Tweeting Political Dissent: Retweets as Pamphlets in #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution and #OccupyWallSt

Marco Toledo Bastos
University of São Paulo
m.toledo-bastos@lse.ac.uk

Rodrigo Travitzki
University of São Paulo
travitzki@usp.br

Rafael Raimundo
University of São Paulo
rafael.raimundo@usp.br

Abstract

This paper examines Twitter hashtags related to political upheavals in the Middle East, North Africa, Europe, United States and the Americas. We monitored the hashtags #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution and #OccupyWallSt and compiled a dataset of over two million messages, including 791,968 retweets and billions of users interconnected as follower and following. We calculated the statistically significant correlations between retweets, mentions, followers, following and volume of messages. We also looked into the number of hashtag-related messages tweeted by single users regardless of their position in the network topology, providing a hypothesis to explain the emergence of Twitter political hashtags that is consistent with the practice of pamphleteering. The results show statistically significant correlations between retweets rate and number of tweets per user, but also low statistically significant correlations between retweet rate and the first-level network topology (follower and following network). These results suggest that instead of depending on user-hubs that act as opinion leaders, Twitter political upheavals are correlated with the intense activity of users who pamphleteer a cause or idea. We further examined this finding by comparing the hashtags network structure to the presumed communication pattern of pamphleteering, and we found a higher than average cluster of retweets sent and received by interconnected users. We also looked into the growing occurrence of Twitter accounts dedicated to activism and the high frequency of messages designed to increase the chances of a hashtag to spread through Twitter network. Lastly, we analyzed the text corpora of the hashtags looking for political charged concepts across the dataset. The results suggest that pamphleteering is a valid metaphor to the political activity of Twitter users, but also encourage further research using longitudinal networks to represent changes over time in the network.

Keywords

Twitter; Protests; Hashtags; Arab Spring; *Indignados*; Pamphleteering

1. Information Channel or Social Network

Since 2009 a burgeoning field of microblog research has been examining the information diffusion in Twitter and establishing different metrics to determine the influence of users (Bakshy et al. 2011; Kwak et al. 2010), the comparative impact of message content (Chew and Eysenbach 2010; Goel et al. 2012) and Twitter's topological features (Kwak et al. 2010; Wu et al. 2011). On the one hand, Twitter's highly skewed distribution of followers and low rate of reciprocated ties seems to depict an information channel instead of a social network. On the other hand, there is no agreement on whether influence is determined by network topology or by message content (Bakshy et al. 2011; Grabowicz et al. 2012; Huberman et al. 2009).

These controversies can be found in Bakshy et al. (2011), who investigated the distribution of retweet cascades on Twitter and determined that although users with large followers counts and past success in triggering cascades were on average more likely to trigger large cascades in the future, these features were in general poor predictors of future cascade size. Kwak et al. (2010) and Wu et al. (2011, 3) found that Twitter does not conform to the usual characteristics of social networks, which exhibit much higher reciprocity and far less-skewed degree distributions, but instead better resembles a mixture of mass communication and face-to-face communication.

Kwak et al. (2010) also encountered a short average path length that might be a symptom of Twitter's role as an information mechanism, as users follow users not for social networking, but for information. The investigation of Wu et al. (2011) highlighted that Twitter more closely resembled an information sharing network than a social network, and Quercia et al. (2012) again contrasted Twitter network metrics to information channels and social networks. Twitter's ambivalent network features are important to understand how political movements flourish in the Twittersphere; that is, whether the projection of grassroots movements into political society depends on the spanning of structural holes by social brokers, much like individuals in real-life communities, or depend on the performance of opinion-leader hubs, much like information channels.

This is of special interest in view of the similarities between online activism and printed pamphlets, which emerged as the most effective means of persuasion and communication during the 16^{th} century. The pamphlet allowed the emergence of influential moral and political communities of readers and shaped the public sphere of popular, political opinion. The pamphlet changed the cycles of communication between author and reader, incorporating a diversity of genres and imaginative devices employed by pamphleteers (Raymond 2003). The single combination of mass and direct communication, which stimulated the emergence of pamphlets, was again essential to the emergence of Twitter political hashtags fueled by a medium that is both an information channel and a social network.

2. Talking Politics on Twitter

Romero et al. (2011) delivered a major investigation into the dynamics of information diffusion in Twitter and found structural differences across topics. The differences were not a result of stickiness, that is, the probability of adoption based on one or more exposures. Hashtags with high persistence tended to continue having relatively significant effects even after repeated exposures. Perhaps not surprisingly, Romero et al. (2011) found that hashtags for politically controversial topics were particularly persistent, with repeated exposures continuing to have unusually large marginal effects on adoption.

Still according to Romero et al. (2011), the distinctive network structure of Twitter political hashtags—the unusually large effect relative to the peak after successive exposures—not only corresponded with the sociological principle of complex contagion, but also depicted the first large-scale validation of the principle. In a different approach, Tumasjan et al. (2011) found that Twitter mirrors the political landscape off-line and can be used to predict election results, suggesting that the mere number of party mentions can accurately reflect the election results. Similarly, Ratkiewicz (2010) investigated the spread of astroturf political campaigns on Twitter, presenting a machine learning framework capable of identifying centrally-controlled accounts that supported a candidate or opinion.

Political hashtags tend to exhibit a higher internal degree, a greater density of triangles and a larger number of nodes (Romero et al., 2011). The subgraphs on which political hashtags initially spread have high degrees and extensive clustering. These characteristics were noticeably pronounced in the political hashtag #FreeIran, which emerged in connection to the protests following the 2009 Iranian presidential election. This pattern was again found in the rising number of posts in the hashtag #FreeVenezuela, which related to the socio-economic situation in Venezuela during February 2010. Both hashtags evolved in contexts of economic crisis and political uncertainty.

But it was in 2011 when political tweets gained momentum, as Twitter became instrumental during the Arab Spring uprisings. We followed these political events on Twitter and created a database of political hashtags that spans from 2009 to 2011. Establishing the retweet messages (RT-messages) as the fundamental metric for message diffusion on Twitter, we evaluated whether the replication of political messages in Twitter was mostly reliant on the network topology, hence in the strength of the ties, or in the message itself, thus not depending on the connections of follower and following to spread.

The debate on online activism comprehends on the one side scholars arguing that the use of communication technologies is instrumental in increasing the size, speed, and reach, but not the process of harnessing political power (Van

De Donk 2004). On the other side of the spectrum we find scholars arguing that communication technologies change the actual processes of organizing and participating in activism to what has been referred to as technoactivism (Kahn and Kellner 2004). The empirical results of this investigation are consistent with the observation of Earl and Kimport (2011), who understand the two perspectives as not mutually exclusive, but likely attached to the differing types of protest scholars study.

3. Word of Mouth and Mass Communication

Results of previous investigations (Bastos et al. 2012a) suggested that the retweet rate in political hashtags is strongly connected to the number of tweets per user in the examined datasets. Perhaps even more surprising, we have not found a strong correlation between the retweet rate and the user's number of followers and following (first-level network topology). Based on these early findings, we hypothesized that political action in social networks emerges from the high frequency of messages posted by ordinary users with lower-than-average connectivity.

The action of these users resembles the historical engagement to pamphleteering that employed an organized group of citizens with the objective of broadcasting opinions on an issue or ideology to a larger audience. While Twitter distinctively combines mass-distributed with peer-to-peer communication, the pamphleteer relied on analogous platforms that integrated mass communication output and a peer-to-peer interaction (or rather face-to-face) in approaching the public. Though political freedom varies tremendously across different contexts, pamphleteering in social networks also presents a considerable level of risk, mostly because political hashtags involve the publicly aligning of a user with a position that might alienate her from other peers within the social circle.

