Multimedia Tools and Applications

Influential users in Twitter: detection and evolution analysis

--Manuscript Draft--

Manuscript Number:	
Full Title:	Influential users in Twitter: detection and evolution analysis
Article Type:	1094 -- Smart Technologies for Social Good
Keywords:	Graph analysis; social media; Twitter graph; retweet graph; graph dynamics; centrality.
Corresponding Author:	Paola Vocca Universita degli Studi della Tuscia Dipartimento delle Scienze Umanistiche, della Comunicazione e del Turismo Viterbo, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita degli Studi della Tuscia Dipartimento delle Scienze Umanistiche, della Comunicazione e del Turismo
Corresponding Author's Secondary Institution:	
First Author:	Giambattista Amati
First Author Secondary Information:	
Order of Authors:	Giambattista Amati
	Simone Angelini
	Giorgio Gambosi
	Gianluca Rossi
	Paola Vocca
Order of Authors Secondary Information:	
Funding Information:	

Noname manuscript No.
(will be inserted by the editor)

Influential users in Twitter: detection and evolution analysis

Giambattista Amati · Simone Angelini · Giorgio Gambosi · Gianluca Rossi · Paola Vocca

Received: date / Accepted: date

Abstract In this paper, we study how to detect the most influential users in the microblogging social network platform Twitter and their evolution over the time. To this aim, we consider the *Dynamic Retweet Graph (DRG)* proposed in [3] and partially analyzed in [4,2]. The model of the evolution of the Twitter social network is based on the retweet relationship. In a DRGs, the last time a tweet has been retweeted we delete all the edges representing this tweet are deleted. In this way we model the decay of tweet life in the social platform.

To detect the influential users consider the central nodes in the network with respect to the following centrality measures: *degree, closeness, and pagerank-centrality*. These measures have been widely studied in the static case and we analyze them on the sequence of DRG temporal graphs with special regard to the distribution of the 75% most central nodes.

We derive the following results: (a) in all cases the closeness measure produces many nodes with high centrality, so it is useless to detect influential users; (b) for the other measures almost all nodes have null or very low centrality and (c) the number of vertices with significant centrality are often the same; (d) the above observations hold also for the the whole DRG and, (e)

This work was conducted in the Laboratory of Big Data of ISCOM-MISE (Institute of communication of the Italian Ministry for Economic Development).

G. Amati · S. Angelini
Fondazione Ugo Bordoni, Rome, Italy
E-mail: gba,sangelini@fub.it

G. Gambosi · G. Rossi
University of Rome "Tor Vergata", Rome, Italy
E-mail: giorgio.gambosi, gianluca.rossi@uniroma2.it
Partially supported by ISIDE project.

P. Vocca
University of Tuscia, Viterbo, Italy
E-mail: vocca@unitus.it

central nodes in the sequence of DRG temporal graphs have high centrality in static graphs.

Keywords Graph analysis · Social media · Twitter graph · Retweet graph · Graph dynamics · Centrality.

1 Introduction

One of the fundamental and most studied features in a social network is the detection of central nodes, which can usually be considered as the *most important* nodes [7,8,13]. Centrality is widely-used for measuring the relative importance of nodes within a graph and it has many applications: in social networks to determine the most influential or well-connected people; in the Web graph to rank pages in a search; in a terrorist network, to detect agents that are critical for facilitating the transmission of information; for the dissemination of information in P2P Networks, Decentralized Online Social Networks and Friend-to-Friend Network [11].

There is a plethora of centrality definitions: degree centrality [18], closeness centrality [5], graph centrality [15], stress centrality [19], betweenness centrality [12], each one of them useful to detect specific properties and with significantly different computational costs. Here we consider four of them: the *degree*, *closeness*, *betweenness*, and *PageRank*-centrality.

Degree centrality, i.e. the degree d_v of a vertex v, is the simplest measure of centrality: it just takes into account how many direct, "one hop" connections each node has to other nodes of the network, hence it can be applied to detect popular individuals, agents who are likely to hold most information or individuals who can quickly connect with the wider network. The degree centrality is very cheap to compute but, being a purely local notion, it is often unable to recognize the relevance of certain nodes.

