TRank: ranking Twitter users according to specific topics

Manuela Montangero
Dipartimento di Fisica, Informatica e Matematica
Università di Modena e Reggio Emilia
Email: manuela.montangero@unimore.it

Marco Furini
Dipartimento di Comunicazione ed Economia
Università di Modena e Reggio Emilia
Email: marco.furini@unimore.it

***Abstract*—Twitter is the most popular real-time micro-blogging service and it is a platform where users provide and obtain information at rapid pace. In this scenario, one of the biggest challenge is to find a way to automatically identify the most influential users of a given topic. Currently, there are several approaches that try to address this challenge using different Twitter signals (e.g., number of followers, lists, metadata), but results are not clear and sometimes conflicting. In this paper, we propose TRank, a novel method designed to address the problem of identifying the most influential Twitter users on specific topics identified with hashtags. The novelty of our approach is that it combines different Twitter signals (that represent both the user and the user's tweets) to provide three different indicators that are intended to capture different aspects of being influent. The computation of these indicators is not based on the magnitude of the Twitter signals alone, but they are computed taking into consideration also human factors, as for example the fact that a user with many active followings might have a very *noisy* time lime and, thus, miss to read many tweets. The experimental assessment confirms that our approach provides results that are more reasonable than the one obtained by mechanisms based on the sole magnitude of data.**

I. Introduction

The social media scenario is increasingly used to communicate, collaborate and to share information: people discuss about brands, products, services, personal preferences, issues, etc. [1], [2]. Not surprisingly, social media are used to find signs for future trends, society interests, emerging big changes [3], [4], [5], [6]. Within this scenario, micro-blogging is an emerging form of communication that allows users to publish brief message updates through different channels, from Web to mobile. Twitter is the most popular real-time micro-blogging service: it is a social network used to provide and obtain information at rapid pace. The service allows users (named Twitterers) to publish messages (named tweets) that cannot be longer than 140 characters and thus they are fast to read and write. There is a large variety of users posting messages (from ordinary individuals to important people, from public to private organizations, etc.) and tweets contain a wide variety of information, ranging from relevant to useless [7], [8], [9]. As of 2014, Twitter has more than 250 million monthly active users, who post around 500 million tweets per day [10].

Twitter main difference with other social networks is the idea that tweets are not meant to be uniquely directed to another specific user or group of users, but they are available to all the people who are willing to *listen* to what a user has to say. These users are called *followers* and the action they are performing is the one of *following* a user. As a consequence, followers can see all tweets of those they are following in their profile timeline in the order of their arrival. Hence a tweet can be thought as a broadcast to all (uninvited and uncontrolled) followers.

In this scenario, one of the biggest challenge is to find a way to determine the most influential user(s) in Twitter. Indeed, the credibility of the sources affects the user behaviors and opinions [11]. Several fields may receive benefits from solving this challenge. For instance, in the business scenario, a company may want to contact the most influential users in order to spread information about a product; in marketing, an advertising agency may want to find the most influential users able to increase the efficiency of a campaign; in politics, a politician may want to engage those Twitter users that are more influential about certain topics.

Since it is not realistic to think there are universally context independent authoritative users, the challenge is to automatically identify authoritative user(s) given a specific topic. Currently, most approaches interpret a twitterers influence as the number of followers he/she has (the more the followers, the higher the popularity), assuming that any tweet sent by a user with a large number of followers will reach a large number of other users [12]; other approaches rely on the information provided by the users in their Twitter profile (*e.g.*, in the bio or short autobiography or personal information) and others analyze the characteristics of the user relationship network and the tweeting activity of users [13].

