

Article

# Authority-Based Conversation Tracking in Twitter: An Unattended Methodological Approach

Marçal Mora-Cantalops <sup>\*,†</sup> , Salvador Sánchez-Alonso , Elena García-Barriocanal and Miguel-Ángel Sicilia

Computer Science Department, Universidad de Alcalá, 28801 Alcalá de Henares, Madrid, Spain; salvador.sanchez@uah.es (S.S.-A.); elena.garciab@uah.es (E.G.-B.); msicilia@uah.es (M.-Á.S.)

\* Correspondence: marcal.mora@uah.es

† Current address: Plaza de San Diego, s/n, 28801 Alcalá de Henares, Madrid, Spain.

Received: 22 April 2020; Accepted: 2 May 2020; Published: 8 May 2020



**Featured Application:** We propose a method capable of dynamically identifying and monitoring the most relevant discussion topics of a set of authorities, without human intervention. It is robust and interruption proof, not needing any set of seed words to start the analysis.

**Abstract:** Twitter is undoubtedly one of the most widely used data sources to analyze human communication. The literature is full of examples where Twitter is accessed, and data are downloaded as the previous step to a more in-depth analysis in a wide variety of knowledge areas. Unfortunately, the extraction of relevant information from the opinions that users freely express in Twitter is complicated, both because of the volume generated—more than 6000 tweets per second—and the difficulties related to filtering out only what is pertinent to our research. Inspired by the fact that a large part of users use Twitter to communicate or receive political information, we created a method that allows for the monitoring of a set of users (which we will call authorities) and the tracking of the information published by them about an event. Our approach consists of dynamically and automatically monitoring the hottest topics among all the conversations where the authorities are involved, and retrieving the tweets in connection with those topics, filtering other conversations out. Although our case study involves the method being applied to the political discussions held during the Spanish general, local, and European elections of April/May 2019, the method is equally applicable to many other contexts, such as sporting events, marketing campaigns, or health crises.

**Keywords:** Twitter; topic tracking; data extraction; information retrieval; data mining

## 1. Introduction

The widespread use of social networking services (SNS) and the fact that users increasingly trust the news shared by their contacts or friends over those in traditional media has significantly changed our information and communication habits. In particular, when contacts considered “opinion leaders” [1] publish a story within their sphere of influence, their contacts/followers tend to give it more credibility than they would give the mainstream media.

User activity in SNS somehow reflects actual news and everyday-life trends. Analyzing trends is highly relevant for understanding public opinion [2], as it allows researchers to examine public conversations and debates in detail, evaluate the behavior of a product after a marketing campaign, or analyze reactions to a particular event, among others. Since trending topics to some extent describe the opinion of a community and provide the means to analyze it, knowing where public attention is at a certain point in time becomes a topic of interest for researchers and professionals.

Not all SNS serve the same purpose in any case. Twitter, which according to Gadek [3] is the reference in microblogging platforms, allows users to post short messages called tweets, and interact with the tweets posted by others replying to them, quoting or retweeting them. User messages often include hashtags as a way of explicitly marking the relevant topics, easing such monitoring and analysis. Such a simple mechanism makes Twitter especially convenient for information retrieval and automatic processing purposes, as vast amounts of data are generated every day; data which provides valuable information for many different domains, such as political communication, consumer behavior, or disaster management, just to name a few. These data are freely and publicly available through Twitter's application programming interface (API), allowing for real-time monitoring of users' preferences, opinion, and behavior.

Twitter is a window open to the spontaneous communication model that takes place in the real world, a model rooted in the fact that user behavior and occurrences are unpredictable and dynamic. Therefore, as messages in Twitter reflect real-time news and everyday-life trends, and given that many events cannot be accurately foreseen—e.g., natural disasters or accidents—it is really hard for analysts to anticipate the wording that will be used by users in their hashtags. Of course, if manual monitoring was carried out, new hashtags could be added as they appear as qualifiers of hot topics of interest, and yet this approach would have important shortcomings:

- It would introduce a significant delay between the moment when a new topic emerges and the starting point of its tracking. Such a delay would translate into losing relevant information or interactions;
- Human supervision, which is not always possible (e.g., late at night), would be necessary. Unfortunately, depending on humans can cause the introduction of failures derived from fatigue, incorrect interpretation of the information, inability to detect and track relevant changes, and others.

Moreover, it is also worth noting that, through hashtags, users pretend to label their content in a word or expression that summarizes what they are talking about; this is, however, a linguistic construct, and therefore, it carries its limitations. For an averagely informed person, the meaning of a hashtag can be relatively easy to understand or infer, but for an automatic process, this can become a complex task, as even the most famous hashtags such as #MeToo or #StayAtHome are not self-explanatory, and thus, need to be considered in their own context. Hashtag wording is part of the game in Twitter and often corresponds to specific (and deliberate) communication strategies.

The mechanism we propose consists of two separate parts, each governed by an algorithm. The first algorithm identifies the hottest topics of conversation of a group of authorities. It does this by continuously monitoring any tweet from the authorities, producing as an output an always-updated list of hashtags that includes the most interesting topics of conversation. Simultaneously, a second algorithm uses the current hashtag list as input, extracting and storing all the tweets tagged with any of the hashtags in the list. The whole mechanism must be initialized by setting up the set of authorities as well as some other parameters such as the time window frame (more details in Section 3). Once running, the software—which aims to minimize the loss of information—works in unattended mode without human intervention.

It is well known that political information is one of the most shared types of information on social media. In fact, literature on the use of Twitter for political activities abound, such as those studies on the effect of social media, especially Twitter, as a facilitator in political campaigns [4] and protests [5] worldwide. The fact is that two-thirds of social media users show some kind of political engagement by, for instance, following candidates, posting thoughts about political issues, or pressing friends to vote [6]. This behavior is especially evident in Twitter opinion leaders, who consistently show a higher involvement in political processes [7]. Besides, politically engaged young people integrate social media use into their existing organizations and political communications [8]. This prominence of the use of Twitter for politics, the interest that the study of political information awakens, and the rising concerns

about the effect of false stories (or “fake news”) on social media [9] are what inspired us to apply our model to the analysis of political discussions during the Spanish general elections as a use case, which will be further detailed in Section 4.

The remainder of this paper is organized as follows. The next section outlines related works, while the proposed approach is detailed in the methodology section. Section 4 presents a case study where our model is applied to the analysis of the political debate in Twitter: the Spanish general elections of April 2019. Finally, conclusions are discussed in Section 5.

## 2. Background

### 2.1. Social Media, Topics, and Trends

Social media (or SNS) are web-based services that allow individuals to create a profile, articulate a list of connections to other users, view and traverse such connections, and share content [10]. Although social media are today the primary source of information for many, the spread of information is still irregular, difficult to predict in nature, and incidental. Methods that aim to track and retrieve given conversations are required to be flexible and dynamic enough to be able to follow them with the minimum information loss.

To be able to accurately track opinion over time has been one of the main concerns of analysts for a long time [11–14]. With the advent of Twitter, public opinion can be tracked continuously and in real time. In fact, Twitter continuously publishes real-time produced lists of the most popular topics under discussion either locally or globally (also known as trending topics). Many users consider something to be news when it becomes a Twitter trending topic, even though the way Twitter produces these lists of trending topics is unknown, as the algorithm used remains unpublished. It is relevant, however, that although Twitter is assumed to classify trends according to every tweet published during a period, it only provides (via its API) a reduced sample of the entire stream for free. This limitation makes it difficult for those trying to extract trending topics from the whole data stream. However, our work does not aspire to replicate the way Twitter produces its trending topic lists, and assumes working on a subset of the complete data to be reasonable.

Thus, Twitter produces trending topic lists from all users’ tweets. However, the way information spreads in online social networks is more reminiscent of a complex contagion model where information diffusion is affected not only by the number of exposures to a piece of information but also by the exposure to multiple sources and their social influence [15]. Hence, users tend to follow other users on topics of their interest in order to acquire information on those topics. These later users, leaders of opinion or “authorities”, are sources of information for many individuals in this media environment, and consequently, become more influential, and the information diffused by them more “viral” [1].

As mentioned, our work aims not at detecting global trending topics but instead at obtaining the relevant topics of discussion among a set of opinion leaders that we call “authorities”, and adapting the tracked keywords accordingly to retrieve the most relevant information throughout the duration of the monitoring. As a result, we will be able to: (a) observe how the topics of interest addressed by a set of authorities evolve; and (b) to extract the conversations linked to those topics, discarding any others, which will tell us how authorities’ public discourse has changed over time.

### 2.2. Topic Tracking

Previous work has pointed out the difficulties of implementing reliable, precise, and fast trend detection [16]. The high volume of information, the myriad different topics under discussion in any particular time, and the significant variations in the time and volume scales of social datasets stand in the way of direct tracking. Many authors have addressed the problem of detecting new events from a stream [17,18]. The work by Petrovic et al. [19] described a method applicable to a stream of Twitter posts, but although their system outperformed previous approaches, the results invariably include spurious information and not only news events as one would expect. One way

to overcome the limitations related to the analysis and classification of corpora is to look at the Twitter hashtags (keywords or terms starting with “#”), which are the most common feature for users to connect and relate to within a larger networked discourse [20]. D’heer et al. have shown how messages that include such hashtags have, in general, more informational value than tweets without them; these labeled messages are also often longer, connected with other topics, and enriched with hyperlinks [21]. Additionally, as Enli and Simonsen [22] note, using hashtags is not an individual action: many actors, such as politicians, make use of them to reach outside their own connections and contacts. This is an example of how the use of social media by many sectors is closely-related to their professional practice.

