

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/281427547>

Measuring user influence on Twitter: A survey

Article in *Journal of Information Processing and Management* · April 2016

DOI: 10.1016/j.ipm.2016.04.003 · Source: arXiv

CITATIONS

56

READS

956

2 authors:



Fabián Riquelme

Universidad de Valparaíso (Chile)

17 PUBLICATIONS 97 CITATIONS

[SEE PROFILE](#)



Pablo González-Cantergiani

University of Santiago, Chile

6 PUBLICATIONS 58 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Measuring influence on social networks [View project](#)

Measuring user influence on Twitter: A survey^{***}

Fabián Riquelme and Pablo González-Cantergiani

CITIAPS - Universidad de Santiago, Chile.
{fabian.riquelme.c, pablo.gonzalezca}@usach.cl

Abstract. Centrality is one of the most studied concepts in social network analysis. There is a huge literature regarding centrality measures, as ways to identify the most relevant users in a social network. The challenge is to find measures that can be computed efficiently, and that can be able to classify the users according to relevance criteria as close as possible to reality. We address this problem in the context of the Twitter network, an online social networking service with millions of users and an impressive flow of messages that are published and spread daily by interactions between users. Twitter has different types of users, but the greatest utility lies in finding the most influential ones. The purpose of this article is to collect and classify the different Twitter influence measures that exist so far in literature. These measures are very diverse. Some are based on simple metrics provided by the Twitter API, while others are based on complex mathematical models. Several measures are based on the PageRank algorithm, traditionally used to rank the websites on the Internet. Some others consider the timeline of publication, others the content of the messages, some are focused on specific topics, and others try to make predictions. We consider all these aspects, and some additional ones. Furthermore, we include measures of activity and popularity, the traditional mechanisms to correlate measures, and some important aspects of computational complexity for this particular context.

Keywords: Twitter, User influence, Opinion leaders, Centrality measure, PageRank

1 Introduction

Since the origins of social network analysis, there has been interest in identifying the most relevant actors of a social network. With the rise of the Internet and technology, online social networks (OSNs) turned into challenging cases of study, on which big data content and complex interpersonal ties among actors converge. Knowing the influence of users and being able to predict it can be useful for many applications, such as viral marketing [34,64,110], information propagation [49,45], searching [1], expertise recommendation [121], social customer relationship management [78], percolation theory [91], etc.

A social network can be represented by a graph, as in Figure 1, whose nodes can represent the users, and the edges the interpersonal ties among them. The *centrality* of a node refers to its relative importance within the network to which it belongs. There is a vast literature about centrality measures to identify the most important actors on a social network [140,128]. Each measure is based on different relevance criteria. For example, two of the most traditional measures are *degree* and *closeness* [39]. The first one considers for each node the number of its adjacent edges, while the second one considers the minimum sum of the shortest paths from a node to all the other nodes within the network. In Figure 1 nodes *c* and *e* have the higher degree, with three adjacent edges, followed by node *d*, with two adjacent edges. However, for the same graph, the higher closeness is for node *d*, with a sum of shortest paths equals 10, followed by nodes *c* and *e*, with a sum equals 11. For both measures, nodes *a*, *b*, *f* and *g* are the last in the ranking. Depending on the topology of the network, some other traditional measures are *betweenness* [39] and *eigenvector* [10]. There are also centrality measures based on how much information can be dispersed through the nodes of a network [40,46]. Such measures, as well as other derivatives, can also be used in the context of OSNs like Twitter, although do not take into account all the features of these particular services [93].

Besides the centrality of a node, it is important to know how a small set of initially active nodes can propagate some “germs” (ideas, trends, fashions, rules, ambitions, etc.) to other distant nodes in

* ©Published version: <http://dx.doi.org/10.1016/j.ipm.2016.04.003>

** ©2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

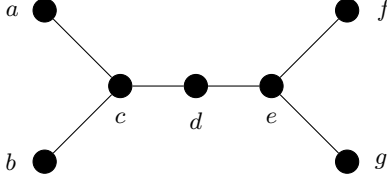


Fig. 1: An undirected graph representing interpersonal ties (edges) between users (nodes).

the network. For instance, in Figure 1, node f can only be directly influenced by node e , but node b could also propagate his influence until f , in a cascade behaviour through the nodes c , d and e . This phenomenon, known as the *spread of influence*, has been empirically verified on social networks several times [159]. Furthermore, there exist centrality measures explicitly based on this phenomenon [88].

Only in Twitter, spread of influence has been used for different purposes, such as political sciences [85], human mobility [47], transportation [79], rumor spreading [12], epidemiology [125], among many others. Twitter¹ was created in 2006, and since 2010 is being studied extensively in the contexts of social network analysis, computer science, and sociology. One of the most studied problems in Twitter is the identification of influential users. This problem is especially important, considering the high percentage of users who are often inactive or do not provide additional information [3]. Furthermore, the criteria for identifying influential users are as many as the growing number of techniques to rank them.

In this article we review the numerous investigations that have been carried out regarding the definition of influence measures on Twitter. Although the earliest studies date just from 2009, the amount of literature on the subject is extensive. Table 1 classifies the studied measures under different criteria. First column splits the measures by activity, popularity and influence criteria. The reason for this classification is explained in Section 4. The second column presents each measure and its corresponding reference. Note that *General activity* is the only new measure proposed in this paper. In the third group of columns we mark as yes (✓) or no (✗) if the measures use metrics based on follow-up relationships (F), retweets (RT), mentions (M), replies (RP), favorites or likes (FT). All these kind of Twitter relationships are explained in Section 2. Furthermore, we distinguish those measures that are based on the PageRank algorithm, those that consider timeline, and those that require content analysis. The latter column shows the computational complexity for each measure. This column is explained later in Section 8.1.

As far as we know, this is the first attempt to bring together all centrality measures used in the Twitter network. However, there are also other interesting partial surveys in the area. For example, in 2013, Gayo-Avello [44] explained a dozen frequently used measures. These measures have been included here, along with other more than fifty. In a more recent article, Kardara et al. [62] also mention some few metrics and measures. However, their main contribution is to propose a two-dimensional taxonomy to classify centrality measures: On the one hand, the authors classify the measures at *global*, *local*, or *glocal*, where *local* is our topical-sensitive criterion (see Table 1); on the other hand, they classify the measures in *graphical* (those based on topological criteria), *contentual* (those based on content analysis) and *holistic* (an hybrid of the previous two). That classification is not as complete as the one given for Twitter in this paper, but is also suitable for non-Twitter measures. More common are the surveys related with centrality measures for general OSNs. Probst et al. [102] try to classify different types of centrality criteria. In addition to our criteria of activity (“how active one is”) and popularity (“how known one is”), they consider other two: “who one is” and “what one knows”. This classification is not clear enough to distinguish the influence criterion, which is our main interest. Vogiatzis [139] also considers a few centrality measures aimed at different OSNs. Regarding Twitter, the author focuses on the same kind of relationships we consider: retweets, mentions, replies, etc. Finally, Sun and Tang [128] make a general summary about the social influence analysis for social networks. Regarding centrality measures, they only mention the most traditional ones, such as degree, closeness and betweenness.

In the next section we present the main concepts, characteristics and metrics regarding the Twitter network. In Section 3 we explain how to obtain the data through the Twitter API, and we present the anatomy of a tweet. In Section 4 we talk about the ambiguity of the influential user’s definition, and we

¹ www.twitter.com

	Name	Ref.	Kind of metric used					PageRank	Timeline	Content analysis	Time complexity
			F	RT	M	RP	FT				
Activity	Tweet count score	[95]	X	✓	X	X	X	X	X	X	$O(T)$
	General activity		X	✓	✓	✓	✓	X	X	X	$O(T)$
	Topical signal	[99]	X	✓	X	✓	X	X	X	✓	$\Omega(T)$
	Signal strength	[99]	X	✓	X	✓	X	X	X	✓	$O(T)$
	Effective readers	[75]	X	✓	X	✓	X	X	✓	✓	limited request
	ActivityScore	[152]	X	✓	X	✓	X	X	✓	✓	limited request
	DiscussRank	[60]	X	✓	✓	X	X	X	✓	✓	limited request
	Competency	[2]	X	✓	X	X	X	X	✓	✓	? (high)
Popularity	IP Influence	[111]	X	✓	X	X	X	X	✓	✓	limited request
	FollowerRank	[92]	✓	X	X	X	X	X	X	X	$O(1)$
	Popularity	[2]	X	✓	X	X	X	X	X	X	$O(1)$
	Paradoxical discounted	[44]	X	✓	X	X	X	X	X	✓	limited request
	Network Score	[99]	X	✓	X	X	X	X	✓	✓	limited request
	A Score	[123]	✓	✓	✓	✓	X	X	✓	X	$O(T \cdot k)$
	AA Score	[123]	✓	✓	✓	✓	X	X	✓	X	$O(T \cdot k^2)$
	AAI Score	[123]	✓	✓	✓	✓	X	X	✓	X	limited request
Influence measures	Action-Reaction	[124]	✓	✓	✓	✓	X	X	✓	✓	$O(T \cdot k^2)$
	Starrank	[66]	X	✓	X	✓	X	✓	✓	X	$O(T + n^{2.3727})$
	Closeness	[50]	✓	X	X	X	X	X	X	X	limited request
	Betweenness	[61]	✓	X	X	X	X	X	X	X	limited request
	H-index	[55]	X	✓	X	✓	✓	X	X	X	$O(T + n)$
	Velocity	[43]	✓	X	✓	X	X	X	✓	X	$O(T)$
	Retweet Impact	[99]	X	✓	X	X	X	X	X	X	$O(T \cdot k)$
	Mention Impact	[99]	X	✓	X	X	X	X	X	X	$O(T \cdot k)$
	SNP	[3]	✓	✓	✓	✓	X	X	X	X	$O(T \cdot k)$
	Content & conversation	[51]	✓	✓	✓	✓	X	X	X	X	limited request
	IP Influence	[111]	✓	✓	X	X	X	✓	✓	✓	limited request
	TunkRank	[133]	✓	✓	X	X	X	✓	✓	X	limited request
	TrueTop	[155]	X	✓	X	✓	X	X	X	X	$O(T + n^{2.3727})$
	UserRank	[84]	X	✓	X	X	X	✓	X	X	limited request
	DIS	[58]	✓	✓	X	X	X	✓	X	X	limited request
	Influence Rank	[50]	✓	✓	X	X	✓	✓	X	X	limited request
	By similarity factor	[80]	✓	✓	✓	✓	X	✓	X	✓	limited request
	InfRank	[60]	✓	✓	X	X	X	✓	X	X	limited request
	LeadRank	[60]	✓	✓	✓	X	X	✓	X	X	limited request
	SpreadRank	[31]	✓	✓	X	X	X	✓	✓	X	$O(T + n^{2.3727})$
	ProfileRank	[117]	✓	✓	✓	✓	X	✓	✓	X	$O(T + n^{2.3727})$
	MultiRank	[30]	✓	✓	X	✓	X	✓	✓	✓	? (high)
	TURank	[149]	✓	✓	X	X	X	✓	X	✓	limited request
	By TAC	[81]	✓	✓	X	✓	X	✓	✓	X	limited request
	Alpha centrality	[97]	X	✓	X	X	X	X	✓	✓	$\Omega(T + n^{2.3727})$
	T-index	[67]	X	✓	X	X	X	X	X	✓	$\Omega(T + n)$
	Information Diffusion	[99]	✓	X	X	X	X	X	✓	✓	limited request
	Topic-Specific Author	[70]	✓	✓	✓	X	X	X	✓	✓	limited request
	By effective audience	[127]	✓	✓	X	✓	X	X	✓	✓	limited request
	TRank	[90]	✓	✓	X	X	✓	X	X	✓	limited request
	RetweetRank	[147]	X	✓	X	X	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	MentionRank	[147]	X	✓	X	X	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	TwitterRank	[144]	✓	✓	X	X	X	✓	X	✓	limited request
	InterRank	[129]	✓	✓	X	X	X	✓	X	✓	limited request
	Topic-Entity PageRank	[16]	X	✓	X	X	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	TIURank	[83]	X	✓	X	X	X	✓	X	✓	limited request
	ARI	[54]	X	✓	✓	✓	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	Twitter user rank	[95]	X	✓	✓	✓	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	TS-SRW	[63]	X	✓	X	X	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	Topical Authority	[57]	X	✓	X	X	X	✓	X	✓	$\Omega(T + n^{2.3727})$
	IARank	[17]	✓	✓	✓	✓	X	X	X	✓	$\Omega(T \cdot k^2)$
	SNI	[53]	✓	✓	✓	X	X	✓	✓	✓	$\Omega(T + n^{2.3727})$
	By polarity & others	[8]	✓	✓	✓	✓	X	X	X	✓	limited request
	By suceptibility...	[76]	✓	✓	X	X	X	X	✓	✓	limited request
	By tweets graph	[126]	✓	✓	✓	✓	X	X	✓	✓	? (high)
	WRA	[152]	✓	✓	X	✓	X	X	✓	✓	limited request
	FLDA	[5]	✓	X	X	X	X	X	✓	✓	? (high)
	Leadership	[2]	✓	X	X	X	X	X	✓	✓	? (high)
Predictive	AWI model	[151]	X	✓	X	X	X	X	✓	X	?
	ACQR Framework	[21]	✓	X	✓	✓	X	X	✓	✓	?
	TNIM	[32]	✓	X	X	X	X	X	✓	✓	limited request
	Author Ranking	[138]	✓	✓	✓	✓	X	X	✓	✓	?
	ReachBuzzRank	[120]	✓	✓	✓	✓	X	X	✓	✓	?
	IDM-CTMP	[78]	✓	✓	✓	✓	X	X	✓	✓	?
	Parameterless mixture	[14]	✓	✓	✓	✓	X	X	✓	✓	?

Table 1: Classification of measures in order of appearance in the article. F refers to follow-up relationships, RT to retweets, M to mentions, RP to replies, and FT to favorite or liked tweets. It is also indicated if the measure uses the PageRank algorithm, timeline, or topical analysis. In the latter column, T is the number of tweets, n the number of users, and k the length of an auxiliary vector (see Section 8.1).

divide the different measures according to user activity, popularity and influence. Section 5 is devoted to activity measures, Section 6 to popularity measures, and Section 7 to influence measures. In Section 8 we address the computational complexity of the measures, as well as some ways to correlate them with each other. Here we also introduce a case study in order to illustrate the variation of results for different centrality criteria. In Section 9 we mention the main open problems related to this topic. We finish this work with some conclusions.

	user	tweet
user	follows / is followed by	posts
	mention	retweets
	replies to	likes
	retweets to	replies
tweet	posted by	replies / is replied from retweets / is retweeted from
	retweeted by	
	liked by	
	replied by	

Table 2: Twitter relationships between users and tweets.

2 A brief description of the Twitter network

Nowadays, Twitter is one of the main online social networking services on the world.² It is a microblog service where users can send and read short 140-character messages called *tweets*. Tweets may include plain text, URL pages, images, mentions to other users (preceded by the symbol “@”) and *hashtags*, which are words that one decide to highlight by placing the symbol “#” in front of them. A *trending topic* is a term (one or more words) that in some moment appears in a huge amount of tweets of some place. These “places” and “moments” are known due to additional metadata contained in each tweet, which are related with geolocation, time and broadcaster account information, etc. Many times, trending topics are written as hashtags.

In Twitter there are four types of public relationships: user-to-user, user-to-tweet, tweet-to-tweet and tweet-to-user. The valid actions for each kind of relationship are shown in Table 2. Note that both the user-to-tweet and the tweet-to-user relationships are symmetric. The *mention* action was already described before. In Twitter, when a user A is following another user B , we say that A is a *follower* of B , and B is a *followee* of A . We can also *reply* a tweet of another user or, in other words, reply to another user through one of his tweets. Therefore, we can consider that the reply action belongs to the four kinds of relationships. Two very powerful actions are sharing, or *retweet*, the tweet of another user, and marking a tweet as *favorite* or *like*.³ Retweets and likes also involve an implicit interaction between the user’s action and the author of the original tweet. Since 2015, Twitter also provides the possibility of including text in retweets, so now we can both retweet and mention another user in the same tweet.

