Emotion-based topic impact on social media

Fernando H. Calderon Alvarado*

Social Networks and Human Centered Computing Program,
International Graduate Program,
Institute of Information Science,
Academia Sinica, 128, Academia Road,
Sec. 2. Nankang, Taipei, 115, Taiwan
Email: fhcalderon87@gmail.com
*Corresponding author

Yi-Shin Chen

Institute of Information Systems and Applications,
Department of Computer Science,
National Tsing Hua University,
Hsinchu, 30010, Taiwan
Email: yishin@gmail.com

Abstract: The increasing use of micro-blogging sites have made them very rich data repositories. Information generated is dynamic by nature, tied to temporal conditions and the subjectivity of its users. Everyday life experiences, discussions or events have a direct impact on the behaviours reflected in social networks. It has become important to asses to which degree are these interactions affecting a social group. A possibility is to analyse how impactful a topic is according to the behaviour presented on a social network over time. It is then necessary to develop methods that can contribute towards this task. Having identified a topic in social media, we can obtain a general summary of the emotions it is generating over a social network. We then propose a topic impact score which will be given to each topic based on how this emotions transition, for how many time they span and how many users they reach. This lays ground to quantify how impactful a topic is over a social group, specifically regarding events detected on twitter.

Keywords: social impact; influence; social media; emotion analysis; microblogs.


Biographical notes: Fernando H. Calderon Alvarado is currently a PhD student in the Taiwan International Graduate Program of Social Networks and Human Centered Computing provided by National Tsing Hua University and Academia Sinica. His research interests include social behaviours reflected on social media such as emotions, influence, context and impact of events.

Yi-Shin Chen is currently an Associate Professor in the Department of Computer Science at National Tsing Hua University. She is passionate about increasing the benefits of society through her research efforts. Due to her fear of a media monopoly, she has focused her research efforts on web intelligence and integration. Her goal is to create a social media interface that explores and visualises web data easily.

This paper is a revised and expanded version of a paper entitled ‘Influence value: quantifying topic influence on social media’, presented at The Second IEEE International Conference on Multimedia Big Data, Taipei, 20–22 April 2016.

1 Introduction

Recent years have witnessed an expansion in traffic and with it, content in social media sites. Amongst them, we have specific micro-blog sites, the most popular being ‘twitter’. Everyday more and more people turn to this kind of content instead of traditional media formats. Information shared and available on these sites have made them come to be considered rich data repositories. The networks that result from the implicit social process are themselves effective platforms for information diffusion and influence (Alvanaki et al., 2011). At a first glimpse micro-blog data in its most raw form could appear very random, unstructured and unrelated. Nevertheless, data scientists are conscious
about the importance and valuable content it can bear; therefore, they are in a constant attempt of ordering this data in ways that can facilitate its usage for research and analysis. These efforts have led to diverse fields of interest being developed, amongst them event identification and emotion detection; and other trends on how to better handle, manipulate and analyse data such as the popular big data realm or several approaches for data visualisation. These techniques can be integrated and open doors to further studies for example that concerning the topic impact mentioned previously.

Social network users usually turn to these platforms on a daily basis, this generates a dynamic automatic information update. These continuous streams of data have opened a door for a field of research defined as event identification. Event identification is a domain of study which consists of identifying events contained in a dataset, based on different features found within micro-blog posts; for practical purposes and specifically related to work dealing with twitter data, event is defined as “a real-world occurrence happening somewhere, on a specific time and followed by a stream of time-ordered twitter messages” (Becker et al., 2011). Events occurring under one same domain or category and with similar temporal characteristics generate a topic. The applications for these studies are of great value, information available on social streams can often reflect real-world events as they happen, sometimes even ahead of the news-wire. This has led to an expansion of micro-blogs as a preferred news source. When viewed in retrospective, it can provide insights on multiple circumstances and conditions that led to an event, all through the proper analysis of user posts. Several methods have been developed and found successful performance regarding this task. Having said this, the next concern is the usage of these events and topics combined with other tools or the development of new ones to enable a progressive extraction of more valuable information.

An alternative is to perform emotion detection analysis on the obtained topics. Emotion detection is an area of study in which based on a multiple range of features we can extract the emotion portrayed by a content, in this case text. Understanding how people feel about something can be very empowering information, it can help understand reactions to a new product launch, public opinion towards new government policies or politicians, reflect the worldwide feeling towards and international conflict or determine necessities of different social groups. The combination of these two techniques can provide us with a scope of how people feel regarding the occurrence of an event. If studied overtime, a broader view of how emotions toward a topic evolve can be obtained. This information is of great value for it can help us identify how changes or actions along time are affecting the public perception towards a subject.

When talking about impact of a topic we refer to the influence an event or occurrence has over a social group, under the assumption that this is quantifiable and measurable in terms of how long was the topic mentioned in social media, how much traffic it generated, what distribution it followed over time and other features. The main motivation for this type of analysis is that if we understand how events that already occurred behaved, taking some variables as reference we could later attempt to predict this impact based on initial behaviours. The applications of this would be many, for example governments having quicker and better responses to social manifestations as in the recent sunflower and umbrella movements. If they were able to predict the crowds behaviour based on their initial reactions, they could perhaps try to find a quicker solution, and not expose themselves or the people to an issue that lasted weeks. It could also involve commercial applications, when launching an advertising campaign, based on the initial response a company could determine to continue through a path or have to come up with a plan b in order to not lose momentum. Media could also be benefited by knowing how impactful a topic will be they can determine the amount of coverage they are willing to give or in another case know when a topic will grow old and therefore anticipate when to stop giving it coverage. To achieve this, we propose three measurements which combined will provide us an impact value:

- Lifespan: to determine if the topic is time specific or long-term, how long did it actually last.
- Emotion transition: illustrates did the emotions towards the subject evolve and change over time.
- Reach: quantify how many different users got involved in the discussion.

