Knowledge and Information Systems (2020) 62:3481–3508
https://doi.org/10.1007/s10115-020-01459-y

REGULAR PAPER

Measuring time-sensitive user influence in Twitter

Behzad Rezaie[1] · Morteza Zahedi[1] · Hoda Mashayekhi[1]

Received: 13 April 2018 / Revised: 1 March 2020 / Accepted: 7 March 2020 / Published online: 28 March 2020

Abstract
Identification of the influential users is one of the most practical analyses in social networks. The importance of this analysis stems from the fact that such users can affect their followers "/friends" viewpoints. This study aims at introducing two new indices to identify the most influential users in the Twitter social network. Four sets of features extracted from user activities, user profile, tweets, and actions performed on tweets are deployed to create the proposed indices. The available methods of detecting the most influential Twitterers either consider a limited set of features or do not accurately measure the effect of each feature. The indices proposed in this paper consider a comprehensive set of features and also provide a time-sensitive rank which can be used to measure the dynamic nature of influence. Moreover, the relative impact of each feature is computed and considered in the indices. We employ the indices to discover the influential Twitter users posting on Paris attacks in 2015, in a comprehensive analysis. The influence trend of users' tweets in a 21-day period discloses that 76% of the users do not succeed in posting a second influential tweet. Results reveal that the proposed indices can detect both the publicly recognized sources (like celebrities) and also the less known individuals which gain credit by posting several influential tweets after a specific event. We further compare the proposed indices with other available approaches.

Keywords Influential Twitter users · Social network analysis · Time-sensitive ranking · AHP · Knowledge discovery

1 Introduction

Social networks allow individuals to construct a personal profile within a bounded system and collaborate with other users [1]. YouTube, MySpace, Flickr, and Facebook are some of the prominent online social networking Web sites which facilitate different interfaces for online content sharing [2]. Such Web sites are being widely used every day by millions of people from all over the globe. Twitter, as one of the most popular social networks, started its services in 2006 and enables users to interact with messages known as tweets. Over 313 million active Twitterers use the services every

✉ Hoda Mashayekhi
hmashayekhi@shahroout.ac.ir

1 Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

month.[1] In this social network, after creation of a user account, the Twitterer can perform different types of activities to enrich her network. Of such activities are posting tweets on different topics, following other users, mentioning users (@username), and using hashtags (#phrase) in tweets [3]. The popularity of Twitter has urged researchers from various fields, such as computer science [4, 5], politics [6, 7], psychology [8], and finance [9], to focus on this social network and study its impact on the society.

One of the issues that have raised particular attention in the literature is the identification of the most influential users in social networks. It has been a leading topic in various applications such as marketing [10] and presidential election [11]. Influence in social media is known as the impact of contents such as text messages, images, and videos shared by users in social networks [12–14]. This impact can be measured in various ways, including the number of times the content is marked as favorite, re-shared, etc., and depends on the specific features provided by each network.

A variety of works studied the issue of identifying the most influential users on Twitter. To this end, some studies consider the features from tweets and ignore the features related to users [7, 15]. Another approach counts only for the number of followers, which is a feature extracted from the users [6]. While some studies take both types of features to build an influence model in Twitter, they do not consider a comprehensive set of features [16, 17]. In addition, these studies do not apply the features properly in their proposed models. For instance, a negative feature like the number of followings is deployed positively and increases the influence of the user [17]. Also, most previous studies do not offer a dynamic measurement to consider the effect of time on influence [6, 16–18].

In this paper, we propose two new indices for measuring influence of users posting about an event in Twitter. The proposed indices employ an inclusive set of features related to both the Twitterers and their tweets, which, to the best of our knowledge, have not been jointly considered in previous studies. We also consider a number of features like *time*, which is a significant feature omitted in previous studies. It enables dynamic measurement of the indices at any given time interval after an event. The temporal dimension of a tweet highlights the fact that if a tweet is influential today, it might lose its popularity some time later. Therefore, the proposed indices can yield a more accurate ranking of users' influence. The process of building the proposed indices consists of several steps as follows:

- (1) The preprocessing step in which spammers and tweets with low similarity relations are removed.
- (2) The weight calculation step in which the weight assigned to each feature is calculated using the analytic hierarchy process (AHP) algorithm [19, 20].
- (3) The model construction step in which tweet influence and user influence indices are built, and their combination forms the general indices.

The approach by which the features and their associated weights are applied in the model is selected carefully to overcome the limitations and flaws of previous studies. The weights in the model reflect the relative importance of each feature and, once determined by the AHP process, are independent of any particular event and its dynamics. The introduced indices can be deployed in other contexts such as marketing, or for users of other social networks. Most features considered in the proposed indices are not bound to an event, but some of them such as distance or time may not be applicable in other contexts. Therefore, in other use cases, irrelevant features may be eliminated, and other feasible features may be plugged in.

[1] https://about.twitter.com/company [January 20, 2018].

Applying the proposed indices in a different social network like Facebook requires exchange of concepts such as followers and friends, or retweeting and sharing.

One of the most important events of 2015 and the most influential topic in Twitter in the same year was the Paris attacks.[2] The attacks occurred in November 13, 2015, in six places of Paris. In the days following the attacks, *#PrayForParis* and *#ParisAttacks* were the popular hashtags in Twitter to express sympathy with people of Paris. Due to the importance of the event in terms of both the number of casualties and its popularity on Twitter, which caused it to be the 2015 hottest topic on the platform, this study is oriented around identification of the most influential Twitterers posting about Paris attacks. The data are collected using the Twitter API.

The previous models of identification of the most influential Twitterers suffer from having a limited set of features. In this study, we took into account both user and tweet features and built our indices upon these comprehensive features. After deploying the indices proposed in this study and the one introduced by Metra [17] on the collected dataset, results reveal at least a 95% difference between the two studies for the rank of the top 20 influential users. This shows the important role of features and the weight of their corresponding variables in our proposed indices. The same comparison is held between our proposed indices and the model introduced by Asadi and Agah [18] revealing at least a 90% difference in ranking of the users.

The paper is organized as follows: Sect. 2 introduces the related studies performed on the problem. The proposed indices are discussed in Sect. 3. Section 4 is oriented around the experimental results of the research.

2 Related work

Available approaches of influence calculation in social networks either directly focus on characteristics of social networks, or borrow concepts from similar body of research such as graph theory and Web page ranking. We first briefly point to most popular measures applied in similar fields and then move to studies which consider influence in the context of social media. The research focusing specifically on Twitter is discussed next by considering user influence, and also influence in context of specific events.

Considering a social network as a graph, several number of centrality measures can be used to identify the important nodes of a graph. Node in-degree, which is defined as the number of links incoming from other nodes, is such a measure. Another measure is betweenness [21], which measures the number of shortest paths passing through a particular graph node. The nodes possessing high betweenness values act like bridges in the graph. They connect the communities to each other [22]. Closeness [23] is the other centrality measure demonstrating the average distance from each node to every other node in the graph. The nodes connecting to the others through many intermediaries have near zero closeness. Importance of lines (DIL) [24] calculates the importance of the edges in a graph, and since each edge connects two nodes, it believes that such importance is conveyed to the corresponding nodes. Efficiency centrality [25] is a recent graph-based measure to identify the most influential nodes. Eigenvector [26] is another measure deployed to detect the influential nodes in a graph. It states that the nodes connected to other well-connected nodes are more important. Although ranking of Web pages is particularly important for search engines, the applied methods may be adopted for influence measurement as well. PageRank (PR) [27], based on eigenvector centrality, is one of the most

[2] https://2015.twitter.com/most-influential [January 20, 2018].

prominent algorithms deployed in Google search engine[3] to present the top-ranked Web sites to users. The other link structure analysis solution in ranking the Web pages on the Internet was proposed by Jon Kleinberg, and named as hyperlink-induced topic search (HITS) [28]. The algorithm introduces two concepts, *authority* and *hub*. An authority is the page holding the authoritative information about the terms queried, and a hub is the Web page pointing to the authorities.

