A Cross-Sectional and Temporal Analysis of Information Consumption on Twitter

Completed Research Paper

Srikar Velichety
Eller College of Management,
University of Arizona
1130 E Helen St, Tucson, AZ
srikarv@email.arizona.edu

Sudha Ram
Eller College of Management,
University of Arizona
1130 E Helen St, Tucson, AZ
sram@email.arizona.edu

Abstract

We report on an exploratory analysis of the similarities and differences among three different forms of information consumption on Twitter viz., following, listing and subscribing. We construct a cross- sectional and temporal framework to analyze the relationships among these three forms. Our analysis reveals several interesting patterns of information consumption on Twitter. First, we find that people not only consume information by following others explicitly but also by listing and subscribing to lists and that the people they list or subscribe to are not the same as the ones they follow. Second, we find that listing and following are more similar to each other than listing and subscribing or subscribing and following. Using temporal analysis, we find that initially, people prefer to use following as a form of information consumption while subscription is a more volatile form of information consumption than following or listing.

Keywords: Microblogging, Twitter, Lists, Subscription, Membership.

Introduction

Twitter has revolutionized the way we communicate online. Using short text restricted to 140-characters, people tweet about their daily activities, seek information (Java et al. 2007), communicate brand sentiments (Jansen et al. 2009) and also communicate during disasters such as earthquakes (Qu et al. 2009). Recent research by Rui et al. (2011) shows that the sentiments of Tweets about movies can be used to accurately predict their sales.

In September 2009, Twitter introduced a feature called Lists that allows users to group other users and follow their activities separately. A user can follow tweets of specific users by adding them to a list. When a user adds a user to a list, that user becomes a list member. Similarly, when a user finds a list created by someone else, the user can subscribe to that list and thus follow the tweets of its members. Hence there are two types of users with respect to lists viz., members – those who are added to a list created by a user and subscribers – those who subscribe to a list voluntarily because of their interest. Thus any user on Twitter can consume information produced by others in three ways viz., by directly following them, by creating a list and adding them, and/or by subscribing to a list in which other users are members (Another way of consuming information is to go to the list in which a specific user is a member and consume tweets of other members in the list. We do not consider this here since it doesn't involve explicit action on part of the concerned user). A user can combine these three forms in any way. For instance, a user can add someone to her list without explicitly following her. Similarly, she can subscribe to a list without following or listing any of the members of that list.

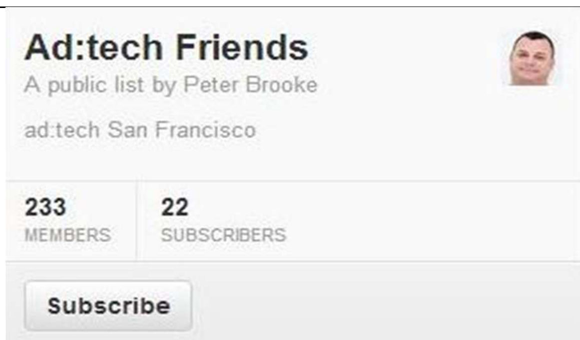


Figure 1. Sample List

Figure 2. Home Page

Figure 1 shows a part of the home page of a list titled "Ad:tech Friends". Note that the list is curated by a user "Peter Brooke" and has a description "ad:tech San Francisco". The list has 223 members and 22 subscribers. As shown in figure 2, the account with twitter handle "peterbrooke" is in fact following only 153 people. Assuming that all the 153 people he is following belong to the list, there are at least 80 people whom he is not following explicitly but he is still consuming information from them via his list. This further implies that this user considers following and listing as two different forms of information consumption.

While there is some research on lists to infer latent user characteristics (Ghosh et al. 2012; Pochampally and Varma 2011; Sharma et al. 2012; Y. Yamaguchi et al. 2011) and a few studies on the importance of follower and followee patterns (Krishnamurthy et al. 2008; Kwak et al. 2010), surprisingly there is very little research so far on the similarities and differences between these forms of information consumption. An understanding of this phenomenon will not only help us understand information consumption in a better way, but can also provide valuable insights into user behavior and microblogging usage in general. Velichety and Ram (2013 b) propose a framework to compare following with listing and subscribing and show that following is different from listing and subscribing. However they use a relatively small sample of 100 Twitter users. Moreover they do not examine differences between listing and subscribing.

In this research, we substantially extend the framework proposed by Velichety and Ram (2013 b). Using a larger sample of 907 Twitter users, we examine differences among these three forms of consumption. In addition, we propose a temporal framework to examine how these forms are related across different time periods to infer consumption behaviors of Twitter users. Our work also provides a robust framework for developing a combined information consumption and production model on Twitter.

Past Work on Twitter Lists

Lists allow Twitter users to distinctly categorize users from whom they wish to receive tweets into distinct categories. When a user creates a list and adds users to it, he/she can receive tweets of only those accounts by going to the list. Also when creating a list, the user can provide a short (100 character) description for it. This in many cases serves as an indicator of the interests of the focal user and those of the members of the list. Leveraging this information, Sharma et al. (2012) developed a framework for identifying characteristics of Twitter users. Their results showed that this information helps in accurately identifying latent user characteristics as well as their popular perceptions on Twitter. Ghosh et al. (2012) developed a search system leveraging the list feature to identify topic experts on Twitter. Pochampally and Varma (2011) combined list information with user interactions to identify topics of interest to a user. Yamaguchi et al. (2011) showed that user interests can be more accurately identified using lists to which she belongs to rather than her tweets. Hence, methods using list-based approaches to identify user interests outperform those that use either tweet based or profile information based approaches. Zhao and Ram (2011) examined the patterns of triadic closures in lists. Velichety and Ram (2013 a) used a network analysis approach to examine Twitter lists with a view of developing a list based recommendation system. While all this research looks at lists as a potential source of identifying user interests and characteristics, very little is known about the similarities and differences between following, listing and subscribing.

Twitter allows users to add someone -whom she may or may not be following explicitly- to a list thus allowing the focal user to consume information via a different mode. Similarly a user is free to subscribe to any public list without explicitly following any of the members of that list. Velichety and Ram (2013 b) provided preliminary evidence to show that following is different from listing or subscribing. In this research we extend their framework to compare how subscribing and listing are different from each other. Also considering that a Twitter user can first follow another user explicitly but then decide to list her or subscribe to a list to which she belongs or vice versa, it would be interesting to know how and why Twitter users switch between different modes of information consumption. Such an analysis is important for two reasons. First, it can help us understand the most common changes in information consumption modes. Second, when combined with an analysis of tweets, it can also help us understand the role of each of these forms of information consumption and the triggers for changes in consumption patterns. To this end, we develop a temporal framework that can help us understand the relationship among these three forms across time.

