IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 2, 2019, accepted May 28, 2019, date of publication June 10, 2019, date of current version June 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922155

Identifying Opinion Leaders on Social Networks Through Milestones Definition

FABIÁN RIQUELME[1,2], PABLO GONZALEZ-CANTERGIANI[3], DAVID HANS[3], RODOLFO VILLARROEL[4], AND ROBERTO MUNOZ[1,2] (Member, IEEE)

[1]Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Valparaíso 2362735, Chile
[2]Centro de Investigación y Desarrollo en Ingeniería en Salud, Universidad de Valparaíso, Valparaíso 2362735, Chile
[3]Independent Researcher
[4]Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile

Corresponding author: Fabián Riquelme (fabian.riquelme@uv.cl)

This work was supported in part by the CORFO 14ENI2-26905 "Nueva Ingeniería para el 2030" - Pontificia Universidad Católica de Valparaíso, Chile.

ABSTRACT A key problem in social network analysis is identifying influential users within a social network. To address this problem, numerous centrality measures have been defined to automatically state rankings of the users. In this article, we define the *MilestonesRank*, a new measure to detect opinion leaders, an important type of influential users focused on specific topics. This measure considers two parameters that can be freely adjusted depending on the needs of the analyst, namely, the *interest* and the *exclusivity* of the users regarding some specific topic. Every topic is bounded by a list of milestones over a period of time of several weeks or even months. We compare this measure with other classic measures to find opinion leaders in a real case study using the Twitter network. Our experiments show that the new measure allows us to find relevant opinion leaders that other measures are not able to detect.

INDEX TERMS Centrality, influential user, opinion dynamics, opinion leader, social network, Twitter.

I. INTRODUCTION

A key problem in social network analysis is identifying influential users within the network. This problem has important applications in viral marketing [1], information propagation [2], search strategies [3], expertise recommendation [4], community systems [5], social customer relationship management [6], percolation theory [7], criminology [8], among others. To address this problem, numerous centrality measures have been defined, providing different criteria to automatically rank the users of the network [9]. Some influence measures are based on simple metrics related to user activity within the network; some others are based on the PageRank algorithm [10], which is in turn inspired on the eigenvector centrality [11]; some others in the influence spread phenomenon [12], [13]; and others in predictive algorithms [9].

In online social networks developed to share and exchange content (e.g., blogs and microblogs), an important type of influential users are *opinion leaders* [14]. These users can be experts or be very interested in a subject, and are noted for their interventions on specific topics. This type of users was originally defined in 1944 in the context of presidential campaigns [15]. However, there can be opinion leaders of any subject, such as politics, sports, science, technology, arts, culture, entertainment, etc. In two-step flow theory, opinion leaders are a synonym of influential users [16]. In social networks like Twitter or Sina Microblog, *leaders* are usually considered as advanced influential users, capable to motivate other users and thus stimulate community movements, visualized by actions like mentions, replies and retweets [17], [18]. In the same line, leaders have also been considered top disseminators and engagers [19]. From a sociological point of view, some researchers have stated that opinion leaders in microblogs like Twitter are more associated with users with high self-esteem [20]. When opinion leaders are well connected and have a certain popularity (that is, a large number of followers), they are also often called *evangelists* [21].

Due to the above, we consider that the algorithms used to identify opinion leaders should focus on specific topics, and therefore be based on search spaces defined by content. Thus, besides the profile data extraction from social networks, we can take advantage of the data provided by the content

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

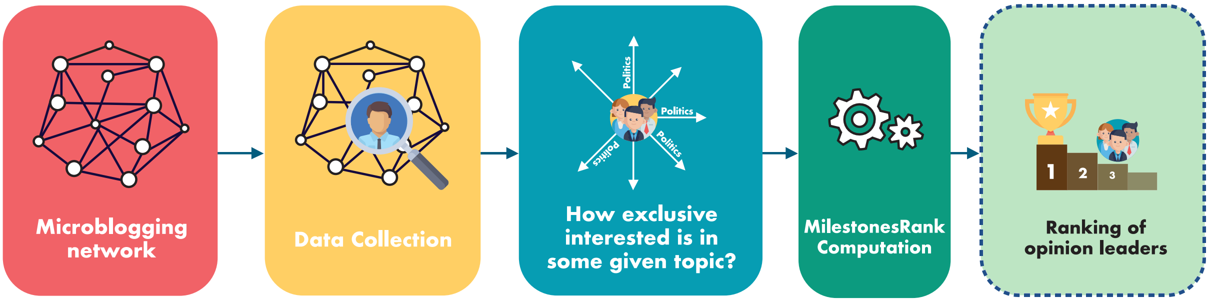

FIGURE 1. General outline of the MilestonesRank measure. From a microblogging network, data of the actors and their messages are collected. Then, in the context of a specific topic defined from a list of milestones, the opinion leaders are detected and ranked according to their exclusive interest in those milestones.

(tweets, posts, etc.) in order to obtain more information from the users. In this article we propose a new centrality measure to identify opinion leaders, called *MilestonesRank*, based on users' interest in a specific topic, as well as the users' exclusiveness to participate on that topic and not another. A novelty of this measure is that unlike others, it does not focus exclusively on structural aspects of the network, but also on the content of the shared information. On the other hand, it is not based on machine learning or predictive algorithms, so the results obtained will always be the same, for the same parameters setting.