Each of which will be combined with the historical context of events withing their same domain.

In summary, we propose a system that can automatically find emotions toward a topic detected on twitter. Then, considering the variables mentioned above obtain a score of how impactful a topic is from a social point of view. In the process identify which kind of events are better at generating influence and information diffusion as well as which emotions are more effective for this tasks.

2 Related work

For some years now, event identification has been a popular area of study in the information extraction domain. Several works tried to observe topic trends in historical news and journal data with different approaches mainly statistics based. Blei and Lafferty (2006) used topic models on journal collections to infer latent topics based on state space models and multinomial distributions. The work by Kumaran and Allan (2004) combines text classification techniques and named identities to develop a vector space-based event detection system focusing on news stories. Then, gradually other forms of written expression began to grow in popularity and parallel to internet growth different content and data volume leaving these methods obsolete but opening doors to new needs and opportunities.
Social networks then became a preferred platform for users to reflect their attention towards events. With this in mind, several studies (Weng and Lee, 2011; Valkanas and Gunopulos, 2013; Macdonald et al., 2013; Zhou and Chen, 2014) empirically demonstrated that the correct manipulation and mining of a stream of tweets can lead to the identification of real-world events. Engineers at HP laboratories represented words as wavelet signals to later cluster similar words in order to detect events (Weng and Lee, 2011). The work by Valkanas and Gunopulos (2013) combines spatio-temporal information with emotional theories to identify events from a live stream. To provide a comparison with mainstream media, Macdonald et al. (2013) found that twitter provides a better coverage when dealing with minor events as compared to traditional newswires. Graph approaches have also proven a successful alternative to the task with the location-time constrained topic model (LTT) proposed by Zhou and Chen (2014). Other studies have tried to take advantage of other characteristics of social media, for example, temporal features. Sliding window approach proved to be effective for the detection of bursty topics by providing an analysis over time (Alvanaki et al., 2011; Diao et al., 2012). Becker et al. (2011) considered other features such as information sharing to distinguish messages as real-world events or non-event messages. Geographic location can also play a role in event identification as experimented by Hong et al. (2012). Given the nature of the streams of data, Gupta et al. (2014) introduce outliers detection in temporal data to focus the data on the specific days an event was occurring or a topic was trending.

Pak and Paroubek (2010) demonstrated twitter can be an effective corpus for sentiment analysis and opinion mining. Liu (2012) highlighted the importance of sentiment analysis in several research communities and how social networks could provide a very interesting data stream for the task. As interest and techniques evolved the depth of the analysis performed grew with them in order to evolve into a more detailed emotion detection. Considering Ekman’s emotions (anger, joy, fear, surprise, disgust, sadness) plus love, Roberts et al. (2012) used a supervised approach as SVM classification, combined with online dictionaries such as WordNet to classify micro-blog posts into those categories. This being a supervised approach makes it dependent on big training datasets and requires external resources such as human knowledge, dictionaries or ontologies, these could limit it expansion towards other applications or languages. In order to avoid this dependence, an approach is to extract patterns as features to identify opinions data. Patterns can be automatically extracted which releases a system from any dependency on external resources. By understanding semantic categories, it has been found co-occurring words can be more accurately related to a sentiment (Choi and Cardie, 2008). Additionally, psychological studies propose that emotional patterns can be found in a person’s utterances and texts. Recent discussions are considering the importance or affective computing within the artificial intelligence development and its possible applications (Cambria, 2016). It is our belief that being able to quantify emotional reactions to real world occurring events is one of the important applications.

Social networking has opened way to a series consideration which did not exist until their appearance (Bruns and Burgess, 2012). Amongst these researchers began to question matters of how individuals in a network influence each other (Bentley et al., 2011). In a similar way, topics discussed on a network can have an impact on its individuals. Acemoglu et al. (2013) studied how opinions fluctuate based on the dynamic nature of information in social networks. Other studies (Ye and Wu, 2013) consider to measure influence by assigning value to followers, replies or retweets. Kim et al. (2012) begin to relate emotion influence and transitions using semi-supervised machine learning methods on unannotated data. Recent works have defined a variety of features which are fed to a neural network to attempt to predict user influence (Mei et al., 2015; Zhang et al., 2016a). The evidence found regarding social influence led to the question “who creates trends in social media?”, the work by Zhang et al. (2016b) seeks to determine if its the crowd or some relevant influential leaders.

Previous work only considers user to user influence. This work attempts to integrate new methods on the domains of event identification and emotion detection to provide an automated approach to score topic impact over a social group based on the analysis of their emotional reactions over a period of time. To our knowledge, these techniques have not been combined with this purpose on previous works.