Data Collection

We generated an initial dataset of lists and their members from listorious.com[1]. First we generated a random sample of lists using keywords derived from the categorization of news articles on NY Times. We used the Tweepy module[2] in python to collect data on these lists. We initially examined the data for a sample of 100 users and found that there were almost no changes in list memberships, subscriptions or following on a weekly basis. We therefore collected the data once a month. Collecting the data once on following, members of the lists they are curating and the members of the lists they subscribed to for 907 curators in our sample takes a month given the restrictions of the Twitter REST API. We therefore collected the data continuously twice over a two month period. Table 1 provides the descriptive statistics of our dataset:

Table 1 Descriptive Statistics

	Count		Average (Per Curator)		Median		Standard Deviation	
	Period 1	Period 2	Period 1	Period 2	Period 1	Period 2	Period 1	Period 2
Lists Curated	10146	10030	11.18	11.05	11	11	6.23	6.23
Members*	623232	622736	687.13	686.58	787.5	785	1305.75	1312.65
Curator Friends** (Following)	2378401	2407438	2622.27	2654.28	1554	1577	18212.45	18590.66

Curator Subscriptions***	9906	9587	20.19	12.97	4	4	20.19	19.96
Subscribed Members****	757905	739530	835.61	815.35	707	677.5	3216.88	3193.72

*indicates the number of unique members in all the lists of all the curators.
** indicates cumulative number of users followed by all curators.
*** indicates cumulative number of lists to which all the curators have subscribed.
****indicates the cumulative number of people in all the lists to which curators have subscribed.

Cross Sectional Analysis

In this section we develop a framework to understand the similarities and differences between Following, Listing and Subscribing by extending the framework and metrics proposed by Velichety and Ram (2013 b) to compare following with listing and subscribing. Since we have three modes of information consumption we have 3 pairwise combinations for comparison. Note that n(X) defines the cardinality of the set X and C, L, F & S represent Curator, Listing, Following and Subscribing respectively in the names of the metrics we define. For example CF Ratio means Curator Following Ratio.

First we examine the relationship between following and listing. For any user on Twitter, we can divide the rest of the Twitter population into the following 4 disjoint sets.

		Followed	
		No	**Yes**
Listed	**No**	W	X
	Yes	Y	Z

Figure 3. Listing and Following Matrix

Let W, X, Y, Z represent the corresponding sets and n (W), n(X), n(Y) and n(Z) their cardinalities. For any curator i, W will be the have the highest cardinality. We define three metrics to understand the relationships between listing and following.

$$CF_{Ratio} = \frac{n(Z)}{n(Y) + n(Z)}$$

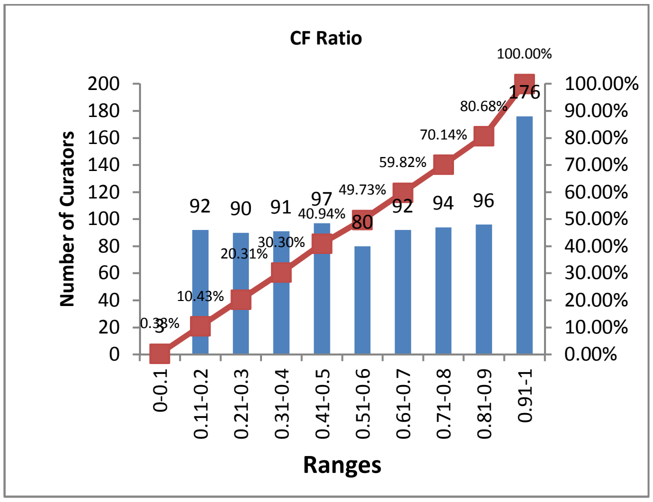
$$CM_{Ratio} = \frac{n(Z)}{n(X) + n(Z)}$$

$$MF_{Ratio} = \frac{n(Y)}{n(X)}$$

[1] http://www.listorious.com

[2] https://github.com/tweepy/tweepy

The CF Ratio indicates "Of all the people that a Twitter user has listed across one or more of her public lists, how many of them is she explicitly following?" while the CM Ratio indicates "Out of all the people that the curator is explicitly following, how many of them has she listed at least once across her public lists?". The MF Ratio compares the relative strength in numbers of those who are only listed to those who are only followed.



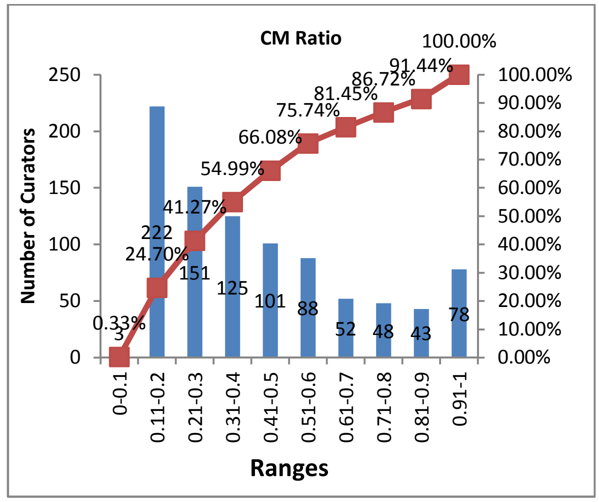


Figure 4. CF Ratio	Figure5. CM Ratio

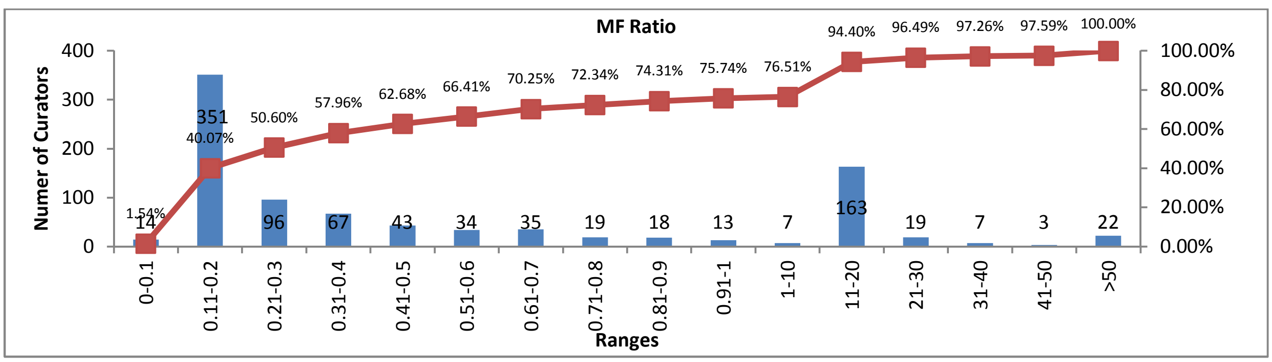


Figure 6. MF Ratio

Figures 4, 5 and 6 show the distributions of CF, CM and MF ratios for the 907 curators in our dataset. From Figure 4 it is clear that there are a substantial number of curators who have listed people without following them. The mean CF Ratio for the sample was 0.49 indicating that curators follow only half of the people they list on average. However the standard deviation was 0.29 indicating that there is sufficient heterogeneity in this case. Similarly, the average CM Ratio was 0.32 indicating that curators only list a third of the people they follow explicitly. The standard deviation in this case was also substantial (σ=.27) indicating heterogeneity here as well. Finally, the fact that MF Ratio had an average of 10.5 and standard deviation of 168.9 indicates that there are certain people who prefer to consume information primarily through listing while there are others who prefer to do so primarily through following. All three metrics together provide sufficient evidence for the fact that following and listing are two forms of information consumption that differ from each other substantially. Figure 7 shows a three dimensional representation of the cardinalities of the sets X, Y and Z for all the curators in our sample.