Some studies identify the most influential users in the broad social media context not limited to any specific network. The method proposed by Du et al. [29] is based on TOPSIS,[4] as a multi-attribute decision-making (MADM) technique. Four networks are employed to evaluate the model. The features extracted from the networks are betweenness, closeness, in-degree, and eigenvector. The other MADM technique deployed in identification of the most influential users is AHP. Like [29], centrality measures, such as in-degree, betweenness, closeness, and PageRank, have been treated as features and calculated for each node. Next, such features are fed to the AHP to rank the most influential nodes [30]. Leung et al. [31] introduced the DIFSoN model to detect the influential groups in social networks. The model is based on frequent pattern trees and identifies the prominence or popularity of each user. Prominence of a group is the average of prominence of group users. Frequency is the other concept deployed in DIFSoN, which counts the number of times a user is listed in a group. Multiplication of the two variables reveals the importance of a group. In another research, a user's influence is modeled as direct and indirect influence [32]. The influences are calculated using the entropy of friend nodes and the entropy of interaction frequencies. The sum of such influences builds the model.

Influence in Twitter can be measured in various ways. For instance, when looking for the most famous Twitterer, a simple search in related work might point to the number of followers or tweets as the best features to look for. While such attributes can indicate the other users' interest in one's tweets, it cannot guarantee the Twitterer's influence. Moreover, being influential in a single topic might not result in the same trend in other topics. Along with the development of social networks, specifically Twitter, the research in the field of identifying the most influential users has experienced notable advances. Inspired by the PR algorithm, a recent study [33] focuses on the Turkish tweets posted on six different topics, including politics, sports, religion, etc. It takes into account the number of tweets, retweets, and also the date on which the tweet is posted and creates four types of measures, focus rate (the focus of a user on a specific topic), activeness (relation between being active and user influence), authenticity (the originality of a user's tweets), and speed of getting reaction (the average time passed after a posted tweet gets its first retweet). Each of the introduced measures acts as the damping factor in the PR algorithm. Thus, four types of PRs are defined, and each PR is run on each topic to extract the influential users of each topic. The research states that each of the introduced measures performs well for a (set of) specific topic(s). For instance, for the tweets related to politics, the authenticity outperforms the other measures. In another recent research to detect the most influential users in Twitter [18], the number of followers, lists, ratio of followers/followees, and the number of tweets per year are introduced as the model features. Next, the Twitterers are divided into three categories of active, famous, and passive users and the characteristics of each category are analyzed. The research carried out by Metra [17] extracts random Twitter user profiles. Features are divided into two types: user features and tweet features. The first type includes the number of followers, followees, tweets, etc., and the second one consists of the number of retweets, hashtags, links, and so

[3] https://www.google.com.

[4] Technique for order performance by similarity to ideal solution.

on. AHP algorithm is employed to assign weight to each feature. To measure influence, the sum of weighted feature values is calculated.

Measuring influence in the context of important events is among the active fields of research. “Arab Spring,” as one of the most important events in recent years, has received considerable attention. Overbey et al. [6] employed the alpha algorithm [34] to identify the influential Twitter users in Egypt Revolution in 2011. The tweets containing #Egypt, #Jan25, #mubarak and #tahrir hashtags between December 10, 2010, and October 11, 2011, are collected, and the results uncover that the most influential Twitterers for each hashtag are the news agencies. In another research, the most influential Arab Spring Twitter users are studied [7]. The relevancy of tweets to Arab Spring and their geolocation are the features used to determine the rank of Twitterers due to their impact on the event. Results indicate the people tweeting from the countries experiencing Arab Spring are more influential. The research done by Cha et al. [16] identified the influential users for Twitter trends in 2009. The number of followers, retweets and mentions is taken into account for the users twitting about Iranian Presidential Election, Swine Flu, and Michael Jackson. Results demonstrate the influential Twitterers in a single trend have been important users in the other two trends, as well. News agencies create the majority of such users. The research also stresses the importance of twitting on a single trend in enhancing the users’ influence. Aside from employing the features extractable from Twitter, Servi and Elson [15] have used sentiment analysis of tweets to identify the most influential tweets and, thus, the most influential users. This research identifies and scores the feelings of tweets using linguistic inquiry and word count (LIWC) ranging from − 1 to 1. Next, it searches for the time when the feelings experience a dramatic shift. The tweet causing the most sweeping change in feelings is identified as being the most influential.

The above-introduced models for identification of the most influential Twitterers have two shortages: (1) lack of comprehensiveness in the proposed model; i.e., all the possible Twitter features are not included in the models. Such models are made upon only user-related or tweet-related features [6, 15]. Considering a combination of the features by Cha et al. [16] and Metra [17], they still do not cover all the possible features presented by the Twitter API. This causes the models to suffer from lack of accuracy. In contrast to previous researches, in our indices, we have deployed a comprehensive set of both user and tweet features extracted from the Twitter API. (2) Ignoring the role of each feature in the model: Taking the model introduced in [17] as the most complete one among the previous researches in Twitter, it does not implement the model with the appropriate role of each feature. For example, in the model, the more people a user follows, the more influence score she gains. Knowing the fact that not all Twitter features can cause an increase in influence, we have tuned each feature according to its role. For instance, placing links in a tweet represents the user as a person lacking the creativity [35]. This makes the tweet to be less influential compared with the same tweet with no links.

3 The proposed influence indices

In the proposed indices, the features of users and tweets are considered to achieve a precise ranking. It also takes into account the weights of both set of features, and their negative/positive effect. Figure 1 illustrates the steps taken in this study to build the influence indices. After collection of data, three steps of preprocessing, weight calculation, and index construction are performed to compute the influence indices. The features and their corre-

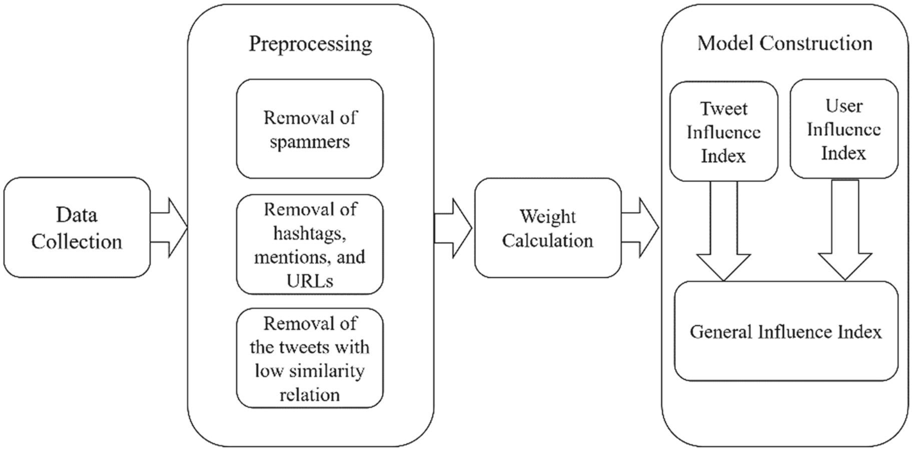


Fig. 1 The steps taken to build the proposed influence indices

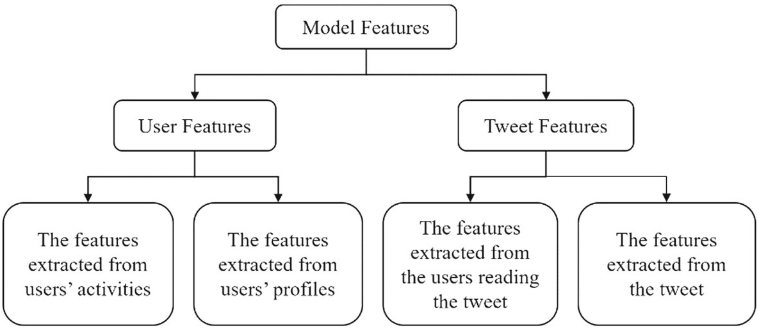


Fig. 2 The hierarchy of features in the proposed index

sponding variables along with the steps taken to build the model are described in detail in the next subsections.

3.1 Features and variables

The features considered for the proposed indices are categorized into two groups: the features related to users and the ones related to tweets. Each group is also divided into two subgroups. Figure 2 shows the hierarchy of index features, and Table 1 shows the features of each subcategory in details and their equivalent variables in the index formulation.

