Noname manuscript No.
(will be inserted by the editor)

Language Agnostic Meme-Filtering for Hashtag-based Social Network Analysis

Dimitrios Kotsakos · Panos Sakkos · Ioannis Katakis · Dimitrios Gunopulos

the date of receipt and acceptance should be inserted later

Abstract Users in social networks utilize hashtags for a variety of reasons. In many cases hashtags serve retrieval purposes by labeling the content they accompany. More often than not, hashtags are used to promote content, ideas or conversations producing viral memes. This paper addresses a specific case of hashtag classification: meme-filtering. We argue that hashtags that are correlated with memes may hinder many valuable social media algorithms like trend detection and event identification. We propose and evaluate a set of language-agnostic features that aid the separation of these two classes: meme-hashtags and event-hashtags. The proposed approach is evaluated on two large datasets of Twitter messages written in English and German. A proof-of-concept application of the meme-filtering approach to the problem of event detection is presented.

1 Introduction

On-line social networks analysis recently attracted attention from various scientific fields like Social Psychology (Quercia et al 2011), Political Science (Grant et al 2010), Media and Communication (Boyd et al 2010), Marketing (Burton and Soboleva 2011), Health Care (Hawn 2009), and, naturally, Computer Science (Kouloumpis et al 2011). In many cases, research on social data is interdisciplinary. This constantly raising interest is certainly expected, since social network data are easy to access and reflect multiple aspects of human behaviour and community dynamics. Probably the most well studied social network, is the Twitter micro-blogging platform.

Address(es) of author(s) should be given

From a data-science perspective, new mining tasks have recently appeared in micro-blog environments presenting interesting research challenges as well as commercial value. Sentiment Analysis (Kouloumpis et al 2011), Event Recognition (Valkanas and Gunopulos 2013), Trend Identification (Parker et al 2013), Community Recognition (Qi et al 2014), Influence Propagation (Gupta et al 2013) are just a few characteristic examples.

Tagging thrives in Internet platforms with user-submitted content where tags are voluntarily assigned for information retrieval purposes: Users can do tag-based searches or browse objects of a particular tag. Tags are currently utilized in many different types of content such as, images (Flickr), videos (YouTube) and music (Last.fm). Twitter is a tag-rich service. Users annotate their posts by inserting keywords marked with the hash (`#`) character.

Hashtags in Twitter are considered very important keywords since they add valuable meta-knowledge to a particular piece of text that is by nature limited to 140 characters. In order to track certain events and to annotate them properly users indirectly agree to hashtag Tweets with a predefined keyword (e.g. `#eqnz` - the hashtag citizens of New Zealand used to annotate content related to earthquakes - see Figure 1). Many micro-blog analysis tasks, like the ones mentioned in the previous paragraph, are exploiting tagging behaviour in multiple ways. Hence, hashtag quality plays an important role not only to information organization within Twitter but also to the efficiency of the state-of-the art tools for social network analysis.

Unfortunately, in social media users use hashtags not only to annotate specific events and topics but also to promote certain ideas or discussions known as *internet memes* (Bauckhage 2011). Many times memes

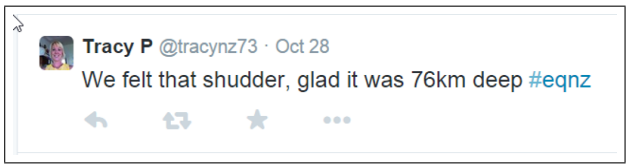


Fig. 1 A Tweet that utilizes hashtags to annotate content

Fig. 2 Hashtags used to promote celebrities.

Fig. 3 A hashtag used to promote a discussion.

arise when a group of individuals, e.g. celebrity fans try to promote a topic related to their pop idol (Figure 2). Other types of memes include internet hoaxes, marketing material or just spreading discussions and ideas (Figure 3).

Motivation. Memes are not inherently detrimental. However, since their data volume is many times significant, they can interfere with other tasks like trend analysis or event detection. In these cases memes are considered noise and must be filtered-out.

Event detection - A motivational example. Social networking platforms can benefit from discriminating between different types of trending or popular topics. For example by providing different landing pages or different advertising options for memes and events. Moreover, most event discovery approaches for social media are based on burst-detection mechanisms, assuming that a bursty behavior of a term or an n-gram may indicate that something important or extra-ordinary is happening in the world, and thus it is triggering popular discussions in social media. However, this is not always the case, as social media dynamics often lead to the creation of topics-of-interest that are internal to the network. As a result, most methods that attempt to discover events in social networking platforms do not take into account the fact that different types of trending topics - that is, topics whose popularity dramatically and unexpectedly increases - have different patterns of behavior in the network. The behavior of a topic can be characterized by a variety of factors, such as the community that is interested in the topic or the type of messages that are relevant to the topic and their characteristics, e.g. the number of hyperlinks to external sites, the number of attached pictures, the presence of hashtags, etc. In this paper, we attempt to disambiguate between the different kinds of trends and reason about what characterizes them.

Contribution. The contributions of this work can be summarized in the following points:

- We provide a definition of *meme* and *event* and discriminate between them in social networks, recognizing the fact that not all trends behave in the same way.
- A set of language-agnostic features to aid the classification of hashtags into *event* or *meme* is proposed. A variety of attributes and classification models are evaluated.
- We evaluate the proposed approach in terms of filtering accuracy using two large real world datasets of Twitter messages written in the English and German language.
- A proof-of-concept application of meme-filtering on the task of event detection demonstrates the utility of our approach.

The rest of this paper, is structured as follows: In Section 2 we review the related work, in Section 3 we define the problem and provide definitions of *memes* and *events* in the context of this paper and in Section 4 we describe our approach. Section 5 contains our experimental evaluation and Section 6 concludes the paper.

2 Related Work

In this section, we present recent and representative work that is related to the research challenges dealt with in this paper. More specifically, we discuss a) research efforts that study meme phenomena, b) papers tackling the problem of trend and event detection, and c) studies on hashtag analysis.

Memes. Bauckhage (2011) defines internet memes as "evolving content that rapidly gains popularity or notoriety on the Internet". Moreover, the author states that memes are spread voluntarily rather than in a compulsory manner, which although partially true, does not describe the full picture. Very often, memes are produced by advertising or community campaigns. In this scenario, they are expected to present different behavior in comparison to organically (i.e. not strategically)

created memes. For example, fans of groups or celebrities organize petitions in order to ask their idol to visit their country or say something about them. In these cases, the goal is to make a hashtag so popular that it appears in the *Trending Topics* list of the platform, affecting the initial goal of the list. The related bibliography lacks methods for recognizing these campaigns. In this context, our work offers the first approach towards this direction. Leskovec et al. define memes as "short, distinctive phrases that travel relatively intact through on-line text" (Leskovec et al 2009). They prove that information propagates from news sites to blogs. In their experiments there is an average lag of 2.5 hours between peaks of attention in news sites and blogs. However, with the spread of social networks and microblogging platforms, like Twitter and Tumblr, this claim has to be re-examined. Kamath et al. study the spatio-temporal properties of online memes, by specifically limiting their research to the propagation of hashtags across Twitter, arguing that hashtags may associate statuses with particular events or with memes and conversations (Kamath et al 2013).

