Degrees of separation on a dynamic social network

André Domingos
ISEL, Poly Inst of Lisbon
A24503@alunos.isel.pt

Hugo Ferreira
ISEL, Poly Inst of Lisbon
A32983@alunos.isel.pt

Pedro Rijo
IST, Tech Univ of Lisbon
INESC-ID at Lisbon
pedro.rijo@ist.utl.pt

Cátia Vaz
ISEL, Poly Inst of Lisbon
INESC-ID at Lisbon
cvaz@cc.isel.pt

Alexandre P Francisco
IST, Tech Univ of Lisbon
INESC-ID at Lisbon
aplf@ist.utl.pt

ABSTRACT

Social networks are nowadays the most prominent online environments for communication and interaction among individuals, that allow people to communicate more often. In fact, in recent years, most studies have concluded that we are more close now, with the number of degrees of separation becoming smaller for most social networks. These networks are nevertheless continuously changing and evolving, being in fact highly dynamic, with many relations among users not being always active, or even becoming dead links. In this paper, we describe the reconstruction of interaction networks from Twitter data, along a given period of time. We report detailed statistics for these networks and results about observed degrees of separation. Our results point out that, taking into account just real time interactions, the degrees of separation are higher than those reported for traditional contact networks.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Experimentation, Measurement

Keywords

Degrees of Separation, Social Network Analysis, Twitter

1. INTRODUCTION

The massive number of online social interactions that are made at each second through social networks can be seen as an opportunity to study people communication patterns, propagation of ideas, viral marketing, among other phenomena. Microblogs, a relatively new phenomena in online social networking, provide a new communication channel for people to broadcast information in a different way when compared to other online social platforms. Microblogging refers to the activity where users broadcast brief text updates about small little things happening in daily life, such as working activities, news on the media, and experiences [19].

Within the interaction models of these networks, specially within microblogs, people tend to communicate more often. Moreover, studies on degrees of separation within social networks have emerged and most of them have concluded that social network users are more close now, since the number of degrees of separation is becoming smaller [16, 1].

Since networks are a communication channel of excellence for the diffusion of information, one of the applications in studying the average number of intermediaries between two people, *i.e.* the degrees of separation, is in influence propagation and diffusion information models. However, as pointed out by Rogers [21], the four main elements that influence the spread of a new idea are: innovation, communication channels, time, and a social system. Therefore, the evolution of the networks over time, namely the dynamics underlying the interaction of network users through time, can influence the average number of intermediaries between users as well as influence its change thought time. Even more, we may observe the presence of *weak ties* [15] that hardly correspond to active interaction links and that would probably not be continued with the evolution of the network over time. In fact, they may even became *dead ties*.

Therefore, to better understand the dynamic structure of interaction patterns underlying social networks, we propose to study the degrees of separation on a social network considering only user interactions that occur in the observation period, ignoring pre-defined relations and contact lists. With this approach we mitigate the presence of *weak ties* and we believe to be more close of comprehending the dynamic nature of social networks.

Our work explores the Twitter network, which being a microblogging service, is not only used as a social network but also as medium for information diffusion, since the communication model underlying Twitter captures the publish-subscribe messaging pattern. Namely, Twitter users declare other Twitter users that they are interested in following, in which case they get notified when the followed user posts a new message (tweet). Unlike most social online networks, the user who is being followed by another need not to follow back. To capture the dynamic nature of Twitter network, we rely our study on tweets exchanged during an observation

Eleventh Workshop on Mining and Learning with Graphs. Chicago, Illinois, USA

period. In fact, we focus on *active ties*, *i.e.*, links between pairs of users that exist whenever one of the users explicitly refers to, or mentions, the other user. By relying on this approach, we obtain significant differences on several well known network metrics when comparing with previous studies of Twitter network [16]. In particular, we observe a higher number of degrees of separation.

2. PROBLEM

Several studies, in recent years, have discussed the relevant properties of several networks underlying well known online social systems, providing sometimes diverging results. In what concerns the number of degrees of separation, by taking a quick look on some of those studies, we observe that later studies [16, 1] conclude that the value is smaller than in previous studies [20, 17]. We should however be careful about this conclusion. As discussed in previous section, we are accumulating data over the years and we may be considering interactions and links that are no longer active. Moreover, as is the case with recent studies about Twitter [16], contact networks are usually considered instead of interaction networks, which are also subject to the existence of dead links. Thus, how interaction and communication networks have changed over time? And, moreover, what may we conclude by observing such networks for a limited period of time?

These questions have been addressed by some well known studies [20, 17] and, given Twitter public nature and data availability, we study interaction networks underlying Twitter and we compare our results with other social networks. We focus on the distance distribution for these interaction networks and we try to answer to the question: how different is the number of degrees of separation, which is given by the average distance minus one, compared with other interaction and communication networks? Note that in this case the average distance considers only reachable pairs, being also known as *average connected distance* [10].

3. METHODS AND RESULTS

We collected data through Twitter streaming API by tracking users for four months, obtaining about 325 million of tweets produced by 19 million of users. Then we reconstruct underlying interaction networks and we analyze them using the Webgraph framework and related tools [7, 6]. We describe here our reconstruction approach and underlying principles, we present our results concerning the number of degrees of separation and other related metrics, and we compare them with results from published studies.