As any social network, Twitter can be represented by a digraph $G = (V, E)$, where V is the set of nodes and E is the set of directed edges that represent how the nodes are related. The most traditional view is to consider that V is the set of Twitter users, and the edges represent the follow-up relationship between users, so that an edge $(u, v) \in E$ means that user u is a follower (or a followee, depending on the model) of user v . In such a case, the interactions among users like retweets, replies, mentions, etc., are part of the implicit dynamics of the network. A second possibility is to consider a digraph with the same set of nodes, but where the edges now represent a dynamic interaction between users, like for instance a retweet from one user to some tweet of another user. A more complete way of representation that may include the previous ones, requires a partition of the set of nodes into two sets V_1 and V_2 , where V_1 is the set of users of the network and V_2 is the set of all tweets posted by the users. In this case, edges can represent the four kind of relationships [149]. The edges on the last two kinds of graphs are usually weighted, in order to represent the volume or frequency of the different interactions. In what follows, we denote these three kinds of graphs as \mathcal{G}_1 , \mathcal{G}_2 and \mathcal{G}_3 , respectively. For simplicity, when a graph of kind \mathcal{G}_2 only considers, for instance, edges for retweets and mentions, we say that it is a “ \mathcal{G}_2 graph of retweets and mentions”. In Figure 2 we illustrate these three kinds of graphs. Of course, there can be many other variations. For instance, Xie et al. [148] use a \mathcal{G}_2 graph of reciprocal retweets, i.e., one that only has bidirectional edges that represent mutual retweets among users. Furthermore, a Twitter network can be

² There are many other microblog systems on the web. After Twitter, the most popular is the chinese Sina Weibo: www.weibo.com.

³ *Likes* are an evolution of the old *favorites* that were replaced at 2015. The differences between these two actions are very subtle, so for the purposes of the measures considered in this paper, we can consider them as analogous terms.

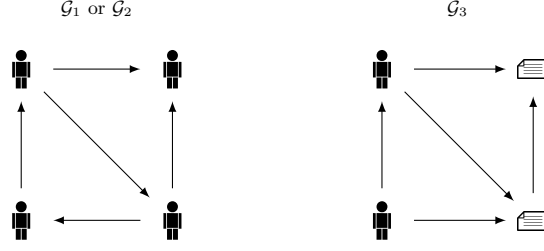


Fig. 2: Three different ways to represent Twitter as digraphs. The different kinds of relationships are described in Table 2. \mathcal{G}_1 considers only follow-up relationships, while \mathcal{G}_2 considers the more dynamic actions of mentions, replies and retweets. All the possible relationships can be represented by \mathcal{G}_3 .

represented by many other mathematical models, like multigraphs [60], Markov chains [120], edge-colored graphs as a particular kind of multilayer networks [96], among others.

2.1 Twitter metrics

We consider a *metric* as a simple mathematical expression that helps us to provide basic information about the social network in the form of a numerical value. In turn, metrics can be combined to define a (*ranking*) *measure*, i.e., either a formula or an algorithm that provides a criterion to rank each user of a network. Of course, there exist measures which are more complex than a simple combination of metrics.

Pal and Counts [99] suggest a set of metrics that involve original tweets, replies, retweets, mentions and graph characteristics. After that, other researchers have continued developing this metric system [153]. We list these metrics in Table 3, by adding new metrics based on favorites or likes. First column of the table represents the metric identification, and the second column describes the computation of the metric for one particular Twitter user. Third column is the time complexity required to compute each metric, considering as input a Twitter database. This latter column is explained later in Section 8.1.

Some of these metrics are relatively correlated. For instance, according to Ye and Wu [150], if $RT2$ has a high value, then $RT3$ usually tends to be high; however, $M4$ tends to be low, even if $M3$ is high.

Note that $F2$ and $F4$ introduce topical analysis that involves analyzing the content of the tweets. This can be done by using machine learning techniques and natural language processing. In Section 7.3 we survey the main influence measures related with topical analysis. The same metrics also refer to “active” users. As we shall see in Section 4, “activity” is an ambiguous concept, like “influence” and “popularity”. How often a user must tweet to be considered active? It depends on the context. As we mentioned, each tweet has its date of creation associated as metadata. Hence, we can construct a sequential timeline to see the order of the user’s posts. Many other metrics of Table 3 can also be restricted by topic and time [82].

According to Chorley et al. [23], metrics of retweets are the best quantitative indicators to prefer reading a tweet over another. This means that for a reader, a tweet being retweeted several times is more attractive than, for instance, a tweet with a lot of mentions. Despite of this, the most important indicators are qualitative, e.g., the friendship between the reader and the tweet’s author. These considerations further hinder the definition of influential users.

3 Twitter API

All the information required to compute these metrics is accessible through the Twitter API ⁴. An *API* (*Application Programming Interface*) is a set of functions, protocols and tools that are used to build an application, or to facilitate the communication with services. Twitter provides three kinds of API to developers (two of them are public) that together provide access to information from the social network. This information is obtained through different kind of requests, and it is useful for research, to search for historical or real-time information, and to develop unofficial clients.

⁴ <https://dev.twitter.com/>

ID	Metric description	Time complexity
<i>OT1</i>	Number of original tweets (OTs) posted by the author.	$O(T)$
<i>OT2</i>	Number of URL links shared by their OTs.	$O(T)$
<i>OT3</i>	Number of hashtags included in their OTs.	$O(T)$
<i>RP1</i>	Number of replies posted by the author.	$O(T)$
<i>RP2</i>	Number of OTs posted by the author and replied by other users.	$O(T)$
<i>RP3</i>	Number of users who have replied author's tweets.	$O(T \cdot k)$
<i>RT1</i>	Number of retweets accomplished by the author.	$O(T)$
<i>RT2</i>	Number of OTs posted by the author and retweeted by other users.	$O(T)$
<i>RT3</i>	Number of users who have retweeted author's tweets.	$O(T \cdot k)$
<i>FT1</i>	Number of tweets of other users marked as favorite (liked) by the author.	$O(T)$
<i>FT2</i>	Number of author's tweets marked as favorite (liked) by other users.	$O(T)$
<i>FT3</i>	Number of users that have marked author's tweets as favorite (likes).	$O(T \cdot k)$
<i>M1</i>	Number of mentions to other users by the author.	$O(T)$
<i>M2</i>	Number of users mentioned by the author.	$O(T \cdot k)$
<i>M3</i>	Number of mentions to the author by other users.	$O(T)$
<i>M4</i>	Number of users mentioning the author.	$O(T \cdot k)$
<i>F1</i>	Number of followers.	$O(1)$
<i>F2</i>	Number of topically active followers.	limited request
<i>F3</i>	Number of followees.	$O(1)$
<i>F4</i>	Number of topically active followees.	limited request
<i>F5</i>	Number of followers tweeting on topic after the author.	limited request
<i>F6</i>	Number of followees tweeting on topic before the author.	limited request

Table 3: List of some valuable Twitter metrics with their respective time complexity. T is the total number of tweets in the input, and k the length of an auxiliary vector of unrepeated elements (see Section 8.1).

3.1 API REST

API REST allows to read or write data on Twitter through simple HTTP primitives like *GET* and *POST*.⁵ For instance, if a GET is performed to the “<https://api.twitter.com/1.1/trends/place.json?id=1>” resource, then the API will return 50 global trending topics. We show in Listing 1.1 an extract of the answer to the previous query. The answer is presented in JSON format and it consists of key-value pairs. Note that the value for a key can be a list containing more key-value pairs.

Listing 1.1: Extract of JSON response

```

1  [
2    {
3      "as_of": "2012-08-24T23:25:43Z",
4      "created_at": "2012-08-24T23:24:14Z",
5      "locations": [
6        {
7          "name": "Worldwide",
8          "woeid": 1
9        }
10     ],
11     "trends": [
12       {
13         "tweet_volume": 3200,
14         "events": null,
15         "name": "#UserAccount",
16         "promoted_content": null,
17         "query": "%23UserAccount",
18         "url": "http://twitter.com/search/?q=%23UserAccount"
19       }, ...

```

The main limitation for obtaining data through the API REST is the limited number of times we can do each kind of request. These numbers are controlled by tokens that are reduced for each kind of request,

⁵ <https://dev.twitter.com/rest/public>

whenever a request of that kind is made. The amount of tokens to be performed may differ depending on whether we connect to the API with a user authentication or with an application authentication. Twitter provides to each developer a working window of 15 minutes, so if the tokens are depleted, we must wait to finish the current window, so that the number of tokens is restarted for another 15 minutes, and so on.

Twitter provides to the developers a list with the different types of requests and their associated tokens.⁶ For instance, the request `followers/ids` has 15 tokens, each of which returns a maximum of 5000 identifiers. Therefore, if we want to get the followers list of the Twitter account with the maximum number of followers to date (@KatyPerry, with 78,822,185 followers) we would need 15765 tokens, or 262.75 hours in the best case (Twitter does not ensure that the returned identifiers are always unique). After that, if we want more information about each follower, we must perform a GET to the resource `users/show`, which returns complete information about a user. An extract of the data available through the resource `users/show` is shown in Listing 1.2.

Listing 1.2: Extract of JSON user response

```

1  "favourites_count": 757,
2  "follow_request_sent": false,
3  "followers_count": 143916,
4  "following": false,
5  "friends_count": 1484,
6  "geo_enabled": true,
7  "id": 2244994945,
8  "id_str": "2244994945",
9  "is_translation_enabled": false,
10 "is_translator": false,
11 "lang": "en",
12 "listed_count": 516,
13 "location": "Internet",
14 ...
15 "statuses_count": 1279,
16 "time_zone": "Pacific Time (US & Canada)",
17 "url": "https://t.co/66w26cua10",
18 "utc_offset": -25200,
19 "verified": true

```

Given the limitations of the API REST, it is possible that the representation of the network as a graph turns out to be prohibitive, because of the time needed for each request. For instance, let us consider to build a \mathcal{G}_1 graph, assuming that each user has 5000 followers (so that for each user we only require one token to obtain all their identifiers). In this case, to obtain the second searching depth level, more than 83 hours are needed; for the third level, more than 17361 hours are needed, and so on, tending to an exponential curve. According to studies conducted by the consulting Beevolve and published in The Telegraph (2012) [131], an average user has only 208 followers. Even so, to obtain the follow-up relationships between users in a third depth level we require more than 721 hours.

3.2 Streaming API

The *Streaming API*, also known as *Firehose*, provides developers access to the global flow of tweets and other events in real time (deletions, retweets, replies, etc.).⁷ Unlike API REST, the Streaming API generates a persistent communication between the client and Twitter, which allows a continuous data stream. This API also provides various filters that allow to obtain data that only meet certain features (e.g., tweets sent from a certain geographical area, tweets that contain a specific bag of words, tweets containing hashtags, among others). Listing 1.3 shows an extract of a Streaming API response.

⁶ <https://dev.twitter.com/rest/public/rate-limits>

⁷ <https://dev.twitter.com/streaming/overview>

Listing 1.3: Extract of Streaming API response

```

1  "created_at": "Mon Nov 16 18:23:16 +0000 2015",
2  "id": 666320647509921794,
3  "id_str": "666320647509921794",
4  "text": "RT @ZeikBaard: Het is de schuld van de wiet mensen. We zijn er. #
      dwdd",
5  ...
6  "in_reply_to_status_id": null,
7  "in_reply_to_status_id_str": null,
8  "in_reply_to_user_id": null,
9  "in_reply_to_user_id_str": null,
10 ...
11 "user": {
12   "id": 39451098,
13   "id_str": "39451098",
14   "name": "Pika",
15   "screen_name": "TeshhCrack",
16   "location": "Somewhere over the rainbow",
17   ...

```

Years ago, Firehose allowed access to all Twitter data stream. However, since 2012 the new versions of the API began to restrict access to data [118]. Nowadays, the Streaming API gives access to about 1% of the total data.

Twitter originally sold tweets through three resellers: Gnip, Datasift and NTT Data. In 2014, Twitter acquired Gnip, and since then it sells tweets directly to its customers, exclusively through this service [56]. Currently, the cost of tweets is very high, and there are profitable companies outside Twitter that are engaged in selling tweets. As an approach, from a price quote requested by the authors at the end of 2015, we observed that an annual subscription for downloading paid tweets can reach \$39,000 dollars.

These restrictions have changed the case studies of researches after 2012. The articles before that date, could easily obtain large databases of tweets that today are very expensive. On the one hand, due to the huge daily flow of tweets, this restriction of 1% is still significant for case studies. On the other hand, these restrictions have led many researchers to turn their attention to Twitter analysis in real time. The latter, as discussed in Section 7, has also motivated new ways to measure the influential actors within the network.

3.3 Gnip

Gnip⁸ is the business platform provided by Twitter to access their data. Unlike the two APIs aforementioned, this API is not public and prices vary according to the needs of each client.

This API allows to work with the total flow of events taking place in the social network (full details in real time). It ensures reliability data capture, through different fault tolerance mechanisms, such as resuming flow in case of a shutdown, the ability to use redundant connections across replicated servers, among others. Furthermore, it also offers an API oriented to requests, which is able to obtain historical information (through the use of complex rules), and also gets full access to the public tweets file.

3.4 Anatomy of a tweet

Despite the difference between Gnip and the public APIs, the latter also provide a large amount of data. The data accessible through the Streaming API can be classified into three groups. The first group is directly related to the tweets received. It provides information such as their unique identifiers, date of creation, sources (mobile application, web application, unofficial client, etc.), references to other social network features (if the tweet is a reply, retweet, mention, etc.), language, location, among others. The complete list of attributes for a tweet is the following:

created_at	id	id_str	text
source	truncated	in_reply_to_status_id	in_reply_to_status_id_str
in_reply_to_user_id	in_reply_to_user_id_str	in_reply_to_screen_name	geo
coordinates	place	contributors	is_quote_status
retweet_count	favorite_count	favorited	retweeted
filter_level	lang	timestamp_ms	

⁸ <https://gnip.com/>

The second group refers to the user performing the action. The API returns a user object that provides his name, screen name, location, profile description, number of followers and followees, among others. The different attributes for a user are the following:

id	id_str	name
screen_name	location	url
description	protected	verified
followers_count	friends_count	listed_count
favourites_count	statuses_count	created_at
utc_offset	time_zone	geo_enabled
lang	contributors_enabled	is_translator
profile_background_color	profile_background_image_url	profile_background_image_url_https
profile_background_tile	profile_link_color	profile_sidebar_border_color
profile_sidebar_fill_color	profile_text_color	profile_use_background_image
profile_image_url	profile_image_url_https	profile_banner_url
default_profile	default_profile_image	following
follow_request_sent	notifications	

The third group refers to the categorization of entities and considers the tweet's content:

hashtags symbols urls user_mentions media

This anatomic description is based on the Streaming API. However, the information obtained from the API REST is quite similar. The main difference is the ability to get more information by performing historical searches according to some criteria and filters.

4 What is an influential user?

The problem of measuring the influence of a user in a social network is, in the first place, a conceptual problem. There is no agreement on what is meant by an influential user. Therefore, new influence measures are constantly emerging, each of which offers different measurement criteria.

The wide variety of influence criteria also involves the definition of new kinds of users who are closely related to influential users. Sometimes, influential users are also called *opinion leaders*, *innovators* [21], *prestigious* [44] or *authoritative actors* [14]. Occasionally, they have been associated with *topical experts* for specific domains [82,116]. In addition, there are other user classifications related to spread of influence. For example, some researchers distinguish between *opinion leaders*, *influencers* and *discussers* by type of activity and impact [60]. We can also distinguish between *inventors* (users who start a new topic) and *spreaders* (users who are responsible for disseminating that topic) [78]. Another classification considers *disseminators* (users who spread their influence and prevent the formation of structural holes), *engagers* (users who facilitate relationships with third parties) and *leaders* (top disseminator-engagers) [41]. We can also recognize *idea starters* (users with many followers) and *connectors* (users connecting starters) [126], *amplifiers*, *curators*, *commentators* and *viewers* [132]. Influential users can also be classified according to their content and authority [147].

Other relevant users are the *celebrities* [123], who meet somewhat different criteria from those of influencers. Users can be classified according to their popularity into *broadcasters* or *passive users* [109] (many followers and few followees), *acquaintances* (similar amount of followers and followees) and *miscreants* or *evangelists* (few followers and many followees, like spammers and bots) [71]. For some researchers, the previous notion of evangelists corresponds to *mass media* users, keeping the name *evangelists* to well connected opinion leaders, in contrast to normal users or *grassroots* [19]. Besides the above, some researchers distinguish between *popular*, *influential*, *listener*, *star* and *highly-read* users, considering not only the existing metrics but also the tweets' content, as well as both their cognitive and psychological behaviour [105].

