Examining Lists on Twitter to Uncover Relationships between Following, Membership and Subscription

Srikar Velichety
srikarv@email.arizona.edu

Sudha Ram
ram@eller.arizona.edu

University of Arizona, Tucson
Department of MIS,
Eller College of Management

ABSTRACT
We report on an exploratory analysis of pairwise relationships between three different forms of information consumption on Twitter viz., following, listing and subscribing. We develop a systematic framework to examine the relationships between these three forms. Using our framework, we conducted an empirical analysis of a dataset from Twitter. Our results show that people not only consume information by explicitly following others, 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. Our work has implications for understanding information propagation and diffusion via Twitter and for generating recommendations for adding users to lists, subscribing and merging or splitting them.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: Data Mining; G.3 [**Probability and Statistics**]: Correlation and Regression Analysis; H.4.m **[Information Systems]**: Miscellaneous

Keywords
Twitter, Lists, Subscription, Membership, Descriptive Modeling.

1. INTRODUCTION
Twitter is one of the most popular microblogging platforms in the world used by individuals and organizations. Using text restricted to 140-characters, individuals share information about their about various interests [7], or express preferences for brands [6]. Organizations such as news agencies, government and private companies use Twitter to disseminate information[2]. Twitter has been used successfully to track flu trends [3], and to gauge people's sentiments and moods [4].

In 2009, Twitter introduced a feature called Lists that allows users to group other users and follow their tweets. A user on Twitter can consume information produced by others in three ways viz., by directly following, by creating a list and adding them or by subscribing to a public list created by someone else. A user may add someone to his list without explicitly following him. Similarly, he can subscribe to a list without explicitly following or including any of the members into his own list. Any combination of following, listing and subscribing is also possible.

While there is preliminary research on the use of lists to infer latent user characteristics [5, 8, 9, 11], surprisingly there is little research so far on how these three forms of information consumption are related to each other. An examination of lists will not only help us understand information consumption in a better way, but can also provide insights into user behavior and microblogging usage in general. In earlier work we used a network analysis approach to examine Twitter lists with a view toward developing a list recommendation system [13].

In this research, we follow up by developing a framework to understand the relationships among membership, subscription and following on Twitter. The rest of the paper is organized as follows. In section 2 we provide a short review of current research on Twitter lists to create an appropriate context for our work. In section 3 we describe a framework used for analyzing the relationships between different forms of information consumption at a specific point in time. In section 4 we provide a description of the dataset used for an empirical analysis and discuss some preliminary results. In section 5 we conclude with a description of our ongoing work.

2. SUMMARY OF PREVIOUS WORK
When a user creates a list and adds users to it, he/she is able follow the tweets of the list members by going to the list. While creating a list, the user may provide a short -100 character – description of the list. This is many cases indicates the interests of the list creator (hereafter known as curator) and those of the members of that list. [9] developed a framework, using this information, for identifying latent user characteristics. Their results showed that user interests and characteristics along with their popular perceptions on Twitter can be identified using the information on the lists to which they belong. [5] developed a system for identifying topic experts on Twitter using information on lists to which they belong. [8] leveraged the interactions of a user along with his/her list information to identify interests. [11] showed that methods that use list-based information outperform those that use tweet-based information, for identifying user interests. [12] examined patterns of closures in lists. While all this research looks at lists as a potential source for identifying user interests and characteristics, and examining homophily, very little is known about the relationships between following, listing and subscribing, which the focus of our work is.

3. A FRAMEWORK FOR EXAMINING LISTS
In this section we develop a systematic framework to understand pairwise relationships between Following, Listing and Subscribing. n(X) represents the cardinality of the set X, and C, L, F & S represent Curator, Listed, Following and Subscribed, respectively in the names of the metrics we define. For example CF Ratio means Curator Following Ratio.

First, to examine the relationship between following and listing, we can divide the universe of Twitter users into 4 disjoint sets from the perspective of a specific Twitter user (Figure 1).

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

		Followed	
		No	Yes
Listed	No	W	X
	Yes	Y	Z

Figure 1. Relationship between Listing and Following

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

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

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

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

Thus, CF Ratio is the fraction of people a curator is following out of all the people that she has listed across one or more of her public lists. On the other hand, CM Ratio is the fraction of people that a curator has listed at least once (across all his public lists) out of all the people that she is explicitly following. Finally, the MF Ratio is a measure that compares the relative strength in numbers of those who are only listed to those who are only followed.

To examine the relationships between subscribing and following, we similarly identify 4 disjoint sets in the universe of Twitter users (Figure 2).