3 Methodology
3.1 Overview

The process begins with the collection of the required datasets all of which are obtained from twitter. The topic identification will require a continuous stream of a couple of months at least to identify multiple occurrences within that time range. Our emotion detection stage will require three additional datasets which characteristics will be later mentioned. Each of these needs to undergo a pre-processing stage to fit their specific requirements. Once the data is ready it will first undergo the process of event identification which will return multiple sets of keywords, each set related to a topic detected. Once we have our list of events we will proceed to filter our corpus to obtain tweets corresponding only to this events, to achieve this MongoDB database is created with the crawled data then via its querying function, tweets containing the pertinent keywords can be filtered out for our usage. Resulting events and their content will then be separated into pertinent topics, since each will have different analysis requirements. Having obtained a reduced corpus for each event we will proceed to run our emotion detection system on each of the remaining posts. Outlier detection on temporal data will be user to further filter our events to the specific days it lasted.
Emotion-based topic impact on social media

Emotion distributions will be calculated for each day in question then altogether using a sliding window approach an emotion transition value will be obtained. To obtain our final topic impact value, we will require to determine the users involved in each discussion to obtain our reach. The result for each event will work as feedback to the system so variables can train themselves according to historic context. The complete framework of the system can be observed graphically in Figure 1.

3.2 Event identification

The main objective of this stage of our process is to identify an event set \( E \) which occurred during the time period \( t_i \) to \( t_j \) from the twitter dataset \( T_w = \{ T_{w_1}, \ldots, T_{w_n} \} \), where each \( T_{w_i} \) is a tweet posted between time period \( t_i \) and \( t_j \). In order to achieve this, our data must first fit into a specific format before being processed. A diagram for the framework of this specific process is presented in Figure 2.

3.2.1 Pre-processing

We first submit text to a part-of-speech tagger (POS tagging), this process tags each word to a corresponding category (noun, verb, adjective, etc.). The purpose of this is to enable the next step which would be lemmatisation. Lemmatisation groups together all the different inflected versions of a word so it can be analysed as a single item, for example conjugated verbs will be transformed into their base form. A filter is then applied to remove words that do not contain alphabet characters or consist uniquely of punctuation marks. Finally some characters specifically related to twitter such as ‘RT’, ‘@’ and ‘#’ are removed. Given the time consideration we will have, the content is then separated into files containing the tweets for a specific day and labelled accordingly.

3.2.2 Keyword extraction

One of the basic yet most important concepts to this method is keyword selection. Keywords are the best representation to summarise a tweet. A keyword extraction method is adapted so that for each tweet we can retrieve a set of keywords \( K_t \) that can best represent a tweet \( T_{w_t} \). Assuming our event is happening within a specific time period, as we have defined before, keywords reflection attention towards this event should present a noticeably high occurrence within this same time period. A basic approach could be to use frequency to determine keyword candidates. However, different kinds of events would have different standards, for example the count for international events would be much higher than that of a local community event, hence, it would be difficult to define a global parameter that could be applied to all conditions. Taking this into consideration, a sliding-window-based statistical approach is selected to extract different types of event candidates. The sliding window will enable to take into consideration the evolution of a keyword over time as compared to others and hence determine its quality. The keywords are then categorised based on their corresponding statistics. Let \( \text{Category} = \{ \text{category}_1, \ldots, \text{category}_n \} \) be a set of categories, each category consists of keywords that satisfy its corresponding statistic criterion. For example, let us assume a category named \( Q4 \) containing high frequency words where in the past sliding window the sum of occurrences is higher than 75% of other keywords. The proposed framework can design and extract different keyword categories such as this.

Figure 1 Framework overview (see online version for colours)
3.2.3 Event graph generation

After having selected the keywords from our twitter set, if we properly group them they can proceed to generate event candidates. The ideal and first approach to group them would be considering their co-occurrence, this is how often do the keywords appear together in a tweet. However, such a naive approach cannot be adapted in reality, since keywords are likely linked together by other keywords ultimately leading to a large set of keywords being identified as one single event. To address this issue, some works have opted to link these keywords via graphs. In these studies, keywords are considered vertices and edges between keyword pairs are constructed if the co-occurrence probability is above an experimentally defined threshold. To avoid terms in this keyword set that are linked to many other immediate keywords (which would make them not representative of an event), we use betweenness centrality to determine all pair of vertices that pass through that vertex. Edges with a high ranking are removed from the graphs for they are frequent but not representative, the remaining connected keywords are considered the summary of the event.

**Definition 1 (Event graph):** An event graph \( G(t) \) is a directed graph for the time frame \( t \), where the vertex set \( V \) represents keywords and the edge set \( E = \{e(w_i, w_j) \mid w_i, w_j \in K\} \) denotes the co-occurrence relationship between keyword pairs in the same tweet. Let \( \omega_{e(w_i, w_j)} \) denote the weight for directed edge \( e(w_i, w_j) \). \( \omega_{e(w_i, w_j)} = \frac{\omega_{e(w_j, w_i)} \omega_{e(w_i, w_j)}}{\omega_{e(w_j, w_i)}} \) for time frame \( t \). The weight of a directed edge represents the fraction of source words \( w_i \) that appear together with the target word \( w_j \).

For this graph-based approach, edges are undirected and unweighted. After constructing the edges how strong a connection is between to keyword pairs is not really considered during the summary graph generation process. However, we do believe this is an important value and should not be disregarded. To strengthen this feature, we take into account the edge weight and use a text-rank (Mihalcea and Tarau, 2004) based approach to cluster keywords. This approach uses the co-occurrence and frequency of keywords to generate the hierarchical concept graphs. Since the keywords in an event graph are all frequent but not representative, the remaining connected keywords are considered the summary of the event candidates. The vertex weights of the event graph \( G \) are assigned as indicated by Definition 2.