On the negative side, *meat-puppets* are a negative evolution of *sock puppets* in OSNs [141]. Furthermore, *social capitalists* are users who only seek followers and retweets, but differ from spammers and bots [87] (see also Dugué and Perez [35]; Danisch et al. [27]). The spammer detection is a separate issue that has also received much attention on Twitter. Although this is not the main focus of the article, it is interesting to mention that some metrics can be used to recognize spammers on Twitter. For instance, if a user is retweeting or marking as favorite his own tweets, this could be considered an act of spam. Moreover, Pal and Counts [99] proposed an additional metric (not included in Section 2.1, because it is

more complex than the others) called the self-similarity score, that computes the similarity between the recent and previous tweets of a users. If the self-similarity is high, the user can be a spammer; if it is low, the user speaks about many topics and uses a large vocabulary. According to Gayo-Avello [44], a good measure should place the most relevant users at the top of the ranking, without caring about their exact location in the ranking. At the same time, that measure should punish the spammers, placing them in lower positions.

It is important to note that an influential user is not necessarily one who writes influential tweets [106]. According to Francalanci and Hussain [38], while influential users are typically associated with hub nodes, the influence can be spread through multi-layered peripheral node clusters. The tweet’s influence can be measured in terms of *extension* (number of users “affected” by the message), *intensity* (cognitive, emotional and conductual impact) [86], content, sender’s popularity, etc. In a sociological sense, some researchers argue that opinion leaders in Twitter are more associated with people with high self-esteem [59].

Li et al. [78] define a framework to observe the behavior of the influence between users. This observation considers three dimensions: the degree of similarity of the initiated topics, the latency between one tweet and the next on the same topic, and the tendency of a user to create new topics or spread already existing topics through the network.

The present list of user types is not exhaustive, but it allows us to realize that there is no standard definition to measure influence. Moreover, it seems reasonable to ensure that some of these types of users, such as celebrities and social capitalists, have no direct relation with influence criteria. In what follows, we propose to split the existing measures into three different kinds of criteria: *activity measures*, *popularity measures* and *influence measures*. Thus, for instance, the *RT2* metric seems to be much more proper to measure influence than *F2* [143], while *F1* and *F3* are useful metrics to measure popularity [120]. This classification is valid because there is no direct correlation between active, popular and influential actors [111].

5 Activity measures

We say that users are *active* when their participation in the social network is constant and frequent in a period of time, regardless of the attention they receive for their participation. Note that there may be very active Twitter’s readers, who however cannot be observable by any metric, because they do not leave any trace on the network. Therefore, by “participation” we mean to do actions that can be measured, like doing tweets, retweets, mentions, replies, etc. In this sense, Yin and Zhang [151] define the activity of a user as the probability of a user seeing a tweet. We cannot determine when a user has seen a tweet, but if a user retweets a tweet, we can assume that the tweet has been read. Therefore, more active users are more capable to see new tweets, and thus interact with them.

Perhaps the simplest activity measure is the *TweetRank* [92], which is just a metric that counts the number of tweets of the user. Slightly more sophisticated is the *Tweet count score* [95], that counts the number of original tweets plus the number of retweets. Following this approach, a reasonable activity measure could be the sum of the visible actions of each user. Thus, for every user i we define the *General Activity* as follows:

$$General\ Activity(i) = OT1 + RP1 + RT1 + FT1$$

We assume that metric $M1$ is contained in the other metrics. We can normalize this measure by dividing the value by the total number of the considered tweets. Furthermore, we can restrict our attention to the tweets related with some specific topic. This is what the *topical signal* (TS) [99] does, in spite of it omits the likes:

$$TS(i) = \frac{(OT1 + RP1 + RT1) \Big|_{\text{specific topic}}}{\#tweets}$$

The same authors define a *signal strength* (SS) [99], that indicates how strong is the author’s topical signal:

$$SS(i) = \frac{OT1}{OT1 + RT1}$$

This expression measures the originality of the author’s tweets, in such a way that a greater authorship means values closer to 1.

There are also measures that, although were defined as influence measures, under our classification they fit better as activity measures. Such is the case of the sum of *effective readers* for all user tweets [75], where an effective reader of a tweet is a follower who still has not tweeted on any trending topic when the user sent the tweet. In some way, it measures the speed of a user to tweet about new topics. Note that despite being one of the oldest measures (it was published in 2010), it already takes into account the timeline of tweets. Note also that, unlike previous measures mentioned above, it cannot be calculated immediately, so it is more suitable for offline analysis. This is also the case of all the remaining measures in this section. We address the computational complexity aspects of measures in Section 8.1.

The *ActivityScore* [152] also considers the timeline. This measure counts the number of followers, followees and tweets on a \mathcal{G}_3 graph for each user during a period of time. The *DiscussRank* [60] determines how active a user is, in the sense of initiating conversations around a topic. This algorithm differs from the others in that it is based on multigraphs (i.e., graphs with multiple edges between two nodes), whose nodes are the users, and the edges are based on followers, retweets and mentions among users. An additional topical-sensitive measure is the *Competency* [2], which ranks the actors according to his ability to post relevant tweets according to hot topics. The topical detection is given by a Latent Dirichlet Allocation (LDA) algorithm, and the topical relevance of each tweet is determined by a Divergence From Randomness retrieval model. One problem of Twitter measures based on LDA is that the traditional LDA algorithms are used for larger texts than tweets, so it is necessary to use specific variations [158].

Finally, the *IP Influence* [111] not only measures the users' influence (see Section 7), but also their passivity. The *passivity* of a user is defined as the difficulty for the user to be influenced by another in some period of time. It is measured by considering metrics of retweets, followers and followees. Most social networking users are not active, but passive, in the sense that they do not cooperate with the spread of content over the network. According to the authors, most users with very high passivity tend to be spammers or bots. This notion of passivity is also used in the *influence graphs*, where it is displayed on the labels of the nodes [89]. The IP Influence was criticized by Gayo-Avello et al. [43] because it does not consider the followers, and it was also disproved that it really correlates with the number of clicks on URLs.

6 Popularity measures

Briefly speaking, we can say that a user is *popular* when he is recognized by many other users on the network. An example of a popular user is a celebrity, who does not necessarily have an active and influential account. For instance, the Clint Eastwood's account (@Eastwood_) has more than 60,000 followers, but no followees, no tweets.

The simpler popularity measures just count the follow-up relationships between users. The *Followers-Rank* [92], also known as *Structural Advantage* [17], is the normalized version of the traditional *in-degree* measure [50,61] for social networks in general:

$$FollowerRank(i) = \frac{F1}{F1 + F3}$$

There are variations of this measure, as the *Twitter Follower-Followee ratio (TFF)* [8] that corresponds to $F1/F3$. The disadvantage of these measures is that $F1$ and $F3$ metrics may differ a lot, and the number of followers for some Twitter users is too high compared to the rest. To mitigate these differences, the *Popularity* [2] was defined as follows:

$$Popularity(i) = 1 - e^{-\lambda \cdot F1}$$

where λ is a constant that by default can be equals 1. As both metrics $F1$ and $F3$ can be obtained directly from the API REST, all the previous popularity measures can be computed in real time. However, note that with few changes, these measures lose their efficiency and become on measures restricted to offline environments. This is the case of the *Followers to followee ratio with paradoxical discounted reciprocity* [44], that introduces an additional metric, namely the number of reciprocal actors of a user, i.e., the number of followers who are also followees:

$$Paradoxical\ discounted(i) = \begin{cases} F1/F3 & \text{if } F1 > F3 \\ \frac{F1 - \text{reciprocal}(i)}{F3 - \text{reciprocal}(i)} & \text{otherwise} \end{cases}$$

The purpose of this measure is to punish the spammers (users with many followees and few followers). However, computing the $\text{reciprocal}(i)$ metric is not as simple as computing $F1$ and $F3$. Indeed, it considerably increases the computational costs, as we mention in Section 3.1. A simple topic-sensitive popularity measure, also based on complex metrics (see Table 3), is the *Network Score* (NS) [99], which is based on the active non-reciprocal followers of the user:

$$NS(i) = \log(F2 + 1) - \log(F4 + 1)$$

All popularity measures seen so far consider metrics based on follow-up relationships. However, popularity measures may also consider other kind of metrics. A good example is given by three measures defined by Srinivasan et al. [123]. Each one of these measures is based on the previous one, in addition to other metrics. Despite the variety of metrics used for the first two, only the third one becomes more complex, because of the API limitations discussed above. The first one is the *Acquaintance Score* $A(i)$ that measures how well-known user i is. Let n be the number of considered user accounts, it is defined as:

$$A(i) = \frac{F1 + M4 + RP3 + RT3}{n}$$

The second measure is the *Acquaintance-Affinity Score* $AA(j)$ that measures how dear user j is, by considering how well known are those who want him:

$$AA(j) = \sum_{i \in E_{RP}} A(i) \cdot \frac{\# \text{replies of } i \text{ to } j}{\# \text{replies of } i} + \sum_{i \in E_M} A(i) \cdot \frac{\# \text{mentions of } i \text{ to } j}{\# \text{mentions of } i} + \sum_{i \in E_{RT}} A(i) \cdot \frac{\# \text{retweets of } i \text{ to } j}{\# \text{retweets of } i}$$

where E_{RP} , E_M and E_{RT} are the set of users who reply, mention and retweet the tweets of j , respectively. Finally, the *Acquaintance-Affinity-Identification Score* $AAI(j)$ measures how identifiable user j is, by considering how dear are those who identify him:

$$AAI(j) = \sum_{i \in F_r} \frac{AA(i)}{\# \text{followees of } i}$$

where F_r is the set of followers of j . The AAI Score is well correlated with the $F1$ metric, and was used to identify celebrities in the “real world”, i.e., outside the Twitter network [124]. After that, the same authors defined the *Action-Reaction* (AR) [124] to identify celebrities within the Twitter network. It is a combination of two measures: the *Action score* that measures the loyalty of the user’s “fans”, and the *Reaction score* that measures the attention aroused by the celebrity for his actions. This measure uses conditional probability variables based on replies, mentions and retweets, and successfully avoids the complexity limitations of the AAI Score.

We finish this section with a measure based on the *PageRank*, a widely used algorithm to define centrality measures, and specially influence measures, as we will see in the following section. The *Star-rank* [66] uses PageRank to make a daily dynamic analysis in a \mathcal{G}_2 graph of mentions. It considers metrics such as the acceleration of mentions over time. This popularity measure is interesting, because it does not consider explicitly the follow-up relationships, but the times the user is mentioned by other Twitter accounts.

7 Influence measures

In this section we survey the different influence measures for the Twitter network, which is the main purpose of this paper. We say that a user is *influential* whether his actions in the network are capable to affect the actions of many other users in the network. Since we are focused on a microblog service, we can also understand that influential users are the most able to spread information within the network [91]. The influential users tend to be active [20,66], but only a few active users are influential. Furthermore, it is important to understand that influential users in a social network are not necessarily influential in the real life [24,25].

According to Quercia et al. [105], there are two broad paradigms of social influence. The first point of view believes that the influence is often massively exerted by a small number of very persuasive or connected users. The second point of view believes that many users can be influential accidentally,

depending on many unpredictable factors. Despite the second paradigm, the authors argue that for social networks like Twitter it is possible to develop several influence criteria that can be quantified. Morone and Makse [91] are inclined to the first of these two points of view, although they suggest that for complex networks, the influential actors also respond to unpredictable criteria, and usually are not those who are better connected. Borge-Holthoefer et al. [13] also agree with the first point of view, but in addition, they support the possibility that the influence accumulated in smaller subnetworks, formed by less influential users, could explode with a cascading behaviour, affecting the whole network.

7.1 Traditional measures used on Twitter

In Section 2.1 we mentioned several metrics that can be determined from the particular dynamics of the Twitter network. The Twitter influence measures in literature usually consider metrics related to retweets, mentions and, to a lesser extent, followers [20]. However, some researchers have used traditional centrality measures that are not based on Twitter metrics at all, but only on the passive topology of the \mathcal{G}_1 graphs. That is the case of *Closeness* (C_c) and *Betweenness* (C_B) [50,61]. Closeness centrality is based on the length of the shortest paths from a node i to everyone else. It measures the visibility or accessibility of each node with respect to the entire network. Let D be the distance matrix of a network with n nodes, it is defined as follows [74]:

$$C_c(i) = \frac{n-1}{\sum_{i \neq j} (D)_{ij}}$$

If there is no path from i to j , then we assume that $(D)_{ij} = n$. Betweenness centrality considers for each node i all the shortest paths that should pass through i to connect all the other nodes in the network. It measures the ability of each node to facilitate communication within the network. Let b_{jk} be the number of shortest paths from node j to node k , and b_{jik} the number of those shortest paths that pass through node i , then [74]:

$$C_B(i) = \frac{1}{(n-1)(n-2)} \sum_{j \neq k} \frac{b_{jik}}{b_{jk}}$$

A relatively recent measure is the *Hirsch Index* or *H-index* [55], originally defined to measure the productivity of a person in the scientific community, according to the citations of his articles. In the context of Twitter, it can be defined as the maximum value h such that h tweets of the user have been replied, retweeted, or liked, at least h times. A simplified version for Twitter was proposed by Romero et al. [111], where instead of replies, retweets and likes, they only consider retweets of tweets that contain URLs. Razis and Anagnostopoulos [109] consider two additional versions: one based on retweets, and the other on favorites or likes. In order to do daily analysis reports, the authors take, for both measures, only the last one hundred daily tweets from each user.

Several influence measures that we consider in the next section are based on *PageRank* [98], a well known algorithm used to measure both the relevance and presence of websites on the Internet. This algorithm considers the hub users as important nodes within the network. PageRank is a variation of another traditional centrality measure known as *eigenvector* [73], that favors users who are well connected with other well connected users within the network [128]. An alternative algorithm to PageRank is *HITS* [69]. While the former is executed at indexing time, the latter is executed at query time. Moreover, HITS provides two closely related scores: an authority score and a hub score. Both algorithms have been used repeatedly in the context of Twitter [133,72,22,70,60]. PageRank has been applied on \mathcal{G}_2 graphs of retweets and replies, for dynamic analysis over different periods of time [156]. Both PageRank and HITS have been applied on \mathcal{G}_2 graphs of retweets, replies and mentions for networks with sparse followers [104]. In this latter case, authors also use, as an additional method, a decision tree classification model that classifies the nodes after a supervised training. Another somewhat less known algorithm is *NodeRanking* that, unlike PageRank, it can be used on graphs with weighted edges [103]. For graphs without weights, NodeRanking and PageRank are almost equal [44]. As PageRank, HITS and NodeRanking are recursive algorithms, and the Twitter network is so dynamic, measures based on these algorithms do not allow real-time analysis [17]. Despite of this, the smart idea behind the PageRank algorithm has been applied on several Twitter influence measures for offline analysis. The HITS algorithm, instead, is not advisable to be used on OSNs, since it gives too much power to spammers [44].

Besides the traditional centrality measures, some authors have been inspired on other disciplines different from social network analysis and computer science. Thus, Gayo-Avello et al. [43] proposed an influence measure for Twitter based on the variable-mass system of physics. Here, the user score is represented by the velocity of an object over a period of time:

$$v_t = v_{t-1} + \frac{F_a}{m} - c$$

where time t is taken per each 1 hour (although in practice it can be calculated almost in real time), the force applied F_a is the number of mentions made to the user during the last hour, the mass m is the number of followers of the user (this value can be smoothed with a logarithm), and c is a constant that can be zero. Negative velocities are also considered zero. The authors found a positive correlation with the number of clicks on URLs, analogous to IP Influence, PageRank and TunkRank (a direct adaptation of PageRank that we consider later). This correlation increases when the number of followers is adjusted.