Trends and Events. In (Yang and Leskovec 2011) the authors employ time series clustering in order to uncover temporal patterns in popular content and focus on the propagation of hashtags on Twitter. Leskovec et al (2009) claim that mainstream media accounts (CNN, BBC, etc) produce content and push it to other contributors, including Twitter's news-related accounts and professional bloggers. However, in (Petrovic et al 2013), Petrovic et al. use classes of content, i.e. *sports, politics, business, tv, etc.* to study the time aspect in Newswire and Twitter data and argue that Twitter covers most events that are mentioned by major news providers like CNN, BBC etc. Moreover, they suggest that it covers even smaller events that are not mentioned elsewhere. In their study they show that Twitter reports first sports events and unpredictable disaster-related events. In this sense, in the real time world, the highly credited news accounts are not always the ones that produce the *important content* first and the definition of what *important content* is still open. Thus, real-time event detection in Twitter and other social media has been a hot research topic in the last few years. In (Yang and Leskovec 2011) the authors use a time series shape similarity approach to find temporal patterns and form clusters. They show that media agency news present a very rapid rise followed by a relatively slow decay. In (Tsur and Rappoport 2012) the authors try to predict the popularity of a hashtag in a given time frame using linear regression.

Platakis et al. used a burstiness-based method to address the problem of identifying events in the Blogosphere (Platakis et al 2009). They applied Kleinberg's burstiness-detection automaton (Kleinberg 2003) on titles of blog posts to discover bursty terms and evaluated their method by correlating bursty topics with a real life event that took place in the same time period. Lappas et al. explored how burstiness information can be used to address event detection in the context of timestamped document collections and presented an approach to model the burstiness of a term, using discrepancy theory concepts (Lappas et al 2009).

Hashtag Analysis. Many approaches in social network mining research have been devoted to the analysis of the role of tags and hashtags in social networking platforms. The hash symbol (`#`) is used to indicate special meaning to a single or multiple words in social networks like *Twitter, Instagram* or *Facebook*. Apart from tagging, social network users utilize hashtags for various other reasons like creating viral conversations (*memes*). As opposed to traditional web search, queries in Twitter search that contain the hash symbol account for a significant portion of the total queries issued to the system (Teevan et al 2011). Thus, special effort has to be devoted to further utilizing hashtag-related to enhance the search process in social media. Moreover, many Twitter queries reference words used in hashtags, but without the preceding `#` in the query. Since the amount of possible hashtags a user can use to either tag content or search for results is essentially unlimited, both these tasks would benefit if users were aware of tags used by other users for the same or similar purposes (Sen et al 2006). Kamath et al. study the spatio-temporal dynamics of hashtags, proving that the spatial distance between locations affects the propagation of hashtags (Kamath and Caverlee 2013).

To conclude, even though there are many applications of hashtag analysis that might be affected by meme-hashtags, related work has not yet covered the problem of filtering them out and distinguishing them from event-hashtags. This is the gap filled in this work.

3 Memes and Events: Problem Definition

A proper formulation of the problem under study requires the definition of some basic concepts that in this setting can be rather abstract and ubiquitous. A clear distinction of the concepts "meme" and "event" in the context of social media is not an easy task. However, in the rest of this section, we underline their common elements and highlight their differences.

Definition 1 Social stream (s): *An infinite stream of content $c_i, c_{i+1}, \ldots$, where content c_i could be created or submitted by users of a single or multiple social networks.*

Definition 2 News stream (n): *An infinite stream of news items $n_i, n_{i+1}, \ldots$, where news item n_i can be generated from one or multiple online news sources (e.g. electronic editions of newspapers, magazines, etc).*

Based on the above elementary concepts we define the distinction between events and memes.

Definition 3 Memes and Events. *Events* and *memes motivate users to generate content (text, images, etc) and submit it to a social network. Hence, memes and events can be observed in s by identifying an excessive and bursty appearance of content related to them. The difference between an event and a meme is that an event can be traced back in the news stream n of the same period (as in s) whereas a meme only appears in s.*

Informally, events refer to something that happens in the real world, whereas a meme is initiated in the cyber-space. An event could be identified by observing messages and discussions in the social stream regarding the recent Elections in Germany, a soccer match between the teams Barcelona and Manchester United, an earthquake, or the Oscars ceremony. On the other hand, memes consist of messages related to a celebrity fan group requesting their idol to give a concert in their location, a discussion about why people cannot sleep at that time, etc.

For the rest of this paper we will refer to features computed over *relevant* documents or *relevant* authors. The definition of *relevance* in our setting is provided below.

Definition 4 Relevant document. *A document d_i is relevant to a topic t_j, if it contains the term describing t_j in the raw text or in its meta-information fields (e.g. the attached hashtags in the case of Twitter, the categories in the case of blog posts, etc). We define the document-relevance function as $rd(d_i, t_j) = 1$ if d_i is relevant to t_j, and $rd(d_i, t_j) = 0$ if not.*

Definition 5 Relevant author. *An author u_i is relevant to a topic t_k if $a(d_i) = u_j, i \in [1, n]$ and $rd(d_i, t_k) = 1$. We define the author-relevance function as $ra(d_i, t_k) = 1$ if a_i is relevant to t_k, and $ra(a_i, t_k) = 0$ if not.*

The Time Arrow. There might be a case that memes could also appear in the news stream. For example, if a meme appears in great volumes, reference in news sources (electronic and printed) will occur. For example, in *April, 2014* Twitter users were asked by the New York Police Department to tweet a photo of themselves with officers and use the hashtag `#myNYPD` as part of a social media campaign. However, several people used the hashtag to annotate tweets containing photos in more hostile situations than originally expected by the NYPD, resulting in the appearance of `#myNYPD` in the *Trending Topics* list of Twitter (Fig. 4).

Fig. 4 A social media campaign that resulted to a newsworthy event.

The distinction between memes and events in such cases can be spotted in the order of appearance in the two streams. A real-world event will first be identified in the news stream and then will have an effect on the social streams[1]. On the other hand, a virtual-world event will first appear on the social stream and then (in some extreme cases) might have an effect on the news stream.

Nevertheless, the distinction line is thin and this fact is reflected to the disagreement rate of our human annotators (see Section 5). This is the main intuition behind our language-agnostic approach: Since human expertise can not be represented in a set of rules, a set of social and document features are defined and a machine learning classifier is trained to identify the most informative ones.

Memes and Events Hashtags. In both cases, memes and events, as with any other content, are annotated with hashtags. For example, hashtags of the events described above could be: `#GermanyElections`, `#BarcaVsMancester`, `#earthquake`, `#Oscars2014`,

[1] Obviously, there are exception to this rule. In our times, information about earthquakes appear on social media first. However, this is still related to a real-world event.

whereas for the internet memes mentioned in the previous paragraph, example hashtags could be the following: `#loveit`, `#insomnia`, `#20ReasonsIAmCute`, `#WeWantJustInIreland`. As we can observe, there are not any structural characteristics that can aid in separating an event-hashtag from a meme-hashtag.

