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

# Discovering Sentiment Polarity Alternations in Twitter Conversations with Hilbert-Huang Spectrum

**Georgios Drakopoulos** · **Andreas Kanavos** ·
**Phivos Mylonas** · **Spyros Sioutas**

Received: date / Accepted: date

**Abstract** Does a tweet with specific emotional content posted by an influential account have the capability to shape or even completely alter the opinions of its readers? Moreover, can other influential accounts further enhance its original emotional potential by retweeting it and, thus, letting their followers read it? Real Twitter conversations seem to imply an affirmative answer to both questions. If this is indeed the case, then what is the key for not only successfully reaching to a large number of accounts but also for convincingly offering an alternative perspective via affective means, therefore triggering a large scale opinion change in an ongoing Twitter conversation? This work primarily focuses on determining which tweets cause multiple sentiment polarity alternations to occur based on a window segmentation approach. Moreover, an offline framework for discovering affective pivot points in a conversation based on its Hilbert-Huang spectrum, which has close ties to the Fourier transform. Finally, given that it is highly desirable to track the sentiment shifts of a Twitter conversation while it unfolds, an adaptive scheme is presented for approximating the window sizes obtained by the offline methodology. As a concrete example, the abovementioned methodologies are applied to three recent long Twitter discussions and the results are analyzed.

**Keywords** Opinion polarity · Functional analytics · Emotional influence · Social media analytics · Topic sampling · Signal processing for social media · Fourier spectrum · Hilbert-Huang transform

**PACS** 02.30.Gp · 02.30.Nw

**Mathematics Subject Classification (2000)** 42A16 · 46E20 · 62H30 · 62P25 · 91C20 · 91C99 · 91D10 · 91D30 · 91D99

---

Georgios Drakopoulos and Phivos Mylonas
Department of Informatics, Ionian University
E-mail: {c16drak, fmylonas}@ionio.gr

Andreas Kanavos and Spyros Sioutas
Computer Engineering and Informatics Department, University of Patras, Greece
E-mail: {kanavos, sioutas}@ceid.upatras.gr

**CR Subject Classification** C.3 · F.2.1 · G.1.1 · H.3 · J.4

## 1 Introduction

Microblogging platforms like Twitter are today the constantly changing melting pot of opinions typically shaped from and expressed in conversations about a broad array of subjects. Sharing thoughts and sentiments may well lead to a viral tweet, especially by accounts highly regarded by their respective communities, which may be able to change the collective sentiment of a conversation despite the restriction placed on the length of tweets. In fact, the latter may well be a major driver behind highly sentimental tweets, as there is barely sufficient space available for long, articulate arguments. Instead, terse and laconic tweets, conveying a substantial amount of information nonethless, frequently appear as stated in [5]. Therefore, harnessing the emotional content in this enormously continuous and volatile Twitter stream is bound to reveal trending opinions about and reactions, which ultimately shape public attitude, to a wide array of phenomena ranging from online marketing campaigns to political events as shown among other in [36].

A key factor towards discovering the dynamics of online public sentiment lies in identifying the evolving set of emotionally influential accounts. This set may be evolving over time and depends heavily on the conversation topic as observed in [31]. Typically, candidate influential accounts include corporate accounts, verified accounts, and individual persons who are accomplished in their field or are fluent. However, this not need be the case and the criteria for ascertaining emotional influence over a conversation or a segment thereof are more complex. Consequently, numerous techniques for assessing the affective potential of an account have been developed. A large class of white box methodologies rely on providing Twitter features to knowledge discovery algorithms. Alternatively, schemes from the emerging field of signal processing for social media can be employed. The latter typically treat social media data as time domain signals and perform on them signal processing operations such as noise reduction, signal modeling, and harmonic analysis.

The three primary research objectives of this article are the following. First, a methodology is proposed for assessing the emotional content of a given tweet based on its potential to trigger collective sentiment shifts during an ongoing Twitter conversation. This is accomplished by comparing the affective dynamics between successive conversation windows of fixed size. Second, a benchmark offline framework is presented in order to evaluate the effectiveness of the fixed window approach and to assess the degree of sentimental volatility in a given Twitter conversation based on its Hilbert-Huang spectrum. Finally, an adaptive window scheme is developed based on the offline methodology.

The remainder of this work is structured as follows. Section 2 presents presents background topics in sentiment analysis and digital influence estimation. Section 3 explains the methodologies for assessing tweet affective content and for creating an evolving set of emotionally influential accounts. Section 4 shows the offline baseline framework for evaluating the effectiveness of the above schemes and describes the adaptive step selection mechanism. Section 5 describes the dataset collection, the experiments, and the associated results. Finally, section 6 recapitulates the principal conclusions and enumerates possible directions for future work. Table 1 summarizes article notation. Finally, two notes about terminology:

- The term *account* tends to displace the less generic term *user*, since entities such as organizations, states, government agencies, and corporations may well have Twitter presence.
- The terms *graph* and *network* are not interchangeable in the text. The former refers only to the structural properties of a social network, whereas the latter to its functionality.