The *flerCnt*, *flingCnt*, *userFavCnt*, *tweetCnt*, *listCnt*, *hashCnt*, *menCnt*, *urlCnt*, *len*, *retweetCnt*, and *tweetFavCnt* are integer features. The five variables *imgUrl*, *des*, *mimgUrl*, *loc*, and *lang* are binary, and their existence increases the trust on the user and thus promotes the influence. *sim* represents the similarity of other tweets to the selected tweet. The more similar tweets found, the more influential the tweet is. It must be noted that the tweets are first sorted by the posting time, and the similarity is calculated for the tweets posted after the selected tweet. *dist* is a feature inversely proportional to the distance between the place where the tweet is posted from, to the location of the target event. The Twitterers observing

Table 1 The index features

Feature category	Feature	Description	Equivalent variable
Extracted from users' activities	TweetsCount	The number of tweets the user has posted	$tweetCnt$
	UserFavoritesCount	The number of times the user marks other Twitterer' tweets as favorite	$userFavCnt$
	FollowersCount	The number of users following the selected user	$flerCnt$
	FollowingsCount	The number of users the user is following	$flingCnt$
	ListedCount	The number of lists in which the user is grouped in	$listCnt$
Extracted from users' profiles	MiniProfileImageURL	Existence of the user mini profile image URL	$mimgUrl$
	ProfileImageURL	Existence of user profile image URL	$imgUrl$
	Description	Existence of user description	des
	Language	Existence of the language of the user	$lang$
	Location	Existence of the location where the user lives	loc
Extracted from the tweet	HashtagsCount	The number of hashtags in a tweet	$hashCnt$
	MentionsCount	The number of mentions in a tweet	$menCnt$
	URLsCount	The number of URLs in a tweet	$urlCnt$
	Length	The length of tweet message (in characters)	len
	Time	The date and time of the tweet	$time = \frac{1}{t_{CurrentDate} - t_{TweetCreation} + 1}$
	Distance	The distance between the place the tweet is posted from, to the event location (km)	$dist = \frac{1}{Distance}$
Extracted from the users reading the tweet	RetweetsCount	The number of times the tweet is republished in Twitter	$retweetCnt$
	TweetFavoritesCount	The number of times the tweet is marked as favorite	$tweetFavCnt$
	Similarity	The sum of the similarities of other tweets to the selected tweet	sim

the event in its actual location are the eyewitnesses. The closer a user is to the event location, the more influential and trustworthy she is [7]. Distance is measured in two steps: First, Google Maps Geocoding API[5] is employed to gain the altitude and latitude of the place mentioned in the tweet. Next, the Haversine formula [36] is used to calculate the distance from event location. *time* as another important variable will be described in Sect. 3.4.3. $t_{CurrentDate}$ and $t_{TweetCreation}$ are the date of computing the indices and the date when the tweet has been published, respectively.

flingCnt, *userFavCnt*, and *menCnt* act like out degrees [37] in graph theory [38, 39]. When a tweet mentions a user, it is like linking to that user and thus reduces the influence of the tweet. Therefore, *menCnt* is treated negatively in our model. The same philosophy goes for *flingCnt* and *userFavCnt*, as the user is following another user or marking her tweet as favorite (outgoing edge). Existence of URLs in a tweet reveals lack of novelty of the tweet [40]. Thus, *urlCnt* is also treated as a negative feature.

3.2 Preprocessing

The preprocessing step is performed to remove useless/adverse contents from data and to extract desired features from profiles and tweets. Spammers in social networks are the users posting messages within a periodic and regular timing [41]. Since such users can directly affect the process of identifying the most influential users, it is strongly required to remove them from the dataset. According to [17], spammers are the Twitterers posting an average of fewer than 2 messages daily. The tweets from spammers are removed from the dataset.

One of the variables deployed in the indices is the similarity between each pair of tweets. Due to the large amount of tweets in the dataset, measuring similarity of all pairs of tweets is not feasible. Therefore, tweets are, first, clustered into 50 groups using the suffix tree clustering (STC) algorithm [42]. Next, within each cluster, pairs of tweets are compared for similarity detection. Hashtags, mentions, and URLs in tweets can negatively affect the results when messages are compared for similarity. In other words, such items, which are of little importance when scoring the similarity between two tweets, may highly affect the similarity. These items are removed only for similarity calculation. Having a low similarity score between tweets means a weak relation between them. Weak relations make the graph of the tweets sparse and also increase the index computation time. To remove tweets with low similarity relations, the Cosine similarity algorithm is employed for similarity calculation. The algorithm takes two texts as input and outputs a number between 0 and 1, indicating the similarity between the given texts. The higher the score, the more similar the texts. A threshold of 0.6 is set to identify the similar tweets. To keep a tweet in dataset, it must have a similarity score higher than 0.6 with at least one other tweet.

3.3 Weight calculation

After extracting and introducing the features and index variables, we believe that all the variables are not of the same importance and they should be ranked according to their role in the index [17]. Because of this fact, the AHP algorithm, as a hierarchical multiple-criteria decision-making (HMCDM) method, is employed to assign weights to index variables. The algorithm is based on pairwise comparisons between parameters. Each comparison is marked within the scale of 1/9 for "least valued than," to 1 for "equal," to 9, for "absolutely more

[5] https://developers.google.com/maps/documentation/geocoding/intro [January 20, 2018].

Table 2 The decision matrix for user variables (each number represents the importance of the row variable to the column variable)

	tweetCnt									
tweetCnt	1	*flingCnt*								
flingCnt	6	1	*flerCnt*							
flerCnt	8	3	1	*userFavCnt*						
userFavCnt	3	1/3	1/7	1	*listCnt*					
listCnt	6	2	1/2	3	1	*lang*				
lang	1/4	1/7	1/9	1/4	1/8	1	*des*			
des	1/3	1/6	1/8	1/3	1/7	2	1	*loc*		
loc	1/3	1/6	1/8	1/3	1/7	2	1	1	*mimgUrl*	
mimgUrl	1/2	1/4	1/6	1/2	1/5	8	4	7	1	*imgUrl*
imgUrl	1/2	1/4	1/6	1/2	1/5	8	4	7	1	1

Table 3 The decision matrix for tweet variables (each number represents the importance of the row variable to the column variable)

	len						
len	1	*tweetFavCnt*					
tweetFavCnt	8	1	*retweetCnt*				
retweetCnt	9	2	1	*hashCnt*			
hashCnt	2	1/8	1/9	1	*menCnt*		
menCnt	3	1/8	1/9	½	1	*urlCnt*	
urlCnt	3	1/8	1/9	½	2	1	*dist*
dist	2	1/9	1/9	1/3	½	½	1

important than" [43]. The comparison between parameters depends on the variable and the application that variables are deployed in. Different judgments make different influential users, as well. Having no reference document about the importance of variables, in this study we use logical judgment to identify relative importance of variables. This approach is previously adopted in other studies as well [17, 44, 45]. For example, we consider that the number of times a tweet is marked as *favorite* is more important than the number of hashtags used in it. Tables 2 and 3 represent the decision matrices for user and tweet variables. The matrices are constructed using the definitions given in Sect. 3.1. For user variables, the comparison reveals the importance of *flerCnt*, as a representative for in-degree in graph theory, which dominates other variables. *listCnt* is ranked as the second most important variable improving the influence of a user. The next two important user variables are *flingCnt* and *userFavCnt*, both putting a negative effect on the user's influence.

The scores assigned to *retweetCnt* and *tweetFavCnt* in the matrix of tweets disclose the notable importance they impose compared to other variables. On the other hand, the length of tweet (*len*) is taken as the variable with the least importance in the comparison.

