Measuring node importance on Twitter microblogging

Leila Weitzel
Federal University of Pará
Marabá, Pará, 68501-970, Brazil
+559421017114
lmartins@ufpa.br

Paulo Quaresma
University of Évora
Évora, 7000, Portugal
pq@uevora.pt

José Palazzo M. de Oliveira
Federal University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, 91501-970, Brazil
palazzo@inf.ufrgs.br

ABSTRACT

Social Networks (SN) are created whenever people interact with other people in online social networks, such as Twitter, Google+, Facebook and etc. Twitter is a social networking and micro-blogging service; it creates several new interesting social network structures. In this sense, our main goal is to investigate the power of retweet mechanism. The findings suggest that relations of "friendship" at Twitter are important but not enough. Still, the centrality measures of a node importance do not show how important users are. We uncovered some other principles that must be studied like, homophily phenomenon, the tendency of individuals to associate and bond with similar others.

Categories and Subject Descriptors

E.1 [**Data Structures**]: Graphs and network; G.2.2 [Graph Theory]: graph labeling.

General Terms

Measurement, Human Factors, Verification.

Keywords

Social Network Analysis, Twitter, Retweet, Node Importance.

1. INTRODUCTION

The recent proliferation of web applications and mobile devices has made online Social Network - SN more accessible than ever before. People connect with each other beyond geographical and timeline barriers, diminishing the constraints of physical boundaries in creating new ties [1].

The recent proliferation of Internet social media applications and mobile devices has made social connections more accessible than ever before. In the last few years the number of users of online social networks like Facebook, MySpace and Twitter gained considerable popularity and grown at an unprecedented rate [14]. Twitter is a social networking and micro-blogging service. Twitter allows users to communicate and stays connected through the exchange of short messages, called tweets. These posts are brief (up to 140) and can be written or received with a variety of computing devices, including cell phones.

Twitter creates several interesting social network structures. The most obvious network is the one created by the "follows" and "is followed by" relationships without approval, these create a different type of ties, where the directionality of tie is important (i.e. who is following whom) [12]. When a user posts a message, if other users like it, they repost it (or "Retweet" - RT), and a large number of users can be potentially reached by a particular message. Based on this context, we looked at the problem through two perspectives: first, studying topological structure of user´s RT alter and ego-network, second, ranking nodes based on strength of RT ties. In particular, we investigate the influence of "retweeting" mechanism in health context.

The outline of this paper is as follows: Section 2 presents the background of the research in the context of social network analysis; Section 3 we explain the data extraction technique and network modelling approach and data analysis; Section 4 explain the methodological approach; Section 5 we discuss the results and future works and Section 6 we present the acknowledgment, and the last Section the references.

2. RELATED WORK

Human beings have been part of Social Networks - SN since our earliest days. We are born and live in a world of connections. The SNs are created whenever people interact with other people. These ties can characterize any type of relationship, friendship, authorship, etc. For further details see "Social Network Analysis: Methods and Applications", by Wasserman and Faust, the most usually used reference book [26].

One common type of social analysis is the identification of communities of users with similar interests, and within such communities the identification of the most "influential" users. Efforts have been made to measuring the influence and ranking users by both their importance as hubs within their community and by the quality and topical relevance of their post. Some of these efforts are: [2, 3, 5–9, 11, 13, 15, 17, 18, 21, 23, 24, 27–29]. Most of these researches are based on: follower, tweet and mention count, co-follower rate (ratio between follower and following), frequency of tweets/updates, who your followers follow, topical authorities. Centrality measures such as Indegree/Outdegree, Eigen Vector, Betweenness, Closeness, PageRank [20] and others have been used to evaluate node importance too.