An interesting feature of meme and event hashtags is that they share a common time-series pattern. Both of them present a "bursty" element, meaning that they suddenly appear in great volume in a limited time window. This element is the main motivation of our work since memes, having similar time-footprint with real-world events, hinder a lot of hashtag-based mining algorithms.

In this work, we aim to automatically build a model that distinguishes between those two cases. We provided the above definitions and examples to the human annotators we employed in the experimental section. The annotators used the above guidelines for labeling hashtags as events or memes.

Problem Definition: *Given a limited part of the social stream $s_T \subset s$ (training set), build a model that can assign a label (`event, meme`) to a hashtag h, given a specific set of information (statistics) for this hashtag.*

This set of information, as we discuss later on, should be able to be calculated *incrementally* since this feature is crucial for data streaming environments. Statistical information for each hashtag can be represented as a vector of features $\mathbf{h_x}$. The requested model is actually a function $f(\mathbf{h_x}) \rightarrow \{event, meme\}$, where $\mathbf{h}_x = \{g_1(s_T, h_x), \ldots, g_n(s_T, h_x)\}$ and g_i $(i = 1, \ldots, n)$ are the functions that incrementally can calculate the features i for the hashtag h. Note that with h we represent the keyword expressing the hashtag whereas $\mathbf{h}$ is its feature representation and n is the number of features. As we discuss later on, we formulate this problem as data classification problem by training machine learning classifiers to identify the features that separate the two classes.

4 Meme Filtering

In the previous sections we presented the motivation of meme filtering as well ass the intuition behind our approach. In this section, we go into detail description of the initial feature seed used as input for the machine learning classifiers.

4.1 Overview

We computed 15 features, resulting in 15-dimensional vectors representing the the hashtags in the set $\mathcal{T}$. Some of the features are Twitter specific (e.g. retweets or favorites), but can be extended to other social networks that support sharing or promoting functionality (e.g. *Share* or *Like* in *Facebook*). For many of the features we had an initial intuition regarding their information value when it comes to separating memes from event hashtags. However, as stated in the previous sections the idea is to let the classifier and the feature selection measures estimate the predictive value of each attribute.

For each hashtag $t_i \in \mathcal{T}$, we take its *hashtagLength* in characters into account. Communities aiming to promote a meme or an phrase representing an advertising campaign often try to collapse a whole phrase into a single word in order to save characters for their messages. In this sense, often memes are longer than event-representing hashtags, because people in the social network try to embed in them as much information as possible. For example, not many English words are as long as `#WeWantOneDirectionInLondon`, which happened to appear as a trend for London sometime in March, 2013. For each hashtag t_i we computed the following features:

4.1.1 Content features

The following features are computed over the set of all documents (messages) that were annotated with the hashtag t_i. We tried to capture the significance of rich content accompanying tweets, e.g. links, pictures, or videos and hashtag statistics.

- *tokensPerTweet*: The average count of distinct tokens per relevant tweet
- *hashTagsPerTweet*: The average count of distinct hashtags per relevant tweet
- *urlsPerTweet*: The average count of included hyperlinks to external sites per relevant tweet
- *mediasPerTweet*: The average count of attached media objects per relevant tweet. Media objects can be photos, videos, songs. As of *December, 2014* Twitter supports only photos and videos.
- *favoritesPerTweet*: The average count of favorites per relevant tweet. Twitter offers the functionality of *favoriting* a tweet. This action serves as a means to either expressing approval or bookmarking a tweet for future reference. In the Twitter language, *favorite* or *fav* is the equivalent to *Like* in Facebook.
- *retweetsPerTweet*: The average count of retweets per relevant tweet. Twitter offers the functionality of re-posting a tweet, in order to express agreement with it. *Retweet* serves as a means for dissemination of popular content.

4.1.2 Interaction features

With the interaction features we try to capture the importance of conversations about a topic in the social network. Twitter offers the ability to reply directly to a specific tweet or to mention other users (not necessarily friends or followers) inside a tweet, by adding the '@' sign before the name of the user to be mentioned. In our analysis we use the following two features to capture the interaction properties of the hashtags:

- *tweetsWithReplies*: The *tweetsWithReplies* feature reflects the percentage of relevant tweets to hashtag t_i that were replies to other tweets.
- *mentionsPerTweet*: The *mentionsPerTweet* represents the average number of mentions to other users over all tweets relevant to hashtag t_i.

4.1.3 Community features

Memes are expected to come from groups of connected users, whereas events interest a broader user base. This idea is encoded into the following features.

- *statusesPerUser*: The *statusesPerUser* feature represents the average number of total posted status updates from the set of unique users that posted a tweet relevant to hashtag t_i. This feature aims to capture the activity of the community that used the hashtag under consideration.
- *uniqueUsersCount*: This feature captures the size of the community that was interested in the corresponding hashtag.
- *userFollowersPerUser, userFriendsPerUser, listedCountPerUser*: These three features capture the popularity and the social activity of the users that appeared to be interested in a particular hashtag. They represent the avg. number of *Followers*, *Following* and *Lists* the users appeared in.
- *avgVerifiedUsers*: In order to have a measure of the credibility of the users interested in each hashtag, we utilize the *Verified* feature of the Twitter platform, and we compute the avg. number of users that are verified by Twitter.

5 Experiments

In this section we present the experimental evaluation including dataset description as well as the annotation process.

Table 1 Dataset Statistics

Description	UK	Germany
Unique Tweets	27,868,183	6,826,709
Tweets with at least one hashtag	4,432,052	950,739
Unique Users	721,644	237,344
Unique Hashtags	1,102,320	491,043
Hashtags appearing only once	806,160 (73.1%)	246,311 (66.6%)
Average occurrences per hashtag	6.1	7.47

5.1 Data Description

We collected data using the Twitter Streaming API using two different bounding boxes. Specifically, we collected tweets from the bounding box of the United Kingdom for the period between *February 16, 2014* and *April 6, 2014* and tweets from the bounding box of Germany for the period between *April 1, 2014* and *October 10, 2014*. The two datasets contain 27.8 and 6.8 million tweets respectively. We split the datasets into single day periods and computed features for the top-20 most popular hashtags for each day. Table 1 presents detailed information regarding the two datasets. An important point regarging the motivation towards the usage of hashtags for this work is that in both datasets the percentage of tweets that have at least one hashtag is quite high, reaching 16% in United Kingdom and 13.9% in Germany.

Figure 5 illustrates the distribution of the hashtags across the dataset. It is apparent that most hashtags are used only once, which indicate that (i) users utilize them for annotating content *but*, (ii) they are not aware of which hashtags are used by other people. In fact, Figure 5(right) shows that most hashtags in United Kingdom appear less than 80 times in the period of study, which spans 50 days. Figure 6 displays the wordcloud of the top-100 most popular hashtags in our dataset. It is apparent that even the set with the top-100 most popular hashtags contains both *memes* (e.g. `#georgesnapchatme`, `#100happydays`, etc.) and *events* (e.g. `#brits2014`, `#bbcqt` tagging tweets about the BRIT Awards 2014 and the 'BBC Question Time' television program respectively, etc.), as long as some hashtags belonging to more *general* categories, like weekday or location names.

