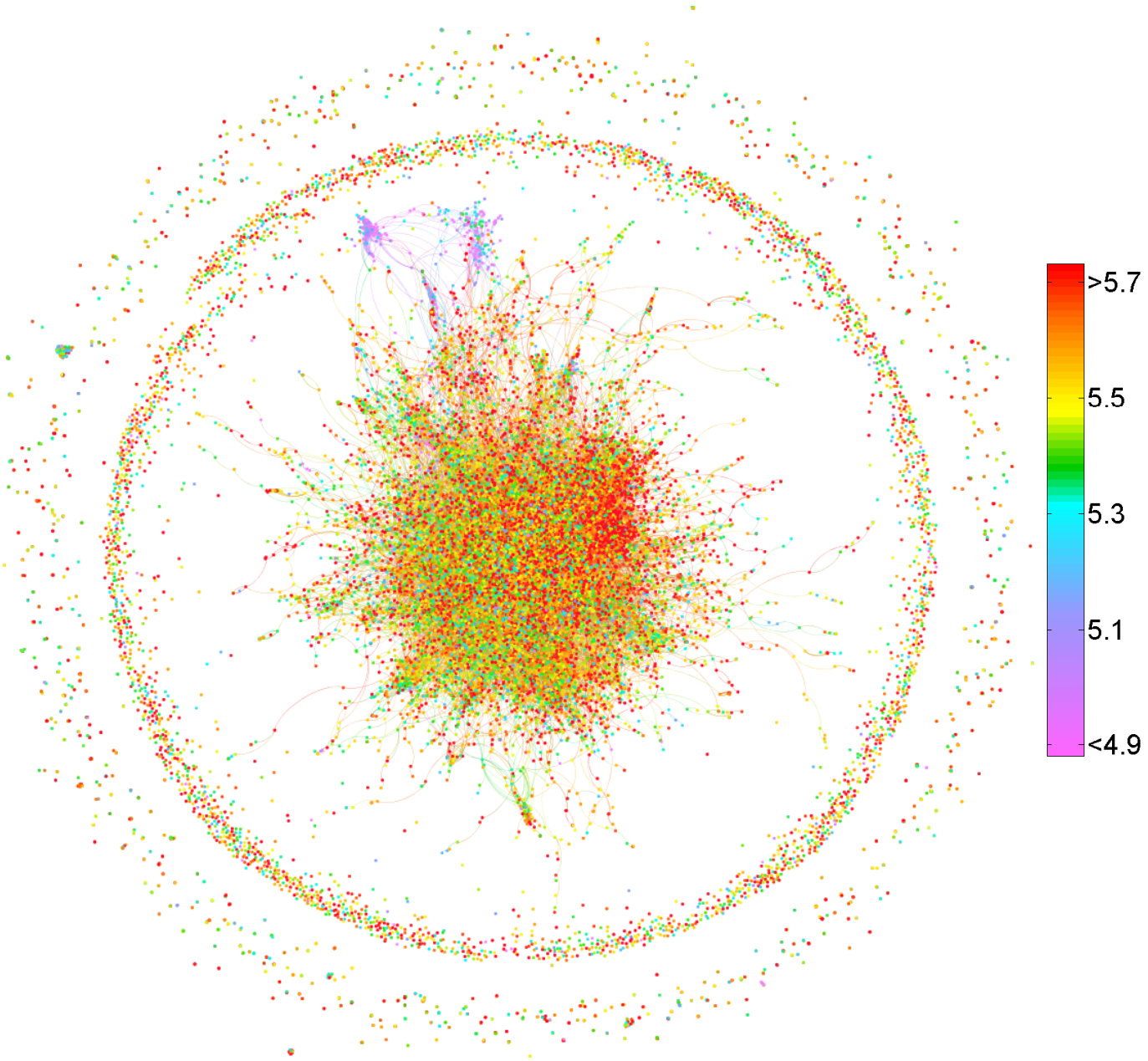


Figure A5: A visualization of the reciprocal reply network for the week beginning September 9, 2008 (Week 1). Colors represent happiness scores for nodes with greater than $\alpha = 50$ labMT words (57% of all nodes in the week). The visualization was produced using Gephi [39]. The algorithm employed by the software clusters nodes according to their connectivity. Collections of nodes with similar colors provide a visualization of the happiness is assortativity finding.