**Table 1** Notation of this article.

| Symbol | Meaning |
|---|---|
| $\triangleq$ | Definition or equality by definition |
| $\{s_1,\ldots,s_n\}$ | Set with elements $s_1,\ldots,s_n$ |
| $(s_1,\ldots,s_n)$ | Tuple with elements $s_1,\ldots,s_n$ |
| $\lvert S \rvert$ | Cardinality of set or tuple $S$ |
| $\tau_{S_1,S_2}$ | Tanimoto similarity coefficient for sets $S_1$ and $S_2$ |
| $\Phi\left(c[i]\right)$ | Set of accounts following $c[i]$ |
| $\Psi\left(c[i]\right)$ | Set of accounts followed by $c[i]$ |
| $\langle p \Vert q \rangle$ | Kullback-Leibler divergence for distributions $p$ and $q$ |
| $x_1(t) \star x_2(t)$ | Linear convolution between signals $x_1(t)$ and $x_2(t)$ |
| $\mathscr{F}\left[x(t)\right]$ | Fourier transform of signal $x(t)$ |
| $\mathscr{H}\left[x(t)\right]$ | Hilbert transform of signal $x(t)$ |

## 2 Related Work

Sentiment analysis has garnered considerable interdisciplinary interest as social media are an excellent multimodal source of emotionally polarized text, hashtags, images, music, and video as stated among others in [21] and in [9]. For instance, movie reviews can be identified as positive or negative as in [28]. As a result, various methodologies have been developed to harvest emotional potential in social media as described in [24]. Aspects such as text objectivity as in [2], opinion polarity as in [27] and clustering as in [39], text mining on word and sentence level as in [20] or on phrase level as in [42], sentiment mining in mutilingual Web texts as in [4], discovering multilingual communities as in [12], and linguistic styles for various arguments as in [31] have been examined. Emotional information diffusion in Twitter based on the Ekman model is examined in [18]. The relationship between decsion making and emotional tweet content is explored in [41]. For a thorough review see [25].

Signal processing for graphs has been proposed as an alternative to the combinatorial approach. Fundamental notions such as graph frequency, graph shifting, and graph Fourier transform are defined and explained in [32]. The applications of Kronecker and strong product graphs to big data processing are described in [33]. Most of these concepts come from the graph Laplacian matrix as described in [3]. The Laplacian matrix can be used for spectral graph clustering as shown in [6], hashing as shown in [26], and regularization as shown in [34]. Many of the graph Laplacian properties including those of the Fiedler eigenvalue are explored in [22]. The combination of functional and structural Twitter features to a multilayer graph and its interpretation as a signal are discussed in [10].

The applications of the above approaches to both political and commercial campaigns are many as shown in both [19] and [35]. Governments seek ways to reach specific target groups in social media as stated in [38]. Converesely, social media have recently been the platforms for massive protests or even revolutions, resulting in more detailed examination of their content as shown in [16]. A system for real time Twitter sentiment analysis during the US 2012 presidential election cycle was presented in [40] and improved in [8]. The public sentiment of a community, a city, or even a country is examined in [30]. In [1] is argued that substantial information about beliefs and emotional states can be inferred from a person's tweets. In [29] and in [43] tweets about Hollywood films are driven to classifiers including Naive Bayes, Maximum Entropy, and SVM. In [7] the effect of community feedback to the decisions of authors is examined, with negative and positive feedback resulting in vastly different decisions.

Finally, the significance of digital influence is disussed in detail in [13]. In [44] a two-way friendship model is proposed and used to study the influence between Twitter accounts. A methodology for creating a higher order Twitter from first order ones is shown in [11]. The notion of influence is extended from accounts to communities and networks and, moreover, communities are built based on user personality traits in [15]. Machine learning models for predicting whether a mention will be made to a verified account are proposed in [23]. Finally, account influence based on Twitter functionality and the effects of inaccurate model selection is also studied in [17].

## 3 Emotionally Influential Tweets And Accounts

### 3.1 Topic Sampling

The architecture of the system implementing the proposed Twitter emotional metrics and the associated data flow are shown in figure 1. Its main components are the Twitter crawler, the tweet emotion recognition, and the influence computation.

Initially, a large volume of tweets about a specific topic are collected through the Twitter crawler as shown in algorithm 1. The latter traverses the Twitter graph and extracts tweets relevant to the query keywords and additionally:

- The tweet timestemp.
- The account posted that tweet.

Moreover, for each account $c[k]$ are collected:

- The number of followers $|\Phi(c[k])|$ and followees $|\Psi(c[k])|$.

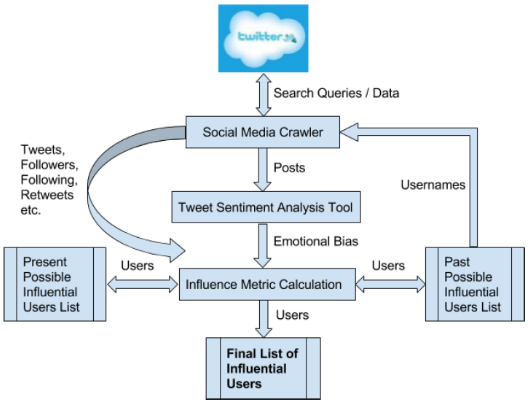


**Fig. 1** System architecture.

- The number of tweets $\sum_j t[j;k]$, retweets $\sum_j r[j;k]$, and direct messages $d[k]$. Notice that $t[j;k]$ and $r[j;k]$ denote the $j$-th tweet and retweet of $c[k]$ respectively.
- Mentions $m[k]$, profile clicks $n[k]$, favorites $v[k]$, and replies $s[k]$.

These two distinct feature sets allow the construction of metrics of both account and tweet affective influence in accordance with the guidelines set forth in [15], [17], and [18]. In fact, the tweet affective metrics can be seed as the dual of the account ones, since there is a mapping between the tweets and the accounts posting them. Also, the number of tweets is deliberately large in order to ensure sufficient statistical diversity.

### 3.2 Activity and Affective Metrics

A popular sentiment analysis tool titled SentiStrength [1] and described in [37] extracted tweet emotional content by analyzing each word based on a sentiment strength algorithm. The main reason for choosing SentiStrength is its procedures for decoding non-standard spellings and methods for boosting the strength of words, which accounted for much of its performance. The key elements of SentiStrength are listed below:

- The algorithmic core is the sentiment word strength list; this is a collection of 298 positive and 465 negative terms classified for either positive or negative strength.
- A term is randomly selected and its strength is increased or decreased by one and classification is performed again until no accuracy change occurs for all strengths.
- Spelling correction identifies words that have been misspelled by repeated letters.
- A booster list has words boosting or reducing the emotion of subsequent words.
- A negating list contains words inverting subsequent emotion words.
- Emoticons with associated strengths supplement the sentiment word strength.