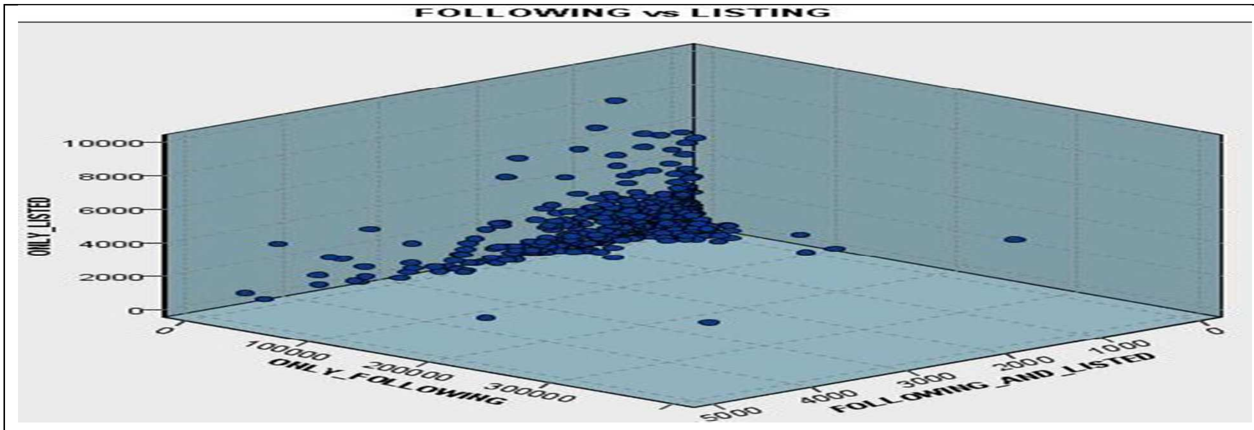


Figure 7. Following and Listing Triplet

Since each of three sets viz., only following, only listing and following, and only listing are disjoint for any particular curator, if we find a data point that lies on or close to one of the corresponding axes, we can conclude that the curator corresponding to that point prefers to consume information primarily in one specific form. Since we find data points that lie either on or close to all the axes, we can conclude that there is sufficient heterogeneity in terms of either listing or following.

Besides listing and following explicitly, Twitter users have a third option for information consumption viz., subscribing to a list that someone has already created to see the tweets of all the members of that list. This creates a third dimension for our analysis. Hence we investigate the relationship between subscribing to a list and following users explicitly. Similar to the matrix defined above, we describe the relationship between these two forms of following in this matrix.

		Followed	
		No	**Yes**
Subscribed	**No**	A	B
	Yes	C	D

Figure 8. Subscribed and Following Matrix

*Subscribed in this case means that they belong to the lists to which the curator has subscribed to

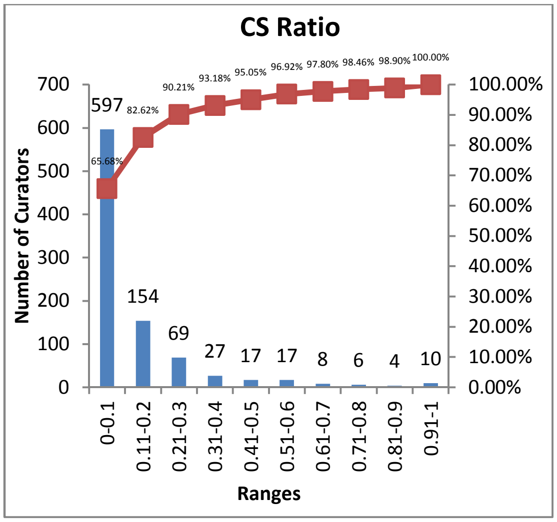
Similar to the metrics defined in the previous case, we define the following ratios

$$\mathrm{CS}_{\mathrm{Ratio}} = \frac{\mathrm{n(D)}}{\mathrm{n(C)} + \mathrm{n(D)}}$$

$$\mathrm{CFS}_{\mathrm{Ratio}} = \frac{\mathrm{n(D)}}{\mathrm{n(B)} + \mathrm{n(D)}}$$

$$\mathrm{SF}_{\mathrm{Ratio}} = \frac{\mathrm{n(C)}}{\mathrm{n(B)}}$$

These three metrics together define the relationship between following explicitly and following by subscribing to a list.