5.2 Preprocessing

Before annotating the extracted hashtags, we performed some preprocessing steps. Specifically:

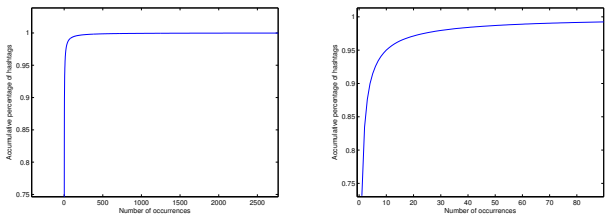


Fig. 5 Cumulative distribution function of unique hashtags over number of occurrences (total and zoomed-in)

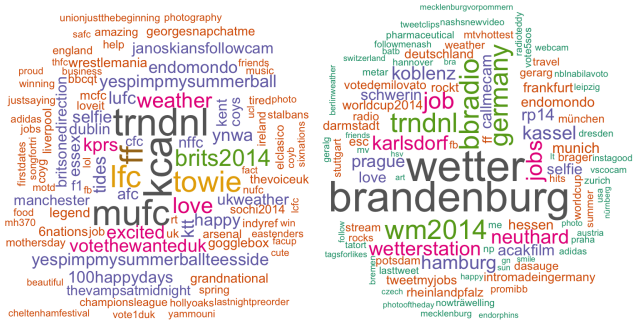


Fig. 6 The wordclouds of top-100 most popular tags in *United Kingdom* and *Germany*

- We lowercased all hashtags, in order to collapse to one entity hashtags representing the same thing but written in different ways, e.g. `#WeWantJustinInIreland` and `#wewantjustininireland`
- We filtered out location names, e.g. `#London`, `#Dublin`, `#Berlin`, `#Frankfurt` etc. In the case of significant events, e.g. an earthquake, the event would show up in other popular hashtags too. In all other cases location hashtags are vague with respect to whether they represent a meme or an event. Table 2 lists the occurrence counts for the most popular hashtags in our datasets. In comparison, as shown in Table 1, the average number of occurrences per hashtag was 6.1 in *United Kingdom* and 7.47 in *Germany*.
- We filtered out hashtags obtained from messages posted by automated systems like Spotify, Facebook, Instagram or bot accounts, e.g. `#nowplaying`, `#ukweather`, `#trdnl`, etc.
- We filtered out day and month names, e.g. `#friday`, `#sunday`, `#january` etc.

5.3 Annotation Process

After the initial preprocessing we asked independent people to manually tag all remaining preprocessed hashtags into one of the two classes, *meme* and *event*, while there wasn't any option not to tag an example. The annotators were not exposed to the feature vectors that corresponded to the hashtag examples, in order to avoid bias towards any of the classes. The number of the independent annotators was 5 for the *United Kingdom* dataset and 3 for the *Germany* dataset. We used majority voting in order to specify the class of each hashtag. In the *United Kingdom* dataset we ended up having 1100 tagged examples and vectors, among of which 558 where tagged as *events* and 542 as *memes*. In the *Germany* dataset we ended up having 800 tagged examples and vectors, among of which 358 where tagged as *events* and 442 as *memes*. The agreement among the two sets of annotators for the *United Kingdom* and *Germany* datasets is illustrated in Tables 3 and 4 respectively.

Table 2 Occurrence Counts for very popular hashtags

hashtag	Dataset	Count
`#nowplaying`	Germany	96,261
`#ger`	Germany	64,430
`#berlin`	Germany	51,561
`#nowplaying`	United Kingdom	43,478
`#london`	United Kingdom	35,837

In order to quantitatively measure the annotators' agreement in the labeling process we used the Fleiss kappa statistic. Fleiss kappa is an extension of the well known Cohen's kappa, which is a measure of the agreement between two raters, where agreement due to chance is factored out. In our case, for both datasets the number of raters was more than two, so Fleiss' extension was used. For the *Germany* dataset the annotators agreed with $\kappa = 0.301$, whereas for the *United Kingdom* dataset the corresponding value was $\kappa = 0.202$.

Whereas in the case of the *Germany* dataset the *kappa* value constitutes for a fair agreement among the annotators, when the number of the annotators increases, as is the case with the *United Kingdom* dataset, *kappa* decreases. This can be attributed to the following fact: Since Twitter users in United Kingdom are more active, there is a larger diversity of the top hashtags, which makes it more difficult for the annotators to reason on whether a hashtag represents an actual event or a social network generated meme. On the other hand, in Germany, the distinction of the two classes is clearer, since Twitter users in this area tend to post updates about a more narrow variety of topics, including mainly football- and celebrity-related tweets. Thus, most hashtags that belong to the former category are characterized by the annotators as *events*, since a real football match is held, shortly before or shortly after the time of the posting. Hashtags that belong to the

Table 3 Annotator Agreement for the *United Kingdom* dataset

Majority	Meme	Event
3 to 2	29.5%	44%
4 to 1	65.4%	51.6%
5 to 0	5.1%	4.4%

Table 4 Annotator Agreement for the *Germany* dataset

Majority	Meme	Event
2 to 1	36.5%	68.5%
3 to 0	63.5%	31.5%

latter category are annotated as *memes*, since most of the time nothing important has happened concerning the respective celebrity.

Discussion. In Germany most events are related to soccer. On the other hand, most memes are about celebrity and television. Memes thrive during the weekends. In both datasets, the most popular hashtags included tags obtained from automated messages for weather, running and music playing + locations (berlin, frankfurt, london). These hashtags were exluded from the annotation and testing process, since they don't contain important information about significant events or memes. Moreover, in the *Germany* dataset, we observed significantly less diversity in the usage of hashtags, which can be seen in Figure 7. More specifically, we computed

In our own evaluation, we computed the Jaccard similarity of the top-20 hashtags between consecutive days for a whole month in the two datasets and compare the resulting values. The Jaccard similarity between two sets is computed as the quotient of the overlap and the union of their respective vocabularies. Formally, given the set of top hashtags th_i corresponding to day i and the set of hashtags th_{i+1} corresponding to day $i+1$:

$$Jaccard(th_i, th_{i+1}) = \frac{|th_i \cap th_{i+1}|}{|th_1 \cup th_{i+1}|} \quad (1)$$

As Fig. 7 illustrates, the Jaccard coefficient of the top-20 hashtags in Germany is quite hiqh when comparing the sets day over day during a whole month, which indicates that the most frequently used hashtags are more or less the same every weekday with the exception of football and Champions League days. As a result, the average Jaccard Coefficient has a value of 0.48. In comparison, in United Kingdom, the top hashtags change quite significantly as time goes by, resulting to an average Jaccard Coefficient of 0.21.