To be able to track hashtags accurately is, thus, critical to be able to analyze the discourse in Twitter. However, difficulties arise when the researcher needs to predict what hashtag or word is going to be used by the public in a particular situation, and whether it is going to change in the process, alternating with modified or customized hashtags, so a flexible approach to the tracking becomes a requirement. In spite of this, many previous works follow a static tracking approach. For instance, Fano and Slanzi [23] used Twitter to monitor the discussion around a constitutional referendum held in Italy for five weeks. Even with a potentially disputed debate, they opted for a manual selection of five hashtags. Similarly, Reyes-Menendez et al. [24] tracked a single hashtag, #WorldEnvironmentDay, to assess the public opinion around such an event. However, such opinion included only users that used hashtags in English, and therefore, discarded variants or hashtags written in different languages that could, potentially, be even more widely used, individually or altogether, than the official tag. Takahashi et al. [25] also used four static hashtags to analyze the messages in Twitter in a natural disaster, even when the typhoon went on for five days. In their work about the EU 2014 election trends, Tsakalidis et al. [26] admit that when the selection of hashtags is static, missing data and losing track of the conversation is inevitable, even when they “aggregated tweets written in the respective language that contained a party’s name, its abbreviation, its Twitter account name and some possible misspells” and “excluded several ambiguous keywords in an attempt to reduce the noise”. These examples reflect how when only the expected generic terms or the mainstream hashtags are used, relevant information is lost as they leave little space to unexpected events, shortly lived relevant topics or unusual wordings. In this line, it is also worth noting how different languages, abbreviations, misspellings, or ambiguous keywords can be problematic when using a static approach.

Within the general analysis of trending topics on Twitter, some authors have tried to analyze what makes a topic trend [27], what characterizes these topics [28], what different types of emerging topics exist [29], and even what characteristics are shared by users who started or had a greater influence in the dissemination of trending topics—what has been referred as Twitter trend demographics [30]. Nevertheless, only a limited number of works have tackled real-time topic detection in Twitter. Choi and Park [31] proposed a method to detect emerging topics on Twitter using high utility pattern mining (HUPM), which takes the frequency of appearance and the utility of words into account. Although their approach gives good results when detecting topics in known datasets, it is not designed to dynamically use the resulting topics for extraction. With a different approach, the work of Adedoyin-Olowe et al. [32] aims to detect relevant events from a set of twitter posts, but it displays a similar problem as the analysis is conducted in postprocessing, instead of detecting events in near real time and tracking them to extract all the relevant tweets produced by the users. Such adaptation is found in Gaglio et al. [33], who proposed a system able to progressively refine its query to include new relevant terms, reflecting the emergence of new topics or trends. In their conclusions, they also noted how “other systems were unable to capture the social aspects of the observed events [...] every time the users left the main topic and started to talk about unexpected events”. Their work, however, presents a couple limitations that could have a relevant impact in many contexts, as an initial set of hashtags must be provided in advance, and afterwards, the adaptation is solely based on the extracted tweets, without considering the relevancy (or authority) of their authors. This could easily become an issue, as irrelevant hashtags can get into the system, gradually drifting to diverging topics. Our proposed

solution aims to overcome the previous shortcomings by proposing a method which works in real time, provides precise outputs, and is adjustable against spurious emerging topics.

It is important to clarify that the problem our model solves is a simplification of the general problem of trend detection and monitoring. As previously mentioned, other works monitor hashtags or topics in the tweets published by any user, while we analyze those tweets produced by a known set of users only. This set of users is not necessarily fixed—this can be customized—but it is quite stable and constrained. In any case, the size of the corpus is still enormous, and the variations in time and volume scale of our datasets are still there. To control these two aspects, and to do so dynamically in one pass is still challenging. As we will see in the next section, our solution provides excellent results without the need for any preprocessing or the creation/use of additional corpora.

### 3. Method Description

The method that follows aims to provide a tool that, once a context has been chosen, is able to follow and extract the online conversation that is generated among Twitter users in the most automatic manner possible, while keeping the possibility of manual adaptations whenever needed. This solution is loosely inspired by the resource management systems (RMS) [34] found in computing systems; as RMS manage resources and processes and prioritizes them according to a set of criteria (such as relevancy or aging), the algorithm found in the current work aims to manage the pool of hashtags (or topics) that emerge from relevant sources in a particular subject and prioritizes them according to given criteria but also periodically updates and renews them following the pace of a rolling time window.

In summary, a first script monitors the Twitter stream, following all the posts by the “authorities” (a previously determined set of users that are deemed as relevant by the researchers in the studied context). Their tweets are then examined to extract the hashtags they use, and a weight can be added to them (giving, for instance, more relevance to authorities with more followers). As a result, for each time period, an ordered list of hashtags (that are equivalent to conversation topics) is produced and passed to a separate script, that receives the top  $T$  terms, updates its tracking filter for the stream, and immediately begins to extract (and store) tweets from any user that match those terms. Figure 1 summarizes the process, which can be divided into four main steps: (1) set of authorities; (2) activity monitor; (3) list of hashtags; and (4) tweet collection.

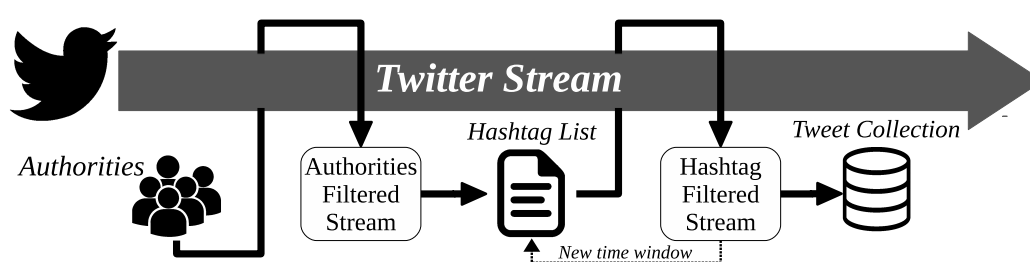


Figure 1. General overview and flow of the proposed tool.

#### 3.1. Set of Authorities

The first step of this method is to determine a set of relevant users (considered “authorities”) for the topic or context that is being analyzed. These are users that lead or start the conversation, such as political candidates in a political campaign or health authorities in a health crisis. The criteria can be set by experts in the field or by the researcher themselves, but ensuring the quality of the set of authorities is key for accurate tracking. Through the Twitter API, the authorities are monitored, preventing an undesired drifting of our tracking to scopes outside the relevant context. Such an approach is key to allow for unattended tracking, as the hashtag selection will change synchronously with the changes in the authorities’ public discourse.

When studying political campaigns, for instance, one would follow all the relevant politicians or candidates, and might consider including the media too. In disasters or health crises, one could use the official accounts from the related institutions, an approach that could also be used to study virtually any other domain.

### 3.2. Activity Monitor

The defined set of authorities in the previous step is used to filter the Twitter API stream. The output is a constant flow of messages (tweets) where: (1) the authority is the author of the post or retweet; (2) the authority post has been retweeted; or (3) a previous tweet from the authority is replied to.

Algorithm 1 describes the flow that is followed by each tweet obtained from the “authorities” tracking. The aim of this first script is double: first, it accumulates the hashtags contained in the “authorities” messages and weights them; second, it is in charge of the time window, effectively setting the beat to which the system updates itself.

---

#### Algorithm 1: Tweet Monitor.

---

```

Input: A Tweet coming from the “authorities” stream and  $C_t \leftarrow$  Set of hashtags for current
time window  $t$ . ( $C_t$  is empty at the start of every time window  $t$ ).
Output:  $C'_t \leftarrow$  An updated set of hashtags for the same window  $t$  (in case the period has not
changed) or a new set of hashtags for the new window  $t + 1$  ( $C_{t+1}$ ) (in case a new
period has started)
begin
  /* Check for a new time window (and renew if true) */
  if new time window then
    | Save  $C_t$  to file /* Store  $C_t$  to use it later */
    | Clear  $C_t$  /*  $C_t$  becomes  $C_{t+1}$  */
  end
  /* Go over the Tweet that was obtained from the stream */
  for each Hashtag in Tweet do
    | if Hashtag not in  $C_t$  then
    | | Add Hashtag to  $C_t$  with weight  $w$ 
    | else
    | | Increase  $C_t[\text{Hashtag}]$   $w$  units
    | end
  end
end

```

---

The collection  $C$  of hashtags in time window  $t$  ( $C_t$ ) contains the hashtags adopted by the authorities and included in their tweets in a single time period. The script stores all the hashtags from each post (they are recognized using the # character). Here, it is possible to compute a weight  $w$  following a user-defined weight function, introducing a certain degree of flexibility in the prioritization, as the weight function can be adapted to the desired application. Some examples include a simple repetition count (number of messages that include the hashtag in question), weighting by the user follower base (e.g., a larger potential impact of a hashtag if tweeted by a user with more followers), limiting the influence of large users by using the logarithm of the same number, or any other mathematical function that fits the application.