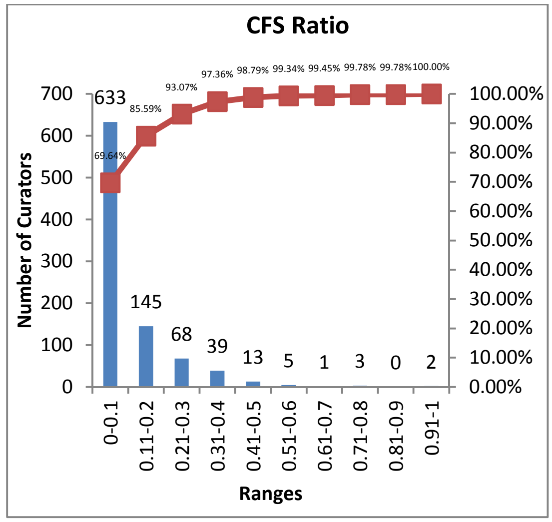


Figure 9. CS Ratio | **Figure 10. CS Ratio**

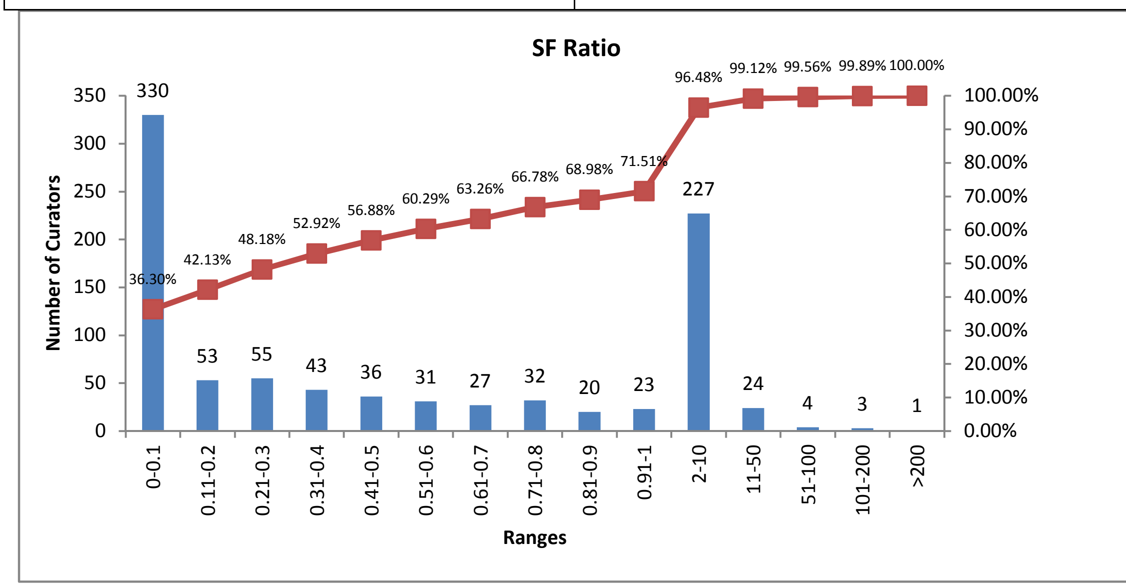


Figure 11. SF Ratio

Figures 9, 10 and 11 show the relationships between following and subscription. It is clear that the distributions here are more skewed than the previous ones. The average CS ratio was 0.11 (σ=0.17) indicating that curators follow a meager 11% of the people in the lists they subscribe to. Further, the average CFS ratio was 0.08 (σ=0.12) indicating that a mere 8% of the people that a curator follows are also members in the lists to which they subscribe. These two ratios provide evidence for the fact that the people followed by Twitter users are substantially different from members of the lists to which they subscribe. Finally the fact that SF Ratio had a mean of 2.49 (σ=12.18) shows that some users prefer to consume information primarily through following while others through subscribing. Figure 12 shows the three dimensional representation of the cardinalities of the sets B, C and D. This indicates heterogeneity in terms of following and subscription.

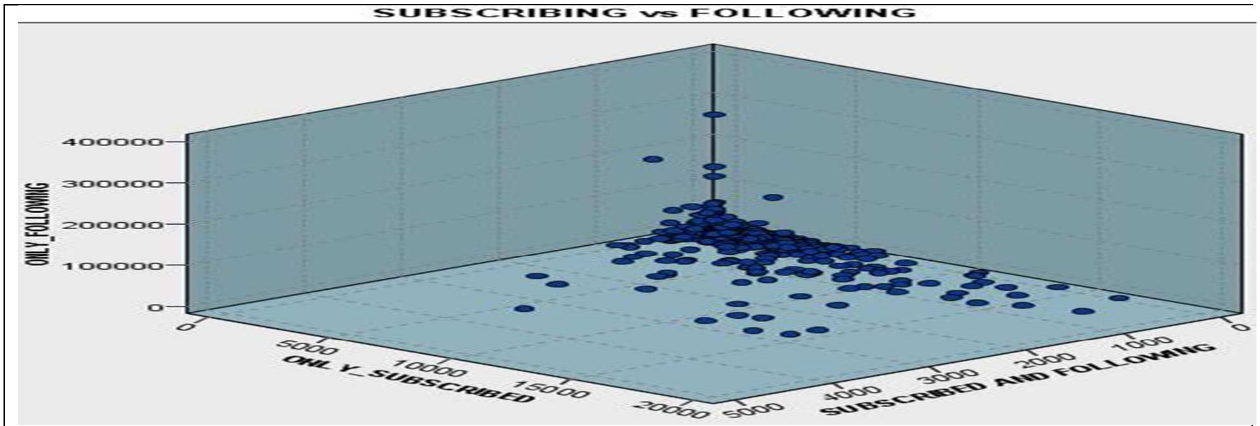


Figure 12. Subscribing and Following Triplet

Finally to understand the relationship between creating one's own list and subscribing to an already created list, we have the following matrix and the relevant metrics.