During the 16th century, pamphlets came to refer to a short, vernacular work engaged with social, political or ecclesiastical issues. Generally printed in quarto format, pamphlets soon became the most common medium for conveying news and the predominant means of influencing public opinion. According to Raymond (2003, 26), the pamphlet became a pre-eminent model of public speech, a way of conceiving of the power of the world, and even if English was not the language of literary culture, pamphleteers wrote in English in order to reach a broad audience. Still according to Raymond (2003, 44), in the period of 1580-1640 pamphleteers abandoned formal rhetorical structures and developed a range of plain modes adapted to polemic and reportage.

The production of pamphlets fed a circuit in which readers actively participated in writing, verbally or by action, and forced writers to respond to the reactions, forming a network of communication that included an array of vendors, editors, compositors, press-men, book-sellers, post men, authors, and readers (Raymond 2003, 55-56). The production of printed news in France was designed to respond to a demand for political information that depended on the emergence of larger circuits of diffusion with major towns constituting the chief nuclei, from which information spread out unequally through commercial channels that penetrated suburbs and surrounding villages (Vittu 1999, 174).

The diffusion of pamphlets relied, on the one hand, on light volumes distributed by itinerant vendors, and on the other hand, on the word of mouth among readers. Despite the significant amount of paper in circulation, most news would still spread by word of mouth and form a cycle of information and misinformation that only news could end (Levy 1982, 24). Even though some pamphlets were less newsbooks than vehicles for political expression, with pamphleteer being singularly effective during political crisis, news was from any perspective and by any definition the hub of the pamphlet culture (Mendle 1999, 58-61), thus perpetuating the circle of pamphlets, word of mouth and news.

Despite the remarkable differences in prose and style, pamphlets and tweets comprehend the prolific expressions on an individual level made possible due to low costs of production and distribution. While tweets might include hashtags that gather wider attention, pamphlets could reach villages or towns and be posted for greater consumption. Like tweets, pamphlets offered no financial rewards to individual authors, who were driven by the piquant power of the printed text. Moreover, the interactive nature of pamphlets, often written in response to other pamphlets (Raymond 2003, 209), resembles the reciprocal nature of AT and RT-messages that form a network of voices in Twitter.

Pamphleteering and tweeting are particularly analogous in the context of lowering the obstacles to participation, housing a wide variety of styles and genres and allowing for anonymity and the use pseudonyms (Moe 2010). Similarly to pamphleteers, Twitter political upheavals are correlated with the intense activity of users with relatively few connections and not necessarily dependent on user-hubs that act as political decision makers or opinion leaders. We will refer to these users as clickers, in opposition to hubs or elite users, and our results suggest that the pamphleteering performed by individual users is capable of generating highly replicated messages and pushing trending topics without the direct influence of elite users or media pundits.

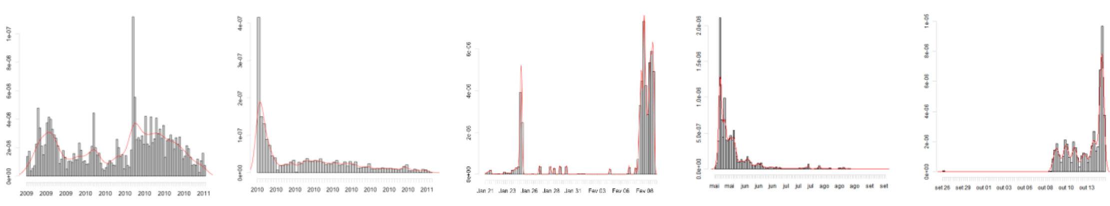
In order to test this hypothesis, we investigated six Twitter hashtags related to political upheavals in Middle East, North Africa, Europe, United States and the Americas that emerged as a string of protests targeting public policy and social inequalities. Twitter hashtags #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution, and #OccupyWallSt evolved in contexts of economic crisis and/or political uncertainty in Iran, Venezuela, Egypt, Spain and the United States during the years of 2009, 2010 and 2011.

4. Dataset

For this investigation we compiled a dataset of approximately two million Twitter messages, with 791,968 RT-messages and 83,024 AT-messages. Due to computational limitations related to data analysis in present systems, we sampled the hashtags #SpanishRevolution from 747,303 to 31,854 messages, and the #OccupyWallSt hashtag from 553,510 to 67,620 Twitter messages. The final dataset includes all messages from #FreeIran, #FreeVenezuela and #Jan25, but only a portion of the hashtags included in the original datasets of #SpanishRevolution and #OccupyWallSt.

The number of messages in the final dataset is distributed across the five hashtags as following: 45,535 in #FreeIran; 246,736 in #FreeVenezuela; 195,155 in #Jan25; 31,854 in #SpanishRevolution; and 67,620 in #OccupyWallSt. The dataset thus totalizes 586,900 tweets, 258,800 RT-messages and 28,593 AT-messages. The data spans from 9 July 2009 to 14 October 2011, presenting pronounced peaks in September 2009 (#FreeIran), February 2010 (#FreeVenezuela), January 2011 (#Jan25), May 2011 (#SpanishRevolution), and October 2011 (#OccupyWallSt).

Figure 1 shows the distribution of messages over time and depicts two different patterns. The first pattern, shown in #FreeVenezuela and #SpanishRevolution, presents a peak during the first days followed by a gradual decrease in the amount of messages. These hashtags emerged abruptly and disappeared just as suddenly when the bubble burst. The second pattern, shown in #Jan25 and #OccupyWallSt, presents the inverse curve, with messages accumulating in a crescendo until they burst in the last days of archiving. The hashtag #FreeIran is an exception, as it shows a more stable pattern notwithstanding the great number of peaks and troughs. The patterns also reflect the period when archiving processes have started and ended.



freeiran freevenezuela jan25 spanishrevolution occupywallst

Figure 1 Number of tweets over time in the dataset

#FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution, and #OccupyWallSt are political hashtags that trended in Twitter's Trending Topics across the world through messages in up to 27 languages. Eighty-three percent of the datasets from #FreeIran and #FreeVenezuela contain unique messages, while in #Jan25 the proportion is 55%. Due to the sampling performed in the hashtags #SpanishRevolution and #OccupyWallSt, they present a much higher than average number of RT-messages, being 45% in #OccupyWallSt and 48% in #SpanishRevolution, while the percentage of retweets in #FreeIran and #FreeVenezuela is respectively 30% and 32%.

Nonetheless, the five hashtags are reasonably symmetrical in regard to the percentage of messages that include a mention (AT-Network): 4.1% in #FreeIran, 5.7% in #FreeVenezuela and #OccupyWallSt and 3.8% in #Jan25 and #SpanishRevolution. Mentions from a unique user to another unique user also show minor variations, but network topology is not so symmetrical. #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution and #OccupyWallSt have, respectively, 19,599; 626,023; 268,575; 58,734; and 268,643 users interconnected as follower and following. The social graph included in the five hashtags consists of over one million interconnected users from a total of more than 3 billion users.

We analyzed the impact of individual users generating highly replicated messages based on three of the main networks of each hashtag. The first two networks refer to message diffusion, comprehending the network of mentions (AT-Network) and the network of retweets (RT-Network). The third network refers to Twitter's follower (FF) and following (FR) network (FF-Network), which we expected to play an important role in message replication. The dataset was retrieved using the GNU/GPL application yourTwapperKeeper, a web-based application that stores social media data in MySQL tables. Additional server-side PHP scripts were designed to retrieve the social graph.