		Followed	
		No	Yes
Subscribed	No	A	B
	Yes	C	D

Figure 2. Relationship between Subscribing and Following

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

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

$$CS_{Ratio} = \frac{n(D)}{n(C)+n(D)}$$

$$CFS_{Ratio} = \frac{n(D)}{n(B)+n(D)}$$

$$SF_{Ratio} = \frac{n(C)}{n(B)}$$

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

4. DATASET AND ANALYSES

A dataset for analysis was collected during December 2012. We identified a random sample of 100 Twitter curators –and collected the set of users they are following, the members of the lists they have created and members of the lists to which they have subscribed. The initial data about lists and their members was collected from listorious.com. We used the Tweepy module in python to collect the membership and following data. All together our dataset has 1183 lists that were curated and 984 unique lists to which there were subscriptions.

The CF Ratio had an average value of 0.43 with a median of 0.42 (σ = 0.28) meaning users on an average follow only 43% of the people they list. The CM Ratio had an average value of 0.28 (σ = 0.23) with a median of 0.21 meaning users list a mere 28% of the people they follow. However the fact that the standard deviations were substantial indicates that there is a large user specific heterogeneity. We also found that the MF Ratio had an average of 58.46 (σ = 501.26) with a median of 0.27 meaning for some users listing is the primary form of information consumption while for others it is following explicitly. Together these results point out that members in a curators list are not the ones they follow and vice versa.

The average CS ratio was 0.11 (σ=0.13) with a median of 0.06 indicating that curators follow a meager 11% of the members in the lists they subscribe to. A paired t-test on the CF and CS Ratios for these users indicated that (t-value =11.73, $p<0.001$) the CF Ratio was significantly higher meaning curators follow a greater fraction of the people they list than members of the lists to which they subscribe. Further, the average CFS ratio was 0.09 (σ=0.10) with a median of 0.05 indicating that a mere 9% of the people that a curator follows are also members in the lists to which they subscribed. We also found significant differences between the CM and CFS Ratios (t-value=7.9, $p<0.001$) meaning that out of all the people that a user follows, a greater fraction are members of the lists they have created than the lists to which they have subscribed. These two ratios provide evidence for the fact that the people users follow are substantially different from the members of the lists they subscribe to. Finally the fact that the SF Ratio had a mean of 5.03 (σ = 25.2) with a median of 0.54 shows that some users prefer to consume information primarily through following while others through subscribing.

5. CONCLUSION IMPLICATIONS AND ONGOING WORK

We have investigated relationships between three forms of information consumption on Twitter viz., following, listing and subscribing by developing a systematic framework and defining specific measures for pairwise analysis. Our results show that these forms of information consumption are significantly different from each other. We also show that Twitter users follow a greater fraction of the people they list than the people in the lists to which they subscribe. Similarly, users list a greater fraction of the people they follow as compared to the fraction of the users in the lists to which they subscribe. Finally we show that there is considerable user specific heterogeneity in terms of preference for each form of information consumption. Our framework has implications for developing improved models of information propagation and diffusion on Twitter. Most current models consider following as the only form of information consumption[1, 10].It also has implications for list recommendations i.e. for adding members to lists, subscriptions to lists, follower recommendations, and for merging/splitting of lists.

While our research points to interesting results about forms of information consumption, we are continuing extend it in several directions. First, this paper reports on an analysis of a sample of 100 Twitter list curators and it is primarily a static analysis. In our ongoing work we are collecting and analyzing a larger sample to provide stronger evidence for the results of this research. Second, our work examines the relationship between 3 forms of information consumption at a specific point in time. There may be a temporal relationship between these forms. For example, it is possible that a user follows someone first and then decides to list him or vice versa. In our ongoing work we are building a framework for temporal analysis of such relationships while leveraging our static framework. Finally, based on the fact that some curators primarily prefer one form of information consumption over others, it would be interesting to see how these patterns are manifested in the tweeting activity. Our ongoing work is to examine the differences between tweeting, retweeting and mentioning behaviors of these users.

6. REFERENCES

[1] 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 (2012).

[2] Bhattacharya, D. and Ram, S. 2012. Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter. *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (2012), 966–971.

[3] Culotta, A. 2010. Towards detecting influenza epidemics by analyzing Twitter messages. *Proceedings of the First Workshop on Social Media Analytics* (2010), 115–122.

[4] Diakopoulos, N.A. and Shamma, D.A. 2010. Characterizing debate performance via aggregated twitter sentiment. *Proceedings of the 28th international conference on Human factors in computing systems* (2010), 1195–1198.

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

[6] 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 (2009), 2169–2188.

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

[8] Pochampally, R. and Varma, V. 2011. User context as a source of topic retrieval in Twitter. *Workshop on Enriching Information Retrieval (with ACM SIGIR)* (2011).

[9] Sharma, N.K., Ghosh, S., Benevenuto, F., Ganguly, N. and Gummadi, K. 2012. Inferring who-is-who in the Twitter social network. *ACM SIGCOMM Computer Communication Review*. 42, 4 (2012), 533–538.

[10] 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.

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

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

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