Different studies highlighted that these approaches do not provide satisfactory results for many different reasons: the mere followers number is not such a good indicator because a significantly large percentage of users follow back their followers just for courtesy and therefore the following relation does not come with meaningful insight on influence, as there is a high chance that many tweets will never be read [8]. Furthermore, Twitter has become a target for link farming, where users, especially spammers, try to acquire large numbers of follower links in the social network [12], [14]; another important reason is that the information provided in Twitter (e.g., the bio data) are not completely reliable: many people provide fake information, fake contents or fake images [15]. Finally, another important reason is that Twitter has rapid changes in the underlying graph: users not even existing prior to an event might become very popular in a very short time, or remain very popular for a very short time and graph based algorithms are computationally too demanding to be up to date

978-1-4799-6390-4/15/$31.00 ©2015 IEEE

in such dynamic scenarios, even if they succeed in detecting "long term" popular users.

In this paper we propose TRank, a novel method designed to address the problem of identifying the most influential twitterers on specific topics, where the topics are specified by hashtags. The novelty of our approach is that it is not based on the simple magnitude of data like the absolute number of followers or of retweets. Conversely, different Twitter signals are used to compute three different indicators that are intended to capture different aspects of *being influent*: the follower influence, the retweet influence and the favorite influence. It is to note that none of these indicators are based on the simple magnitude of data, but they are computed taking into consideration human factors, as for example the fact that a user with many active followings might have a very *noisy* time lime and, thus, miss to read many tweets. It is also to note that TRank uses Twitter signals that represent both the user (e.g., the number of active followers and the number of active followings of those users that marked as favorite one tweet of the user) and the user's tweets (e.g., the number of retweets and of favorites).

The remainder of this paper is organized as follows. In Section II we briefly review related works; in Section III we present details and characteristics of our proposal, whereas the evaluation process is described in Section IV. Conclusions are drawn in Section V.

II. Related work

In the literature, several studies attempt to measure the influence of Twitter users and thereby identify influential (or expert) users. One of the first studies that attempted to find influential Twitter users was the one of Kwak et al. [2]: they compared three different measures of influence (number of followers, page-rank, and number of retweets) finding that the ranking of the most influential users by the number of followers and by PageRank are similar and the ranking by retweets differs. A subsequent study of Weng et. al. [8] showed that the absolute number of followers is not a good indicator of the influence a user can have in the Twitter scenario. Similarly, Stringhini et al. [12] showed that a growing industry of Twitter follower markets provides followers for sale and some markets use fake accounts to boost the follower count of their customers. Cha et al. [16] compared three different measures of influence (number of followers, number of retweets, and number of mentions) and found that the most followed users did not necessarily score highest on the other measures. Weng et. al. [8] proposed a Page-Rank like algorithm, named TwitterRank, that uses both the Twitter graph and the processed information from tweets to identify experts in particular topics. The study assumed that the more a friend publishes, the higher portion of tweets the follower will read from it and this increase the Twitter influence. Pal et. al. [13] used 15 different features to rank Twitter users and, to speed-up the computation, instead of using network analysis techniques, they use probabilistic clustering to identify a small set of authors with a desirable configuration across the set of proposed metrics. Ghosh et a. [17] proposed a different approach: instead of focusing on some Twitter features, they followed a semantic approach focusing on the Twitter Lists. The key assumption was that Twitter Lists of a specific Twitter users contain enough metadata to infer the user's topics of expertise, which in turn enables them to identify topical experts. This approach might be reasonable in a scenario where all the Twitter users are real, and where all the users are honest, but, unfortunately, in the Twitter scenario there are a lot of fake contents, as shown in by Gupta et al. [15].

In addition to the above studies, there are different services available on the Web that aim at identifying the most influential users in Twitter. One of this services is the Twitter "who to follow" service, where one can search for people according to the people one already follows [18]. Another service is the Klout score[1], that rates users according to different signals (e.g., number of followers, mentions, lists, etc) a user has. It is to note that the Klout score is not related to a specific topic, but is a general score that aims at measuring how influential a user is in the social media scenario.