In order to define a specific topic, the usual method is to collect messages, posts, tweets, etc., from a bag of words, keywords, or hashtags. However, to define the interest and exclusivity of each user, the MilestonesRank also considers a set of milestones that occurred during the data collection period. As a result, we obtain a measure that is able to identify opinion leaders that other measures may overlook. The scheme of Figure 1 summarizes the general operation of the new measure.

This article continues as follows. In the next section, we focus on the related work, including the definition of some well-known centrality measures used to identify influential and active users. These measures will be compared later with the MilestonesRank. Section III is dedicated to the definition and detailed explanation of the MilestonesRank. In Section IV we explain the associated methodology for data collection and application of the measure. In Section V we compare the different measures in a real case study for the Twitter microblogging network, and in Section VI we present the obtained results. Finally, we present the main conclusions of this work.

II. RELATED WORK

To study opinion leaders in social networks, there is a wide range of techniques and approaches, such as multi-agent systems [22], metaheuristics [23], discrete dynamic systems [24], collective decision-making theory [25], [26], or centrality measures [9]. Recently, inspired on the word-of-mouth marketing and using numerical simulations, a bounded confidence-based opinion dynamics model was defined to classify opinion leaders and followers in e-commerce networks [24]. Other recent approaches prefer to analyze social networks as big data challenges, instead of focusing on the network dynamics. In this context, clustering and data mining techniques have been used to efficiently find opinion leaders within the network [27].

Regarding centrality, although there is much theory about opinion leaders, as a particular type of influential users [19], in practice there are very few centrality measures dedicated to identifying only this type of users. In most cases, centrality measures tend to treat opinion leaders as synonym of influential users [21]. For a methodological review about opinion leader detection, see [28].

Among the measures used to specifically identify opinion leaders, we can mention the following. The *Influence Rank* [29] is a variation of PageRank that combines follow-up relationships, mentions, favorites and retweets, in order to identify opinion leaders who are capable of influencing other influential users. The *LeaderRank* [30] is another variation of PageRank applied on microblogs represented as follow-up relationship networks. The *CompetenceRank* [8] is a recent modification of the LeaderRank, applied to networks with star graph topology, such as the Facebook pages of political organizations. Other centrality measures have also been studied on this graph topology [31]. In addition, variations of the classic centrality measures of degree and closeness have also been defined to identify opinion leaders [32], [33]. Finally, the *LeadRank* [17] is an iterative method that starts from a current state, calculated through the number of followers, retweets and mentions of the user. These measures are useful if we can access the follow-up relationships of the network, and we do not consider the dynamics of the network through a timeline. However, the set of followers of a user can vary greatly over time. The Twitter APIs, for example, do not keep record of it's users' historical variation. Therefore, these centrality measures are not useful for our context, since we are interested on specific milestones that have occurred over a long period of time.

The measure that we propose is based on the definition of milestones on specific topics. In social networks like Twitter it is common for milestones to be viralized in the form of "trending topics". Hence, it is important to note that the trending topics do not last forever, nor disappear to never return. In an experiment conducted between 2009 and 2010 [34], it was found that 73% of the trending topics collected during almost four months were only one day long; 15% lasted two days, and 5% three days. There were a few trending topics that lasted more than three days, with the longests becoming less frequent. Besides, the authors discovered that certain trending topics could come back later in time. 31% of the trending topics came up only once, 7% remained intermittent for more than 10 days, and there were even isolated cases of topics that kept coming back for several months [34].

III. IDENTIFYING OPINION LEADERS

In this section we define the new centrality measure called MilestonesRank. We defined a *milestone* as a particular day within the sampling period, in which something especially relevant to the considered case study occurred. Let D be the set of total days considered in the sample, and $M \subseteq D$ the set of all the milestones considered.

The MilestonesRank is a linear combination of two other measures, namely, the Exclusivity and the Interest. As we will see below, the Exclusivity is based on the general activity of each user, while the Interest depends on the weights assigned to each milestone.

A. EXCLUSIVITY MEASURE

The general activity of a user can be calculated by summing up all of his or her traceable activity during a given time [9]. In this case, we define the *(total) general activity* of a user i, denoted $G_T(i)$, as the sum of his tweets, retweets and replies. Formally,

$$G_T(i) = OT1 + RT1 + RP1$$

where $OT1$ (number of original tweets posted by the author), $RT1$ (number of retweets accomplished by the author) and $RP1$ (number of replies posted by the author) are well-known classic metrics [35]. We denote $G_T(i)|_M$ the general activity of the user i restricted to the milestones in M. The *daily general activity* $G_d(i)$ of a user i is the sum of their tweets, retweets and replies collected by the sample on the day $d \in D$, i.e.,

$$G_d(i) = G_T(i)|_d = (OT1 + RT1 + RP1)|_d$$

Furthermore, we define the *average general activity* of user i during the entire period of the sample as

$$G(i) = \frac{G_T(i)}{|D|} = \frac{\sum_{d \in D} G_d(i)}{|D|}$$

We say that a user is more active than normal on a day d if $G_d(i) > G(i)$.

Given the above, the *exclusivity* of a user i for a topic defined from a set M of milestones is defined as

$$\text{Exclusivity}(i) = \frac{|\{d \in M \mid G_d(i) > G(i)\}|}{|\{d \in D \mid G_d(i) > G(i)\}|}$$

Note that this measure returns values between 0 and 1, where 1 means that the user's activity is focused exclusively during the milestones of interest, and 0 means that his or her activity focuses on the least important days of the sample. In the unlikely event that the denominator is 0 (in which case the numerator will also be 0), then the measure returns 0.