[1] http://sentistrength.wlv.ac.uk/

**Algorithm 1** Generation of Twitter subgraph with topic sampling

**Require:** Query with hastag #$q$
**Ensure:** $Users$, $Followers[]$, and $Newnodes[]$ are computed

**identify** set of tweets for given #$q$, $T = \{t_1, t_2, \ldots, t_i\}$
**for all** $t_i \in T$ **do**
    $u_i \leftarrow$ user of tweet $t_i$
    $Followers[i] \leftarrow$ Followers of $c[i]$
    **for all** $t_i \in T$ **do**
        $Users \leftarrow Users \cup u_i$
    **end for**
    **identify** the followers of $u_k$, $Followers[u_k] = \{f_1, f_2, \ldots, f_j\}$
    **for all** $u_k \in$ Users **do**
        **for all** $f_j \in Followers[u_k]$ **do**
            **if** $f_j \in$ Users **then**
                **link** $f_j$ with $u_k$
            **else**
                **for all** $u_l \in$ Users **and** $u_l \neq u_k$ **do**
                    **if** $f_j \in Followers[u_l]$ **then**
                        $Newnodes \leftarrow Newnodes \cup f_j$
                        **link** $f_j$ with $u_k$ **and** link $f_j$ with $u_l$
                    **end if**
                **end for**
            **end if**
        **end for**
    **end for**
    $Users \leftarrow Users \cup Newnodes$
**end for**
**return** $Users$, $Followers[\,]$, and $Newnodes[\,]$

– Any sentence with an exclamation mark or even repeated punctuation including at least one exclamation mark is given a corresponding sentiment strength.

Initially, the sentences were split by line breaks or after punctuation. Then, the abovementioned elements were separately applied to each tweet to derive the final affective strength of each word, with each tweet receiving the sum of the individual such strengths as shown in algorithm 2. It should be noted that the sentiment analysis tool disambiguates of equivocal phrases. Such phrases exhibit a contradiction between the emotional bias hinted by their words and the actual emotions their authors intend to convey. This analysis is independent of this tool, since any emotional analysis algorithm mapping the affective strength of a tweet to a scalar can work.

**Algorithm 2** Tweet Emotional Bias Computation

**Require:** Account $c[k]$ and tweet $t[j;k]$
**Require:** ComputeBias($t[j;k]$, $term$)
**Ensure:** Emotional bias is in the form of an integer is computed
1: **for all** $term \in t[j;k]$ **do**
2: $integer += $ ComputeBias($t[j;k]$, $term$)
3: **end for**
4: **return** $(c[k], t[j;k], integer)$

In our experiments the activity impact $J[k;w]$ and sentiment impact $Q[k;w]$ metrics evaluate the influence of $c[k]$ from different perspectives over a window $w$, with $0 \leq w \leq W-1$, consisting of $L$ tweets each sorted in ascending order based on their timestamps. The former metric quantifies the online activity of an account, whereas the latter acts as an indicator of its affective potential. The intuition behind relying on both a functional and an affective metric is that although any account can potentially post a tweet of high affective impact, truly influential accounts will not only post many such tweets but also they will have a substantial online presence. Thus digital influence should be checked by both an activity metric and an affective one, or alternatively by a metric combining both such aspects.

The PostImpact metric is a product of the online activity of $c[k]$ and the LFtF factor. The rationale is that an influential account should not only be active but also this activity should be diffused through a high number of followers. The followers-to-following (FtF) ratio for a given Twitter account $c[k]$ is an important feature, whereas the logarithmic FtF (LFtF) reveals the order of magnitude of the FtF as:

$$\mathrm{LFtF}[k] \triangleq \begin{cases} \log_{10}\left(1 + \frac{|\Phi(c[k])|}{|\Psi(c[k])|}\right), & |\Psi(c[k])| \neq 0 \\ 0, & |\Psi(c[k])| = 0 \end{cases} \tag{1}$$

Thus, the online activity $J[k;w]$ during window $w$ is computed as follows:

$$J[k;w] \triangleq \frac{\left(\sum_j t[j;k] + \sum_j r[j;k] + 1\right)}{d[k]+1} \times \\ \times (m[k]+1)(n[k]+1)(v[k]+1)(s[k]+1)\mathrm{LFtF}[k] \tag{2}$$

One important aspect of equation (1) is that the base 10 logarithm of the FtF ratio is taken in order both to avoid outlier values and to take into account the order of magnitude of this ratio. Also, the FtF ratio is augmented by one in the logarithm argument in order to remedy numerical instabilities from very small values of this ratio. In addition, factors in equation (2) is added by one so as to avoid side effects with zero values. Along a similar line of reasoning, the five features are also added by one so as to avoid the metric being equal to zero in cases that retweets, replies, favorites, mentions, or clicks received are zero.

As stated earlier, consider a window of $L$ tweets of a given conversation ordered in ascending timestamps. Let the cumulated bias of a window $w$ be the sum of the emotional polarity of the tweets of that window. For every $c[k]$ and for every $t[j;k]$ the following two steps are repeated:

- The bias signs $b$ and $b'$ of two successive windows $w$ and $w+1$ are computed. Both windows contain tweets sorted based on their timestamp, meaning that the wallclock time during which the conversation is unfolded is ignored, and may well include multiple tweets of the same account $c[k]$.
- If the above signs are equal, the tweets in $w$ are not considered influential enough to alter the discussion sentiment. Otherwise, if a tweet $t[j;k]$ in $w$ has the same polarity with that of $w+1$, then it may have been the cause of the change and $c[k]$ is marked as potentially influential. Thus the status of $c[k]$ for window $w$ is updated as:

$$c[k] \triangleq \begin{cases} \text{NotInfluential}, & b = b' \\ \text{PotentiallyInfluential}, & (b \neq b') \wedge (\operatorname{sign}(t[j;k]) = b') \end{cases} \tag{3}$$

Moreover, two counters are updated. Counter $p[k;w]$ contains the times $c[k]$ is marked as potentially influential, whereas $q[k;w]$ contains the number of tweets $c[k]$ has posted in window $w$.

