www.e-compos.org.br
| E-ISSN 1808-2599 |

Perspectivism as a Foundation for Methodologies of Analysis of Social Networks: some results with Twitter

Rodrigo Travitzki, Marco Toledo Bastos e Rafael Luís Galdini Raimundo

Abstract

This article presents a methodology for analysis of variables that influence the dissemination of information on the social network Twitter. We analyze data from 595,240 messages produced by 261,757 users and we found three local variables related to the degree of replication of messages (retweet): 1) the number of responses received (which explained an average of 21% of the retweets); 2) the number of followers in the sub-network (17%) and 3) the number of tweets with the hashtag (7%). Comparing *a priori* variables (global) and *posteriori* ones (local), there was a significant difference in the number of followers of messages ($p<0.001$) and also in the number of friends ($p<0.01$). The higher explanatory power of the local variables suggests a self-organized character of the Trending Topics, indicating paths for future research based on perspectivist epistemology. These results can also help to delineate methods for ranking and segmentation of users.

Keywords
Social network. Twitter. Perspectivism. Influential people. Software.

Rodrigo Travitzki | rodrigo@rizomas.net
Doctoral candidate in Education at the University of São Paulo (USP).

Marco Toledo Bastos | opus@usp.br
Post-doctoral grantee from Fapesp at the University of São Paulo (USP).

Rafael Luís Galdini Raimundo | rafael.raimundo@gmail.com
Doctoral candidate in Ecology at the University of São Paulo (USP).

1 Introduction

1.1 Perspectivism in social networks

Digital media allow the convergence of other media. The Internet, a global space that grew with this principle, has become the largest source of information available to individuals. In this space, technological conditions arose for the flourishing of a new type of social temporality, which has been establishing itself around a "real time" (LÉVY, 1993, p. 117).

Recently, the speed in the creation and reception of information has suffered a great leap with the rise of social networks, which are sub-networks within the global network that are characterized by particular dynamics. In the case of microblog platforms, such as Twitter,[1] users can contribute with zero cost and minimum effort (BAKSHY et al., 2011), generating news, concepts and other ideas that can rapidly have broad repercussions. In addition, there is evidence that the simple number of mentions on Twitter is capable of reflecting the result of a choice (TUMASJAN et al., 2011), which opens new methodological possibilities for opinion surveys that can support social communication strategies, marketing or even government decisions.

Nevertheless, amid so much information, much winds up becoming noise, both to users and to researchers. How is it possible to filter the noise and maximize the information? The social network appears and functions differently for each user. That is, there is a type of "personalized filter" that allows the intelligibility and operationality of the social networks. Each person sees the network from his or her own perspective. But for researchers, is there something similar in methodological terms?

Perspectivism is not an invention of modernity, it can be found, for example, in ancient Amerindian cultures (VIVEIROS DE CASTRO, 1998). According to McGuire (1989, p. 217), perspectivist epistemology is based on the idea that any knowledge is a valid representation only in limited contexts, and therefore multiple theories are needed. Moreover, perspectivism enriches empiricism, not limiting it to a test of hypotheses:

> [...] empiric confrontation is used more effectively, not to test if a certain hypothesis is true, but to discover its meaning [...]Empiric confrontation is best used as an *a posteriori* continuation of the discovery process that began in the *a priori* conceptual analysis.

In a perspectivist approach, the concepts of truth and objectivity do not operate as in logical positivism (ANDERSON, 1998). It does not simply involve testing if an affirmation is true or false, but of striving to better know the meaning or meanings of the affirmation.

The Internet, which is in a constant process of self-organization, is perhaps a more fertile ground for perspectivist than essentialist approaches (GRANIC; LAMEY, 2000).

This study investigates this hypothesis and presents the first results of a methodology for analysis of the users on Twitter, the principles of which can be extrapolated to other social networks. More "essentialist" (global *a priori*) variables were compared, as well as more "perspectivist" ones (local).

1.2 Information on Twitter

What characteristics or behavior of people are more associated to their ability to disseminate information? On social networks, there are at least two non-exclusionary types of cause for this communicative effectiveness, which we can call the "topological factor" (relating to

1 http://twitter.com

| E-ISSN 1808-2599 |

more complex metrics about the position of the user in the network) and the "celebrity factor" (related to more simple metrics, such as the number of followers or of mentions for example) (BOYD; GOLDER; LOTAN, 2010). Users can be classified in three types: information producers, sharers and readers (MAZZOCATO, 2011). In this study we mainly investigate the first type, which is given potential by the second.

Twitter has become consolidated as a base for quantitative research about communication (WU et al., 2011). The main mechanism for dissemination of information in this social network is the retweet, and its dynamic has already begun to be studied (SUH et al., 2010). Most of these studies are based on this analysis of (virtually) all the tweets in a certain period of time (KWAK et al., 2010; ROMERO; MEEDER; KLEINBERG, 2011), which would allow the comparability of tweets, as well as an analysis of the global impact of a certain event.

On the other hand, this method requires computers with high processing capacity, in addition to "special" access to data from Twitter.[2] Moreover, the focus on global characteristics can hide local patterns that refer to each theme. Thus, in this study we use a different methodology, independently collecting and analyzing each Trending Topic.

We thus look to identify local factors related to communication on social networks that can support methodologies for ranking and segmentation of users, for low-cost spontaneous opinion studies and others.