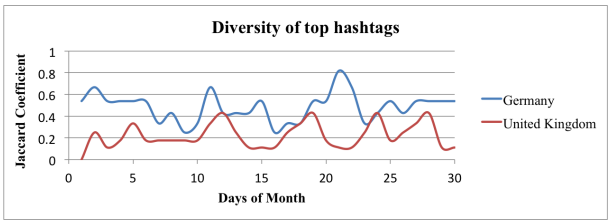


Fig. 7 Jaccard Coefficient of top-20 hashtags as they appeared in *United Kingdom* and *Germany* during *March, 2014* and *May, 2014* respectively

Table 5 Confusion Matrix of the four classifiers for the *United Kingdom* dataset (M=Meme, E=Event)

	Predicted							
	N Bayes		**R Forest**		**SVM**		**k-NN**	
Actual	M	E	M	E	M	E	M	E
M	.64	.36	.91	.09	.73	.27	.87	.13
E	.08	.92	.11	.89	.13	.87	.14	.86

Table 6 Confusion Matrix of the four classifiers for the *Germany* dataset (M=Meme, E=Event)

	Predicted							
	N Bayes		**R Forest**		**SVM**		**k-NN**	
Actual	M	E	M	E	M	E	M	E
M	.77	.23	.90	.10	.79	.21	.85	.15
E	.11	.89	.13	.87	.13	.87	.13	.87

5.4 Classifiers - Results

We experimented with four classifiers representatives of four different learning paradigms: Naive Bayes (Probabilistic Learning), Random Forest (Ensemble Learning), Support Vector Machines and the k-Nearest Neighbor classifier (Lazy Learning) (Aha et al 1991). Implementation of the above algorithms in the Weka platform (Hall et al 2009) were utilized.

The Random Forest classifier was able to reach an accuracy of 89.2%, with an average precision and recall of 89.2%. The confusion matrices of all classifiers for the United Kingdom and Germany datasets are illustrated in Tables 5 and 6 respectively. Figure 8 illustrates how the four classifiers compare against each other in terms of accuracy as a function of the size of the training set. Random Forest classifier has been more accurate than the other classifiers for all values of the size of the training set. This result confirms recent research on the accuracy of the Random Forest algorithm (Fernández-Delgado et al 2014).

Figure 9 compares the achieved accuracy of the four classifiers when using a 10-fold cross-validation scheme. Again, Random Forest outperforms Naive Bayes, k-NN and SVM, reaching an accuracy of 89.66%.

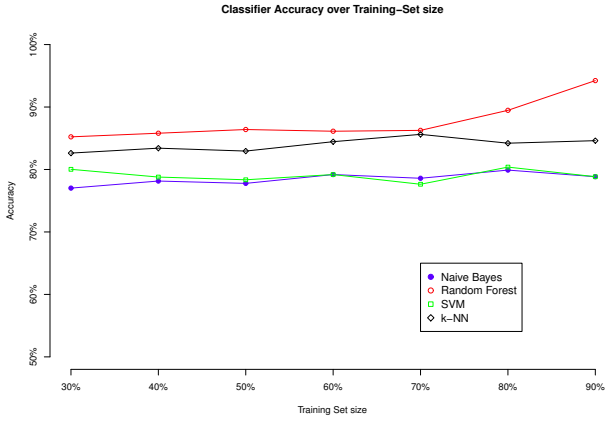


Fig. 8 Accuracy of Naive Bayes, Random Forest, SVM and k-NN classifiers as a function of training set size

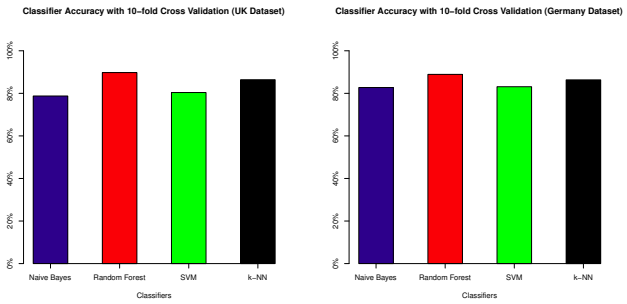


Fig. 9 Accuracy of Naive Bayes, Random Forest, SVM and k-NN classifiers with 10-fold cross-validation for the two datasets

5.5 Feature Evaluation

In order to argue about which features are the most important for the classification of hashtags we ranked them in decreasing Gain Ratio with respect to the two classes. Table 7 lists the features according to this ranking. We then repeated the classification process with the four classifiers, starting with the first feature in Table 7 (in the United Kingdom part) and incrementally adding the remaining features one by one, in order to inspect the benefits in classification accuracy. Figure 10 depicts the results of this experiment. Here again, the Random Forest classifier outperforms all others for all feature subsets. Interestingly, when we used only the *community* features, the Random Forest classifier was able to reach an accuracy of 70.8%, while when we used only the *document* features the classifier reached an accuracy of 86%. The behavior was the same for the Germany dataset, so we only show the United Kingdom results here.

In order to further investigate relationships between individual features that serve as indicators and the

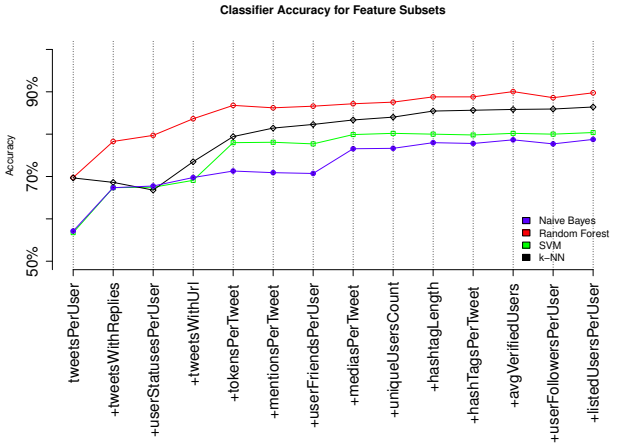


Fig. 10 Accuracy of the four classifiers with different feature subsets, incrementally adding the next feature w.r.t to Gain Ratio

hashtag classes we study the Figures 11, 12 and 13. By looking carefully at Figure 11(a) we can identify a tendency indicating *events* are being discussed by more unique and less active users than *memes*, while Figure 12(a) shows that that users who follow few others but have many followers themselves, tend to mostly write about events. On the other hand, in Figure 13(b), it is apparent that the average number of the unique relevant users to hashtags classified as *memes* is not large, while these users tend to be very active in the network, having posted far more tweets than users writing about *events*. Similar conclusions can be derived from Figure 11(b), where users posting content about *events* tend to be included in considerably more lists than users promoting or contributing to *memes*.