One of the most popular measures, but computational expensive for large graphs, is the betweenness-centrality. It detects nodes which act as "bridges" between other nodes in a network. It does this by identifying all the shortest paths and then counting how many times each node falls on one. Betweenness centrality is suitable for finding vertices who influence flows (such as information flow) in the network.

A third measure considered below is closeness-centrality, which, after computing the set of all-pairs shortest paths, assigns each node a score based on the number of shortest paths to which it belongs. This definition of centrality is useful for quickly finding the agents who are in good position to influence the entire network but in a highly connected network often most nodes have a similar score.

Finally, PageRank-centrality was introduced in [9] and it recursively quantifies a "value" or the PageRank of a node based on: (i) the number of links it receives, (ii) the link propensity of the linkers (that is, the number of outgoing links of each in-going node), and (iii) the centrality of the linkers, that is their PageRank.

To study how influential users evolve over the time we analyze the distribution of the centrality measures on an temporal evolutionary model of the Twitter network, the *Dynamic Retweet Graph (DRG)* proposed in [3] and partially analyzed in [4,2].

This model has two major features: (i) we consider the retweet graph since allows to better represent relationships among users and the information flow in Twitter [17,16] and (ii) once a tweet has been retweeted for the last time all the edges representing this tweet are deleted, to model the decay of relevance of the tweet content.

The temporal model we consider coincides with other temporal models in the growing phase [6,14], that is, a new vertex is added when a new user starts or retweets a tweet, and a new directed edge (a, b) is inserted when an user a retweets for the first time a tweet of b, if already an edge exists then a timestamp is added to it. Conversely, the decreasing stage happens when a tweet is no more retweeted. Then, all the vertices and the edges, not involved in other retweeting processes, are deleted at once. As shown in previous experimentation [2,4], this evolutionary model better captures the information flow in Twitter. DRGs seem to better represent the double nature of the Twitter platform: social network or news media [17,16].

For what concerns the use of centrality measure to assess influential or authoritative users Kwak et al. [16] compared three measures of influence: in-degree centrality, PageRank centrality in the following/follower network and the number of retweets on Twitter. In Cha et al. [10] compared three different measures of inuence: in-degree centrality, the number of retweets and mentions on Twitter. The results indicate that users with high in-degree were not necessarily influential.

In this paper we study the evolution of the most influential users in the microblogging social network platform Twitter with respect to the above four centrality measures (betweenness, degree, closeness, and PageRank) and we analyze their behavior on the DRG evolutionary model of the retweet social networks proposed in [3].

We consider two different kind of data sets, first introduced in [1] and updated and refined in [3]: the *event driven* retweet graphs based on the events *Black Friday 2015* and the *World Series 2015* and the *Italian Sampling* that is the *firehose* retweet graph, filtered by language (i.e. Italian) from the whole Twitter stream.

The four centrality measures are analyzed on three levels: (i) with respect to the sequence of DRG temporal graphs; (ii) with respect to the static cumulative graph, that is the graph that contains all the nodes and edges and (iii) with respect to the kind of networks considered, that is *event driven* or the firehose.

We derive that the model proposed allows to detect the most authoritative users, since:

1. in all cases the closeness centrality provides too many central nodes, hence it is useless to detect influential users;

2. with regard the other measures, almost all nodes have null or very low centrality;
3. vertices with centrality values above 75% of the maximum is a small set and they are often repeated in the three centrality measures;
4. the above observations hold also for the static graphs (the whole DRG);
5. central nodes in the sequence of DRG temporal graphs have high centrality in static graphs.

2 DRG temporal graphs

In this paper we will use a definition of Dynamic Retweet Graph (DRG) slightly different from the one in [4].