All the above studies and approaches primarily rely on the magnitude of numbers or on the information provided by the user itself (e.g., account name and bio, tweets posted, etc.). In contrast, our approach does not just consider the magnitude of the signals, but combines different Twitter signals (that represent both users and user's tweets) taking also into consideration reasonable assumptions on the way humans use Twitter (e.g., humans can not possibly constantly read thousands of tweets per day, silent users do not contribute to the number of tweets their followers receive).

III. Our proposal

In this paper we propose TRank, a novel method designed to address the problem of identifying the most influential twitterers on specific topics, where the topics are specified by hashtags. Our novel approach relies on a combination of data that accounts for human behaviors, limitations and abilities. Such data are derived from the user we want to rank and from other users being in some kind of relation with him/her (not necessarily followers and following), and on the user's tweets.

In details, let us consider a user U that sent tweet T containing the hashtag h under consideration (e.g., "#toys"). To estimate the influence of U with respect to topic h we take into consideration the following Twitter signals:

- the number of *followers* of user U;
- the number of *followings* of all the U followers;
- the number of retweets that T got in a given interval of time $\mathcal{I}$;
- the number of *favorite* the tweet got;
- the number of *followers* of those user that put T among their favorites.

These signals are used to compute the three different indicators that are intended to capture different aspects of *being influent*: the follower influence, the retweet influence and the favorite influence. The computation of the indicators is based on the following observations.

[1] https://klout.com/corp/score

Followers influence (FI): As mentioned in the previous sections, it is commonly assumed that the more followers a user has, the more influence his/her tweets have. While this might have been correct in the past, nowadays this measure is no longer sufficient to measure the influence. For example, there exists a follower market, where spammers try to acquire large number of followers in order to impact the ranking of the user's tweets by search engines. Furthermore, we observe that the more followings a user has, the smaller the chance that the user reads all tweets of those he/she is following. Indeed, if the tweet scrolls far down in the time line it might be easily missed. Our indicator tends to penalize users with too many active followings.

Retweet influence (RI): The number of retweets is also an important indicator of the influence a user has. Needless to say, the higher this number the more viral the tweet was. However, the magnitude of this number may be misleading. If the tweet has been written by user U, and if the number of retweets of a specific tweet T is much higher than the number of users following U, this is a clear indication that the tweet has reached far beyond the close neighborhood of user U and this is a good indicator that the tweet content was considered important by many users. Our indicator tends to favor such tweets, as it has been observed that people often retweet their followings messages also for "friendship" or "homophily", that is, a user retweets just to do a favor and not because of the tweet content [8].

Favorite influence (FVI): Twitter gives users the ability to mark a tweet as *favorite*. When this occurs the marked tweet is automatically placed in a *Favorite List* owned by the user that marked the tweet and such list can be seen by all his/her followers. There are several and different reasons people make a tweet as favourite [19], but the action is intentional and conscious, meaning, with high probability, the tweet has been read and considered interesting. On the other hand, the favorite list is not often visited (i.e., it is not straightforward that all the followers will visit the Favorite List, as this operation implies a follower to explicitly visit the user profile page). Hence, being placed in a favorite list gives a tweet authority, but it does not automatically guarantee visibility. We try to mediate between these two conflicting aspects of marking a tweet as favorite, by measuring visibility using the log of the number of followers of those users that marked the tweet as favorite.

According to the above considerations, TRank works in three main steps:

1) Given an hashtag h and time interval $\mathcal{I}$, select all tweets sent in interval $\mathcal{I}$ that contain hashtag h.
2) For every such tweet, determine the user that wrote it and score the user with three indicators FI, RI and FVI.
3) Select and output the best scoring users.

Before presenting details of the indicators computation, for the sake of presentation, we summarize all the used notations in Table I and we observe that when talking about an *active* Twitter user, we refer to a user that has sent at least one tweet in the time interval $\mathcal{I}$.