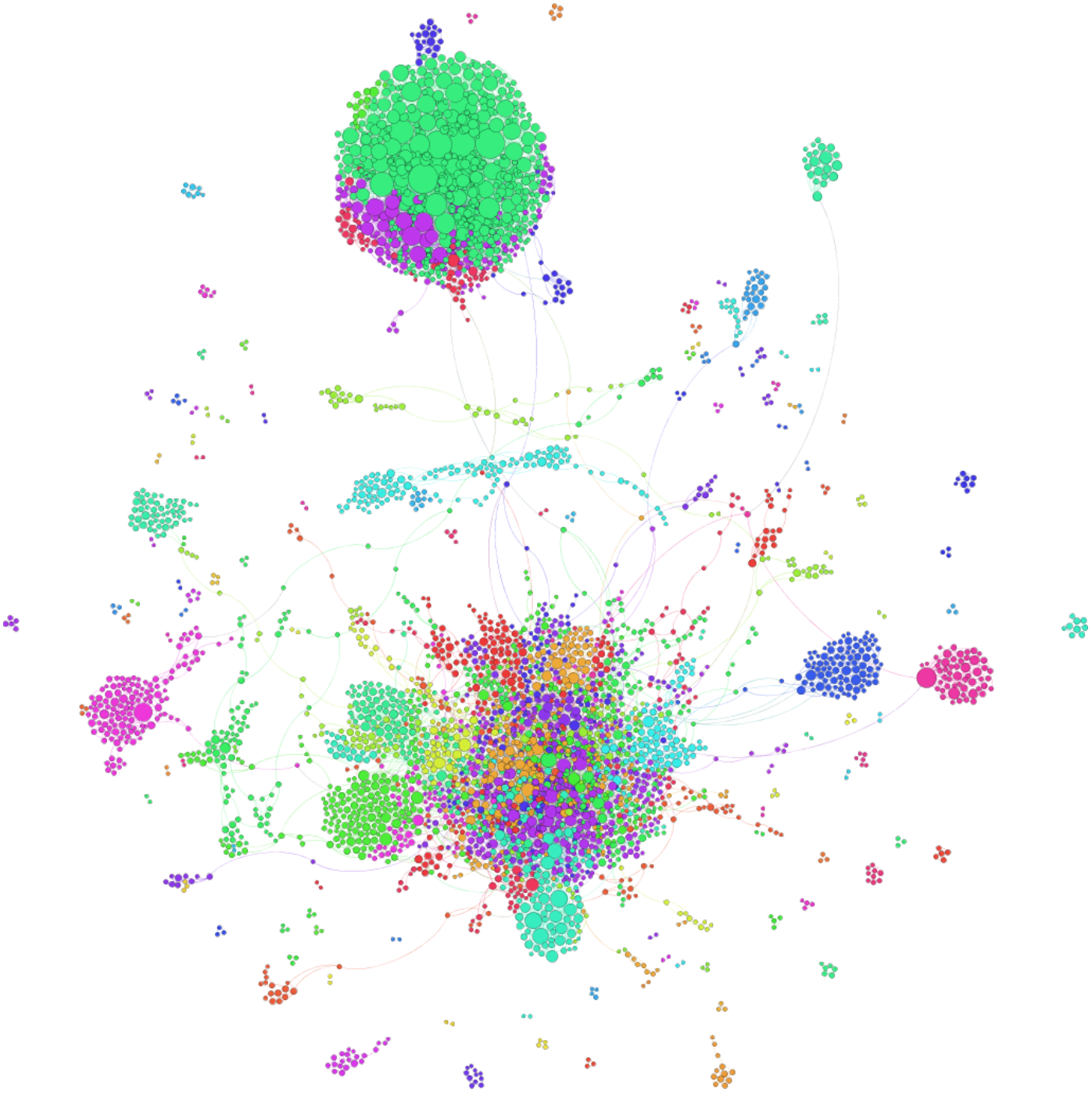


Figure A6: A visualization of the reciprocal reply network for the day of October 28, 2008. The size of the nodes is proportional to the degree and colors indicate communities detected by Gephi's implementation of the community detection algorithm suggested by [49].