We finish this section by mentioning three traditional methods used to deal with complex networks. The first method is the k-shell decomposition algorithm, that has been applied to identify influential actors in epidemic dynamics [68]. Although this algorithm has been used in Twitter directly [101], Borge-Holthoefer and Moreno [12] suggest that it does not work to measure influential spreaders in rumor dynamics. Through numerical simulations, they notice that well connected users may already know a recent news but are not willing to spread it anymore. A variation of this algorithm to identify influential users in a \mathcal{G}_1 graph was proposed by Brown and Feng [15]. The second method is the particle swarm optimization (PSO) algorithm, that has been used with metrics of followers and retweets [154]. Finally, the *F-measure* [137], used in statistical analysis for binary classification, can be applied to study the power of retweets, by comparing the follow-up relationships and different centrality measures, such as betweenness, closeness, eigenvector and PageRank [142].

7.2 Measures based on Twitter metrics and PageRank

Two simple measures based on the metrics of Table 3 are *Retweet Impact (RI)* and *Mention Impact (MI)* [99]. The first one estimates the impact of the user tweets, in terms of the retweeted tweets:

$$RI(i) = RT2 \cdot \log(RT3)$$

The logarithms moderate the impact of overly enthusiastic users who retweet the same content many times. This measure could be easily generalized, by considering the same expression for the *FT2* and *FT3* pair of metrics. The second measure estimates the impact of the user tweets, in terms of the mentions received by other users:

$$MI(i) = M3 \cdot \log(M4) - M1 \cdot \log(M2)$$

The subtraction ensures that the mentions are given by their own merit, and not because the user has been mentioning other users. A third measure based on simple metrics is the *Social Networking Potential (SNP)* [3], defined as follows:

$$SNP(i) = \frac{Ir(i) + RMr(i)}{2}$$

where the *Interactor Ratio*, $Ir(i)$, and the *Retweet and Mention Ratio*, $RMr(i)$, are defined as:

$$Ir(i) = \frac{RT3 + M4}{F1} \quad \text{and} \quad RMr(i) = \frac{\# \text{tweets of } i \text{ retweeted} + \# \text{tweets of } i \text{ replied}}{\# \text{tweets of } i}$$

For each user, $Ir(i)$ measures how many different users interact with that user, while $RMr(i)$ measures how many of his tweets imply a reaction from the audience.

Note that the SNP measure considers all kind of actions on Twitter, except the favorites or likes. The same occurs for a measure based on *content* and *conversation*, informally defined by Hatcher et al. [51]. The content criterion considers the number of published tweets and follow-up relationships, and it has the 25% of the total importance. The conversation criterion considers the number of replies, as well as the number of followers related with the user through mentions and retweets, and it has 75% of the total importance of the measure.

Besides the previous measures, which are based directly on simple metrics, some authors have studied user influence in terms of their ability to post websites that are propagated through the network. This can be implemented using PageRank [4] and HITS. The *IP Influence* [111] (also mentioned in Section 5) uses HITS in this context, in order to analyze the influence and passivity (based on retweets) of users over time.

The remainder of this section is devoted to several influence measures based on PageRank. The first direct adaptation of this algorithm into the context of Twitter was *TunkRank* [133]:

$$TunkRank(i) = \sum_{j \in \text{followers}(i)} \frac{1 + p \cdot TunkRank(j)}{\# \text{followees of } j}$$

where, analogously to PageRank, $0 \leq p \leq 1$ is the probability that a tweet is retweeted. This probability is assumed to be equal for all users. In the literature, the authors normally use $p = 0.5$, but in fact this value should vary from case to case. For instance, for experiments considered by Gayo-Avello [44], the ideal value was $p = 0.287$. A variation of TunkRank is *UserRank* [84], defined to measure the influence of a user according to the relevance of his tweets:

$$UserRank(i) = \sum_{j \in \text{followers}(i)} \frac{1 + \frac{\# \text{followers of } i}{\# \text{tweets of } i} \cdot UserRank(j)}{\# \text{followers of } j}$$

Note that these measures changes the denominator $\text{followees}(j)$ by $\text{followers}(j)$. As PageRank, TunkRank is correlated to the number of clicks on URLs [43]. To penalize the ranking of spammers, Gayo-Avello [44] proposed, for each user i , to multiply $\text{PageRank}(i)$ by $\text{Paradoxical_discounted}(i)$ (see Section 6). This could also be done with the TunkRank. Indeed, as TunkRank is an adaptation for Twitter, this measure tends to penalize spammers more than PageRank [44].

Besides spammers, there are users using multiple false accounts. These accounts can be used to interact among themselves, and thus to improve their influence scores. The *TrueTop* [155] measure aims to be resilient to sybil attacks. It applies weighted eigenvector centrality (recall that PageRank is an evolution of eigenvector centrality) in a \mathcal{G}_2 graph of retweets, replies and mentions, where the edge weights represent the number of such user interactions. The algorithm allows to split the graph in a region with sybil users and another region without them, so that the users in the second region receive a higher score in the ranking.

The following measures strongly depend on follow-up relationships among users, so that the acquisition of necessary data is restricted to the limitations of Twitter API. The *diversity-dependent influence score (DIS)* [58] considers a variation of PageRank on a \mathcal{G}_2 graph of retweets and following. To favor a deeper influence spread, it gives more power to those users that are capable to influence (according to their follow-up relationships) the less dense users, i.e., those with a “high diversity”. The authors also proposed a variation in the definition of “diversity”, allowing a dynamic computation of the measure. The *Influence Rank (IR)* [50] combines follow-up relationships, mentions, favorites and retweets, to identify opinion leaders who are capable of influencing other influential users. Once the graph is constructed, it can be implemented as a polynomial approximation algorithm based on PageRank. At last, Li et al. [80] proposed that a user A can exert more influence over another user B if A writes tweets more related with those that B often writes. With this idea, the authors use a variation of PageRank based on a similarity factor between published tweets, on a \mathcal{G}_2 graph of following, retweets, mentions and replies. This latter idea brings new considerations to state the influence of users. However, this by itself does not seem to be enough. Due to the spontaneous phenomenon of viral content, it makes sense to think that a user can be influential without reading constantly what others write. For some authors, the user profile and the style of the tweets can be an additional issue to be considered. In this line, Ramírez-de-la-Rosa et al. [107] proposed 23 features of user profile (presence of hashtags, URLs, self-mentions, number of followers and tweets, etc.) and 9 features of tweet’s style (extension, frequency, quality, number of retweets, etc.) that for them are involved in influence capacity.

We continue with two measures called *InfRank* and *LeadRank* [60] InfRank is a variation of PageRank that measures the user influence in terms of his ability to spread information and to be retweeted by other influential users. The algorithm operates on a multigraph whose nodes are the users and the edges are based on retweet relationships. LeadRank measures the leadership of a user, in terms of his ability

to stimulate retweets and mentions from other users on the network, especially other leaders. This algorithm operates on a similar multigraph, but here the edges can represent retweets and mentions. Although these measures do not consider the follow-up relationships, they are also restricted by the Twitter API limitations.

Until now we have not seen any measure based on PageRank that avoids the limitations of the Twitter API. Two measures that bypass these limitations are *SpreadRank* [31] and *ProfileRank* [117]. The first one measures the spreadability of users on a \mathcal{G}_2 graph of retweets, whose edges are weighted by the proportion of retweets with respect to the total tweets of the retweeted user. The authors suggest that spreading influence is greater the faster their tweets are retweeted. They also consider information cascades, which can be described as a tree that grows over time, and whose nodes are the retweets of an original post. The closer the user is to the root node (earlier in time), the greater his influence spread. Therefore, the measure takes into account both the time interval between retweets and the location of users in the information cascade. In turn, the *ProfileRank* considers the influencers as users that can generate relevant content to other users. Inspired on PageRank, it is computed by random walks on a bipartite \mathcal{G}_3 graph whose edges represent generation and consumption of content by users over time.

Some ideas mentioned above have also been used by other authors. Valiati et al. [136] adapt PageRank on bipartite \mathcal{G}_3 graphs of retweets. Ding et al. [30] define *MultiRank*, a more complex measure inspired on PageRank that considers random walks on \mathcal{G}_2 graphs of retweets, replies, and two additional relationships, namely “reintroduce” and “read”. “Reintroduce” refers to post tweets which are similar to others previously posted by another user, but without acknowledging that user as the source. “Read” refers to the probability to read tweets published by other users, according to the appearance of such tweets in their timelines. In addition, the authors characterize the spammers in order to rank them below the common Twitter accounts.

Finally, there are influence measures that extend the PageRank algorithm. *TURank* [149] is based on *ObjectRank*, which is indeed an extension of PageRank. The algorithm uses a \mathcal{G}_3 graph of following, tweeting and retweets. From a similar graph that also allows replies, Liu et al. [81] add a “time-effectiveness attenuation coefficient” (TAC) that returns the time on which tweets are published, so that the earliest tweets gradually lose relevance in the measure. Again, these measures are restricted by Twitter API limitations.

7.3 Topical influential users

Several authors deal with the problem of identifying influential users regarding some specific topic. For these cases we need to analyze the content of the tweets. Traditional LDA algorithms and topical modeling tools are often more helpful in larger texts than tweets, so for this context it is necessary to use specific variations of the algorithms [158]. These computational techniques are not the purpose of this survey, so they will not be detailed here. Somehow, topical-sensitive measures are a kind of local measures, not in a topological sense but in a semantic level. According to Kardara et al. [62], these measures are more effective and functional than global measures, i.e., those that are not restricted to a topical domain.

These measures are usually used in offline analysis. A volume of tweets is collected by streaming over a certain period of time. During or after obtaining a database, tweets can be filtered to leave only those related to a specific topic. By using this method, Schenk and Sicker [114] analyzed the variation of followers who posted tweets about a specific natural disaster. For the authors, the most influential users are those with a greater positive variation.

At least two traditional centrality measures can be easily adapted as topical-sensitive measures for Twitter. The *alpha centrality* [11], based on eigenvector and the time as an additional parameter, has been adapted to consider retweets [97]. In turn, the *T-index* [67] is a topical variation of *H-index* and also considers retweets.

We found only one measure defined as a combination of metrics of Table 3. The *Information Diffusion* [99] estimates the possible influence of the user’s tweets, among his followers who are non-followees:

$$ID(i) = \log(F5 + 1) - \log(F6 + 1)$$

The “+1” in the logarithms avoids divisions by zero, in case we want to normalize the measure. This measure only considers follow-up relationships, but it is independent of the number of followers and followees on the network. Besides *F5* and *F6* metrics, *F2* and *F4* are also topical-sensitive. Therefore, every measure based on some of these metrics is affected by Twitter API limitations.

There are other measures affected by these limitations. The *Topic-Specific Author Ranking* [70] considers the quality of the tweets posted on a given topic. The quality of each tweet is measured in terms of its interest in the network, by using metrics of mentions and retweets. The algorithm is applied on a \mathcal{G}_3 graph of following, tweets and retweets (similar to that used in TURank). The implementation uses the MapReduce framework, and it is compared with some rankings based on PageRank and HITS. On the other hand, Sun et al. [127] define six measures based on effective audience for each user. The effective audience is quantified in terms of follow-up relationships, retweets and replies. Two of these measures consider a topical analysis, and another provides a dynamic value that depends on time. The first one has a high correlation with PageRank on \mathcal{G}_1 and \mathcal{G}_2 graphs of retweets, with the $F1$ metric, and especially with the $RT1$ metric.

Instead of consider the full text of the tweets, some measures restrict their attention to the hashtags. Such is the case of *TRank* [90], that ranks users in three dimensions by considering follow-up relationships, retweets and favorites. They also consider human factors, such as the fact that a user with many active followers may simply not be able to read all the tweets they publish. Other examples, without API limitations, are *RetweetRank* and *MentionRank* [147], restricted to hashtags of news events. Under the assumption that interaction with influential users makes you more influential, they extend PageRank on \mathcal{G}_2 graphs of retweets and mentions, respectively.

As expected, PageRank can also be applied in this context. Cataldi and Aufaure [18] proposed a direct topic-oriented adaptation of PageRank on \mathcal{G}_2 graphs of retweets, that requires the number of users of the network.

The *Topic-sensitive PageRank (TSPR)* [52] was created as an extension of PageRank for the original purpose of ranking websites. The *TwitterRank* [144] uses a similar idea as TSPR for the context of Twitter. It is applied on \mathcal{G}_1 graphs and considers criteria of similarity between users, restricted to the topics of their tweets. The TwitterRank algorithm is relatively slow, because it requires not only an iterative process like PageRank, but also a preprocessing of topical analysis. Gayo-Avello [44] proposed a more efficient version that does not consider topical analysis, but it has the problem of ranking spammers with high scores. As an alternative to TwitterRank, the *InterRank* [129] also uses PageRank (indeed, both measures are highly correlated) on a \mathcal{G}_1 graph with similarity criteria, but it does not require predefined topics.

There also exist extensions of TSPR and TwitterRank. Furthermore, TSPR has been extended to study the semantics of retweet and mention relationships [143]. The *Topic-Entity PageRank* [16] looks for influential users by using both topics and entities (e.g., a location, a person, a product, etc.) on a \mathcal{G}_2 graph of retweets. The disadvantage of this measure is that the entities are not always available for all users. Another measure is the *TIURank* [83]. Besides retweets, it also considers the frequency of retweets and strength relationship between users. This strength relationship is estimated by its similarity and a poisson regression-based latent variable model. As it is expected, this measure is also restricted by API limitations. Indeed, it was tested only with small databases. An additional extension is the *Author-Reader Influence (ARI)* [54] model that uses \mathcal{G}_2 graphs, whose edges represent the probability that a tweet posted by a user can be read or shared by another user (by mean of retweets, mentions or replies). The ARI measure considers the influential users as authors of attractive content (relevant and unique), which in theory will be read and replicated by others.

We continue with three additional measures that, like the ARI measure, use PageRank in an efficient way, together with other tools that already exist. The first one is the *Twitter user rank* [95], that uses Twitter keyword search (TURKEYS) with retweets, replies and mentions relationships. It is a combination of the Tweet count score activity measure (see Section 5) and the *User influence score (UI)*, which is a mechanism based on PageRank that can also use other algorithms such as HITS or SALSA [77]. Alternatively, this measure can also use the tweets dispersion in the network. Then we have the *Topic-Sensitive Supervised Random Walks (TS-SRW)* [63], that considers a \mathcal{G}_2 graph of mentions and random walks. Unlike the ProfileRank activity measure (see Section 7.2) that also uses random walks, TS-SRW uses similarity of content to support specific topics. Thirdly, we have the *Topical Authority* [57], calculated on a \mathcal{G}_2 graph of retweets, where the weights of the edges represent the relevance of the content of each retweeted tweet with respect to the chosen topic. This relevance can be determined by any text retrieval model. For their experiments, the authors used the Okapi BM25 ranking function. Additionally, we must mention the *Information Amplification Rank* or *IARank* [17], an algorithm slightly weaker than PageRank. In addition to supporting specific events, this is one of the few influence measures that can run in near real time. It corresponds to a linear combination of the FollowerRank popularity measure

(see Section 6) and the *Buzz*, given by the number of mentions to the user, divided by its total activity (tweets, retweets and replies) related with a particular event.

In the remainder of this section we include measures that require more complex algorithms. Only the following avoids the Twitter API limitations. For these measures, the algorithms are not explicitly given. The *Social Network Influence (SNI)* [53] is a measure comparable with *Klout*, an online influence measure that we shall see in Section 7.5. It is a mixture of several tools. It considers timeline and specific topics, and is calculated from a \mathcal{G}_2 graph of retweets and mentions. Furthermore, it is defined by a formula with adjustable parameters, and considers the probability that a user can influence his followers. In addition, it uses betweenness centrality and PageRank. All in all, it considers the network topology, some activity criteria, and parameters of the user’s profile.

Bigonha et al. [8] measured user influence as a combination of centrality, polarity of the users, and quality of their tweets. They used a centrality measure that combines normalized versions of betweenness, eigenvector, in-degree and Twitter Follower-Followee ratio (TFF) on a \mathcal{G}_1 graph and a \mathcal{G}_2 graph of mentions, replies and retweets. Polarity is measured by a sentiment analysis that classifies users into evangelists (those with more positive tweets), detractors (those with more negative tweets) and irrelevants (those with more neutral tweets). Finally, the content quality can be determined by using the *Kinkaid factor* [7], a particular formula based on the number of words, sentences and syllables. Lee et al. [76] also use sentiment analysis, but in this case for the retweeted tweets. With this and the follow-up relationships, the authors can determine the user influence in terms of their susceptibility and cynicalness.