		Subscribed	
		No	**Yes**
Listed	**No**	P	Q
	Yes	R	S

Figure 13. Listing and Subscribing Matrix

$$\mathrm{CL_{Ratio}} = \frac{\mathrm{n(S)}}{\mathrm{n(S)+n(R)}}$$

$$\mathrm{CLS_{Ratio}} = \frac{\mathrm{n(S)}}{\mathrm{n(S)+n(Q)}}$$

$$\mathrm{LS_{Ratio}} = \frac{\mathrm{n(R)}}{\mathrm{n(Q)}}$$

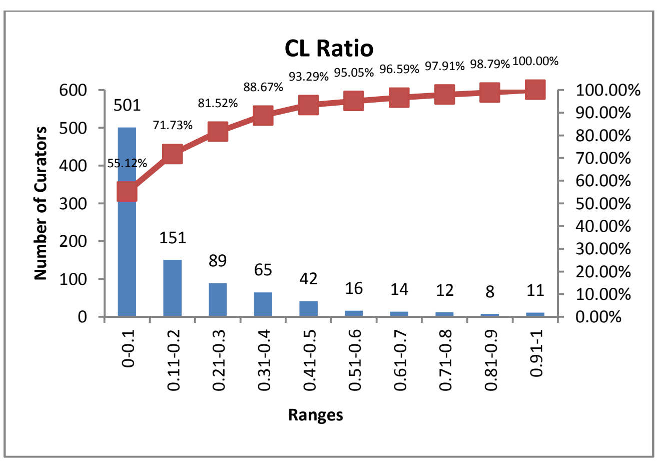


Figure 14. CL Ratio

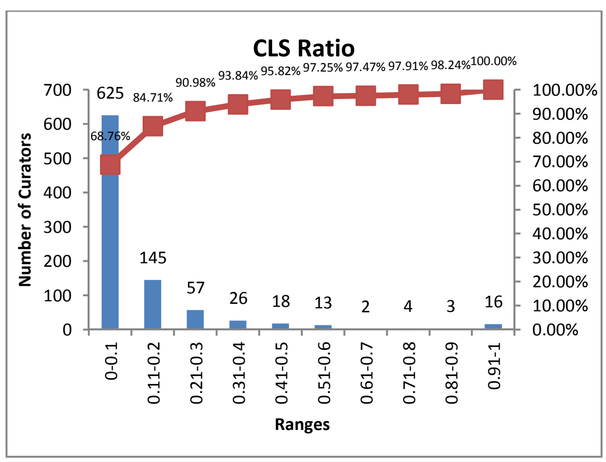


Figure 15. CLS Ratio

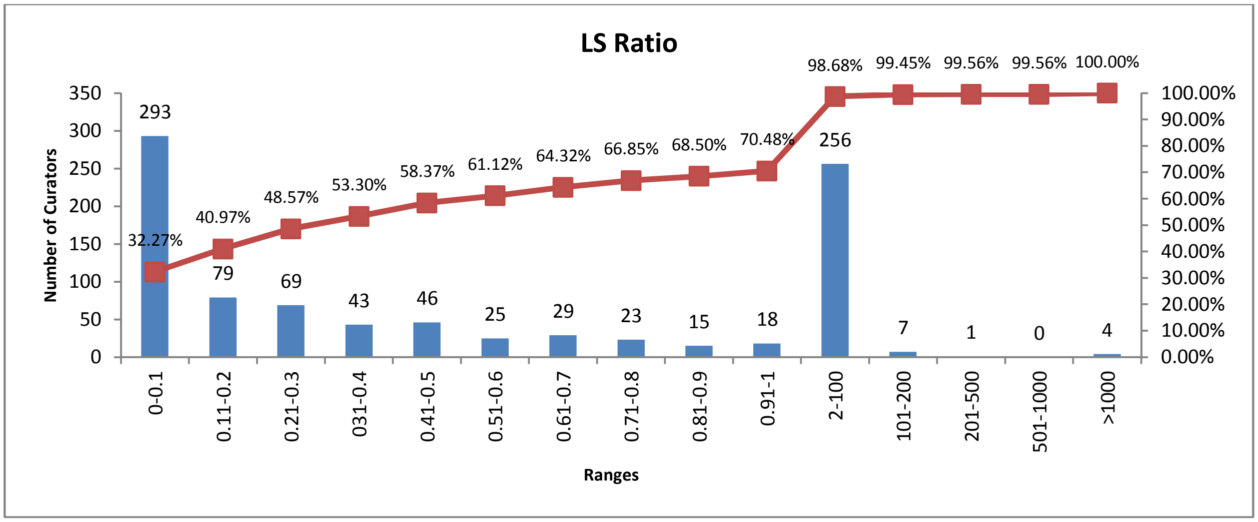


Figure 16. LS Ratio

Figures 14, 15 and 16 show the relationship between listing and subscribing. The CL Ratio had a mean of 0.15 (σ=0.20) indicating that out of all the people that the user has listed across her public lists, a mere 15% are also members of the lists to which she subscribes. The CLS Ratio had a mean of 0.109 (σ=0.17) meaning that out of all the members of the lists that a user has subscribed to, only 10% are also members of the public lists of the user. The distributions of these two ratios provide sufficient evidence for the fact that listing and subscribing are two different forms of information consumption. Further the fact that LS Ratio had a mean of 10.9 (σ=109) indicates that some users prefer to consume information by creating their own lists and adding people to them, while others primarily subscribe to the lists that other Twitter users have created. Figure 17 shows the three dimensional representation of the cardinalities of the sets Q, R and S. This indicates heterogeneity in terms of listing and subscription.

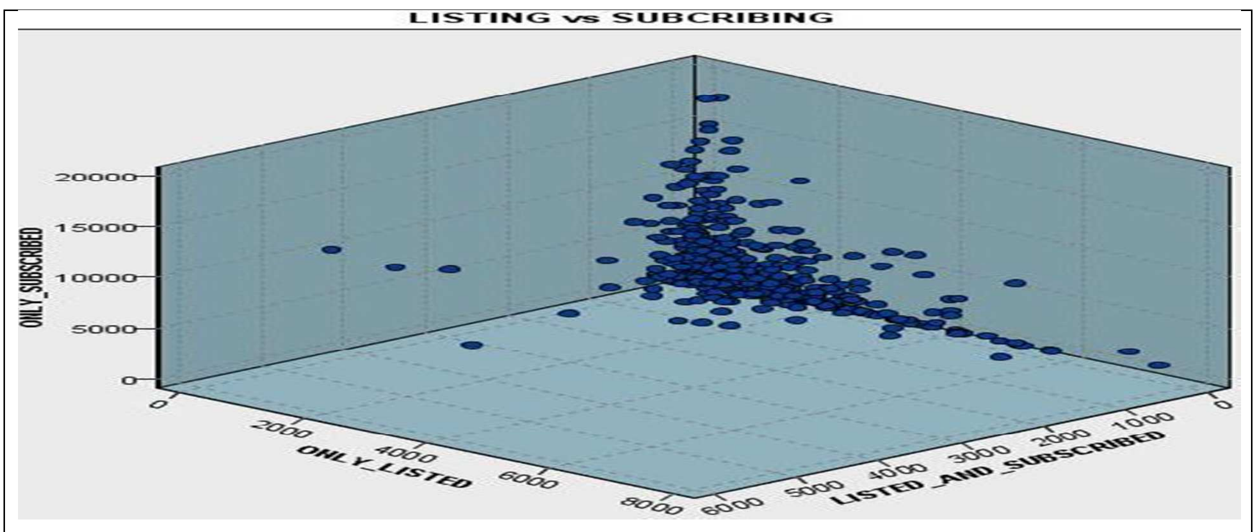


Figure 17. Listing and Subscribing Triplet

We also found that all these ratios remain consistent across different time periods for each curator suggesting the possibility of strong user specific heterogeneity. Table 2 gives a description of all the ratios we have defined. In order to understand the similarities and differences between pairs of information consumption forms, we conducted a series of paired t-tests and two sample Kolmogorov-Smirnov tests the results of which are documented in table 3. For example it would be useful to know if listing and following are more similar to each other than listing and subscribing or subscribing and following. In order to know this, we use the paired t-test on CF and CS Ratios and on CM and CFS Ratios.

Table 2 Metrics and Descriptions	
Metric	**Description**
CF Ratio	Out of all the people a curator has listed, how many is she following.
CM Ratio	Out of all the people a curator is following, how many has she listed.
MF Ratio	The ratio of the cardinalities of people a curator is “only following” to those she has “only listed”.
CS Ratio	Out of all the members in the subscribed lists of a curator, how many is she following.
CFS Ratio	Out of all the people a curator is following, how many belong to the subscribed lists.
SF Ratio	The ratio of the cardinalities of the people who only belong to subscribed lists to those who are only being followed for the curator.
CL Ratio	Out of all the people a curator has listed in her across her lists, how many belong to her subscribed lists.
CLS Ratio	Out of all the people who belong to the subscribed lists of a curator, how many belong to the lists she is curating.
LS Ratio	The ratio of the number of users who belong only to the curated lists of a curator to those who only belong to her subscribed lists.