A DRG graph $G = (V, E, \ell)$ is defined as follows: the set V of nodes are Twitter accounts and a direct edge $e \in E$ represents an interaction (a retweet) between two accounts. In particular, there is a directed edge from an account a towards an account b, if a has retweeted at least one tweet of b, that can be itself already a retweet. Observe that user a may retweet more tweets of b. This edge information is implemented with a list $\ell(e)$ associated to every edge $e = (a, b)$ that contains pairs (i, t) where i is the id of a tweet and t is the timestamp in which a retweets i from b. The pairs of $\ell(e)$ are sorted for timestamps in non-decreasing order.

From the data that we have collected in G we define, for all tweets i, the *date of death* of i (in short, $\texttt{dod}(i)$) as the timestamp of the last retweet of i. Formally,

$$\texttt{dod}(i) = \max_{e \in E} \{t : (i, t) \in \ell(e)\}.$$

Consequently we define the *expiration date* of an edge e (in short, $\texttt{ed}(e)$) as the timestamp from which all tweets associated to e will be dead. Formally,

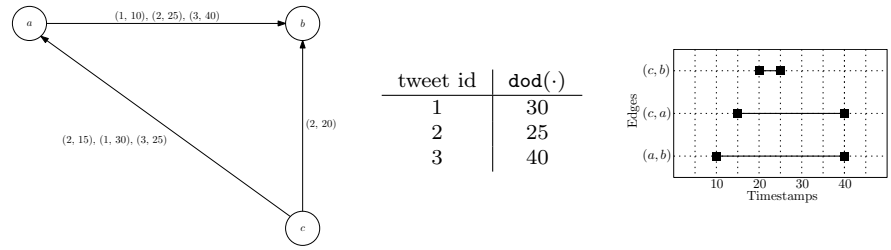
$$\texttt{ed}(e) = \max\{\texttt{dod}(i) : (i, t) \in \ell(e)\}.$$

On the contrary, the *creation date* of an edge $e = (a, b)$ (in short, $\texttt{cd}(e)$) is the timestamp b retweets a for the first time, formally:

$$\texttt{cd}(e) = \min\{t : (i, t) \in \ell(e)\}.$$

Let t be a timestamp, we define a *DRG temporal graph* at time t the subgraph $G_t = (V_t, E_t)$ of the DRG G at time t as follows: E_t contains any edge e such that $\texttt{cd}(e) \leq t \leq \texttt{ed}(e)$; V_t is the set of nodes induced by E_t.

For example if G is the retweet graph represented in the left part of Fig. 1, G_{30} contains edges (a, b) and (c, a) and the induced vertices since (c, b) expires at timestamp 25. For all $20 \leq t \leq 25$, G_t contains all edges of G.



tweet id	dod(·)
1	30
2	25
3	40

Fig. 1 On the left side, an example of a DRG retweet graph. Edges are labelled by pairs with the id of the tweet and the timestamp of the retweet. The center table shows the date of death of all tweets in the graph. On the right side, for each edge of G is represented its creation and expiration date.

Table 1 Dimensions of the dataset

	World Series	**Black Friday**	**Italian Sampling**
Vertices	$4.74 \cdot 10^5$	$2.7 \cdot 10^6$	$2.541739 \cdot 10^6$
Edges	$8.40 \cdot 10^5$	$3.8 \cdot 10^6$	$1.3708317 \cdot 10^7$
Tweets/edges	2.3	2.603	5.45
Tweets/vertices	4	3.66	29.4

3 Data sets

For the experiments we use the dataset of [3] that consists in two different classes of retweet graphs: the event driven retweet graph, filtered by topics about specific events (i.e. the Black Friday 2015 and the World Series 2015) and the Sampling retweet graph, filtered by the Italian language from the whole Twitter stream. To obtain the Italian Twitter sample we use a list of the most used Italian stop words and the Twitter native selection function for languages. In Table 1 the dimensions of the three graphs are shown. In Figure 2 we show the evolution of the dimensions of the three datasets over the period of observation. Note that the event-driven datasets (World Series and Black Friday) show a rapid growth close to the events, and then a slow decline. Differently, the Italian Sampling show a smooth and stable behavior, ignoring the border effects.