.

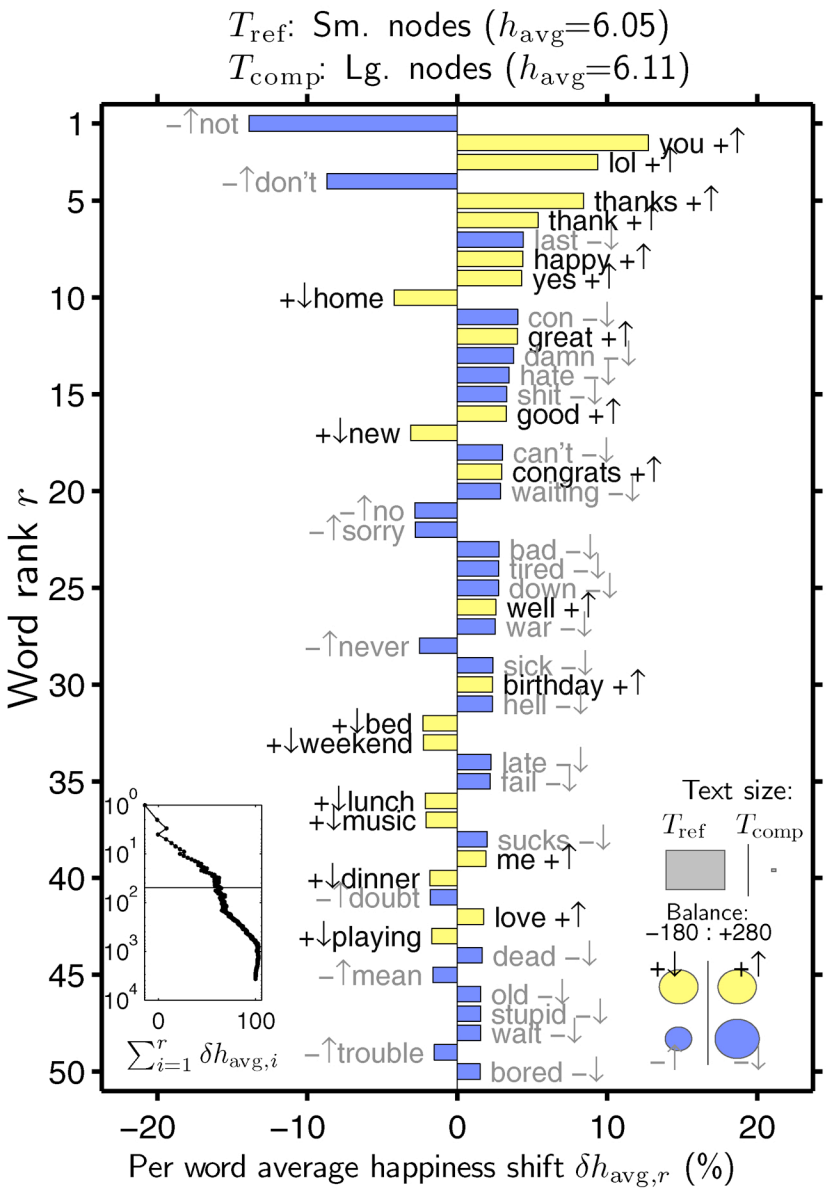


Figure A7: The collection of words used by nodes with degree $k \geq 100$ ($T_{\rm comp}$) is compared to words written by users with degree $k < 100$ ($T_{\rm ref}$). Words "award" and "awards" were excluded because their usage dominated the wordbag of one high degree node (Twitters ShortyAwards). The word "die" was also excluded to eliminate the possibility of the hedonometer incorrectly being applied to German tweets. We note that removal of these words resulted in a negligible change to all values of r_s reported in the paper. The horizontal bars on the right side of the plot represent words which raise the happiness score of $T_{\rm comp}$. The symbols of +/− and ↑ / ↓ combine to convey whether a positive/negative word appears more/less frequently in the $T_{\rm comp}$ as compared to the $T_{\rm ref}$. Notice that an increase in the usage of positive words (e.g., "you"), as well as a decrease in the use of a negative word (e.g., "last") will contribute to $T_{\rm comp}$ having a higher happiness score. On the left hand side of the word shift plot are words which contribute to lowering the happiness score of $T_{\rm comp}$. Such examples include an increase in the usage of negative words (e.g., "not") as well as a decrease in the usage of positive words (e.g., "home"). The magnitude of the bars indicate the relative contribution of each word to these effects. In summary, we see that $T_{\rm comp}$ has a higher happiness score than does $T_{\rm ref}$. In the lower right, the relative text sizes are depicted as rectangles proportional to the number of words. The reference text, $T_{\rm ref}$, has considerably more words in its collection than does $T_{\rm comp}$. The circle plots depicted the relative amount of positive vs. negative words contained in $T_{\rm ref}$ and $T_{\rm comp}$. While both collections are similar in terms of positive word usage, the collection of words used by larger nodes contains fewer negative words and thus, this contributes to the slightly higher happiness score for this collection of words. The lower left inset shows the cumulative sum of individual word contributions as a function of $\log_{10} r$, where r is the rank of the 3,686 labMT words. See [11] for the full details of the wordshift graph.