Thus, the sentiment impact $Q[k;w]$ is computed as:

$$Q[k;w] \triangleq \frac{1}{2}\left(\frac{p[k;w]}{\sum_k p[k;w]} + \frac{q[k;w]}{\sum_k q[k;w]}\right) \tag{4}$$

Finally, the total influence $I[k;w]$ of $c[k]$ marked as potentially influential is estimated by computing the harmonic mean of normalized $J[k;w]$ and $Q[k;w]$. The normalization of the former term is necessary in order to keep both terms of the harmonic mean to the same range:

$$I[k;w] \triangleq \frac{2}{\frac{\max_k(J[k;w])}{J[k;w]} + \frac{1}{Q[k;w]}} \tag{5}$$

The harmonic mean is robust to outliers and has the tendency to be closer to the smaller of its arguments. Therefore, a truly and consistently influential account must achieve high scores in both activity and sentiment influence.

## 4 Offline Conversation Evaluation

### 4.1 Polarity Intrinsic Mode Analysis

Since sentiment polarity alternations in a Twitter conversation cannot be known in advance and, moreover, no prior knowledge or ground truth is available, it makes sense to extract these alternations from raw data. One way to achieve that is the

Hilbert-Huang transform (HHT) or empirical mode decomposition (EMD) which decomposes a time domain signal $x[n]$ to a set of $C$ intrinsic mode functions (IMFs) $c_j[n]$ and a possible data residual $r_C[n]$ as:

$$x[n] = \sum_{j=1}^{C} c_j[n] + r_C[n] \tag{6}$$

The form of (6) reveals oscillations expressed in the IMFs full with local patterns inherent in the original data and also they factor in latent higher order dynamics.

Consider the discrete signal $u[n]$, where $0 \leq n \leq N-1$, formed by the polarity sign of the $N$ tweets of the conversation, where tweets are sorted in ascending order based on their timestamps. The HHT of $u[n]$ can be computed in either of two ways. The first is to treat $u[n]$ as a sampled continuous time signal $u[t]$ and compute the transform of equation (7):

$$U(\tau) = \mathscr{H}[u(t)] \triangleq u(t) \star \frac{1}{\pi t} = \int_{-\infty}^{+\infty} \frac{u(t)}{\pi(\tau - t)} dt \tag{7}$$

The integral in equation (7) should be interpreted in the sense of the Cauchy principal value or, equivalently, as:

$$U(\tau) = -\lim_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \frac{u(\tau + t) - u(\tau - t)}{\pi t} dt \tag{8}$$

The second way it to extract directly the HHT coefficients according to algorithm 3. The stopping criterion of standard deviation, as proposed in [14], is given in equation (9) and it mandates that it should remain higher than a given threshold $\eta_0$:

$$\mathrm{sdev} \triangleq \sum_{n=0}^{N-1} \frac{\left| h_{j,k-1}[n] - h_{j,k}[n] \right|^2}{h_{j,k-1}^2[n]} \geq \eta_0 \tag{9}$$

Note that for our purposes only the first IMF $c_1[n]$ is necessary in order to determine the sentiment polarity changes in $u[n]$. That is because $c_1[n]$ is considered to be the primary IMF deriving directly from the original data. Other IMFs may well capture higher order dynamics as stated earlier, but right now they will not be considered.

Finally, it should be also noted that the Hilbert spectrum is closely related to the Fourier spectrum:

$$U\left(e^{i\omega}\right) \triangleq \int_{-\infty}^{+\infty} u(t) e^{-i\omega t} dt \tag{10}$$

Specifically, the HHT spectrum is an instantaneous snapshot of the Hilbert spectrum of $x[n]$, where the latter is the Fourier spectrum of $x[n]$ with the positive frequencies shifted by $\frac{\pi}{2}$. Thus, from the Hilbert spectrum the Fourier spectrum can be recovered and vice versa. The HHT spectrum has the important property that the frequencies of the IMFs are inherent in the original data and, hence, they reflect significant changes in $x[n]$. This is the reason for selecting $c_1[n]$ as a natural representation for sentiment polarity, as its intrinsic zero crossings reflect actual polarity changes.

Still, since this an offline analysis, these crossings are computed once the conversation is complete. Nonetheless, this type of analysis is useful for the following reasons:

- It can establish a benchmark for comparing window size selection policies.
- The differences in zero crossings can be used to compare Twitter conversations.

**Algorithm 3** Hilbert-Huang Transform (HHT) with standard deviation termination

**Require:** Signal $x[n]$, number of IMFs $C$, and stopping criterion threshold $\eta_0$
**Ensure:** The IMFs are computed
1: **for** $j \leftarrow 1$ **to** $C$ **do**
2:     **interpolate** between local maxima of $x[n]$ to obtain $x_u[n]$
3:     **interpolate** between local minima of $x[n]$ to obtain $x_l[n]$
4:     $m_{j,0}[n] \leftarrow \frac{1}{2}(x_u[n] + x_l[n])$ **and** $h_{j,0}[n] \leftarrow x[n] - m_{j,0}[n]$
5:     **repeat**
6:         **apply** recursively steps 2 to 4 to $h_j[n]$ to obtain $\{m_{j,k}[n]\}$, $\{h_{j,k}[n]\}$, $k$
7:     **until** sdev $\geq \eta_0$
8:     $c_j[n] \leftarrow h_{j,k}[n]$ **and** $x[n] \leftarrow x[n] - c_j[n]$
9: **end for**
10: **return** $\{c_j[n]\}$

## 4.2 Influential Set Evolution

From the discussion of the section 3 it follows that the window length $L$ is a crucial parameter in uncovering local patterns and sentiment dynamics, including sentiment alternations. Since the optimal length $L^*$, which may well be variable, is unknown, one approach lies in analyzing a conversation offline in order to discover patterns which can be used to construct mechanisms capable of approximating it. By optimal it is meant that $L^*$ leads to the same conversation segments with those obtained by the zero crossings of $c_1[n]$.