B. INTEREST MEASURE

Let $d \in M$ be a milestone, we define its weight $w_d > 0$ as the interest that this milestone has for our sample. The weight of a milestone can be determined automatically (e.g., as a number proportional to the total activity generated by all users on that day), manually (e.g., given by an expert user) or semi-automatically (e.g., as a hybrid of the previous two).

To understand the usefulness of weights, let's consider some examples. For example, if we want to analyze the interest of the users in an electoral campaign, the televised debates of the candidates are undoubtedly important milestones to consider, but they must have a lower weight than the voting day. Also, in a sports championship, the day of the semifinals will have a lower weight than the day of the final.

Thus, the *Interest* of a user i for a topic defined from a set M of milestones is defined as

$$\text{Interest}(i) = \frac{\sum_{d \in M} \{w_d \mid G_d(i) > G(i)\}}{\sum_{d \in M} w_d}$$

Note that this measure also returns values between 0 and 1. Here, 1 means that the user has a greater activity during all the milestones, so it can be assumed that he or she has a special interest in them; on the contrary, 0 means that the user is not interested at all by the milestones considered.

C. THE MILESTONESRANK

With all the above, we define the *MilestonesRank* (MR) as a centrality measure to identify opinion leaders as follows:

$$\text{MR}(i) = \text{Interest}(i) \cdot c_I + \text{Exclusivity} \cdot c_E$$

where $c_I + c_E = 1$.

We denote $\text{MR}_{I=a}^{E=b}$ as the application of MR with the parameters $c_I = a$ and $c_E = b$. Thus, $\text{MR}_{I=0.5}^{E=0.5}$ means that the measure is giving the same importance to both criteria. Note also that for any user i, $\text{MR}_{I=0}^{E=1} = \text{Exclusivity}(i)$ and $\text{MR}_{I=1}^{E=0} = \text{Interest}(i)$.

This measure varies between 0 and 1. A user has a high MilestonesRank if he or she is more interested (or active) during the milestones considered, and if he or she tends to be active exclusively during those milestones.

IV. METHODOLOGY

Next, we explain the methodology to apply the MilestonesRank measure to any online social network based on the transmission and viralization of content. Since some steps can be performed in parallel or sequentially (depending on whether the activity is carried out manually or automatically, respectively), the flow of steps can vary in each case