**Definition 2 (Vertex weights):** Let \( w(v_i) \) denote the vertex weights in the event graph \( G \).

\[
w(v_i) = (1-d) + d \sum_{w_j \in \text{in}(w_i)} \frac{\omega_{e(w_j, w_i)}}{\sum_{w_k \in \text{out}(w_j)} \omega_{e(w_j, w_k)}} w(v_j)
\]

where \( \text{in}(w_i) \) is the set of keywords that point to \( w_i \), \( \text{out}(w_j) \) be the set of keywords which \( w_j \) points to, and \( d \) is a damping factor between 0 and 1.

Were \( \text{in}(w_i) \) is the set of keywords that point to \( w_i \), \( \text{out}(w_j) \) the set of keywords which \( w_j \) points to and \( d \) is a damping factor between 0 and 1. The resulting clusters and weights together provide a better defined event candidate, automatically identified and given the nature of the techniques used, it does it at low computational cost. Once the generated candidates list is generated, the keywords defining it will be used to query and filter our twitter corpus.

### 3.3 Emotion detection

Taking advantage of NoSQL databases which allow storage of elements with different features, such as tweets, we then go back to our original unprocessed dataset and perform a query-filter task using the keyword-based events obtained from the previous step. These results in multiple subsets of tweets separated into its corresponding events. The next step is to process the resulting tweets for each event with the emotion classifier developed by Argueta and Chen (2014). The classification will provide up to eight distinct emotion labels which will be suitable for procedures later introduced.

### 3.4 Topic impact

The main goal of this work is to propose a metric that can help asses the impact a topic has over society. **Topic impact score** is introduced with this in mind, as a score that can help compare, rank or summarise this idea so it can serve many different purpose. Having an understandable measure of impact and the conditions met to achieve it will strengthen the management of information diffusion. Multiple sectors from traditional media, government entities and even individual users can leverage on this concept to try and make the ideas they share as impactful, contagious or reachable as can be. By focusing on the emotional side, it can help understand which behaviours or trends can favour the spread and sharing of information and concepts. This could help market analysis consultants in determining the focus of their next advertisement, or assist political advisors in planning the upcoming campaign. It can also have an application in how to improve the coverage of socially aware campaigns, first in determining the needs of different sectors second in finding the proper way to change the status through the analysis of their emotions and which
events can help change them. The proposed value can also lay grounds for further studies in influence prediction. With a better understanding of the topic in hand, additional case studies that can improve the algorithms and other statistical considerations, impact will not only be determined and quantified but can also be predicted based on the behaviour during the initial days an event is occurring. This could lead to a great development in areas as sensible as crisis management.

The topic impact score attempts to take into consideration some of the most relevant and unique features of twitter data in order to return a score that can work as a standard reference. Amongst these features, the most important are its timely characteristics, the subjective nature of its posts and the network structure that allows spread of information. These variables in turn become the major components of our analysis and altogether collaborate and add up to one final score. The degree of interaction, how are they quantified and handled will be detailed in the following subsections.

### 3.4.1 Lifespan

Both topics and emotions are intrinsically bound to time. Topics and their related events appear over timelines and span over defined periods of time. The correct interpretation of this time relation can provide valuable insights when it comes to social media analysis. Additionally, emotions are also dynamic, and in constant fluctuation, sometimes correlated to the occurrence of the mentioned events. Time holds an important role in the proposed analysis specifically on two tasks:

- Definition on the specific time period over which an event spans.
- Classification of topics as time-specific or long-term.

The first is required because bursty topics can not be dealt with the same way as those who imply discussions lasting longer periods of time. The statistics bound to each are very different and therefore the methods to deal with them must be independent. The second will relate more directly to our concept of lifespan. The intuition is an event that spread for a longer time period (as compared to those within its same category) will have a greater chance to gather more attention hence impact more people. An alternative for this task is to treat it an outlier detection problem in order to reduce the data to the actual size of the event and from there proceed. The details will be provided in the following paragraphs.

#### 3.4.1.1 Outlier detection for temporal data

An outlier is observation point that is distant from other observations, it can be due to the variability of the experiment or may indicate experimental error. In our scenario and data formats, the observation points would be the days and the content of posts collected for that day. The experimental error is a consequence of the naive collection of data related to events which just queried tweets containing the related keywords. But for example for a set ‘{north, korea, bomb, threaten, CNBC}’ the query could return a tweet ‘last nights party was the bomb’ which could be contained in or close to the time window but totally unrelated. It is therefore necessary to consider these days and the tweets contained as outliers and filter them out from our event sets. In general, it is a user-action sequence scenario for outlier detection, were changes, sequences or temporal patterns in data can be used to model outliers and time is the contextual variable.

In order to achieve this desired outcome, the first stage is to find a better metric of similarity between our tweets and the keywords. Popularity used vector representations together with Cosine or Jaccard similarity are not helpful in this scenario due to the variable lengths of our text. Other alternatives such as semantic similarity measures are also unusable since what is required is not to match meaning considering a variety of users can be referring to a topic without necessarily trying to repeat the same statement. For this reason, we resort to a text specific implementation of Dice’s coefficient (White, 2004). The algorithm seeks to comply with requirements which go in line with ours: a true reflection of lexical similarity and a robustness to changes of word order, additionally it can be language independence which can be useful for future ventures. The metric is described in Definition 3.