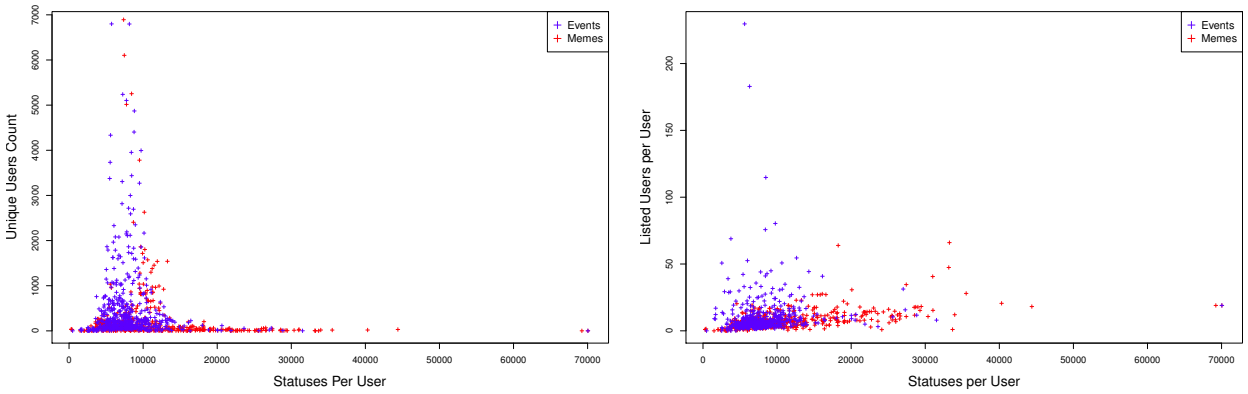
Interestingly enough, Figure 13(a) reveals a rather odd observation. While tweets about breaking and significant events were expected to contain a relatively high number of URLs linking to external sites with the source of the information, this appears not to be true. In the Figure, there is a clear separation of the spaces covered by *memes*-examples and *events*-examples, showing that *memes* are represented by tweets with fewer tokens - as described above - and more URLs, whereas *event*-related tweets contain on average and on aggregate much fewer URLs and more tokens.

Figures 14 and 15 reveal a number of interesting observations:

– Tweets that contribute to propagation or promotion of *memes* have significantly more videos, photos or hashtags attached to them than tweets discussing real-life *events*. Memes often are parts of campaigns or internet petitions and users try to enrich the content they generate so it ranks higher in search re-

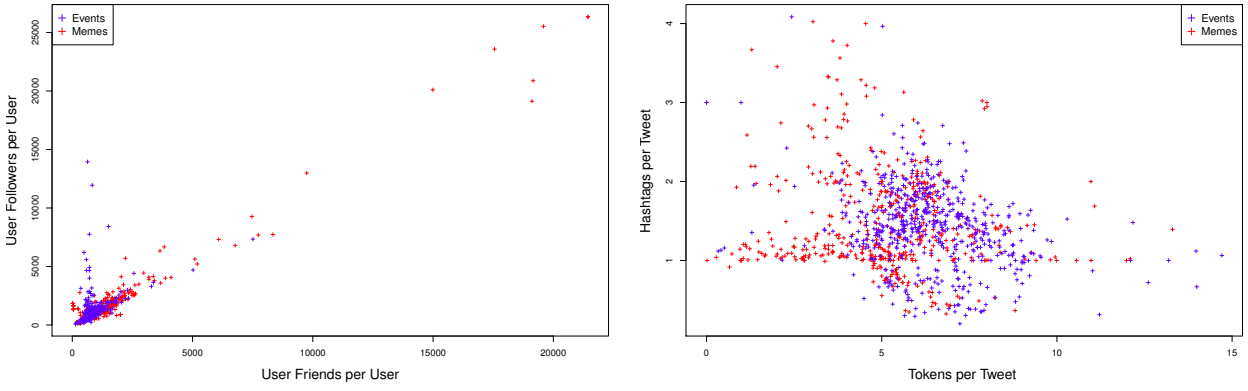
Table 7 Decreasing Gain Ratio Feature Ranking for *United Kingdom* and *Germany* datasets

United Kingdom		**Germany**	
Feature	**Gain Ratio**	**Feature**	**Gain Ratio**
tweetsPerUser	0.1617	mentionsPerTweet	0.1764
tweetsWithReplies	0.1432	tweetsPerUser	0.1566
userStatusesPerUser	0.1181	avgVerifiedUsers	0.1547
tweetsWithUrl	0.0958	tweetsWithReplies	0.1451
urlsPerTweet	0.091	hashTagsPerTweet	0.1422
tokensPerTweet	0.0822	tweetsWithUrl	0.1416
mentionsPerTweet	0.0822	urlsPerTweet	0.1416
userFriendsPerUser	0.0802	listedUsersPerUser	0.1383
mediasPerTweet	0.0778	uniqueUsersCount	0.1294
uniqueUsersCount	0.0574	mediasPerTweet	0.1244
hashtagLength	0.0572	tokensPerTweet	0.1127
hashTagsPerTweet	0.0527	userFriendsPerUser	0.0959
avgVerifiedUsers	0.0461	userFollowersPerUser	0.0883
userFollowersPerUser	0.0355	userStatusesPerUser	0.0728



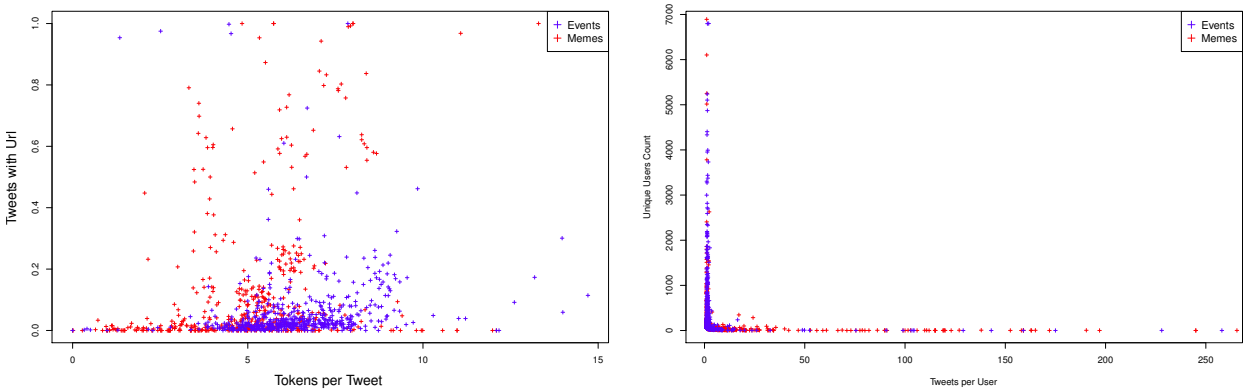
(a) (b)

Fig. 11 Feature Correlations



(a) (b)

Fig. 12 Feature Correlations



(a) (b)

Fig. 13 Feature Correlations

sults, either for a specific hashtag or for a relevant topic. Having more hashtags in the tweet increases the chances of it becoming viral or including a hashtag other users search for.

– Tweets that are relevant to *memes* draw more conversations in the social network than tweets that report a real-life *event*. This is to be expected, since, as described above, the number of unique users who are interested in memes is relatively small and thus communities with similar meme-oriented interests are more easily formed. Such communities consist of people interested in celebrities, jokes, etc.
– Tweets discussing *events* have on average slightly more tokens. This is normal, since these kind of tweets have a less arbitrary structure as they often include quotes or headlines in order to reproduce news reports, thus more words are needed to express something news-worthy.

6 HashTag-Based Event Detection: A Proof of Concept Use Case of Meme-Filtering

In order to show the utility of the proposed methodology we applied a hashtag-based event detection approach to our data. As mentioned in the introduction, due to the limited text length of tweets, hashtag analysis is a common approach for micro-blogs mining.