All matrices were created using the 'network' package for R, which supports arbitrary vertex/edge attributes and provides flexibility for manipulation of large, sparse network class objects. The adjacency matrices are later expanded to include all users that appear across the dataset. The package 'futile.matrix' allows multiple comparisons between matrices of equal size, and the overlapping of RT, AT, FF and FR networks identifies the percentage of RT-messages and AT-messages with users interconnected as follower and following.

Identifying the volume of messages shared by users in a mutual relationship is the first step in determining whether pamphleteering on Twitter relies on word of mouth. Accordingly, RT-messages with senders and receivers interconnected as followers and following rely on Twitter's network to spread the information, while RT-messages without interconnected users should rely on other networks, such as media outlets and peer-to-peer communication. After calculating the percentage of RT-messages with interconnected users, we plotted the network using the 'ggplot2' package for R and Gephi (Bastian et al. 2009).

5. Methodology

In order to evaluate whether message replication in Twitter political hashtags was correlated to the network topology, we performed a statistical correlation analysis considering the following user data: 1. AT-messages to users; 2. AT-messages from users; 3. Number of followers in the hashtag; 4. Total number of followers; 5. Number of following in the hashtag; 6. Total number of following; 7. RT-messages to users; 8. RT-messages from users; 9. Number of tweets in the hashtag; 10. Total number of tweets. Array number 9 refers to the number of tweets posted with a given hashtag, while array number 10 refers to the number of tweets posted since the user started posting.

The correlation analysis is focused on retweets (RT-messages), which is currently considered as the fundamental metric to measure information diffusion on Twitter and the most effective technique for identifying intermediaries that transmit information (Boyd et al. 2010; Cha et al. 2010; Kwak et al. 2010; Mendoza et al. 2010; Suh et al. 2010). Retweeting comprises the relaying of a tweet that has been written by another Twitter user, thus triggering the emergent properties of the Twitter ecosystem. Apart from being an instrument of information diffusion, given that the original tweet is propagated to a new set of audiences, retweet activity can also spot elite users in the network (those generating more retweets than ordinary users).

We tabulated the ten aforementioned arrays and calculated a Pearson correlation coefficient ($p<0.001$) of the components in each hashtag. A correlation coefficient of 0 demonstrates that the variables are independent; a correlation coefficient of 1 means that the two numbers are perfectly correlated. The correlation between variables does not imply that one causes the other. However, the correlation between variables is a necessary condition for linear causation in the absence of any third variable. The correlation analysis was calculated using R, the software for statistical computing, and the results offer insights into the retweet activity determining which array presented the most statistically significant correlation.

Following the results from the correlation analysis, we looked into the network structure of the hashtags to find whether the inner interconnectivity of peers presented features common to the communication pattern of pamphleteering, including a combination of mass-distribution and peer-to-peer sharing. Finally, we used the software Leximancer (Smith 2003) to perform a text analysis of the corpora, extracting concepts and relationships from the five hashtags in the dataset. Leximancer displays a visual map that indicates the concepts and themes embedded within the dataset.

6. Results

The correlation analysis indicated that Hashtag Tweets—the volume of messages posted by single users—presented the most statistically significant correlation with the number of retweets. Table 1 shows the correlation between RT-message to users and the remaining arrays. Table 2 shows the correlation between RT-message from users and the remaining arrays. In both tables RT-messages (from and to users) present a statistically significant correlation with the number of tweets per user (Hashtag Tweets).

RT-messages to users					
	FreeIran	FreeVenezuela	#Jan25	#SpanishRevolution	#OccupyWallSt
AT to Users	0.53	0.19	0.14	0.09	0.00
AT from Users	0.22	0.31	0.17	0.09	0.05
Hashtag Followers	0.16	0.09	0.16	0.20	0.11
Hashtag Following	0.18	0.06	0.14	0.06	0.04
Account Followers	0.00	0.00	0.00	0.00	0.00
Account Following	0.00	0.00	0.00	0.00	0.00
RT from Users	0.56	0.39	0.15	0.12	0.02
Hashtag Tweets	**0.79**	**0.63**	**0.66**	**0.51**	**0.58**
Account Tweets	0.12	0.07	0.11	0.12	0.11

Table 1 Correlations with retweets to users (p<0.001). High correlations highlighted

RT-messages from users					
	FreeIran	FreeVenezuela	#Jan25	#SpanishRevolution	#OccupyWallSt
AT to Users	0.62	0.33	**0.82**	0.47	**0.25**
AT from Users	0.22	0.33	0.13	0.10	0.05
Hashtag Followers	0.08	0.07	0.09	0.08	0.04
Hashtag Following	0.15	0.08	0.40	0.19	0.19
Account Followers	0.00	0.00	0.00	0.00	0.07
Account Following	0.00	0.03	0.00	0.00	0.02
RT from Users	0.56	0.39	0.15	0.12	0.02
Hashtag Tweets	**0.83**	**0.73**	0.28	**0.49**	0.14
Account Tweets	0.09	0.07	0.06	0.03	0.03

Table 2 Correlations with retweets from users (p<0.001). High correlations highlighted

Figure 2 shows each pair of the ten arrays in the dataset qualitatively on a matrix, colored in yellow for high correlation and blue for low. High correlations indicate a relationship between units, while low correlations indicate that the arrays do not vary together. Results indicate that the user's number of tweets in the hashtag—the volume of messages posted by single users—presents on average the most statistically significant correlation with retweet rate. Figure 2 depicts high correlations as dark yellow squares.