3.1 Twitter social network

Online platforms have been used for information diffusion as well as for personal communication. Diffusion actions are usually implemented in the underlying system in the form of sharing and/or reposting actions, which allow users to broadcast information for their own contacts. Personal communication, or conversation when bidirectional communication occurs, implies interaction between people that mention, implicitly or explicitly, the recipient of the message. Both information diffusion and personal communication are possible, and are present, in Twitter.

Twitter allows *follower-following* relationships, *i.e.* users may follow others and/or might be followed by others. When a user posts a message, Twitter broadcasts it to its followers. Posts are known as *tweets* and can be classified as: *normal tweets* or just *tweets*, being messages with fewer than 140 characters; *mentions* which are tweets addressing a specific user; *replies* which are tweets addressing another user as a reply to another tweet; and *retweets* which are reposts by a user of tweets posted by other users. When a user is retweeted, replied, or just mentioned in a tweet, it appears in the tweet as @ followed by the user id. Besides mentioning the user that made the initial tweet, *retweets* are also marked with the string "RT". When using the Twitter API for accessing data, all this information is provided as part of a JSON document for each tweet.

Table 1: The details for all data collected, with 22,343,103 users and 123,621,864 pairs of users with at least one mention between them.

Data	Total	Authors	Mentioned
Tweets	325,333,833	19,558,917	10,250,087
Mentions	322,204,140	19,548,896	10,250,087
Replies	86,004,326	7,579,946	6,033,393
Retweets	105,361,317	15,895,165	5,135,943

The different kinds of tweets and the publish-subscribe messaging model underlying Twitter make its interaction network much more richer and dynamic than other interaction networks, that privilege communication among users in contact lists, usually established *a priori* as friendship relations. In Twitter, given the possibility of retweeting others' tweets, interactions among users become strongly dynamic. Moreover, users are always able to mention or reply directly to a user, even if that user is not in their contact list. In both cases, and specially with respect to retweeting, an user mention can be seen as a manifestation of influence as the tweet author is mentioning that user. We can clear discuss influence orientation, as it can be positive or negative, even though it will be a kind of influence.

In this paper we will use mentioning events to infer interaction networks. Note that this approach is rather different from follower-following based approaches, where a "static" network is studied instead, and where interactions among users are not being captured. We will study three different, but related, interaction networks. The *mention* network, the *influence* network, and the *conversation*. The first one considers a link from user a to user b whenever a mentions b, either in reply, retweet or a simple mention. The *influence* network is the transpose of the *mention* network and reflects influence among users as discussed above. Finally, the *conversation network* corresponds to the symmetric subgraph of previous networks.

3.2 Data collection

Our dataset was collected for four months, comprising 325,333,833 tweets authored by 19,558,917 users. Since our approach relies on following such many users, we developed a distributed application for this task that establishes several independent connections to the statuses/filter Twitter public streaming and stores all collected tweets in a large MongoDB instance. While collecting data, our application follows a set of users and extracts all mentioned users in collected tweets, adding new users to a following list. The data collecting process operates through that list giving priority

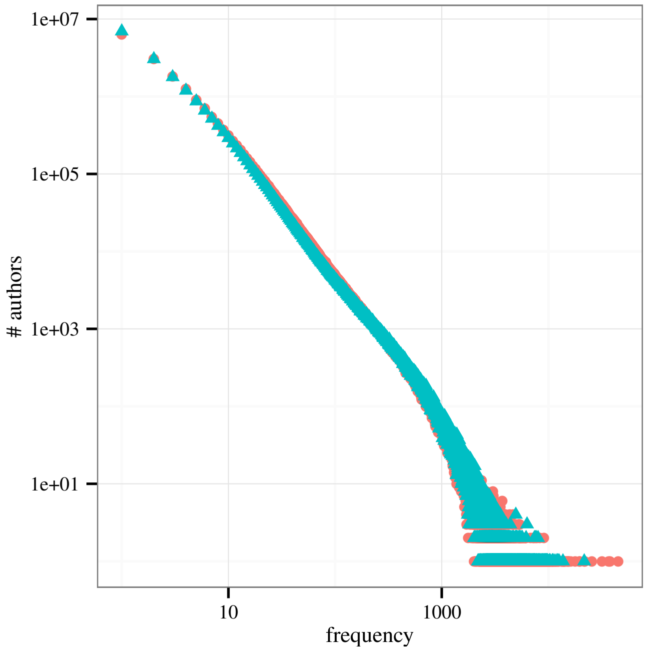

Figure 1: The distribution of the frequency of tweets (in red) and of tweets with mentions (in blue), per author.