For example, most event detection methods in social media rely on time-series analysis of hashtags, inspecting terms that appear bursty for specific time-periods or generally popular terms. The assumption is that the bursty keywords/hashtags will be related to emerging events.

In this section, we argue that these event detection methods can lead to mixed results, since memes are also popular and bursty. In fact, memes appear to have a very well defined popularity period, just like events, so time-series approaches will fail distinguishing one from the other.

We applied a burstiness algorithm for event detection in order to study the insufficiency of this type of approach. At a high level, a time-frame is considered bursty if the term exhibits atypically high frequencies for its duration. Bursts in terms of frequency capture the trends in vocabulary usage during each corresponding time-frame and can thus prove useful in event detection. When an event takes place in real life (e.g. an earthquake, sports finals), the event's characteristic terms (e.g. *earthquake, shooting, overtime*) appear more frequently in social media. Unfortunately, memes demonstrate a similar behaviour.

6.1 Finding the bursty intervals

We use the `GetMax` algorithm, introduced in (Lappas et al 2009), as a representative burstiness approach. However, this experiment is compatible with any method that can identify bursty intervals given a sequence of frequency measurements. Given a discrete time series of frequency measurement for a given term, `GetMax` returns the set of non-overlapping bursty intervals, each associated with a score indicating how significant the burst is. In other words, how unexpected and dramatic the deviation from the baseline is. A brief description of the algorithm is provided below, where scores represent observed frequencies.

The `GetMax` algorithm computes a set of bursty intervals, after reading the time series consisting of frequency values for a term. A *burst* on the timeline is marked whenever the popularity of a specific

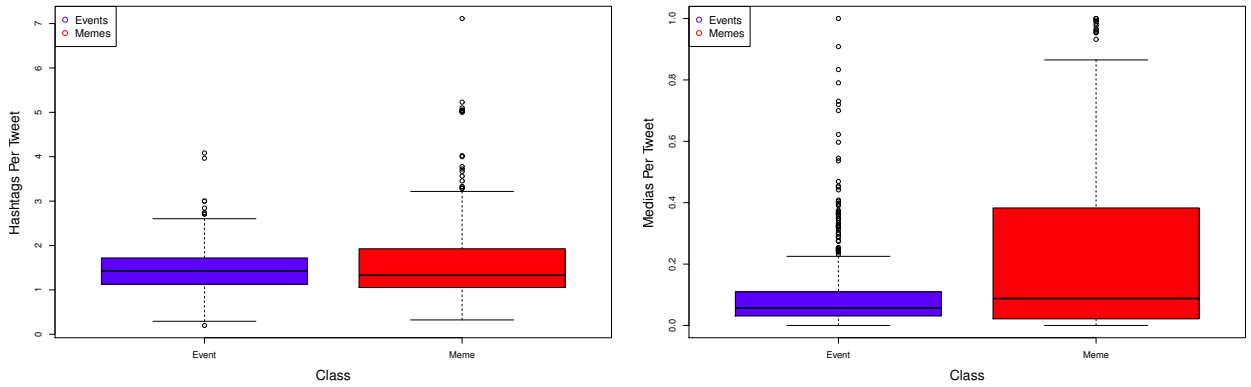


Fig. 14 Boxplots for various features

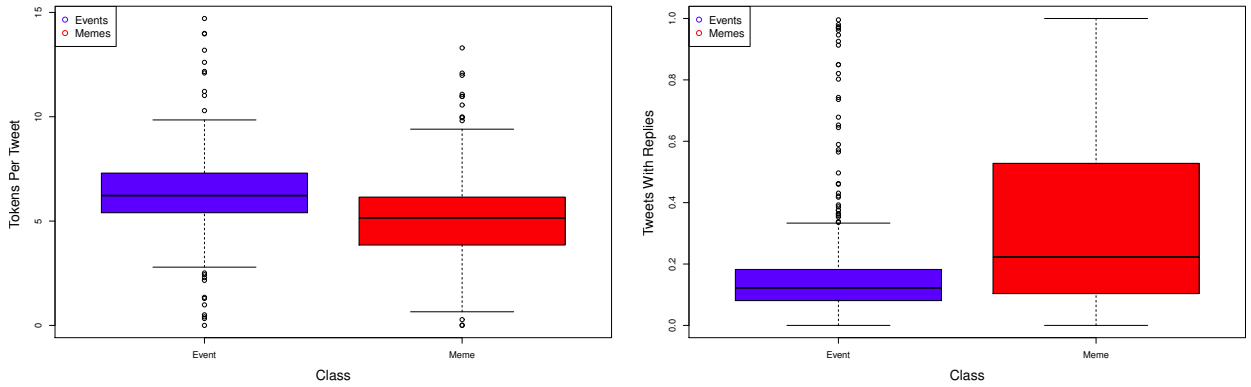


Fig. 15 Boxplots for various features

term dramatically and unexpectedly increases. Segments that are candidates for maximality, and thus candidate bursty intervals, are kept in a list L. For each candidate $l_j \in L$, we record the sum $l_j.L$ of all scores up to the leftmost score of l_j (exclusive) and the sum $l_j.R$ up to the rightmost score of l_j (inclusive). Non-positive scores require no special handling. If a positive score is read, a new sequence l_k containing only this score is created and processed as follows:

1. Search the list L, from right to left, for the maximum value of l_j satisfying $l_j.L < l_k.L$.
2. If there is no such l_j, then append l_k to the list L.
3. If there is such a l_j, and $l_j.R \geq l_k.R$, then append l_k to the list L.
4. Otherwise (i.e., there is such a l_j, but $l_j.R < l_k.R$), extend l_k up to the leftmost score in l_j (inclusive). Remove candidates $l_j, l_{j+1}, ..., l_{k-1}$ from L (none of them is maximal) and reconsider the newly extended segment l_k (now numbered l_j) from step 1.

After the entire input has been processed, the candidates left in the list L are the maximal segments representing the bursty intervals on the timeline (Lappas et al 2009; Ruzzo and Tompa 1999).

6.2 Burstiness Results

In our experiment we split the *Germany* dataset in two sets, one including months *April, June, July* and *August* which served as the training set and one including only *September* which was our testing set. Table 8 lists some bursty intervals computation examples along with a short description for the corresponding hashtags. The last column shows the classification result when using meme filtering with the Random Forest classifier, trained over labeled data from the first four months of the dataset. It is apparent that while the bursty intervals computed by `GetMax` algorithm precisely match the actual dates of excessive popularity of the corresponding hashtags, it is not enough to reason about significant

real life events that affected the Twitter community. Hence `GetMax` identifies memes and events.

A closer look at Fig. 16 reveals an even further similarity between the different types of popular hashtags in terms of behavior in time. On *Friday, September 19* four hashtags exhibit similarly bursty behavior, being simultaneously and unexpectedly popular. Two of them, namely `#iPhone6` and `#iphone6Plus`, correspond to the event of the release of the new iPhones, `#eaia2014` is the hashtag used to annotate discussions and reports from the European Association for International Education held in Prague, while the `#ff` is a viral Twitter meme with the aim of suggesting people for other users to follow. While the reader would argue that the distinction between an event and a meme in this case is rather trivial, since `#ff` is a periodically popular hashtag, this is not the case with hashtags like `#nominateavrillavigne` that have similarly bursty behavior, but only not periodic.