**Definition 3 (Dice’s similarity coefficient):** A string $S_i$ of size(# of characters) $n$ can be separated into a set of ordered by word adjacent character pairs $A_i = \{p_0, …, p_{n-1}\}$. Then given a pair of strings $S_1$, $S_2$ we can find the size of their intersection relative to the size of the original strings as:

$$similarity(S_1, S_2) = \frac{2 \times \text{pairs}(S_1) \cap \text{pairs}(S_2)}{\text{pairs}(S_1) + \text{pairs}(S_2)}$$

This approach penalises tweets with only one keyword appearing, and will favour those with multiple of the related keywords present in the text. An alternative to improve the performance is manually assigning a title to the detected topic and using this as a reference for comparison, or adding additional keywords more related to the specific event.

Once a similarity value is obtained for every tweet a distribution can be generated for each day as well as for the whole event. Tweets having the lowest scores can be filtered out to keep the more highly related. The resulting similarity is a value from 0 to 1, ten equidistant (0.1 each step) intervals are initially predefined to generate the distribution, once more information is gathered regarding each topic category it can become adaptable. The outcome will be a similarity distribution generated for each collected day. We can then resort to statistical distribution comparison methods to determine which days do not share the general distribution for the event. A graphical representation of this process is depicted by Figure 3, we can observe a different trend begins at day 4 this can be an indication of the start of the event since it means the word distribution in that day is more related to our dataset.
Together with volume counts this measures can help us determine which days are not really part of the core topic discussion.

A Z-test-based approach is then used to determine how similar (or not) the distributions are. A sliding window together with well noticed criterion is used to determine which point in the timeline defines the outlier segment. The distribution of the current time window is compared to accumulated of the previous days if the Z-statistic condition is not met it would imply they do not belong to the same set, in this case to the same event. In a timeline series data this implies the breaking point and generation or culmination of a topic. The outcome of this stage will allow us to reduce our set to the specific dates we need to focus on and open way for other considerations. Figure 4 provides a graphical representation of the outcome of this process.

Figure 3  Sample similarity distributions for initial 4 days in a set (see online version for colours)

Figure 4  Highlighted are the days with traffic actually related to the event (see online version for colours)
3.4.1.2 Time-based classification

Through experimentation, it is observed the resulting sets behave the following way: time-specific topic do not last longer than seven days, long-term events span from three weeks to a month which is the maximum collection we hold for analysis, but can expand to several months. Therefore, their classification will be done according to Definition 4.

Definition 4 (Time-based classification): Given a topic \( X \) lasting from \( t_s \) to \( t_e \) where every \( t \) is a date and each step between \( s \) and \( e \) is a day. \( D_X \) = \( t_e - t_s \) represents the topics duration in days. Let \( Class \) be the corresponding label for each topic \( i \).

\[
Class(i) = \begin{cases} 
\text{timespecific} & \text{if } D_X \leq 7 \\
\text{long-term} & \text{otherwise}
\end{cases} \tag{3}
\]

Starting from here topics will not only be classified according to their time extension, but manual categories will also be assigned, and volume considered as detailed in the following section.

3.4.1.3 Topic duration

As observed in the previous step a topic duration \( D_X \) can be easily obtained, nevertheless other variables have to be taken into consideration. Together with the volume of traffic generated they will define the lifespan component of our impact score. Detected events are manually assigned one of seven major categories: sports, news, entertainment, technology, seasonal or other occurrence. Some of these will be more prone to be time-specific and others long-term events. Grouping events into these categories will allow us to compare among them in order to assign a proper value to their behaviour, the more events are added the better results will turn out. The obtained value for each event must be normalised in order to serve its purpose properly. Comparing integers would not make much sense, it is required to come up with a value which can also be later combined with the remaining topic impact score components. Taking these factors into consideration Algorithm 1 is introduced to come up with an adequate value.

By normalising the components, the returned value will be between 0 and 1 as required. Additionally, by taking reference in the maximum values for duration and volume the algorithm will update according to the datasets added. Duration is given double the priority for it is considered to be the main component of this metric. As the topic collection grows, more categories can be added to ensure the proper evaluation of each dataset.

3.4.2 Emotion transition

Emotions and their resulting expressions are very dynamic and in constant transition. Much of changes come depending on things happening in our surroundings. Social networks being a platform that allows us to access such a massive amount of opinions, lay the ground for users to be affected by an event and the discussion this leads to throughout its occurrence. Relating the transitions to the surrounding events can be of much value on social network analysis. Therefore, we introduce the emotion transition metric which will contribute in a very significant manner to our intended impact score. The logic behind it being, that a topic that generates a significant change in public perception must be having a greater degree of influence over that group. This analysis can also lead to deeper insights on emotions, for example which are more prone to change, and which are the most effective means to make them change. The details on how to come up with this metric will be presented in the following paragraphs.