4 Experimentation

For each graph G in our dataset, we consider the sequence of DRG temporal graphs $(G_{t_i})_{i\geq 0}$ where $t_{i+1} - t_i$ is 4 hours. For each G_t we compute the four centrality values (betweenness, closeness, degree, and PageRank centrality) of each vertex of the graph.

Given the centrality measure c, the *relative centrality value* with respect to c of a vertex u is the ratio $c(u)$ and the maximum value of $c(\cdot)$.

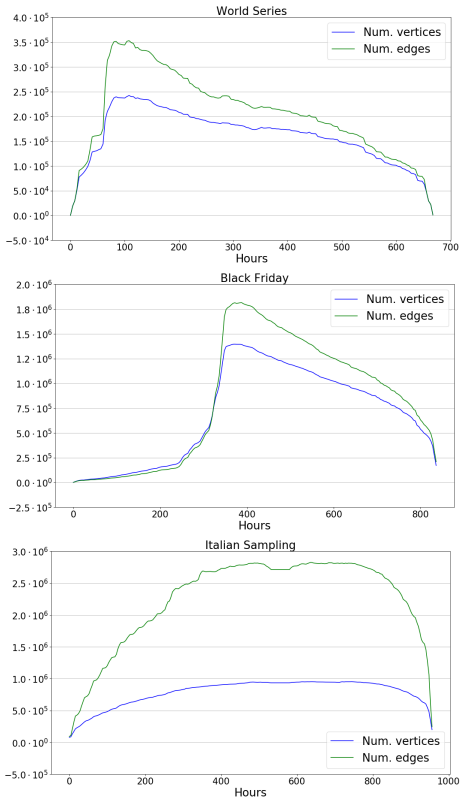


Fig. 2 Number of vertices (blue) and number of edges (green) of: World Series, Black Friday, and Italian Sampling, as functions of hours.

Preliminary considerations. First of all, for each centrality measure $c(\cdot)$ and for each G_t, we consider the number of nodes with centrality values above the 90% of the maximum. Fig. 3(a) shows the behavior of the closeness centrality: observe that this value is almost always greater than 30%. This means that closeness centrality is not very suitable to determine the more influential nodes in the graph. Conversely, the other centrality measures (degree, betweenness, and PageRank) reveal an opposite behavior: excluding the first and last timestamp, 99.9% of vertices always have centrality values below the 20% of the maximum. This is shown in Figure 3(b) which shows the evolution over time of the three centrality values below which the 99.9% of all values fall (99.9-th percentile). Observe that, from Fig. 3(b) it results that the highest values are at the very beginning of time sequences, when there is still much instability. After that, values fall below 0.05.

Analysis of temporal graphs. From the previous observations it follows that if we restrict ourselves to the betweenness, degree and PageRank measures, the