Notation	Description
T	Tweet containing hashtag h
U	User that wrote T
$\mathcal{I}$	time interval
u_i	other user
aFw_i	number of active users that u_i is following
aR_U	number of active followers of user U
Fs_i	number of followers of user u_i
Rt_T	number of retweets T got
Fv_T	number of favorites T got
FI_T	Tweet T followers influence indicator
FI_U	User U followers influence indicator
RI_T	Tweet T Retweet influence indicator
RI_U	User U Retweet influence indicator
FVI_T	Tweet T Favorite influence indicator
FVI_U	User U Favorite influence indicator

TABLE I. NOTATION USED IN SECTION III.

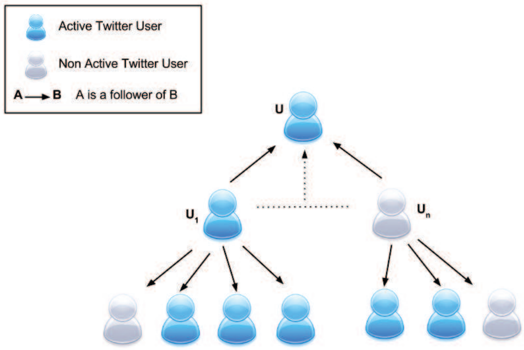

Fig. 1. The Twitter scenario: a user may have followers, but some of these may be non active.

- **Followers Influence (FI):** Let $u_1, \ldots, u_n$ be the followers of U and let aFw_i be the number of users that u_i is following and that have been active in time interval $\mathcal{I}$. We assume that the chance that u_i actually reads a tweet coming from U is proportionally inverse with aFw_i. Hence, we estimate user U influence with respect to the number n of the user followers as:

$$FI_U = \sum_{i=1}^{n} \frac{1}{aFw_i}.$$

Observe that $aFw_i \geq 1$, for all $i = 1, \ldots, n$, because user u_i has at least user U among his/her followings. As an example, with respect to Figure 1 and user U, we have that $FI_U = 1/4 + \ldots + 1/3$.
We have that $FI_U \geq 0$ and the higher FI_U, the higher the user influence. A large FI_U score might be reached by a user with a very large number of followers each having possibly large number of active following, or by a user with a smaller number of followers each having a possibly small number of active followings.
Observe that we are interested only in counting u_i active followings and not in the fact that u_i is active or not. This because the number of aFw_i influences the number of tweets u_i receives in the considered time interval. A large number of tweets lowers the chance that u_i actually reads a specific tweet. On the contrary, the fact that u_i did not tweet in that time interval does not imply that he/she did not read tweets.

- **Retweet Influence (RI):** Let T be one (possibly *the* *one*) of the tweets written by U during time interval $\mathcal{I}$. We estimate the tweet influence with respect to the number of retweets Rt_T of tweet T as:

$$RI_T = \frac{Rt_T}{aF_U},$$

where aF_U is the number of followers of U that have been active in the time interval $\mathcal{I}$. If $aF_U = 0$, we set RI_T to zero.
Observe that, whenever $0 \leq RI_T < 1$, the tweet had less retweets that active followers; conversely, if $RI_T > 1$ the tweet had been retweeted also by users that are not U direct followers and this indicates that the tweet content was considered interesting. Therefore the user who wrote the tweet can be considered influential with respect to the tweet topic.
Since, in the considered time interval, a user might have written more than one tweet, the retweet influence score of the user is computed as the average of the retweet influence scores of his/her tweets. Hence, let $T_1, \ldots, T_p$ be the p tweets of user U, then we define the user retweet influence as:

$$RI_U = \frac{\sum_{i=1}^{p} RI_{T_i}}{p}.$$

Obviously, if $p = 1$, we have that $RI_U = RI_{T_1}$. Observe that $RI_U \geq 0$ and the higher RI_U, the higher the user influence derived by twitting $T_1, \ldots, T_p$.