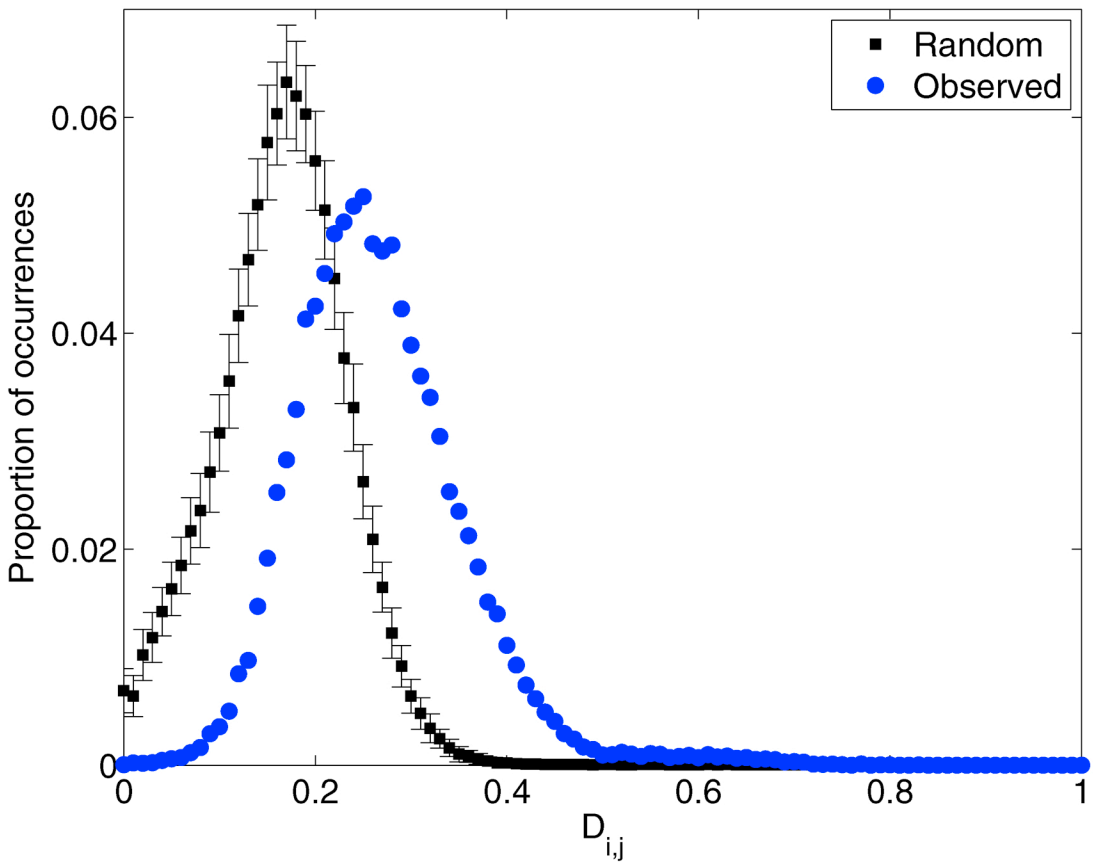


Figure A8: The similarity of word bags for pairs of users connected in a week reciprocal reply network is computed as follows: For users i and j, we compute $D_{i,j} = 1 - \frac{1}{2}\sum_{n=1}^{3686} |f_{i,n} - f_{j,n}|$, where $f_{i,n}$ represents the normalized frequency of word usage of the nth labMT word by user i. The value of $D_{i,j}$ ranges from 0 (dissimilar word bags) to 1 (similar word bags). The proportion of occurrences of user-user pairs in the reciprocal reply network for a sample week (Sept. 16, 2008) having word similarity indices between 0 and 1 are shown (blue dots), with $\alpha = 50$ and $\Delta h = 1$. The majority of user-user similarity indices are less than 0.4, indicating that users and their nearest neighbors use dissimilar collections of words in their tweets. We then perform 100 random permutations of word vector assignments to users, while holding the network topology intact (black squares). The resulting distributions show that while users are using more similar words than would be expected by chance, this shift is small. The mean score for randomized user-user paired word collections is $\overline{D_{i,j}} = .167$. This value is not zero, since users are using a common language (English). The mean score for our observed network data is $\overline{D_{i,j}} = .267$, which is slightly higher than the randomized value due to conversations occurring between these users.

Rank	Word	Frequency ($\times 10^5$)	Happiness	Rank	Word	Frequency ($\times 10^5$)	Happiness	Rank	Word	Frequency ($\times 10^5$)	Happiness
1	you	103.55	6.24	51	sure	6.82	6.32	101	music	4.24	8.02
2	my	94.91	6.16	52	done	6.81	6.54	102	found	4.23	6.54
3	me	56.35	6.58	53	show	6.73	6.24	103	doesn't	4.23	3.62
4	not	39.98	3.86	54	awesome	6.72	7.60	104	online	4.23	6.72
5	up	36.04	6.14	55	check	6.51	6.10	105	party	4.20	7.58
6	no	34.40	3.48	56	bed	6.42	7.18	106	soon	4.20	6.34
7	new	34.03	6.82	57	sleep	6.33	7.16	107	thinking	4.15	6.28
8	like	31.75	7.22	58	cool	6.32	7.20	108	snow	4.14	6.32
9	all	30.71	6.22	59	live	6.29	6.84	109	give	4.13	6.54
10	good	30.20	7.20	60	big	6.28	6.22	110	movie	4.12	6.84
11	will	23.58	6.02	61	free	6.18	7.96	111	ha	4.09	6.00