Each one of this metrics evidences a class of issue. For instance, Betweenness Centrality represents a node that occurs in many shortest paths among other nodes; this node is called "gatekeeper" between groups node. Closeness Centrality is the inverse of Average Distance (geodesic distance). Closeness reveals how long it takes information to spread from one node to others. Eigen Centrality measures take into account Hub-centrality (out links) and Authority-Centrality (in links). According Bonacich [4], "Eigenvector Centrality can also be seen as a weighted sum of not only direct connections but indirect connections of every length, thus, it takes into account the entire pattern in the network.

WIMS'12, June 13-15, 2012 Craiova, Romania

These measures are especially sensitive to situations in which a high degree position is connected to many low degree or vice-versa." Nevertheless, sometimes we must take node importance into full consideration based on several criteria that incorporate more global information. Evaluating node importance with a single metric can be considered incomplete and limited as it couldn't capture the specific differences among nodes

The "follower" concept, in Twitter perspective, represents the user who is following you. The "following" concept represents the user who you follow. Unlike most other online social networking sites such as Facebook, Google+, and etc social relationships are as binary: two people are either "friends" or they are not. The following relationship on Twitter is not a mutual relationship. Any user can follow you and you do not have to follow back. Relationships that are reciprocated on Twitter are different and perhaps stronger (at Twitter) than those that are not, and they are called "friendships". Twitter users follow someone, mostly because they are interested in the topics the user publishes in tweets, and they follow back because they find they share similar topic interest.

According to Macpherson's approach [16], homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. Homophily suggests that people with similar backgrounds with regard to their socio-demographic, behavioral and intrapersonal and others characteristics tend to established ties. The probability of a newly tie is higher among individual who are similar to each other.

In taking Macpherson's approach [16], we considered Kwak et al. [15] working paper and extended friendship similarity to "retweeting" mechanism. We regarded that RT mechanism may work to increase ego-network in this way: a user A post an interesting "Tweet", you like this post and then forwarding to your ego-network. Your followers or other user from your alter-network discover and maybe follow the user who "Tweet", or perchance, they forward to their own ego-network. These can potentially increasing the size and reach of user´s "Tweet" ego-network.

Kwak et al. [21] performed an extensive study about Twitter follower-following topology analysis. He crawled 41.7 million user profiles and 1.47 billion social relations, and constructed a directed network based on the follower/following relationship and analyzed its basic characteristics. They noticed that there are a few users with more than a million followers, and all of them are celebrities or mass media and most of them do not follow back. The majority of users who have fewer than 10 followers never tweet or did once. Still, according to Kwak et al. [21], Twitter shows a low level of reciprocity (22.1%) and 67.6% of user are not followed by any of their following.

3. WEIGHTED TIES APPROACH

In this section we present the background of the research in the context of SN analysis, in particular the Twitter case of study and its main aspects which are essential for this work.

We compute the basic statistic of data sample. The Figure 1 plots the frequency per source user over the spam of our dataset. Source User - $\boldsymbol{u_i}$ is the user that reposts the tweet, and Target User $\boldsymbol{u_j}$, is the object of study, the user who had his tweet replayed.

In addition, we suppressed the Twitter "screen name". The three major percentages were: UC99 at 34%, UC2 at 37%, UC5 at 54%.

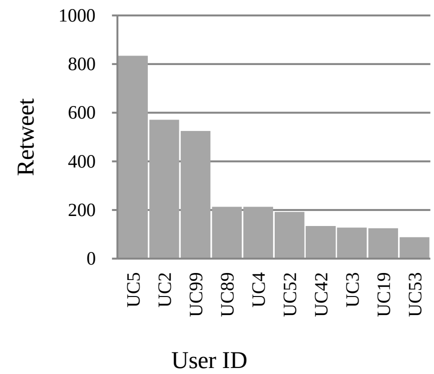

Figure 1. RT Freqeuncy per source user

The measures of network-level can be seen in Table 1. The Density is low, i.e., do not have a dense "in" and "out" ties to one another. In contrast, a higher density score reflects more ties, which is generally interpreted as more coordinate network with more opportunities for sharing of information among nodes. This indicates that maybe exist potentials relationships. Conversely, Fragmentation shows that nodes are highly connected, as pointed out in Table 1 by Isolate Count Measure. The Transitivity represents the idea: "if friends of my friends are my friends", it is not quite the reality at RT network. That can be confirmed by low value of transitivity measure.