- **Favorite Influence (FVI):** Let $u_1, u_2 \ldots, u_m$ be the list of the m users that marked tweet T (written by U) as their favorite, and let Fs_i be the number of followers user u_i has. Observe that, even if unrealistic, we can not exclude the case in which $Fs_i = 0$. Hence, we estimate the tweet influence FVI with respect to the number of T favorites as:

$$FVI_T = \sum_{\substack{i=1 \\ Fs_i \neq 0}}^{m} \log(Fs_i).$$

Similarly to the retweet indicator, also in this case more than one tweet might have been written by the same user and, analogously, we compute the favorite influence score for user U as the average of those of his/her r tweets $T_1, \ldots, T_r$ written in time interval $\mathcal{I}$:

$$FVI_U = \frac{\sum_{i=1}^{r} FVI_{T_i}}{r}.$$

We have that $FVI_U \geq 0$, the higher FVI_U the higher the user influence, and that if $r = 1$ then $FVI_U = FVI_{T_1}$.

We pointed out that some of the tweets selected in the first step may have been written by the same user. In such case, we compute RI_U and FVI_U as the average of the user corresponding tweets scores (observe that FI is tweet independent). In this way we intend to capture the overall influence of the user during the whole time interval: if all his/her tweets had a high success his/her influence is high; if some of his/her tweets had high success but others were ignored, his/her influence is (reasonably) affected by the latter.

Observe that the three indicators are computed taking into account information concerning small (not necessarily connected) portion of the Twitter graph. Indeed, FI explores users at distance two from U, while RI and FVI use data retrieved by users that are not necessarily followers and/or following of user U.

All needed data can be quickly retrieved and indicators can be computed with limited computational powers.

To select and output the best scoring users, TRank operates in the following way: let $U_1, \ldots, U_\ell$ be the list of users that sent a tweet containing the hashtag h during time interval $\mathcal{I}$. For each such user we have computed the triplet $(FI_{U_i}, RI_{U_i}, FVI_{U_i})$ as described before. Then we order the users in decreasing order by means of each of the three indicators. We obtain three lists and we select the first k top scoring users of each list (where k is a parameter that may vary according to the specific need of finding influence user on Twitter). If one of the score is zero, the users is canceled from the corresponding list. Selected users are thus divided into three categories: (1) *Highly influential users*, containing those users with all three indicators in the top scoring positions; (2) *Influential users* containing those users with two out of three indicators in the top scoring positions; (3) *Potential influential users*, containing those users with only one indicator in the top scoring positions.

IV. Experimental Assessment

In this section we present results obtained while applying our method to a Twitter driven datasets: we perform different hashtag searches to identify the most influential users on specific topics. As already discussed in literature (see for example [8]), it is difficult to compare different approaches because it is difficult to interpret the results. However, in the following we compare TRank with results obtained while analyzing influential users based on the absolute number of user's followers.

In the following, we report three different examples: #winxclub (a popular animated television series), #oliodipalma (palm oil, a topic that is becoming popular for health and ethical issues related to the large use of this product) and #thesimpsons (a popular animated television series).

For each search, we retrieve the most recent tweets (published up to one week before[2]), and for each tweet we compute how the sender is influential with respect to the metric defined in the previous section. This means that we compute: i) the followers influence (FI) by computing the number of his/her followers and by checking the number of active users that are following this user; ii) the retweet influence (RI) by calculating the ratio between the number of retweets obtained by the considered tweet and the number of active followers of the user who wrote the tweet, and iii) the favorite influence (FVI) by computing the log of the number of active followers that marked the tweet as favoirte in their account.

Table II and Figure 2 show the results obtained when searching for influential users of the topic #winxclub. Observe that if one uses the absolute number of followers, then user X_8

[2]As Twitter limits the number of developer requests per day, we selected a time interval that guaranteed experiments to end in a short time.