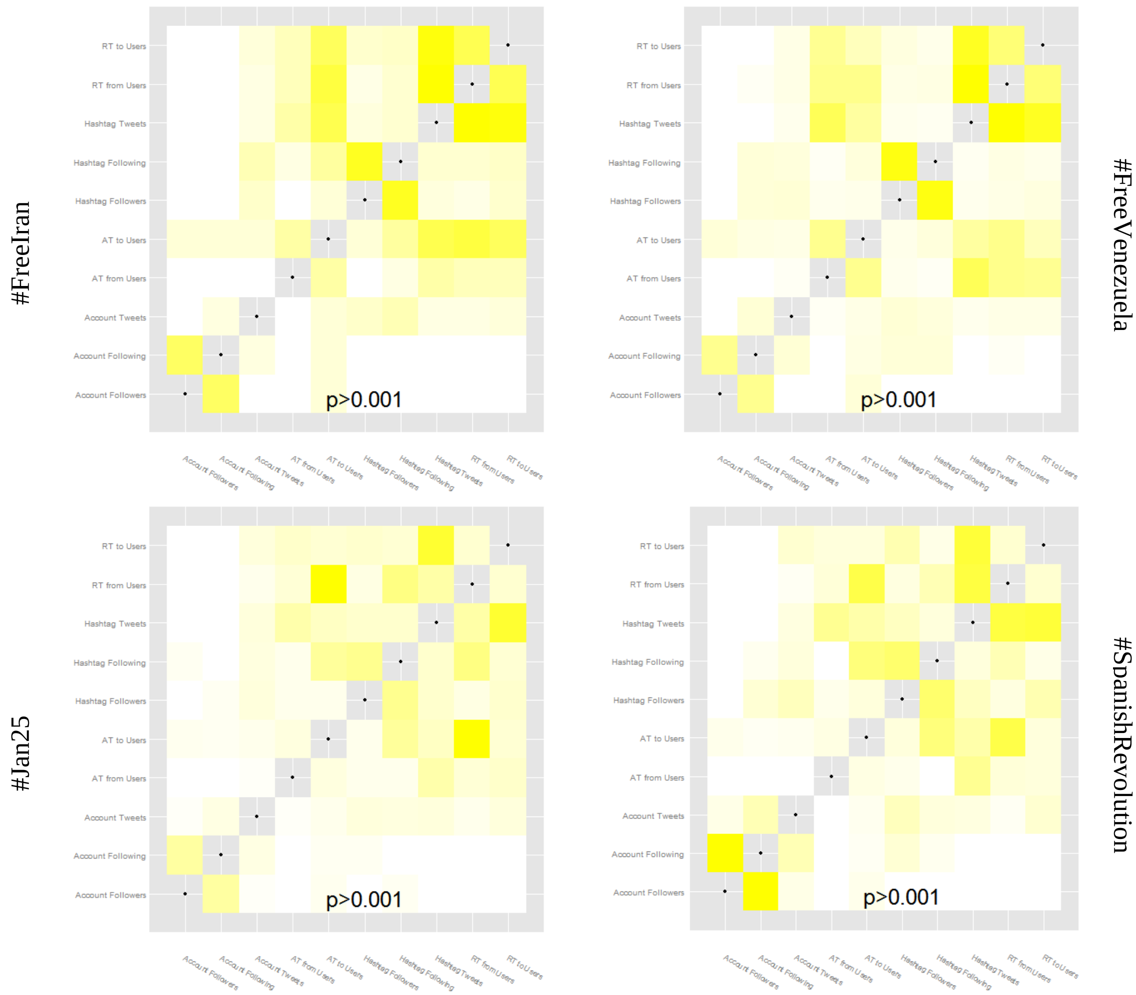


Figure 2 Pearson correlation (p<0.001) for the five hashtags in the dataset

#FreeIran, #FreeVenezuela, and #SpanishRevolution show a statistically significant correlation between retweet rates from users and the number of tweets per user (r=0.83, r=0.73, and r=0.49, respectively, p<0.001). As shown in Table 2, the sheer volume of messages per user is positively and statistically related to retweet activity in the hashtags, followed by messages that mentioned another user (r=0.62, r=0.33 and r=0.47, respectively, p<0.001). That means the growing number of retweets is not correlated to network properties such as a high number of followers and following, but mostly to the assiduous activity of users that are not necessarily hubs or elite users.

In #Jan25 and #OccupyWallSt, messages that mentioned another user (AT-messages) presented the strongest correlation with retweet rates (r=0.83, and r=0.25, p<0.001), thus suggesting that messages were retweeted at the same time that they mentioned another user. #Jan25 data shows that network topology played an important role in retweetability, as retweets have a significant correlation with the followers-and-following network ($r=0.40$, $p<0.001$). Retweet rates and the number of tweets per user is only the third statistically significant correlation in #Jan25, as retweeted messages show a correlation coefficient of 0.28 ($p<0.001$) with the number of tweets per user.

Based on these results we performed a linear regression analysis for each of the five datasets. The results indicate 69%, 54% and 24% variance result to #FreeIran, #FreeVenezuela, and #SpanishRevolution, respectively (Adjusted R Square), thus supporting the results of the correlation analysis and indicating that half of the RT-messages in the three datasets can be explained by the user's number of tweets in the hashtag, while less than one out of one hundred RT-messages can be explained by the number of user's followers participating in the hashtag (1%, 0%, and 0%, respectively).

In addition to that, it is interesting to note that nowhere in the five analyzed hashtags was network topology the array most statistically correlated to the retweeting of messages. This preliminary analysis provided two important evidences about Twitter political hashtags. The first one refers to the low incidence of statistically significant correlations between retweet rate and the first-level network of followers and following, thus suggesting that the emergence of political topics on Twitter is not dependent on the engagement of users with high connectivity positioned at the first-level network.

The second conclusion indicates a statistically significant correlation between retweets and the number of tweets per user, thus suggesting that the emergence of political hashtags relates to the engagement of an unidentified number of users succeeding at sending out messages at high frequency. These users start or "click" a topic that otherwise would not have been featured in the trend topics session or covered by media outlets. This result is consistent with findings of previous investigations showing that Twitter offers an opportunity of outreach for minor, partly marginalized actors, while major political parties and actors find it difficult to adapt to the reciprocal nature of Twitter's AT-messages and RT-messages (Larsson and Moe 2012; Murthy 2011).

7. Pamphleteering in Social Media

In this section we investigated whether the act of tweeting and retweeting messages of political content corresponded to communication strategies associated with pamphleteering. The analysis draws a distinction between social networks, based on relationships within communities, and information channels, based on hubs that broadcast and enforce relevant information. Retweet messages that relied on Twitter's social network include sender and receiver interconnected as follower and following, while retweets that spread without relying on Twitter's social network are similar to mass-distribution media.

The results are broadly consistent with the communication pattern associated with pamphleteering. We found a comparatively higher percentage of retweets sent and received by users who follow each other in the Twittersphere, therefore matching the direct and often confrontational contact common to pamphleteering. We also found a lower than average attendance of opinion leaders or social brokers in the dataset, which is also consistent with the perceived communication strategies of pamphleteering.

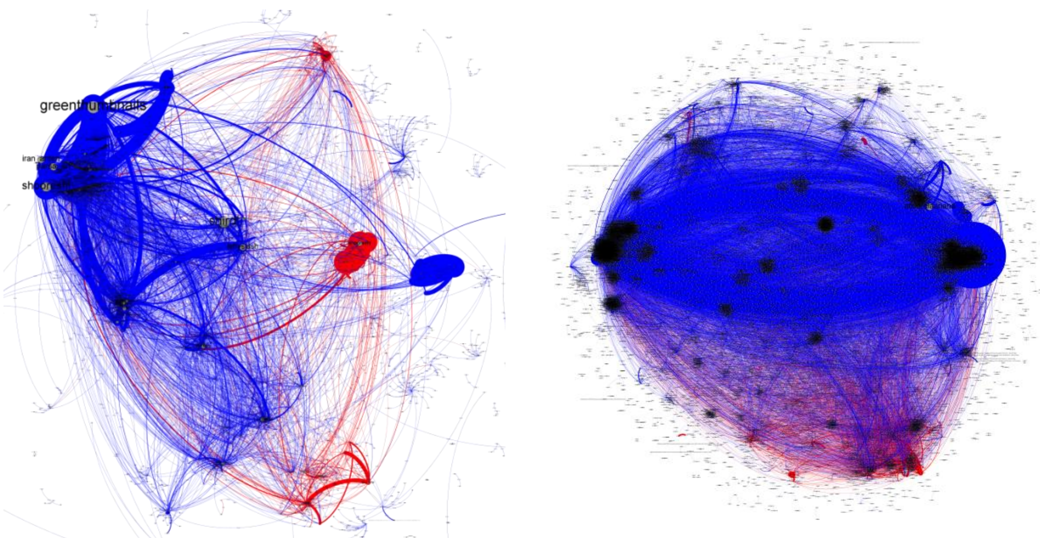
We first looked into the percentage or retweets sent and received by users following each other in the five hashtags and compared the results to hashtags that have no political content. The results indicate that on average 10% of retweets (RT-messages) in Twitter hashtags are sent and received by users interconnected as follower and following. In non-political hashtags this number is significantly lower than average, ranging from 3%-5% in hashtags about music and media personalities, while political hashtags presented a much higher than average percentage of retweets sent and received by users interconnected as followers and following (Bastos et al. 2012b).