A common form of research of mechanisms for dissemination of information is based on the identification of the different degrees of influence of people on a network. There are, however, different variables related to the degree of influence, and in recent years various studies have mapped these relationships (BAKSHY et al., 2011; CHA et al., 2010; WENG et al., 2010; WU et al., 2011; YE; WU, 2010). According to these authors, the main indicators of influence of users on Twitter are:

a) number of followers, or (*);
b) number of retweets (*);
c) distribution of tweet cascades;
d) number of mentions;
e) number of responses received (*) or of responses received by single users;[3]
f) ranking of the site ().

To better understand the dissemination of information on social networks we study here the factors related to the number of retweets of each user (b). There is evidence that this variable is

2 Like the *whitelists* or the *Phirerose*.

3 The variables marked (*) were measured in this work.

among the most stable as an indicator of user influence (YE; WU, 2010).

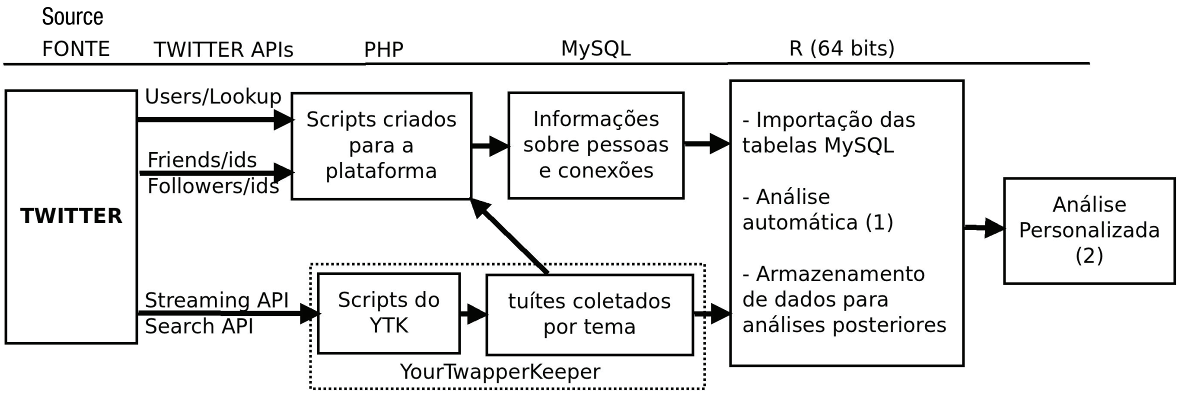
2 Methodology

2.1 T Platform

To (virtually)[4] collect all the tweets that mention a hashtag during a certain period, we used the software ,[5] licenced in . However, to obtain detailed information about the users who write, receive or were mentioned in each tweet, we created some scripts in PHP and MySQL. The analysis of the data, in turn, was conducted in R, using the scripts articulated to the first. We call this system the T platform, represented in Figure 1.

The objective of the PHP is to collect data uninterruptedly from the Internet, storing it in MySQL tables. These are imported to R, where they can be analyzed in various manners, some of which we describe below. Observe the diversity of the APIs and articulation among different languages and programming environments. Much pertinent information can already be obtained with the automated analysis (1). If necessary, some semi-ready

Figure 1 – T Platform - a set of scripts capable of automatizing the analysis of themes on Twitter – integrated to the program YourTwapperKeeper and to different Twitter APIs.



Scripts created for the platform > Information about people and connections > Importation of MySQL tables
Automatic Analysis Personalized Analysis
YTK Scripts > tweets collected by theme > Storage of data for later analysis

4 The program YourTwapperKeeper uses two APIs in a complementary manner to collect all the tweets containing a certain term, but overloads the Twitter servers. Among other technical problems, it can cause one or another message to fail to be collected.

5 http://github.com/jobrieniii/yourTwapperKeeper

scripts help to conduct different types of analysis, customized according to interest (2). Our sample is composed exclusively of hashtags that became Trending Topics in 2011, with a minimum of two thousand and a maximum of 50 thousand tweets. Each one of these sets of tweets that mention a hashtag in a certain period is constituted in a unit of analysis, and the various data collected about these tweets will be called here hashdata. Table 1 shows the 32 Trending Topics analyzed in this work. We construct three different networks for each one: one for retweets (RT), one for responses (RP) and another of friends and local followers – within hashdata – (LF), all with connections directed toward the dissemination of information. The indegree and outdegree edges of each user in each network were counted and we calculated some simple metrics. This technique allowed working with local variables () of relatively difficult access, given that they are not directly accessible by the APIs.

In addition, we collected various types of information about the users (who wrote the tweets of the sample or are in them in retweets or responses). Those without a public profile were removed from the analysis, for both technical and ethical reasons. For each user, the following numeric variables were obtained:

1) opening of the account: time since the date of creation of the Twitter account;
2) followers (global): number of followers on Twitter;
3) friends (global): number of friends on Twitter;
4) status (global): total number of messages written on Twitter;
5) status (local): number of messages in hashdata;
6) <u>followers (local)</u> - LF out: number of followers in hashdata;
7) friends (local) - LF in: number of friends in hashdata;
8) <u>retweets - RT out</u>: number of times retweeted in hashdata;
9)retweets sent - RT in: number of times retweeted in hashdata;
10) responses sent - RP out: number of personal messages sent in hashdata;
11) <u>Responses received- RP in</u>: number of personal messages received in hashdata.[6]