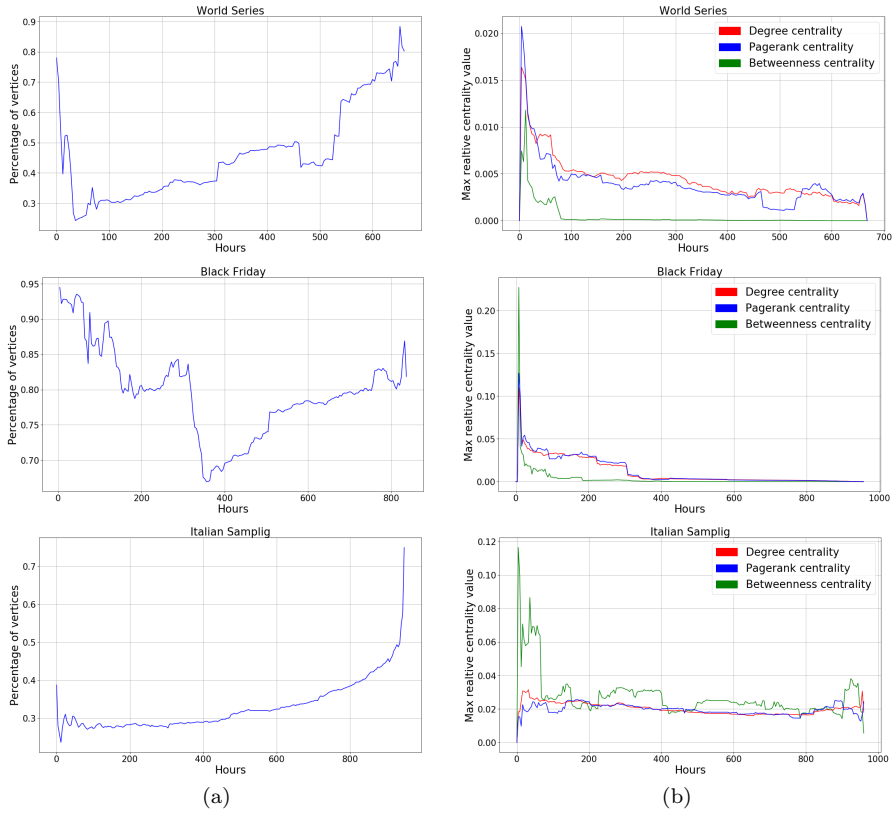


Fig. 3 (a) Trend over time of the ratio of nodes whose closeness centrality is above the 90% of the maximum. (b) The 99.9-th percentile evolution over time of the three relative centrality measures.

Table 2 Number of central nodes for dataset and centrality measure.

	World Series	**Black Friday**	**Italian Sampling**
Betweennes	15	44	31
Degree	4	11	10
PageRank	12	16	11

number of vertices for which the centrality value is meaningful is so small that we can study them one by one.

We say that a node is *central* (with respect to a centrality measure) if its centrality value is at least 75% of the maximum. Let G be a DRG, c be a centrality and t be a timestamp, we define $A_{G,c,t}$ as the set of central node of G_t with respect to c. Table 2 shows the number of central nodes for each dataset and centrality measure.

In Fig. 4, are shown the sets $A_{G,c,t}$ for the three datasets. The x-axis represent the time and in the y-axis are reported the vertex ids. In the same plot are collected the informations regards the three centrality measures each of which is represented by a color: green for the betweenness; red for the degree; and blue for the PageRank centrality. An horizontal segment in correspondence to node u that intersects timestamp t means that $u \in A_{G,c,t}$ where c is the centrality measure associated to the segment color. Nodes that are central for

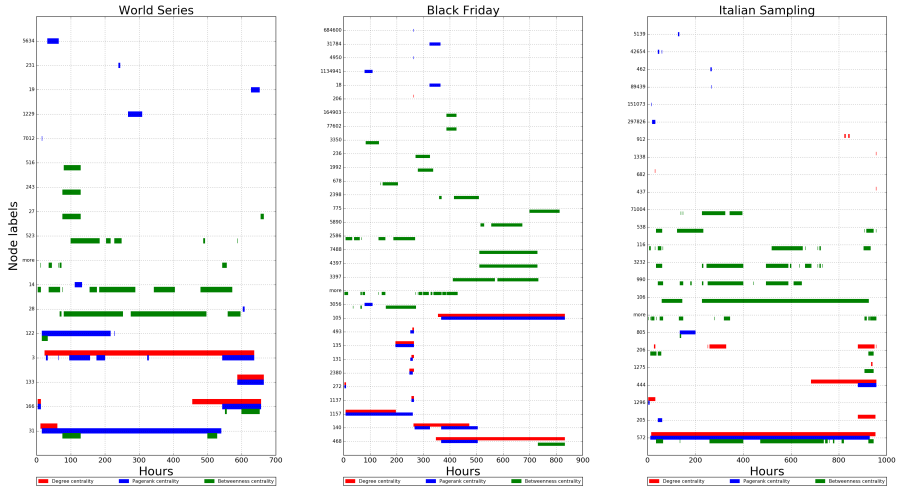


Fig. 4 Temporal evolution of $A_{G,c,t}$ for the three datasets with respect to the betweenness (in green), degree (in red), and PageRank (in blue) centrality measure.