Sun and Ng [126] constructed a graph whose nodes are tweets, such that the edges are weighted according to explicit relationships (retweets and replies) and implicit relationships (similarity of content). After that, the most influential tweets are measured on different topics using degree centrality, shortest path-cost, and *graph entropy*, the latter based on Dehmer [28]. From the first graph, a second graph of users is extracted, and from here the corresponding most influential users are chosen through algorithms that are able to distinguish between starters and connectors. These algorithms have been applied in online networks with relatively few nodes.

Besides the above, there are more measures that reuse other measures that already exist. The *Weighted Ranking Algorithm (WRA)* [152] is a linear combination of the *ActivityScore* and another measure called *QualityScore*, which considers the tweets (filtered by some topic) posted by users that follow another user, the number of tweets of these users, and the probability that tweets are retweeted. WRA is applied on a \mathcal{G}_3 graph of following, tweeting, retweets and replies. Another measure is the *Leadership* [2], which is just a linear combination of the *Competency* and the *Popularity* measures defined before. Finally, the *Followship-LDA (FLDA)* [5] model is a slow and particularly complex algorithm that implements a Bayesian Bernoulli-Multinomial mixture model trained by Gibbs sampling. Unlike other measures like TwitterRank, FLDA performs the topical analysis at the same time of computing the influence. It considers offline content analysis and follow-up relationships. Bi et al. [5] also proposed a framework to rank influential users by topic, under different existing measures such as TSPR and TwitterRank. Note that the last two mentioned influence measures do not consider metrics of retweets, mentions nor replies. Actually, they try to rank the referents in a topic, rather than users capable of spreading influence on the network.

We finish with two interesting frameworks. The first one corresponds to Rübiger and Spiliopoulou [106]. By using supervised learning on specific topics, it was designed to distinguish users according to their influence. The authors do not define a new influence measure. Instead, they observe the real perception of people to classify an actor as influential. They use follow-up relationships, retweets, replies and mentions, as well as the structure and activity network, the centrality of users, and the quality of the tweets. In contrast with the unsupervised environments we have seen above, the authors conclude that for this supervised environment the follow-up relationships are the most important ones to identify influential users. The second framework was design by Kardara et al. [62] to evaluate topical-sensitive measures. They propose a two-dimensional taxonomy classification: by scope (global, local and glocal measures) and by metrics (graphical, contentual and holistic measures). Graphical measures are those that consider topological aspects; contentual measures consider some aspects like content quality, retweets and mentions; and holistic measures consider some aspects shared by the previous two. Furthermore, the authors propose three conditions that a group of real influencers should meet to be compatible with the framework, and two other to determine the relative effectiveness of the measure. This is an interesting attempt to establish an overall classification of influence measures.

7.4 Predicting influencers

All the measures seen so far allow ranking the users from known data at a certain instant or period of time. These rankings may change significantly over time, leaving a trail of historical information that can be used to make predictions. To predict influential users, we can extract information from metrics such as the ones given in Table 3. Xiao et al. [146] take advantage of additional information supported by *Bitly*⁹, referred to the number of short URLs shared by the users, as well as to the number of visits after clicking on them. For most of the remaining measures presented in this section, the details of the algorithms are unknown. In general, these measures have a high complexity, and they use advanced computational tools.

The *AWI model* [151] is a user interaction model that considers the (A)ctivity and (W)illingness of users to retweet through time, in order to measure the (I)nfluence among pairs of users. This model also predicts retweet ratios and influential users. Although the authors apply this model in the context of Sina Microblog, it can also be applied to Twitter. Similarly, the *ACQR Framework* [21] uses data mining to detect activity (original tweets, retweets and replies), centrality (in-degree and another measure based on Euclidean distance and out-degree), and user reputation (mechanism to distinguish between real users and spammers). It also considers the quality of tweets through the number of replies and retweets, and the reputation of users that reply and retweet. This framework was used to identify and predict the influential users in a relatively small network that was restricted to a specific topic. Soon after, other researchers were able to make online predictions on a small number of user accounts, by combining the ratio of followers and followees with the ratio of tweets written during a certain period of time [109].

As in the previous section, there also exist predictive influence measures that only use the follow-up relationships, leaving aside the other Twitter actions. Thus, the *Time Network Influence Model (TNIM)* [32] uses a probabilistic generative model to make an offline estimation of the influence power between users. This measure takes into account the time intervals between messages, follow-up relationships, and the relationships of similarity in the content of the tweets. Furthermore, the *Author Ranking* [138] uses the style of the tweets (words, hashtags, websites, references to other accounts) and user behavior (profile information, following ratios, number of tweets, and main user activity, previously determined by a text classification task). This measure considers two steps: A supervised approach and an unsupervised approach, based on Markov Random Field.

In contrast, the following measures take advantage of the different Twitter actions. The *ReachBuzzRank* [120] uses a predictive algorithm based on Hidden Markov Models (HMM) to measure and predict influential users through a statical analysis (based on PageRank) and a temporal analysis (with the new *BuzzRank* measure). ReachBuzzRank considers the network topology and the dynamic interactions between users (followers, retweets, replies and mentions). It is restricted to a particular topic, and it was applied on a small network with less than 10,000 nodes. On the other hand, the *IDM-CTMP (Information Diffusion Model based on Continuous-Time Markov Process) model* [78] is a dynamic information propagation model, based on continuous-time Markov process, to predict influential users in a dynamic and continuous way over time. It is applied on graphs whose nodes are users sending tweets, and whose edges relate users that interact (by posting tweets, mentions, retweets or replies) about the same topic. Thus, the paths in the graph generate a succession of tweets over time, sorted by time of emission. The IDM-CTMP model was applied on a database much larger than that used by the ReachBuzzRank. The authors contrast their results with several measures, such as degree and PageRank on \mathcal{G}_1 and \mathcal{G}_2 graphs of mentions and retweets. This measure encourages users who write on various topics, users who are dedicated to spread topics already initiated by others, and users whose tweets are able to generate a rapid response (low latency adoption topic) in other users.

Finally, Bouguessa and Romdhane [14] used a parameterless mixture model-based approach. Firstly, each user is associated with a feature vector that contains information related to their behavior and activity (based on metrics of following, retweets and mentions). After that, through a statistical framework based on multivariate beta mixtures over an offline database, a probability density function is estimated to obtain the users with greater authority.

⁹ <http://bitly.com>

7.5 Web applications

In this section we mention several web applications that put into practice several measures developed with a more applied than academic goal. Although almost all were designed to general purposes, Sameh [113] described an analytical tool for the study of presidential elections. This application runs in real time and uses data-mining, text-mining, algorithmic graph theory, and sentiment analysis. These techniques have also been used to measure the influence of users on other OSNs [157].

There are websites like *Klout* [108], *PeerIndex* [115], *InfluenceTracker* [109], *Twitter Grader*, *Favstar*, *BehaviorMatrix*, *Kred* and *Twitalyzer* (besides other that no longer exist, like *TurnRank* [15]),¹⁰ that rank the most relevant Twitter users according to activity, popularity, or influence. They use the Twitter API, other OSNs, timeline, and sometimes more complex criteria that are not transparent. At least until 2013, the Klout and PeerIndex measures were well correlated, while the simpler Twitter Grader measures were independent of the previous two. By using machine learning and reverse engineering, we can learn to partially predict the increment of these measure values [29]. Like previous measures, we must understand that the rankings given by these websites like Klout do not necessarily correspond with the actual influence of the actors in real life [24,25].

Most of these applications measure global influencers. A more recent web application is *tweetStimuli* [130],¹¹ that allow us to rank *local influencers*, a concept that comes from Bakshy et al. [4]. Given a user i , the *in-local influence* is calculated with the number of followees of i who have been retweeted at least once by i ; and the *out-local influence* is calculated with the number of followers of i that have retweeted to i at least once. Besides retweets, these definitions also consider tweets marked as favorites or likes. Since the idea is to measure influence at a local level, the offline networks used by tweetStimuli are quite small.

Finally, *NavigTweet* [37]¹² is another application that, among other functions, allows to visualize influential users. It implements a variation of the Analytical hierarchy process (AHP), by considering different metrics as parameters, regarding following, likes, number of tweets, lists, retweets, mentions, hashtags, and URLs mentioned in tweets.

8 Some additional aspects

8.1 About computational complexity

From a computational complexity point of view, it is more desirable to find algorithms that can run in polynomial time in terms of input size. In algorithmics, the challenge is to build more efficient algorithms in terms of both execution time and the amount of computational memory required. When we consider an input as a data stream, the “input size” turns into a less clear concept. The huge amount of information and the high dynamism of Twitter network make a polynomial time algorithm often not enough to capture the current reality of the system, because that reality already have changed. In these cases, if we want to improve the efficiency, the only option is to ignore some of the information we have.

Broadly speaking, there are two types of algorithms to compute measures of influence (or activity, or popularity): Algorithms that run in offline mode and algorithms that run in real time. The algorithms of the first type can leverage more network and user information, and thus to get results more faithful to reality, at the time the data were extracted. In contrast, although the algorithms of the second type cannot use all the information provided by the network, they offer in exchange a quick view of the users status that can be constantly updated.

Almost all measures considered in this article run in offline mode. For instance, all measures based on PageRank and HITS require to know the network graph representation in order to make their calculations. In an offline mode, the biggest problem is whether we use \mathcal{G}_2 or \mathcal{G}_3 graphs, which explicitly represents the high network dynamics. If we consider \mathcal{G}_1 graphs, the problem is the huge amount of Twitter users, together with the Twitter API limitations explained in Section 3.1. For all these cases, the usual practice is to download a large database with Twitter activity within a certain period of time. After that, the

¹⁰ See <http://klout.com>, <http://www.peerindex.net>, <http://www.influencetracker.com>, <https://mokumax.com>, <http://favstar.fm>, <http://www.behaviormatrix.com>, <http://www.kred.com>, <http://www.twitalyzer.com>.

¹¹ <http://tweetstimuli.com>

¹² <http://home.deib.polimi.it/hussain/navigtweet/>

required information is filtered and stored in a smaller database that is used as the input to compute the measures. This modus operandi can be very useful in several contexts: If we want to replicate experiments from an already stored database; if we want to study a particular event during a period of time (a natural disaster, a presidential election, etc.); or if both the storing process and the algorithmic execution are efficient enough to yield results relatively frequently (e.g., once daily).

Regarding real-time algorithms, they typically use simple metrics, and avoid working with graphs. Such measures are useful in dynamic websites, because they allow us to display attractive and current network information, e.g., a ranking of the most active current users, or the information spread by some specific set of key users.

The measures designed for Twitter usually do not include a complexity analysis. However, there are some researchers who care about these computational restrictions. For example, some authors are aware of the temporal and spatial complexity of PageRank [17,117,58]. The time complexity of several simple metrics is shown in Table 3 (for a standard textbook related with computational complexity theory, see Garey and Johnson [42]). Suppose we have a database of Twitter activity over a period of time. In total there are T tweets chronologically ordered. The cost of access to tweets data and metadata in the database is assumed to be constant. Recall that the retweets, mentions and replies are identifiable by patterns described in Section 2. Since tweets are limited to a length of 140 characters, we also assume that those patterns can be detected in constant time. It is also negligible the cost of comparing two strings, such as user names. Therefore, metrics $OT1$, $OT2$, $OT3$, $RP1$, $RP2$, $RT1$, $RT2$, $FT1$, $FT2$, $M1$ and $M3$ run in $O(T)$ time, because we only need to do a linear search through the list of tweets. Instead, for the metrics $RP3$, $RT3$, $FT3$, $M2$ and $M4$, we also need an auxiliary list to avoid repeated elements. Let be k the size of these auxiliary lists, then these metrics run in $O(T \cdot k)$ time. Metrics $F1$ and $F3$ run in $O(1)$ time, because they just require some simple API requests. Finally, metrics $F2$, $F4$, $F5$ and $F6$ depend on the API limitations presented in Section 3.

The time complexity for the different considered measures is shown in Table 1. Several results are obtained directly from the complexity results of the metrics presented above. As metrics $F2$, $F4$, $F5$ and $F6$, those measures that require to know explicit followers or followees are restricted to API limitations. The $F1$ and $F3$ metrics results should be stored ideally when the database is built, because they are subject to future changes. Note that for several topical-sensitive measures we use a lower bound (Ω) besides an upper bound (O). This is because the upper bound depends on the complexity of the algorithm used to verify whether a tweet belongs to the topic or not.

As we have seen in previous sections, several measures use the PageRank algorithm. Given a graph with n nodes, by using Coppersmith-Winograd method variations for fast matrix multiplication, both PageRank and eigenvalue decomposition can run in $O(n^{2.3727})$ time [26,145], and maybe little faster. There are also fast iteration methods like the power method, that allows to compute the dominant eigenpair of a matrix [6]. Furthermore, given a Twitter database, to build a graph \mathcal{G}_2 or \mathcal{G}_3 can be done in $O(T)$ time. That is why measures like Starrank, TrueTop and others can run in $O(T + n^{2.3727})$ time. H-index only takes $O(T + n)$ time, because it is well known that given a graph, this measure can be computed in linear time. Measures like the Acquaintance-Affinity Score and the Action-Reaction run in $O(T \cdot k^2)$ time because they require an auxiliary matrix instead of an auxiliary vector. The time complexity for some complex measures remains unknown.

8.2 About correlation of measures

The centrality measures are usually correlated with each other to determine similarities or dependencies between them [9,112,36,135]. Two similar measures provide similar rankings and thus might be redundant, while two very different measures provide different classification criteria and may be useful for different purposes. Although traditional measures like degree, betweenness, closeness and eigenvector are conceptually related, from a correlation analysis it can be shown that they are different [134]. The correlation mechanisms come from statistics, and are widely used in centrality measures for general networking.

In the context of Twitter, researchers often use, reasonably, the more traditional correlation coefficients, namely the *Spearman's rank correlation coefficient* (ρ) [122] and the *Kendall Tau rank correlation coefficient* (τ) [65]. Let w and y be two lists of n users each, we have [150]:

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)} \quad \text{and} \quad \tau = \frac{n_c - n_d}{0.5n(n - 1)}$$

where x_i and y_i are the rankings of the users i in the lists x and y , respectively. Furthermore, n_c is the number of concordant pairs (i, j) (i.e., such that either $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$) and n_d is the number of discordant pairs, i.e., those that are not concordant.

The values of both ρ and τ are in the $[-1, 1]$ interval, where 1 means that both measures are equal, 0 that they are completely independent, and -1 that one is the inverse of the other. In our context, where we can have thousand and even millions of users, it is usual to compare only the first hundreds or thousand places of the rankings. This usually generates two lists with different users. To solve this we can compute the following:

$$\frac{\rho_k(x, y) + \rho_k(y, x)}{2} \quad \text{and} \quad \frac{\tau_k(x, y) + \tau_k(y, x)}{2}$$

where $\rho_k(x, y)$ and $\tau_k(x, y)$ are obtained by matching with the following steps [150]:

1. Choose the first k users of list x .
2. Determine the rankings in list y , establishing a new list y' that only ranks the users considered in x .
3. Correlate the first k users of x with those of y' .

Other known coefficients are the *Pearson correlation coefficient* [100] and the *Goodman-Kruskal gamma* [48]. If the rankings are known in advance (e.g., through a system of surveys) we can use *precision and recall*, two measures used in information retrieval for evaluating systems where users have a binary classification (relevant / not relevant, influential / not influential, etc.) [17].

8.3 Some experiments

In this section, we present the results of a few number of measures on the same network, in order to show how the rankings change depending on the different considered centrality criteria. The network dataset is provided by the Stanford Large Network Dataset Collection (SNAP) [119].

The dataset represents all the tweets related with the Higgs boson experiment that were posted between 1st and 7th July 2012. It allows to generate a \mathcal{G}_1 graph with 456,626 users nodes, and it also contains interactions of retweets, replies and mentions between the users [33]. With this data the following metrics were computed: $RP1$, $RP3$, $RT1$, $RT2$, $RT3$, $M1$, $M2$, $M3$, $M4$, $F1$ and $F3$.

We computed nine measures (four of popularity, and five of influence) that have been mentioned in this work. The activity measures could not be computed because the dataset does not provides activity information like the total number of tweets. To illustrate each measure, we created a \mathcal{G}_1 subgraph from the union of the thousand top nodes of each measure. This results in a subgraph with 2842 nodes and 62637 edges. The different measures are illustrated in Figure 3. Note that some measures that are not normalized present a few nodes with values significantly higher than others. Furthermore, it seems from the figures that the FollowerRank is very similar to the A Score, and the TFF is very similar to the Paradoxical discounted. This kind of conclusions can be partially explained from their respective formulas. However, a more formal demonstration requires the use of correlation coefficients. We leave this correlation analysis as future work.