Table 1. RT graph-level measures

Measures [min =0; max =1]	Values
Density	0.0009
Fragmentation	0.2567
Efficiency (the degree to which each component in a network contains the minimum links possible to keep it connected.)	0.063
Isolate Count (The number of isolate nodes in a unimodel network)	0.000
Transitivity (The percentage of link pairs {(i,j), (j,k)} in the network such that (i,k) is also a link in the network.)	0.070

We have crawled with NodeXL [19] about 152 users in accordance with link "how to follow" and later "browse interests", and then we searched for topic "health" from Twitter Microblogging during March 2011. Afterward, we selected 100 users that have a website or a blog associated to health subject. From each of 100 initial users, we extracted about 200 RT per user. Kwak et al. [21] demonstrated that the median of tweets/user stays between 100 and 1000, thus, the ratio of RT/user indicates that the size sample is suitable, since not all tweets are replayed.

The Figure 2 illustrates the Retweet Network (RT-Network). The RT-Network was modeled as a direct graph $\mathbf{G_{RT}} = \mathbf{(V, A)}$ where each node $u \in \mathbf{V}$ (totalling 1237 nodes) represents the users and each edge $a_k = (u_i, u_j) \in \mathbf{A}$ represents RT relationship (totalling 1409 edges), i.e., an edge a_k from u_i (source) to u_j (target) stands that user u_i "Retweet" user u_j.

These edges a_k are weighted according the equation 1. Thus, let the weight w_{ak} be defined by:

$$w_{a_k} = \beta + \alpha \qquad \text{Equation (1)}$$

Where $\beta = \frac{\sum \mathrm{RT}}{RT_{max}}$ with $\sum RT$ is the retweet count of target user, RT_{max} is the total number of reweet of source user. This parameter is a normalized retweet.

The parameter α represents the relationships such as:

- *following,*
- *follower,*
- *who are reciprocally connected and*
- *no relationships – (where follower or following are absent).*

The parameter α was calculated in accordance with pie chart showed in Figure 2. Thus, α = 0.64 for absent relationships, α = 0.15 for both (follower or following), α = 0.14 for following and α = 0.07 for follower relationships. This parameter is a sort of discount rate. Furthermore, α intends to discount the weight of the follow phenomenon, since many celebrities and mass media have hundreds of thousands of followers; it defines smaller values to relationships that are "follower" or "following" (or both) and higher values when there is no relationship between users.

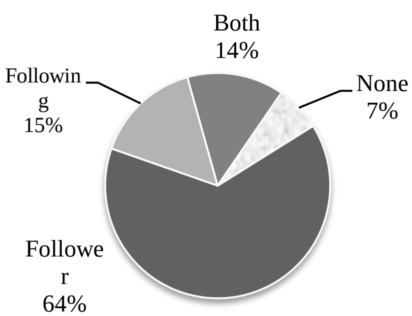

Figure 2. Pie chart of RT relationships.

As we expected, the major percentage belongs to "follower" relationship, in agreement with the main idea of Twitter.

We use ORA, an analysis tool developed by CASOS - Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University to develop the network model. The Figure 3 and 4 illustrates the retweet network RT-network.

The Girvan and Newman´s grouping algorithm [10] is based on betweenness centrality measure. The algorithm is state as follow: calculate the betweenness for all edges in the network. Afterward, the edges with highest betweenness are removed, then recalculate the betweenness for all edges affected by the removal; repeat this procedure until no edges remain.

It must be stressed that, RT are marked with characters RT or via @ + "screenname". Therefore, we extracted either both replay tweets and mention.