Algorithm 1 Get lifespan

<table>
<thead>
<tr>
<th>Input:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_X ): The duration for topic ( X )</td>
</tr>
<tr>
<td>( V_X ): The tweet volume generated for topic ( X )</td>
</tr>
<tr>
<td>( Class(X) ): The time class for topic ( X )</td>
</tr>
<tr>
<td>( Cat(X) ): The assigned category for topic ( X )</td>
</tr>
<tr>
<td>( D_{\text{similar}} ): Set of durations for topics which share ( Class ) and ( Cat ) with ( X )</td>
</tr>
<tr>
<td>( V_{\text{similar}} ): Set of volumes for topics which share ( Class ) and ( Cat ) with ( X )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Lifespan}(X) ): The resulting Lifespan value for topic ( X )</td>
</tr>
</tbody>
</table>

1: \( \text{if } (D(N – X) = 0) \text{ AND } (Class(X) = \text{timespecific}) \text{ then} \)
2: \( \text{MaxD} = 7 \)
3: \( \text{MaxV} = V_X \)
4: \( \text{else if } (D(N – X) = 0) \text{ AND } (Class(X) = \text{longterm}) \text{ then} \)
5: \( \text{MaxD} = 30 \)
6: \( \text{MaxV} = V_X \)
7: \( \text{else} \)
8: \( \text{MaxD} = \text{max} \{D_{\text{similar}}\} \)
9: \( \text{MaxV} = \text{max} \{V_{\text{similar}}\} \)
10: \( \text{end if} \)
11: \( d = 0 \) // Value for duration component.
12: \( v = 0 \) // Value for volume component.
13: \( \text{if } \text{MaxD} = D_X \text{ then} \) // Normalise duration.
14: \( d = 1 \)
15: \( \text{else} \)
16: \( d = \frac{D_X}{\text{MaxD}} \)
17: \( \text{end if} \)
18: \( \text{if } \text{MaxV} = V_X \text{ then} \) // Normalise volume.
19: \( v = 1 \)
20: \( \text{else} \)
21: \( v = \frac{V_X}{\text{MaxV}} \)
22: \( \text{end if} \)
23: \( \text{Lifespan}(X) = \frac{dv}{d+v} \)
3.4.2.1 Emotion detection

The first step required towards a final goal is to have all our resulting filtered tweets labelled with an emotion. The pattern-based approach is suitable for our intentions since the patterns are obtained from the same data source, twitter. This allows this classifier to perform properly even under informal ways of speech such as those present in social media platforms.

Additionally, the classifier considers eight emotion labels, derived from the centre ring of Plutchik’s wheel. These being anger, anticipation, disgust, fear, joy, sadness, surprise and trust. This variety of labels will allow us to have a more reflective distribution and better insights on the data. A sample for some patterns and the corresponding emotions is presented in Table 1.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Pattern examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>“you. + &lt;hashtag&gt;”, “stupid. +”, “shut up. +”</td>
</tr>
<tr>
<td>Anticipation</td>
<td>“expecting. +. +”, “. + wish. +”, “. + for. +”</td>
</tr>
<tr>
<td>Disgust</td>
<td>“. +. + &lt;hashtag&gt;”, “urgh. +”, “never. +. +”</td>
</tr>
<tr>
<td>Fear</td>
<td>“little. +”, “alone. +”, “hate. +”</td>
</tr>
<tr>
<td>Joy</td>
<td>“to be. +”, “. + on the”, “like. +”</td>
</tr>
<tr>
<td>Sadness</td>
<td>“never. +”, “being. +”, “am. +”</td>
</tr>
<tr>
<td>Surprise</td>
<td>“&lt;usermention&gt;. + &lt;hashtag&gt;”, “what. +”, “really. +”</td>
</tr>
<tr>
<td>Trust</td>
<td>“believe. +. +”, “. + sure. +”, “. + know. +”</td>
</tr>
</tbody>
</table>

3.4.2.2 Emotion distribution

As mentioned in Section 3.3, the emotion classifier is able to assign eight different labels to a string of text. Meaning that for every topic, every day there could be up to eight emotions being expressed depending on how the public is feeling. The emotion distribution for each day will be obtained as indicated by Definition 5.

Definition 5 (Emotion distribution): For every emotion $E_m$, $C_{Em}$ is the count of tweets corresponding to that specific emotion. $C_T$ is the total count of tweets for the day. Therefore, the corresponding percentage for each emotion is obtained by:

$$Perc_{Em} = \frac{C_{Em}}{C_T}$$

where the condition $\sum Perc_{Em} = 1$ must be met.

3.4.2.3 Transition analysis

To understand how or why emotions are changing, it is initially required to have an idea on how much they are changing. Therefore, we first generate a table of day to day differences in the distribution for each emotion and for the total. This values can give us a broader perspective, additionally with graphical methods it can help us determine how stable (or not) the feeling towards an event was. This stage is also required in the process of coming up with a value that represents the transition in an adequate manner. Figure 5 provides an example of the visual representation of an events emotion distribution transition.

Figure 5  Topic: Jose Mourinho is a renowned football coach who at the time was training a team in Spain (see online version for colours)
To directly favour those events with a greater emotion transition, the component will be the overall average of adding up all the transitions for every day in the event. This way a higher transition will be directly reflected, while those who did not vary much will be penalised. The effect will be later normalised in the final score computation.

3.4.3 Reach

The assumption that leads to the inclusion of reach in our formula is that an event that generated 1,000 tweets between 100 users is not the same as an event generating those same 1,000 but among 200 users. The greater amount of different users involved in the discussion represents a greater chance of this topic being impactful. A common approach towards this is extracting the spread of information over the social network. This task is computationally expensive and requires the detail crawling of the social network in order to be effective. For practical purposes, we will consider reach as the count of unique users who collaborated to the discussion according to the data present in our collected sets.