9 Open research problems

Identifying the influential actors on Twitter is a research under continuous development. The check marks in Table 3 put in evidence the open problem of defining a measure covering all types of Twitter metrics. In this sense, there are some measures easily generalizable to metrics still not considered. For example, it does not seem complicated to generalize the SNP measure, in order to include also the favorites or likes. These new measures, of course, must continue to be efficient, in the context of Section 8.1.

Despite of the above, there are already many measures, and new ones are constantly appearing. We believe the problem is not necessarily creating new influence measures (unless they really provide new ranking criteria, so far not considered), but classifying these measures so that we can choose from the most useful for our purposes. In this sense, the classification presented in Table 3 could still be improved. Note from the question marks in the chart that the time complexity for some measures remains unknown. Furthermore, new classification criteria could be added, such as the ones suggested by Kardara et al. [62] for general topical-sensitive measures.

In the same vein, a very interesting study that we leave open is to comprehensively correlate all measures contained in this paper, in order to identify similarities and differences, redundancies and independencies. Note that the similarity of check marks in Table 3 does not represent a correlation between the measures. Indeed, as we have seen, there is no direct correlation between active, popular and influential actors [111]. Furthermore, there exist measures that use several kind of metrics, which are well correlated with other measures that only use a strict subset of those metrics. Such is the case of the AAI measure, for instance, that is well correlated with the number of followers [123]. The experiments of Section 8.3 suggest that the FollowerRank and the A Score could be another example.

In addition, there are other ways to detect influential users without the use of centrality measures. For instance, Morone and Maske [91] study the problem of finding the minimal set of nodes (which they called the *optimal influencers*) necessary to spread information to the whole network. Removing the optimal influencers from the network leads to its collapse. The authors propose to solve this problem by solving the mathematical problem of finding an optimal percolation. They conclude that under this criterion, for \mathcal{G}_1 graphs, the influencers are not usually users with too much followers. It would be interesting to know what happens if we consider \mathcal{G}_2 and \mathcal{G}_3 graphs.

Finally, besides influential users, there are several measures that rank the most influential *tweets* in certain period of time. A survey on research of these kind of measures could also be performed. In that case, we must consider some relevant aspects, such as the content quality of tweets, or the influence exerted by a tweet over the readers [94].

10 Conclusions

In this survey we have classified the various measures of activity, popularity and influence in the literature for the specific context of Twitter. As far as we know, this is the first comprehensive survey of this type. The main results can be seen in Table 1. The diversity of measures is very high, especially considering that the oldest ones were defined only in 2009. From this chart we can see that almost all popularity measures are associated with follow-up relationships, while most activity measures consider actions of replies. In turn, for influence measures we highlight the use of retweets. Of all the kinds of metrics considered in Table 3, the least used are related to the favorites or likes. This may be because this functionality took more years to be implemented and be known by Twitter users. Despite of this, the new measures should begin to consider such interactions, which are so well supported by the Twitter API as the others.

The check marks in Table 3 do not necessarily reflect the quality of the measures. They just realize the variety of considered criteria. Note that no current measure uses all the different kind of metrics. The few measures that use metrics of favorites (or likes), always dispense with some another kind of metric. On the other hand, we have seen that using metrics like $F2$, $F4$, $F5$ and $F6$ considerably increases the time response, because they require a precomputing period to provide implicit follow-up relationships among the actors. These metrics can be useful to study particular events, when data is collected in a complete and fixed database that does not need to be updated. Otherwise, it is best to avoid them.

Interestingly, almost one half of the existing influence measures are based in one way or another on the PageRank algorithm. This algorithm, originally created to assess the relevance of websites on the Internet, has been settled as a powerful tool to rank influential users in online social networks. Several efforts have also been devoted to the definition of topical-sensitive measures, which must necessarily include a content analysis of the published tweets. Instead, predictive influence measures are few. Some algorithms for these measures are not explicitly given, so they cannot be used by other researchers.

We hope this survey will help researchers and developers to know the variety of centrality measures for Twitter network. This variety of criteria to identify influential actors is a proof that the concepts related with the spread of influence in social networks have not yet reached a consensus.

Acknowledgements

This work was partially funded by project PMI USA1204.

We thank the anonymous referees for their comments and suggestions that helped us to improve the content and the presentation of the paper.

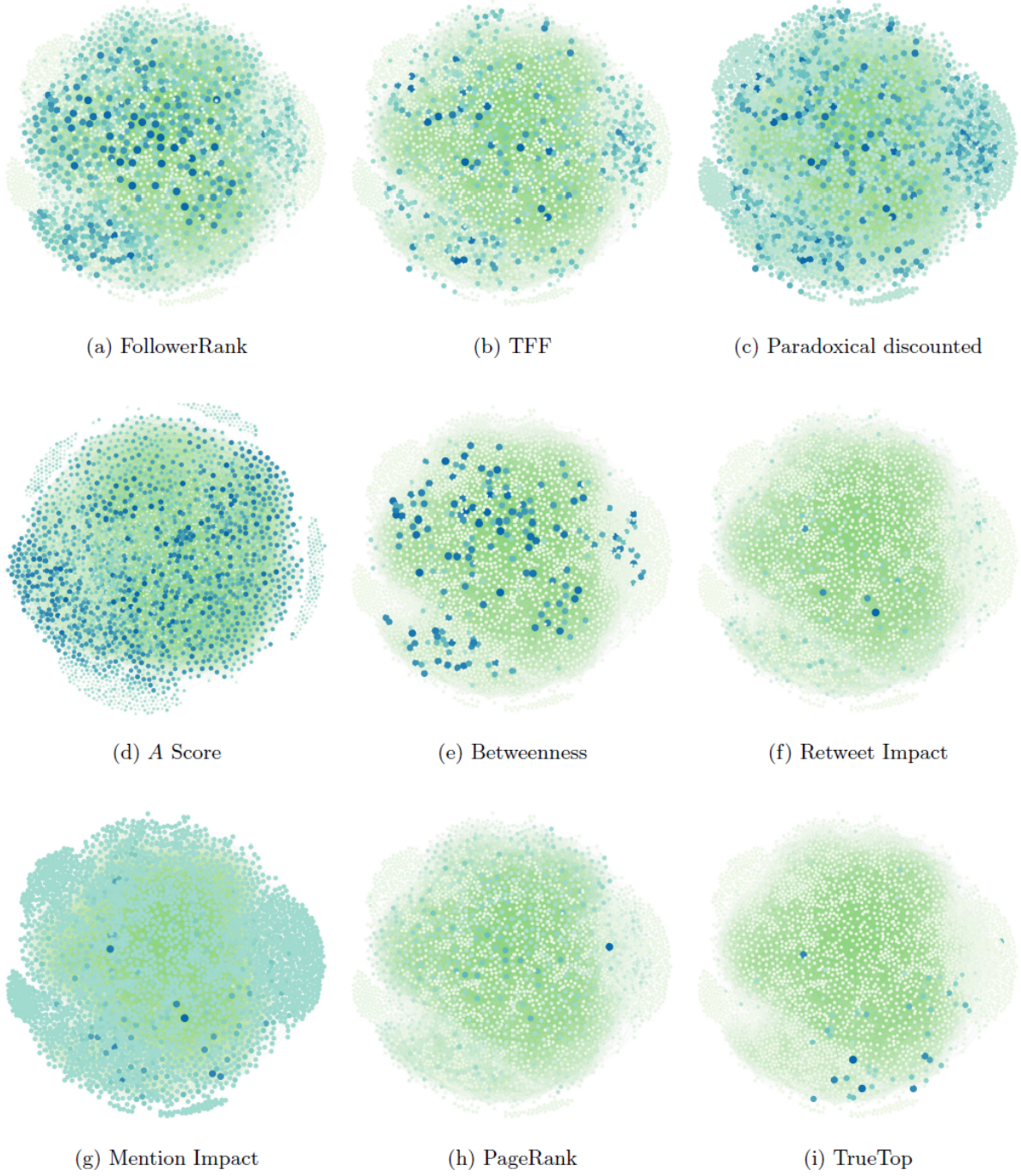


Fig. 3: Different measures computed on the same Twitter network. The most central (popular or influential) user nodes are larger and more blue, while the less central ones are smaller and more white.

References

1. Lada A. Adamic and Eytan Adar. How to search a social network. *Social Networks*, 27(3):187–203, 2005.
2. Abolfazl Aleahmad, Payam Karisani, Maseud Rahgozar, and Farhad Oroumchian. OLFinder: Finding opinion leaders in online social networks. *Journal of Information Science*, 1–16, 2015.
3. Isabel Anger and Christian Kittl. Measuring influence on Twitter. In Stefanie N. Lindstaedt and Michael Granitzer, editors, *I-KNOW 2011, 11th International Conference on Knowledge Management and Knowledge Technologies, Graz, Austria, September 7-9, 2011*, page 31. ACM, 2011.
4. Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone’s an influencer: quantifying influence on Twitter. In Irwin King, Wolfgang Nejdl, and Hang Li, editors, *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011*, pages 65–74. ACM, 2011.
5. Bin Bi, Yuanyuan Tian, Yannis Sismanis, Andrey Balmin, and Junghoo Cho. Scalable topic-specific influence analysis on microblogs. In Ben Carterette, Fernando Diaz, Carlos Castillo, and Donald Metzler, editors, *Seventh ACM International Conference on Web Search and Data Mining, WSDM 2014, New York, NY, USA, February 24-28, 2014*, pages 513–522. ACM, 2014.
6. Monica Bianchini, Marco Gori, and Franco Scarselli. Inside PageRank. *ACM Transactions on Internet Technology*, 5(1):92–128, 2005.
7. Carolina Bigonha, Thiago N. C. Cardoso, Mirella M. Moro, Virgílio A. F. Almeida, and Marcos A. Gonçalves. Detecting evangelists and detractors on Twitter. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia), 2010, Belo Horizonte, Brazil*, 2010.
8. Carolina A. S. Bigonha, Thiago N. C. Cardoso, Mirella M. Moro, Marcos André Gonçalves, and Virgílio A. F. Almeida. Sentiment-based influence detection on Twitter. *Journal of the Brazilian Computer Society*, 18(3):169–183, 2012.
9. John M. Bolland. Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, 10(3):233–253, 1988.
10. Phillip Bonacich. Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.
11. Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.
12. Javier Borge-Holthoefer and Yamir Moreno. Absence of influential spreaders in rumor dynamics. *Physical Review E*, 85(2), 2012.
13. Javier Borge-Holthoefer, Raquel A. Baños, Sandra González-Bailón, and Yamir Moreno. Cascading behaviour in complex socio-technical networks. *Journal of Complex Networks*, 1(1):3–24, 2013.
14. Mohamed Bouguessa and Lotfi Ben Romdhane. Identifying authorities in online communities. *ACM TIST*, 6(3):30, 2015.
15. Phil Brown and Junlan Feng. Measuring user influence on Twitter using modified K-shell decomposition. In *The Social Mobile Web, Papers from the 2011 ICWSM Workshop, Barcelona, Catalonia, Spain, July 21, 2011*, volume WS-11-02 of *AAAI Workshops*. AAAI, 2011.
16. Amparo Elizabeth Cano, Suvodeep Mazumdar, and Fabio Ciravegna. Social influence analysis in microblogging platforms - A topic-sensitive based approach. *Semantic Web*, 5(5):357–372, 2014.
17. Rafael Cappelletti and Nishanth Sastry. IARank: Ranking users on Twitter in near real-time, based on their information amplification potential. In *2012 International Conference on Social Informatics (Social-Informatics), Washington, D.C., USA, December 14-16, 2012*, pages 70–77. IEEE Computer Society, 2012.
18. Mario Cataldi and Marie-Aude Aufaure. The 10 million follower fallacy: audience size does not prove domain-influence on Twitter. *Knowledge and Information Systems*, pages 1–22, 2014.
19. Meeyoung Cha, Fabrício Benevenuto, Hamed Haddadi, and P. Krishna Gummadi. The world of connections and information flow in Twitter. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 42(4):991–998, 2012.
20. Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. Measuring user influence in Twitter: The million follower fallacy. In William W. Cohen and Samuel Gosling, editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*. The AAAI Press, 2010.
21. Wen Chai, Wei Xu, Meiyun Zuo, and Xiaowei Wen. ACQR: A novel framework to identify and predict influential users in micro-blogging. In Jae-Nam Lee, Ji-Ye Mao, and James Y. L. Thong, editors, *17th Pacific Asia Conference on Information Systems, PACIS 2013, Jeju Island, Korea, June 18-22, 2013*, page 20, 2013.
22. Chun Chen, Feng Li, Beng Chin Ooi, and Sai Wu. TI: an efficient indexing mechanism for real-time search on tweets. In Timos K. Sellis, Renée J. Miller, Anastasios Kementsietsidis, and Yannis Velegrakis, editors, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011*, pages 649–660. ACM, 2011.

23. Martin J. Chorley, Gualtiero B. Colombo, Stuart M. Allen, and Roger M. Whitaker. Human content filtering in Twitter: The influence of metadata. *International Journal of Human-Computer Studies*, 74:32–40, 2015.
24. Jean-Valère Cossu, Nicolas Dugué, and Vincent Labatut. Detecting real-world influence through Twitter. In *2015 Second European Network Intelligence Conference, ENIC 2015, Karlskrona, Sweden, September 21-22, 2015*, pages 83–90. IEEE Computer Society, 2015.
25. Jean-Valère Cossu, Vincent Labatut, and Nicolas Dugué. A review of features for the discrimination of Twitter users: Application to the prediction of offline influence. *CoRR*, abs/1509.06585, 2015.
26. Balázs Csanád Csáji, Raphaël M. Jungers, and Vincent D. Blondel. PageRank optimization in polynomial time by stochastic shortest path reformulation. In Marcus Hutter, Frank Stephan, Vladimir Vovk, and Thomas Zeugmann, editors, *Algorithmic Learning Theory, 21st International Conference, ALT 2010, Canberra, Australia, October 6-8, 2010. Proceedings*, volume 6331 of *Lecture Notes in Computer Science*, pages 89–103. Springer, 2010.
27. Maximilien Danisch, Nicolas Dugué, and Anthony Perez. On the importance of considering social capitalism when measuring influence on Twitter. In *Behavioral, Economic, and Socio-Cultural Computing, Oct 2014, Shanghai, China. 2014*, 2014.
28. Matthias Dehmer. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201(1-2):82–94, 2008.
29. José del Campo-Ávila, Nathalie Moreno Vergara, and Mónica Trella-López. Bridging the gap between the least and the most influential Twitter users. In Elhadi Shakshuki, Karim Djouani, Michael Sheng, Mohamed Younis, Eduardo Vaz, and Wayne Groszko, editors, *Proceedings of the 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013), the 3rd International Conference on Sustainable Energy Information Technology (SEIT-2013), Halifax, Nova Scotia, Canada, June 25-28, 2013*, volume 19 of *Procedia Computer Science*, pages 437–444. Elsevier, 2013.
30. Zhaoyun Ding, Yan Jia, Bin Zhou, and Yi Han. Mining topical influencers based on the multi-relational network in micro-blogging sites. *China Communications*, 10(1):93–104, 2013.
31. Zhaoyun Ding, Yan Jia, Bin Zhou, Yi Han, Li He, and Jianfeng Zhang. Measuring the spreadability of users in microblogs. *Journal of Zhejiang University - Science C*, 14(9):701–710, 2013.
32. Zhaoyun Ding, Yan Jia, Bin Zhou, Jianfeng Zhang, Yi Han, and Chunfeng Yu. An influence strength measurement via time-aware probabilistic generative model for microblogs. In Yoshiharu Ishikawa, Jianzhong Li, Wei Wang, Rui Zhang, and Wenjie Zhang, editors, *Web Technologies and Applications - 15th Asia-Pacific Web Conference, APWeb 2013, Sydney, Australia, April 4-6, 2013. Proceedings*, volume 7808 of *Lecture Notes in Computer Science*, pages 372–383. Springer, 2013.
33. Manlio De Domenico, Antonio Lima, Paul Mougél, and Mirco Musolesi. The anatomy of a scientific rumor. *Scientific Reports*, 3(2980), 2013.
34. Pedro M. Domingos and Matthew Richardson. Mining the network value of customers. In Doheon Lee, Mario Schkolnick, Foster J. Provost, and Ramakrishnan Srikant, editors, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, August 26-29, 2001*, pages 57–66. ACM, 2001.
35. Nicolas Dugué and Anthony Perez. Social capitalists on Twitter: detection, evolution and behavioral analysis. *Social Network Analysis and Mining*, 4(1):178, 2014.
36. Katherine Faust. Centrality in affiliation networks. *Social Networks*, 19(2):157–191, 1997.
37. Chiara Francalanci and Ajaz Hussain. NavigTwee: A visual tool for influence-based Twitter browsing. In Brian Donnellan, Markus Helfert, Jim Kenneally, Debra E. VanderMeer, Marcus A. Rothenberger, and Robert Winter, editors, *New Horizons in Design Science: Broadening the Research Agenda - 10th International Conference, DESRIST 2015, Dublin, Ireland, May 20-22, 2015. Proceedings*, volume 9073 of *Lecture Notes in Computer Science*, pages 183–198. Springer, 2015.
38. Chiara Francalanci and Ajaz Hussain. A visual analysis of social influencers and influence in the tourism domain. In Iis Tussyadiah and Alessandro Inversini, editors, *Information and Communication Technologies in Tourism 2015*, pages 19–32. Springer International Publishing, 2015.
39. Linton C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1979.
40. Linton C. Freeman, Stephen P. Borgatti, and Douglas R. White. Centrality in valued graphs: a measure of betweenness based on network flow. *Social Networks*, 13(2):141–154, 1991.
41. Miguel del Fresno García, Alan J. Daly, and Sagrario Segado Sánchez-Cabezudo. Identifying the new influencers in the Internet Era: Social media and social network analysis. *Revista Española de Investigaciones Sociológicas*, 153:23–40, 2016.
42. Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.
43. Daniel Gayo-Avello, David J. Brenes, Diego Fernández-Fernández, María E. Fernández-Menéndez, and Rodrigo García-Suárez. De retribus socialibus et legibus momenti. *EPL (Europhysics Letters)*, 94(3):38001, 2011.