Tweet	User	Followers	FI_U	RI_T (Rt_T)	RI_U	FVI_T (Fv_T)	FVI_U
#1	X_1	1270	**7.64**	.0071 (7)	**.0071**	8.10 (4)	**8.10**
#2	X_2	52	3.02	.0295 (1)	**.0295**	2.84 (2)	**2.84**
#3	X_3	2292	**8.41**				
#4	X_4	129	3.66				
#5	X_5	285	1.71				
#6	X_6	814	**8.51**	.0035 (4)	**.00175**	5.07 (1)	**2.535**
#7	X_7	23	0.49				
#8	X_8	4066	1.49				
#9	X_6	814	8.51	0	.00175	0 (0)	2.535
#10	X_9	0	0				

TABLE II. STATISTICS AND INDICATORS FOR THE DATASET CONCERNING THE HASHTAG #WINXCLUB. EMPTY ENTRIES ARE ZERO ENTRIES. IN BOLD WE HIGHLIGHT THE FIRST 3 TOP SCORING VALUES FOR EACH INDICATOR. REAL ACCOUNT NAMES HAVE BEEN OBSCURED FOR PRIVACY. OBSERVE THAT USER X_6 APPEARS TWICE AS HE/SHE WROTE TWO TWEETS.

Tweet	User	Followers	FI_U	RI_T (Rt_T)	RI_U	FVI_T (Fv_T)	FVI_U
#1	X_1	1208	**5.85**				
#2	X_2	217	4.83				
#3	X_3	55	**6.77**				
#4	X_1	1208	5.85				
#5	X_4	235	1.68				
#6	X_5	73	0.85	.0185 (1)	**.0185**	2.65 (1)	**2.65**
#7	X_6	384	**8.07**				
#8	X_7	335	2.34				
#9	X_8	809	2.58	.0240 (14)	**.0240**	11.66 (4)	**11.66**
#10	X_9	40	0.67				

TABLE III. STATISTICS AND INDICATORS FOR THE DATASET CONCERNING THE HASHTAG #OLIODIPALMA. EMPTY ENTRIES ARE ZERO ENTRIES. IN BOLD WE HIGHLIGHT THE FIRST 3 TOP SCORING VALUES FOR EACH INDICATOR. REAL ACCOUNT NAMES HAVE BEEN OBSCURED FOR PRIVACY. OBSERVE THAT USER X_1 APPEARS TWICE AS HE/SHE WROTE TWO TWEETS.

is the most influent one, user X_3 is the second one and user X_1 is the third one. Looking at the FI_U score, instead, the most influential is user X_6, the second one is user X_3 and the third one is user X_1. In particular, the most influencer user is different. The other two indexes identify one extra influential user (X_2) that was not found by FI_U.

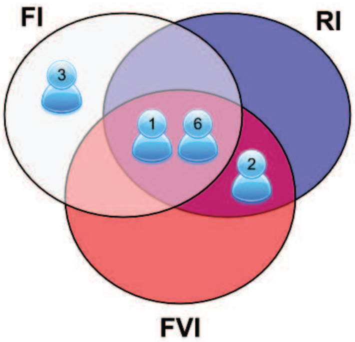

Fig. 2. Selected influential users classification for the hashtag #winxclub considering the first 3 top scoring users per indicator. Users X_1 and X_6 are classified as highly influencer users, X_2 as influential user, while X_3 as potentially influencer one.

Table III and Figure 3 show the results obtained when searching for influential users of the topic #oliodipalma. Also in this case the ranking computed according to the absolute number of followers (X_1, X_8, X_6) differs from the ranking computed by considering the FI_U indicator (X_6, X_3, X_1). In this case the set of users identified by the two indicators RI_U and FVI_U is the same (X_5, X_8), because these users wrote the only two tweets that have been retweeted and favored. Interestingly enough, this set has no intersection with that of indicator FI_U.

Table IV and Figure 4 show the results obtained when searching for influential users of the topic #thesimpsons. As this dataset has a larger number of tweets, we selected the first 4 top scoring users per indicator. Again, the number of followers and FI_U differ in their top scoring users. This time, the ranking computed according to the absolute number of number of followers $(X_{12}, X_8, X_{13}, X_3)$ differs from the ranking computed by considering the FI_U indicator (X_4, X_8, X_5, X_7). In this case the set of users identified by the two indicators RI_U and FVI_U are different: RI_U identifies X_{10} and X_{15}, whereas FVI_U identifies X_{15}, X_{10}, X_{12} and X_7.