The user must define the time window  $t$ , which again, allows for customization, as the tracking of different events might have different needs. When following a political election period or a public health crisis, for example, the rate of messages by the authorities is relatively slow, as are the user reactions, which often last for hours or days. In these contexts, the time window could be set at one hour or even one day. On the other hand, briefer events such as a football match or a song contest, where users basically comment during the event itself and react to what happens in them, would also

need a briefer window. Whatever time period is chosen, the script stores an arranged collection of key terms ( $C_t$ ) for each of them.

### 3.3. List of Hashtags

The collection derived from the previous step might contain a large number of hashtags; therefore, a subset must be chosen. Multiple approaches are possible, but the one proposed here combines a certain number of “hot” topics with a few “emerging” ones, that will be used as an example for the selection (although this is also customizable by the researcher), as described in Algorithm 2.

---

#### Algorithm 2: Tweet Extractor

---

**Data:**  $C_t$  and  $C_{t-1}$   
*Stopwords* is a list of undesired or spurious hashtags  
**Result:** Twitter Stream tracking (at most) the top  $T$  Hashtags, composed by  $H$  hot topics and  $E$  emerging trends

```

begin
  /* Generate the list of Words(Hashtags) to follow          */
  Remove Stopwords from  $C_t$  and  $C_{t-1}$ 
  Hot_hashtags = Top  $H$  Topics in  $C_t$ 
  Emerging_hashtags = Top  $E$  Topics in  $(C_t - C_{t-1})$  and not in Hot_hashtags

  Track_hashtags = Hot_hashtags + Emerging_hashtags

  /* Note that  $(H + E) \leq T$                                */
  /*  $T$  is the maximum number of hashtags to be tracked     */

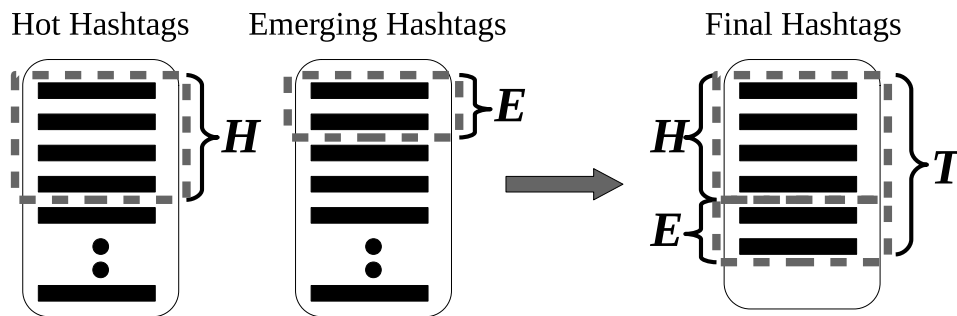
  /* Generate Stream 2                                       */
  Create Stream 2 tracking Track_hashtags
end

```

---

The following steps are carried out in order to establish the hashtags that will be tracked in the extraction script:

1. In each window, the top  $H$  hashtags (referred as “hot”) in the list derived from the monitor step are selected according to their weight. The reasoning is that these are the keywords that the authorities are most strongly using on their posts;
2. Additionally, for each window, the current list ( $C_t$ ) is compared to the previous period ( $C_{t-1}$ ); from the difference between them, a list of emerging terms (labeled as  $E$ ) is obtained (as a customization, this method could look at the  $n$  previous windows in case it was appropriate for the application). The emerging hashtags are a list of keywords that represent topics that could easily become “hot” in the following windows (as they exhibit a fast growth) but they have not reached their peak yet. This way, the method is able to implement an early detection and tracking of newer topics;
3. The combination of hot and emerging hashtags from the previous steps forms a final set of  $T$  total hashtags (of at most  $H + E$  elements). This step is graphically described in Figure 2;
4. Adding a set of stop words is another customization that might be useful for most of the applications in order to avoid capturing common daily spurious expressions used widely by Twitter users (e.g., #happysunday, #goodnight or #followfriday).



**Figure 2.** The final set contains  $T$  hashtags, which is a combination of the “hottest” ( $H$ ) hashtags and the ones that are “emerging” ( $E$ ) or rapidly rising.  $H$  and  $E$  are customizable for each application.

### 3.4. Tweet Collection

A separate script is then set up to track the final set of hashtags in a new stream. This second script is in charge of extracting all the tweets that match the filter (in this case, formed by the obtained  $T$  hashtags) in the Twitter stream, what must be noted is a sample or subset of all the tweets that are being generated by the public users in Twitter [35]. Case is ignored, so differences in the use of lower- or uppercase letters are avoided. This script is updated with a new list of terms in each window, which makes an unattended following of the public conversation around the relevant topics possible, while adapting the hashtags as they change in relevancy. Finally, data processing can be done at a later stage or in (near) real time.

## 4. Results

The methodology described in the preceding section was applied to the period around the Spanish general elections held on 28 April 2019, as the subject of our case study.

### 4.1. Political Context and Background

The Spanish general elections of April 2019 were relevant from multiple perspectives. Pedro Sanchez—leader of the Socialist Party (PSOE)— had been governing in a minority since June 2018, when he successfully used the no-confidence vote against the previous president and Popular Party (PP) leader Mariano Rajoy. Sanchez defeated Rajoy with the help (and votes) of left populists Podemos as well as with several regional (Catalan and Basque) nationalist parties; however, he had to call for early elections after Catalan nationalists withdrew support for the government’s budget. The situation was further complicated by the trial of Catalonia independence leaders that began on 12 February 2019.

Talks on possible coalitions were delayed due to the local and European elections that were held on 26 May. Although conversations lasted for weeks and were featured prominently in the press, the term of office finished early as a result of the failure in government formation negotiations, culminating in Pedro Sanchez’s failed investiture voting on 23–25 July 2019.

In Spain, it is not legal to request votes outside the official election campaign period, whose duration varies from election to election but which always ends two days before the election day. The day before the election, the so-called “reflection day” and the election day, explicitly asking for votes is not permitted. Taking all this into account, our algorithm tracked the selected authorities (see Section 4.2) from 00:00 CEST on 1 April 2019 (four weeks before the election day and two weeks before the official start of the campaign) until the end of 31 July 2019 (with the final result of the failed investiture attempt).



#### 4.2. Selection of Authorities

A total of 238 user accounts were monitored. Most of them (225) were individual accounts of the head candidate for every party in each province. Given that Spain is divided into 50 provinces and two so-called autonomous cities, at most any party could have 52 candidates. The distribution of the authorities is listed in Table 1; these are all the candidates that have Twitter accounts. Additionally, the thirteen official and verified accounts from each party are also added (there is an extra account due to the Catalan branch of Podemos). Note how nation-wide parties such as PSOE or PP have a number of authorities that is close to the maximum (43) while regional parties have only candidates in their respective provinces (e.g., Esquerra Republicana de Catalunya (ERC) has four candidates on Twitter, one heading each province).

**Table 1.** Distribution of authorities among parties.

Party	#Authorities	Example	Party	#Authorities	Example
Coalición Canaria	2	@anioramas	ERC	4	@junqueras
Compromís	3	@joanaldovi	JxCat	4	@jorditurull
Ciudadanos	44	@albert_rivera	Podemos	42	@PabloIglesias
EAJ-PNV	3	@AITOR_ESTEBAN	PP	43	@pablocasado_
EH Bildu	2	@OskarMatute	PSOE	43	@sanchezcastejon
En Marea	4	@baranauskas_ana	Vox	26	@Santi_ABASCAL
Others	5				

As Table 1 shows, the selection was exhaustive and systematic, as it contained the full set of leaders per party (as well as the official account of each party) across all Spanish provinces. Remarkably, all those accounts combined have 17,982,172 followers—Spain has a population of approximately 47 million people—and although many followers follow more than one account, it is worth noticing the order of magnitude, and thus, the deep reach that these 238 accounts can achieve with a single tweet.

#### 4.3. Experimental Settings

Here, the logarithm of the author's number of followers was chosen as the weight ( $w$ ) function in order to reflect that "authorities" with a larger number of followers are considered more relevant (as they have a wider reach) but the order of magnitude was reduced to prevent the existence of disproportionately dominant authorities such as the parties' accounts or the president himself. On the other hand, the time window ( $t$ ) was set at one hour, allowing for more granularity and a faster response in extraction than a daily setting and being more representative than a few minutes window. In regard to the number of hashtags selected for extraction, a  $T$  of 10, with eight hot ( $H$ ) and two emerging ( $E$ ) hashtags was assumed as the optimal setting.

Additionally, a list of common stopwords derived from previous experiments and containing a list of spurious terms (such as #buenosdias or #felizlunes, the Spanish equivalents of #goodmorning or #happymonday) was used to filter the resulting terms.