44. Daniel Gayo-Avello. Nepotistic relationships in Twitter and their impact on rank prestige algorithms. *Information Processing and Management*, 49(6):1250–1280, 2013.
45. Jennifer Golbeck and James A. Hendler. Inferring binary trust relationships in web-based social networks. *ACM Trans. Internet Techn.*, 6(4):497–529, 2006.
46. Daniel Gómez, José Rui Figueira, and Augusto Eusébio. Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems. *European Journal of Operational Research*, 226(2):354–365, 2013.
47. Sandra González-Bailón, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. The dynamics of protest recruitment through an online network. *Scientific Reports*, 1(197), 2011.
48. Leo A. Goodman and William H. Kruskal. Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268):732–764, 1954.
49. Daniel Gruhl, David Liben-Nowell, Ramanathan V. Guha, and Andrew Tomkins. Information diffusion through blogspace. *SIGKDD Explorations*, 6(2):43–52, 2004.
50. Behnam Hajian and Tony White. Modelling influence in a social network: Metrics and evaluation. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 497–500. IEEE, 2011.
51. Don Hatcher, Gurpreet Singh Bawa, and Barry de Ville. How you can identify influencers in SAS® Social media analysis (and why it matters). In *SAS Global Forum 2011, April 4-7, 2011 - Las Vegas, Nevada*, number 319 in Social Media and Networking, 2011.
52. Taher H. Haveliwala. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796, 2003.
53. Reynaldo Herrera, Juan Pablo Cárdenas, and Rodrigo Alfaro. User influence in on-line social networks. In *WACS 2014, IV Workshop on Agents and Collaborative Systems. Jornadas Chilenas de Computación (JCC). November 8-14, 2014, Talca, Chile, 2014*.
54. Jonathan Herzig, Yosi Mass, and Haggai Roitman. An author-reader influence model for detecting topic-based influencers in social media. In Leo Ferres, Gustavo Rossi, Virgilio Almeida, and Eelco Herder, editors, *25th ACM Conference on Hypertext and Social Media, HT '14, Santiago, Chile, September 1-4, 2014*, pages 46–55. ACM, 2014.
55. Jorge E. Hirsch. An index to quantify an individual’s scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 85(3):741–754, 2010.
56. Zach Hofer-Shall. Working Directly With the Twitter Data Ecosystem. Gnip, April 10, 2015. <https://blog.gnip.com/twitter-data-ecosystem/>.
57. Juan Hu, Yi Fang, and Archana Godavathy. Topical authority propagation on microblogs. In Qi He, Arun Iyengar, Wolfgang Nejdl, Jian Pei, and Rajeev Rastogi, editors, *22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013*, pages 1901–1904. ACM, 2013.
58. Pei-Ying Huang, Hsin-Yu Liu, Chun-Ting Lin, and Pu-Jen Cheng. A diversity-dependent measure for discovering influencers in social networks. In Rafael E. Banchs, Fabrizio Silvestri, Tie-Yan Liu, Min Zhang, Sheng Gao, and Jun Lang, editors, *Information Retrieval Technology - 9th Asia Information Retrieval Societies Conference, AIRS 2013, Singapore, December 9-11, 2013. Proceedings*, volume 8281 of *Lecture Notes in Computer Science*, pages 368–379. Springer, 2013.
59. Yoosun Hwang. Does opinion leadership increase the followers on Twitter. *International Journal of Social Science and Humanity*, 5(3):258–264, 2015.
60. Lamjed Ben Jabeur, Lynda Tamine, and Mohand Boughanem. Active microbloggers: Identifying influencers, leaders and discussers in microblogging networks. In Liliana Calderón-Benavides, Cristina N. González-Caro, Edgar Chávez, and Nivio Ziviani, editors, *String Processing and Information Retrieval - 19th International Symposium, SPIRE 2012, Cartagena de Indias, Colombia, October 21-25, 2012. Proceedings*, volume 7608 of *Lecture Notes in Computer Science*, pages 111–117. Springer, 2012.
61. Xin Jin and Yaohua Wang. Research on social network structure and public opinions dissemination of micro-blog based on complex network analysis. *Journal of Networks*, 8(7):1543–1550, 2013.
62. Magdalini Kardara, George Papadakis, Athanasios Papaoikonomou, Konstantinos Tserpes, and Theodora A. Varvarigou. Large-scale evaluation framework for local influence theories in Twitter. *Information Processing and Management*, 51(1):226–252, 2015.
63. Georgios Katsimpras, Dimitrios Vogiatzis, and Georgios Paliouras. Determining influential users with supervised random walks. In Aldo Gangemi, Stefano Leonardi, and Alessandro Panconesi, editors, *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*, pages 787–792. ACM, 2015.
64. David Kempe, Jon M. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Lise Getoor, Ted E. Senator, Pedro M. Domingos, and Christos Faloutsos, editors, *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003*, pages 137–146. ACM, 2003.

65. M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
66. Alexy Khrabrov and George Cybenko. Discovering influence in communication networks using dynamic graph analysis. In Ahmed K. Elmagarmid and Divyakant Agrawal, editors, *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SocialCom / IEEE International Conference on Privacy, Security, Risk and Trust, PASSAT 2010, Minneapolis, Minnesota, USA, August 20-22, 2010*, pages 288–294. IEEE Computer Society, 2010.
67. Dominic King, Daniel Ramirez-Cano, Felix Greaves, Ivo Vlaev, Steve Beales, and Ara Darzi. Twitter and the health reforms in the English National Health Service. *Health Policy*, 110(2-3):291–297, 2013.
68. Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.
69. Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
70. Shoubin Kong and Ling Feng. A tweet-centric approach for topic-specific author ranking in micro-blog. In Jie Tang, Irwin King, Ling Chen, and Jianyong Wang, editors, *Advanced Data Mining and Applications - 7th International Conference, ADMA 2011, Beijing, China, December 17-19, 2011, Proceedings, Part I*, volume 7120 of *Lecture Notes in Computer Science*, pages 138–151. Springer, 2011.
71. Balachander Krishnamurthy, Phillipa Gill, and Martin F. Arlitt. A few chirps about Twitter. In Christos Faloutsos, Thomas Karagiannis, and Pablo Rodriguez, editors, *Proceedings of the first Workshop on Online Social Networks, WOSN 2008, Seattle, WA, USA, August 17-22, 2008*, pages 19–24. ACM, 2008.
72. Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue B. Moon. What is Twitter, a social network or a news media? In Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti, editors, *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 591–600. ACM, 2010.
73. Amy Nicole Langville and Carl Dean Meyer. Survey: Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2003.
74. Vito Latora and Massimo Marchiori. A measure of centrality based on network efficiency. *New Journal of Physics*, 9(6), 2007.
75. Changhyun Lee, Haewoon Kwak, Hosung Park, and Sue B. Moon. Finding influentials based on the temporal order of information adoption in Twitter. In Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti, editors, *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 1137–1138. ACM, 2010.
76. Roy Ka-Wei Lee and Ee-Peng Lim. Measuring user influence, susceptibility and cynicalness in sentiment diffusion. In Allan Hanbury, Gabriella Kazai, Andreas Rauber, and Norbert Fuhr, editors, *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29 - April 2, 2015. Proceedings*, volume 9022 of *Lecture Notes in Computer Science*, pages 411–422, 2015.
77. Ronny Lempel and Shlomo Moran. The stochastic approach for link-structure analysis (SALSA) and the TKC effect. *Computer Networks*, 33(1-6):387–401, 2000.
78. Jingxuan Li, Wei Peng, Tao Li, Tong Sun, Qianmu Li, and Jian Xu. Social network user influence sense-making and dynamics prediction. *Expert Syst. Appl.*, 41(11):5115–5124, 2014.
79. Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. TEDAS: A Twitter-based event detection and analysis system. In Anastasios Kementsietsidis, and Marcos Antonio Vaz Salles, editors, *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012*, pages 1273–1276. IEEE Computer Society, 2012.
80. Xiang Li, Shaoyin Cheng, Wenlong Chen, and Fan Jiang. Novel user influence measurement based on user interaction in microblog. In Jon G. Rokne and Christos Faloutsos, editors, *Advances in Social Networks Analysis and Mining 2013, ASONAM '13, Niagara, ON, Canada - August 25 - 29, 2013*, pages 615–619. ACM, 2013.
81. Dong Liu, Quanyuan Wu, and Weihong Han. Measuring micro-blogging user influence based on user-tweet interaction model. In Ying Tan, Yuhui Shi, and Hongwei Mo, editors, *Advances in Swarm Intelligence, 4th International Conference, ICSI 2013, Harbin, China, June 12-15, 2013, Proceedings, Part II*, volume 7929 of *Lecture Notes in Computer Science*, pages 146–153. Springer, 2013.
82. Nian Liu, Lin Li, Guandong Xu, and Zhenglu Yang. Identifying domain-dependent influential microblog users: A post-feature based approach. In Carla E. Brodley and Peter Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, pages 3122–3123. AAAI Press, 2014.
83. Xinyue Liu, Hua Shen, Fenglong Ma, and Wenxin Liang. Topical influential user analysis with relationship strength estimation in Twitter. In Zhi-Hua Zhou, Wei Wang, Ravi Kumar, Hannu Toivonen, Jian Pei, Joshua Zhexue Huang, and Xindong Wu, editors, *2014 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2014, Shenzhen, China, December 14, 2014*, pages 1012–1019. IEEE, 2014.

84. Tomáš Majer and Marián Simko. Leveraging microblogs for resource ranking. In Mária Bielíková, Gerhard Friedrich, Georg Gottlob, Stefan Katzenbeisser, and György Turán, editors, *SOFSEM 2012: Theory and Practice of Computer Science - 38th Conference on Current Trends in Theory and Practice of Computer Science, Špindlerův Mlýn, Czech Republic, January 21-27, 2012. Proceedings*, volume 7147 of *Lecture Notes in Computer Science*, pages 518–529. Springer, 2012.
85. Misa Maruyama, Scott P. Robertson, Sara K. Douglas, Bryan Semaan, and Heather A. Faucett. Hybrid media consumption: how tweeting during a televised political debate influences the vote decision. In Susan R. Fussell, Wayne G. Lutters, Meredith Ringel Morris, and Madhu Reddy, editors, *Computer Supported Cooperative Work, CSCW '14, Baltimore, MD, USA, February 15-19, 2014*, pages 1422–1432. ACM, 2014.
86. Andrew R. McNeill and Pam Briggs. Understanding Twitter influence in the health domain: a social-psychological contribution. In Chin-Wan Chung, Andrei Z. Broder, Kyuseok Shim, and Torsten Suel, editors, *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume*, pages 673–678. ACM, 2014.
87. Johnnatan Messias, Lucas Schmidt, Ricardo Oliveira, and Fabrício Benevenuto. You followed my bot! Transforming robots into influential users in Twitter. *First Monday*, 18(7), 2013.
88. Xavier Molinero, Fabián Riquelme, and Maria J. Serna. Power indices of influence games and new centrality measures for agent societies and social networks. In Carlos Ramos, Paulo Novais, Céline Ehrwein Nihan, and Juan M. Corchado Rodríguez, editors, *Ambient Intelligence - Software and Applications - 5th International Symposium on Ambient Intelligence, ISAmI 2014, Salamanca, Spain, June 4-6, 2014*, volume 291 of *Advances in Intelligent Systems and Computing*, pages 23–30. Springer, 2014.
89. Xavier Molinero, Fabián Riquelme, and Maria J. Serna. Cooperation through social influence. *European Journal of Operational Research*, 242(3):960–974, 2015.
90. Manuela Montangero and Marco Furini. TRank: Ranking Twitter users according to specific topics. In *Consumer Communications and Networking Conference (CCNC), 2015 12th Annual IEEE*, pages 767–772, 2015.
91. Flaviano Morone and Hernán A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524:65–68, 2015.
92. Rinkesh Nagmoti, Ankur Teredesai, and Martine De Cock. Ranking approaches for microblog search. In Jimmy Xiangji Huang, Irwin King, Vijay V. Raghavan, and Stefan Rueger, editors, *2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010, Toronto, Canada, August 31 - September 3, 2010, Main Conference Proceedings*, pages 153–157. IEEE Computer Society, 2010.
93. Alan Neves, Ramon Vieira, Fernando Mourão, and Leonardo Rocha. Quantifying complementarity among strategies for influencers’ detection on Twitter. *Procedia Computer Science*, 51:2435–2444, 2015. International Conference On Computational Science, ICCS 2015 Computational Science at the Gates of Nature.
94. Thi-Tuoi Nguyen, Tri-Thanh Nguyen, and Quang-Thuy Ha. Applying hidden topics in ranking social update streams on Twitter. In *2013 IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future, RIVF 2013, Hanoi, Vietnam, November 10-13, 2013*, pages 180–185. IEEE, 2013.
95. Tomoya Noro, Fei Ru, Feng Xiao, and Takehiro Tokuda. Twitter user rank using keyword search. In Peter Vojtáš, Yasushi Kiyoki, Hannu Jaakkola, Takehiro Tokuda, and Naofumi Yoshida, editors, *Information Modelling and Knowledge Bases XXIV, 22nd European-Japanese Conference on Information Modelling and Knowledge Bases (EJC 2012), Prague, Czech Republic, June, 4-9, 2012*, volume 251 of *Frontiers in Artificial Intelligence and Applications*, pages 31–48. IOS Press, 2012.
96. Elisa Omodei, Manlio De Domenico, and Alex Arenas. Characterizing interactions in online social networks during exceptional events. *Frontiers in Physics*, 3(59), 2015.
97. Lucas A. Overbey, Benjamin Greco, Christopher Paribello, and Terresa Jackson. Structure and prominence in Twitter networks centered on contentious politics. *Social Network Analysis and Mining*, 3(4):1351–1378, 2013.
98. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library, 1999.
99. Aditya Pal and Scott Counts. Identifying topical authorities in microblogs. In Irwin King, Wolfgang Nejdl, and Hang Li, editors, *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011*, pages 45–54. ACM, 2011.
100. Karl Pearson. Notes on regression and inheritance in the case of two parents. In *Proceedings of the Royal Society of London*, volume 58, pages 240–242, 1895.
101. Sen Pei, Lev Muchnik, José S. Andrade Jr., Zhiming Zheng, and Hernn A. Makse. Searching for super-spreaders of information in real-world social media. *Scientific Reports*, 4(5547), 2014.
102. Florian Probst, Laura Grosswiele, and Regina Pfleger. Who will lead and who will follow: Identifying influential users in online social networks - A critical review and future research directions. *Business & Information Systems Engineering*, 5(3):179–193, 2013.