One policy for selecting $L$ is to keep it constant in a value $L_0$ based on an average of a large number of similar conversations. Although given the plethora of available Twitter features it is fairly easy to reasonably define when two conversations are similar, this can be achieved only when the conversation or at best a large part of it is over. The approach proposed here relies on approximating $L^*$ indirectly based on measurable outcomes of a sentiment polarity change. The intuition behind our approach is that as a given sentiment continues to drive the conversation, then the core set of influential accounts will be roughly the same. On the contrary, when the sentiment changes polarity, then a new set of accounts supporting a different viewpoint will dominate the conversation. Moreover, if the set of potentially influential account changes, then this may be an indication that a shift is going to happen. In other words, changes in that set are considered as an estimate of the future sentiment status of the conversation. Thus, the following rules should hold:

– Each new window has a starting size $L_i$ which is relatively long in order to prevent instabilities but also sufficiently short in order to capture a polarity shift.
– As long as both the polarity sign and the set of potentially influential accounts maintains a relative similarity, then the overall sentiment is considered to be steady and to the window length is added a big quantum $\Delta L$.
– If the polarity is the same but the set of potentially influential accounts changes, then a sentiment polarity shift may be imminent. Therefore, to the window length is added a small quantum $\Delta L'$.
– Finally, if the sentiment polarity changes, then the current window is terminated and a new window starts.

It remains to see how set coherence is measured. One way to measure the similarity of two sets is the Tanimoto coefficient defined as:

$$\tau_{S_1,S_2} \stackrel{\triangle}{=} \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{|S_1 \cap S_2|}{|S_1| + |S_2| - |S_1 \cap S_2|} \tag{11}$$

The second form of equation (11) follows directly from the Venn diagram for two sets and it is more efficient for large sets. This will be used in order to measure the coherence between sets with potentially influential accounts.

As a conversation progresses in time, it is reasonable to ask how the set of potentially emotionally influential accounts evolves in the long term. This is another way to understand how a conversation unfolds as more tweets are posted by more accounts. Assuming that a conversation is segmented to $W$ windows, possibly of variable size, then the geometric mean of the $W-1$ successive Tanimoto coefficients shown in equation (12) can reveal how coherent a conversation is -or not:

$$\bar{\tau} \stackrel{\triangle}{=} \sqrt[n-1]{\prod_{k=0}^{W-2} \tau_{V_{k+1},V_k}} = \left(\tau_{V_1,V_0} \cdot \tau_{V_2,V_1} \cdot \ldots \cdot \tau_{V_{W-1},V_{W-2}}\right)^{\frac{1}{n-1}} \tag{12}$$

## 5 Results

### 5.1 Dataset Synopsis

Three Twitter datasets have been created with Twitter4j[2], a Java based platform for interacting with the Twitter API. Additionally, the lexical and emotional analysis of the words of each tweet was done with SentiStrength. The Twitter subgraphs were collected over a time interval of two months, namely between 01/06/2019 and 31/07/2019. As stated earlier, a topic-based sampling approach was employed where tweets were collected via a keyword search query. Specifically, three discussions have been collected based on the hashtags #BigData, #GermanWings, and #Node.js.

[2] http://twitter4j.org/en/index.html

The first hashtag, #BigData, is reflecting a discussion topic with mostly scientific and business interest. It is quite sparse but also almost linear in activity with time. The second hashtag, #GermanWings, deals with a tragic plane crash and the associated online discussion which had an initial burst of activity spanning a few days but then faded to low activity levels with some occasional resparking after an official announcement or a post regarding the progress of the investigation. Finally, the #Node.js hashtag is a long running technological topic which has become extremely popular during the past few years. In this discussion participate software companies, organizations, and professional developers.

It should be noted that the abovementioned datasets have been preprocessed in order to remove tweets with were irrelevant to the conversation even though they were posted by accounts participating to it. Moreover, "egg" accounts were removed since they are typically considered less reliable. The structural properties of the three datasets are shown in table 2, whereas the functional ones in table 3. The former contains fundamental properties such as connectivity patterns like weak and strong triangles. The latter has Twitter specific properties such as the average tweet length and the average number of followers. Note that the vertices are accounts and the directed edges represent the *follow* relationships.