12	we	22.59	6.38	62	life	6.17	7.32	112	sorry	4.08	3.66
13	day	21.80	6.24	63	old	6.07	3.98	113	real	4.06	6.78
14	know	19.45	6.10	64	didn't	6.04	4.00	114	kids	3.98	7.38
15	more	19.32	6.24	65	find	6.00	6.00	115	phone	3.91	6.44
16	don't	18.29	3.70	66	die	6.00	1.74	116	tv	3.91	6.70
17	today	18.24	6.22	67	video	5.99	6.48	117	stop	3.89	3.90
18	love	17.66	8.42	68	house	5.99	6.34	118	play	3.88	7.26
19	think	17.45	6.20	69	christmas	5.89	7.96	119	waiting	3.88	3.68
20	see	15.28	6.06	70	playing	5.77	7.14	120	lunch	3.81	7.42
21	great	14.60	7.88	71	world	5.76	6.52	121	food	3.79	7.44
22	lol	13.35	6.84	72	game	5.54	6.92	122	reading	3.76	6.78
23	thanks	13.09	7.40	73	wow	5.54	7.46	123	god	3.74	7.28
24	home	13.05	7.14	74	ready	5.53	6.58	124	top	3.65	6.76
25	people	12.71	6.16	75	iphone	5.53	6.54	125	buy	3.60	6.28
26	night	12.70	6.22	76	listening	5.41	6.28	126	book	3.56	7.24
27	blog	12.26	6.02	77	pretty	5.40	7.32	127	car	3.56	6.72
28	last	11.89	3.74	78	always	5.39	6.48	128	idea	3.52	7.06
29	well	11.70	6.68	79	help	5.27	6.08	129	friend	3.51	7.66
30	make	11.27	6.00	80	read	5.07	6.52	130	family	3.51	7.72
31	right	11.04	6.54	81	google	5.05	7.20	131	yay	3.47	6.10
32	can't	10.93	3.42	82	everyone	5.03	6.12	132	glad	3.47	7.48
33	morning	10.38	6.56	83	most	4.95	6.22	133	least	3.46	4.00
34	very	10.10	6.12	84	wait	4.88	3.74	134	nothing	3.44	3.90
35	first	9.69	6.82	85	start	4.87	6.10	135	late	3.43	3.46
36	our	9.26	6.08	86	please	4.79	6.36	136	internet	3.39	7.48
37	better	8.89	7.00	87	con	4.78	3.70	137	amazing	3.38	7.66
38	us	8.82	6.26	88	try	4.77	6.02	138	mean	3.38	3.68
39	tonight	8.79	6.14	89	thought	4.69	6.38	139	myself	3.37	6.30
40	down	8.73	3.66	90	school	4.66	6.26	140	facebook	3.34	6.08
41	happy	8.40	8.30	91	thank	4.64	7.40	141	funny	3.32	7.92
42	tomorrow	7.88	6.18	92	weekend	4.56	8.00	142	tired	3.29	3.34
43	nice	7.80	7.38	93	hey	4.48	6.06	143	talk	3.29	6.06
44	best	7.61	7.18	94	wish	4.44	6.92	144	damn	3.26	2.98
45	she	7.57	6.18	95	hate	4.42	2.34	145	interesting	3.26	7.52
46	yes	7.42	6.74	96	haha	4.41	7.64	146	own	3.24	6.16
47	fun	7.37	7.96	97	friends	4.41	7.92	147	friday	3.23	6.88
48	hope	7.34	7.38	98	making	4.40	6.24	148	open	3.18	6.10
49	bad	6.98	2.64	99	dinner	4.27	7.40	149	lost	3.16	2.76
50	never	6.92	3.34	100	coffee	4.27	7.18	150	guys	3.16	6.22