	FreeIran	FreeVenezuela	Jan25	SpanishRevolution	OccupyWallSt
Total Nodes	11499637	789602298	1016448421	145988491	1643385121
Connected Nodes	19599	626023	268575	58734	268643
Connectivity	0,001704315%	0,000792833%	0,000264229%	0,000402319%	0,000163469%
RT Total	13985	80912	60047	15428	30390
RT in FF Network	2133	10090	7851	1767	3139
RT out FF Network	11852	70822	52196	13661	27251
RT Connectivity	**15%**	**12%**	**13%**	**11%**	**10%**

Table 3 Percentage of retweets with users interconnected as follower and following (mutual relationship)

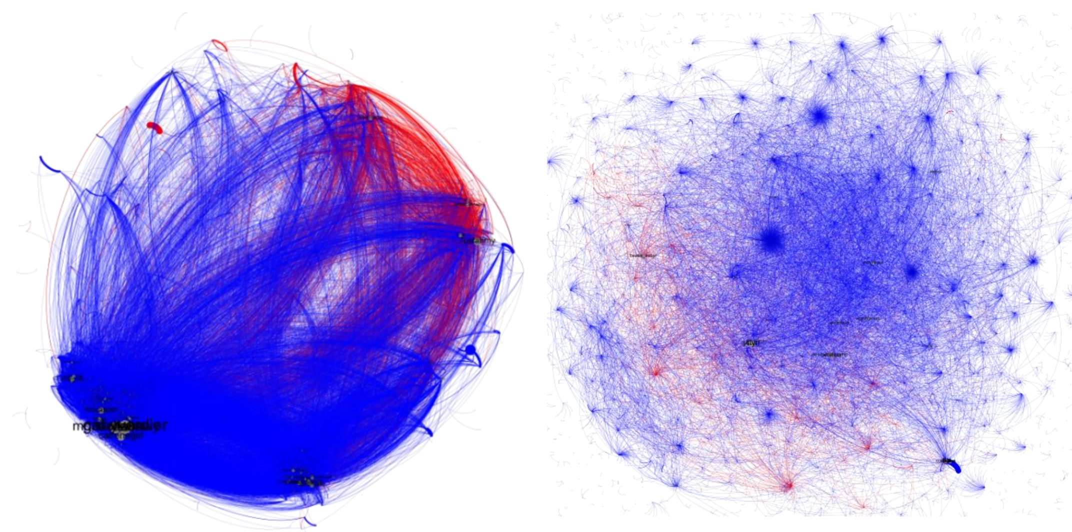
Table 3 shows the percentage of users interconnected as followers and following in the five examined hashtags. These hashtags have on average 12.51% of users interconnected as follower and following, while the average for political hashtags according to previous investigations is of 15% (Bastos et al. 2012b). The results indicate a significantly higher occurrence of users retweeting messages from users they follow and are followed back, thus conforming a community of users who are familiar to each other.

After that we plotted the network of retweets separating RT-messages sent and received by users who do not follow each other from RT-messages sent and received by interconnected users. Figure 3 and 4 show in blue RT-messages from non-interconnected users, while retweets from interconnected users are depicted in red. The plots show the presence of multiple points of broadcasting, but also a higher than average occurrence of red hubs depicting the spread of messages within a community, thus resembling the mechanics of word of mouth.



#FreeIran #FreeVenezuela

Figure 3 Red edges show retweets with users interconnected as follower and following, while blue edges show retweets with users who are not mutually connected (OpenOrd layout)



#Jan25 #OccupyWallSt

Figure 4 Red edges show retweets with users interconnected as follower and following, while blue edges show retweets with users who are not mutually connected (OpenOrd layout)

For political activists, word of mouth is a critical component in the effort to spread political messages. We also found that the presence of professional activists is a consistent trend throughout the whole dataset, being relatively low in hashtags that trended in 2009 (#FreeIran), and expanding further towards a scenario in which the presence of professional activists is substantial, as shown in the hashtags that trended in 2011 (#SpanishRevolution and #OccupyWallSt).

Table 4 was created calculating all Twitter accounts mentioned in the first 30,000 messages of the hashtag. It shows the growing reference to Twitter accounts dedicated to activism and political campaigning throughout the dataset. From the 15 most mentioned accounts, 33% of them are dedicated to political activism in FreeIran; 45% in FreeVenezuela; 75% in SpanishRevolution; and 45% in OccupyWallSt. While FreeIran and FreeVenezuela include a larger number of politicians and media pundits in the top 15 accounts, SpanishRevolution and OccupyWallSt contain a majority of accounts dedicated to political activism.

	2009		**2010**		**2011**		**2011**	
	FreeIran	**Sum**	**FreeVenezuela**	**Sum**	**SpanishRevolution**	**Sum**	**OccupyWallSt**	**Sum**
1	addthis	1633	*presopoliticove*	641	*acampadavlc*	1176	salmanrushdie	798
2	*freeeiran*	1582	*megaresistencia*	375	*acampadaparis*	480	*occupywallst*	412
3	*shary20*	1312	1vzlano	357	*acampadabcn*	438	theonion	398
4	*shooreshi*	1205	chavezcandanga	331	*acampadasol*	347	mmflint	337
5	hambastegimeli	606	globovision	326	*acampadavigo*	324	*avaaz*	286
6	*piroozi1389*	535	albertoravell	285	*acampasevilla*	291	*thinkprogress*	246
7	maryam_rajavi	528	*jocarva*	277	*democraciareal*	281	rainnwilson	200
8	iranaryan	496	gusade	244	*acampadacoruna*	246	amandapalmer	167
9	bluekaruna	346	*no_al_comunismo*	222	eraser	237	*moveon*	164
10	*freedommesenger*	296	*superchavezman*	216	youtube	237	brianstelter	160
11	annedanmark	280	*orvex*	160	*acampadazamora*	182	youtube	139
12	shirdl	255	nelsonbocaranda	157	*acampadagranada*	150	senatorsanders	135
13	lissnup	199	*acuarelaariana*	155	querelle28	130	*occupytogether*	125
14	gloriahere	187	asingusa	148	*alex_riveiro*	123	*occupywallstnyc*	121
15	jefryslash	183	savevenezuela	137	raph83	119	*naomiaklein*	117

Table 4 Top 15 user accounts by total number of mentions in the first 30,000 tweets. Accounts in italics are dedicated to political activism

8. Pamphletweets

Another indication of pamphleteering in Twitter political hashtags was found in the high frequency of a message containing only the hashtag in the body text. These messages are retweets that present no content other than the keyword prefixed with a hash symbol (#) used to create the hashtag. It is important to note that hashtags can be clicked as a link to a global search of tweets using that same keyword, thus increasing the chances of the hashtag to spread through Twitter network. While the occurrence of Twitter messages with no content but the hashtag is consistent through the dataset, it does vary significantly from hashtag to hashtag.

In the top list of FreeIran RT-messages, the retweet containing only the hashtag #FreeIran ranked 54 out of 3035 retweets, while in FreeVenezuela the retweet containing only the hashtag #FreeVenezuela was the most retweeted message out of 17,211 retweets, having being retweeted 896 times against 438 of the second most retweeted message. In Jan25 the iconic message appears in the 117th position out of 15,897 messages with 66 retweets. In SpanishRevolution the message ranked the 101st position, though in related hashtags the iconic message appeared in the 42nd position (#AcampadaBCN) and in the 37th position (#AcampadaSol). In the OccupyWallSt hashtag the non-textual, but persuasive statement appears in the 12th position out of 41,001 retweets, having being retweeted 744 times.