Table 3 Pair Wise Comparison		
Pairs	**t-statistic**	**Kolmogorov-Smirnov statistic**
CF Ratio, CS Ratio	38.5309***	0.4763***
CS Ratio, CLS Ratio	1.4231	0.0572
CF Ratio, CL Ratio	28.5561***	0.5281***
CM Ratio, CFS Ratio	26.1521***	0.6293***
MF Ratio, SF Ratio	0.1414	0.2002***
SF Ratio, LS Ratio	0.3689	0.2464***
LS Ratio, MF Ratio	1.3017	0.2376***

***p<0.0001

The results in the table above reveal several interesting facts about the relationships among listing, subscribing and following. First, the difference between CF and CS Ratios was significant meaning people usually follow a greater fraction of people in their own lists than in the lists to which they subscribe. Also the significant difference between CF and CL Ratios suggests that if curator has listed a certain number of people, there is a greater chance of her following them rather than subscribing to the lists in which they are members. Together, these two facts point to a greater similarity between listing and following rather than between listing and subscribing. The statistically significant difference between CM and CFS ratios adds strength to this fact. Also the fact that the difference between CS and CLS Ratios was not statistically significant and that their distributions are similar –as evidenced by the KS Statistic- further exacerbates these differences. Second, as indicated by the KS-statistic, the distributions of the MF, SF and LS ratios are different though their absolute values do not differ significantly indicating the difference between these relevant forms of information consumption.

To summarize, we find evidence of substantial differences among three different forms of information consumption on Twitter. We also find that listing and following are more similar to each other than listing and subscribing or following and subscribing. Finally, we find that there is a large heterogeneity among Twitter users in terms of preference for a specific form of information consumption. The next section discusses the temporal relationship among these three forms of information consumption.

Temporal Analysis

We first create a framework for temporal analysis by partitioning the set of users - from whom a focal user in our sample is consuming information - into eight pairwise disjoint sets as shown in table 4. Figure 18 is a visual representation of these sets.

Table 4 Information Consumption Sets

Category	Description
A	Only Followed
B	Only Listed
C	Only Subscribed
D	Followed and Listed but not Subscribed
E	Listed and Subscribed but not Followed
F	Followed and Subscribed but not Listed
G	Listed, Followed and Subscribed
H	Neither Listed, Nor Followed and Nor Subscribed

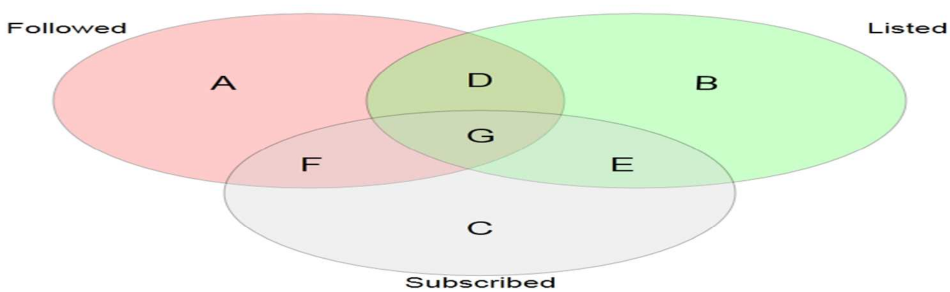


Figure 18. Information Consumption Set

Note that it is not possible to get the exact cardinality of set H. However, in the context of this research, we consider set H as the prospective information consumption set for the focal user, meaning, the user is not consuming information in any of the three forms at present but may decide to do so in the future. Accordingly, we set the cardinality of this set to be three times the cardinality of the union set created from sets A-G.

For any two consecutive time period's t and t+1, we can construct a transition probability matrix as follows

t \ t+1	A	B	C	D	E	F	G	H
A	X_{AA}	X_{AB}	X_{AC}	X_{AD}	X_{AE}	X_{AF}	X_{AG}	X_{AH}
B	X_{BA}	X_{BB}	X_{BC}	X_{BD}	X_{BE}	X_{BF}	X_{BG}	X_{BH}
C	X_{CA}	X_{CB}	X_{CC}	X_{CD}	X_{CE}	X_{CF}	X_{CG}	X_{CH}
D	X_{DA}	X_{DB}	X_{DC}	X_{DD}	X_{DE}	X_{DF}	X_{DG}	X_{DH}
E	X_{EA}	X_{EB}	X_{EC}	X_{ED}	X_{EE}	X_{EF}	X_{EG}	X_{EH}
F	X_{FA}	X_{FB}	X_{FC}	X_{FD}	X_{FE}	X_{FF}	X_{FG}	X_{FH}
G	X_{GA}	X_{GB}	X_{GC}	X_{GD}	X_{GE}	X_{GF}	X_{GG}	X_{GH}
H	X_{HA}	X_{HB}	X_{HC}	X_{HD}	X_{HE}	X_{HF}	X_{HG}	X_{HH}

Figure 19. Consumption Activity Characterization Matrix

*Each of the elements is calculated as the fraction of number of people in a particular state k Є {A, B, C, D, E, F, G, H} at time t who belong to another state p Є {A, B, C, D, E, F, G, H} at time t+1. For example X_{BC}= n ($C_{t+1} \cap B_t$)/n (B_t). Other elements are defined similarly.

We call this matrix "Consumption Activity Characterization Matrix". Note that each of the rows in the matrix sum to 1. Table 5 shows how the elements of this matrix can be used to infer specific "information consumption behaviors" of people.

Table 5 Information Consumption Behaviors

Elements	Behavior Name	Description
X_{AA},X_{BB},X_{CC},X_{DD},X_{EE},X_{FF},X_{GG},X_{HH}	Status Quo	The consumption mode is the same for the two consecutive time periods. High values indicate that the status quo is preserved.
X_{AH},X_{BH},X_{CH},X_{DH},X_{EH},X_{FH},X_{GH}	Disengagement	From consuming in one or more of the three modes, the user completely stops consuming information from these users. The elements hence indicate the disengagement

		rates for each type of consumption.
X_{HA},X_{HB},X_{HC},X_{HD},$X_{HE}$$X_{HF}$,$X_{HG}$	Engagement	Indicates the preferences of a user for consuming information for the first time.
X_{BA},X_{CA},$X_{DA}$$X_{EA}$,$X_{FA}$,$X_{GA}$,$X_{HA}$	Following Influx	Indicates the influx to following form from all other forms. We can similarly define influx to all other forms using the corresponding elements in the column.
X_{AB},X_{AC},X_{AD},X_{AE},X_{AF},X_{AG},X_{AH}	Following Outflux	Indicates a transition from an initial state of "only Following" to any of the other states. Hence all these elements characterize the outflux from Following. We can similarly define outflux for other states using corresponding elements in a row.
X_{DA},X_{DB},X_{DC},X_{EA},X_{EB},X_{EC},X_{FA},X_{FB},X_{FC},X_{GA},X_{GB},X_{GC},X_{GD},X_{GE},X_{GF}	Redundancy Avoidance	The user has moved from either two or three states of information consumption to one or two states respectively

		meaning that he is avoiding redundancy in information consumption across different forms.
X_{AD},X_{AE},X_{AF},X_{AG},X_{BD},X_{BE},X_{BF},X_{BG},X_{CD},X_{CE},X_{CF},X_{CG},X_{DG},X_{EG},X_{FG}	Redundancy Seeking	The user has moved from one or two states of information consumption to two or three states respectively meaning that he is consuming information of the same users from more sources now than before.