Table A1: The top 150 most frequently occurring words from the labMT list in our Sept 2008 through Feb 2009 data set, where stop words ($4 < h_{\text{avg}} < 6$) have been removed.

Rank	Word	Frequency ($\times 10^5$)	Happiness	Rank	Word	Frequency ($\times 10^5$)	Happiness	Rank	Word	Frequency ($\times 10^5$)	Happiness
1	the	295.60	4.98	51	an	22.73	4.84	101	el	11.76	4.80
2	to	249.91	4.98	52	we	22.59	6.38	102	well	11.70	6.68
3	i	221.28	5.92	53	some	22.32	5.02	103	oh	11.69	4.84
4	a	218.13	5.24	54	que	22.26	4.64	104	who	11.64	5.06
5	and	135.23	5.22	55	day	21.80	6.24	105	should	11.48	5.24
6	is	127.94	5.18	56	how	21.64	4.68	106	over	11.34	4.82
7	in	122.94	5.50	57	going	20.64	5.42	107	make	11.27	6.00
8	of	121.79	4.94	58	am	20.60	5.38	108	then	11.15	5.34
9	for	114.41	5.22	59	go	20.03	5.54	109	right	11.04	6.54
10	you	103.55	6.24	60	has	19.68	5.18	110	can’t	10.93	3.42
11	on	96.97	5.56	61	or	19.55	4.98	111	way	10.84	5.24
12	my	94.91	6.16	62	know	19.45	6.10	112	only	10.72	4.92
13	it	91.09	5.02	63	more	19.32	6.24	113	getting	10.63	5.68
14	that	69.81	4.94	64	la	18.77	5.00	114	his	10.56	5.56
15	at	58.51	4.90	65	don’t	18.29	3.70	115	morning	10.38	6.56
16	with	56.42	5.72	66	today	18.24	6.22	116	very	10.10	6.12
17	me	56.35	6.58	67	too	18.15	5.22	117	after	9.82	5.08
18	just	50.25	5.76	68	they	18.09	5.62	118	watching	9.76	5.84
19	have	49.86	5.82	69	work	17.95	5.24	119	her	9.73	5.84
20	be	46.10	5.68	70	got	17.91	5.60	120	them	9.71	4.92
21	this	45.75	5.06	71	love	17.66	8.42	121	first	9.69	6.82
22	de	44.38	4.82	72	think	17.45	6.20	122	e	9.66	4.72
23	so	40.93	5.08	73	back	17.37	5.18	123	that’s	9.55	5.28
24	not	39.98	3.86	74	twitter	17.18	5.46	124	rt	9.52	4.88
25	i’m	39.89	5.74	75	when	16.84	4.96	125	y	9.47	4.48
26	are	39.03	5.16	76	there	16.39	5.10	126	than	9.42	4.74
27	but	37.78	4.24	77	had	15.30	4.74	127	its	9.36	4.96
28	was	37.74	4.60	78	see	15.28	6.06	128	our	9.26	6.08
29	up	36.04	6.14	79	en	14.97	4.84	129	better	8.89	7.00
30	out	35.20	4.62	80	really	14.93	5.84	130	us	8.82	6.26
31	now	35.12	5.90	81	off	14.89	4.02	131	tonight	8.79	6.14
32	no	34.40	3.48	82	great	14.60	7.88	132	down	8.73	3.66
33	new	34.03	6.82	83	need	14.45	4.84	133	i’ve	8.59	5.58
34	do	33.96	5.76	84	he	14.34	5.42	134	u	8.40	5.52
35	from	33.78	5.18	85	still	13.74	5.14	135	happy	8.40	8.30
36	like	31.75	7.22	86	been	13.43	5.04	136	again	8.34	5.42
37	your	31.43	5.60	87	lol	13.35	6.84	137	could	8.34	5.52
38	all	30.71	6.22	88	would	13.15	5.38	138	un	8.15	4.64
39	good	30.20	7.20	89	thanks	13.09	7.40	139	into	8.08	5.04
40	get	30.04	5.92	90	home	13.05	7.14	140	i’ll	8.05	5.38
41	what	29.46	4.80	91	want	12.81	5.70	141	man	7.99	5.90
42	about	28.97	5.16	92	people	12.71	6.16	142	tomorrow	7.88	6.18
43	it’s	27.14	4.88	93	night	12.70	6.22	143	nice	7.80	7.38
44	if	25.21	4.66	94	here	12.28	5.48	144	any	7.70	5.22
45	by	24.66	4.98	95	o	12.26	4.96	145	take	7.63	5.18
46	as	24.50	5.22	96	blog	12.26	6.02	146	best	7.61	7.18
47	time	24.19	5.74	97	why	12.10	4.98	147	she	7.57	6.18
48	one	23.73	5.40	98	much	11.92	5.74	148	even	7.42	5.58
49	will	23.58	6.02	99	last	11.89	3.74	149	yes	7.42	6.74
50	can	23.57	5.62	100	did	11.84	5.58	150	little	7.38	4.60

Table A2: The top 150 most frequently occurring words from the labMT word list in our Sept 2008 through Feb 2009 data set including stop words.