We also looked into the most frequent terms of the dataset. In FreeIran, the first large-scale experience in using Twitter for political goals, the most frequent words include 'OnlyDemocracy4Iran,' a subsidiary hashtag related to #FreeIran, 'addthis,' a social bookmarking website, and 'neda,' referring to the death of Nedā Āghā-Soltān, whose footage filmed by bystanders became a rallying point for the opposition. Other frequent terms include the Twitter accounts of Iranian political activists 'shary20,' 'shooreshi,' and 'piroozi1389.' There is also a recurrent reference to 'sakineh,' referring to Sakineh Mohammadi Ashtiani, who was convicted of adultery and sentenced to death by stoning, and to '22bahman,' referring to the date of the Islamic revolution in the Persian calendar.

Freevenezuela mirrors the struggle between government and anti-government forces in Venezuela. The wide variety of words that refer to the government include 'Chavez' (*Hugo Chávez*), 'government' (*gobierno*), 'president' (*presidente*), 'chavistas,' and 'vivachavez.' The keywords for the anti-government forces are clustered around the words 'freedom' (*libertad*), 'protest' (*protesta*), 'free' (*libre*), 'dictator' (*dictador*), and 'fight' (*lucha*). The most frequent words in the hashtag Jan25 portray the direct confrontation common to traditional activism, including terms related to rallies, street marches, strikes and sit-ins. Among the most frequent words are 'Tahrir,' 'protesters,' 'people,' 'revolution,' 'police,' 'square,' 'freedom,' 'streets,' 'army,' 'thousands,' 'security,' and 'missing.'

SpanishRevolution is an example driving the trend towards stronger activism and the pursuit of political goals. The list of 60 most frequent terms in the dataset includes 25 Twitter accounts managed by political activists. OccupyWallSt is another example of intense activism, presenting a plethora of battle cries or slogans and including highly frequent terms that belong to the vocabulary of pamphleteering, such as 'message,' 'movement,' 'liberty,' 'police,' 'think,' 'stop,' 'protest,' 'media,' 'watch,' 'system,' 'demand,' 'join,' and 'solidarity.' The text analysis of the datasets supports the thesis of a growing use of Twitter as a platform for political goals. Table 5 shows the most frequent words throughout the dataset. The figures indicate the total number and the percentage of occurrence in each hashtag. Figures 5 and 6 show the numbers presented in Table 5 in a chart.

	FreeIran		FreeVenezuela		Jan25		SpanishRevolution		OccupyWallSt	
	Total	%	Total	%	Total	%	Total	%	Total	%
Democracy	505	1.7%	1657	0.7%	95	0.9%	1357	3.2%	NA	0.0%
Fight	321	1.0%	815	0.3%	NA	0.0%	NA	0.0%	NA	0.0%
Freedom	**1377**	**4.6%**	**9005**	**3.9%**	185	1.8%	NA	0.0%	2167	3.5%
News	**1406**	**4.7%**	1688	0.7%	262	2.6%	NA	0.0%	1497	2.3%
Opposition	258	0.8%	1070	0.4%	NA	0.0%	NA	0.0%	NA	0.0%
People	**1597**	**5.3%**	2481	1.1%	**686**	**6.9%**	174	0.5%	1686	2.7%
Prison	471	1.5%	4246	1.8%	NA	0.0%	NA	0.0%	NA	0.0%
Protest	875	2.8%	2133	0.9%	**1355**	**13.6%**	NA	0.0%	**6268**	**10.0%**
Regime	**1468**	**4.9%**	1121	0.5%	181	1.8%	NA	0.0%	NA	0.0%
Revolution	NA	0.0%	NA	0.0%	**532**	**5.3%**	**3726**	**11.9%**	252	0.4%
Stop	457	1.5%	0	0.0%	0	0.0%	0	0.0%	1056	1.7%
Support	552	1.8%	0	0.0%	167	1.6%	0	0.0%	**5965**	**9.7%**
Time	332	1.1%	0	0.0%	141	1.4%	0	0.0%	508	0.8%
Today	367	1.2%	5285	2.3%	350	3.5%	888	2.8%	923	1.5%
Twitter	0	0.0%	6133	2.7%	111	1.1%	0	0.0%	681	1.1%
Video	881	2.9%	0	0.0%	221	2.2%	246	0.7%	1586	2.5%
World	604	2.0%	3184	1.7%	149	1.5%	481	1.5%	460	0.7%

Table 5 Most frequent words in the dataset by number of occurrences and percentage of occurrences in each hashtag. Highlights indicate the most common words.