#### 4.4. Experimental Results

After the authorities were selected (see Sections 3.1 and 4.2 for reference), they were continuously monitored (see Section 3.2), obtaining a list of hashtags in each time window (in this case, each hour). The total number of tweets by the authorities in the four months (122 days/2928 h) tracked was 250,636; an average of between 85 and 86 tweets were obtained per hour (although this figure is obviously higher during the day and lower during the night).

The total number of different hashtags (not case sensitive) produced by the authorities was 61,043, although only a portion (1991) reached the top ten list at least once in the 2928 one-hour windows. Table 2 shows the most frequently found hashtags among the hourly top ten as a sample.

The extraction using the captured topics of discussion among the authorities (see Section 3.4) resulted in 25,941,397 tweets over 122 days, an average 212,634 tweets per day (again, although a few exceptional days in April reach more than one million tweets and some in July go down to a few ten thousands).

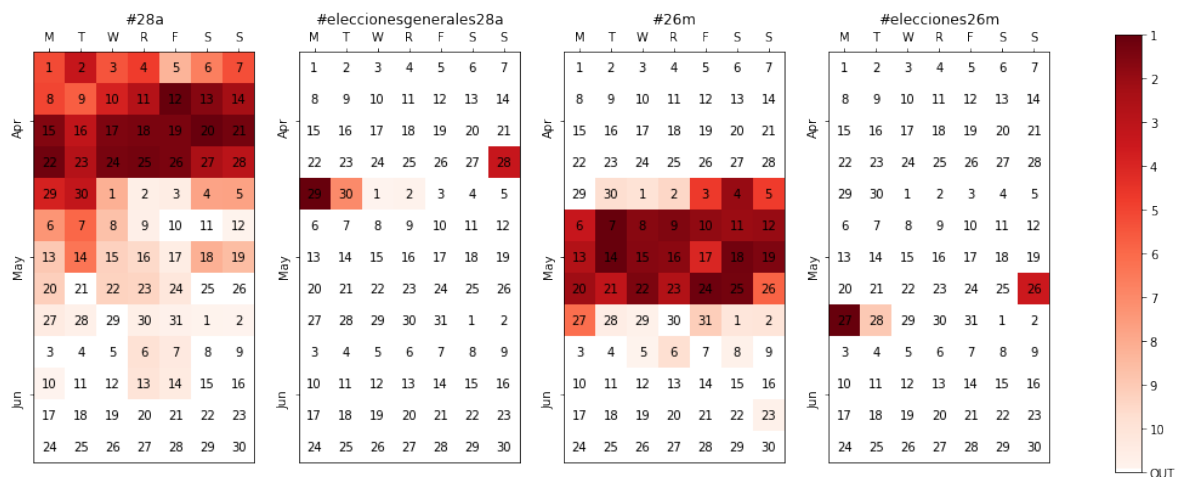
**Table 2.** Hashtags that appear in the largest number of hourly extraction lists.

Hashtag	Periods among the Top 10
#españaviva	1142 out of 2928 (39.0%)
#28a	952 (32.5%)
#absolució	686 (23.4%)
#psoe	657 (22.4%)
#freetothom	656 (22.4%)
#26m	654 (22.3%)
#porespaña	645 (22.0%)
#votapsoe	593 (20.2%)
#hazquepase	566 (19.3%)
#siemprehaciadelante	474 (16.2%)

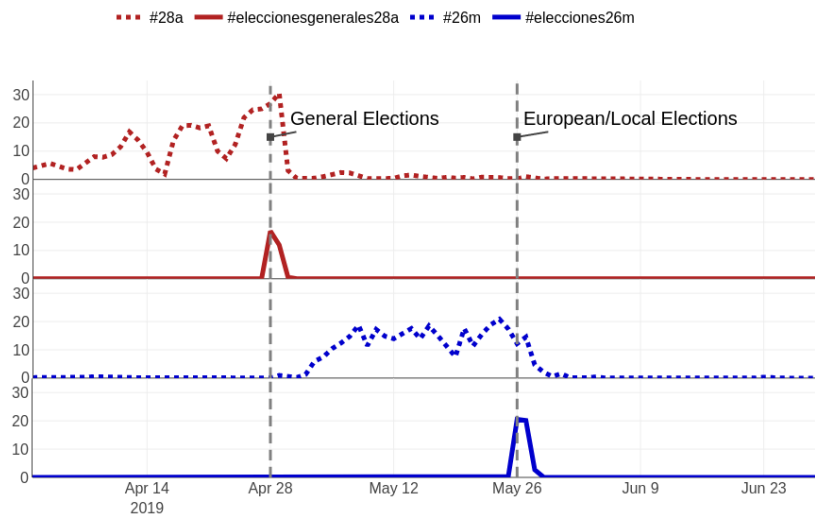
In order to illustrate how our approach follows the conversation topics, a sample of the most relevant hashtags is presented as a specific case study in the following subsections. In them, calendar heatmaps are used to show the relevance of such hashtags over time; for the sake of clarity, the hourly classification has been averaged daily. Thus, for instance, a hashtag that ranks first (1) during 12 h and fifth (5) during the other 12 is represented with a three (3) in the heatmap. In addition, note that even though the relevancy of each topic in the top ten tracked hashtags obtained from the authorities is different (due to the assigned weight), they are all tracked equally in the extraction script.

#### 4.4.1. General Tracking

First, we focus on the two most general hashtags, adopted by politicians, media, and general public to tag their messages related to both electoral periods. As the Spanish general elections were held on 28 April 2019, the adopted hashtag was #28a. On the other hand, the European and local elections were held on 26 May, using an equivalent hashtag (#26m). Figure 3 shows how our method tracked both hashtags consistently in their relevant periods (April and May, respectively), resulting in 3.9 million tweets being extracted. Noticeably, two related hashtags were also captured in very specific subperiods: both right at the end of the respective voting day: #eleccionesgenerales28a (general elections 28A) and #elecciones26m (elections 26M). Upon further inspection, we found out that #eleccionesgenerales28a and #elecciones26m were the hashtags used in media for the TV shows that tracked the recount as well as for the subsequent talk-shows that analyzed the results. Opposite to what happened with the previous tags—#28a and #26m, probably easier to predict, short, and thus fast to type—here there is not a complete consistency between both hashtags, hindering the chances of predicting the second hashtag once the first one (#eleccionesgenerales28a) is known. Naively, one could expect something like #elecciones europeas26m (European elections 26M), which did not happen. In any case, the relevancy of these unexpectedly captured hashtags (#eleccionesgenerales28a and #elecciones26m) should not be underestimated, as in the day after both electoral processes, they reach a higher relevancy than their general counterparts (more details in Figure 4). For reference, the translation to absolute figures means that a total of 662,413 tweets—442,379 discussing the results of the general elections and 220,034 for the European/local ones—would have been lost had it not been for the automated approach.



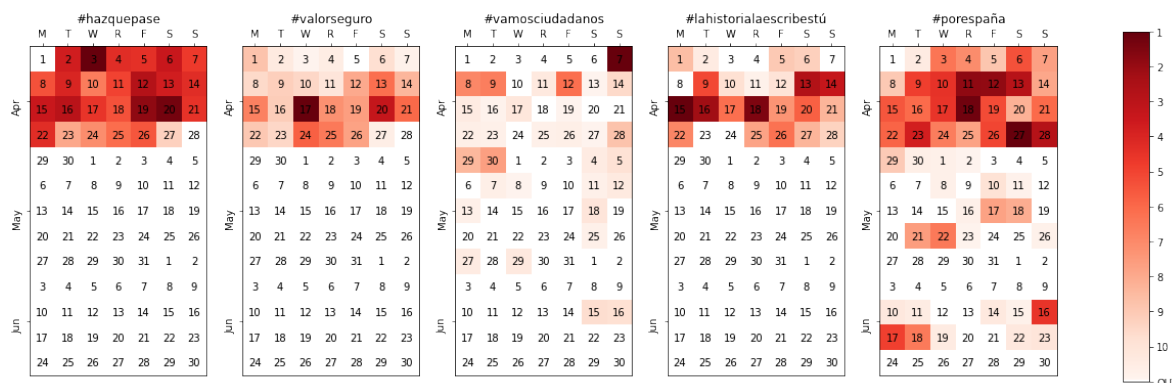
**Figure 3.** Tracking the relevancy of the general hashtags in the election period; notice how #28a (campaign) is followed by #eleccionesgenerales28a (results) and the interest in #26m surges right afterwards, also ending in a result-related topic (#elecciones26m) after the 26M vote.