Week	Start date	N	$< k >$	$k_{\max}$	C_G	Assort	# Comp.	S
1	09.09.08	95647	2.99	261	0.10	0.24	10364	0.71
2	09.16.08	99236	2.95	313	0.10	0.24	11062	0.71
3	09.23.08	99694	2.90	369	0.09	0.13	11457	0.70
4	09.30.08	100228	2.87	338	0.09	0.13	11752	0.69
5	10.07.08	78296	2.60	241	0.09	0.21	11140	0.63
6	10.14.08	122644	3.20	394	0.09	0.14	12221	0.74
7	10.21.08	130027	3.30	559	0.08	0.09	12420	0.75
8	10.28.08	144036	3.56	492	0.08	0.14	12319	0.78
9	11.04.08	145346	3.54	330	0.08	0.19	12597	0.78
10	11.11.08	136534	3.35	441	0.08	0.12	12972	0.76
11	11.18.08	153486	3.46	444	0.08	0.13	13594	0.77
12	11.25.08	155753	3.46	1244	0.06	0.00	14122	0.77
13	12.02.08	165156	3.44	1245	0.06	0.01	14496	0.78
14	12.09.08	162445	3.33	1456	0.05	0.01	15342	0.76
15	12.16.08	148154	3.12	730	0.06	0.04	15645	0.73
16	12.23.08	140871	3.22	575	0.07	0.07	15216	0.72
17	12.30.08	143015	3.30	519	0.07	0.15	15272	0.73
18	01.06.09	170597	3.19	253	0.07	0.18	17234	0.74
19	01.13.09	188429	3.29	477	0.07	0.13	18403	0.75
20	01.20.09	196038	3.16	680	0.06	0.04	19927	0.74
21	01.27.09	203852	3.04	973	0.05	0.01	21537	0.73
22	02.03.09	212513	2.92	1718	0.04	-0.01	24387	0.71
23	02.10.09	213936	2.83	828	0.06	0.02	25854	0.70
24	02.17.09	215172	2.65	437	0.06	0.07	28742	0.67
25	02.24.09	170180	2.27	320	0.06	0.04	28388	0.58

Table A3: Network statistics for reciprocal-reply networks by week. As Twitter popularity grows, so does the number of users (N) in the observed reciprocal-reply network. The average degree ($< k >$), degree assortativity, the number of nodes in the giant component (# Comp.), and the proportion of nodes in the giant component (S) remain fairly constant, whereas the maximum degree ($k_{\max}$) shows a great deal of variability from month to month. Clustering (C_G) shows a slight decrease over the course of this period.

Week	Start date	# Obsvd. Msgs. $\times 10^6$	# Total Msgs. $\times 10^6$	% Obsvd. $\left(\frac{\#Obsvd.}{\#Total} \times 100\right)$	# Replies $\times 10^6$	% Replies $\left(\frac{\#Replies}{\#Obsvd.} \times 100\right)$
1	09.09.08	3.14	7.26	43.2	0.88	28.1
2	09.16.08	3.36	8.31	40.4	0.90	26.9
3	09.23.08	3.43	8.89	38.6	0.90	26.2
4	09.30.08	3.33	9.06	36.8	0.89	26.6
5	10.07.08	2.33	9.38	24.8	0.64	27.5
6	10.14.08	4.39	9.87	44.4	1.24	28.3
7	10.21.08	4.70	10.01	47.0	1.35	28.8
8	10.28.08	5.74	10.34	55.5	1.64	28.5
9	11.04.08	5.58	11.14	50.1	1.63	29.3
10	11.11.08	4.70	9.88	47.6	1.42	30.2
11	11.18.08	5.48	11.34	48.3	1.67	30.5
12	11.25.08	5.71	11.47	49.8	1.73	30.2
13	12.02.08	5.54	12.85	43.1	1.80	32.4
14	12.09.08	5.41	13.54	39.9	1.72	31.7
15	12.16.08	4.57	12.72	35.9	1.45	31.8
16	12.23.08	4.80	11.62	41.3	1.46	30.5
17	12.30.08	4.61	13.48	34.2	1.50	32.5
18	01.06.09	5.16	16.11	32.0	1.72	33.3
19	01.13.09	5.73	17.33	33.1	1.97	34.4
20	01.20.09	5.82	18.87	30.9	1.98	34.1
21	01.27.09	5.75	20.79	27.6	1.98	34.5
22	02.03.09	5.78	22.42	25.8	2.01	34.8
23	02.10.09	5.66	23.39	24.2	1.99	35.1
24	02.17.09	5.43	25.71	21.1	1.91	35.1
25	02.24.09	3.80	20.75	18.3	1.34	35.1

Table A4: The number of "observed" messages in our database comprise a fraction of the total number of Twitter message made during period of this study (September 2008 through February 2009). While our feed from the Twitter API remains fairly constant, the total # of tweets grows, thus reducing the % of all tweets observed in our database. We calculate the total # of messages as the difference between the last message id and the first message id that we observe for a given month. This provides a reasonable estimation of the number of tweets made per month as message ids were assigned (by Twitter) sequentially during the time period of this study. We also report the number observed messages that are replies to specific messages and the percentage of our observed messages which constitute replies.