**Table 2** Structural features.

| Property | #BigData | #GermanWings | #Node.js |
|---|---|---|---|
| Vertices | 22718 | 15619 | 23557 |
| Edges | 149152 | 40081 | 72903 |
| Density | 6.5653 | 2.5661 | 3.0947 |
| Log-completion | 0.5938 | 0.5487 | 0.5561 |
| Triangles | 1213 | 804 | 1135 |
| Squares | 772 | 517 | 693 |
| 4-cliques | 472 | 355 | 594 |
| Stars | 35 | 21 | 26 |
| Components | 1 | 1 | 1 |
| Diameter | 11 | 7 | 9 |

**Definition 1 (Completion)** The completion $\sigma$ of a directed graph is defined as the number of edges to the number of the edges of a complete directed graph with the same number of vertices:

$$\sigma \stackrel{\triangle}{=} \frac{|E|}{2\binom{|V|}{2}} = \frac{|E|}{|V|(|V|-1)} \approx \frac{|E|}{|V|^2} = \gamma_1 |V|^{\gamma_0 - 2} \tag{13}$$

The last step to equation (13) is due to the fact that in scale free graphs the number of edges $|E|$ and the number of vertices $|V|$ are connected as:

$$|E| = \gamma_1 |V|^{\gamma_0} \tag{14}$$

**Definition 2 (Log-completion)** The log-completion $\sigma'$ of a directed graph is defined as the logarithm of the number of edges to the logarithm of the number of edges of a complete directed graph with the same number of vertices:

$$\sigma' \triangleq \frac{\log |E|}{\log \left(2\binom{|V|}{2}\right)} \approx \frac{\log |E|}{2\log |V|} = \frac{\log \gamma_1 + \gamma_0 \log |V|}{2\log |V|} \approx \frac{\gamma_0}{2} \tag{15}$$

The entries of table 2 suggest that all three conversations come from highly interconnected social subnetworks. Therefore, it is reasonable to assume that information, especially tweets and retweets, are visible to the participants to each conversation. Thus, a tweet with rich emotional potential can alter the course of a conversation as it will not only influence the thoughts and actions of the followers of its creator.

In table 3 some Twitter functional features related to the evolution of a conversation over time are shown. From its entries, it can be deduced that all three datasets contain rather active conversations. The latter is an indication that the results of this section have statistical validity.

**Table 3** Functional features.

| Property | #BigData | #GermanWings | #Node.js |
|---|---|---|---|
| Hashtags | 24 | 17 | 53 |
| Tweets | 21315 | 11717 | 25932 |
| Retweets | 9133 | 10881 | 8356 |
| Mentions | 5117 | 7114 | 8221 |
| Favorites | 89082 | 99540 | 123881 |

The sentiment distribution throughout each conversation can show some inherent tendencies. #BigData topic appears to be a mostly positive topic with participants mostly talking enthusiastically about new technologies or algorithmic breakthroughs, although certain objections or doubts about the efficiency of some proposed technology are also present. #GermanWings is clearly a negatively charged conversation, which is understood given the tragic event and the subsequent revelations about it. Finally, the #Node.js conversation is a balanced one, which can be explained by the technical nature of the topic. Additionally, it may be explained by the fact that many professionals seeking networking, looking for special exclusive events about Node.js and devops, get frequent updates about events such as the *Nodeconf*, or interacting online with technology companies select a wording for their post which is neutral or very mildly positive.

Observe that in all three conversations, all sentiment categories are present in various degrees. Positive polarity describes the emotions that change the affective stance towards a better situation, while in contrast, negative polarity tends to affect human psychology towards a more unpleasant direction. Additionally, the three conversations have different characteristics which heavily depend on how the Twitter community reacts emotionally as the conversation topic unfolds over time.

**Table 4** Sentiment distribution in each conversation.

| Conversation | Positive | Negative | Neutral |
|---|---|---|---|
| #BigData | 71% | 16% | 13% |
| #GermanWings | 24% | 65% | 11% |
| #Node.js | 19% | 14% | 67% |

### 5.2 Results

The following window length policies for segmenting a conversation will be compared. Notice that in every case the window length is an odd number. This is done on purpose in order to break possible ties in the total sentiment polarity.

- P0: The windows obtained by the first IMF of the HHT of the three conversations with $\eta_0 = 0.1$.
- P1: An adaptive policy with the following parameter tuple:

$$\left(L_0, \Delta L, \Delta L', \eta_1\right) = (71, 51, 21, 1 - \ln 2) \tag{16}$$

- P2: A constant window length $L_1 = 51$
- P3: A constant window length $L_2 = 71$

The threshold $\eta_0$ for the $P_0$ has been selected based on recommendations found in [14]. The rationale behind the selection of the parameters of $P_1$ is that the initial window size $L_0$ is long enough to create a robust estimation of the initial community sentiment. Then, the regular window increment $\Delta L$ contains a small batch of tweets so that the list of potentially emotionally influential accounts can be updated, reflecting the underlying conversation dynamics. Finally, the small window increment $\Delta L'$ provides a finer granularity so that an imminent sentiment polarity alternation will be tracked. Finally, the threshold under which a change to the set of potentially influential accounts is considered high is when two of three such accounts are replaced.

In order to evaluate the above window selection policies, or any two such policies for that matter, one possible criterion will be the Kullback-Leibler divergence between the window sizes, treated as a distribution. Specifically, the divergence between a test distribution $p_t$ and a baseline distribution $p_b$ is defined as:

$$\langle p_t || p_b \rangle \stackrel{\triangle}{=} -\sum_{k=1}^{n} p_t[k] \log\left(\frac{p_t[k]}{p_b[k]}\right) \tag{17}$$

In equation (17) $p_b$ will always be the one obtained by the HHT. The latter has been selected as the ground truth, since the IMFs reflect frequencies inherent in the data themselves.

A deterministic way to compare the sequences of the window sizes $l_b[j]$, $1 \leq j \leq W_b$, and $l_t[j]$, $1 \leq j \leq W_t$, obtained by a baseline and a test policy respectively is to compute the modified mean square error (MMSE) as shown in equation (18):

$$\text{MMSE} \triangleq \frac{1}{\min(W_b, W_t)} \sum_{j=1}^{\max(W_b, W_t)} (l_b[j] - l_t[j])^2 \tag{18}$$

When the length of the two window sequences is different, then the shortest is padded with zeros in order to match the longest one in length. However, the MMSE is divided by the shortest length, which acts as an additional penalty factor when $W_t$ is either much shorter or much longer than $W_b$.