To verify the consistency of the judgment, we calculated the consistency ratio [19, 20]. The consistency values for tweet and user variables are measured as 0.09 and 0.10, respectively, revealing the consistency (less than or equal to 0.1) for both sets of variables. After tuning the importance of pairs of model features, the AHP algorithm calculates the weight of each

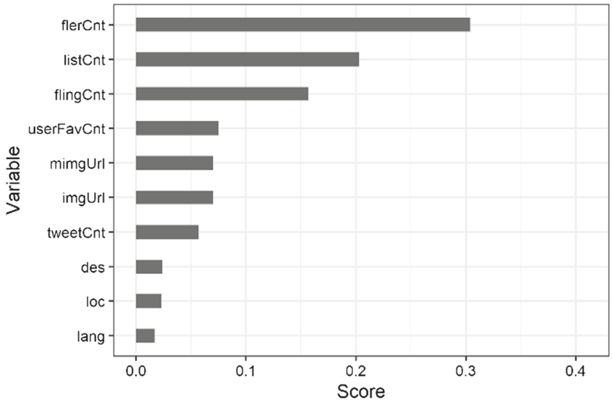


Fig. 3 The weights assigned to user variables according to the AHP algorithm

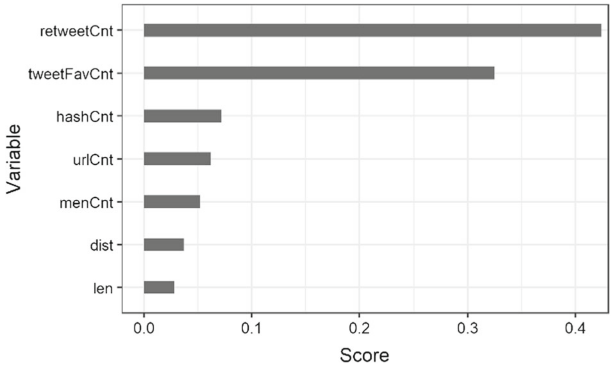


Fig. 4 The weights assigned to tweet variables according to AHP algorithm

parameter within the range of [0–1]. Figures 3 and 4 demonstrate the weights computed for user and tweet variables (from Tables 2, 3), respectively.

3.4 Influence indices

The proposed indices are composed of two components: (1) User Influence Index (UII) and (2) Tweet Influence Index (TII). The former is built upon the features extracted from a user. This index takes into account the number of tweets, favorites, followers, followings, lists and also the existence of the features related to user's profile, including profile mini image URL, profile image URL, description, language and location. The UII is presented to rank the most influential users, based only on users' features. The TII, on the other hand, focuses on the features extracted from the user's tweets. This index is created upon the features included in the tweet (the number of hashtags, mentions, URLs, length and time of the tweet, and its geographical distance from event location) and the ones shaped by the users reading it (the number of retweets, favorites, and its similarity to other tweets). The TII aims at ranking the tweets according to the influence they make.

When the two components are created, the general influence indices (GIIs) can be established. The GIIs are deployed to rank the most influential users based on not only her own characteristics, but also her tweets' influence.

3.4.1 User Influence Index

User Influence Index is based on the attributes introduced for users. As elaborated in Sect. 3.1, the number of followings and favorites and their corresponding variables must be considered as the ones reducing the influence of the user. Therefore, these variables have negative impact on influence, whereas they are applied positively in [17]. Equation (1) presents the mathematical formula deployed to calculate the influence index for users, where w_i represents the weight of each variable i according to the AHP (Sect. 3.3).

$$\begin{aligned} UI = {} & w_{flerCnt} * flerCnt + w_{tweetCnt} * tweetCnt + w_{listCnt} * listCnt \\ & + w_{imgUrl} * imgUrl + w_{des} * des + w_{mimgUrl} * mimgUrl \\ & + w_{loc} * loc + w_{lang} * lang - w_{flingCnt} * flingCnt - w_{userFavCnt} * userFavCnt. \end{aligned} \tag{1}$$

3.4.2 Tweet Influence Index

Tweet Influence Index is built upon the variables listed in Table 1. The *hashCnt*, *menCnt*, *len*, *urlCnt* and *dist* are the variables that preserve their effect on influence as time passes by. These variables create the *Tweet Static Influence Index* (*TSII*) [Eq. (2)]. The *retweetCnt*, *tweetFavCnt*, sum of sim_i, and *time* are the other features creating another index named *Tweet Dynamic Influence Index* (*TDII*) [Eq. (3)] in which *time* plays an important role. To elaborate, given a specific time after an event, if two tweets have equal numbers of *retweetCnt*, *tweetFavCnt*, and sum of sim_i, the one which is more recent is considered more influential. Otherwise, the older tweet should have sufficiently larger values in all three mentioned features to be superior in terms of TDII, despite the negative effect of time. Therefore, the index puts more emphasis on recent tweets with larger values in all features. Nevertheless, it allows tweets which, despite passage of time, still receive high attention from other users, to receive a high score.

$$\begin{aligned} \mathrm{TSII} = {} & w_{hashCnt} * hashCnt + w_{len} * len + w_{dist} * dist - w_{menCnt} * menCnt \\ & - w_{urlCnt} * urlCnt \end{aligned} \tag{2}$$

$$\mathrm{TDII} = \left(\sum_{i=1}^{n} sim_i + w_{retweetCnt} * retweetCnt + w_{tweetFavCnt} * tweetFavCnt \right) * time. \tag{3}$$

The sum of TSII and TDII builds the TII as in Eq. (4).

$$\mathrm{TII} = \mathrm{TSII} + \mathrm{TDII}. \tag{4}$$

3.4.3 General Influence Indices

Considering the UII and TII, two general influence indices ($\mathrm{GII_c}$ and $\mathrm{GII_a}$) are introduced in Eqs. (5) and (6), where c and a stand for the cumulative and averaged, respectively, and M

Username: andrewmfc
UserId: 167053099
TweetCreatedAt: Tue Nov 10 19:28:16 UTC 2015
Message: @PBruniges @MrMartinMorrish @MLong94 thanks Paul, I'm not sure other fans are taking this seriously.I'm going to light a fart #PrayforParis
TweetLenght: 139
TweetFavoritesCount: 1
RetweetCount: 0
TweetLanguage: en
GeoLocation: null
Place: null

Fig. 5 A sample tweet accompanied with its information extracted from Twitter API

is the number of tweets posted by a user. The cumulative GII (GII_c) takes into account the sum of TIIs, whereas the averaged GII (GII_a) measures the average score of TIIs.

The GIIs measure users' influence based on both their own features and their tweets features. The indices, in contrast to previous approaches, which generally consider only a number of features from a user and/or her tweets, take into account a larger number of the features from a user and her tweets in Twitter. This makes it more accurate to rank the users tweeting on a specific topic.

The difference between UII and TII is the effect of *time*. If all the parameters are similar in a period of time, tweets lose their influence as time goes by, while users' influence remains constant over the period.

$$\mathrm{GII_c} = \mathrm{UII} * \sum_{i=1}^{M} \mathrm{TII}_i \tag{5}$$

$$\mathrm{GII_a} = \mathrm{UII} * \frac{\sum_{i=1}^{M} \mathrm{TII}_i}{M}. \tag{6}$$

4 Experimental results and discussion

In this section, we first describe the collected dataset and then move on to presenting the actual results.

4.1 Dataset

Our analysis on identification of the most influential users in Paris attacks is based on the data collected from Twitter API. Twitterers used to post messages about the event by placing *#PrayForParis* and/or *#ParisAttacks* hashtags in their tweets. Since Paris attacks occurred in November 13, 2015, we collected the tweets containing the two hashtags for a 21-day period until December 3, 2015. The dataset is collected in two phases:

a. Phase I: In this phase, the tweets containing one of the *#PrayForParis* or *#ParisAttacks* hashtags posted in the period from November 13, 2015, to December 3, 2015, are collected in five intermittent days from November 17 to December 5. Figure 5 shows the sample data from a single tweet.
b. Phase II: The data related to the users whose tweets were gathered in Phase I are collected in this phase. A sample of the data is revealed in Fig. 6.