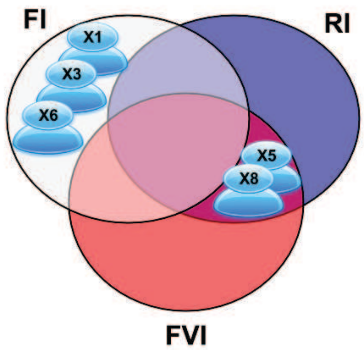

Fig. 3. Selected influential users classification for the hashtag #oliodipalma considering the first 3 top scoring users per indicator . There are no user classified as highly influencer, users X_5 and X_8 are classified as influential user, while X_1, X_3 and X_6 as potentially influencer ones.

Finally, we make some general observations. First, even if most of the tweets were not retweeted or marked as favorite, for those that had, difference in such numbers reflected in an interesting RI and FVI indicator variability. Second, ranking according to FI_U is sensibly different from ranking according to the absolute number of followers, either because the set of user is different or because their order in the ranking is different. This is a positive result, as indicators based on the sole number of followers proved not to be accurate [8]. Third, influential users computed according to FI_U are often distinct from influential users computed according to the other

Tweet	User	Followers	FI_U	RI_T (Rt_T)	RI_U	FVI_T (Fv_T)	FVI_U
#1	X_1	21	4.03				
#2	X_2	320	0.76				
#3	X_3	717	3.39				
#4	X_4	581	**9.67**				
#5	X_5	209	**6.83**				
#6	X_6	118	6.1				
#7	X_7	320	**4.76**			2.17 (1)	**2.17**
#8	X_8	1664	**7.53**				
#9	X_9	27	2.65				
#10	X_{10}	208	2.81	.0054 (1)	**.0054**	3.00 (1)	**3.00**
#11	X_{11}	154	2.29				
#12	X_{12}	2468	4.49			2.38 (1)	**2.38**
#13	X_{13}	1222	4.74			2.16 (1)	2.16
#14	X_{14}	22	1.21				
#15	X_{15}	372	1.92	.0041 (1)	**.0041**	4.83 (2)	**4.83**

TABLE IV. STATISTICS AND INDICATORS FOR THE DATASET CONCERNING THE HASHTAG #THESIMPSON. EMPTY ENTRIES ARE ZERO ENTRIES. IN BOLD WE HIGHLIGHT THE FIRST 4 TOP SCORING VALUES FOR EACH INDICATOR. REAL ACCOUNT NAMES HAVE BEEN OBSCURED FOR PRIVACY. OBSERVE NO USER APPEARS MORE THAN ONCE.

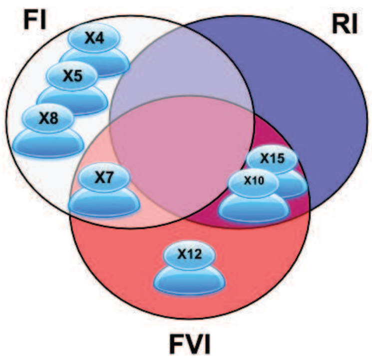

Fig. 4. Selected influential users classification for the hashtag #thesimpsons considering the first 4 top scoring users per indicator. There are no user classified as highly influencer, users X_7, X_{10} and X_{15} are classified as influential user, while X_4, X_5, X_8 and X_{12} as potentially influencer ones.

two indicators, that seem to be more correlated. In general, the three indicators never detected the same set of influential users. This is a positive result as well, meaning that the three indicators are able to capture different aspects that can influence twitter users. If the three indicators always generated the same user ranking, two out of three would be redundant and, thus, useless. Finally, the classifications derived in the three datasets showed sensible differences, meaning that the indicators do not derive just trivial results.