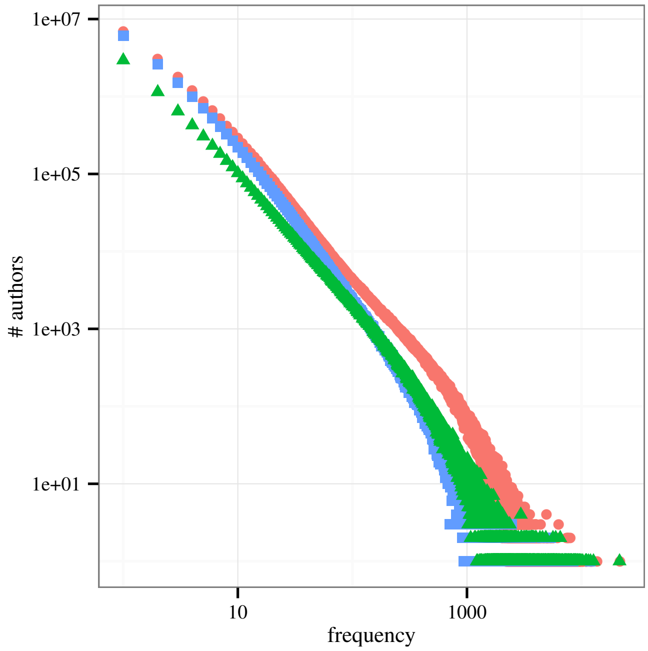

Figure 2: The distribution of the frequency of tweets (in red), of replies (in green), and of retweets (in blue), per author.

to more active authors. In fact we distribute the workload uniformly by simultaneously allocating 10% of processing capacity in each worker to more active users and scheduling less active users in a round-robin fashion within the remaining 90% workers capacity. The application can start with an initial list of users or with a set of topics to track.

Note that our approach can be seen as a breadth first search over Twitter network, where we identified 22,343,103 different users. This and other details are provided in Table 1. It is interesting that both the frequency of tweets and the frequency of tweets with mentions, per author, follow a similar distribution, as depicted in Figure 1, with 99.04% of tweets containing at least a mention to some user. Figure 2 depicts the frequency distribution per author for tweets, replies and retweets. As detailed in Table 1 we have that 40.61% and 32.70% of the tweets are retweets and replies, respectively. Figure 3 depicts the frequency distribution per user for tweets, replies and retweets containing mentions to that user. If we try to fit these distributions to a power law, we obtain for the distributions in Figure 2 the coefficients 3.97, 3.24 and 3.57, respectively. And, for the distributions in Figure 3, the coefficients 2.04, 3.10 and 2.03, respectively. The most interesting difference is observed for retweets distributions, as the per author frequency distribution (Figure 2) has a coefficient of 3.57 and the per retweeted user frequency distribution (Figure 3) has a coefficient of 2.04, providing and interesting insight on retweeting dynamic on Twitter.

3.3 Network construction

We start by identifying the pairs of users for each mention, namely the author and the mentioned user. In order to get some insight on the data, we analyzed mention frequency per pair of users, identifying also different kinds of mentions

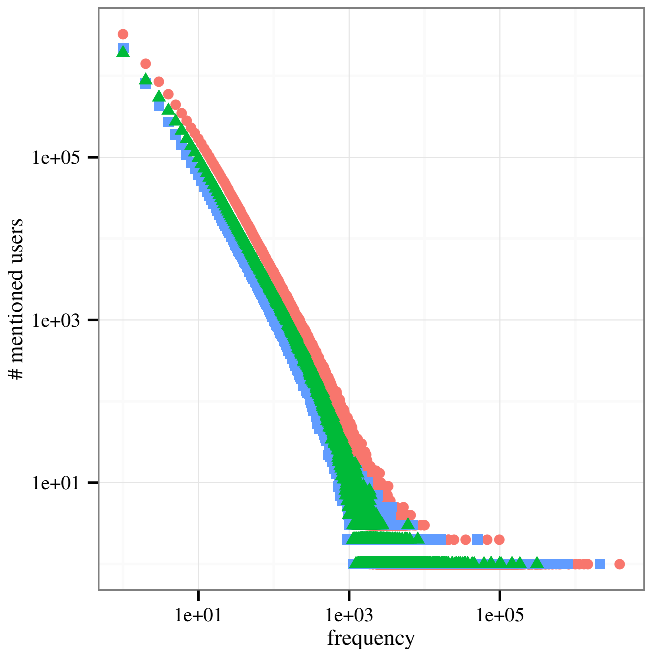

Figure 3: The distribution of the frequency of mentions (in red), of replies (in green), and of retweets (in blue), per mentioned user.

(retweet or reply), which distributions are provided in Figure 4. Our first network, the mention network, is obtained by adding each observed user as a vertex and a link per each observed mention (reply, retweet or just a simple mention). Note that this network is directed and unweighted. Although we can derive weights from mention frequency,

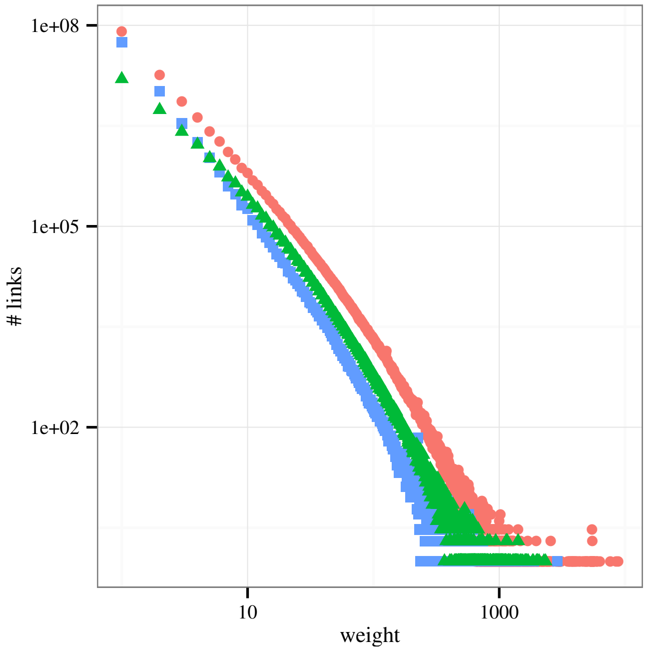

Figure 4: The distribution of weights for mention links (in red), for reply links (in green), and for retweet links (in blue), where the weight is the number of times that a given link was observed.

and we have analyzed it to some degree, we will skip that analysis in our discussion.