When none of the elements in the matrix has a zero value, we can calculate the steady state probabilities which in this case are the long term chances of the curator consuming information in a certain mode. For any state j Є {A, B, C, D, E, F, G}, we can calculate the steady state probability π_j as follows

$$\pi_j = \sum_{k=A}^{G} \pi_k X_{kj}$$

However, it is not possible to calculate the steady state values if any of the elements in the matrix has a zero value. Figure 20 shows the number of curators for whom that particular element in the matrix has zero value meaning there is no transition. The ones marked in red are zero for more than 95% of the curators in the sample and the ones marked in yellow are zero for 90-95% of them. Therefore we cannot calculate the steady state probabilities.

	A	B	C	D	E	F	G	H
A - Only Followed	1	880	900	724	907	818	900	261
B - Only Listed	898	4	905	679	869	908	897	509
C - Only Subscribed	893	904	15	904	867	714	866	448
D -Followed and Listed	823	646	907	1	906	904	838	645
E - Listed and Subscribed	907	827	882	905	15	907	821	873
F - Subscribed and Followed	820	907	803	906	903	15	867	845
G - Listed, Subscribed and Followed	901	894	894	847	832	886	14	876
H- Neither listed, nor subscribed and nor followed	201	688	498	654	888	859	882	0

Figure 20. Transition Numbers Matrix

Figure 21 shows the transition volumes meaning the number of Twitter accounts that move from one form of information consumption to the other. Note that transition volumes here are the numerators in each of the X values of the "Consumption Activity Characterization Matrix". The numbers outside the parentheses represent the averages while the ones inside represent standard deviations. We can clearly see that there are a substantial number of people being moved from one state to other even though the corresponding transitions are zeros more than 90 percent of the time.

	A	B	C	D	E	F	G	H
A - Only Followed	4403.8(17770.24)	0.08(0.72)	0.01(0.09)	2.22(13.81)	0(0.03)	0.71(5.38)	0.02(0.3)	110.04(743.61)
B - Only Listed	0.03(0.39)	519.41(845.97)	0(0.08)	1.19(5.69)	0.24(3.42)	0(0)	0.01(0.12)	4.38(43.88)
C - Only Subscribed	0.31(5.38)	0.04(0.94)	1631.9(2703.35)	0.01(0.16)	0.3(4.21)	1.68(10.84)	0.18(2.51)	51.87(554.47)
D -Followed and Listed	1(7.61)	2.56(10.95)	0(0)	501.74(613.34)	0(0.05)	0(0.07)	0.42(3.46)	0.73(2.65)
E - Listed and Subscribed	0(0)	1.43(20.67)	0.26(3.24)	0.01(0.28)	93.85(306.69)	0(0)	0.35(2.36)	0.13(1.75)
F - Subscribed and Followed	2.49(23.35)	0(0)	1.07(7.3)	0.02(0.44)	0.01(0.07)	124.98(299.36)	0.2(2.08)	0.12(0.61)
G - Listed, Subscribed and Followed	0.02(0.37)	0.03(0.28)	0.02(0.16)	2.52(32.56)	0.52(4.06)	0.3(5.13)	100.99(192.27)	0.05(0.28)
H- Neither listed, nor subscribed and nor followed	137.59(851.6)	4.5(31.6)	17.28(94.14)	4.01(15.5)	0.13(2.13)	1.17(29.95)	0.5(11.47)	22524.14(56479.75)

Figure 21. Transition Volumes

In order to explore the transitions listed above, we conducted a set of F-Tests for the behaviors listed in table 5 the results of which are documented in table 6. For each of the properties listed in table 4, we used the appropriate X values where each of them was treated as a group. For example, for the status quo property, we treat each of X_{AA} - X_{HH} as a group and in each group we have 907 entries. We tested for the normality assumption in ANOVA and found that this was violated. However, one-way ANOVA is robust against normality. We assume that each of the transitions within in a particular property is independent of the other. For example in status quo property, X_{AA} is independent of X_{BB}. We used the Levene's test for checking the homogeneity of variances assumption and found that this was violated. Therefore we used the Welch's F-test to obtain the values. To get the transitions that have the highest and least values in a particular property, we used the Games-Howell post hoc test (because the homogeneity assumption was violated).

Table 6 F-Tests on Consumption Behaviors

Behavior	F-Value	Highest	Least
Status Quo	3.39**	X_{DD}	X_{FF}
Disengagement	18.22***	X_{CH}	X_{FH}
Engagement	29.019***	X_{HA}	X_{HE}
Following Influx	17.44***	X_{FA}	X_{EA}
Following Outflux	20.77***	X_{AH}	X_{AC}
Listing Influx	19.18***	X_{EB}	X_{FB}
Listing Outflux	27.60***	X_{BD}	X_{BF}
Subscribing Influx	16.00***	X_{FC}	X_{DC}
Subscribing Outflux	11.99***	X_{CH}	X_{CD}
Redundancy Avoidance	13.32***	X_{EB}	X_{GC}
Redundancy Seeking	4.72***	X_{EG}	X_{BF}

**p<0.002 ***p<0.00001

The results point to some useful findings about the nature of state transitions. First, we found that on an average 97% of the people who are classified into one of the consumption forms continue to stay in that form i.e., there are no transitions. However as we have seen in figure 21, the volume of transitions that happen are substantial. The results also indicate that that the "listed and followed" form of consumption has the highest retention while "subscribed and followed" has the least. This can be partially explained by the fact that we later found the disengagement rates to be highest for "only subscribed" form. Second, the disengagement rate is highest for "only subscribed" followed by "only following". This shows that list subscriptions are more volatile than following. This can be attributed to the fact that curators can disengage with a lot of people just by unsubscribing from a list while in case of following, they have to stop following each account individually. We also found that in general, dual and triple forms of information consumption have lower disengagement rates than the single ones. Third, following was the most common form of engagement behavior. This suggests that people prefer to follow explicitly first and then list/subscribe. On the other hand following also has the second highest disengagement rate. Fourth, we found that "listed and subscribed" to "only listed" is the most common form of redundancy avoidance behavior. This can be partially attributed to the fact that "only subscribed" is the most volatile form of information consumption. Finally, we found "listed and subscribed" to "listed, subscribed, and followed" is the most common form of redundancy seeking behavior. We attribute this to the fact that on Twitter, to get the tweets of people one has listed or subscribed to requires him/her to navigate through 3 or 4 links to reach the concerned list, while the tweets of a user being followed can be obtained on the user's home page right after logging in.

