

Received March 20, 2020, accepted April 3, 2020, date of publication April 10, 2020, date of current version May 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987207

# On the Statistical and Temporal Dynamics of Sentiment Analysis

MARGARITA RODRÍGUEZ-IBÁÑEZ<sup>1</sup>, FRANCISCO-JAVIER GIMENO-BLANES<sup>2</sup>,  
PEDRO MANUEL CUENCA-JIMÉNEZ<sup>3</sup>, SERGIO MUÑOZ-ROMERO<sup>3,4</sup>,  
CRISTINA SOGUERO-RUIZ<sup>3</sup>, AND  
JOSÉ LUIS ROJO-ÁLVAREZ<sup>3,4</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Business, Universidad Rey Juan Carlos, 28933 Madrid, Spain

<sup>2</sup>Department of Communication Engineering, Universidad Miguel Hernández, 03202 Elche, Spain

<sup>3</sup>Department of Signal Theory and Communications, Telematics, and Computing Systems, Universidad Rey Juan Carlos, 28933 Madrid, Spain

<sup>4</sup>Center for Computational Simulation, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: José Luis Rojo-Álvarez (joseluis.rojo@urjc.es)

This work was supported in part by the Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación, through a research project FINALE under Grant TEC2016-75161-C2-1-R, in part by the KLINILYCS under Grant TEC2016-75361-R, in part by the FEDER, in part by the Young Researchers through a Research and Development Project under Grant 2020-656, in part by the Community of Madrid, in part by the Rey Juan Carlos University, in part by the KERMES under Grant TEC2016-81900-REDT, and in part by the MAPAS from Spanish Government under Grant TIN2017-90567-REDT.

**ABSTRACT** Despite the broad interest and use of sentiment analysis nowadays, most of the conclusions in current literature are driven by simple statistical representations of sentiment scores. On that basis, the generated sentiment evaluation consists nowadays of encoding and aggregating emotional information from a number of individuals and their populational trends. We hypothesized that the stochastic processes aimed to be measured by sentiment analysis systems will exhibit nontrivial statistical and temporal properties. We established an experimental setup consisting of analyzing the short text messages (tweets) of 6 user groups with different nature (universities, politics, musicians, communication media, technological companies, and financial companies), including in each group ten high-intensity users in their regular generation of traffic on social networks. Statistical descriptors were checked to converge at about 2000 messages for each user, for which messages from the last two weeks were compiled using a custom-made tool. The messages were subsequently processed for sentiment scoring in terms of different lexicons currently available and widely used. Not only the temporal dynamics of the resulting score time series per user was scrutinized, but also its statistical description as given by the score histogram, the temporal autocorrelation, the entropy, and the mutual information. Our results showed that the actual dynamic range of lexicons is in general moderate, and hence not much resolution is given within their end-of-scales. We found that seasonal patterns were more present in the time evolution of the number of tweets, but to a much lesser extent in the sentiment intensity. Additionally, we found that the presence of retweets added negligible effects over standard statistical modes, while it hindered informational and temporal patterns. The innovative Compounded Aggregated Positivity Index developed in this work proved to be characteristic for industries and at the same time an interesting way to identify singularities among peers. We conclude that temporal properties of messages provide with information about the sentiment dynamics, which is different in terms of lexicons and users, but commonalities can be exploited in this field using appropriate temporal digital processing tools.

**INDEX TERMS** Sentiment analysis, machine learning techniques, sentiment dictionaries, social networking, public opinions, Twitter, entropy, autocorrelation, sentiment score.