more than one centrality measure are grouped together in the lower portion of the plots. We have observed that there are nodes central only with respect the betweenness centrality measure and for a very short period. For sake of clarity we have grouped these nodes together and represented them by a pseudo-node denoted as `more`. In the World Series the `more` node represents the union of the segments from 6 nodes; in the Black Friday 29; and in the Italian Sampling 21.

From the above analysis we get the following observations:

- For all datasets, the degree centrality always produces a total number of central nodes lower than the other measures. Conversely, betweenness centrality is the one that produces more.
- For all datasets and all the centrality measures, there are nodes that are central for long periods: this trend is more prominent for degree and pagerank centrality.
- Another important result that turns out is a significant overlap between the central vertices with respect the three measures. For example vertex 572

Table 3 Percentage of vertices whose relative centrality value is at most 0.01.

	Betweenness	**Degree**	**PageRank**
World Series	99.93%	99.95%	99.93%
Black Friday	99.97%	99.96%	99.97%
Italian Sampling	99.93%	99.78%	99.84%

Table 4 Pearson correlation between the centrality values in the cumulative DRGs and the aggregated centrality values in the temporal DRGs.

	Betweenness	**Degree**	**PageRank**
World Series	0.59	0.89	0.91
Black Friday	0.16	0.84	0.97
Italian Sampling	0.72	0.75	0.88

in Italian Sampling is central for most of the time over the three measures (see the third plot of Fig. 4).

Comparison with the static cumulative DRGs. This last analysis involves the centrality measures of the static cumulative DRGs G representing the three datasets. Like DRGs temporal graphs, a large portion of vertices, varying form 28% (for World Series) to 50% (for Black Friday), have closeness centrality above 90% of the maximum, hence, we discard it.

On the contrary for the betweenness, degree, and PageRank centrality, almost all the nodes have centrality below 1% of the maximum. Table 3 shows, for each dataset and for each measure the percentage of vertices whose relative centrality value is at most 0.01.

Our goal is to compare the centrality measures in the cumulative DRGs with the ones in the temporal DRGs. At this moment given a dataset and a centrality measure c we have for each node u:

- a single centrality value $c(u)$ in the case of the cumulative DRGs;
- a sequence of centrality values $c_0(u), c_1(u), \ldots$ in the case of temporal DRGs, one value for timestamp.

In order to make the two data comparable, we aggregate the sequence of centrality values $c_0(u), c_1(u), \ldots$ into a single value $s(u)$ given by the sum of all $c_i(u)$. Finally we can compare the sequence of centrality values of the cumulative DRGs with the sequence of the $s(\cdot)$ values of the temporal DRGs. Table 4 shows the Pearson correlation coefficients between these observations. It turns out that there is a strong correlation in the case of degree and PageRank centrality. Instead, regarding Betweenness centrality the correlation coefficient varies considerably.

Analyzing deeply, we will discover that also for the betweenness centrality there is a strong relationship between nodes that are central in both the cumulative DRGs and the temporal DRGs. To this aim we will focus on the relative centrality on cumulative DRG of vertices with high relative centrality

Table 5 Relative betweenness centrality in the cumulative Black Friday dataset of nodes that are central in the temporal graphs.

Vertex id	**Relative centrality**
236	0.53
2398	0.39
5890	0.16
7780	0.14
12605	0.12
17414	0.11
16426	0.10
3397, 2607	0.09
9451	0.07
2586, 16417, 4397	0.06
7488, 3056	0.05
7082, 5806, 2542, 146487	0.04
3350, 678, 56750, 56759, 4946, 6118	0.03
37990, 37982, 56760, 164903	0.02
77602, 682	0.01
8714, 9450, 4846, 24191, 22726, 16411, 25530, 118159, 1992, 468, 775, 6077, 170197	< 0.01

Table 6 Relative betweenness centrality in the cumulative World Series dataset of nodes that are central in the temporal graphs.