7 Conclusion

In this paper we defined the problem of distinguishing a popular topic of interest in a social network between network-generated topics of discussion, denoted as *memes* and real-life events that triggered the interest of the social network users, denoted as *events*. We provided a detailed study of the features that affect the classification, applying our experiments on the Twitter network using two different real-life datasets with 27.8 and 6.8 million tweets each and 1.1 million 491,043 unique hashtags respectively. We evaluated multiple classification methods, among of which the Random Forest classifier performed always best, having been able to reach an accuracy of 89% in its prediction on whether a topic is a *meme* or an *event*. Our study reveals interesting characteristics of the two hashtag categories. To demonstrate the utility of our approach we enhance a hash-tag based event detection with meme-filtering and comment on the improved results.

Acknowledgments

The authors would like to thank the data annotators. This work has been co-financed by EU and Greek National funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Programs: Heraclitus II fellowship, THALIS - GeomComp, THALIS - DISFER, ARISTEIA - MMD" and the EU funded project INSIGHT.

References

Aha DW, Kibler DF, Albert MK (1991) Instance-based learning algorithms. Machine Learning 6:37–66

Bauckhage C (2011) Insights into internet memes. In: ICWSM

Boyd D, Golder S, Lotan G (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, IEEE Computer Society, Washington, DC, USA, HICSS '10, pp 1–10, DOI 10.1109/HICSS.2010.412, URL http://dx.doi.org/10.1109/HICSS.2010.412

Burton S, Soboleva A (2011) Interactive or reactive? marketing with twitter. Journal of Consumer Marketing 28(7):491–499

Fernández-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research 15(1):3133–3181

Grant WJ, Moon B, Busby Grant J (2010) Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter. Australian Journal of Political Science 45(4):579–604

Gupta A, Sycara KP, Gordon GJ, Hefny A (2013) Exploring friend's influence in cultures in twitter. In: ASONAM, pp 584–591

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update. SIGKDD Explor Newsl 11(1):10–18

Hawn C (2009) Take two aspirin and tweet me in the morning: how twitter, facebook, and other social media are reshaping health care. Health affairs 28(2):361–368

Kamath KY, Caverlee J (2013) Spatio-temporal meme prediction: learning what hashtags will be popular where. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, ACM, pp 1341–1350

Kamath KY, Caverlee J, Lee K, Cheng Z (2013) Spatio-temporal dynamics of online memes: A study of geo-tagged tweets. In: Proceedings of the 22nd international conference on World Wide Web, International World Wide Web Conferences Steering Committee, pp 667–678

Kleinberg J (2003) Bursty and hierarchical structure in streams. Data Mining and Knowledge Discovery 7(4):373–397

Kouloumpis E, Wilson T, Moore J (2011) Twitter sentiment analysis: The good the bad and the omg! ICWSM 11:538–541

Lappas T, Arai B, Platakis M, Kotsakos D, Gunopulos D (2009) On burstiness-aware search for document sequences. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009, pp 477–486, DOI 10.1145/1557019.1557075, URL http://doi.acm.org/10.1145/1557019.1557075

Leskovec J, Backstrom L, Kleinberg J (2009) Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 497–506

Parker J, Wei Y, Yates A, Frieder O, Goharian N (2013) A framework for detecting public health trends with twitter. In: ASONAM, pp 556–563

Petrovic S, Osborne M, McCreadie R, Macdonald C, Ounis I, Shrimpton L (2013) Can twitter replace newswire for breaking news. In: Seventh International AAAI Confer-

Table 8 Bursty Intervals for popular hashtags in *Germany* during September, 2014

hashtag	Bursty Intervals	Description	Meme Filtering
#ff	Sep 5, 12, 19, 25	"Follow Friday" Twitter meme	meme
#eaie2014	Sep 16 - Sep 19	Conference held in Prague during Sep 16 - Sep 19	event
#jaykingslandto60k	Sep 11 - Sep 12	Bot account posting thousands of tweets	meme
#nominateavrillavigne	Sep 11, 15	Celebrity fan campaign	meme
#h96hsv	Sep 14	Soccer match: Hannover 96 vs. Hamburger SV	event
#iphone6	Sep 9, Sep 19	Announcement and release of iPhone 6	event
#iphone6plus	Sep 9-10, 19	Announcement and release of iPhone 6 Plus	event

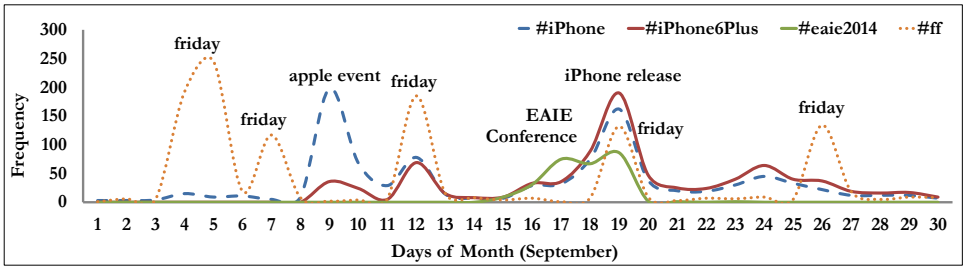


Fig. 16 Frequency curves of popular hashtags of various kinds as they appeared in *Germany* during *September, 2014*

ence on Weblogs and Social Media

Platakis M, Kotsakos D, Gunopulos D (2009) Searching for events in the blogosphere. In: Proceedings of the 18th international conference on World wide web, ACM, pp 1225–1226

Qi X, Tang W, Wu Y, Guo G, Fuller E, Zhang CQ (2014) Optimal local community detection in social networks based on density drop of subgraphs. Pattern Recognition Letters 36:46–53

Quercia D, Kosinski M, Stillwell D, Crowcroft J (2011) Our twitter profiles, our selves: Predicting personality with twitter. In: Privacy, security, risk and trust (passat), 2011 IEE third international conference on social computing (socialcom), pp 180–185

Ruzzo WL, Tompa M (1999) A linear time algorithm for finding all maximal scoring subsequences. In: ISMB, vol 99, pp 234–241

Sen S, Lam SK, Rashid AM, Cosley D, Frankowski D, Osterhouse J, Harper FM, Riedl J (2006) Tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, ACM, pp 181–190

Teevan J, Ramage D, Morris MR (2011) # twittersearch: a comparison of microblog search and web search. In: Proceedings of the fourth ACM international conference on Web search and data mining, ACM, pp 35–44

Tsur O, Rappoport A (2012) What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In: Proceedings of the fifth ACM international conference on Web search and data mining, ACM, pp 643–652

Valkanas G, Gunopulos D (2013) How the live web feels about events. In: He Q, Iyengar A, Nejdl W, Pei J, Rastogi R (eds) CIKM, ACM, pp 639–648

Yang J, Leskovec J (2011) Patterns of temporal variation in online media. In: Proceedings of the fourth ACM international conference on Web search and data mining, ACM, pp 177–186