Conclusion and Future Research

In this research, we have explored the relationship between three different forms of information consumption on Twitter viz., following, listing and subscribing by developing and testing a cross sectional and temporal framework for analysis. We discovered that usage of these three forms of consumption is different from each other. In addition, using the temporal framework we discovered that the disengagement rates are highest for subscription and that when users decide to consume information for the first time from another user; following is the preferred mode of consumption. We also discovered that "following and listing" are the most stable form of information consumption. To our knowledge, our study is one of the first to explore the relationships among three different forms of information consumption on Twitter.

While our analysis points to very interesting findings – not known in literature previously - about Twitter users, they also have profound implications for further research in this area. First, recent models of information diffusion such as Achananuparp et al. (2012) assume that people receive tweets only by following others explicitly. Also Rui and Whinston (2012) measure the attention that people gain on Twitter using the number of followers. More recent research by Shi et al. (2011) uses the number of followers for a user to determine the probability of retweeting. However, in our cross sectional analysis we have provided substantial evidence for the fact that people also receive information by listing and subscribing and the users they list/subscribe to are not the same as the ones they follow. This also means that people can gain the attention of others by either getting listed or by gaining subscribers to the lists they belong to. The results of this research should be used to modify existing models of diffusion on Twitter to better reflect information consumption patterns. This should in turn lead to the development of more sophisticated metrics for measuring the size of a user's audience. Such studies can not only help gain a granular understanding of information diffusion and attention dynamics but also improve our general understanding of microblogging usage.

Second, while our study provides an initial exploratory analysis of the temporal relationship between forms of information consumption, it is yet not clear why people choose to switch from one form of information consumption to another. Future research can look into this phenomenon by directly leveraging the temporal framework in this research across different consecutive time periods. Also measures of user specific heterogeneity in terms of information consumption preferences can be embedded into such models to improve our analysis. Moreover networks and their properties- (Such as those that have curators and the users they are following/listed/subscribed as nodes with relationships between them) – may help provide an in depth understanding of the transitions. For example it is possible that users with a high degree centrality in the follower network have a high chance of being listed subsequently.

Third, we looked at the sampled users as information consumers only, in that, we collected the set of people they are following, the ones they have listed and the members of the lists to which they have subscribed. However, on Twitter, users play the dual roles of information producers and consumers simultaneously. In our future research we plan to extend our framework to devise a comprehensive model of information consumption and production on Twitter. Such a model can serve to identify the exact audience of a user, communities in Twitter, and reciprocal relationships.

Finally, drawing upon the evidence we showed for heterogeneity in information consumption across different curators, future research should delve deep into the reasons for preference for a certain form of information consumption. Such an investigation should look more into the user specific characteristics like tweets, profiles, images, and history of interaction with people the user is following, has listed and subscribed to. Also considering the recent evidence for demographic differences in the usage of Twitter, these can be used to explain the user heterogeneity. This can lead to more in depth insights into the usage of Twitter and identification of communities of users with preferences for specific forms of information consumption.

Acknowledgments

We thank the track chairs and the two anonymous reviewers for their valuable insights on improving the content and presentation of this paper. We are also grateful to the participants of the MSND workshop at the WWW Conference, 2013 for their feedback. Our sincere thanks to Paulo Goes and Ming Feng Lin, University of Arizona for their valuable suggestions. Finally, we also thank Michael A. Bruck, Assistant Director of Research Computing at the University of Arizona for his help on using High Performance Computing (HPC) systems without which dealing with datasets of this volume and veracity would not have been possible. All the errors and omissions in this paper are a sole responsibility of the authors.

References

Achananuparp, P., Lim, E. P., Jiang, J., and Hoang, T. A. 2012. "Who is Retweeting the Tweeters? Modeling, Originating, and Promoting Behaviors in the Twitter Network," *ACM Transactions on Management Information Systems (TMIS)* (3:3), pp. 13:1-13:30.

Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., and Gummadi, K. 2012. "Cognos: crowdsourcing search for topic experts in microblogs," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pp. 575–590.

Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. 2009. "Twitter power: Tweets as electronic word of mouth," *Journal of the American society for information science and technology* (60:11), pp. 2169–2188.

Java, A., Song, X., Finin, T., and Tseng, B. 2007. "Why we twitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65.

Krishnamurthy, B., Gill, P., and Arlitt, M. 2008. "A few chirps about twitter," in *Proceedings of the first workshop on Online social networks*, pp. 19–24.

Kwak, H., Lee, C., Park, H., and Moon, S. 2010. "What is Twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web*, pp. 591–600.

Pochampally, R., and Varma, V. 2011. "User context as a source of topic retrieval in Twitter," in *Workshop on Enriching Information Retrieval (with ACM SIGIR),* pp. 1-3.

Qu, Y., Wu, P. F., and Wang, X. 2009. "Online community response to major disaster: A study of Tianya forum in the 2008 Sichuan earthquake," in *42nd Hawaii International Conference on System Sciences, 2009. HICSS'09*, pp. 1–11.

Rui, H., Liu, Y., and Whinston, A. 2011. "Whose and What Chatter Matters? The Impact of Tweets on Movie Sales," *NET Institute Working Paper* (11-27).

Rui, H., and Whinston, A. 2012. "Information or attention? An empirical study of user contribution on Twitter," *Information Systems and E-Business Management* , pp. 1–16.

Sharma, N. K., Ghosh, S., Benevenuto, F., Ganguly, N., and Gummadi, K. 2012. "Inferring who-is-who in the Twitter social network," in *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pp. 55–60.

Shi, Z., Rui, H., and Whinston, A. 2011. “Content Sharing in a Social Broadcasting Environment: Evidences from Twitter,” Working Paper, University of Texas at Austin.

Velichety, S. and Ram, S. 2013 a. "The Science of Twitter Lists: Understanding Membership and Subscription through Network Analysis", Working Paper, University of Arizona.

Velichety, S., and Ram, S. 2013 b. “Examining Lists on Twitter to Uncover Relationships between Following, Membership and Subscription,” in *proceedings of the MSND workshop at WWW 2013,* pp.673-675.

Yamaguchi, Y., Amagasa, T., and Kitagawa, H. 2011. “Tag-based user topic discovery using twitter lists,” in *proceedings of 2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* pp. 13–20.

Zhao, J., and Ram, S. 2011. “Examining the evolution of networks based on lists in twitter,” in *proceedings of 2011 IEEE 5th International Conference on Multimedia Systems Architecture and Application (IMSAA)*, pp. 1–8.