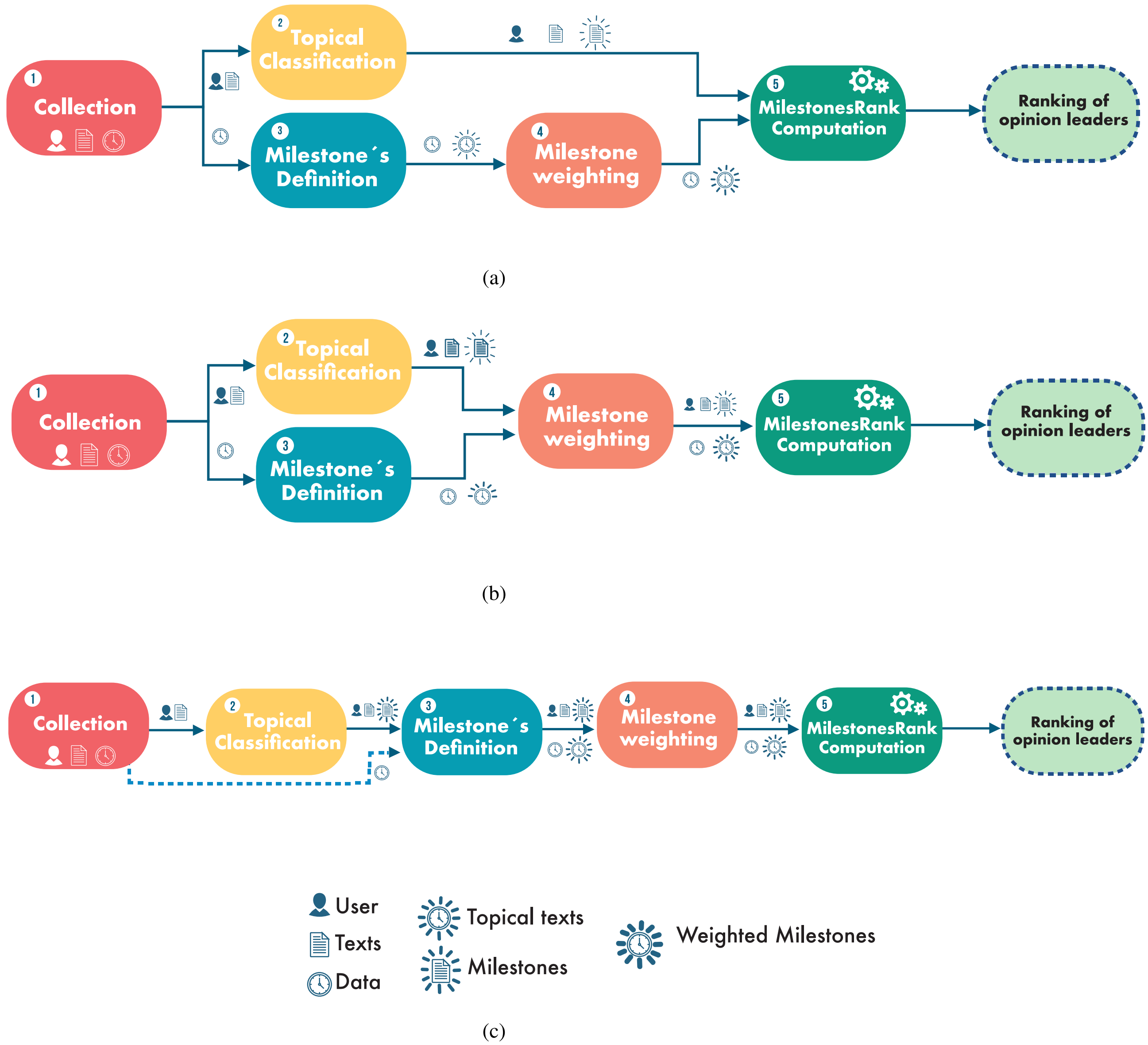

FIGURE 2. Methodology used to compute the MilestonesRank. The legend below the schemas represents the types of data that circulate through the different sequential processes. Note that "milestones" can only appear after "texts", as well as "weighted milestones" after "topical texts". (a) Manual milestones' definition and manual weighting. (b) Manual milestones' definition and authomatic weighting. (c) Automatic milestones' definition and manual/automatic weighting.

(see Figure 2). The general process distinguishes three sequential activities:

1) **Data collection.** Usually, messages transmitted through online social networks have a large amount of associated metadata. However, for each message (tweet, post, etc.) issued, it is only necessary to collect the user that issues the message, its text body, and its date of issue.
2) **Data post-processing.** In this stage, three different sub-processes are carried out:
   - *Topical classification.* Classification of messages according to the topic of study. Specifically, a Boolean value is associated to each message, assigning a value of "true" if the message belongs to the topic, or "false" if it does not belong to the topic. The usual way to define a topic is through keywords, bag of words or hashtags.
   - *Milestones' definition.* A set of milestones is defined (set M of the sample). This can be done manually (through an expert user), automatically (using computational algorithms) or semi-automatically (a hybrid of the previous two). Figures 2a and 2b show the flow for the manual case, and Figure 2c for the authomatic case.
   - *Milestones weighting.* The weights w_d are defined for each milestone $d \in M$ stated in the previous

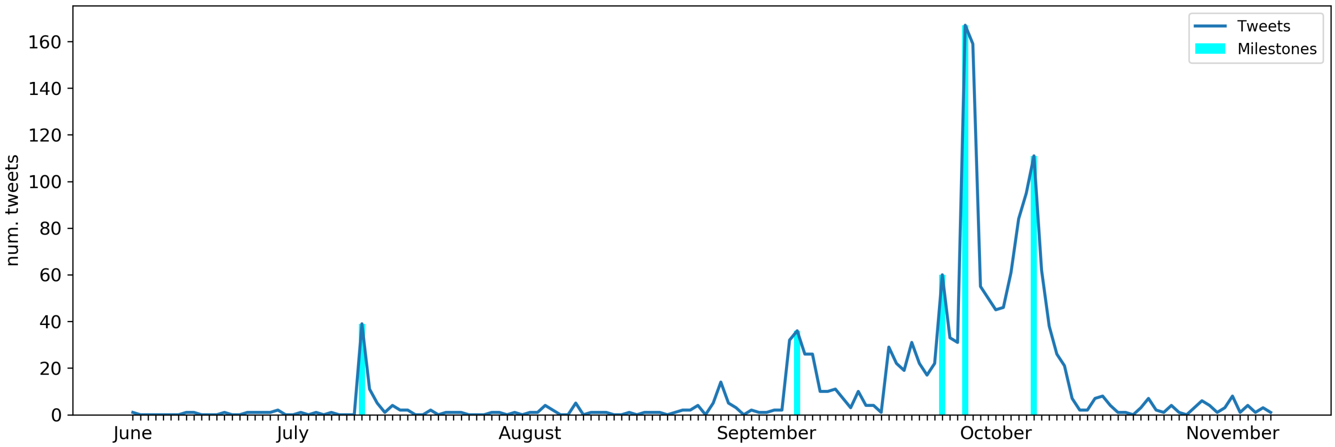

FIGURE 3. Peaks detected automatically. The horizontal axis includes the time and the vertical axis the number of tweets. The weights of the milestones are based on the fourth peak.

step. Again, this can be done manually, automatically, or semi-automatically. Figure 2a shows the flow for the manual case, and Figures 2b and 2c for the authomatic case.

3) **MilestonesRank computation.** Calculation of the MilestonesRank measure. Note that for users who do not participate in any milestone, the value of the measure is 0, so it is not necessary to calculate the measure for them.

V. EXPERIMENT

As a case study, we set out to find opinion leaders within the Twitter network, regarding the controversial nomination of Brett Kavanaugh to the U.S. Supreme Court on 2018, under the presidency of Donald Trump. The data was collected via the Twitter Streaming API. The total sample contains 12,910,703 tweets geolocated in the United States, written between June 10, 2018 and November 6, 2018 (as we will see, one month before the first milestone and one month after the last milestone).