Totally, a number of 1,792,454 tweets in Phase I and the data from 59,133 users in Phase II were accumulated. Of the collected tweets, less than two percent contained the place they

Username: PBruniges
UserId: 849586333
TweetCount: 11588
FollowingCount: 1179
FollowersCount: 917
UserFavouritesCount: 526
ListedCount: 7
UserLanguage: en
UserCreatedAt: Thu Sep 27 16:53:11 UTC 2012
Description: I Own my own company support Millwall / Napoli I have 2 boys and granddaughter family naples via bermondsey
Location: South London
MiniProfileImageURL: http://pbs.twimg.com/profile_images/3306948881/e3bdb834bba6b8fd427f108327da8171_mini.png
ProfileImageURL: http://pbs.twimg.com/profile_images/3306948881/e3bdb834bba6b8fd427f108327da8171_normal.png
TimeZone: Amsterdam

Fig. 6 A sample user accompanied with its information extracted from Twitter API

Table 4 The correlation between the UII score and user variables

	TweetCnt	*FlingCnt*	*flerCnt*	*userFavCnt*	*listCnt*	*des*
UII	0.19	– 0.16	0.82	– 0.16	0.83	0.44

Table 5 The correlation between the TII score and tweet variables

	len	*dist*	*hashCnt*	*menCnt*	*urlCnt*	*tweetFavCnt*	*retweetCnt*
TII	0.13	– 0.04	0.65	– 0.16	– 0.67	0.22	0.24

are posted from. After preprocessing, discussed in Sect. 3.2, the number of tweets and users dropped to 31,815 and 867, respectively. Before calculations, values for each variable are normalized by dividing them by the maximum value of the corresponding variable. After calculation of the distance between the location the tweet has been sent from to Paris, the longest and shortest distances were 16,959 km and 8 km for Sydney, Australia and Saint Denis, France, respectively. For the tweets with no mentioned location, the average distance is assigned (4535 km).

4.2 Correlation between variables

Tables 4 and 5 represent the correlation between user and tweet variables with their corresponding scores, calculated as in Eq. (6). Since all the users collected after the preprocessing phase have filled their location, language and (mini) profile image, the corresponding variables of these binary features are excluded for calculation. As expected, UII possesses the strongest positive correlation with *flerCnt* and *listCnt* and negative correlation with the two negative variables, *userFavCnt* and *flingCnt*.

Table 5 represents that *hashCnt*, as a static variable, has created the strongest positive correlation with TII. The next positive correlations with it are measured as 0.24 and 0.22 for *retweetCnt* and *tweetFavCnt*, respectively. The strongest negative correlation with TII is demonstrated by *urlCnt*.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}. \tag{6}$$

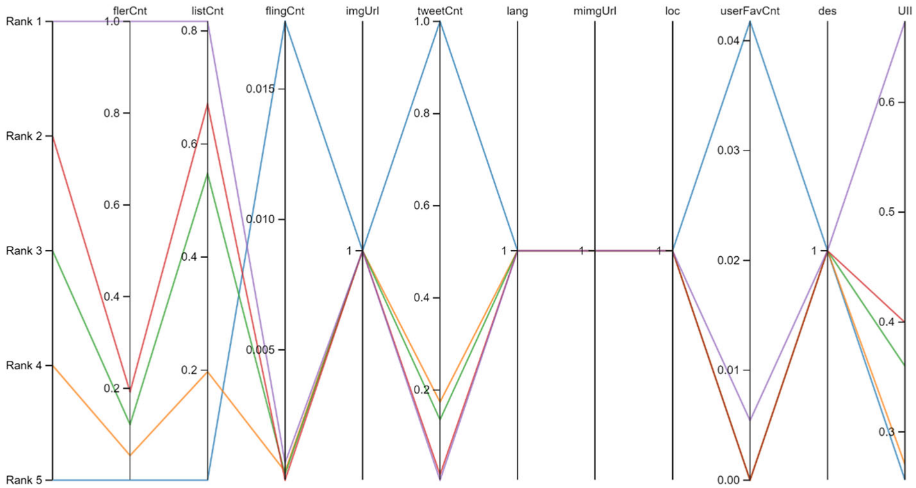


Fig. 7 Top five influential users and their index variables according to the UII (due to privacy of the users, their real usernames are not shown. Rank 1 points to the top-ranked user, etc.)

4.3 Study of User Influence Index

The index introduced in Sect. 3.4.1 is applied to the information of users tweeting on Paris attacks. The Twitterers are ranked based on the UII, and the five most influential users are shown in Fig. 7. While user *Rank 5* has posted the highest number of tweets (*tweetCnt*), due to the relative lower weight of this feature, it has not helped her to improve her ranking compared to other four Twitterers. She also has the highest number of followings (*flingCnt*) and marked favorites (*userFavCnt*) making her less influential. Having the highest number of followers (*flerCnt*) and lists (*listCnt*), and the low numbers of followings (*flingCnt*) and favorites (*userFavCnt*), *Rank 1* has been able to prove herself as the champion with a UII score being 1.7 times more than the runner-up's. Results also depict the fact that the top five users are trustful, too; i.e., they have set their location (*loc*), language (*lang*), description (*des*), and (mini) profile image (*mimgUrl* and *imgUrl*).

4.4 Study of Tweet Influence Index

This section covers the analyses over the influential tweets. First, the influential tweets during the 21-day period are detected and studied. Then, daily influential tweets are identified and the trend of users' daily influential tweets is analyzed. Next, the hashtags of the daily influential tweets are explored.

4.4.1 Influential tweets

Tweets are ranked according to the TII. After normalizing the variables, results of the top five influential tweets are presented in Table 6. The interesting point of the highest influential tweets is that all of them contain no mentions (*menCnt*) as a negative parameter. Also, their users have not posted the location where the tweet is sent from.

Table 6 The top five influential tweets and their variables according to the TII

Rank	Text	*retweetCnt*	*tweetFavCnt*	*dist*	*hashCnt*	*len*	*menCnt*	*urlCnt*	$t_{CreationDate}$	Score
1	Guys, it's time to #PrayForParis right now	1	1	0.26	0.06	0.17	0	0	2015-11-13 22:33:49	0.39
2	During difficult times, we need to remember the wise words from Dumbledore. #PrayForParis, https://t.co/dfzrf5aWNu ; https://t.co/ztswPJUq2J	0.15	0.11	0.26	0.06	0.55	0	0.40	2015-11-17 18:09:33	0.27
3	Thoughts and prayers with everyone in Paris. #prayforparis FR	0.22	0.11	0.26	0.06	0.25	0	0	2015-11-16 07:49:34	0.22
4	A survivor of the Paris attacks shares their story. #PrayForParis, https://t.co/KYzou6bOKV	0.38	0.26	0.26	0.06	0.36	0	0.20	2015-11-14 23:41:05	0.19
5	Just pray…. FR #PrayForParis	0.44	0.30	0.26	0.06	0.13	0	0	2015-11-13 23:26:20	0.18

According to the weights calculated for the TII variables, it was expected that the number of retweets and favorites make the most difference: having the highest number of retweets and favorites, the top-ranked tweet has made its way for being crowned as the champion. It should also be noted that the first ranked tweet is posted by a well-known celebrity. While the fifth tweet has higher number of retweets and favorites compared to the second, third, and fourth tweets, the latter tweets achieve a better rank as they are more recent. Another interesting point, as the results reveal, is that the *#ParisAttacks* hashtag does not appear in the top five influential tweets. The links used in the runner-up tweet point to a quotation taken from Albus Dumbledore, a fictional character in the Harry Potter series, depicting the importance of alliance among people in dreadful situations. The URL mentioned in the fourth tweet shares the story of an eyewitness in the Paris attacks.

4.4.2 Influence trend

Identifying the top ten influential tweets for each day, the influence trend is studied. A number of 121 users have tweets ranked as the top ten influential tweets on daily basis, among which only 29 users have posted at least two influential tweets in the 21-day period. Table 7 shows the influence trend of tweets posted by these users in each day. Users are sorted by decreasing order of number of influential tweets. In Table 7, 22 users have posted influential tweets on at least two different days. Also, 20 users have posted more than one influential tweet in a single day. As the results reveal, user 1 has the largest number of influential tweets (21 tweets) in 7 days. For 2 days (November 19 and November 28), in each single day, she has posted five influential tweets. Taking November 23 as the divider of the 21-day period into two halves, the champion user has posted influential tweets in both halves. Of the 121 users whose tweets are ranked as the top ten influential tweets on daily basis, 76% of them could not manage to keep posting influential tweets for a second day (and not shown in Table 7).

Of the users posted the top five influential tweets in the 21-day period, introduced in Sect. 4.4.1, only one of them has posted another tweet in the period and her tweet is not placed in the top ten daily influential tweets.