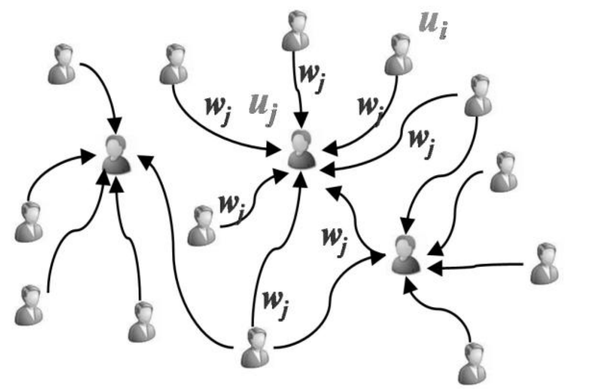

Figure 3. RT-Network

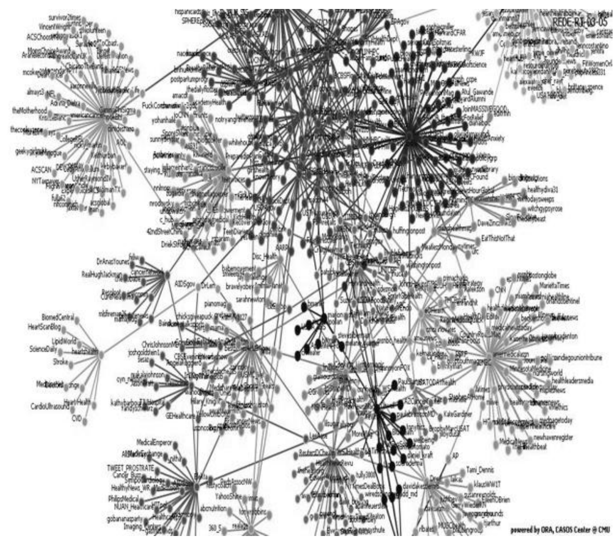

Figure 4. Retweet network clustered by Girvan and Newman Grouping algorithm [3], the modularity values measures the degree to which grouping has found community structure.

3.1 Dataset Analysis

The sample of RT has a mean of 3.0 per user target u_j and standard deviation of 15.23 per user target u_j. The Twitter "screen names" were suppressed. The major's frequencies of RT/user target are: UC99 = 34%, UC2 = 37%, UC5 = 54%.

Approximately 65% had only one RT, the remaining was split between 2 and 523 retweets. Approximately 96% have 0.00 of Betweenness Centrality, this means that, exist only a few nodes that occur in shortest path. There are 1236 strongly nodes connected, and 8 weakly connected, i.e., if they are removed consequently breaks the remainder of the nodes into many small, disconnected clusters, these nodes are necessary to keep the network connected.

4. RANKING NODE METHODOLOGY

One common type of social analysis is the identification of communities of users with similar interests, and within such communities the identification of the most "influential" users. A simple notion of influence is the number of connections, and influential users act as hubs within their community.

The Centralities Measures of a node importance proposed by Albert & Barabási [1] are only based on: ties (ingoing and outgoing counting edges) and topological structure of network. Hence, "edges counts" doesn't show how important users are. It can be treated only as the "popularity" measure, Kwak et al. [21].

The measure the node importance has become a worth studying issue in the field of Complex Network Analysis - CNA. Several works are based on: follower count, co-follower rate (ratio between follower and following), frequency of tweets/updates, who your followers follow, and etc. Others are based on centralities measures such as: Degree, Betweenness, Closeness, Eigenvector and PageRank [20] each of them is proposed in order to tackle with a class of issue.

Betweenness Centrality represents a node that occurs in many shortest paths among other nodes; this node is called "gatekeeper" between groups node. Closeness Centrality is the inverse of Average Distance (geodesic distance); it reveals how long it takes information to spread from one node to others. Eigen Vector Centrality takes into account out links and in links. Eigen Vector Centrality can also be seen as a weighted sum of not only direct connections but indirect connections of every length [4]. All these measures are especially sensitive to situations in which a high degree position is connected to many low degree or vice-versa.