The influence network and the conversation network are easily obtained from the mention network. In the first case we transpose it and, in the second case, we keep only the links for which we have the reciprocal one. All three networks were compressed using the Webgraph framework [7].

3.4 Network characterization

We computed several statistics for each network, including compression ratio through network hierarchical clustering followed by vertex reordering [5] and the neighborhood function values, that allow us to compute several other statistics as provided in Table 2.

The neighborhood function values were computed with the HyperANF algorithm [6]. This algorithm, which is a diffusion-based algorithm, provides an approximation of the neighbourhood function subject to a relative standard deviation. This error depends on the number m of registers per counter and is less or equal than $1.06/\sqrt{m}$. Hence, with $m = 512$, we obtain a relative standard deviation for all statistics less than 5%. As noted by HyperANF authors, the relative error for the neighbourhood function becomes however an absolute error for the distance distribution. Therefore, as recommended by HyperANF authors, we run the algorithm several times and we estimated the standard error of our measurements with the Jacknife method [13].

From measurements in Table 2 and degree distributions in Figures 6 and 7, we observe that studied graphs have the common characteristics of scale-free small-world networks. We note nonetheless some interesting characteristics. Since our approach to collect data mimics a BFS, the giant component is almost 100% for both the mention and the influence

Table 2: The details for studied networks, where the mention network contains a link from u to v whenever u mentions v, the influence network is the transpose to the mention network, and the conversation network is the symmetric subgraph.

Networks	Mention	Influence	Conversation
Vertices	22,341,855	22,341,855	5,807,158
Links	121,705,381	121,705,381	28,031,968
Avg. outdg	5.45	5.45	4.78
Max. outdg	30,908	518,075	4,534
Min. outdg	0	0	0
Outdg. exp	3.49 (>268)	2.13 (>57)	3.70 (>222)
Avg. indg	5.45	5.45	4.78
Max. indg	518,075	30,908	4,534
Min. indg	0	0	0
Indg. exp	2.13 (>57)	3.49 (>268)	3.70 (>222)
Term. nodes	12.51%	54.28%	0.85%
Giant comp.	99.98%	99.98%	96.36%
Comp. ratio	73.6%	55.3%	55.1%
Avg. dist	6.64 ±0.04	6.63 ±0.03	7.67 ±0.05
Efect. dmt	7.71 ±0.05	7.73 ±0.05	8.95 ±0.05
Diameter	16	19	22
Harm. dmt	17.53 ±0.92	18.43 ±0.57	7.56 ±0.19
SPID	0.20 ±0.01	0.21 ±0.01	0.26 ±0.01

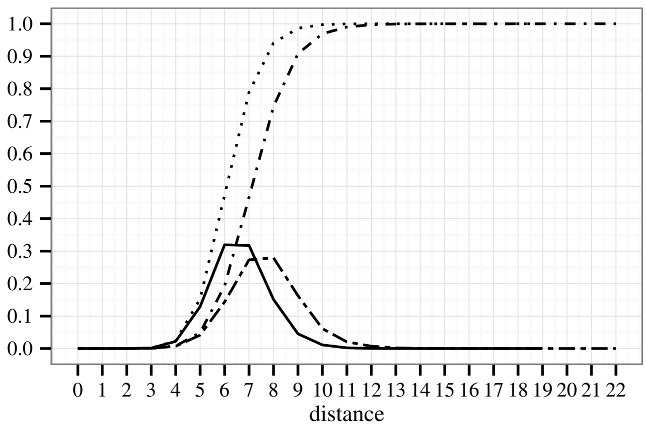

Figure 5: The distance probability mass function for the influence network (solid line) and for the conversation network (twodash line), and cumulative distribution functions (dotted and dotdash lines, respectively).

networks. Note also that, after hierarchical clustering followed by vertex reordering, we obtain high compression ratios which denote the presence of rich clustering structures.

The neighborhood function approximation allows us to compute the average distance, which is equal to the number of degrees of separation plus one, the effective diameter, which is the 90th percentile of distance distribution, the harmonic diameter, and the shortest path index of dispersion (SPID). If we compare our values with the values in Table 3 for known studies, we may observe that our values are more close to those of dynamic networks. In particular, it is interesting that we obtain a number of degrees of separation close