The local variables were not found in the literature, which is normally restricted to the information received directly from the APIs (as a global number of followers, obtained by the API users/lookup). Perhaps this predilection for information from the API is due to the multiple methodological difficulties that we had in preparing the scripts of the T Platform, or perhaps because this type of information is only

6 The variables underlined correspond to those marked (*) in the introduction. Variables 1 - 4 were obtained directly from the APIs of the Twitter, while the others were calculated with specific scripts for the T platform. Finally, variables 6 - 11 correspond to simple metrics (number of connections - indegree and outdegree edges – of each vertex, with the vertexs being the users) of the three networks mentioned above.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

valid within its own context, limiting its use as a predictive model. It is important to highlight at least two peculiarities of our method: 1) we chose to focus on characteristics of the users related to the general structure of Twitter – the connections, the retweets and responses (which correspond to the three networks constructed) – more specific to a certain theme, given that they mention the same hashtag. In this way, it was possible to reach the second peculiarity: 2) we obtained some qualitatively similar measures, although in two different levels – that is, on Twitter as a whole or only in the sample chosen as a specific topic (which we call global and local variables, or variables and , respectively). They are three variables of this type, totaling six among the eleven measured for each user: the number of messages (statuses), the number of friends and followers.

2.2 Influence of the users

We conducted an analysis of the influence of the users in terms of number of retweets. This was done independently in each theme, so that we could later compare the values obtained in the 32 samples. We evaluate the importance of the various characteristics of the user in the probability that he or she be retweeted, seen as an indicator of influence on the network. That is, it involves a particular element of the more general question: in what conditions does information tend to be most disseminated?

In this analysis of user influence we began calculating the correlations (Pearson, $p<0,05$) among the 11 variables collected about each user in each hashdata. Based on these results, we chose three variables that could potentially explain the number of retweets and we adjusted some regression models. We thus verified the correlations (Pearson, $p<0,01$) of these three variables among themselves in order to respect the presumptions of the multi-varied regression, and we opted to adjust only one model of this type. There were four different models in our regression analysis, all had the number of retweets as a variable response: three of them were uni-dimensional, corresponding to the three indicators selected, and one multi-varied, combining status (local) and followers (local) as explanatory variables (Figura 6). Therefore, 128 models were adjusted, four for each one of the 32 hashtags. For each model, in turn, we calculated three values – the AIC, the p and the adjusted coefficient of determination – totaling 384 measures for comparison of models.

For the comparison of the local and global variables, we use the Student's T test with some of the correlations mentioned above. Before this, we substituted the data missing with zeros, based on the principle that when R was not able to calculate the correlation, it can be considered null.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

2.3 Segmentation

A final technique that deserves to be mentioned is that which is used as the basis for the construction of the dendrogram (Figure 2) or, in a more general manner, the segmentation of users and themes, depending on what is desired, based on the same set of data. This is a reasonably simple procedure for tabulation and grouping, based on the overlapping of users, or that is, the occurrence of certain users in more than one hashtag of the set analyzed.

The first step of this technique consists in collecting a certain set of hashtags (or words or expressions), according to defined objectives. Later, for each one of them, it identifies the users (using regular expressions or information from the APIs, for example), storing them in column A of a table. Column B must be completely filled in with the same word, which corresponds to the hashtag. Once this is done for all the hashtags in the group, the tables are added vertically, resulting in a single table of two columns. Based on this general table it is possible to compare both users and hashtags.

In the first place, we can rank the users by simply tabulating column A. This can even be another local variable to be included in the models of influence of the user. For segmentation of the hashtags, however, more procedures are needed. Based on the same data, users per hashtag are tabulated, creating a matrix composed of zeros and ones, of U lines per H columns (with U the number of unique users and H the number of hashtags in the previously selected set). In graph theory, this would formally be an adjacency matrix of a , a bipartite graph (HARARY, 1969). A distance matrix is constructed based on this matrix, which in turn could permit the segmentation. For the dendrogram of Figure 2, we respectively used the and methods, although others are available for calculation of both the distances (BORG; GROENEN, 1997) and segmentation (HARTIGAN, 1975). Future studies will be needed to compare these methods. Finally, based on this segmentation of hashtags, a segmentation of the users present in them can also be conducted. This would establish possible uses for the data found in the general table mentioned: ranking users, segmenting themes and consequently segmenting users.

3 Results

We analyzed 595,240 messages – of which 217,344 were retweets and 51,817 responses – distributed in 32 with size varying from two thousand to 50 thousand tweets each. Most were in Portuguese, with the exception of three (#notenemosmiedo, #DemiYouAreBeautiful and #misstwitter). In spite of this, the hashdata had about ten languages each, reflecting a certain internationality of Twitter. Table 1 shows the 32 themes and their basic variables.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

| E-ISSN 1808-2599 |

Table 1 – Summary of the hashdatas. Each hashdata corresponds to a series of variables obtained from a set of tweets about a certain Trending Topic