Nevertheless, sometimes, we must take node importance into full consideration based on several criterions that incorporate more global information. Thus, evaluating node importance with a single metric can be considered incomplete and limited as it couldn't capture the specific differences among nodes [25].

Hence, we propose using F-measure in order to estimate node importance. The F-measure is generally accepted at Information Retrieval as evaluation performance methods. It is by far, the most widely used and first introduced by van Rijsbergen [22].

F-measure (**F**) combines Recall (**R**) and Precision (**P**) in the following form:

$$F(R,P) = \frac{(\beta^2+1)P * R}{\beta^2 P + R} = \frac{1+\beta^2}{\frac{\beta^2}{R}+\frac{1}{P}} \qquad \text{Equation (2)}$$

where $(0 \leq \beta \leq +\infty)$

Where β is a parameter that controls a balance between **P** and **R**. When β = 1 F comes to equivalent to the harmonic mean of **P** and **R**. If β > 1, F becomes more recall-oriented and if β < 1, it becomes more precision oriented $\mathbf{F_0 = P}$.

Thus, our methodological approach is based on combining standard metrics with adjustable weighted parameters, considering not only the topological importance of a node, but also the strength of ties of retweet expressed in Equation 1. The modified F-measure, named Rank is a linear combination of metrics with associated weight defined by:

$$\text{Rank} = \frac{\sum_{k=1}^{m} w_k}{\sum_{k=1}^{m} \frac{w_k}{x_k}} \qquad \text{Equation (3)}$$

The $(\sum_{k=1}^{m} w_k) = 1 \iff (\delta + \beta + \theta + \gamma) = 1$ is the weighted parameter.

The x_k where k =1..4 is a set of four measures:

- **BC** is Betweenness Centrality,
- **CC** is Closeness Centrality,
- **EC** is Eigen Vector Centrality, and
- **PRANK** is the PageRank [8].

The first hypothesis is all of parameters have same value (line two in Table 2): δ = 0.25; β = 0.25; θ = 0.25; γ = 0.25, the Equal weighted approach and afterward is weighted according remaining lines of Table 2.

Table 2. Weighted parameter

Measure / Weight	δ	β	θ	γ
Equal weighted = W(EQUAL)	0.25	0.25	0.25	0.25
BC weighted = W(BC)	**0.7**	0.1	0.1	0.1
CC weighted = W(CC)	0.1	**0.7**	0.1	0.1
EC weighted = W(EC)	0.1	0.1	**0.7**	0.1
Prank weighted = W(PRANK)	0.1	0.1	0.1	**0.7**

The Table 2 shows only the top 10 ranked nodes using our approach.

In order to gain insight we compute the top 20 recurring nodes based on the sum of its position for the four measures, W(BC), W(CC), W(EC), W(PRANK). The results are displayed in Figure 5.

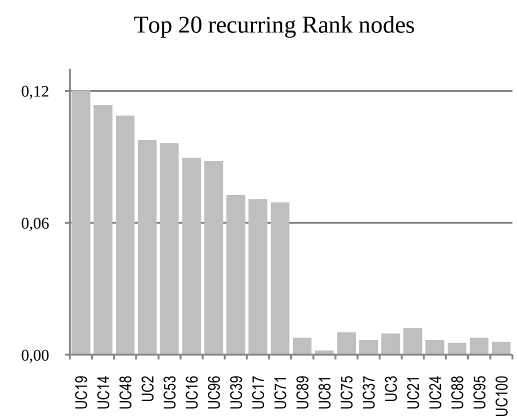

Figure 5. Bar chart of recurring top 20 nodes (target user)