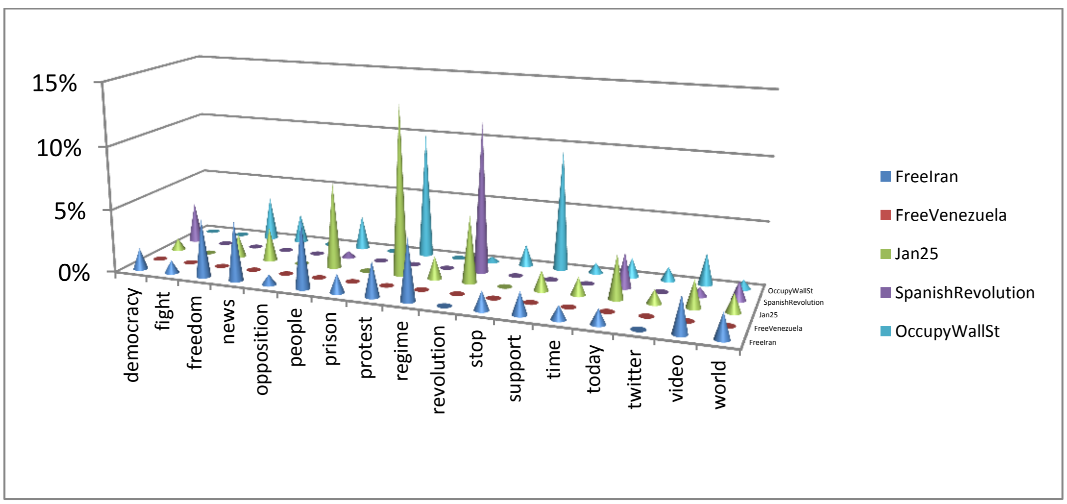


Figure 5 Most frequent words in the datasets by percentage of occurrences

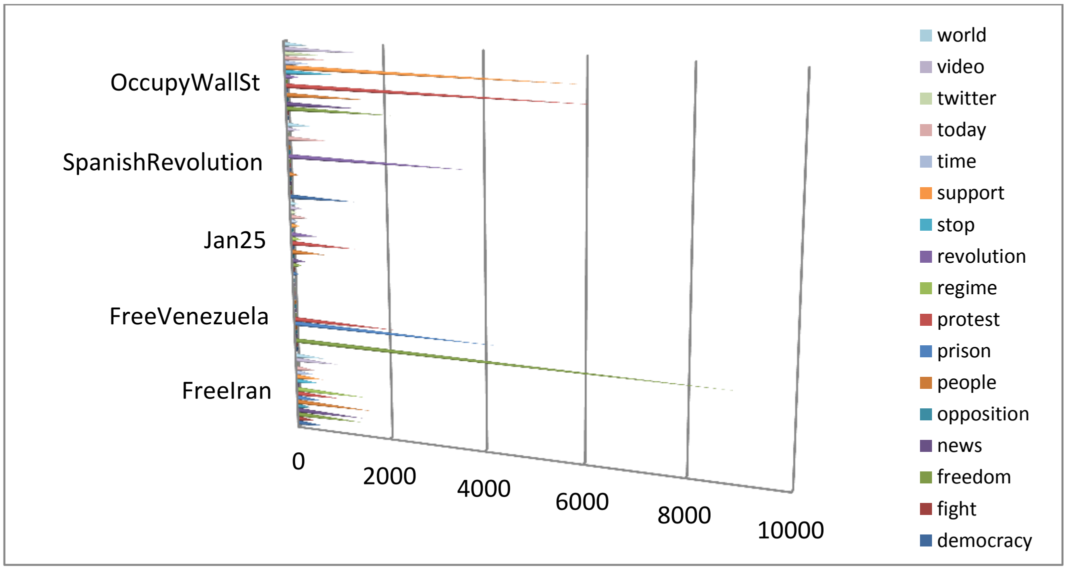
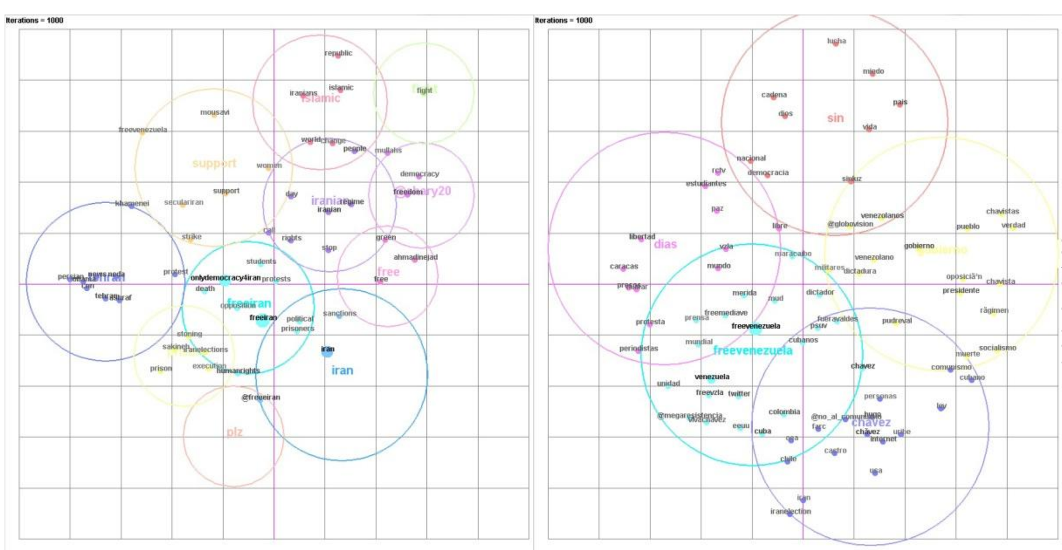


Figure 6 Most frequent words in the datasets by absolute number of occurrences

Figure 5 indicates a high percentage of occurrences for the words ‘revolution,’ ‘support,’ ‘regime,’ ‘protest,’ ‘people,’ and ‘freedom.’ Figure 6 shows a much higher than average number of occurrences for the words ‘regime,’ ‘protest,’ and ‘support.’ These words are recurrent in the universe of political activism, and the relationship between these words sheds further light into the mechanics of activism in Twitter. Figure 7 shows a visual map with concepts extracted from the relationship among these words.

#FreeIran



#FreeVenezuela

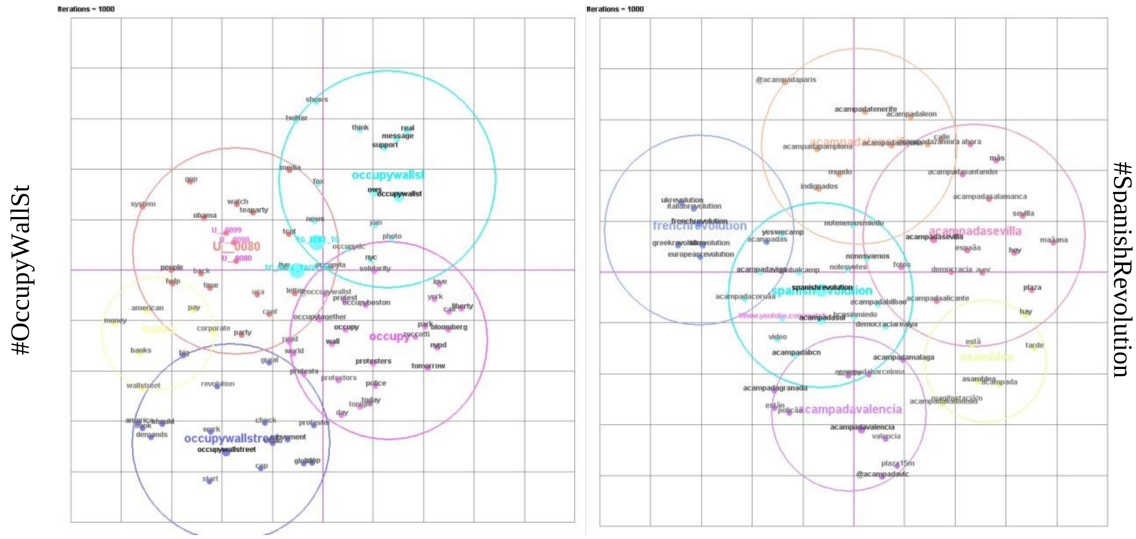


Figure 7 Visual maps of concepts and relationships between words extracted from the text corpora of hashtags FreeIran, FreeVenezuela, OccupyWallSt and SpanishRevolution

The text included in the hashtag #Freeiran presents six axes strictly connected to the slogan 'FreeIran,' encompassing a variety of concepts such as 'death,' 'opposition,' 'political prisoners,' 'fight,' and 'sanctions.' The concept 'support' appears connected to 'secularism,' 'strike,' 'woman,' and 'Mousavi,' a candidate for the 2009 presidential election whose campaign sparked the protests against the Government. The concept 'Islamic' has a connection with the words 'republic,' 'world,' 'people,' and 'Iranians.' The map shows a large number of activist-related concepts, such as 'protest,' 'freedom,' 'prison,' and 'fight.'

FreeVenezuela is organized around anti and pro-government concepts. Pro-government messages are clustered around the concept 'Chávez,' the Venezuelan president, and to a certain extent around the concept 'government.' Anti-government messages are highly connected to the concept 'no' (*sin*), expressing demands such as 'no electricity,' 'no democracy,' 'no life,' or 'no fear'. There is also a high recurrence of words related to news coverage and Venezuelan media outlets, given that the protests emerged as a reaction to the government's decision to ban six TV stations from broadcasting.

Jan25 resembles the picture portrayed at FreeIran, presenting concepts strictly connected to the dramatic events of 2011's Arab Spring. The word 'Cairo' is connected to the words 'protests' and 'police,' while 'Tahrir Square' is connected to 'people,' 'support,' 'protest,' 'people,' and 'revolution.' SpanishRevolution presents a combination of multiple hashtags related to 2011-2012 Spanish protests (*Indignados*), a pattern also present in the hashtag OccupyWallSt. However, the latter includes the concept 'bank' clustering the words 'pay,' 'American,' 'money,' 'Wall Street,' and 'corporate.'

Common to all hashtags is the recurrence of activism-related concepts such as ‘fight’ (*lucha*), ‘freedom’ (*libertad*) and ‘protests.’ The word ‘protest’ appears regularly connected to the word ‘police,’ possibly in reference to the direct and often confrontational tactics common to political activism. There are also repeated cross-references within the dataset, possibly due to the activity of international activists. FreeIran includes the word FreeVenezuela around the concept ‘support,’ while the hashtag SpanishRevolution also appears in the text corpus of OccupyWallSt.