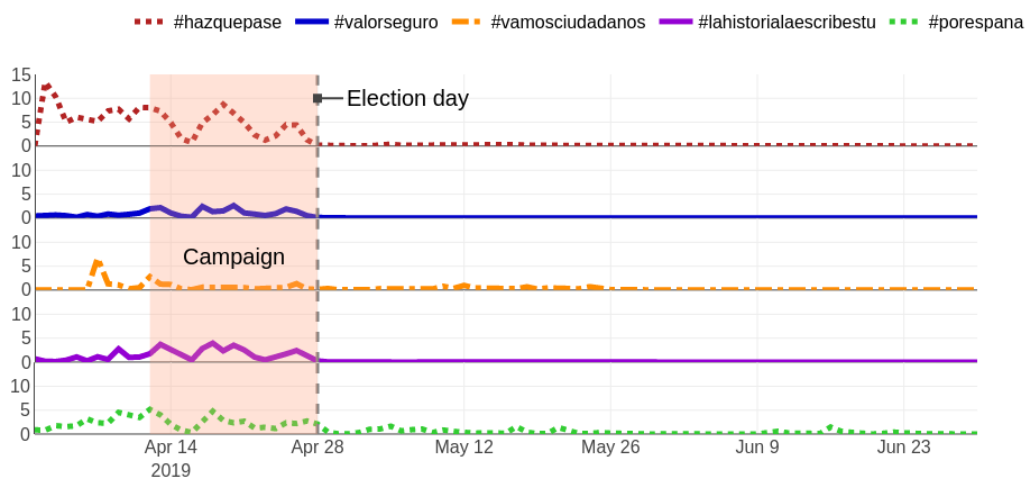
**Figure 4.** Daily relative amount of tweets extracted using #28a, #26m, #eleccionesgenerales28a, and #elecciones26m (in percentage). Notice the spikes at the end of each electoral period for the latter two.

#### 4.4.2. Electoral Campaign

Similar observations can be made when tracking both electoral processes changing the perspective to that of each political party (Figure 5). The monitoring of authorities was conducted from 1 April, twelve days before candidates kick-started the campaign and unveiled the official slogans of the parties. PSOE’s and VOX’s slogans, #hazquepase (make it happen) and #porespaña (for Spain), however, were not used until 2 April, while Ciudadanos’ one (#vamosciudadanos, let’s go ciudadanos) was not presented until Sunday 7 April, where it had the most activity. The relative activity of each of the five nationwide parties is shown in Figure 6, where the simultaneous and sudden fall of activity right after the campaign ends also becomes visible.



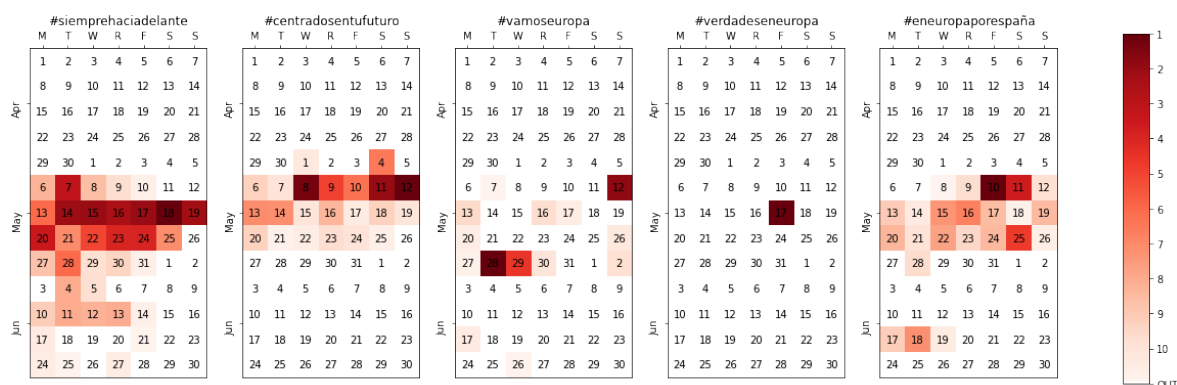
**Figure 5.** Calendar heatmap for the official slogan per party (top 5 parties) during the Spanish general elections campaign.



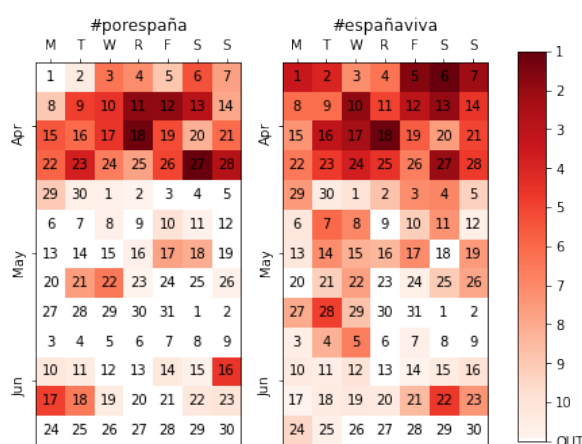
**Figure 6.** Daily relative amount of tweets extracted using each party’s slogan (#hazquepase, #valorseguro, #vamosciudadanos, #lahistorialaescribestú, and #porespaña) in percentage. Notice how the activity ceases after the election day.

Immediately after the end of the 28A campaign, there is remarkable surge of interest in the next campaign, as previously shown in Figures 3 and 4. Our unattended tracking method eliminates the need to intervene; as soon as the new slogans are revealed, they are captured (Figure 7) just like in the previous campaign. A couple of things are worth mentioning here in any case. First, there was, in general, less interest in the 26M campaign than in the 28A from a communication point of view. Second, the hashtags associated with the 26M process had, in general, little to do with those used during the 28A, which again, would have complicated early tracking of these.

Another case of interest is depicted in Figure 8, with a side-by-side comparison between the official slogan of the party (#porespaña, for Spain) and their preferred hashtag on social media (#españaviva, the living Spain), which became the number one hashtag in 39% of the time windows (Table 2). This basically implies that, if only the “official” slogans had been followed, the extraction would have resulted in 455,954 tweets but failed to extract 536,889 tweets containing the #españaviva hashtag. This is just another example of an unpredictable topic that our method did capture which, in this particular example, coincides with the most used hashtag over the four months.



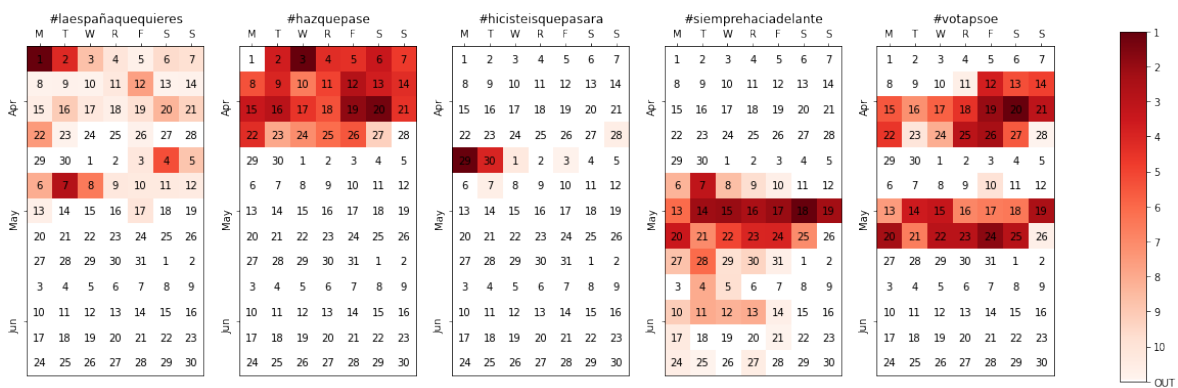
**Figure 7.** Calendar heatmap for the official slogan per party (top 5 parties) during the European/local elections campaign.



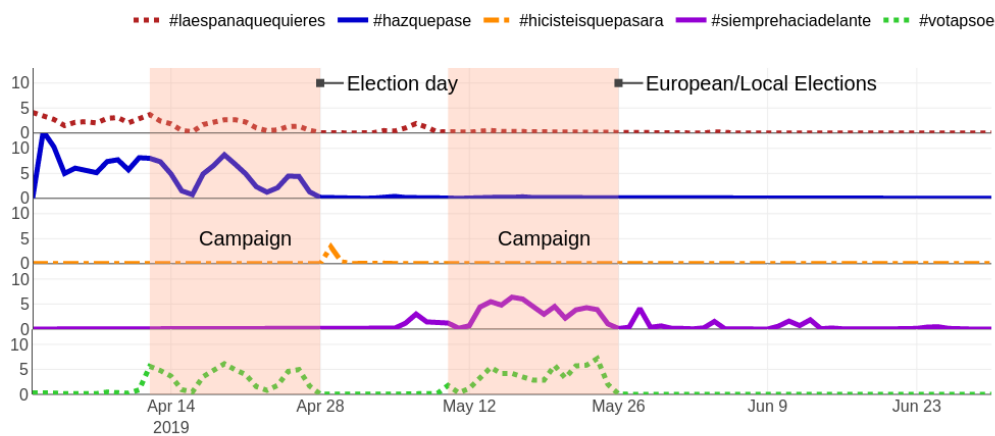
**Figure 8.** Side-by-side comparison between Vox' official slogan (#porespaña) and the most used hashtag on social media during the campaign (#españaviva).