V. Conclusions

In this paper we presented TRank, a novel method designed to find the most influential Twitter users on a given topic, defined through hashtags. Our approach combines different Twitter signals (that represent both the user and the user's tweets) and provides three different indicators that are intended to capture different aspects of being influent. The novelty of our proposal is that the computation of the indicators is not based only on the magnitude of the Twitter signals, but they are computed taking into consideration human factors, as for example the fact that a user with many active followings might have a very *noisy* time lime and, thus, miss to read many tweets. The experimental assessment confirmed that TRank can be used to find the most influential Twitter users on a given topic, as it provides results that are more reasonable than the one obtained by mechanisms based on the sole magnitude of data.

Future work include a more estensive set of experiments and a correlation study on the three indicators.

References

[1] S. Ferretti, M. Furini, C. E. Palazzi, M. Roccetti, and P. Salomoni, "Www recycling for a better world," *Communication of the ACM*, vol. 53, no. 4, pp. 139–143, 2010.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 591–600. [Online]. Available: http://doi.acm.org/10.1145/1772690.1772751

[3] E. Bothos, D. Apostolou, and G. Mentzas, "Using social media to predict future events with agent-based markets," *Intelligent Systems, IEEE*, vol. 25, no. 6, pp. 50–58, Nov 2010.

[4] M. Furini and S. Montangero, "Siweb: understanding the interests of the society through web data analysis," *Open Journal of Web Technologies (OJWT)*, vol. 1, no. 1, pp. 1–14, 2014.

[5] H. Chen and C. C. Yang, "Special issue on social media analytics: Understanding the pulse of the society," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 41, no. 5, pp. 826–827, Sept 2011.

[6] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1 – 8, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187775031100007X

[7] M. Furini, "Users behavior in location-aware services: Digital natives vs digital immigrants," *Advances in Human-Computer Interaction*, vol. 2014, 2014.

[8] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 261–270. [Online]. Available: http://doi.acm.org/10.1145/1718487.1718520

[9] M. Furini and V. Tamanini, "Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions," *Multimedia Tools and Applications*, pp. 1–31, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2151-7

[10] Twitter, "About twitter," in *(https://about.twitter.com/company*, 2014.

[11] M. de Maertelaere, T. Li, and G. Berens, "Social influence: The effect of twitter information on corporate image," in *Proceedings of the 14th Annual International Conference on Electronic Commerce*, ser. ICEC '12. New York, NY, USA: ACM, 2012, pp. 292–293. [Online]. Available: http://doi.acm.org/10.1145/2346536.2346595

[12] G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and B. Y. Zhao, "Follow the green: Growth and dynamics in twitter follower markets," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 163–176. [Online]. Available: http://doi.acm.org/10.1145/2504730.2504731

[13] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, ser. WSDM '11. New York, NY, USA: ACM, 2011, pp. 45–54. [Online]. Available: http://doi.acm.org/10.1145/1935826.1935843

[14] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. P. Gummadi, "Understanding and combating link farming in the twitter social network," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12. New York, NY, USA: ACM, 2012, pp. 61–70. [Online]. Available: http://doi.acm.org/10.1145/2187836.2187846

[15] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy," in *Proceedings of the 22Nd International Conference on World Wide Web Companion*, ser. WWW '13 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 729–736. [Online]. Available: http://dl.acm.org/citation.cfm?id=2487788.2488033

[16] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *in ICWSM 10: Proceedings of international AAAI Conference on Weblogs and Social*, 2010.

[17] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: Crowdsourcing search for topic experts in microblogs," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '12. New York, NY, USA: ACM, 2012, pp. 575–590. [Online]. Available: http://doi.acm.org/10.1145/2348283.2348361

[18] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "Wtf: The who to follow service at twitter," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 505–514. [Online]. Available: http://dl.acm.org/citation.cfm?id=2488388.2488433

[19] K. Rosman, "On twitter, more 'favoriting': Why more people are pushing the "favorite" button," in *The Wall Steet Journal*, 2013.