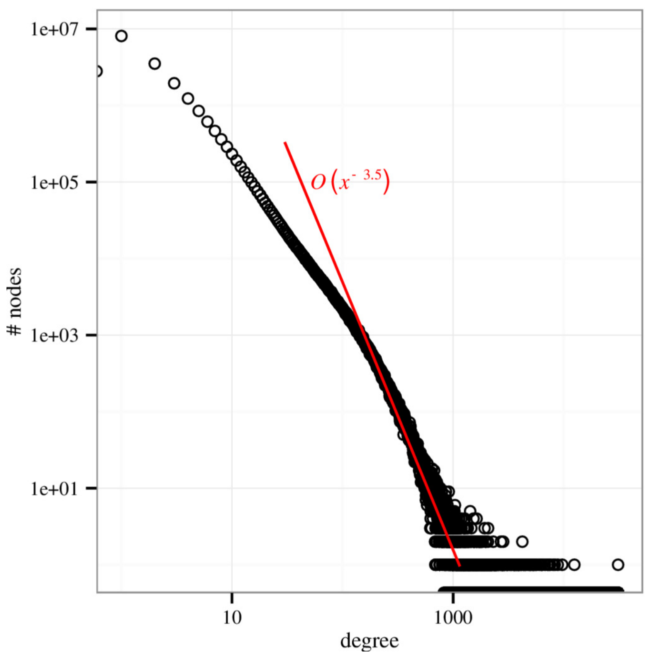

Figure 6: The outdegree distribution for the mention network or, equivalently, the indegree distribution for the influence network.

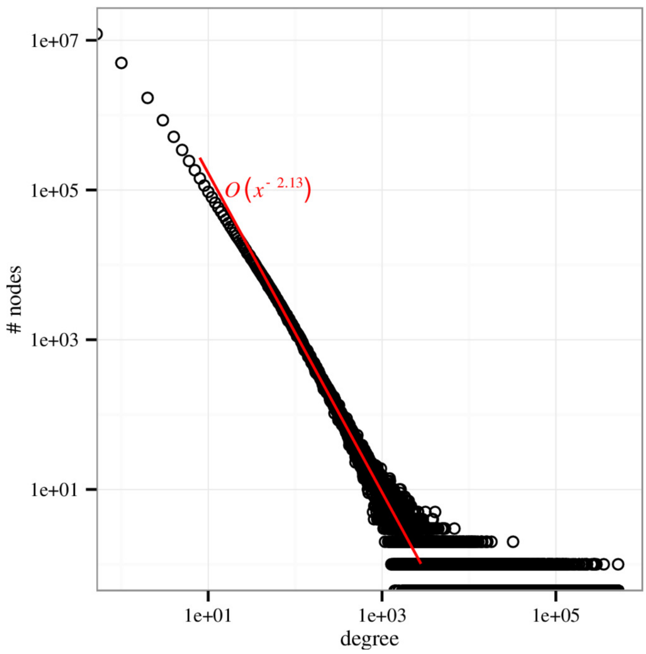

Figure 7: The outdegree distribution for the influence network or, equivalently, the indegree distribution for the mention network.

to six, rather different from results obtained in recent studies of Twitter static networks, but in line with Milgram and MSN experiments. See Figure 5 for the distance probability mass and cumulative distribution functions. The SPID is around 0.2 for the three studied networks, corroborating the conjecture that proper social networks have a SPID

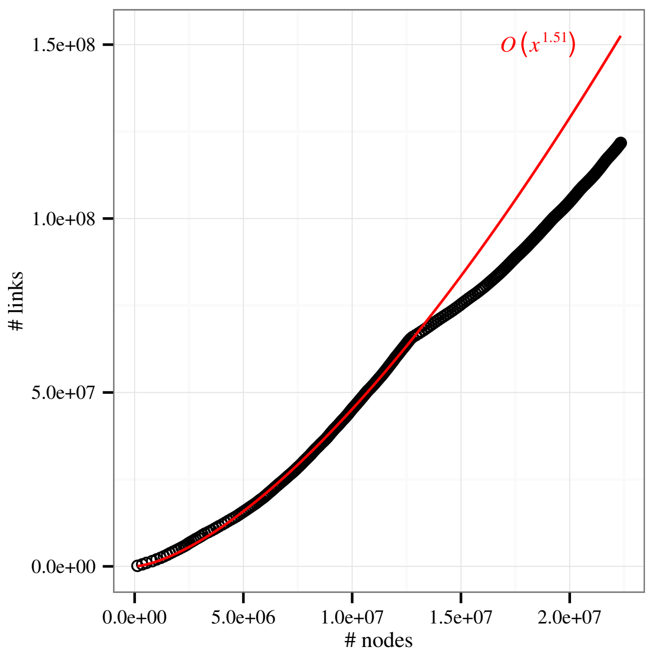

Figure 8: The densification law for the influence network.

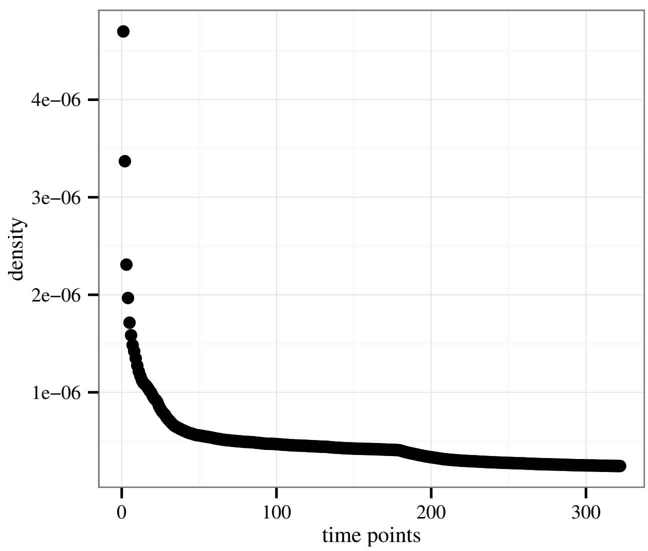

Figure 9: The density variation over time for the influence network.

value bellow one [6]. Finally, we observe that the harmonic diameter is close to the effective diameter and to the average distance for the conversation network, which is interesting as this is also observed for other symmetric networks (Table 3).