Hashtag	Beginning of the collection	Duration (days)	Number of users	Tweets	Retweets	Personal responses
#1bomprofessormeensinou	12/04/2011	78	16855	28697	17498	233
#abaixodecreto	22/02/2011	93	11372	44813	17274	13466
#adoteumanimalabandonado	02/07/2011	62	1879	3131	1448	107
#amorodeio	23/08/2011	8	8079	16725	6325	1631
#battisti	08/06/2011	27	3309	6434	2658	365
#DemiYouAreBeautiful	29/08/2011	5	7736	11971	3915	1283
#diadoadvogado	11/08/2011	17	10513	12689	4595	587
#diadofrevo	09/02/2011	35	2227	3601	1198	367
#diadoreporter	16/02/2011	32	4927	6596	2070	1047
#diamundialsemtabaco	31/05/2011	34	9223	14320	3994	600
#escolhiesperaremdeus	05/07/2011	49	15048	27274	13068	653
#estudarvaleapena	11/08/2011	29	19798	26638	9742	791
#fichalimpa	23/03/2011	69	17034	32255	11659	1597
#forabolsonaro	31/03/2011	76	18030	30526	11038	1819
#frasesqueeikenuncadiz	12/02/2011	38	1821	3929	1034	219
#menoscormaisrock	09/05/2011	119	5113	15472	4354	2649
#misstwitter	27/06/2011	54	12951	35774	8549	12148
#naofoiacidente	01/03/2011	86	5892	15124	8133	710
#notenemosmiedo	19/05/2011	26	14901	33908	19179	1586
#odeiorodeio	23/08/2011	6	13134	26988	12400	1921
#otwittermeensinou	09/01/2011	74	924	2199	1243	15
#pedofilianao	27/06/2011	54	3908	6308	2939	665
#qndomertiolateardia	27/05/2011	27	8455	15839	4772	399
#realengo	08/04/2011	126	19470	33105	11602	2247
#seeufosserico	09/07/2011	50	16418	24440	5874	381
#sosnatal	15/06/2011	16	2495	6318	3136	459
#trabalhoescravo	17/08/2011	19	3496	4315	2020	154
#tuíteeumsonho	15/06/2011	5	19418	26574	6165	591
#tuíteumlivro	27/05/2011	5	28411	43960	6311	577
#tuíteumnomeestranho	18/02/2011	30	5102	8095	1009	393
#tuíteumproverbio	09/05/2011	17	6000	10474	4146	213
#uniaohomoafetiva	06/05/2011	22	11593	16748	7996	1944

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

| E-ISSN 1808-2599 |

Note that the number of responses is systematically lower than that of the retweets, confirming their role as the principal mechanism of Twitter (SUH et al., 2010). The proportion of retweets in the 32 Trending Topics had a normal distribution (X=0,38, SD=0,12), with a minimum of 12.5% and a maximum of 61.0%. This proportion appears greater than that in the previous study (KWAK et al., 2010).

3.1 Segmentation of users and themes

To analyze the factors that influence the dissemination of information in the social networks, we focused the investigation on the characteristics of the users (who tweet, receive responses or were retweeted), of each hashdata. Some users were found in more than one Trending Topic.

Table 1 – Recurrence of the users along the 32 *hashtags*

TTs em que foi citado	Número de Usuários	N.U. (%)
1	218047	83,3
2	31253	11,94
3	8095	3,09
4	2642	1,01
5	949	0,36
6 a 25	771	0,29

A total of 261,757 users were analyzed, of which 16.7% were found in more than one of the 32 hashdatas, which suggests a certain overlapping of the public involved in the themes selected here. It is possible to use this information as a form of grouping the hashtags or the users based on the same set of data, carefully choosing the set of hashtags/words to be collected continuously by the T platform. This technique can be effective for surveys of opinions and segmentation based on the Twitter.

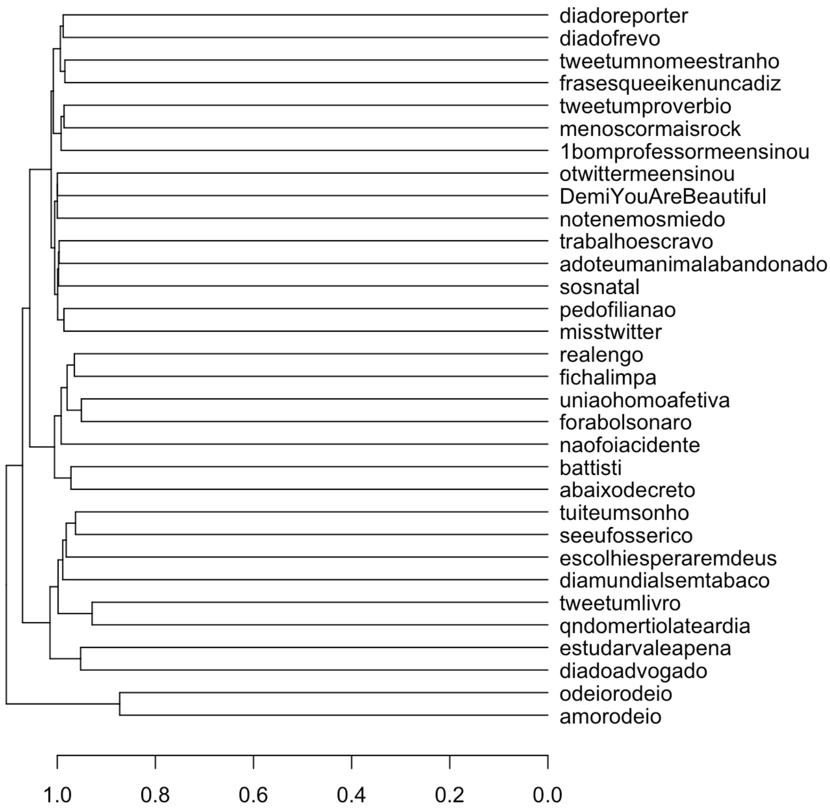
We can begin with the principle that the overlapping of the users would be more common in similar hashtags in some sense. Only to illustrate the possibilities, Figure 2 shows a dendrogram that groups the issues as a function of the competition for users among them.