## I. INTRODUCTION

New technologies have brought in the very last few years new and efficient ways of working, as well as new and

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal<sup>1</sup>.

very straightforward communication methods. The benefits of these technologies as a form of communication are very well known, and instant messaging is boosting the socialization of comments, ideas, and concepts, almost in real time. Sometimes impulsive and some others well sustained, but always interactive and inexpensive, these social media

communications have been pointed as a potential source to identify and measure the underlying sentiment of people about tagged companies, persons, or ideas. Sentiment analysis could be emerging as a solid, interesting, and relevant tool to consolidate individuals themselves, but also their moods, opinions, and feelings [1], [2]. Traditional research in this area used to incorporate extensive and effort intensive social polls and quits. But nowadays, new technology offers a brand-new and powerful novel way of doing so, together with the favor of general affection for social media, websites, forums, blogs, and microblogs, among other sources [3]. Sentiment Analysis has proven to be a remarkable instrument in different research areas and for diverse applications, including examples such as capturing public opinion and investor sentiment about finance issues, stock movements, products, or companies [4], [5]. These efforts are challenging both for researchers and investors [6], [7], as the forecast of stock-price has been a continuum goal for decades, [8], [9], and current always-on availability of social media references seems to be bringing into the scene an unprecedented amount of new available information for its possible evaluation.

But this is not all the landscape. In addition, specific studies have focused on particular and technical aspects of this phenomenon. As an example, the relationship among social media and financial markets has been analyzed through the specific case of a tweet sent from a hacked Associated Press Twitter account [10]. Another specific could be seen on how Twitter broadcasting power has been scrutinized in relations to Corporate Social Responsibility, and market prices [11]. So, it is in the general consensus that Twitter sentiment and stock price evolution are related, in particular, on how the volume of tweets appears to be related to the abnormal growth in the daily return of a stock over a certain period, rather than to feelings and price [12]. Also in this direction, a weighted sentiment measure was recently constructed by using tweet messages from a financial community in Twitter, which allowed the researchers to conclude that this community is a robust predictor of financial markets [13].

The present work was proposed in order to proceed in this direction and to scrutinize the statistical and temporal properties of sentiment when measured in social media, such as in the example of Twitter. We evaluated the nontrivial statistical and temporal properties of the relevant and generally available methods hereafter described for sentiment quantification through an aggregating and benchmarking approach from a number of individuals, industries and their population trends. In particular, we analyzed the existing variables using different statistical views, including histograms distribution, autocorrelation of time-dependent statistical abstract variables, time-domain evolution of all variables, the entropy of significant variables to evaluate the amount of relevant existing information, and the mutual information among said variables. This work analyzed not only the complete set of tweets of the selected users, but also the differential behavior, including and excluding retweets, using a number of different lexicons, as well as the inter-industry and intra-industry.

Additionally in order to consolidate and uniquely quantify the positivity attached to a certain user, a novel Compounded Aggregated Positivity Index is presented and evaluated.

Under this perspective, and even although a vast amount of experiments, databases, and algorithms have already been described, no systematic approach exists, to our knowledge, allowing us to benchmark the different strategies, lexicons, and semantic analysis techniques in the field of sentiment analysis. Therefore, in this work we conjecture whether these available lexicons, databases, or techniques, especially all those labeled from an emotional perspective, could eventually drive a coherent and unbiased handler to qualify or quantify the consolidated underlying mood or sentiment behind all intertwined communications. We understand that a positive response could lead to a wide number of applications, as there are already different commercial and free tools offering the companies graphical and numerical perspective in terms of their social media exposure. Results revealed interesting patterns whether they were general, specific for certain industries and also singular conduct of certain users within industries.

This paper is structured as follows. Section II reviews the background and the literature in this field, including both the general literature as well as the relevant techniques, approaches, and databases used. Section III describes the methods and the experimental setup. Sections IV and V incorporate the experiments with their results, and the conclusions, respectively.

## II. BACKGROUND

A wide diversity of published studies on the subject is present in the literature about the sentiment analysis topic. This is especially true recently, revealing the potential and topicality of the subject in many fields. This kind of analysis has been devoted, for instance, to consumer goods [14], [15], to brand recognition [16], to characterize political sentiment [17], [18], or to even real-time television scheduling of tv-programming [19], among others. But the area of stock markets and companies evaluation is probably the most intensively analyzed, as the titanic economic environment surrounding this specific topic encourages any potential new perspective that may provide tools or support for investor business cases. In this area, exogenous information such as news, political decisions, or customers purchasing expectation, has been historically proven to affect in short- or medium-term valuations of companies. Therefore, it is not uncommon to find that studies in this area are mostly focused in stock price, and especially in an attempt to forecast price and volume evolution of ultimate share value. As a consequence of this extensive coverage, we can find the corresponding bias on literature review, specially in terms of examples and references.