Table 3. Top 10 ranked nodes

W(BC)	**W(CC)**	**W (EC)**	**W (PRANK)**	**BC**	**CC**	**EV**	**PRANK**
UC2	UC43	UC19	UC19	UC2	UC2	UC2	UC57
UC19	UC4	UC48	UC53	UC624	UC624	UC13	UC4
UC14	UC29	UC53	UC48	UC43	UC43	UC38	UC2
UC96	UC25	UC21	UC18	UC4	UC4	UC46	UC89
UC48	UC56	UC115	UC14	UC89	UC89	UC14	UC99
UC39	UC32	UC14	UC21	UC3	UC3	UC134	UC43
UC89	UC40	UC102	UC35	UC776	UC776	UC1175	UC67
UC71	UC100	UC119	UC16	UC19	UC19	UC19	UC624
UC17	UC24	UC114	UC110	UC1145	UC1145	UC48	UC48
UC75	UC74	UC113	UC102	UC48	UC48	UC53	UC40

The Table 3 also shows the top 10 nodes ranked with our methodology (column 1-4 in Table 3) and the top 10 with (column 5-8 in Table 3) without F-measure approach.

Table 4. Top 10 ranked nodes W(EQUAL)

user ID	Followed	Followers	Tweets	RT normalized	Acceptance Rate
UC19	78	27599	797	124	16%
UC14	31	116129	511	0	0%
UC48	9	90389	296	32	11%
UC2	1419	78480	1884	571	30%
UC53	92	88600	599	0	0%
UC16	28	6900	226	0	0%
UC96	1095	174651	2217	0	0%
UC39	269	111390	1341	24	2%
UC17	82	1259595	414	0	0%
UC71	95	4789	524	27	5%
UC89	775	3064	4826	213	4%
UC81	1755	89064	1128	17	2%
UC75	2180	101023	5040	57	1%
UC37	52	17400	1258	5	0%
UC3	134697	136962	1893	127	7%
UC21	585	4913	2995	0	0%
UC24	1618	164211	3168	9	0%
UC88	4143	3984	2900	36	1%
UC95	225	180109	1045	47	4%
UC100	168	85633	2367	0	0%

Table 4 illustrates some main properties of data sample. The first seven records represent American Health Agencies. The last three are mass media. It must be stressed that column RT normalized represents the parameter β in equation 1. We also compute the retweet acceptance rate (column five in Table 4), which mean is equal to 4 percent, i.e, not all tweets are reposted, hence, users employ retwet mechanism with prudence and moderation.

The Figura 6 displays the scatter plot of values for the two variables, followers and retweet of data sample. A scatter plot is used when a variable exists that is under the control of the experimenter. Scatter plots are normally used to analyze patterns in bivariate data, these patterns are described in terms of linearity, slope, and strength, we noticed that there is a: linear, zero slope, strong dependencies. Linearity refers to whether a data pattern is linear (straight) or nonlinear (curved). Slope refers to the direction of change in variable Y when variable X gets bigger. Hence zero slope means that variable follower is parallel to X axis. Strength refers to the degree of “scatter” in the plot. If the dots are widely spread, the relationship between variables is weak. If the dots are concentrated around a line, the relationship is strong, this means that follower variable is strongly associated to retweet variable.

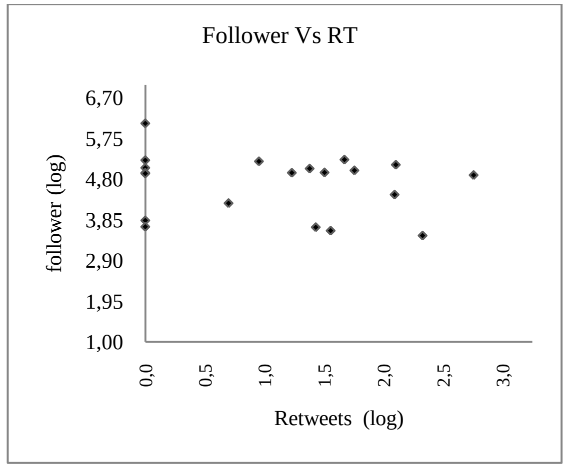

Figure 6. log X log Scatter plot of data sample

Table 5. Correlation Matrix

	PageRank	Eigen vector	Betweenness	Closeness	W(Equal)
PageRank	1				
Eigenvector Centrality	0,69	1			
Betweenness	0,50	0,45	1		
Closeness	0,37	0,35	0,59	1,0	
W(equal)	0,55	0,51	0,96	0,67	1