3.5 Network over time

In order to better understand how these networks change over time, we provide the measurements for the influence network over 323 incremental snapshots. We order the tweets in a non-decreasingly date and we consider for the first snapshot the first one million tweets from our dataset. Then, each snapshot increments the previous one with the next

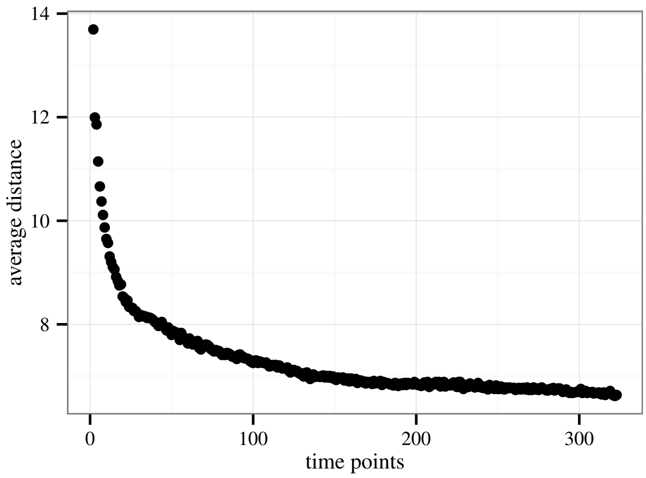

Figure 10: The average distance over time for the influence network.

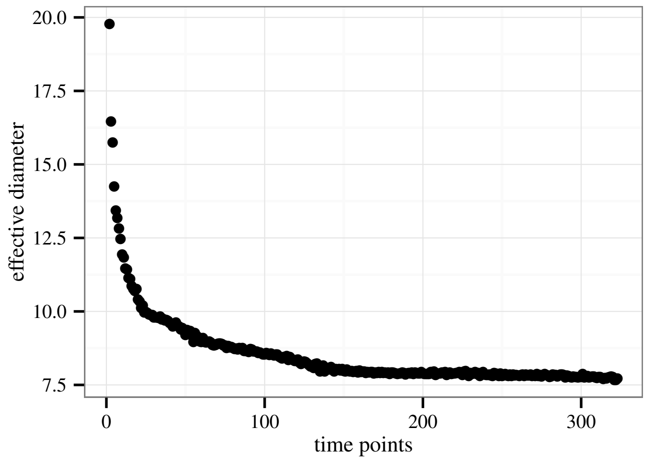

Figure 11: The effective diameter over time for the influence network.

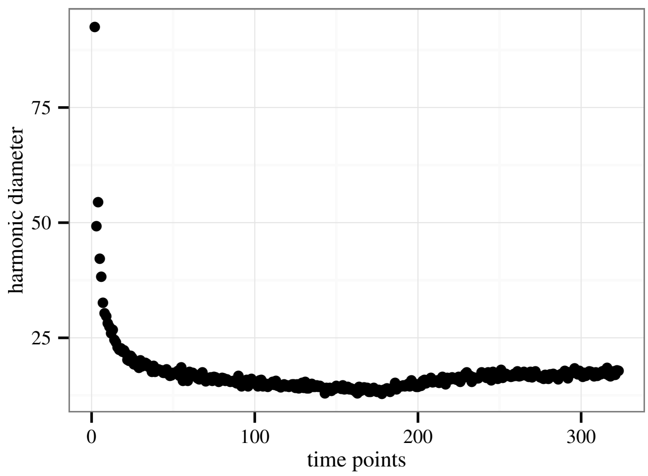

Figure 12: The harmonic diameter over time for the influence network.

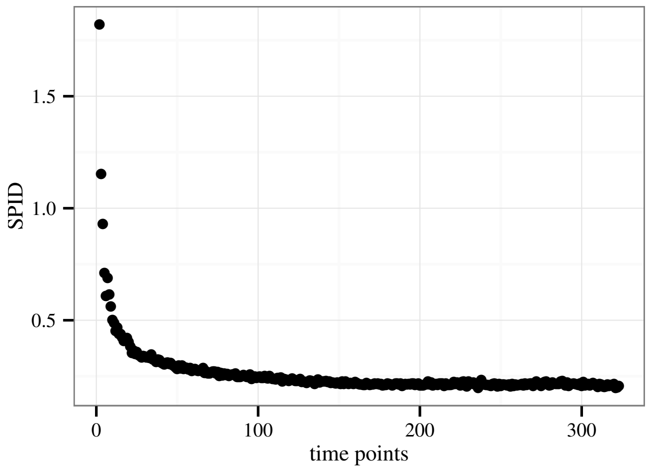

Figure 13: The shortest path index of dispersion (SPID) over time for the influence network.

one million tweets. We prefer this approach instead of splitting data by date, for instance, by weeks, because there were some variations on the amount of data collected per unit of time due, for instance, to holiday periods.

