

Why You Are More Engaged: Factors Influencing Twitter Engagement in Occupy Wall Street

Jilin Chen and Peter Pirolli

Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA
{jichen, pirolli}@parc.com

Abstract

Twitter has been used for engaging with audiences online in several popular political movements. In this paper we explore factors that influence the engagement of Twitter users during the recent Occupy Wall Street movement, where engagement is measured by retweets and hashtag usage related to the movement. Through analyzing Twitter activities of more than 18,000 users, we found that users' general activity level, geographic location, topic interests and interpersonal interactions before the movement all had measurable effects on users' engagement level during the movement.

1. Introduction

Twitter has been increasingly used as an online medium for popular political movements to engage with audiences, strengthen their influences, and obtain support from the general public (Lotan et al., 2011).

Occupy Wall Street – a recent popular movement in U.S. protesting economic inequality in the country – is of no exception: Organizers of the movement created the official Twitter account for the movement on Jul 14th, 2011, two months before the movement itself, and maintained a dedicated media team so as to broadcast live updates to Twitter and other social media platforms (Santo, 2011).

Their effort on Twitter seems well rewarded: After the movement started on Sep 17th, 2011, the official Twitter account, OccupyWallSt, was able to attract more than 30,000 Twitter followers in merely 15 days. By the end of December, the OccupyWallSt account had accrued more than 140,000 followers.

However, not all these followers were equally engaged with the movement on Twitter. At one end of the spectrum, some followers actively retweeted individual posts from the OccupyWallSt account, and exchanged hundreds of tweets using hashtags related to the movement. At the other end of the spectrum, some other followers behaved as silent observers and did little more than become a follower on Twitter. The rest of the followers were in between the

two extremes, posting a few related tweets through the course of the movement.

A series of questions therefore emerge regarding this differing level of engagement: Why are some followers more engaged than others? Is it simply because they are more active Twitter users? Or is it because they are more concerned with political issues? Or is it because they live in New York and therefore care more about local events?

In response to these questions, this paper presents an exploration of factors that influenced the engagement of OccupyWallSt's followers on Twitter, where engagement is measured by related actions on Twitter during the movement, including retweets and hashtag usage.

We limit our exploration on these OccupyWallSt followers, because they have already shown their support of the movement by following the official OccupyWallSt account, therefore allowing us to focus our effort on investigating how engaged they were. We in particular explore the following four hypotheses:

- H1: Followers who are more active in general are more engaged with the movement;
- H2: Followers who exhibited related topic interests before the movement will be more engaged during the movement;
- H3: Followers who had more interactions with other would-be followers before the movement will be more engaged during the movement;
- H4: The geographic locations of the followers affect the engagement with the movement on Twitter.

In exploring these hypotheses, we hope to not only understand how people have engaged in the Occupy Wall Street movement on Twitter, but also extend our general knowledge of online user engagement that may extend to future political movements featured on social media.

2. Background

2.1 Occupy Wall Street

Occupy Wall Street is a protest movement started in Sep 17, 2011 in Zuccotti Park (formerly, Liberty Square) in New York City, against a variety of issues around

	Mean	Min	Lower Quartile	Median	Higher Quartile	Max
<i>RetweetOWS</i>	1.1	0	0	0	1	205
<i>HashtagOWS</i>	21.7	0	0	2	11	1,781
<i>TweetN</i>	3,667.2	100	331	987	3,175	291,227
<i>FollowerN</i>	915.5	0	58	147	417	865,698
<i>FolloweeN</i>	728.6	1	136	290	640	136,276
<i>RetweetN</i>	148.4	0	7	45	166	2903
<i>MentionN</i>	563.9	0	51	231	837	8982
<i>TopicInt</i>	0.031	0.000	0.008	0.025	0.044	0.311
<i>RetweetFL</i>	9.0	0	0	0	4	1240
<i>MentionFL</i>	30.0	0	0	4	25	1347
<i>WithinNY</i>	10.5% Positive (within New York), 72.6% Negative (outside New York), 16.9% Missing					
<i>OutsideUS</i>	24.8% Positive (outside U.S.), 58.3% Negative (within U.S.), 16.9% Missing					

Table 1: Descriptive Statistics of Data Variables

economic and social inequality, undue corporate influence on government (especially from the financial sector, i.e., “Wall Street”), and joblessness. The movement claimed to have spread across over 100 cities in U.S. and 1,500 cities world wide, and has attracted heated discussions from news media and political figures within U.S. and world wide.