It is interesting to observe that two hasthtags with opposite content show themselves to be quite close in terms of recurrence of users (#amorodeio and #odeiorodeio). A certain trend was also noted of grouping of "political" themes (#forabolsonaro, #fichalimpa, #battisti, #abaixodecreto, #naofoiacidente, #realengo). However, these are preliminary results, using one of various grouping methods. New studies can be conducted to investigate the effectiveness of the different methods depending on the finality (segmentation of themes or of users, or ranking of users).

The results and the very effectiveness of this method depend, among others, on the choice of the set of themes, which reflects their perspectivist character.

www.e-compos.org.br
| E-ISSN 1808-2599 |

Figure 2 - Dendrogram illustrating possibility of grouping of hashtags as a function of overlapping of users.



3.2 The dissemination of information

Of the 11 variables obtained from each user, one was the number of retweets in the hashdata, considered here as a main indicator of the influence of the user – his capacity to disseminate information. In this sense, we consider the other ten variables as potentially explanatory. We can look at a sample of the raw data to have a better idea of what was done. Table 2 shows some of the variables obtained from each user in Trending Topic "#diadofrevo".

Note that the number of retweets (RT) appears to be explained by different variables depending on the user. The most retweeted (@marceloadnet0) is a famous television comedian, and can be detected with any one of the indicators (global or local followers, or responses received). The second user (@carnavalrecife), however, is more intimately related to the hashtag (given that Recife is considered the capital of frevo, a traditional Brazilian dance), and would be better detected by local than by global followers. The third, in turn, although it has few followers, receives many responses. And the fourth user

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

| E-ISSN 1808-2599 |

Table 2 - Example of the raw data collected from each user in each *hashdata*

User name	RT	Status (global)	Followers (global)	Status (local)	Followers (local)	Responses received
@marceloadnet0	261	1360	657218	1	330	8
@carnavalrecife	75	2661	3608	19	121	9
@giiorgio_	44	3142	389	1	0	11
@lucas22monart	24	3685	217	42	2	2
@vanessa_violeta	18	566	402	1	12	0
@c_amaraloficial	14	820	261	36	0	6
@frevodepressivo	13	106	769	2	56	1
@edizinhah	9	16756	183	33	5	0

(@lucas22monart), finally, would not be detected with any of the indicators of influence. However, we note that he tweeted many messages with that hashtag. This user can be considered an example of personal effort that generates a certain break in the spontaneous standard of the dissemination of information (strongly influenced by the "celebrity factor"), which can be related to the political role of platforms such as Twitter or social networks in general.

To verify the hypotheses illustrated above, we first calculated the correlations between all the variables at each hashdata. We then separated the results referring to the correlations among the ten potentially explanatory variables and the number of retweets and counted the number of significant correlations ($p<0,05$) in the 32 and we calculated their averages (Figure 3).

Note that the correlations are low and often rare, but there are some potential explanatory variables. To what degree do these results confirm the results of previous studies? Of the 11 variables that we had for each user, three are considered in the specialized literature as potential indicators of influence. Of these three, one is the number of retweets itself, and therefore there are still two others to be tested as possible explanatory variables, or indicators of this. In Figure 3, the number of responses is the variable most correlated to the number of retweets, confirming what was already supposed. Another result that confirms previous studies (SUH et al., 2010) is the irrelevance of the time of existence of the Twitter account for the degree of influence of the user.

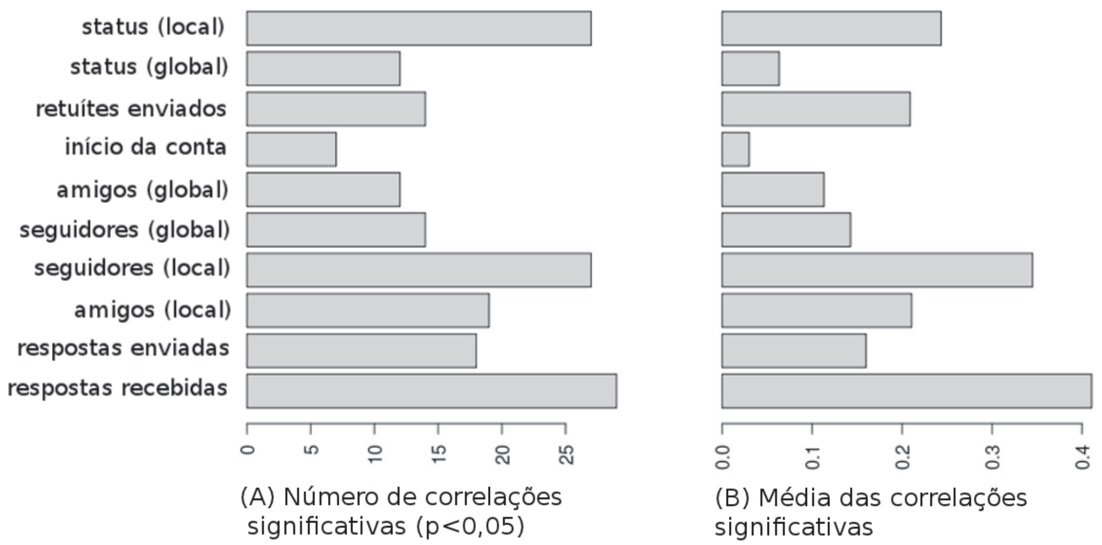
3.3 "Global" and "local" variables

In the second variable potentially indicative of the retweet (the number of followers) the correlations