103. Josep M. Pujol, Ramon Sangüesa, and Jordi Delgado. Extracting reputation in multi agent systems by means of social network topology. In *The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002, July 15-19, 2002, Bologna, Italy, Proceedings*, pages 467–474. ACM, 2002.
104. Hemant Purohit, Jitendra Ajmera, Sachindra Joshi, Ashish Verma, and Amit P. Sheth. Finding influential authors in brand-page communities. In John G. Breslin, Nicole B. Ellison, James G. Shanahan, and Zeynep Tufekci, editors, *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*. The AAAI Press, 2012.
105. Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft. In the mood for being influential on Twitter. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 307–314. IEEE, 2011.
106. Stefan Rübiger and Myra Spiliopoulou. A framework for validating the merit of properties that predict the influence of a Twitter user. *Expert Systems with Applications*, 42(5):2824–2834, 2015.
107. Gabriela Ramírez-de-la-Rosa, Esaú Villatoro-Tello, Héctor Jiménez-Salazar, and Christian Sánchez-Sánchez. Towards automatic detection of user influence in Twitter by means of stylistic and behavioral features. In Alexander F. Gelbukh, Félix Castro Espinoza, and Sofía N. Galicia-Haro, editors, *Human-Inspired Computing and Its Applications - 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part I*, volume 8856 of *Lecture Notes in Computer Science*, pages 245–256. Springer, 2014.
108. Adithya Rao, Nemanja Spasojevic, Zhisheng Li, and Trevor DSouza. Klout score: Measuring influence across multiple social networks. In *2015 IEEE International Conference on Big Data, Big Data 2015, Santa Clara, CA, USA, October 29 - November 1, 2015*, pages 2282–2289. IEEE, 2015.
109. Gerasimos Razis and Ioannis Anagnostopoulos. InfluenceTracker: Rating the impact of a Twitter account. In Lazaros S. Iliadis, Ilias Maglogiannis, Harris Papadopoulos, Spyros Sioutas, and Christos Makris, editors, *Artificial Intelligence Applications and Innovations - AIAI 2014 Workshops: CoPA, MHDW, IIVC, and MT4BD, Rhodes, Greece, September 19-21, 2014. Proceedings*, volume 437 of *IFIP Advances in Information and Communication Technology*, pages 184–195. Springer, 2014.
110. Matthew Richardson and Pedro M. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*, pages 61–70. ACM, 2002.
111. Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. Influence and passivity in social media. In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III*, volume 6913 of *Lecture Notes in Computer Science*, pages 18–33. Springer, 2011.
112. Richard B. Rothenberg, John J. Potterat, Donald E. Woodhouse, William W. Darrow, Stephen Q. Muth, and Alden S. Klodahl. Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. *Social Networks*, 17(3-4):273–297, 1995.
113. Ahmed Sameh. A Twitter analytic tool to measure opinion, influence and trust. *Journal of Industrial and Intelligent Information*, 1(1), 2013.
114. Christopher B. Schenk and Douglas C. Sicker. Finding event-specific influencers in dynamic social networks. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 501–504. IEEE, 2011.
115. Javier Serrano-Puche. Herramientas web para la medición de la influencia digital: análisis de Klout y PeerIndex. *El Profesional de la Información*, 21(3), 2012. In Spanish.
116. Xianlei Shao, Chunhong Zhang, and Yang Ji. Finding domain experts in microblogs. In Valérie Monfort and Karl-Heinz Krempels, editors, *WEBIST 2014 - Proceedings of the 10th International Conference on Web Information Systems and Technologies, Volume 2, Barcelona, Spain, 3-5 April, 2014*, pages 247–254. SciTePress, 2014.
117. Arlei Silva, Sara Guimarães, Wagner Meira Jr., and Mohammed Javeed Zaki. ProfileRank: finding relevant content and influential users based on information diffusion. In Feida Zhu, Qi He, Rong Yan, and John Yen, editors, *Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNAKDD 2013, Chicago, IL, USA, August 11, 2013*, pages 2:1–2:9. ACM, 2013.
118. Michael Sippey. Changes coming in Version 1.1 of the Twitter API. Twitter Blogs, August 16, 2012. <https://blog.twitter.com/2012/changes-coming-in-version-11-of-the-twitter-api>.
119. Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. June 2014. <http://snap.stanford.edu/data>.
120. Donal Stephen Simmie, Maria Grazia Vigliotti, and Chris Hankin. Ranking Twitter influence by combining network centrality and influence observables in an evolutionary model. In Kokou Yétongnon, Albert Dipanda, and Richard Chbeir, editors, *Ninth International Conference on Signal-Image Technology &*

- Internet-Based Systems, SITIS 2013, Kyoto, Japan, December 2-5, 2013*, pages 491–498. IEEE Computer Society, 2013.
121. Xiaodan Song, Belle L. Tseng, Ching-Yung Lin, and Ming-Ting Sun. Personalized recommendation driven by information flow. In Efthimis N. Efthimiadis, Susan T. Dumais, David Hawking, and Kalervo Järvelin, editors, *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6-11, 2006*, pages 509–516. ACM, 2006.
 122. Charles Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15:88–103, 1904.
 123. M. S. Srinivasan, Srinath Srinivasa, and Sunil Thulasidasan. Exploring celebrity dynamics on Twitter. In *Proceedings of the 5th IBM Collaborative Academia Research Exchange Workshop, I-CARE '13*, pages 13:1–13:4. New York, NY, USA, ACM, 2013.
 124. M. S. Srinivasan, Srinath Srinivasa, and Sunil Thulasidasan. A comparative study of two models for celebrity identification on Twitter. In Srikanta Bedathur, Divesh Srivastava, and Satyanarayana R. Valluri, editors, *20th International Conference on Management of Data, COMAD 2014, Hyderabad, India, December 17-19, 2014*, pages 57–65. Computer Society of India, 2014.
 125. Giovanni Stilo, Paola Velardi, Alberto E. Tozzi, and Francesco Gesualdo. Predicting flu epidemics using Twitter and historical data. In Dominik Slezak, Ah-Hwee Tan, James F. Peters, and Lars Schwabe, editors, *Brain Informatics and Health - International Conference, BIH 2014, Warsaw, Poland, August 11-14, 2014, Proceedings*, volume 8609 of *Lecture Notes in Computer Science*, pages 164–177. Springer, 2014.
 126. Beiming Sun and Vincent T. Y. Ng. Identifying influential users by their postings in social networks. In Alvin Chin, Martin Atzmueller, and Denis Helic, editors, *Proceedings of the 3rd international workshop on Modeling social media, MSM 2012, Milwaukee, WI, USA, June 25, 2012*, pages 1–8. ACM, 2013.
 127. Caihong Sun, Lu Zhang, and Qian Li. Who are influentials on micro-blogging services: Evidence from social network analysis. In Jae-Nam Lee, Ji-Ye Mao, and James Y. L. Thong, editors, *17th Pacific Asia Conference on Information Systems, PACIS 2013, Jeju Island, Korea, June 18-22, 2013*, page 25, 2013.
 128. Jimeng Sun and Jie Tang. A survey of models and algorithms for social influence analysis. In Charu C. Aggarwal, editor, *Social Network Data Analytics*, pages 177–214. Springer, 2011.
 129. Juyup Sung, Seunghyeon Moon, and Jae-Gil Lee. The influence in Twitter: Are they really influenced? In Longbing Cao, Hiroshi Motoda, Jaideep Srivastava, Ee-Peng Lim, Irwin King, PhilipS Yu, Wolfgang Nejdl, Guandong Xu, Gang Li, and Ya Zhang, editors, *Behavior and Social Computing*, volume 8178 of *Lecture Notes in Computer Science*, pages 95–105. Springer International Publishing, 2013.
 130. Arturo Tejeda-Gómez, Miquel Sánchez-Marrè, and Josep M. Pujol. Discovering social structures of local influence by using *tweetStimuli*. *International Journal of Computer Mathematics*, 91(2):291–303, 2014.
 131. The Telegraph. Average Twitter user is an an American woman with an iPhone and 208 followers. October 11, 2012.
<http://www.telegraph.co.uk/technology/news/9601327/Average-Twitter-user-is-an-an-American-woman-with-an-iPhone.html>.
 132. Ramine Tinati, Leslie Carr, Wendy Hall, and Jonny Bentwood. Identifying communicator roles in Twitter. In Alain Mille, Fabien L. Gandon, Jacques Misselis, Michael Rabinovich, and Steffen Staab, editors, *Proceedings of the 21st World Wide Web Conference, WWW 2012, Lyon, France, April 16-20, 2012 (Companion Volume)*, pages 1161–1168. ACM, 2012.
 133. Daniel Tunkelang. A Twitter analog to Pagerank. January 2009.
<http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank>.
 134. Thomas W. Valente, Kathryn Coronges, Cynthia Lakon, and Elizabeth Costenbader. How correlated are network centrality measures? *Connections (Toronto, Ont)*, 28(1):16–26, 2008.
 135. Thomas W. Valente and Robert K. Foreman. Integration and radiality: Measuring the extent of an individual’s connectedness and reachability in a network. *Social Networks*, 20(1):89–105, 1998.
 136. Hérico Valiati, Arlei Silva, Sara Guimarães, and Wagner Meira Jr. Uma estratégia baseada em difusão de informação para determinação de conteúdos relevantes e usuários influentes em redes sociais. *RITA*, 20(3):183–208, 2013. In Portuguese.
 137. Cornelius Joost van Rijsbergen. *Information Retrieval*. Butterworth, 1979.
 138. Esaú Villatoro-Tello, Gabriela Ramírez-de-la-Rosa, Christian Sánchez-Sánchez, Héctor Jiménez-Salazar, Wulfrano Arturo Luna-Ramírez, and Carlos Rodríguez-Lucatero. UAMCLyR at RepLab 2014: Author profiling task. In Linda Cappellato, Nicola Ferro, Martin Halvey, and Wessel Kraaij, editors, *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014.*, volume 1180 of *CEUR Workshop Proceedings*, pages 1547–1558. CEUR-WS.org, 2014.
 139. Dimitrios Vogiatzis. Influential users in social networks. In Ioannis Anagnostopoulos, Mária Bieliková, Phivos Mylonas, and Nicolas Tsapatsoulis, editors, *Semantic Hyper/Multimedia Adaptation - Schemes and Applications*, volume 418 of *Studies in Computational Intelligence*, pages 271–295. Springer, 2013.
 140. Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Cambridge University Press, 1994.

141. Benjamin Waugh, Maldini Abdipannah, Omid Hashemi, Shaquille A. Rahman, and David M. Cook. The influence and deception of Twitter: The authenticity of the narrative and slacktivism in the australian electoral process. In *Australian Information Warfare and Security Conference*, 2013.
142. Leila Weitzel, Paulo Quaresma, and José Palazzo M. de Oliveira. Measuring node importance on Twitter microblogging. In Dumitru Dan Burdescu, Rajendra Akerkar, and Costin Badica, editors, *2nd International Conference on Web Intelligence, Mining and Semantics, WIMS '12, Craiova, Romania, June 6-8, 2012*, pages 11:1–11:7. ACM, 2012.
143. Michael J. Welch, Uri Schonfeld, Dan He, and Junghoo Cho. Topical semantics of Twitter links. In Irwin King, Wolfgang Nejdl, and Hang Li, editors, *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011*, pages 327–336. ACM, 2011.
144. Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. TwitterRank: finding topic-sensitive influential twitterers. In Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu, editors, *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010*, pages 261–270. ACM, 2010.
145. Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 887–898. ACM, 2012.
146. Chunjing Xiao, Yuhong Zhang, Xue Zeng, and Yue Wu. Predicting user influence in social media. *JNW*, 8(11):2649–2655, 2013.
147. Feng Xiao, Tomoya Noro, and Takehiro Tokuda. Finding news-topic oriented influential Twitter users based on topic related hashtag community detection. *J. Web Eng.*, 13(5&6):405–429, 2014.
148. Wei Xie, Ai Phuong Hoang, Feida Zhu, and Ee-Peng Lim. Information vs interaction: An alternative user ranking model for social networks. In Adam Jatowt, Ee-Peng Lim, Ying Ding, Asako Miura, Taro Tezuka, Gaël Dias, Katsumi Tanaka, Andrew J. Flanagan, and Bing Tian Dai, editors, *Social Informatics - 5th International Conference, SocInfo 2013, Kyoto, Japan, November 25-27, 2013, Proceedings*, volume 8238 of *Lecture Notes in Computer Science*, pages 227–240. Springer, 2013.
149. Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. TURank: Twitter user ranking based on user-tweet graph analysis. In Lei Chen, Peter Triantafillou, and Torsten Suel, editors, *Web Information Systems Engineering - WISE 2010 - 11th International Conference, Hong Kong, China, December 12-14, 2010. Proceedings*, volume 6488 of *Lecture Notes in Computer Science*, pages 240–253. Springer, 2010.
150. Shaozhi Ye and Shyhtsun Felix Wu. Measuring message propagation and social influence on Twitter.com. *International Journal Communication Networks and Distributed Systems*, 11(1):59–76, 2013.
151. Zibin Yin and Ya Zhang. Measuring pair-wise social influence in microblog. In *2012 International Conference on Privacy, Security, Risk and Trust, PASSAT 2012, and 2012 International Conference on Social Computing, SocialCom 2012, Amsterdam, Netherlands, September 3-5, 2012*, pages 502–507. IEEE, 2012.
152. Jinfeng Yuan, Li Li, and Le Luo Min Huang. Topology-based algorithm for users' influence on specific topics in micro-blog. *Journal of Information and Computational Science*, 10(8):2247–2259, 2013.
153. Yanmei Zhai, Xin Li, Jialiang Chen, Xiumei Fan, and William K. Cheung. A novel topical authority-based microblog ranking. In Lei Chen, Yan Jia, Timos Sellis, and Guanfeng Liu, editors, *Web Technologies and Applications - 16th Asia-Pacific Web Conference, APWeb 2014, Changsha, China, September 5-7, 2014. Proceedings*, volume 8709 of *Lecture Notes in Computer Science*, pages 105–116. Springer, 2014.
154. Biao Zhang, Shuai Zhong, Kunmei Wen, Ruixuan Li, and Xiwu Gu. Finding high-influence microblog users with an improved PSO algorithm. *IJMIS*, 18(4):349–356, 2013.
155. Jinxue Zhang, Rui Zhang, Jingchao Sun, Yanchao Zhang, and Chi Zhang. TrueTop: A sybil-resilient system for user influence measurement on Twitter. *CoRR*, abs/1504.04663, 2015.
156. Meng Zhang, Caihong Sun, and Wenhui Liu. Identifying influential users of micro-blogging services: A dynamic action-based network approach. In Peter B. Seddon and Shirley Gregor, editors, *Pacific Asia Conference on Information Systems, PACIS 2011: Quality Research in Pacific Asia, Brisbane, Queensland, Australia, 7-11 July 2011*, page 223. Queensland University of Technology, 2011.
157. Kang Zhao, John Yen, Greta Greer, Baojun Qiu, Prasenjit Mitra, and Kenneth Portier. Finding influential users of online health communities: a new metric based on sentiment influence. *Journal of the American Medical Informatics Association*, 21(e2):e212–e218, 2014.
158. Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. Comparing Twitter and traditional media using topic models. In Paul D. Clough, Colum Foley, Cathal Gurrin, Gareth J. F. Jones, Wessel Kraaij, Hyowon Lee, and Vanessa Murdock, editors, *Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings*, volume 6611 of *Lecture Notes in Computer Science*, pages 338–349. Springer, 2011.
159. Haiyi Zhu, Bernardo A. Huberman, and Yarun Luon. To switch or not to switch: understanding social influence in online choices. In Joseph A. Konstan, Ed H. Chi, and Kristina Höök, editors, *CHI Conference*

on Human Factors in Computing Systems, CHI '12, Austin, TX, USA - May 05 - 10, 2012, pages 2257–2266.
ACM, 2012.