On Twitter, the official account of the movement has posted more than 5,000 tweets since its creation in July. Uses of movement-related hashtags surged on Twitter since September, and by early October had reached to more than 10,000 tweets a day.

2.2 Prior Research

To date we are aware of little research directly addressing Occupy Wall Street on Twitter; however, research has been devoted into characterizing the use of Twitter during other social and political movements in Iran (Gaffney, 2010), Tunisia and Egypt (Lotan et al., 2011). Our effort differs from these studies in that we are not characterizing the use of Twitter in these events per se; instead, we are exploring factors that may have influenced the use.

We attempt to identify factors predictive for particular actions on Twitter, thus sharing a similar angle with earlier Twitter analyses, such as Golder et al. (2010) on tie formation and Kivran-Swaine et al (2010) on tie break-up.

We formulated our hypotheses around interest, prior interaction and geographic location, because several prior studies have demonstrated the effects of these features on information diffusion (Aral et al., 2009) and platform adoption (Toole et al., 2011) in online social networks.

3. Dataset and Variables

The dataset for this exploration is based on the 36,145 followers of OccupyWallSt on Twitter as of Oct 3rd, 2011. To construct the dataset, we used Twitter API to collect

tweets posted by each follower by Oct 3rd, 2011 and excluded followers who had protected accounts or had posted fewer than 100 tweets by that time.

The resulting dataset contained 18,611 followers, and descriptive statistics of the dataset is shown in Table 1. Note that due to the constraint of Twitter API, we could only collect up to 3,200 recent tweets per user. This limitation affected the 24% of followers who have posted more than 3,200 tweets by October 3rd, as for these followers we had to compute some variables (e.g. the number of retweets) from a partial collection of tweets.

Below we describe all the variables in detail. The dataset included two dependent variables for measuring the engagement of each follower with the movement:

RetweetOWS: The number of the follower’s retweets of OccupyWallSt by Oct 3rd. This included not only direct retweets of the OccupyWallSt account, but also indirect retweets through other users (i.e. the follower retweets another user’s direct retweet) and retweet chains.

HashtagOWS: The number of the follower’s tweets that contain movement-related hashtags produced by Oct 3rd. We included four hashtags as related: #sep17, #ows, #occupywallst, #occupywallstreet.

Comparing the above two measures of engagement, we consider *RetweetOWS* as being narrower and more direct, as by posting retweets a follower is directly reinforcing the voice of the official account of the movement on Twitter – in fact, as shown in Table 1, only about 25% followers ever posted any such retweets in our dataset. In contrast, *HashtagOWS* is a broader measure, as it indicates general engagement in the discussion around the movement.

To test H1, we included five independent variables to represent the general activity level of followers:

TweetN: The number of tweets that the follower had posted as of October 3rd.

	Predicting Retweets of OccupyWallSt (<i>RetweetOWS</i>)				Predicting Related Hashtag Uses (<i>HashtagOWS</i>)			
	Model 1		Model 2		Model 1		Model 2	
	Coefficient	Sig. Level	Coefficient	Sig. Level	Coefficient	Sig. Level	Coefficient	Sig. Level
<i>TweetN</i>	0.273	***	0.258	***	3.583	***	3.225	***
<i>FollowerN</i>	-0.120	***	-0.124	***	-0.533	**	-0.766	***
<i>FolloweeN</i>	-0.112	***	-0.109	***	-0.902	***	-0.760	***
<i>RetweetN</i>	0.143	***	0.050	***	1.998	***	0.311	***
<i>MentionN</i>	-0.240	***	-0.213	***	-3.189	***	-3.001	***
<i>TopicInt</i>			0.012				0.988	***
<i>RetweetFL</i>			0.151	***			2.508	***
<i>MentionFL</i>			-0.001				0.279	**
<i>WithinNY</i>			0.002				0.123	***
<i>OutsideUS</i>			-0.010	***			-0.236	***
Adjusted R ²	.147		.176		.191		.266	