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

| E-ISSN 1808-2599 |

Figure 3 - Correlations between the possible explanatory variables and the number of retweets for each user. Note the number of valid correlations (A) and their average (B).



Retweets sent – opening of account – friends (global) followers (global) – followers (local) – friends (local) – responses sent – responses received

(A) Number of significant correlations
(B) average significant correlations

do not appear as consistent, at least when we observe their "global version." However, when we observe the number of local followers, or that is, those within a specific theme, there appears to be more solid correlation. This takes place in the three variables that have "local" and "global" versions (followers, friends and status – which are the messages – the tweets).

The "global version" of the variables corresponds to the three numbers that are seen when clicking on someone within a twitter. Since this data can be obtained before filing specific tweets of a hashtag, they were considered here as variables. The "local version" of the variable was calculated with specific scripts from the T Platform, based on a certain hashdata, that is, it can only be obtained *a posteriori*. Very well, in what degree are the apparent differences between these three pairs of variables statistically consistent? Comparing the global and local versions of each pair, the student's T test reveals a difference in all of them, which is more significant between number of status (p=2e-5), followers (p=2e-5) and a bit less among friends (p=0,008). That is, the local variables () present, on average, greater correlation with the retweet than with the global variables (), when we compare variables of the same type.

This suggests that the degree of influence of the retweet in a subject is more closely related to the internal dynamic of the issues itself than a general structure preceding or exterior to it.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

In other words, the comparison of the global and local variables (Figure 3) suggests that the dynamic of the retweet is more influenced by factors created during the process than by those data – a characteristic present, for example, in self-organized processes (DEBRUN, 1996) and emerging properties (KIM, 1993, p. 346). The self-organization would be a spontaneous increase of order generated in the points of instability, or of points of bifurcation, and is intimately related to the living phenomenon and their manifestations (CAPRA, 2002). Although it has a certain unpredictability, related to bifurcations, the self-organization can be analyzed , using variables created during the process itself.

This is a possible theoretical background for studying social networks as heterogeneous spaces from which local phenomenon continuously emerge that can be better understood based on their own parameters – that is, from a perspectivist approach (MCGUIRE, 1989). We can say, therefore, that these results confirm the hypothesis that perspectivism can be more useful than essentialism in the analysis of the Internet (GRANIC; LAMEY, 2000) – especially of social networks.

Finally, all of these elements confirm the idea that Twitter is an rapid movement, of continuous creation of sub-networks within the network, and is more explainable by its temporal characteristics than by the isolated properties of the elements that compose it, or by previously known variables. This conception can help to understand the difficulty of adjusting models capable of foreseeing the dynamic of information in social networks. In this sense, the investigation conducted here perhaps has more to contribute to the analysis of the themes than to the development of algorithm to estimate future patterns. This can be useful in the elaboration of opinion and market studies based on information spontaneously provided by people – a new branch of research that arises with the blogs, microblogs and social networks.

3.4 Detecting those who influence public opinion

Based on the analysis of correlations, we selected some variables for regression analysis. Observing the measures in Figure 3, two variables stand out: 1) the local number of followers, which proves to be more significant than the global – which appears to be important (YE; WU, 2010), but less than supposed (CHA et al., 2010); and 2) the number of responses received – a factor considered important both in Twitter (BOYD; GOLDER; LOTAN, 2010) and in other social networks (RECUERO, 2005). Nevertheless, if we also observe the number of significant correlations, a third potentially significant factor appears: 3) the status number in the sample; This third variable, unlike the first two, was not found in the literature as an important factor for the degree of influence of the users.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

Once three factors are identified, we began to verify the relations between them. The average correlations (Pearson, p<0,05) between the three potential indicators along the 32 hashdatas were: 0,27 between responses and followers; 0,22 between responses and status; 0,13 between status and followers. The highest average found in the correlations between responses received and local followers confirms the idea that they can be indicators of the same property, possibly related to the "celebrity factor." The status number (local), in turn, is a bit less correlated with the others, suggesting that it can be an additional source of information (with low redundancy), especially if articulated to the number of followers (local). Based on these observations, we can opt to adjust four types of models for each hashdata, three of which are un-varied (having the three indicators as independent variables and the retweet as a dependent variable) and one that is multi-varied.

By comparing the p values of the models, it is found that the multi-varied model was invalid (p>0,001) in only two of the 32 hashtags, while the model for responses was invalid in three. The other two models were invalid five times each. There is little difference in the results, but they reveal the statistical significance of the first two models.

A comparison of the AICs of only the uni-dimensional models found that in 16 hashdatas the number of responses received better explained the retweet than the other two. In 14 hashdatas the number of local followers better explained the retweet, leaving to the number of local status only two of the 32 TTs. The comparison of the coefficient of determination presented the same result. Finally, when we include the multi-varied model in the comparison, it is considered the best 16 times, followed by the "responses" model (14 times) and by local followers (2 times).