9. Future Work

A major problem with the comparison between tweets and pamphlets is the lack of precision regarding the data on distribution of pamphlets in early modern Europe (Armstrong 1948, 429-30). Even though the pamphlet literature indicates an abnormal value upon the dispatch of news (Armstrong 1948; Mendle 1999; Raymond 2003), the available data is very limited in comparison to the vast and exponentially-growing amount of data generated by social media. Another important difference refers to the difference between tweeting and pamphleteering. While posting a tweet is a relatively trivial action, publishing or delivering a pamphlet involved economic costs and often the possibility of facing legal actions.

Nonetheless, the results presented in this paper support the analogy between tweets and pamphlets previously explored by Moe (2010). The results also suggest the need for further investigation based on longitudinal networks, thus evaluating whether the activity of hubs and edges marked with blue (Figures 3 and 4) is an outcome of the activity of hubs and edges marked with red (Figures 3 and 4). Blue edges depict a broadcasting communication model, as there is no established connection between sender and receiver and the distribution of messages is highly skewed. Red edges depict the spreading of a message within members of a community, thus bearing a resemblance to the underlying determinants of word of mouth.

The statistical modeling of social networks is challenging because of the complicated dependence structures underlying the arrangement of users and messages. This is especially evident when considering the network of retweets. It is necessary to further investigate whether retweets from interconnected users trigger retweets from non-interconnected users, eventually contaminating the whole network. This can be achieved with networks that display a timestamped, continuous chronological axis. This experiment should shed light on whether Twitter political hashtags, and online activism in general, somehow recreate the dynamics of pamphleteering within social networks.

Another key factor to investigate is the relative importance of the second-level network (the network of followers and followings of the followers and/or the network of followers and followings of the followings). The results presented in this paper indicate very low occurrences of statistically significant correlations between retweet rates and the first-level network topology. However, non-significant correlations do not necessarily mean that two variables are not related, as the relationship between variables may be non-linear or non-monotonic.

Finally, the result of the text analysis indicates a noticeable presence of an activism-oriented content. The coordination of additional network and content analysis can describe the reciprocal effects between messages and channels of information diffusion. Even though the findings presented in this study provide preliminary evidence that the content of the messages might be of greater importance than the topological position of the sender, further investigations between the two variables are necessary, though beyond the scope of this paper.

10. References

Armstrong, Charles Arthur John. 1948. "Some Examples of the Distribution and Speed of News in England at the Time of the Wars of the Roses", in Richard William Hunt, W. A. Pantin, and R. W. Southern (eds.), *Studies in medieval history presented to Frederick Maurice Powicke*. Oxford: Clarendon Press.

Bakshy, Eytan, Jake Hofman, Winter Mason, and Duncan Watts. 2011. "Everyone's an Influencer: Quantifying Influence on Twitter", *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. Hong Kong: ACM, 65-74.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. *Gephi: An Open Source Software for Exploring and Manipulating Networks*.

Bastos, Marco Toledo, Rodrigo Travitzki, and Cornelius Puschmann. 2012b. *What Sticks With Whom? Twitter Follower-Followee Networks and News Classification*.

Bastos, Marco Toledo, Rodrigo Travitzki, and Rafael Raimundo. 2012a. "Gatekeeping Twitter: Message Diffusion in Political Hashtags", *Media, Culture & Society*.

Boyd, Danah, Scott Golder, and Gilad Lotan. 2010. "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter". Honolulu: IEEE Computer Society, 1-10.

Cha, Meeyoung, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. 2010. "Measuring User Influence in Twitter: The Million

Follower Fallacy", *International Conference on Weblogs and Social Media*.
Chew, Cynthia and Gunther Eysenbach. 2010. "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak", *PLoS ONE*, 5 (11), e14118.
Earl, Jennifer and Katrina Kimport. 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. Cambridge: MIT Press.
Goel, Sharad, Duncan Watts, and Daniel Goldstein. 2012. "The Structure of Online Diffusion Networks", *Proceedings of the 13th ACM Conference on Electronic Commerce (EC 2012)*.
Grabowicz, Przemyslaw A., José J. Ramasco, Esteban Moro, Josep M. Pujol, and Victor M. Eguiluz. 2012. "Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media", *PLoS ONE*, 7 (1), e29358.
Huberman, Bernardo, Daniel Romero, and Fang Wu. 2009. "Social networks that matter: Twitter under the microscope", *First Monday*, 14 (1).
Kahn, Richard and Douglas Kellner. 2004. "New Media and Internet Activism: From the 'Battle of Seattle' to Blogging", *New Media & Society*, 6 (1), 87-95.
Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. 2010. "What is Twitter, a social network or a news media?", *WWW 2010: Proceedings of the 19th international conference on World Wide Web*. New York: ACM, 591-600.
Larsson, Anders Olof and Hallvard Moe. 2012. "Studying political microblogging: Twitter users in the 2010 Swedish election campaign", *New Media & Society*, 14 (5), 729-47.
Levy, F. J. 1982. "How Information Spread among the Gentry, 1550-1640", *Journal of British Studies*, 21 (2), 11-34.
Mendle, Michael. 1999. "News and pamphlet culture of mid-seventeenth-century England", in Brendan Dooley and Sabrina Alcorn Baron (eds.), *The politics of information in early modern Europe*. London: Routledge, 57-79.
Mendoza, Marcelo, Barbara Poblete, and Carlos Castillo. 2010. "Twitter Under Crisis: Can we trust what we RT?", *1st Workshop on Social Media Analytics (SOMA '10)*. Washington: ACM Press.
Moe, Hallvard. 2010. "Everyone a pamphleteer? Reconsidering comparisons of mediated public participation in the print age and the digital era", *Media, Culture & Society*, 32 (4), 691-700.
Murthy, Dhiraj. 2011. "Twitter: Microphone for the masses?", *Media, Culture & Society*, 33 (5), 779-89.

Quercia, Daniele, Licia Capra, and Jon Crowcroft. 2012. *The Social World of Twitter: Topics, Geography, and Emotions*.

R Development Core Team. 2009. “R: A language and environment for statistical computing.”, *R Foundation for Statistical Computing*. Vienna.

Ratkiewicz, Jacob, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2010. “Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams”.

Raymond, Joad. 2003. *Pamphlets and pamphleteering in early modern Britain*. Cambridge: Cambridge University Press.

Romero, Daniel, Brendan Meeder, and Jon Kleinberg. 2011. “Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter”, *Proceedings of the 20th ACM International World Wide Web Conference*. Hyderabad: India.

Smith, Andrew E. 2003. “Automatic extraction of semantic networks from text using leximancer”, *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Demonstrations - Volume 4*. Edmonton: Association for Computational Linguistics, 23-24.

Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed Chi. 2010. “Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network”, *Proceedings of the IEEE Second International Conference on Social Computing (SocialCom*. Minneapolis, 177-84.

Tumasjan, Andranik, Timm Sprenger, Philipp Sandner, and Isabell Welpe. 2011. “Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape”, *Social Science Computer Review*.

Van De Donk, W.B.H.J. 2004. *Cyberprotest: New Media, Citizens, and Social Movements*. Routledge.

Vittu, Jean-Pierre. 1999. “Instruments of Political Information in France”, in Brendan Dooley and Sabrina Alcorn Baron (eds.), *The politics of information in early modern Europe*. London: Routledge, 160-78.

Wu, S., J.M. Hofman, W.A. Mason, and D.J. Watts. 2011. “Who Says What to Whom on Twitter”. New York: Yahoo! Research.