As we can observe in Figure 8, the number of edges grows superlinearly in the number of nodes, although with a coefficient smaller than 2. Since the number of possible edges grows quadratically with the number of vertices, we can conclude that the density of the influence network decreases over time, as is observed in Figure 9.

In Figures 10, 11, 12, and 13, we can observe that the average distance, the effective diameter, the harmonic diameter and the SPID, although decreasing, seem to be stabilizing.

4. RELATED WORK

Given the large scale proportions that online social networking services have gain in last years, we have today the opportunity to study and analyze large social networks involving the interaction among millions of persons [4, 24]. In this context, several studies have focused the measurement of the average distance on social networks, and other related metrics, in order to replicate the study conducted by Stanley Milgram [20, 22] and to better understand "small world phenomena" in large scale networks. Surprisingly, even though studied networks are rather different, most recent results for interaction and communication networks, including those discussed in this paper, corroborate the estimate by Milgram in 1967: the number of intermediaries on postcard routing is on average 5.2 (and, hence, the average distance is 6.2). We provide in this section, and in particular in Table 3, a briefly comparison of ours and others results.

As we discussed earlier and as is described in Table 3, we obtain in this work an average distance similar to the one obtained in Milgram experiment. In fact, if we consider the influence network, we have a similar setup as messages are propagated from authors to forwarders. Although only few persons were observed in Milgram experiment, we should not forget that the experiment took place over a large real network. Similarly, in our study, we are also observing a subset of users, but we have a much larger and representative sample. The interesting fact is that we are observing similar values for all metrics on this larger network sample.

Table 3: Comparison of large social networks and with the Milgram experiment. Note that the median distance includes also the percentage of pairs of vertices at such distance.

	Milgram Exp.[20]	**MSN** [17]	**Facebook**[1, 8]	**Twitter**[16]	**Tt. Influen.**[d]	**Tt. Conv.**[d]
Vertex	person	user	user	user	user	user
Link	selected person	message exchange[c]	friendship[b]	follower	mention[c]	mention[c]
Link type	dynamic	dynamic	static	static	dynamic	dynamic
Symmetric	no	yes	yes	no	no	yes
Avg. dist	6.2	6.6/6.618[e]	4.74	4.12/4.46[a]	6.63	7.67
Efect. dmt	–	7.8	–	4.8	7.73	8.95
Harm. dmt	18.29[e]	8.935[e]	4.59	5.29[a]	18.43	7.56
Median	∞	6 /7 (78.7 %)[e]	5 (92 %)	5 (73.16%)[a]	∞	8 (69.65%)
SPID	–	–	0.09	0.17	0.21	0.26

[a] Value reported in `http://law.di.unimi.it/webdata/twitter-2010/` for same data.
[b] Friendship connections stablished prior to the date of collecting data.
[c] During the period of collecting data.
[d] Networks studied in this paper.
[e] Value reported in [8].

Leskovec and Horvitz[17] observed also that the average distance within the MSN messenger network is 6.6. Their data corresponds to the interactions among messenger users during June 2006, from which they identified a communication network where a link exists between two users if they exchanged at least a message during that month. The average distance was computed by randomly sampling 1000 users and calculating for each user the shortest paths to all other users in the network. They have also observed that the effective diameter is 7.8. As in our case, this network is highly dynamic since it was built from as messages exchange among users. Note however that the MSN network is symmetric (with a lower harmonic diameter) and, hence, we should compare it with our conversation network, where we have slight larger values, although comparable.

More recently, Backstrom *et al.* computed the average distance for the Facebook network, obtaining a value of 4.74, as well as other metrics (see Table 3). For this study, the authors collected data between 2007 to 2012 and only considered users that were active in May of 2011. In contrast to the networks studied in our paper, the type of link is "static" since it is not capturing users dynamic activity as links are just friendship connections, including passive/weak links. Moreover the Facebook network is symmetric and, comparing with the MSN network and with our conversation network, its values are significantly smaller. This fact may not be so surprisingly given that Facebook is the largest social network analyzed till nowadays and, in this case, authors are not capturing user interactions, just friendship links. It would be interesting to replicate our study considering instead information exchange on Facebook.

In what concerns Twitter, Kwak *et al.* [16] studied a directed network based on *following* and *followed by* links, observing an average distance of 4.12 and an effective diameter of 4.8. Authors conjecture that these results are an evidence that Twitter has another role than a social network. By comparing with the MSN network, the average distance and effective diameter is indeed smaller, but such values are almost the same as those obtained for the Facebook network. Hence, based on this observation and on our study, we believe that the difference observed is more related with the type of links considered than with Twitter role.

Network evolution over time has been also addressed in several studies. Leskovec *et al.* [18] studied the citation graph for U.S. patents and affiliation graphs of authors, by observing snapshots of these network over long periods of time. They observed that most of these graphs densify over time, *i.e.*, the number of edges grows superlinearly in the number of nodes, and that the average distance shrinks over time. Notice that the densification definition used by the authors is not equivalent to the usual density definition (observed edges *vs* the number of possible edges). In what concerns our analysis, although we observe a similar feature for the densification (Figure 8), we do not observe a similar result with respect to the average distance. In fact, the average distance appears to be stabilizing, being in concordance with the results by Backstrom *et al.* [1] for Facebook.