For the topical classification stage, we use "`kavanaugh`" as the only keyword. This search is not case-sensitive, and includes the hashtag "`#kavanaugh`". From the above, a total of 1,720 tweets were obtained, coming from 828 users. These values are much lower than in the original sample, which shows the enormous diversity of topics circulating daily through social networks. For the definition of both the milestones and the milestone weights, automatic methods were used.

The milestones were chosen from peaks detected by a Python library called PeakUtils[1]. The parameters were set to `thres` $= 0.2$ and `min_dist` $= 0$, where `thres` indicates the minimum magnitude of a peak to be detected and `min_dist` the minimum distance required between peaks. In addition to this library, several more were tested, but this was the one that gave the best results regarding validation with the highlights in the online press and Wikipedia. The resulting peaks detection is illustrated in Figure 3. The dates of each peak, with their numbers of tweets and the corresponding milestones, are described in the following list:

1) July 9, 2018 (39 Tweets): U.S. President Donald Trump nominated Judge Brett Kavanaugh to become an Associate Justice of the Supreme Court of the United States.
2) September 4, 2018 (36 Tweets): The Senate Judiciary Committee began Kavanaugh's confirmation hearing.
3) September 24, 2018 (60 Tweets): The Senate Judiciary Committee invited both Kavanaugh and Christine Blasey Ford to provide testimony about the allegation. Kavanaugh agreed to testify.
4) September 27, 2018 (167 Tweets): Both Kavanaugh and Blasey Ford were questioned by members of the Judiciary Committee and Arizona-based sex crimes prosecutor Rachel Mitchell.
5) October 6, 2018 (111 Tweets): The Senate voted 50–48 to confirm Kavanaugh's nomination to the Supreme Court.

From the 828 users tweeting about the topic, a total of 254 users participated in at least one milestone.

Regarding the milestones weighting, a frequency analysis was used, where each milestone is normalized respect to the milestone with the most data. Thus, the obtained weights are $w_1 = 0.233$, $w_2 = 0.215$, $w_3 = 0.359$, $w_4 = 1$ and $w_5 = 0.664$.

VI. RESULTS AND DISCUSSION

For the case study described in Section V, we apply the MilestonesRank measure for the following combinations of parameters: $\mathrm{MR}^{E=0.5}_{I=0.5}$, $\mathrm{MR}^{E=0.25}_{I=0.75}$, $\mathrm{MR}^{E=0.75}_{I=0.25}$, $\mathrm{MR}^{E=0}_{I=1}$, $\mathrm{MR}^{E=1}_{I=0}$. In addition, for the same sample we apply the total general activity measure, G_T, and its restricted version to the days of the milestones, $G_T|_M$.

[1]https://peakutils.readthedocs.io/en/latest/

TABLE 1. Kendall correlation results.

	$MR^{E=0.5}_{I=0.5}$	$MR^{E=0.25}_{I=0.75}$	$MR^{E=0.75}_{I=0.25}$	$MR^{E=0}_{I=1}$	$MR^{E=1}_{I=0}$	$G_T\|_M$	G_T
$MR^{E=0.5}_{I=0.5}$	1	0.901	0.895	0.681	0.654	0.046	-0.217
$MR^{E=0.25}_{I=0.75}$	-	1	0.796	0.791	0.555	0.083	-0.188
$MR^{E=0.75}_{I=0.25}$	-	-	1	0.564	0.760	0.011	-0.244
$MR^{E=0}_{I=1}$	-	-	-	1	0.295	0.120	-0.160
$MR^{E=1}_{I=0}$	-	-	-	-	1	-0.720	-0.267
$G_T\|_M$	-	-	-	-	-	1	0.591
G_T	-	-	-	-	-	-	1

TABLE 2. Standard deviation and number of different values for each measure.

Metric	σ	#diff
$MR^{E=0.5}_{I=0.5}$	0.261	144
$MR^{E=0.25}_{I=0.75}$	0.321	146
$MR^{E=0.75}_{I=0.25}$	0.178	144
$MR^{E=0}_{I=1}$	0.292	11
$MR^{E=1}_{I=0}$	0.120	75
$G_T\|_M$	0.058	11
G_T	0.012	19

In order to compare the rankings obtained by the measures, we used the Kendall rank correlation coefficient (τ). The correlation results are illustrated in Table 1. As can be seen, the three centrality measures have a very low correlation with each other. This means that a user's activity is not enough to be considered an opinion leader, which validates the theory [28]. In addition, it is shown that several of the opinion leaders identified with the MilestonesRank could not have been identified only from their activity. Regarding the variations of parameters for the MilestonesRank, it is observed that the Interest was a stronger criterion than the Exclusivity, since the correlation is higher when increasing the Interest rather than the Exclusivity.

Regarding the usefulness of the centrality measures, there are two important factors that have been considered in other studies to determine if the users are ranked properly, namely, the standard deviation (σ) and the number of different results obtained by the measure [13]. On one hand, a large number of different results implies that the measure generates distinguishable classes of users with different values. On the other hand, a high standard deviation implies that the differences in their values are large enough so that they do not lend themselves to confusion. Table 2 shows the standard deviation and the different values for the different measures considered. Regardless of the chosen parameters, the MilestonesRank returns a standard deviation much higher than the other activity measures. Likewise, the MilestonesRank also returns a greater number of different results, with the exception of $MR^{E=0}_{I=1}$. This means that the Interest, on its own, returns very different values to each other, but many users had an equivalent interest for the topic analyzed. The best results in both aspects were obtained for $MR^{E=0.25}_{I=0.75}$, with a standard deviation of $\sigma = 0.321$ and 146 different results.