3.4.4 The score

The final score corresponding to the Topic Impact is initially computed as:

\[ \text{TopicImpact} = \text{lifespan} \times \text{transition} \times \text{reach} \]  

Then normalised based on the maximum values of the score for each duration class. Therefore, the resulting list will be composed of events with scores between 0 and 1, every new event and its score will be normalised according to and ranked within the list.

4 Experiments and results

A series of experiments were performed to test the proposed method. From traditional precision evaluations to some interesting combinations of results which can provide an insight of the value of this kind of analysis. Other specific require expert opinions to demonstrate the value of the obtained results. The following section will provide the details on these results, the setup that leads to them and a variety of graphical representations that can better portray the meaning of the findings.

4.1 Experimental setup

Multiple twitter datasets were collected from different time periods within the years 2013, 2014 and 2015. After applying the topic detection algorithm, 35 verified relevant topics were handpicked for the presented experiments and analysis. Since the main focus of this study is the further analysis done with the topics, the least ambiguous were selected to improve the quality, clarity and value in the outcome of the procedures applied later. Diversity was another consideration for the selection process, as a result within these are included topics for every combination of category and time class (12 in total). Out of the 35 datasets the initial 12 were used as training models and parameter setting, the remaining 23 were experimental.

The presented results contain both the training and test sets altogether since the more events included, the better the results. Due to space limitations, the majority of the results will be presented based on categories with some exceptions such as the case analysis which will focus on a specific event or the topic impact score and rank which will be presented for all events analysed.

4.2 Lifespan and data refinement

A questionnaire was designed to verify the precision of some of the methods leading to our final outcome, specifically our topic lifespan identification and data filtering stage. Sets of 100 tweets were created for each category (12 sets total), including tweets from every topic contained. For every tweet, a drop down menu containing the topics within that category is created so the user must select the topic to which the tweet belongs or ‘none’ in case it is unrelated. The same kind of sets were prepared for unfiltered data before lifespan and classes being defined. The latter data will enable us to have a reference point.

Test subjects were assigned sets according to areas they considered themselves informed about, to ensure they knew the topics being dealt with. The majority bilingual twitter users living in Central and North America. They are allowed to use external query resources to check details on topics they might not be sure of or specific slangs or ambiguous terms found in the tweets.

Lifespan definition is one of the novel and most significant contributions in the presented work. To provide an evaluation on its result, it is necessary to assess the quality of the collected data. A well defined event range, with content in proper levels of similarity would reflect the proper definition of the lifespan of the event. If the user believes the given content belongs to any of the events in the given category represents a good collection and filtering process. The results for this test are depicted in Figure 6. Additionally, we can observe the value of this process by comparing it to the results if the data was not properly filtered for every topic.

4.3 Topic impact score

Topic impact was calculated for each event according to the presented method. The results for all 35 events are ranked and presented in Table 4. One of the great events of 2014 such as the FIFA World Cup hosted in Brazil leads the ranking in a somewhat logical way. To compare our method against human opinion a survey was created.

Sets of 15 questions were generated; each question randomly paired two topics together and asked the user to determine which they considered more impactful. Results for each question were summarised and compared with the ranking provided on the table. Over 500 answers where collected from both male and female users in their mid 20s,
testers were allowed to verify events on search engines in case they were not so familiar with them. If the public opinion matched the result of the table it is considered correct, if it does not it is incorrect. Additionally, the survey was also compared to the ranking by individual component to demonstrate the combination of these outperforms ranking them independently. Figure 7 presents the result when comparing our method with the summary of each question, this means if the distribution of the users response for a given question was 60% to 40% or 80% to 20% the option selected by the majority will be considered the correct answer. Figure 8, on the other hand, will compare every individual answer to the ranking table.

**Figure 6**  Precision on how related the collected data is to the event with and without applying lifespan and data filters (see online version for colours)

**Figure 7**  The ranking obtained by our method was compared to a summary of human annotation answers on a questionnaire regarding impact of the topics (see online version for colours)
Figure 8 The ranking obtained by our method was compared to the results of human annotation comparing all answers on a questionnaire regarding impact of the topics (see online version for colours)

![Metric Comparison by Answer](image)