Let us focus on the activity related to PSOE—the party that won the elections—and follow the automatic evolution of their hashtags over time. As Figure 9 shows, the 1st of April marks the highest point for #laespañaquequieres (the Spain that you want) where the PSOE candidates asked their followers in Twitter for suggestions on their candidature, supposedly to include them in their electoral program. The relevancy of this first hashtag quickly went down (without disappearing from the top 10 in many time periods), handing over the baton to the official slogan for the 28A elections, #hazquepase (make it happen). However, the most notable effect of our method is clear when analyzing the third hashtag, #hicisteisquepasara (you made it happen). It is not only unlikely to predict the winner of the elections, but it is also nearly impossible to guess the hashtag that would be used to celebrate such event, with more than 30,000 tweets extracted during the 48 h that followed the announcement of the results. In the fourth place, the then-new slogan for the 26M electoral process (#siemprehaciadelante, always going forward) greatly gained relevancy from the moment the 28A elections were over and until the new election process ended. Incidentally, some authorities used a misspelling (#siemprehaciaadelante, with a double “a”) which was also captured during a few time windows, and although it was only used in 4181 extracted tweets (which could be considered anecdotal versus the 197,968 of the correct hashtag), it shows that even in cases where the authorities use misspellings or alternatives to the supposed hashtag, our method would have detected it and tracked it equally. The final calendar heatmap, at the far right in Figure 10, reflects how the explicit message asking for votes—using the hashtag #votapsoe—only has activity coinciding almost perfectly with the campaign periods for both processes (from the 12th to the 26th of April and from the 10th

to the 24th of May) and the day of reflection (the election day eve). For illustrative purposes, the full evolution from the perspective of the extracted tweets is depicted in Figure 10.



**Figure 9.** The evolution of PSOE-related hashtags, since the Spanish general elections pre-campaign to the European post-campaign.



**Figure 10.** Daily relative amount of tweets extracted using the PSOE-related hashtags in percentage (#laespañaquequieres, #hazquepase, #hicisteisquepasara, #siemprehaciadelante, and #votapsoe). Notice how the activity responds to the electoral processes.

#### 4.4.3. Televised Debates

Debates are arguably the most relevant events during electoral processes, and in particular, televised leaders’ debates. The conversation during political debates on TV often migrates to the online plane, so it is also relevant to look at the behavior of the unmanned methodology proposed here in regard to the main televised debates that happened during the Spanish general election process, which are summarized in Table 3.

**Table 3.** Summary of the debates’ hashtags. (\*) Note that #eldebatedecisivo was not the hashtag for the debate but was widely used for the talk-shows that discussed its outcome.

Hashtag	TV Channel	Date	Tweets Extracted
#l6neldebate	La Sexta	13/04/19 22:00	72,749
#eldebateenrtve	TVE	22/04/19 22:00	952,546
#debateatresmedia	A3/La Sexta	23/04/19 22:00	116,732
#eldebatedecisivo	A3/La Sexta	(*)	971,681
#debattv3	TV3	24/04/19 22:00	90,937

During the first weekend of the campaign, a warm-up debate among female candidates—hashtagged with #L6Neldebate—was held on La Sexta Noche, a political late night show. Then, the two main debates with the leaders of the main parties took place on the 22nd in the public television (TVE) and the 23rd in Atresmedia (a media corporation providing signal to both Antena 3 and La Sexta tv channels). Other regional debates, such as the one on Catalan public television (TV3), were also captured; this one on the 24th of April, in particular, involved the leading candidates for Barcelona.

Besides the detection of the topics themselves (see Figure 11), a few things are worth considering. First, the wording used in the hashtags is inconsistent, so again, even if the TVE debate is tracked with #eldebateenrtve, the Atresmedia one shortens the correct sentence “el debate en Atresmedia” into #debateatresmedia. Second, different languages are captured without the need for adaptation; #debattv3, for instance, which is in Catalan instead of Spanish. Third, all events happen during prime time on the nights of their respective days, but the discussion is as important (or, in some cases, even stronger) on the following day (see Figure 12). Therefore, had we tracked the hashtag only during the TV program, at least half of the information would have been lost. Finally, notice how the Atresmedia debate was actually discussed on talk-shows using a different hashtag (#eldebatedecisivo, the deciding debate); this results in a dual conversation, one much weaker—hashtagged with the “official” #debateatresmedia, only 116,732 tweets—and another one which generates much more conversation—hashtagged with the spontaneous #eldebatedecisivo, 971,681 tweets. Thus, tracking only the official hashtag for the debate would have meant a loss of almost a million tweets containing the bulk of the discussion.

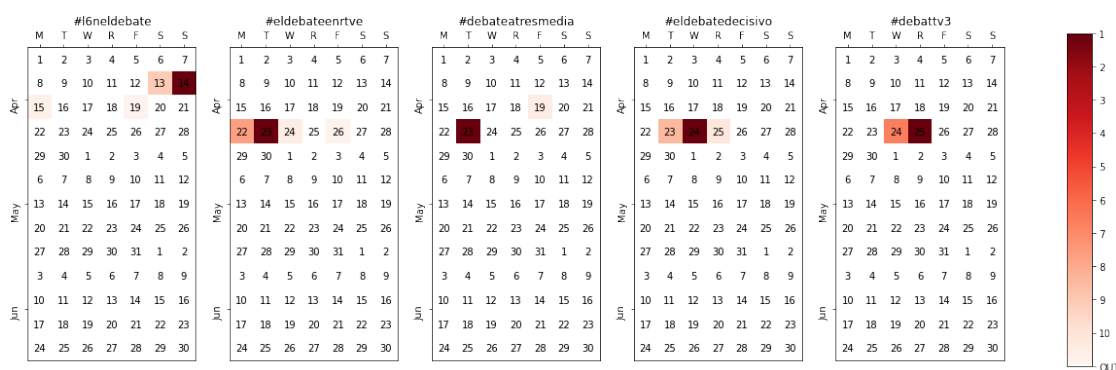


Figure 11. Calendar heatmap for the multiple televised debates with candidates.

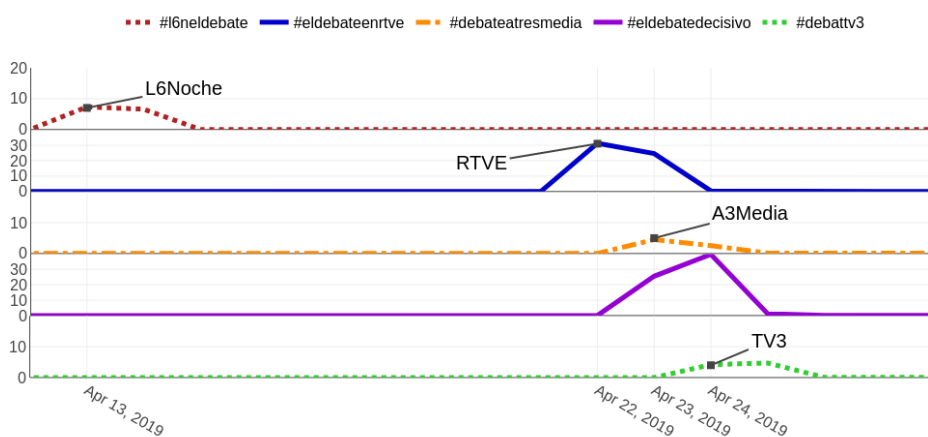
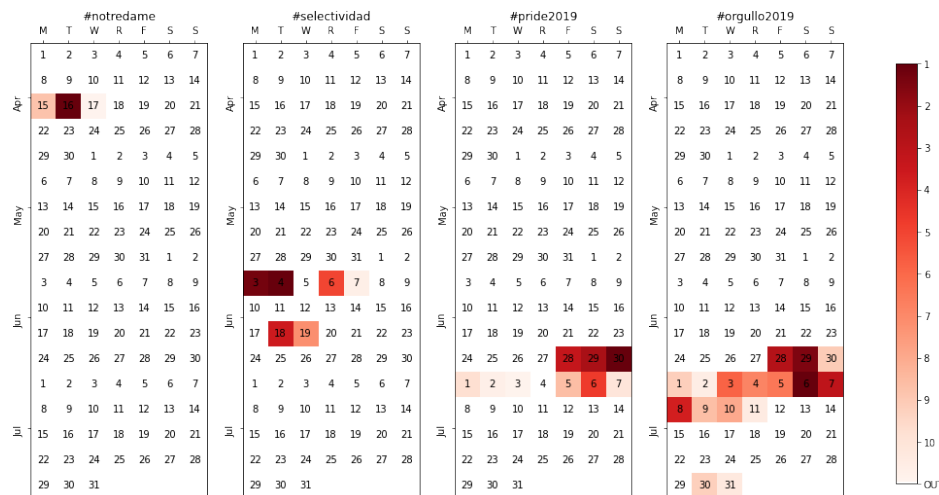


Figure 12. Daily relative amount of tweets extracted using the hashtags for each political debate in percentage (#l6neldebate, #eldebateenrtve, #debateatresmedia, #eldebatedecisivo, and #debattv3). For the purposes of clarity, notice that the y-axis scale is not equal across all hashtags.

#### 4.4.4. Event Detection and Tracking

Detecting events and tracking them is, indeed, one of the objectives of the methodology. In broad terms, the events taking place can either be desired (when they are part of the context under study) or undesired (when they are external phenomena). In Figure 13, three different events are analyzed. The first event, hashtagged with #notredame, refers to the fire that devastated Notre-Dame de Paris cathedral on the afternoon of 15 April 2019. Consternation ensued, and most politicians quickly expressed their lament on social media. Notice, however, how the tracking happens only during the highest point of global dismay (on the 16th, with the images of destruction after the fire was extinguished) and swiftly fades afterwards. As soon as the authorities change the subject, the topic—undesired for the political tracking—disappears. A similar case but with a routine event happens with #selectividad, a hashtag devoted to the university entrance exams that take place in June in Spain. There are a few periods where the hashtag gains some relevancy among the top topics, corresponding to the dates where the test is held in different regions (so, each region’s politicians direct their posts to their own region’s students), but it also fades right afterwards.