In an attempt to make this section independent from the final target pursued by authors in sentiment analysis, it is focused on techniques, tools, and databases used in the sentiment analysis process. For that reason, it is structured

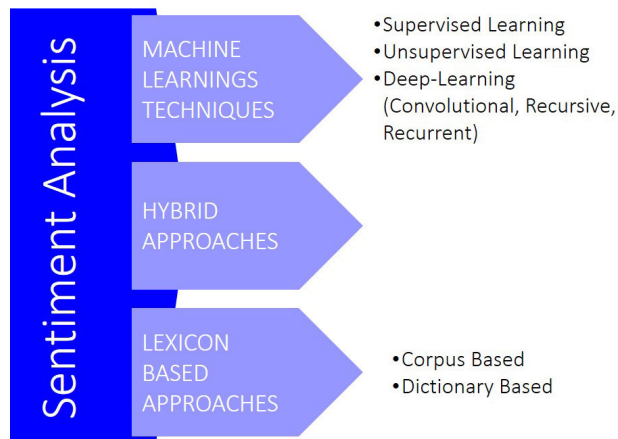


FIGURE 1. An structured description of sentiment analysis.

according to different approaches shown on literature for sentiment analysis, where we review the most common applications using Machine Learning Techniques, Lexicon-Based Techniques, and Hybrid Techniques.

We can define Sentiment Analysis as the required process to identify the underlying attitudes, opinions, or moods, expressed in a certain text data or aggregation of a number of organized characters. Sentiment Analysis could be either quantified by a number in a certain range, or categorized according to the kind of standard subjective statements, such as positive, negative, or neutral. In order to build up a comprehensive description of the different perspectives shown in the literature, we follow the review in [20], which organized the different possible strategies in three main collections, namely, Lexicon-Based, Machine Learning-Based, and Hybrid Approaches (see Fig. 1 for details). But not all three are seamlessly represented in research efforts, and just a few studies have combined Lexicon-Based together with machine learning methods, and achieved relatively better performance [21].

### A. MACHINE LEARNING TECHNIQUES

Entering into what would probably be the most promising area of techniques for sentiment analysis, we should mention that machine learning techniques are commonly used nowadays to build models that can predict sentiment over pieces of text. A number of machine learning techniques have been adopted in this ground.

First, Natural Language Processing (NLP) is an interdisciplinary research field focused on enabling computers to understand and process human language input. NLP is used to parse written texts to infer their syntax and semantics [22]. Second, Case-Based Reasoning (CBR) is designed to solve a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation [23]. Third, Artificial Neural Networks (ANN) are computational models inspired by the structure and elements of natural or biological neural systems. The structure of an ANN is

designed to learn from the data that is presented to it and the output it gives. Using pre-labeled data, the error between the output and the input is minimized through an iterative process. Once the error reaches a certain low threshold, the network is ready to generalize over new data sets [24].

Finally, Support Vector Machines (SVM) correspond to supervised learning methods, and their principle is to improve the generalization capability of learning by seeking the minimum structural risk. As a result, they can obtain the model based on a limited training set and ensure lower error levels in the test, yielding acceptable statistical results even with lower size of statistical samples or with high-dimensional input spaces [25], [26].

### B. LEXICON-BASED TECHNIQUES

Lists of words and their associated sentiment score are commonly referred to as sentiment lexicons, and they are widely used in sentiment analysis [27]. A number of different approaches have been described in the literature, and they could be classified according to: (i) The process followed to create the lexicon, i.e., whether it is manual or automatic; (ii) The elements used for classification, which could be based on single words, n-grams, phrases, or even just certain kinds of words such as adjectives; (iii) The classification model, depending on whether it works with discrete levels of polarity or gradual scales. This strategy often has the problems that the sentiment dictionary doesn't contain enough sentiment words, omits some field sentiment words, polysemic sentiment words and words' polarity. In this perspective we find some current and meaningful contribution for the sentiment recognition of the comment texts [28], [29]. The authors used different methods, some based on a naive Bayesian classifier (to determine the field of the text in which the polysemic sentiment word is), an extended sentiment dictionary and the design of sentiment score rules. While others authors build a Feature Ensemble Model (FEM) and a Convolutional Neural Model (CNN) for tweets containing fuzzy sentiment [30].