Finally, the comparison of the coefficients of determination reveal a large variation (Table 3), while there are significant differences[7] between responses received and status (p<0,001), but mainly, as expected, between the multidimensional model and that of the responses received (p<2e-8), followed by that of status (p<0,001) and less by that of followers (p<0,01).

Table 3 - Coefficients of determination of the models for retweet

Modelo	Média	Desvio Padrão
Respostas recebidas	0,21	0,24
Seguidores (local)	0,14	0,18
Status (local)	0,08	0,13
Seguidores + Status	0.20	0.21

4 Conclusions

This study had two objectives: 1) to present and test the T Platform, which was developed especially for collection and analysis of Twitter

7 Student's T Test with paired samples.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

data; 2) map the indicators of influence of the users, comparing the local and global variables as a form of evaluating perspectivist and essentialist approaches.

In terms of the first objective, the structure of the T platform (which included Twitter API, PHP/MySQL and R) proved to be reliable and flexible, capturing tweets in a selective and continuous form, realizing automatic and customized analyses. The achievement of the results expected confirms the quality of the data collected and the effectiveness of the methods.

In the analysis of the influence of the users, we consider the number of retweets for a certain theme to be a principal indicator, based on which we mapped the importance of the other factors. Of the ten numeric variables of each user (in a specific theme), three others correlated to the number of retweets were identified in order of importance:

1) RP: the number of responses (personal messages) received within the theme (in this case, one hashtag);
2) The number of responses (personal messages) received within a theme (in this case, one hashtag);
3) SE: followers (local), the number of followers among the people who tweet the hashtag;
4) ST: status (local), the number of tweets with the hashtag.

The regression analysis of the three variables revealed that in different Trending Topics the retweet can be better explained by one or another factor, and rarely by neither of them. Notwithstanding this variation, the two best models among the four tested were significant ($p<0,01$) in more than 90% of the samples. Confirming the results of the analysis of correlation, the variable most capable of explaining the retweet was RP (21%), followed by SE, which on average explained 17% of the variation in the number of retweets. The third variable (ST) explained an average of only 7% of the retweet, but its low correlation with SE allowed the creation of the multi-varied model, combining local connections (SE) and effort (ST). This model explained on average 20% of the variation of the retweet, with verisimilitude (AIC) subtly better than the uni-dimensional model by RP. These results can help to delineate methods for ranking and segmentation of users.

On this basis we constructed a dendrogram representing different levels of grouping of hashtags, as a preliminary result to illustrate the potential of the method. This analysis was only possible because nearly 17% of the users analyzed occurred in more than one of the 32 hashtags, although they had not been collected as if they were part of a single subject. This overlapping allows the use of the method segmentation of subjects by co-occurrence of users, or the opposite, segmentation of users by co-occurrence of hashtags. Both are based on the same data, referring to a set of themes

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

chosen by the researcher – opening space for a perspectivist approach.

In the three "dual" variables (the status number, and of friends and followers) the local version is more correlated to the number of retweets than the global version, suggesting that the dissemination of the information has a dynamic barely determined , but that can be analyzed , based on the parameters created during the process itself. In this sense, confirming the hypothesis of Granic and Lamey (2000), perspectivism can constitute an important theoretical background for Internet analysis.

These results are important in the establishment of methodologies for mapping the dissemination of information and influence in social networks, which in turn can help studies in this field, and to improve public opinion studies based on the Internet.

References

ANDERSON, R. Truth and Objectivity in Perspectivism. **Synthese**, v. 115, n. 1, p. 1-32, 1 abr. 1998.

BAKSHY, E.; HOFMAN, J.; MASON, W.; WATTS, D. **Identifying "Influencers" on Twitter**. In: ACM INTERNATIONAL CONFERENCE ON WEB SEACH AND DATA MINING (WSDM). 4., 2011. New York. **Proceedings**... New York: ACM, 2011.

BORG, I.; GROENEN, P. **Modern Multidimensional Scaling**. Theory and Applications. New York: Springer, 1997.

BOYD, D.; GOLDER, S.; LOTAN, G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES (HICSS-43), 43., 2010. Kauai. **Proceedings**... Kauai: IEEE, 2010. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5428313>. Acesso em: 4 jan. 2012, 2010.

CAPRA, F. **As conexões ocultas.** São Paulo: Editora Cultrix, 2002.

CHA, M. et al. Measuring user influence in Twitter: The million follower fallacy. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 4,, 2010. Washington. **Proceedings**... Washington: AAAI, 2010. Disponível em: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1538/1826>. Acesso em: 4 jan. 2012, 2010.

DEBRUN, M. A dinâmica da auto-organização primária. Auto-organização: estudos interdisciplinares, Centro de Lógica, Epistemologia e História da Ciência. Campinas: Unicamp, 1996. (Coleção CLE n. 18).

GRANIC, I.; LAMEY, A. V. The self-organization of the Internet and changing modes of thought. **New Ideas in Psychology**, v. 18, n. 1, p. 93-107, 2000.

HARARY, F. Determinants, permanents and bipartite graphs. **Mathematics Magazine**, v. 42, n. 3, 1969.

HARTIGAN, J. A. **Clustering Algorithms**. New York: Wiley, 1975.