Since we based our analysis on mentions, that translate influence among users, we point out some work on influence supporting our previous observations. There are several recent studies about influence in social networks and, in particular, in Twitter [25, 2, 11, 23]. However, current influence metrics may be fooled, for instance, by spammers [3], by bots [9], by social capitalists [14], and may not capture the temporal dynamics of Twitter. Cha *et al.* [11] observed also that the most followed users are not the necessarily the more influential ones. Moreover, Counts and Fisher [12] found evidences that there are some reply patterns that reflect attention interest. These observations were the main motivation for our work, since friendship and subscription links do not necessarily correspond to active links and, moreover, it may exist even contextual interactions beyond such "static" links.

5. CONCLUSIONS AND FURTHER WORK

We analyzed different interaction networks from Twitter social platform. Contrary to previous studies of Twitter, that define underlying networks based on following and followed relations among users, we observed higher values for the average distance and for the effective diameter, 6.6 and 7.73 respectively. Moreover, by comparing these results with published ones for different social networks, even though they are different in nature, observed differences seem to be related with particular network characteristics. In particular, network symmetry and link nature seem to be the main distinguishing features, where by link nature we mean if links are inferred from observed interactions among users or if links are just friendship or following relations.

In our case we studied networks inferred from user interactions on Twitter, namely through the analysis of mentions, ignoring existing following relations. This is an important difference with respect to previous analyses of Twitter since we try to capture possible influence among users beyond static links and to avoid the existence of weak/dead links. Moreover, we considered a similar setup to the one in Milgram experiment, where messages are propagated from authors to forwarders. The observed values for all metrics are similar to both the Milgram experiment and the MSN experiment, which is an interesting fact given that both our study and the MSN experiment consider a much larger set of users than the Milgram experiment.

We analyzed also networks through time and, as in previous studies, we observed densification over time, *i.e.*, the number of edges grows superlinearly in the number of nodes, and that the average distance appears to be stabilizing.

Given our conclusions, and in particular the fact that we obtained different results for Twitter interaction networks than those known for Twitter static networks, but similar to those obtained for other different dynamic networks underlying social platforms, we consider that it would be interesting to replicate our study to other well known social networks. It is our belief that this kind of analysis provide new insights on how social interactions should be considered, given their dynamic nature over time, with significant implications on most applications relying on user influence and information propagation models on social systems.

Finally, we leave also as future work the task of modelling this kind of network, as well as the fitting of existing theoretical network models. This is a crucial step to fully understand these complex networks.

Acknowledgments.

This work was partly supported by national funds through FCT – Fundação para a Ciência e Tecnologia, under projects PTDC/EIA-CCO/118533/2010 and PEst-OE/EEI/LA0021 /2013.

6. REFERENCES

[1] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Web Science 2012*, pages 33–42, 2012.

[2] E. Bakshy, J. Hofman, W. Mason, and D. Watts. Everyone's an influencer: quantifying influence on twitter. In *WSDM'11*, pages 65–74. ACM, 2011.

[3] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Detecting spammers and content promoters in online video social networks. In *SIGIR'09*, pages 620–627, 2009.

[4] F. Benevenuto, T. Rodrigues, M. Cha, and A. V. Characterizing user behavior in online social networks. In *ACM SIGCOMM Conference on Internet Measurement*, pages 49–62. ACM, 2009.

[5] P. Boldi, M. Rosa, M. Santini, and S. Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *WWW'11*, pages 587–596. ACM, 2011.

[6] P. Boldi, M. Rosa, and S. Vigna. Hyperanf: approximating the neighbourhood function of very large graphs on a budget. In *WWW'11*, pages 625–634. ACM, 2011.

[7] P. Boldi and S. Vigna. The webgraph framework I: compression techniques. In *WWW'04*, pages 595–602. ACM, 2004.

[8] P. Boldi and S. Vigna. Four Degrees of Separation, Really. In *ASONAM'12*, pages 1222–1227, 2012.

[9] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In *ACSAC'11*, pages 93–102, 2011.

[10] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Comput. Netw.*, 33(1-6), pages 309–320, 2000.

[11] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in Twitter: The million follower fallacy. In *ICWSM'10*.

[12] S. Counts and K. Fisher. Taking it all in? Visual attention in Microblog consumption. In *ICWSM'11*.

[13] B. Efron and G. G. A leisurely look at the bootstrap, the jacknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[14] S. Ghosh, B. Viswanath, F. Kooti, N. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. Gummadi. Understanding and combating link farming in the twitter social network. In *WWW'12*, pages 61–70. ACM, 2012.

[15] M. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[16] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW'10*, pages 591–600. ACM, 2010.

[17] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW'08*, pages 915–924, 2008.

[18] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), 2007.

[19] P. McFedries. Technically speaking: All a-twitter. 44:84–84, 2007.

[20] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[21] E. M. Rogers. *Diffusion of Innovations, 5th Edition.* Free Press, 1995.

[22] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, pages 425–443, 1969.

[23] J. Weng, E. Lim, J. Jiang, and Q. He. TwitterRank: finding topic-sensitive influential twitterers. In *WSDM'10*, pages 261–270. ACM, 2010.

[24] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Zhao. User interactions in social networks and their implications. In *In EuroSys Conference*, pages 205–218. ACM, 2009.

[25] S. Wu, J. Hofman, W. Mason, and D. Watts. Who says what to whom on twitter. In *WWW'11*, pages 705–714. ACM, 2011.