Table 2: Results of Linear Regressions

Coefficients were obtained after scaling all independent variable into [0, 1]. For significance level: *** p<.001, ** p<.01, * p<.05

FollowerN: The number of Twitter users who were following the follower as of Oct 3rd.

FolloweeN: The number of Twitter users whom the follower was following as of Oct 3rd.

RetweetN: The number of retweets that the follower had posted before Jul 14th.

MentionN: The number of @ mentions that the follower had posted before Jul 14th. Multiple @ mentions within one tweet were counted as multiples.

Note that Jul 14th is creation date of the official OccupyWallst account. We chose this date as a cut-off for *RetweetN* and *MentionN* to avoid potential confound with the dependent variables *RetweetOWS* and *HashtagOWS*.

For testing H2, we included one independent variable to represent prior topic interest:

TopicInt: The dot product between two normalized TF-IDF vectors, one computed from the tweets that the follower had posted before Jul 14th, and the other computed from a random sample of roughly 500,000 tweets that were posted in September and October 2011 and contained the four hashtags we deemed as related to the movement. Computing the dot product between TF-IDF vectors is a standard technique for measuring topic relevance between text corpora in information retrieval (Frakes et al., 1992). Intuitively, this *TopicInt* variable reflects how much interest the follower had exhibited before the movement on topics that were later related to the movement.

For testing H3, we included two independent variables to represent prior interactions among would-be followers:

RetweetFL: The number of times that the follower had posted retweets of other would-be followers before Jul 14th.

MentionFL: The number of times that the follower had @ mentioned other would-be followers before Jul

14th. Multiple @ mentions within one tweet were counted as multiples.

For H4 on the effect of geographic location, we used Yahoo! PlaceFinder API to code the text content in the location field of the follower's profile. As location fields on Twitter are usually no more specific than city level and PlaceFinder API does not always provide reliable fine-grained results (Hecht et al., 2011), we only include the following two coarse-grained measures as independent variables for the locations of followers:

WithinNY: A binary variable indicating whether the location of the follower was within the U.S. state of New York.

OutsideUS: A binary variable indicating whether the location of the follower was outside of United States.

Note that the PlaceFinder API was unable to code the location fields of 16.9% of followers in our dataset. The two variables above were therefore marked as missing data for these followers in later analyses.

4. Data Analysis and Results

We tested our four hypotheses by analyzing how the ten independent variables can predict the two measures of engagement in linear regressions. Because all the numerical variables exhibited highly skewed distributions (as shown in the top 10 rows in Table 1), we normalized these skewed distributions using a log-transformation following Golder et al., (2010). We then scaled all independent variables into the range of [0, 1] (i.e. mapping the minimal value of each variable into 0 and mapping the maximal value into 1), so that the effect size in the regression is comparable across independent variables.

The results of the regressions are shown in Table 2. For predicting each engagement measure we present two models: Model 1 contains only the five independent

variables measuring the general activity level, while Model 2 contains all the ten independent variables. Comparing the two models allows us to understand how the five independent variables specific to the Occupy Wall Street movement affects engagement.

We found mixed support for H1 on the effect of general activity level: On one hand, more tweets and retweets indicate higher engagement by both measures, thus supporting H1; on the other hand, however, larger personal social network and more @ mentions indicate lower engagement. This result suggests that the intuition “active people will do everything more” is only partly true: For instance, people writing more @ mentions might be using Twitter primarily for chatting with close friends, and thus not interested in spreading words about a political movement.