Given the above, we will focus on the top results obtained for $MR^{E=0.25}_{I=0.75}$. In order to analyze in more detail the types of opinion leaders found, we manually viewed the descriptions of the user profiles. In most cases, descriptions were complete describing the user's political profile. For those cases in which the description of the user profile was not enough, the tweets issued by the users during the milestones were observed.

Of the top 20 results, there were 9 men, 7 women, 1 queer activist, 1 with unidentified sex, and 2 accounts without sex. The first place was for an independent journalist, contrary to the policies of President Donald Trump. 14 of the 20 accounts were anti-Trump, 4 pro-Trump, and only 2 were moderate, and therefore alien to the topic (a 911 operator and a woman interested in fashion and photography).

Two verified accounts were detected: Adrienne Lawrence, a well-known interviewer and progressive opinion leader, and the official account of the website http://endrapeoncampus.org. In addition, a private account appeared, from a retired military officer. 12 accounts of the top 20 correspond to independent journalists, political activists or people interested in politics.

If we focus on the top 50, we find 26 women and 20 men. In total there are 36 anti-Trump accounts, only 5 pro-Trump (one of them deactivated), 8 moderate and 1 private account from which no more information could be found. Within the anti-Trump accounts, a single account of an explicitly Republican person was observed (a religious promoter, retired Lutheran pastor). In general, it is quite clear when the person is conservative or liberal. Conservative people often mention their religious beliefs and family in their user profile, while liberal people tend to focus on their jobs and personal interests. Of the latter, there are many liberal people related to the world of technology, design and photography, although writers, actors and workers in the world of fashion and marketing were also detected.

Finally, it is interesting to mention that most of the accounts that were found are independent. Official press media such as CNN or ABC News do not appear in the ranking. This is mainly due to two factors. First, because these accounts are not geolocated, and therefore did not appear in the original sample obtained from the data collection. Second, these accounts cover news continuously, without highlighting one milestone over another. Therefore, even if they had been in the sample, they would have a high interest in the milestones considered, but a very low exclusivity on the topic.

VII. CONCLUSIONS

In this article we have defined a new centrality measure to identify opinion leaders in an online social network. It is important to note that although the case study is based on the Twitter network, the measure can be applied to any online social network based on content spread, such as blogs, microblogs, etc. Unlike the few existing measures to identify this particular type of influential users, the MilestonesRank does not need follow-up relationships. To apply the measure, a set of weighted milestones must be defined, which can be stated automatically or manually. In particular, we have seen that the automatic generation of milestones produces good results that can be easily validated.

As a case study we consider a real and recent US policy event. The ranking of opinion leaders obtained by the measure yielded valuable and interesting accounts, such as several independent journalists, political activists and people interested in politics, both liberal and conservative. Many of these users could not be detected by the general activity measures. In fact, we show that there is no correlation between the MilestonesRank and the general activity measures, even if it is restricted exclusively to the activity carried out during the milestones.

Interestingly, the highest correlation (although negative) was obtained between $\mathrm{MR}_{I=0}^{E=1}$ and $G_T|_M$, with a value of -0.720. This could mean that for this sample, users who wrote exclusively during the milestones had a low general activity and, on the contrary, there were other topics that generated much more activity. However, by definition, a user with high exclusivity does not necessarily have low activity the rest of the time. Likewise, it is possible that during a considered period there is no other topic with more activity, but that when comparing the topic studied with the general activity, this topic seems less relevant. Given the above, it would be interesting as future work to define a methodology to analyze different topics, in order to compare their specific relevance within the same period.

ACKNOWLEDGMENT

Rodolfo Villarroel was supported by CORFO 14ENI2-26905 ''Nueva Ingeniería para el 2030'' - Pontificia Universidad Católica de Valparaíso, Chile.

REFERENCES

[1] P. Domingos and M. Richardson, ''Mining the network value of customers,'' in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. D. Lee, M. Schkolnick, F. J. Provost, and R. Srikant, Eds. San Francisco, CA, USA: ACM, Aug. 2001, pp. 57–66. [Online]. Available: http://portal.acm.org/citation.cfm?id=502512.502525

[2] D. Gruhl, D. Liben-Nowell, R. Guha, and A. Tomkins, ''Information diffusion through blogspace,'' *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 2, pp. 43–52, 2004. doi: 10.1145/1046456.1046462.

[3] L. Adamic and E. Adar, ''How to search a social network,'' *Social Netw.*, vol. 27, no. 3, pp. 187–203, 2005. doi: 10.1016/j.socnet.2005.01.007.

[4] X. Song, B. L. Tseng, C.-Y. Lin, and M.-T. Sun, ''Personalized recommendation driven by information flow,'' in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, E. N. Efthimiadis, S. T. Dumais, D. Hawking, and K. Järvelin, Eds. Seattle, WA, USA: ACM, Aug. 2006, pp. 509–516. doi: 10.1145/1148170.1148258.

[5] X. Zhang, J. Zhu, Q. Wang, and H. Zhao, ''Identifying influential nodes in complex networks with community structure,'' *Knowl.-Based Syst.*, vol. 42, pp. 74–84, Apr. 2013. doi: 10.1016/j.knosys.2013.01.017.