4.4.3 Hashtags of daily influential tweets

Considering the top ten daily influential tweets, indicated in the previous section, their hashtags were extracted. Due to the point that hashtags can be placed in the text in both small and capital letters, they were changed to small letters before the analysis. Figure 8 illustrates the hashtags used in the top ten influential tweets of each day from Table 7. #gunsense, #parisattacks, and #prayforparis are the first three mostly used hashtags with 12.1%, 9.7%, and 3.8% usage ratios, respectively. By the rank of the mostly used hashtags in Fig. 8, it is revealed that tweeting about what is heard (#gunsense), reporting the event and its place (#parisattacks) accompanied with expressing sympathy for the people of Paris (#prayforparis) are of high importance for Twitterers.

4.5 Study of General Influence Indices

Building the GII_c, none of the top five influential users in the UII (Fig. 7) are outlined in the top ten influential users of the GII_c. The same has occurred for the users posting the top five most influential tweets presented in Table 6. Figure 9 shows the results of the most

Table 7 The daily rank of influential users (numbers represent tweet rank in the given date)

User	November 13	November 14	November 15	November 16	November 17	November 18	November 19	November 20	November 21	November 22	November 23
1		4, 5, 8	3				1, 2, 3, 4, 6	1, 5			
2								7	3		2, 8
3		6	1, 2, 4, 6, 7								
4				8, 9							
5					5, 9, 10						7
6											
7											
8					3, 7, 8					1	
9											
10											
11											
12											
13									1		
14									6	7	
15											
16											
17				3, 6							
18						1, 2					
19								9, 10			
20									7		
21											3, 10

Table 7 continued

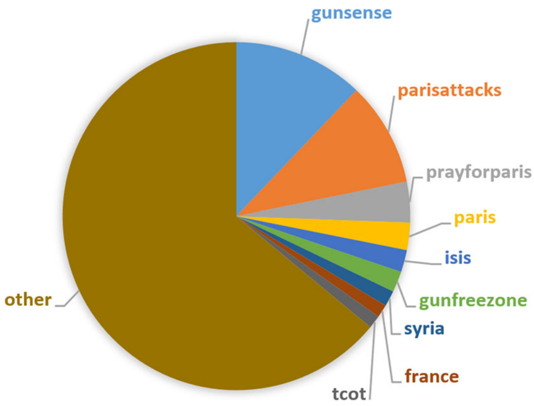
User	November 13	November 14	November 15	November 16	November 17	November 18	November 19	November 20	November 21	November 22	November 23
22											5
23											6
24											
25											
26											
27											
28											
29											

User	November 24	November 25	November 26	November 27	November 28	November 29	November 30	December 1	December 2	December 3
1				4, 9	2, 3, 4, 5, 6					5, 7, 8
2	3	2, 3, 10								
3										
4					1		2	1		2
5								9		
6	7					1, 2	1	10		
7		5	1, 2, 3, 7							
8										
9	2	8					4, 7			
10		1		7				6		1
11		7						7, 8		6
12			4					3, 4	3	
13		6							4	
14						3				

Table 7 continued

User	November 24	November 25	November 26	November 27	November 28	November 29	November 30	December 1	December 2	December 3
15	9								5, 6	
16			5				8, 10			
17										
18										
19										
20							3			
21										
22										4
23									2	
24	1, 10									
25	4, 8									
26			6							10
27				2	7					
28					8			2		
29						4, 8				

Fig. 8 The hashtags used in the top ten daily influential tweets



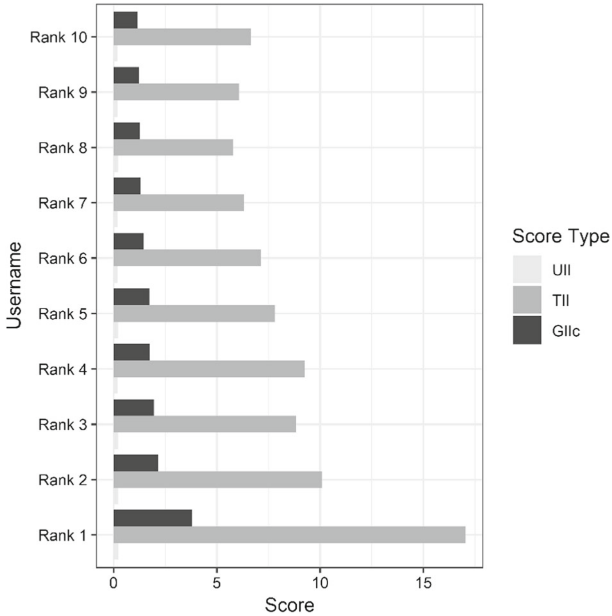


Fig. 9 Users rank in different indices (rankings based on the GII_c score)

influential users based on the GII_c. Results elaborate the fact that TII score dominates the UII and makes a stronger effect in the final ranking of the users.

The relation between the number of tweets during the 21-day period and users' rank, based on the GII_c, is presented in Table 8. Results state the high number of tweets influential users have posted during the period. Comparing the most influential tweet presented in Table 6 and the results illustrated in Table 8, it is disclosed that the most influential user has been able to overcome the TII score of the most influential tweet by posting many non-influential tweets. The sum of her tweets' score is 43 times more than the score of the most influential tweet. It is also shown that 45% of the tweets posted by users presented in Fig. 9 are published in the first 3 days after the event.

On the other hand, building the GII_a, as illustrated in Fig. 10, it is revealed that, in contrast to the GII_c, UII and TII scores do not present a big gap. The Twitterers in the top five list of

Table 8 The number of tweets for the top influential users, based on the GII_c, with each tweet's average influence score

Rank	#tweets in dataset during the 21-day period	Average influence score of each tweet
1	305	0.055
2	267	0.037
3	174	0.050
4	126	0.073
5	248	0.031
6	120	0.059
7	205	0.031
8	191	0.030
9	178	0.033
10	135	0.049

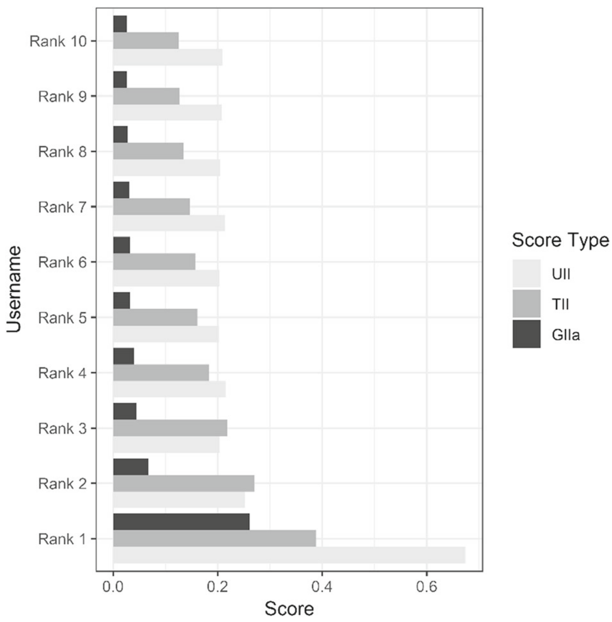


Fig. 10 Users rank in different indices (ranking based on the GII_a score)

TII are all present in the final list of GII_a with only one of them having exactly the same rank (Rank 1). Table 9 represents the total number of tweets each user from Fig. 10 has posted along with her average TII score. It is shown that, in contrast to the results extracted from the GII_c, the top influential users of the GII_a have posted at most two tweets during the 21-day period with much higher quality.

Taking a closer look at the Twitterers illustrated in Figs. 9 and 10, it is disclosed that publicly recognized sources (e.g., celebrities, politicians, etc.) are not included in the list of GII_c. Of its top ten users, 40% of them have introduced themselves as news agencies in their profile description. However, they are either local or not widely known news agencies. On the other hand, the top Twitterers GII_a are shaped by celebrities and well-known users, whose

Table 9 The number of tweets for the top influential users, based on GII_a, with each tweet's average influence score

Rank	#tweets in dataset during the 21-day period	Average influence score of each tweet
1	1	0.388
2	1	0.270
3	1	0.218
4	1	0.183
5	1	0.161
6	1	0.157
7	1	0.147
8	2	0.134
9	1	0.127
10	2	0.126

tweets gain a high number of retweets and favorites. Therefore, both their UII and TII scores are high.