Table 2 Total list of topics rank by resulting topic impact score

<table>
<thead>
<tr>
<th>Topic</th>
<th>Score</th>
<th>Topic</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIFA World Cup 2014</td>
<td>1</td>
<td>Bayern-Barca(UCL 2013)</td>
<td>0.1384</td>
</tr>
<tr>
<td>Oscars 2015</td>
<td>0.9439</td>
<td>Christmas</td>
<td>0.1237</td>
</tr>
<tr>
<td>Golden Globes 2015</td>
<td>0.7399</td>
<td>New Years Eve</td>
<td>0.1224</td>
</tr>
<tr>
<td>UEFA CL Final 2014</td>
<td>0.6291</td>
<td>Indian Mars Rover</td>
<td>0.1081</td>
</tr>
<tr>
<td>Super Bowl 2014</td>
<td>0.4552</td>
<td>UK Independence Referendum</td>
<td>0.1052</td>
</tr>
<tr>
<td>American Idol</td>
<td>0.3411</td>
<td>Obama Tax Release</td>
<td>0.0752</td>
</tr>
<tr>
<td>Ellen’s Oscars 2014 Photo</td>
<td>0.3318</td>
<td>Justin Bieber on Teen Vogue</td>
<td>0.0646</td>
</tr>
<tr>
<td>North Korea Bomb Threat</td>
<td>0.2939</td>
<td>IcantBreathe campaign</td>
<td>0.0552</td>
</tr>
<tr>
<td>NBA Playoffs 2015</td>
<td>0.2852</td>
<td>Iphone 6 Launch</td>
<td>0.0461</td>
</tr>
<tr>
<td>OccupyCentral HK</td>
<td>0.2852</td>
<td>Liang Golf Masters 2013</td>
<td>0.0343</td>
</tr>
<tr>
<td>Spring Break</td>
<td>0.2499</td>
<td>PS4 Launch</td>
<td>0.0322</td>
</tr>
<tr>
<td>Charleston Massacre</td>
<td>0.2422</td>
<td>Margaret Thatcher’s Death</td>
<td>0.0309</td>
</tr>
<tr>
<td>Mourinho Back To Chelsea</td>
<td>0.2409</td>
<td>Movember</td>
<td>0.0300</td>
</tr>
<tr>
<td>Boston Marathon Bombing</td>
<td>0.2338</td>
<td>Justin Bieber and Lil Za hang out</td>
<td>0.0299</td>
</tr>
<tr>
<td>BringBackOurGirls</td>
<td>0.1979</td>
<td>Dark Soul II Box Art Revealed</td>
<td>0.0297</td>
</tr>
<tr>
<td>Feel Good Campaign</td>
<td>0.1525</td>
<td>Naya Rivera Has New Boyfriend</td>
<td>0.0155</td>
</tr>
<tr>
<td>ESA Rosetta Mission</td>
<td>0.1426</td>
<td>Justin Bieber meet and greet</td>
<td>0.0079</td>
</tr>
<tr>
<td>Valentines Day</td>
<td>0.1418</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4 Transition by category

A measure of transition by category can provide clues on which kind of topics can cause a greater change. From the obtained chart presented as Figure 9, we find TimeSpecific entertainment topics generated the greatest transition. When looking at the events contained in this category we find most of them are related to show business and gossip. This can lead to the inference that gossip has a strong impact on people’s opinion. Usually, these topics are brought to light as PR campaigns to generate this effect. This kind of measure can be useful on a long-term scale to better define our parameters and understand some categories tend to change more than the other.
Figure 9  The graph presents the percentage of how much did the emotions change over time for each different category and lifespan (see online version for colours)

![Transition Per Category](image)

Figure 10  The above illustrates the average emotion distribution separated by category and time class (see online version for colours)

![Emotion Distributions Per Category](image)

4.5 Emotion distribution by category

By intuition, it can be initially assumed categories will vary in emotional distribution. Given our process and collected data presents us with both categories and emotions Figure 10 is introduced to reflect how emotions are distributed across categories. Additionally, it can present how some categories behave differently from TimeSpecific to LongTerm classes.

In sports for example, for the TimeSpecific distribution we find joy as the dominant emotion while for LongTerm events trust leads the way. A reason for this could be the TimeSpecific content contains expressions on the exact day an event is taking place which could make people feel happy to watch. While on LongTerm sport events are usually discussions about season performance, or the constant debate on rivalries therefore trust being dominant could mean the fan groups expressions of loyalty to their teams.
4.6 Transition by emotion

Another interesting aspect to observe is which emotions are more prone to change, this could lead to an ongoing deeper analysis. To get a quick superficial glimpse of it, we calculated which emotions fluctuate the most across all events by adding together the absolute value of their transitions. The result is presented in Figure 11. Interestingly enough the two emotions that change the most are joy and sadness this can perhaps be a reflection of societies bipolar tendencies. Another rationalisation could be that these sensitive emotions are usually more on the surface and hence can be reached and modified with greater ease. It is normally easier to become suddenly happy or sad than disgusted for example.

Surprise also presents an interesting behaviour which can be rationalised in that there are less moments or expressions of surprise than the other emotions and additionally surprise is isolated, not accompanied by other emotions and usually found just at the beginning or ending of an event.

5 Conclusions and future work

Assigning an impact score can provide us with an effective way of which topics affect more social groups. We designed and demonstrated effective methods for the extraction of three major components to obtain this score. The importance of combining lifespan, emotion transition and reach is demonstrated in the experimental results, as well as how would a system perform if it only considered them independently. It was also demonstrated that laying the foundation on topic identification and emotion classification can lead to a richer understanding of social phenomena. The process itself has proven to generate collateral content that can as well be valuable for different analysis, for example, which emotions are more recurrent in different event categories or which are bound to change easily, etc. These insights can be of great value when trying to design political or marketing campaigns, prioritising news content, addressing social issues or even choosing how to behave on social networks.

The quantification of an impact score according to these variables is still a novel idea and therefore has significant room for improvement. Future work could include the integration of an automatic categorisation algorithm which can efficiently distribute our events into categories. Some components within our framework could be also developed for example integrating a more complex similarity measure when determining lifespan or enabling the topic detection system to work over a livestream. The time period could also be refined since currently our studies are done on a per day basis, this could be extended to hours. It is our belief this work lays the initial stones towards an ultimate goal which would be topic impact prediction, through which after analysing the initial behaviour of specific values an impact measure can be assigned to a topic before it concludes.

References


Notes


2 https://www.mongodb.org/ MongoDB is one of the leading popular NoSQL solutions on the market, its schema free json based database allows data as complex as twitter data to be stored without much complications.