A third way to evaluate the window selection policy is to consider the Kullback-Leibler divergence between the distribution of the number of accounts marked as potentially influential obtained by a test policy against that obtained by the baseline policy:

$$\langle q_t || q_b \rangle \triangleq -\sum_{k=1}^{n} q_t[k] \log\left(\frac{q_t[k]}{q_b[k]}\right) \tag{19}$$

In table 5 for each conversation the values obtained by each performance metric are shown. Given its entries, it follows that $P_1$ is the best approximation to $P_0$. This can be attributed to its adaptive nature, which can track easier sentiment polarity alternations. The second best approximation is $P_2$, since it provides finer granularity. Finally, $P_3$ has the worst performance as it has a long window which not only yields low resolution, but also systematically an incorrect number of potentially influential accounts. Conversely, the #Node.js is the easier conversation to approximate, whereas the #GermanWings the latter. This can be explained in conjunction with the findings of table 6.

**Table 5** Performance of each policy in each conversation.

| $P_1$ | $\langle p_t \|\| p_b \rangle$ | MMSE | $\langle q_t \|\| q_b \rangle$ |
|---|---|---|---|
| #BigData | 0.4319 | 4.7767 | 0.3998 |
| #GermanWings | 0.5833 | 5.4980 | 0.4646 |
| #Node.js | 0.3721 | 4.2221 | 0.3982 |
| $P_2$ | $\langle p_t \|\| p_b \rangle$ | MMSE | $\langle q_t \|\| q_b \rangle$ |
| #BigData | 0.5833 | 7.4532 | 0.4486 |
| #GermanWings | 0.6011 | 8.7778 | 0.4698 |
| #Node.js | 0.4417 | 7.1902 | 0.4003 |
| $P_3$ | $\langle p_t \|\| p_b \rangle$ | MMSE | $\langle q_t \|\| q_b \rangle$ |
| #BigData | 0.6209 | 9.1132 | 0.6544 |
| #GermanWings | 0.6551 | 9.6312 | 0.6787 |
| #Node.js | 0.4851 | 8.5093 | 0.5999 |

Finally, in table 6 the values of $\bar{\tau}$ from equation (12) are shown. Its entries indicate that #Node.js has a very highly coherent list of influential accounts, implying that the majority of the participants are influenced by the same few accounts, which given the nature of the subject may include companies, prestigious community conferences, lead developers, and technology experts. The #BigData conversation is considerably less coherent, indicating there are much fewer accounts with consistently high influence. This can be attributed to the very open and wide scope of the topic which allow many accounts to contribute to the subject. Finally, the #GermanWinds has very low overall coherence. A possible explanation is the original accounts, mostly news agencies, which announced the incident had a neutral tone. As the conversation unwinds though, accounts such as celebrities, posted more dramatic tweets influencing many participants.

**Table 6** Values of $\bar{\tau}$.

| Conversation | #BigData | #GermanWings | #Node.js |
|---|---|---|---|
| $\bar{\tau}$ | 0.5182 | 0.2258 | 0.9727 |

## 6 Conclusions

This article focuses on discovering emotional influence on Twitter conversations based on the affective potential of a tweet to change the overall sentiment of that conversation. This central idea leads to the study of emotional dynamics of tweets and how should a sequence of tweets be segmented in order to reveal truly influential tweets. The primary contribution of this work the definition of the sentimental potential of a tweet in terms of affective polarity alternations, which translates to the ability to trigger massive emotional shifts to the conversation participants. Second, a framework is developed for assessing offline the emotional changes of a conversation based on its Hilbert-Huang spectrum. Third, an adaptive mechanism inspired by the field of adaptive signal processing is proposed for approximating the intrinsic sentiment changes reflected in that spectrum. This mechanism relies on both changes in the overall sentiment as well as on abrupt changes to the set of influential accounts in order to estimate when an affective alternation is imminent. The Hilber-Huang spectrum is considered to be the ground truth of the affectional dynamics of a conversation since it is extracted directly from the original data. Still, since this can be achieved only when the conversation is complete, it is logical to develop schemes for estimating it while the conversation unwinds.

The research presented here can be extended in many ways. First, the techniques presented here can be applied to more and larger benchmark datasets. Moreover, more adaptive schemes for tracking the emotional dynamics of a conversation can be developed, perhaps as a variant of LMS or based on the spectrum of short time Fourier transform. Additionally, domain transfer methodologies can be used in order to discover and apply affective patterns among conversations. Another possible line of research would be to predict candidate influential accounts when they have not yet changed the affective course of a conversation they participate to. Furthermore, the evolution dynamics of the set of accounts deemed as influential should be investigated. Also, mechanisms tracking the emotional evolution of a conversation based on its cultural content or its topic should be developed. Finally, we would like to ascertain whether a single account can become influential by following a certain methodology which could involve making posts of specific emotional content on discussions of already high emotional potential.

**Acknowledgements** This article is part of project Tensor 451, a long term research initiative whose primary objective is the development of novel, scalable, numerically stable, and interpretable tensor and higher order analytics.