The Table 5 show the correlation matrix of target measures, PageRank, Eigen Vector, Betweenness Centrality, Closeness Centrality and our proposed measure W(Equal). We noticed that W(equal) variable is strongly associated to Betweenness centrality measure.

Kitsak et al [10] showed that the correlation between Degree and Betweenness Centrality of nodes is much weaker in fractal network models compared to non-fractal models. They also found that in fractal networks even small degree nodes can have very large betweenness centrality while in non fractal networks large Betweenness Centrality is mainly attributed to large degree nodes. This finding quite interesting, and may be is an evidence, that should be explored.

5. DISCUSSIONS

Our goal was mostly to analyze the power of retweeting. As a case of study, we chose Twitter microblogging service. Twitter coordinates conversation based on tweet and retweet mechanisms. The reweet mechanism allows us to design a new topological structure of social network, used as a tool to infer the level of online social interactions.

We also presented a new method to measure node importance which is based on control weighted parameters as it appears in F-measure.

The experimental results offer an important insight of the relationships among Twitter users. The findings suggest that relations of "friendship" or follows are important but not enough to find out how important nodes are. Many users judge as sign of politeness to follow back a new user follower, it is considered "good manner", i.e., the "Twitter's etiquette". Then, it appears that follower counting is not to be trusted when trying to infer a user's influence.

The study also gives us a clear understanding of the how measure selection can affect the rank. Choose the most appropriate measure depends on what we want to represent; for example, in/out degree, Eigen-Vector and even PageRank operate look alike "edges counts" as the "popularity" measures. Conversely, closeness and betweenness centrality measures specify the key position that a node occupies in a graph.

The results also shown that centrality measures associated with our weighted ties approach controls suitably the node rank. Moreover, we have observed that in Twitter community, trust plays an important role in spreading information; it motivates a user to reply messages to other users, thus, the culture of "Retweeting" demonstrate the potential to reach trust for dissemination of information.

As stated before, twitter social networkers communicate with each other by posting tweets allowing for public interactive dialogue; We believe that, Twitter's communicative structure is determined by two overlapping and interdependent networks – one based on follower-following relationships, the most obviously; and one relatively short-term and emergent, based on shared interest in a topic or event, often coordinated by a common hashtag.

For Twitter users, following and posting to a hashtag conversation makes it possible for them to communicate with a community of interest around the hashtag topic without needing to go through the process of establishing a mutual follower/following relationship with all or any of the other participants. Using a Hashtag can be seen as an effort to address a community of users following and discussing a specific topic. Therefore, follower/following network must be understood as separate from this shared communication networks. Thus, in Twitter sphere of influence there are several others networks layers, in this sense, for future work we are motivated to explore these short-term networks.

6. ACKNOWLEDGMENTS

Leila Weitzel and José Palazzo M. de Oliveira would thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazilian National Research Council) for the partial financial support of this work. 201931/2010-9.

Leila Weitzel also would thank you Federal University of Pará for the partial financial support.

7. REFERENCES

[1] Albert, R. and Barabási, A.L. 2002. Statistical mechanics of complex networks. *Reviews of modern physics*. 74, 1 (2002), 47.

[2] Balkundi, P. and Kilduff, M. 2005. The ties that lead: A social network approach to leadership. *The Leadership Quarterly*. 16, 6 (Dec. 2005), 941-961.

[3] Bar-Ilan, J. and Peritz, B.C. 2009. A method for measuring the evolution of a topic on the Web: The case of "informetrics." *Journal of the American Society for Information Science and Technology*. 60, 9 (2009), 1730-1740.