**Figure 13.** Calendar heatmap for a few events that happened between April and July 2019, namely #notredame, #selectividad, #pride2019, and #orgullo2019.

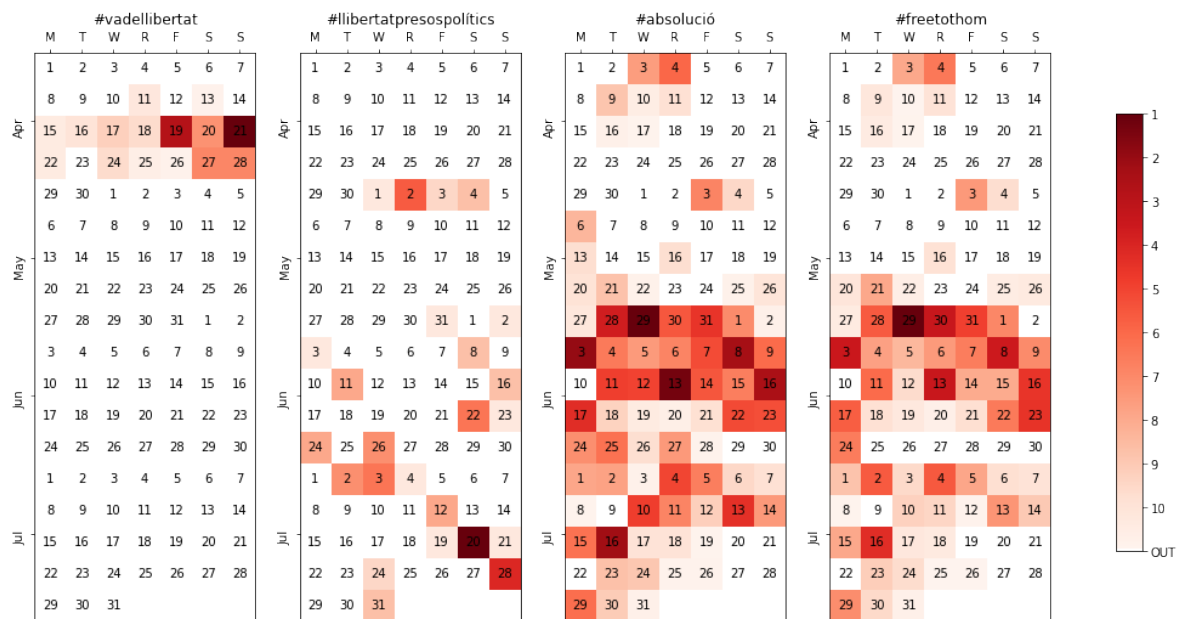
A different case takes place with #pride2019 and #orgullo2019, two linguistic variants (English and Spanish) of the same event, the annual LGBT pride festival. Such a topic is not exempt from political content and discussion on social media, especially between conservative and progressive parties. This is reflected in the length of the topic tracking, which lasts for nine days in the first case and two weeks in the second (as expected, the Spanish version is used for a longer time). These would be two examples of events that are somehow related to the conversation that is being tracked so such lengthy tracking represents the desired behavior.

#### 4.4.5. Languages

Many contexts might include more than one language, especially when a topic is tracked globally. In the case of Spain, besides Spanish, some regions have their co-official languages. Most people in Catalonia, for instance, use Catalan as their first language and consequently “tweet” in Catalan. Two of the five top topics, tracked in more than one-fifth of the time periods, were related to the situation in Catalonia after the 2017 independence referendum and the subsequent trial: #absolució (acquittal) and #freetothom (free everyone), were commonly used together reclaiming the liberation of the pro-independence leaders, as does #llibertatpresospolítics (freedom for the political prisoners). The former two raised their activity at the end of May (Figure 14) following several simultaneous events: three of the accused leaders were elected in the European elections, and the public prosecutor



found the separatist leaders guilty at the end of the trial, all while the winning party hypothetically needed the votes of the nationalists to elect its candidate as president. All these sparked a notable activity among some of the authorities which, even representing only around 5% of the total accounts tracked, often managed to bring the topic to the top. Around 200,000 tweets were extracted using these hashtags in Catalan, a relevant amount taking into account that they are in a language spoken roughly by 15% of the Spanish population (not to forget the fact that they were recovered during June and July, a period of less activity in general). Finally, #vadellibertat (this is about freedom) is also shown as an example of the slogan of a regional party being tracked, in this case, the one used by the Catalan party ERC.



**Figure 14.** Calendar heatmap for #vadellibertat, #llibertatpresopolitics, #absolució, and #freetothom, all related to the Catalan nationalist movement.

### 5. Discussion

Political campaigns have been the subject of multiple previous works, as highlighted previously. However, this is to our knowledge the first attempt to obtain a dynamically adapted dataset over a period of four months of political campaigning and overall conversation. The prominence on the use of Twitter for politics, and the interest that the study of political information awakens is what inspired us to apply our model to the analysis of political discussions in Twitter.

Previous studies analyzed political campaigns using only a few static hashtags or terms; for instance, Yakub et al. [36] used three fixed terms ('Trump', 'Clinton' and 'Election2016') to download tweets in the context of the 2016 US presidential elections, while Lai et al. [37] used a similar method to track the discussion on the 2016 referendum on the reform of the Italian constitution. Others included expected variations or lists of terms without ensuring their exhaustivity [38,39]. In comparison, during our study of the context of the Spanish general and European/local elections, more than 60,000 different hashtags were generated by the authorities tracked, with 1991 different unique hashtags among the top 10 topics in at least one time window. Our research, therefore, proves how previous practices are undesired because, either when choosing a few general terms or when trying to guess potential terms, most of the information can get lost. Furthermore, any unexpected relevant event is also left out, as the system is unable to adapt to the newer circumstances, while any potential wrong term is never removed. We have demonstrated that trying to guess hashtags leads to uncertainty; in our case of study, it would have been improbable to guess the 1991 terms that appeared throughout

the four month tracking, let alone the 61,043 different hashtags that were used at some point by the authorities selected.

From a different perspective, it is also clear that following the local authorities in the context of the Spanish elections has rarely resulted in hashtags or topics that escape the Spanish (political) context; therefore, in the absence of methods or data that limit the extraction to a particular region, country, or community, our method seems to provide a reasonable approximation that can help future studies with similar needs.

## 6. Conclusions

Our work shows how the described method is able to follow the conversation in a community or context of interest, monitoring the labels (hashtags) that the related “authorities” adopt in their scope of activity online, and adjusting such tracking as their messages and topics vary, even if they do so in a dramatic manner. We also demonstrated how the method is robust when faced with potential disruptions while avoiding the need to compile an initial list of “seed” words that could either bias or underestimate the actual conversation taking place. Here, term selection does not depend on what is predicted by the researcher; instead, it is determined by the terms most used by the “authorities” together with the relative relevancy of each of them, a process that requires no external human intervention.

The application of this method could, therefore, enrich any work that seeks to track the conversation around a topic without manual intervention. This approach reduces bias as well as the manual effort required when tracking a particular event—such as a debate—or other situation. The framework provided is applicable not only to politics but also to any other context, and therefore, its implications for research on social media, in general, are vast. Additionally, it must be noted that this mechanism is not limited to Twitter; with the proper adaptation, it could be applied to other SNS with equivalent characteristics.

In summary:

- The proposed methodology is able to follow the conversation around a topic and optimize its tracking, with a quick adaptation and no need for human supervision;
- The methodology is also robust against undesired events, misspellings, and the use of different languages and alternative terms;
- Fixed hashtag selection limits the information that researchers might be able to extract; it is desirable to adopt a flexible and dynamic approach.

A few restrictions that limit this methodology should be noted. The first one is that an accurate and expert selection of the authorities is required. The researcher must, therefore, identify the relevant users in the discussion with care. In the case of politics, which was explored in our case study, selection is arguably straightforward as politicians, candidates, and political parties clearly lead the discourse around such matters. Other contexts, however, might have less evident leaders of opinion. On the other hand, there are a few parameters that give flexibility to the system (length of the time window, weight, stop words, number of topics, etc.), but this is a double-edged sword as their determination is critical for the proper functioning of the tracking. Therefore, testing is suggested before committing to a full-scale extraction, either in other smaller cases or in comparable cases when the uniqueness of the event does not allow for testing in the exact same context.

Dynamically adapting the list of authorities (adding or removing users from that set as things change over time) will be considered in future work, as especially in long-term tracking, some users might lose relevancy (e.g., candidates that have not been elected or have resigned), while others not included in the initial set might rise in relevancy (e.g., influential third parties or surprise elements). Identifying the most relevant users in a particular period [40] could mean adding or removing members from the list. Additionally, and as appropriate parameter setting is critical to generalize the applicability

of the model, we will explore how these parameters can be assessed and their results compared in order to optimize their values in a minimally invasive way.