[6] J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu, ''Social network user influence sense-making and dynamics prediction,'' *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5115–5124, 2014. doi: 10.1016/j.eswa.2014.02.038.

[7] F. Morone and H. A. Makse, ''Influence maximization in complex networks through optimal percolation,'' *Nature*, vol. 524, pp. 64–68, Jul. 2015. doi: 10.1038/nature14604.

[8] M. Spranger, F. Heinke, H. Siewerts, J. Hampl, and D. Labudde, ''Opinion leaders in star-like social networks: A simple case?'' in *Proc. 8th Int. Conf. Adv. Inf. Mining Manage. (IMMM)*, D. Labudde, Ed. Barcelona, Spain: IARIA, Jul. 2018, pp. 33–38.

[9] F. Riquelme and P. González-Cantergiani, ''Measuring user influence on Twitter: A survey,'' *Inf. Process. Manage.*, vol. 52, no. 5, pp. 949–975, 2016. doi: 10.1016/j.ipm.2016.04.003.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd, ''The pagerank citation ranking: Bringing order to the Web,'' Stanford InfoLab, Tech. Rep. 1999-66, Nov. 1999. [Online]. Available: http://ilpubs.stanford.edu:8090/422/

[11] P. Bonacich, ''Factoring and weighting approaches to status scores and clique identification,'' *J. Math. Sociol.*, vol. 2, no. 1, pp. 113–120, 1972. doi: 10.1080/0022250X.1972.9989806.

[12] D. Kempe, J. Kleinberg, and É. Tardos, ''Influential nodes in a diffusion model for social networks,'' in *Proc. 32nd Int. Colloq. Automata, Lang., Program. (ICALP)*, in (Lecture Notes in Computer Science), vol. 3580, L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds. Lisbon, Portugal: Springer, Jul. 2005, pp. 1127–1138. doi: 10.1007/11523468_91.

[13] F. Riquelme, P. Gonzalez-Cantergiani, X. Molinero, and M. Serna, ''Centrality measure in social networks based on linear threshold model,'' *Knowl.-Based Syst.*, vol. 140, pp. 92–102, Jan. 2018. doi: 10.1016/j.knosys.2017.10.029.

[14] F. Riquelme, ''Twitter applications and types of relevant users,'' in *Twitter: Global Perspectives, Uses and Research Techniques*, I. Chiluwa and G. Bouvier, Eds. New York, NY, USA: Nova, 2019, ch. 2, pp. 9–28.

[15] P. Lazarsfeld, B. Berelson, and H. Gaudet, *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York, NY, USA: Duell, Sloan and Pierce, 1944.

[16] W. Chai, W. Xu, M. Zuo, and X. Wen, ''ACQR: A novel framework to identify and predict influential users in micro-blogging,'' in *Proc. 17th Pacific Asia Conf. Inf. Syst. (PACIS)*, J. Lee, J. Mao, and J. Y. L. Thong, Eds. Jeju-do, South Korea: Association for Information Systems, Jun. 2013, p. 20. [Online]. Available: http://aisel.aisnet.org/pacis2013/20

[17] L. B. Jabeur, L. Tamine, and M. Boughanem, ''Active microbloggers: Identifying influencers, leaders and discussers in microblogging networks,'' in *Proc. 19th Int. Symp. String Process. Inf. Retr. (SPIRE)*, in (Lecture Notes in Computer Science), vol. 7608, L. Calderón-Benavides, C. N. González-Caro, E. Chávez, and N. Ziviani, Eds. Springer: Cartagena de Indias, Colombia, Oct. 2012, pp. 111–117. doi: 10.1007/978-3-642-34109-0_12.

[18] X. Liu and C. Liu, ''Information diffusion and opinion leader mathematical modeling based on microblog,'' *IEEE Access*, vol. 6, pp. 34736–34745, 2018. doi: 10.1109/ACCESS.2018.2849722.

[19] M. del Fresno García, A. J. Daly, and S. S. Sánchez-Cabezudo, ''Identifying the new influencers in the Internet era: Social media and social network analysis,'' *Revista Española de Investigaciones Sociológicas*, vol. 153, pp. 23–40, 2016. doi: 10.5477/cis/reis.153.23.

[20] Y. Hwang, ''Does opinion leadership increase the followers on Twitter,'' *Int. J. Social Sci. Hum.*, vol. 5, no. 3, pp. 258–264, 2015. doi: 10.7763/IJSSH.2015.V5.464.

[21] M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi, ''The world of connections and information flow in Twitter,'' *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 42, no. 4, pp. 991–998, Jul. 2012. doi: 10.1109/TSMCA.2012.2183359.

[22] X. Molinero, F. Riquelme, and M. Serna, ''Cooperation through social influence,'' *Eur. J. Oper. Res.*, vol. 242, no. 3, pp. 960–974, 2015. doi: 10.1016/j.ejor.2014.11.006.

[23] L. Jain and R. Katarya, ''Discover opinion leader in online social network using firefly algorithm,'' *Expert Syst. Appl.*, vol. 122, pp. 1–15, May 2019. doi: 10.1016/j.eswa.2018.12.043.

[24] Y. Zhao, G. Kou, Y. Peng, and Y. Chen, ''Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective,'' *Inf. Sci.*, vol. 426, pp. 131–147, Feb. 2018. doi: 10.1016/j.ins.2017.10.031.