Vertex id	**Relative centrality**	**Vertex id**	**Relative centrality**
299	1.00	243	0.19
31	0.69	340	0.18
27	0.67	126	0.10
122	0.62	516	0.07
14	0.49	521	0.05
46	0.25	66050	0.03
28	0.23	166	< 0.01
523	0.20		

in the temporal DRGs. Table 5 shows the relative betweenness centrality value in the Black Friday cumulative DRG of all central nodes in the Black Friday temporal DRGs with respect to the same centrality measure. It is interesting to note that 31 of the 44 listed nodes belong to the 0.03% (=100 − 99.97, see Table 3) of vertices whose relative betweenness centrality is at least 0.01. That is a large majority of nodes that are central in temporal graphs are also central in the whole graph.

Such behavior is even more pronounced in the World Series and Italian Sampling dataset. Table 6 and 7 show the analogue of Table 5 for the World Series and Italian Sampling datasets. In the World Series case there is only one central node (on 15) in the temporal graph with a relative centrality in the cumulative graph less than 0.01. In the Italian Sampling dataset this number is 4 (on 31).

Finally if we consider the degree and PageRank centrality measures the just described behavior is even more evident: all central nodes (but 3) in the

Table 7 Relative betweenness centrality in the cumulative Italian Sampling dataset of nodes that are central in the temporal graphs.

Vertex id	Relative centrality
106	1.00
3232	0.45
572	0.44
206,4853	0.32
990	0.30
3306	0.27
2567	0.22
653	0.21
372	0.16
116	0.15
1125	0.12
538	0.11
1275	0.10
5960, 71004, 493, 1851, 1511	0.08
645	0.07
209	0.06
6039	0.05
805, 8998, 5741	0.04
1849, 34521	0.01
22854, 41134, 273383, 52488	< 0.01

temporal graphs belong to the $\approx 0.2\%$ of nodes with relative centrality in the cumulative graph higher than 0.01. The three exceptions are related the PageRank measure for Black Friday (two nodes) and Italian Sampling (one node).

5 Discussion and Conclusions

In this paper we have studied the evolution of four centrality measures (betweenness, degree, closeness, and PageRank) on the DRG temporal retweet graphs based on three datasets: Black Friday, World Series, and Italian Sampling. Our main results can be summarized as follows: (i) too many nodes are central with respect closeness centrality, hence this measure is useless to detect influential users; (ii) for the other measures, the number of nodes with very low centrality is very high and the sets of central nodes (with centrality values above 75% of the maximum) are very small and quite similar in the three measures; (iii) similar results hold also for the static cumulative graphs where the sets of nodes with relevant centrality contain central nodes in the sequence of DRG temporal graphs.

As pointed out in [4], the DRG temporal graphs derived from our datasets are quite sparse: this could explain the small number of central nodes respect to the three centrality measures.

According to the above analysis the approach based on the DRG temporal graph and the centrality measures represent a promising approach for detecting influencers in the microblogging Twitter platform.