In this study, we identify the most influential users after an event (Paris attacks). While well-recognized users might not have direct knowledge of the event they are tweeting on, this does not change the influence they can make on the society. These users, which may be influential regardless of the topic they are posting on, are identified by GII_a. Other less known users, who can expose their influence on some limited topics they are interested or directly involved in, may tweet more often on an event. Such users are identified using GII_c.

4.6 Comparison with other influence models

The models presented for measuring influence in social networks are shaped intuitively, as accurate and measurable benchmarks are rare for this purpose. For instance, the number of characters is taken into account as a parameter for identification of the most influential bloggers by Akritidis et al. [35], while Moh and Shola [46] have ignored the feature or in case of “Arab Spring,” the location of the tweet is considered as a feature of influence by Kumar et al. [7], while the same feature is not included by Overbey et al. [6].

As mentioned, the data collected for the model presented by Kumar et al. [7] are oriented around Arab Spring. The model takes the relevancy of tweet text and its geolocation into account. This causes to ignore other user and tweet features such as the number of followers or retweets. The methods presented by Cha et al. [16], Zengin Alp and Gündüz Öğüdücü [33] and Asadi and Agah [18] are other ways to identify the most influential Twitter users, which consider only a few parameters including the number of followers, retweets, and mentions. While the introduced features are of high importance, they do not cover the whole properties for identification of the most influential Twitterers. One of the best-presented previous models is the one introduced in [17]. While it uses a large range of features, the following issues are not solved in the model:

- Time, as a parameter affecting tweet influence, is not taken into account.
- The similarities between tweets are not considered.
- Although the number of followings conceptually acts as a negative parameter, it is included as a positive feature in the model, causing the influence to be increased. This issue is the same for the number of retweets.

Table 10 The comparison between the ranks of the top 20 users in [17] with our proposed indices

Rank ([17])	Rank (GII_c index)	Rank (GII_a index)
1	407	1
2	53	108
3	1	90
4	11	201
5	10	322
6	158	16
7	5	781
8	851	7
9	3	105
10	45	816
11	8	802
12	2	517
13	51	647
14	188	814
15	61	833
16	49	120
17	422	646
18	359	839
19	13	603
20	47	827

Table 10 represents the differences in user ranking of the top 20 influential Twitterers between the model introduced by Metra [17] and our proposed indices. According to the results, all of the first 20 influential users in [17] and GII_c have gained different ranks and there is a 60% difference in the list of the top 20 influential users; i.e., 60% of the users included in the top 20 influential of [17] are not in the same list of our proposed GII_c index. There is an average difference of 130 between the ranks assigned to all users by the two models. There is also an 85% difference in the list of the users ranked by [17] and GII_a; i.e., 85% of the users listed in the top 20 list of [17] are not ranked in the same list of GII_a. A bigger gap is also found between the two approaches in the rank of the users; 95% of the users in the two lists do not possess the same rank. There is an average difference of 444 between the users' rank assigned by [17] and the GII_a. The main differences between the two approaches are:

- The weights assigned to each variable in the model presented in [17] and our introduced indices.
- The differences between the features taken into consideration; e.g., in our indices, we have included *time* for building the indices, while this feature has been ignored in [17] or the similarity between tweets is one of the features we count on, whereas it does not exist in the previous model.
- The role of each feature is another difference between the our indices and [17]. For example, we have set the number of followings as a negative feature in our indices, while it plays a positive role in the model presented in [17].

To further study the differences in the two approaches, the users ranked 17 and 18 according to [17] are analyzed in more details. Both users have ranks larger than 350 in both of our proposed indices. The user ranked 17 has posted 33 tweets, among which eight contain a single URL, 21 have two URLs and only four are posted with no URLs. Mentioning URLs is considered negatively in our indices, resulting in a major rank difference. The other difference originates from another negative feature, namely MentionsCount. Seventy-six percent of the user's tweets contain at least one mention. Having such negative parameters has made the user to be ranked 422 and 646 in our proposed GII_c and GII_a indices, respectively. The user ranked 18 has posted 44 tweets, where 80% of them contain at least a single URL and over 72% of them contain at least a mention. She also has a high number of *userFavCnt* and *flingCnt* (155,791 and 51,138, respectively) making a negative effect on her UII score. Having such high number of negative features has dropped her influence to the rank of 359 and 839 based on our proposed GII_c and GII_a indices, respectively.

The model presented in [18] is focused only on user features to detect the most influential Twitterers. With this in mind, a comparison is drawn to highlight the differences between the rank of the first 20 influential users in [18] with our proposed UII, GII_c and GII_a indices. Results are illustrated in Table 11, showing that 8 of the users outlined in the top 20 list of the most influential users, based on the model presented in [18], are not in the same list of our UII ranking. On the other hand, it is revealed that there exists a 100% difference between the top 20 list of our proposed GII_c ranking with the one introduced in by Asadi and Agah [18]. The reason stems from the fact that in our index, the sum of tweets' influence is taken into account to identify the influential users, while this part is ignored in [18]. The comparison of the ranks between the model introduced in [18] and GII_a index shows a 65% difference between the ranks of the users.

5 Conclusion

In this paper, we introduced two new indices for identification of the most influential Twitter users. The adopted features deployed in the index were divided into two categories: user features and tweet features. The AHP algorithm was deployed to assign weights to variables. Next, two subindices for computing users' and tweets' scores in terms of influence were introduced. The final indices combined the two subindices to build unique measures for influence of each user. The results revealed the following conclusions on the proposed indices:

- The cumulative index (GII_c) creates the opportunity for the less recognized users to highlight themselves during an event.
- The averages index (GII_a) focuses on identification of the most influential celebrities, politicians or popular people in a social network.
- The comparison with other models on Twitter shows that the proposed approach covers a wider range of features required to rank and identify the influential users.

Next, the proposed indices were deployed on the 2015 Paris attacks dataset, yielding the following results:

- From the top 20 daily influential Twitterers, 76% cannot manage to keep their influence for a second day.
- *#gunsense* is the most common hashtag in the top ten daily influential tweets and *#parisattacks* is the runner-up.

News agencies have occupied four places among the top ten influential users, based on the GII_c, by posting relevant tweets about the event. In future work, effect of features with many missing values like distance may be studied and different methods of dealing with

Table 11 The comparison between the ranks of the top 20 users in [33] with our proposed UII, GII_c, and GII_a indices

Rank ([18])	Rank (UII index)	Rank (GII_c index)	Rank (GII_a index)
1	1	407	1
2	2	158	16
3	3	53	108
4	7	821	2
5	4	422	646
6	11	858	11
7	50	851	7
8	6	492	649
9	8	51	647
10	14	738	400
11	17	777	594
12	43	843	4
13	9	281	790
14	16	384	650
15	58	780	365
16	29	693	785
17	26	220	724
18	95	832	10
19	34	309	843
20	52	568	860

such features should be analyzed. Also comments of users for a tweet can be considered in computing the tweet influence score. Comments can be studied for their positive, negative, or neutral message in affecting the influence. Using multiple-criteria decision-making (MCDM) methods other than AHP is another option to extend this work and study other ranking measures.