**Author Contributions:** Conceptualization, M.M.-C. and S.S.-A.; Formal analysis, M.M.-C.; Investigation, M.M.-C.; Methodology, M.M.-C. and S.S.-A.; Software, M.M.-C.; Supervision, S.S.-A., E.G.-B. and M.-Á.S.; Validation, S.S.-A., E.G.-B. and M.-Á.S.; Visualization, M.M.-C.; Writing—original draft, M.M.-C.; Writing—review & editing, S.S.-A., E.G.-B. and M.-Á.S. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

SNS	Social Networking Services
API	Application Programming Interface
PSOE	Partido Socialista Obrero Español
PP	Partido Popular
ERC	Esquerra Republicana de Catalunya

## References

1. Turcotte, J.; York, C.; Irving, J.; Scholl, R.M.; Pingree, R.J. News recommendations from social media opinion leaders: Effects on media trust and information seeking. *J. Comput. Mediat. Commun.* **2015**, *20*, 520–535. [[CrossRef](#)]
2. Fang, Y.; Chen, X.; Song, Z.; Wang, T.; Cao, Y. Modelling propagation of public opinions on microblogging big data using sentiment analysis and compartmental models. In *Natural Language Processing: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2020; pp. 939–956.
3. Gadek, G.; Pauchet, A.; Malandain, N.; Vercouter, L.; Khelif, K.; Brunessaux, S.; Grilhères, B. Topological and topical characterisation of Twitter user communities. *Data Technol. Appl.* **2018**, *52*. [[CrossRef](#)]
4. McGregor, S.C.; Mourão, R.R.; Molyneux, L. Twitter as a tool for and object of political and electoral activity: Considering electoral context and variance among actors. *J. Inf. Technol. Polit.* **2017**, *14*, 154–167. [[CrossRef](#)]
5. Jost, J.T.; Barberá, P.; Bonneau, R.; Langer, M.; Metzger, M.; Nagler, J.; Sterling, J.; Tucker, J.A. How social media facilitates political protest: Information, motivation, and social networks. *Polit. Psychol.* **2018**, *39*, 85–118. [[CrossRef](#)]
6. Rainie, L.; Smith, A.; Schlozman, K.L.; Brady, H.; Verba, S. Social media and political engagement. *Pew Internet Am. Life Proj.* **2012**, *19*, 2–13.
7. Park, C.S. Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Comput. Hum. Behav.* **2013**, *29*, 1641–1648. [[CrossRef](#)]
8. Vromen, A.; Xenos, M.A.; Loader, B. Young people, social media and connective action: From organisational maintenance to everyday political talk. *J. Youth Stud.* **2015**, *18*, 80–100. [[CrossRef](#)]
9. Allcott, H.; Gentzkow, M. Social media and fake news in the 2016 election. *J. Econ. Perspect.* **2017**, *31*, 211–36. [[CrossRef](#)]
10. Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput. Mediat. Commun.* **2007**, *13*, 210–230. [[CrossRef](#)]
11. Green, D.P.; Gerber, A.S.; De Boef, S.L. Tracking opinion over time: A method for reducing sampling error. *Public Opin. Q.* **1999**, *63*, 178–192. [[CrossRef](#)]
12. Yang, Y.; Ault, T.; Pierce, T.; Lattimer, C.W. Improving text categorization methods for event tracking. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 24–28 July 2000; pp. 65–72.
13. Yang, Y.; Pierce, T.; Carbonell, J. A study of retrospective and on-line event detection. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, 24–28 August 1998; pp. 28–36.
14. Allan, J.; Carbonell, J.G.; Doddington, G.; Yamron, J.; Yang, Y. Topic Detection and Tracking Pilot Study Final Report. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, VA, USA, 8–11 February 1998; pp. 194–218.

15. Mønsted, B.; Sapieżyński, P.; Ferrara, E.; Lehmann, S. Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PLoS ONE* **2017**, *12*, e0184148.
16. Hendrickson, S.; Kolb, J.; Lehman, B.; Montague, J. Trend Detection in Social Data. Twitter Blog. 2015. Available online: [https://blog.twitter.com/en\\_us/a/2015/trend-detection-social-data.html](https://blog.twitter.com/en_us/a/2015/trend-detection-social-data.html) (accessed on 2 May 2020).
17. Allan, J. Introduction to topic detection and tracking. In *Topic Detection and Tracking*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 1–16.
18. Kleinberg, J. Bursty and hierarchical structure in streams. *Data Min. Knowl. Discov.* **2003**, *7*, 373–397. [[CrossRef](#)]
19. Petrović, S.; Osborne, M.; Lavrenko, V. Streaming first story detection with application to twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 181–189.
20. Bruns, A.; Burgess, J. Twitter hashtags from ad hoc to calculated publics. In *Hashtag Publics: The Power and Politics of Discursive Networks*; Peter Lang Inc.: Frankfurt, Germany, 2015; pp. 13–28.
21. D’heer, E.; Verdergem, P.; De Grove, F. # MissingData: A methodological inquiry of the hashtag to collect data from Twitter. In Proceedings of the AoIR Selected Papers 17th Annual Conference Association Internet Research, Berlin, Germany, 5–8 October 2016; Volume 6.
22. Enli, G.; Simonsen, C.A. ‘Social media logic’ meets professional norms: Twitter hashtags usage by journalists and politicians. *Inform. Commun. Soc.* **2018**, *21*, 1081–1096. [[CrossRef](#)]
23. Fano, S.; Slanzi, D. Using Twitter data to monitor political campaigns and predict election results. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*; Springer: Berlin, Germany, 2017; pp. 191–197.
24. Reyes-Menendez, A.; Saura, J.R.; Alvarez-Alonso, C. Understanding# WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2537.
25. Takahashi, B.; Tandoc, E.C., Jr.; Carmichael, C. Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Comput. Hum. Behav.* **2015**, *50*, 392–398. [[CrossRef](#)]
26. Tsakalidis, A.; Papadopoulos, S.; Cristea, A.I.; Kompatsiaris, Y. Predicting elections for multiple countries using Twitter and polls. *IEEE Intell. Syst.* **2015**, *30*, 10–17. [[CrossRef](#)]
27. Lu, R.; Yang, Q. Trend analysis of news topics on twitter. *Int. J. Mach. Learn. Comput.* **2012**, *2*, 327. [[CrossRef](#)]
28. Naaman, M.; Becker, H.; Gravano, L. Hip and trendy: Characterizing emerging trends on Twitter. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 902–918. [[CrossRef](#)]
29. Lau, J.H.; Collier, N.; Baldwin, T. On-line trend analysis with topic models:# twitter trends detection topic model online. In Proceedings of the COLING 2012, Mumbai, India, 8–15 December 2012; pp. 1519–1534.
30. Cheong, M.; Lee, V. Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In Proceedings of the 2nd ACM Workshop on Social Web Search and Mining, Hong Kong, China, 3–4 November 2009; pp. 1–8.
31. Choi, H.J.; Park, C.H. Emerging topic detection in twitter stream based on high utility pattern mining. *Expert Syst. Appl.* **2019**, *115*, 27–36. [[CrossRef](#)]
32. Adedoyin-Olowe, M.; Gaber, M.M.; Dancausa, C.M.; Stahl, F.; Gomes, J.B. A rule dynamics approach to event detection in twitter with its application to sports and politics. *Expert Syst. Appl.* **2016**, *55*, 351–360. [[CrossRef](#)]
33. Gaglio, S.; Re, G.L.; Morana, M. A framework for real-time Twitter data analysis. *Comput. Commun.* **2016**, *73*, 236–242. [[CrossRef](#)]
34. Krauter, K.; Buyya, R.; Maheswaran, M. A taxonomy and survey of grid resource management systems for distributed computing. *Softw. Pract. Exp.* **2002**, *32*, 135–164. [[CrossRef](#)]
35. Morstatter, F.; Pfeffer, J.; Liu, H.; Carley, K.M. Is the sample good enough? comparing data from twitter’s streaming api with twitter’s firehose. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013; pp. 400–408.
36. Yaqub, U.; Chun, S.A.; Atluri, V.; Vaidya, J. Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Gov. Inf. Q.* **2017**, *34*, 613–626. [[CrossRef](#)]

37. Lai, M.; Patti, V.; Ruffo, G.; Rosso, P. Stance evolution and twitter interactions in an italian political debate. In *International Conference on Applications of Natural Language to Information Systems*; Springer: Berlin, Germany, 2018; pp. 15–27.
38. Howard, P.N.; Kollanyi, B. Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum. *SSRN Electron. J.* **2017**. [[CrossRef](#)]
39. Howard, P.N.; Bolsover, G.; Kollanyi, B.; Bradshaw, S.; Neudert, L.M. Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter? | The Computational Propaganda Project. *Tech. Rep.* **2017**. Available online: <http://275rzy1ul4252pt1hv2dqyuf.wpengine.netdna-cdn.com/wp-content/uploads/2017/07/2206.pdf> (accessed on 2 May 2020).
40. Mahmoudi, A.; Yaakub, M.R.; Bakar, A.A. New time-based model to identify the influential users in online social networks. *Data Technol. Appl.* **2018**, *52*, 278–290. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).