## References

1. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of Twitter data. In: LSM, pp 30–38
2. Barbosa L, Feng J (2010) Robust sentiment detection on Twitter from biased and noisy data. In: COLING, pp 36–44
3. Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. In: NIPS, pp 585–591
4. Boiy E, Moens M (2009) A machine learning approach to sentiment analysis in multilingual Web texts. Information Retrieval 12(5):526–558
5. Bollen J, Mao H, Pepe A (2011) Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: 5th International Conference on Weblogs and Social Media
6. Bühler T, Hein M (2009) Spectral clustering based on the graph p-Laplacian. In: ICML, ACM, pp 81–88
7. Cheng J, Danescu-Niculescu-Mizil C, Leskovec J (2014) How community feedback shapes user behavior. In: 8th (ICWSM)
8. Chikersal P, Poria S, Cambria E (2015) Sentu: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In: 9th International Workshop on Semantic Evaluation (SemEval@NAACL-HLT), pp 647–651
9. Choudhury MD, Gamon M, Counts S (2012) Happy, nervous or surprised? classification of human affective states in social media. In: 6th International Conference on Weblogs and Social Media (ICWSM)
10. Drakopoulos G (2016) Tensor fusion of social structural and functional analytics over Neo4j. In: IISA, IEEE
11. Drakopoulos G, Kanavos A, Mylonas P, Sioutas S (2017) Defining and evaluating Twitter influence metrics: A higher order approach in Neo4j. SNAM 71(1), DOI 10.1007/s13278-017-0467-9
12. Drakopoulos G, Stathopoulou F, Kanavos A, Paraskevas M, Tzimas G, Mylonas P, Iliadis L (2019) A genetic algorithm for spatiosocial tensor clustering: Exploiting TensorFlow potential. Evolving Systems DOI 10.1007/s12530-019-09274-9
13. Hasegawa T, Kaji N, Yoshinaga N, Toyoda M (2013) Predicting and eliciting addressee's emotion in online dialogue. In: 51st ACL Annual Meeting, pp 964–972

14. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences 454(1971):903–995
15. Kafeza E, Kanavos A, Makris C, Vikatos P (2014) T-PICE: Twitter personality based influential communities extraction system. In: International Congress on Big Data, IEEE, pp 212–219
16. Kan Z, Shea JM, Dixon WE (2012) Influencing emotional behavior in a social network. In: American Control Conference (ACC), pp 4072–4077
17. Kanavos A, Perikos I, Vikatos P, Hatzilygeroudis I, Makris C, Tsakalidis A (2014) Conversation emotional modeling in social networks. In: 24th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp 478–484
18. Kanavos A, Perikos I, Vikatos P, Hatzilygeroudis I, Makris C, Tsakalidis A (2014) Modeling retweet diffusion using emotional content. In: Artificial Intelligence Applications and Innovations (AIAI), pp 101–110
19. Kempter R, Sintsova V, Musat CC, Pu P (2014) Emotionwatch: Visualizing fine-grained emotions in event-related tweets. In: 8th International Conference on Weblogs and Social Media (ICWSM)
20. Kim S, Hovy EH (2004) Determining the sentiment of opinions. In: 20th COLING
21. Kim S, Bak J, Oh AH (2012) Do you feel what I feel? social aspects of emotions in twitter conversations. In: 6th International Conference on Weblogs and Social Media (ICWSM)
22. Kim Y, Mesbahi M (2005) On maximizing the second smallest eigenvalue of a state-dependent graph Laplacian. In: Proceedings of the American Control Conference, IEEE, pp 99–103
23. Kyriazidou I, Drakopoulos G, Kanavos A, Makris C, Mylonas P (2019) Towards predicting mentions to verified Twitter accounts: Building prediction models over MongoDB with keras. In: WEBIST, pp 25–33, DOI 10.5220/0007810200250033
24. Leskovec J, Adamic LA, Huberman BA (2007) The dynamics of viral marketing. ACM Transactions on the Web (TWEB) 1(1)
25. Liu B, Zhang L (2012) A survey of opinion mining and sentiment analysis. In: Mining Text Data, Springer, pp 415–463
26. Liu W, Wang J, Kumar S, Chang SF (2011) Hashing with graphs. Tech. rep., Google
27. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: International Conference on Language Resources and Evaluation (LREC)
28. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2):1–135
29. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: Sentiment classification using machine learning techniques. In: ACL Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 79–86

30. Quercia D, Ellis J, Capra L, Crowcroft J (2012) Tracking “gross community happiness” from tweets. In: Computer Supported Cooperative Work (CSCW), pp 965–968
31. Roberts K, Roach MA, Johnson J, Guthrie J, Harabagiu SM (2012) Empatweet: Annotating and detecting emotions on twitter. In: 8th International Conference on Language Resources and Evaluation (LREC), pp 3806–3813
32. Sandryhaila A, Moura JM (2013) Discrete signal processing on graphs. IEEE transactions on signal processing 61(7):1644–1656
33. Sandryhaila A, Moura JM (2014) Big data analysis with signal processing on graphs. IEEE signal processing magazine 31(5):80–90
34. Smola AJ, Kondor R (2003) Kernels and regularization on graphs. In: Learning theory and kernel machines, Springer, pp 144–158
35. Sul H, Dennis AR, Yuan LI (2014) Trading on twitter: The financial information content of emotion in social media. In: 47th Hawaii International Conference on System Sciences (HICSS), pp 806–815
36. Suttles J, Ide N (2013) Distant supervision for emotion classification with discrete binary values. In: 14th International Conference in Computational Linguistics and Intelligent Text Processing (CICLing), pp 121–136
37. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment in short strength detection informal text. Journal of the American Society for Information Science and Technology 61(12):2544–2558
38. Tumasjan A, Sprenger TO, Sandner PG, Welpe IM (2010) Predicting elections with twitter: What 140 characters reveal about political sentiment. In: 4th International Conference on Weblogs and Social Media (ICWSM), pp 178–185
39. Turney PD (2002) Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp 417–424
40. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S (2012) A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. In: ACL System Demonstrations, pp 115–120
41. Wang W, Chen L, Thirunarayan K, Sheth AP (2012) Harnessing twitter ”big data” for automatic emotion identification. In: International Conference on Privacy, Security, Risk and Trust (PASSAT) and International Conference on Social Computing (SocialCom), pp 587–592
42. Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP), pp 347–354
43. Xavier UHR (2013) Sentiment analysis of hollywood movies on twitter. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp 1401–1404
44. Zamparas V, Kanavos A, Makris C (2015) Real time analytics for measuring user influence on Twitter. In: ICTAI