References

1. Giambattista Amati, Simone Angelini, Marco Bianchi, Luca Costantini, and Giuseppe Marcone. A scalable approach to near real-time sentiment analysis on social networks. In CEUR-WS.org, editor, *DART 2014 Information Filtering and Retrieval. Proceedings of the 8th International Workshop on Information Filtering and Retrieval co-located with XIII AI*IA Symposium on Artificial Intelligence (AI*IA 2014)*, volume 1314, pages 12–23, December 2014.
2. Giambattista Amati, Simone Angelini, Francesca Capri, Giorgio Gambosi, Gianluca Rossi, and Paola Vocca. Modelling the temporal evolution of the retweet graph. *IADIS International Journal on Computer Science & Information Systems*, 11(2), 2016.
3. Giambattista Amati, Simone Angelini, Francesca Capri, Giorgio Gambosi, Gianluca Rossi, and Paola Vocca. Twitter temporal evolution analysis: Comparing event and topic driven retweet graphs. In *BIGDACI 2016 - Proceedings of the International Conference on Big Data Analytics, Data Mining and Computational Intelligence, Volume 1, Funchal, Madeira, Portugal, July 2 4, 2016*, 2016.
4. Giambattista Amati, Simone Angelini, Francesca Capri, Giorgio Gambosi, Gianluca Rossi, and Paola Vocca. On the retweet decay of the evolutionary retweet graph. In *Smart Objects and Technologies for Social Good: Second International Conference, GOODTECHS 2016, Venice, Italy, November 30 – December 1, 2016, Proceedings*, pages 243–253, Cham, 2017. Springer International Publishing.
5. Alex Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.
6. Devipsita Bhattacharya and Sudha Ram. *Sharing news articles using 140 characters: A diffusion analysis on twitter*, pages 966–971. 2012.
7. Phillip Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.
8. Stephen P. Borgatti. Centrality and network flow. *Social Networks*, 27(1):55 – 71, 2005.
9. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.
10. Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. *Icwsm*, 10(10-17):30, 2010.
11. M. Conti, A. De Salve, B. Guidi, L. Ricci. Epidemic Diffusion of Social Updates in Dunbar-Based DOSN. In *Proceedings of Parallel Processing Workshops: Euro-Par 2014 International Workshops*, pages 311–322, 2014.
12. Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
13. Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215 – 239, 1978.
14. Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. Information diffusion in online social networks: A survey. *SIGMOD Rec.*, 42(2):17–28, July 2013.
15. Per Hage and Frank Harary. Eccentricity and centrality in networks. *Social Networks*, 17(1):57 – 63, 1995.
16. Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.
17. Seth A. Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. Information network or social network?: The structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 493–498, New York, NY, USA, 2014. ACM.
18. J. Nieminen. On centrality in a graph. *Scandinavian Journal of Psychology*, 15:322–336, 1974.
19. Alfonso Shimbel. Structural parameters of communication networks. *The bulletin of mathematical biophysics*, 15(4):501–507, 1953.

Giambattista Amati is a Senior Researcher at the Fondazione Ugo Bordoni. He graduated "cum laude" in Mathematics at the University of Rome "La Sapienza" in 1983 and in 2003 he received a PhD in Computing Science from the University of Glasgow. In 2002 he devised the DFR (Divergence From Randomness) models of Information Retrieval, founding the University of Glasgow's open source search engine Terrier. Terrier enables rapid development of actual and scalable systems of Information Retrieval applications.

Current interests include sentiment analysis, web, enterprise, microblog, blog and vertical search, information extraction.

Simone Angelini is currently employed at the Fondazione Ugo Bordoni. He received his Bachelor Diploma in Computer Science fronm the University of Rome "Tor Vergata".

Current interests include software development for big data analysis.

Giorgio Gambosi is a full professor at the University of Rome "Tor Vergata". His research interest include the design and analysis of algorithms and data structure with particular reference to their application to networks and distributed systems

Gianluca Rossi received the the Laurea cum laude in Computer Science from the University of Rome "La Sapienza" and the Ph.D. degree in Mathematical Logic and Theoretical Computer Science at the Department of Mathematics of the University of Siena jointly with the Computer Science Department of the University of Florence . Since 2005 he is a researcher in computer science at the University of Rome "Tor Vergata". His research interests include the design and analysis of algorithms for graphs, networks and distributed systems.

Paola Vocca is an associate professor at the University of "La Tuscia", Viterbo. She graduated "cum laude" in Mathematics in 1987 and in 1993 received a PhD in Computer Science both from the University of Rome "La Sapienza".

Current interests include large graphs analysis, microblog, wireless networks, and algorithms and data structure for graphs.

Click here to download Author Photographs FotoGianni.jpg

Click here to download Author Photographs FotoPaola2.jpg