References

1. Boyd D, Ellison NB (2008) Social network sites: definition, history, and scholarship. J Comput Commun 13:210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x
2. Sathik MM, Rasheed AA (2011) Social network analysis in an online blogosphere. Int J Eng Sci Technol 3:117–121
3. Lahuerta-Otero E, Cordero-Gutiérrez R (2016) Looking for the perfect tweet. The use of data mining techniques to find influencers on Twitter. Comput Human Behav 64:575–583. https://doi.org/10.1016/j.chb.2016.07.035
4. Himelboim I, Smith MA, Rainie L et al (2017) Classifying Twitter topic-networks using social network analysis. Soc Media + Soc. https://doi.org/10.1177/2056305117691545
5. Nakov P, Ritter A, Rosenthal S et al (2016) SemEval-2016 Task 4: sentiment analysis in Twitter. In: Proceedings of the 10th international workshop on semantic evaluation
6. Overbey LA, Paribello C, Jackson T (2013) Identifying influential Twitter users in the 2011 Egyptian revolution. In: Social computing, behavioral-cultural modeling, and prediction lecture notes computer science, vol 7812, pp 377–385. https://doi.org/10.1007/978-3-642-37210-0_41

7. Kumar S, Fred, Zafarani R, Liu H (2013) Whom should i follow? Identifying relevant users during crises. In: Proceeding HT'13 proceedings of the 24th ACM conference on hypertext and social media, pp 139–147. https://doi.org/10.1145/2481492.2481507
8. Xu B, Huang Y, Kwak H, Contractor NS (2013) Structures of broken ties: exploring unfollow behavior on Twitter. In: Ideology, politics, and social curation: Recent work Twitter, pp 871–876. https://doi.org/10.1145/2441776.2441875
9. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2:1–8. https://doi.org/10.1016/j.jocs.2010.12.007
10. Zhang M, Guo L, Hu M, Liu W (2017) Influence of customer engagement with company social networks on stickiness: mediating effect of customer value creation. Int J Inf Manag 37:229–240. https://doi.org/10.1016/j.ijinfomgt.2016.04.010
11. Jones JJ, Bond RM, Bakshy E et al (2017) Social influence and political mobilization: further evidence from a randomized experiment in the 2012 US presidential election. PLoS ONE 12:1–9. https://doi.org/10.1371/journal.pone.0173851
12. Francalanci C, Hussain A, Merlo F (2015) Representing social influencers and influence using power-law graphs. Appl Math Inf Sci 9:2453–2465. https://doi.org/10.12785/amis/090529
13. Francalanci C, Hussain A (2016) Discovering social influencers with network visualization: evidence from the tourism domain. Inf Technol Tour 16:103–125. https://doi.org/10.1007/s40558-015-0030-3
14. Francalanci C, Hussain A (2016) Influence-based Twitter browsing with NavigTweet. Inf Syst 64:119–131. https://doi.org/10.1016/j.is.2016.07.012
15. Servi L, Elson SB (2014) A mathematical approach to gauging influence by identifying shifts in the emotions of social media users. IEEE Trans Comput Soc Syst 1:180–190. https://doi.org/10.1109/TCSS.2014.2384216
16. Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in Twitter: the million follower fallacy. In: ICWSM'10: Proceedings of international AAAI conference on Weblogs and social, pp 10–17
17. Metra I (2014) Influence based exploration of Twitter social network. Polytechnic University of Milan, Milan
18. Asadi M, Agah A (2018) Characterizing user influence within Twitter. In: Advances on P2P, parallel, grid, cloud and internet computing, pp 122–132
19. Saaty TL (1980) Decision making for leaders: the analytical hierarchy process for decisions in a complex world. McGraw-Hill, New York
20. Saaty TL (2001) The analytic hierarchy process: planning, priority setting, resource allocation. RWS Publications, Pittsburgh
21. Freeman LC (1979) Centrality in social networks conceptual clarification. Soc Netw 1:215–239. https://doi.org/10.1016/0378-8733(78)90021-7
22. Claros I, Cobos R, Collazos CA (2016) An approach based on social network analysis applied to a collaborative learning experience. IEEE Trans Learn Technol 9:190–195. https://doi.org/10.1109/TLT.2015.2453979
23. Sabidussi G (1966) The centrality index of a graph. Psychometrika 31:581–603. https://doi.org/10.1007/BF02289527
24. Liu J, Xiong Q, Shi W et al (2016) Evaluating the importance of nodes in complex networks. Phys A Stat Mech Appl 452:209–219. https://doi.org/10.1016/j.physa.2016.02.049
25. Hosseini-Pozveh M, Zamanifar K, Naghsh-Nilchi AR (2017) A community-based approach to identify the most influential nodes in social networks. J Inf Sci 43:204–220. https://doi.org/10.1177/0165551515621005
26. Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. J Math Sociol 2:113–120. https://doi.org/10.1080/0022250X.1972.9989806
27. Brin S, Page L (1998) The anatomy of a large scale hypertextual Web search engine. Comput Netw ISDN Syst 30:107–117. https://doi.org/10.1016/S0169-7552(98)00110-X
28. Kleinberg JONM (1999) Authoritative sources in a hyperlinked. Environment 46:604–632
29. Du Y, Gao C, Hu Y et al (2014) A new method of identifying influential nodes in complex networks based on TOPSIS. Phys A Stat Mech Appl 399:57–69. https://doi.org/10.1016/j.physa.2013.12.031
30. Bian T, Hu J, Deng Y (2017) Identifying influential nodes in complex networks based on AHP. Phys A Stat Mech Appl 479:422–436. https://doi.org/10.1016/j.physa.2017.02.085
31. Leung CK-S, Tanbeer SK, Cameron JJ (2014) Interactive discovery of influential friends from social networks. Soc Netw Anal Min 4:154. https://doi.org/10.1007/s13278-014-0154-z
32. Peng S, Yang A, Cao L et al (2017) Social influence modeling using information theory in mobile social networks. Inf Sci (NY) 379:146–159. https://doi.org/10.1016/j.ins.2016.08.023

33. Zengin Alp Z, Gündüz Öğüdücü Ş (2018) Identifying topical influencers on Twitter based on user behavior and network topology. Knowl-Based Syst 141:211–221. https://doi.org/10.1016/j.knosys.2017.11.021
34. Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Soc Netw 23:191–201. https://doi.org/10.1016/S0378-8733(01)00038-7
35. Akritidis L, Katsaros D, Bozanis P (2011) Identifying the productive and influential bloggers in a community. IEEE Trans Syst Man Cybern Part C Appl Rev 41:759–764
36. Inman J (1835) Navigation and nautical astronomy: for the use of British Seamen. Kessinger Publishing, London
37. Chartrand G (1984) Introductory graph theory. Dover Publications, New York
38. Myers SA, Sharma A, Gupta P, Lin J (2014) Information network or social network? The structure of the Twitter follow graph. In: Proceeding WWW'14 companion proceedings of the 23rd international conference on World Wide Web. Seoul, Korea, pp 493–498
39. Grandjean M (2016) A social network analysis of Twitter: mapping the digital humanities community. Cogent Arts Humanit 3:1–14. https://doi.org/10.1080/23311983.2016.1171458
40. Agarwal N, Liu H, Tang L, Yu PS (2008) Identifying the influential bloggers in a community. In: Proceedings of the international conference Web search web data mining—WSDM'08, pp 207–218. https://doi.org/10.1145/1341531.1341559
41. Chu Z, Gianvecchio S, Wang H, Jajodia S (2012) Detecting automation of Twitter accounts: are you a human, bot, or cyborg? IEEE Trans Dependable Secur Comput 9:811–824. https://doi.org/10.1109/TDSC.2012.75
42. Stefanowski J, Weiss D (2003) Carrot 2 and language properties in web search results clustering. In: Lecture notes in artificial intelligence advanced web intelligence proceedings of the first international Atlantic Web intelligence conference, vol 1, pp 1–10. https://doi.org/10.1007/3-540-44831-4_25
43. Vaidya OS, Kumar S (2006) Analytic hierarchy process: an overview of applications. Eur J Oper Res 169:1–29. https://doi.org/10.1016/j.ejor.2004.04.028
44. Choudhary P (2016) Ranking Terrorist Nodes of 26/11 Mumbai Attack using analytical hierarchy process with social network analysis. In: 11th annual symposium on information assurance (ASIA'16)
45. Liebowitz J (2005) Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. J Knowl Manag 9:76–86. https://doi.org/10.1108/13673270510582974
46. Moh TS, Shola SP (2013) New factors for identifying influential bloggers. In: Proceedings—2013 IEEE international conference on big data, Big Data 2013, pp 18–27

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Behzad Rezaie received his B.Sc. and M.Sc. from Shahrood University of Technology in Computer Engineering (Software Engineering) and Artificial Intelligence, in 2012 and 2015, respectively. His main areas of research are data mining and social network analysis.

Morteza Zahedi is currently an Assistant Professor at Shahrood University of Technology. His projects are oriented around human and computer interaction, pattern recognition, image and video processing, and machine vision, in which statistical methods are deployed. He received his PhD from the RWTH Aachen University. He also supervises academic and industrial projects from Iran and European countries.

Hoda Mashayekhi is an Assistant Professor at the Faculty of Computer Engineering, Shahrood University of Technology. She received her PhD from Sharif University of Technology in 2013. Her research interests include parallel and distributed computing, massive data analysis, machine learning, decision making, and recommendation systems.