KIM, J. **Supervenience and mind**: selected philosophical essays. New York: Cambridge University Press, 1993.

KWAK, H. et al. What is Twitter, a social network or a news media? In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB. 19., 2010. New York **Proceedings**... New York: ACM, 2010.

LÉVY, P. **Tecnologias da inteligência**: o futuro do pensamento na era da informática. Rio de Janeiro: Editora 34, 1993.

MAZZOCATO, S. B. Os Papéis do Sujeito com Relação a Conteúdos Midiáticos no Tweetdeck: o produtor, o compartilhador eo leitor. **E-compós**, v. 14, n. 2, 2011.

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.15, n.3, set./dez. 2012.

MCGUIRE, W. J. A perspectivist approach to the strategic planning of programmatic scientific research. **The psychology of science: Contributions to metascience**. New York: Cambridge University Press, 1989. p. 214–245.

RECUERO, R. O capital social em redes sociais na Internet. **Revista FAMECOS**, n. 28, p. 88-106, 2005.

ROMERO, D. M.; MEEDER, B.; KLEINBERG, J. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political , and Complex Contagion on Twitter. In: ACM INTERNATIONAL WORLD WIDE WEB CONFERENCE, 20., 2011. Hyderabad. **Proceedings**... Hyderabad: ACM, 2011.

SUH, B. et al. Want to be Retweeted? {Large} Scale Analytics on Factors Impacting Retweet in {Twitter} Network. In: INTERNATIONAL CONFERENCE ON SOCIAL COMPUTING (SocialCom), 2., 2010. Minneapolis. **Proceedings**... Minneapolis: IEEE, 2010.

TUMASJAN, A. et al. Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. **Social Science Computer Review**, v. 29, n. 4, p. 402-418, 2011.

VIVEIROS DE CASTRO, E. Cosmological Deixis and Amerindian Perspectivism. **The Journal of the Royal Anthropological Institute**, v. 4, n. 3, p. 469-488, 1998.

WENG, J. et al. Twitterrank: finding topic-sensitive influential twitterers. In: ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 3., 2010. New York. **Proceedings**... New York: ACM, 2010.

WU, S. et al. Who says what to whom on twitter. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 20., 2011. New York. **Proceedings**... New York: ACM, 2011.

YE, S.; WU, S. Measuring message propagation and social influence on Twitter. com. **Social Informatics: Lecture Notes in Computer Science**, v. 6430, p. 216-231, 2010.

O perspectivismo como fundamento para metodologias de análise das redes sociais: alguns resultados com o Twitter

Resumo

Neste artigo, apresentamos uma metodologia de análise das variáveis que influenciam a disseminação da informação na rede social Twitter. Analisamos dados de 595.240 mensagens produzidas por 261.757 usuários e encontramos três variáveis locais relacionadas ao grau de replicação das mensagens (retuíte): 1) o número de respostas recebidas (explicou em média 21% dos retuítes); 2) número de seguidores na sub-rede (17%) e 3) o número de tuítes com a *hashtag* (7%). Comparando variáveis *a priori* (globais) e *a posteriori* (locais), houve diferença significativa no número de seguidores, de mensagens (p<0,001) e também no número de amigos (p<0,01). O maior poder explicativo das variáveis locais sugere um caráter auto-organizado dos *Trending Topics*, apontando caminhos para pesquisas futuras com base na epistemologia perspectivista. Tais resultados também podem ajudar a delinear métodos de ranqueamento e segmentação de usuários.

Palavras-chave
Rede social. Twitter. Perspectivismo. Formador de opinião. Software.

El perspectivismo como base para métodos de análisis de redes sociales: algunos resultados con Twitter

Resumen

En este artículo presentamos una metodología de análisis de variables que influencian la diseminación de información en la red social Twitter. Analizamos datos de 595.240 mensajes producidas por 261.757 usuarios y encontramos tres variables locales relacionadas a los niveles de replicación de los mensajes: 1) el número de respuestas recibidas (explicando 21% de los mensajes replicados); 2) el número de seguidores en la subred (17%) y 3) el número de mensajes con la hashtag. La comparación entre las variables a priori (globales) y a posteriori (locales) enseña una diferencia significativa en el numero de seguidores, de mensajes (p<0,001) y también en el número de amigos (p<0,01). El poder explicativo más grande de las variables locales sugiere una propiedad de auto organización de los Trending Topics, indicando caminos para investigaciones futuras desde la epistemología del perspectivismo. Los resultados aquí presentados también pueden apoyar el diseño de métodos de clasificación y segmentación de usuarios.

Palabras claves
Red social. Twitter. Perspectivismo. Formador de opinión. Software.

Received on:
26 April 2012

Accepted on:
25 October 2012

www.e-compos.org.br
| E-ISSN 1808-2599 |

Expediente

A revista E-Compós é a publicação científica em formato eletrônico da Associação Nacional dos Programas de Pós-Graduação em Comunicação (Compós). Lançada em 2004, tem como principal finalidade difundir a produção acadêmica de pesquisadores da área de Comunicação, inseridos em instituições do Brasil e do exterior.

E-COMPÓS | www.e-compos.org.br | E-ISSN 1808-2599

Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação.
E-compós, Brasília, v.15, n.3, set./dez. 2012.
A identificação das edições, a partir de 2008, passa a ser volume anual com três números.