[4] Bonacich, P. 2007. Some unique properties of eigenvector centrality. *Social Networks*. 29, 4 (Oct. 2007), 555-564.

[5] Bongwon Suh et al. 2010. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *Social Computing (SocialCom), 2010 IEEE Second International Conference on* (2010), 177-184.

[6] Boyd, D. et al. 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *Hawaii International Conference on System Sciences* (Los Alamitos, CA, USA, 2010), 1–10.

[7] Cha, M. et al. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. *International Conference on Weblogs and Social Media* (2010).

[8] Detecting Important Nodes to Community Structure Using the Spectrum of the Graph: 2010. *http://arxiv.org/abs/1101.1703*. Accessed: 2011-03-31.

[9] Gayo-Avello, D. 2010. Nepotistic relationships in twitter and their impact on rank prestige algorithms. *Arxiv preprint arXiv:1004.0816*. (2010).

[10] Girvan, M. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*. 99, 12 (Jun. 2002), 7821-7826.

[11] Gruhl, D. et al. 2004. Information diffusion through blogspace. *Proceedings of the 13th international conference on World Wide Web* (New York, NY, USA, 2004), 491-501.

[12] Hansen, D. et al. 2011. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann.

[13] How to Become Famous in the Microblog World: 2010. *http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1505/1887*.

[14] Kim, W. et al. 2010. On social Web sites. *Information Systems*. 35, 2 (Apr. 2010), 215-236.

[15] Kwak, H. et al. 2010. What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web* (Raleigh, North Carolina, USA, 2010), 591-600.

[16] McPherson, M. et al. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*. 27, (2001), 415–444.

[17] Nagarajan, M. et al. 2010. A Qualitative Examination of Topical Tweet and Retweet Practices. *ICWSM 2010* (Washington, DC, 2010).

[18] Nagle, F. and Singh, L. 2009. Can Friends Be Trusted? Exploring Privacy in Online Social Networks. *2009 International Conference on Advances in Social Network Analysis and Mining* (Athens, Greece, Jul. 2009), 312-315.

[19] NodeXL: a free and open network overview, discovery and exploration add-in for Excel: 2007. *http://www.smrfoundation.org*. Accessed: 2011-05-19.

[20] Page, L. et al. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Stanford InfoLab.

[21] Pal, A. and Counts, S. 2011. Identifying topical authorities in microblogs. *Proceedings of the fourth ACM international conference on Web search and data mining* (Hong Kong, China, 2011), 45-54.

[22] Rijsbergen, C.J. van 1979. *Information retrieval*. Butterworths.

[23] Romero, D.M. et al. 2011. Influence and passivity in social media. *Proceedings of the 20th international conference companion on World wide web - WWW '11* (Hyderabad, India, 2011), 113.

[24] Sousa, D. et al. 2010. Characterization of the twitter @replies network. *Proceedings of the 2nd international workshop on Search and mining user-generated contents - SMUC '10* (Toronto, ON, Canada, 2010), 63.

[25] Wang, Y. and Liu, Z. 2007. Automatic detecting indicators for quality of health information on the Web. *International Journal of Medical Informatics*. 76, 8 (Aug. 2007), 575-582.

[26] Wasserman, S. 1999. *Social network analysis : methods and applications.* Cambridge University Press.

[27] Welch, M.J. et al. 2011. Topical semantics of twitter links. *Proceedings of the fourth ACM international conference on Web search and data mining* (Hong Kong, China, 2011), 327-336.

[28] Yamaguchi, Y. et al. 2010. TURank: Twitter User Ranking Based on User-Tweet Graph Analysis. *Web Information Systems Engineering – WISE 2010*. Springer Berlin / Heidelberg. 240-253.

[29] Ye, S. and Wu, S. 2010. Measuring Message Propagation and Social Influence on Twitter.com. *Social Informatics*. Springer Berlin / Heidelberg. 216-231.

[30]