[25] X. Molinero, F. Riquelme, and M. Serna, ''Satisfaction and power in unanimous majority influence decision models,'' *Electron. Notes Discrete Math.*, vol. 68, pp. 197–202, Jul. 2018. doi: 10.1016/j.endm.2018.06.034.

[26] X. Molinero, F. Riquelme, and M. Serna, ''Measuring satisfaction and power in influence based decision systems,'' *Knowl.-Based Syst.*, vol. 174, pp. 144–159, Jun. 2019. doi: 10.1016/j.knosys.2019.03.005.

[27] Y.-C. Chen, Y.-H. Chen, C.-H. Hsu, H.-J. You, J. Liu, and X. Huang, "Mining opinion leaders in big social network," in *Proc. 31st IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, L. Barolli, M. Takizawa, T. Enokido, H. Hsu, and C. Lin, Eds. Taipei, Taiwan, Mar. 2017, pp. 1012–1018. doi: 10.1109/AINA.2017.147.

[28] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, "Opinion leader detection: A methodological review," *Expert Syst. Appl.*, vol. 115, pp. 200–222, 2019. doi: 10.1016/j.eswa.2018.07.069.

[29] B. Hajian and T. White, "Modelling influence in a social network: Metrics and evaluation," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust PASSAT/SocialCom (PASSAT)*, Boston, MA, USA, Oct. 2011, pp. 497–500. doi: 10.1109/PASSAT/SocialCom.2011.118.

[30] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PLoS ONE*, vol. 6, no. 6, 2011, Art. no. e21202. doi: 10.1371/journal.pone.0021202.

[31] X. Molinero, F. Riquelme, and M. J. Serna, "Star-shaped mediation in influence games," in *Proc. 12th Cologne-Twente Workshop Graphs Combinat. Optim. (CTIT)*, K. Cornelissen, R. Hoeksma, J. Hurink, and B. Manthey, Eds. Enschede, The Netherlands: Centre for Telematics and Information Technology, May 2013, pp. 179–182.

[32] L. Yang, Y. Tian, J. Li, J. Ma, and J. Zhang, "Identifying opinion leaders in social networks with topic limitation," *Cluster Comput.*, vol. 20, no. 3, pp. 2403–2413, 2017. doi: 10.1007/s10586-017-0732-8.

[33] L. Yang, Y. Qiao, Z. Liu, J. Ma, and X. Li, "Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm," *Soft Comput.*, vol. 22, no. 2, pp. 453–464, 2018. doi: 10.1007/s00500-016-2335-3.

[34] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web (WWW)*, M. Rappa, P. Jones, J. Freire, and S. Chakrabarti, Eds. New York, NY, USA: ACM, Apr. 2010, pp. 591–600. doi: 10.1145/1772690.1772751.

[35] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proc. 4th Int. Conf. Web Search Web Data Mining (WSDM)*, I. King, W. Nejdl, and H. Li, Eds. Hong Kong, Feb. 2011, pp. 45–54. doi: 10.1145/1935826.1935843.

FABIÁN RIQUELME received the bachelor's degree in computing civil engineering and the master's degree in computer science from Universidad de Concepción, Chile, and the Ph.D. degree in computing from Universitat Politècnica de Catalunya, Barcelona, Spain. In his Ph.D. thesis, he proposed the influence graphs as a model to study influence spread in social systems. He is an Associate Professor with the School of Informatics Engineering and Responsible Researcher with the Centro de Investigación y Desarrollo en Ingeniería en Salud (CINGS) of Universidad de Valparaíso, Chile. His main research interests include social network analysis, social computing, cooperative game theory, and computational complexity.

PABLO GONZALEZ-CANTERGIANI received the B.S. and M.S. degrees in computer science from the University of Santiago, Chile. He has worked in research centers. He is currently the Head of Data Science in InstaGIS and Predictable Media Company. His main research interests include social network analysis, open data, and distributed systems.

DAVID HANS received the B.S. degree in computer science from the University of Santiago, Chile. He worked as a Research Assistant with the Center for Innovation in Information Technologies for Social Applications (Citiaps) from University of Santiago, Chile. He is currently working as a Software Development Engineer in the company Requies. His main current interest include agile web development.

RODOLFO VILLARROEL received his Ph.D. degree from the University of Castilla-La Mancha (2005), Ciudad Real, Spain, and his Master's Degree in Computer Science Engineering at the Universidad Técnica Federico Santa María (2000), Valparaíso, Chile. He is an Associate Professor and the Researcher with the School of Computer Engineering of the Pontificia Universidad Católica de Valparaíso, Chile. He is currently a Board Member of the Chilean Computer Science Society (SCCC). His research interests include software engineering, learning analytics, quality and information security, and computers and education.

ROBERTO MUNOZ received the master's degree in computer engineering, engineering science, and education and the Ph.D. degree in computer engineering. He is an Associate Professor with the School of Informatics Engineering and Adjunct Researcher with the Center of Cognition and Language (CIDCL) and the Centro de Investigación y Desarrollo en Ingeniería en Salud (CINGS) of Universidad de Valparaíso. He is the author of over 70 scientific papers in refereed international conferences and journals. Also, he holds a position as the Editor of the *International Journal on Computational Thinking (IJCThink)*. His research interests include computers and education, multimodal learning analytics, human–computer interaction, and health informatics.