We found partial support for H2: Higher prior topic interest indicates no more retweets of the official account of the movement but significantly more usage of related hashtags. This result suggests that while people with prior topic interest might not help spread the official voice of the movement directly, they were nonetheless more engaged in the discussions around the movement.

We found stronger support for H3: More prior retweets among would-be followers indicate higher engagement by both measures, while prior @ mentions indicate higher engagement by one measure. This positive result can be explained in at least two ways: 1) prior interactions among would-be followers reflect flows of influence, and the higher engagement of these people during the movement was due to higher number of incoming influence flows; 2) prior interactions indicate homophily, and both the interactions before the movement and the higher engagement during the movement are due to particular interests and beliefs shared among these people (e.g. interests in politics, beliefs of economical equality).

We also found support for H4: People in New York were more engaged by one measure, while people outside of U.S. were less engaged by both measures. This result indicates that local populations were more engaged in the movement on Twitter, confirming the effect of location.

5. Conclusion

In this paper we have found that activity level, prior topic interest, prior interpersonal interactions and geographic location all had measurable effects on Twitter engagement of OccupyWallst followers during the movement. We suspect that similar effects may be present in other political movements on Twitter as well, as the factors in this exploration are general and applicable beyond the Occupy Wall Street movement.

One limitation of this work is that we have not directly investigated the mechanism behind these effects, such as

investigating whether the effect of prior interaction is more due to influence or homophily (Aral. et al, 2009). Such investigations will be of great value as future research, as they will provide further insights for designing effective online interventions during political movements.

Meanwhile, we should also note that the prediction strength of the factors we identified in this exploration is fairly limited: the adjusted multiple R^2 of the regression models indicate that they accounted for 14.7% to 26.6% of the variance, suggesting that the models only explain a small portion of the variance. As a result, the models are unlikely to be sufficient for predicting whether a particular follower would be more engaged or not during the movement given their prior activities.

While improvement in prediction power is possible through more sophisticated topic modeling (Ramage et al., 2010) and incorporating more information from the local social network (Golder et al., 2010; Kivran-Swaine et al., 2011), a fundamental challenge will likely remain: many critical factors in determining engagement in political movements (e.g. social status, financial situation, intimate social relationships, political belief) are not readily observable in regular usage of Twitter, and many people may never reveal any information related to some of these factors online. Investigating modeling techniques to meet this challenge can be another promising future direction.

References

- Aral, S., Muchnik, L., and Sundararajan, A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. In Proc. of National Academy of Sciences'09.
- Frakes, W. B. and Baeza-Yates, R. 1992. Information retrieval: Data structures and algorithms. Prentice Hall.
- Gaffney, D. 2010. #IranElection: Quantifying online activism. In Proc of WebSci'10.
- Golder, S. A., and Yardi, S. 2010. Structural predictors of tie formation in Twitter: Transitivity and mutuality. In Proc. of IEEE Social Computing 2010.
- Hecht, B., Hong, L., Suh, B., Chi, E. H. 2011. Tweets from Justin Bieber's heart: The dynamics of the “location” field in user profiles. In Proc of CHI'11.
- Kivran-Swaine., F., Govindan, P., and Naaman., M. 2011. The impact of network structure on breaking ties in online social networks: Unfollowing on Twitter. In Proc. of CHI'11.
- Lotan, G., Graeff, E., Ananny, M., et al. 2011. The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. Int. Journal of Communication, Vol. 5.
- Ramage, D., Dumais, S., and Liebling, D. 2010. Characterizing microblogs with topic models. In Proc of ICWSM'10.
- Santo, A. 2011. Occupy Wall Street's media team. Columbia Journalism Review.
- Toole, J.L., Cha, M., and Gonzalez, M.C. 2011. Modeling the adoption of innovations in the presence of geographic and media influences. arXiv:1110.0535.