### C. HYBRID APPROACHES

Hybrid approaches consist of using both machine learning and lexicons combined, to provide a better understanding of Sentiment Analysis. They are based on the idea that the complementary interleaving of these two very well-known methodologies may eventually lead to a better categorization and benchmark. We present and describe here some representative examples.

A good example of this could be the design of ANN for learning task-specific word embeddings in other NLP tasks [31]. In the same direction, other authors proposed Sentiment-Specific Word Embedding (SSWE) models. In [32], the authors extracted sentiment embeddings from tweets with positive and negative emoticons as distant-supervised corpora without any manual annotation. They verified the effectiveness of sentiment embeddings by applying them to three sentiment analysis tasks.

Empirical results showed that sentiment embedding outperformed context-based embedding on several benchmark datasets of the tasks. This study differs from many others in the field as they used word embedding to capture word similarities in terms of sentiment semantics.

Another example can be the use of linguistic features to detect the sentiment inside Twitter messages [33]. They used three corpora of Twitter messages. For development and training, they used the Hashtagged Dataset (HASH), which they compiled from the Edinburgh Twitter corpus, and the Emoticon Data Set (EMOT) from <http://twittersentiment.appspot.com>. For evaluation, a manually annotated dataset was used, which was produced by the ISieve Corporation (ISIEVE). A different approach in this area tried to automatically collect a corpus for sentiment analysis and opinion mining on Twitter. The corpus contained 300 000 text posts [34]. In this work, an SVM was used to build a sentiment classifier that was able to classify positive, negative, and neutral sentiments for a document, using a previously proposed procedure [35]. Results indicated that this technique was more efficient than previously proposed methods.

One of the newest model of hybrid sentiment analysis is called model-SLCABG, which is based on the sentiment lexicon and combines Convolutional Neural Network (CNN) and attention-based Bidirectional Gated Recurrent Unit (BiGRU) [36]. They showed that this model can effectively improve the performance of text sentiment analysis in the online purchase satisfaction. They used the sentiment lexicon to enhance the sentiment features in the reviews. Then they extract the main sentiment features and context features in the reviews on e-commerce platforms by the CNN and GRU network. And finally they classify the weighted sentiment features.

### III. METHODS AND STATISTICAL DYNAMICS

#### A. TIME SERIES NOTATION

The message event generation and their repetitions can be viewed as temporal random processes, since they can be defined as point-processes. We propose here the following simple signal model. Let  $S(t, \phi)$  denote a stochastic process defining the  $M$ -multidimensional time evolution of the sentiment of a population of  $M$  individuals with respect to some topic  $\phi$ . Let  $s(t, f)$  denote the observable fluctuations from this sentiment evolution, obtained by sampling at a set of times  $t_j$ , and possibly at different users, this is,

$$s(t, f) = \sum_{j=1}^{N_m} s_j(f) \delta(t - t_j), \quad (1)$$

where  $f$  denotes the observable entity that is used as a proxy for the sentiment topic  $\phi$ ,  $t_j$  is the time instant where a sentiment measurement is made through a given measurement instrument, and  $s_j$  is this measurement. If we assume that the measurement instrument is a mathematical transformation made on the text content of a short message, then we can

denote globally this transformation as  $\Gamma$ , so that we can express

$$\hat{s}_j(f) = \Gamma(m_j|f, \ell), \quad (2)$$

where  $m_j$  denotes the text message selected as measured entity, and  $\ell$  denotes the lexicon the operator is linked to. In this case,  $t_j$  is trivially given by the time instant the message is recorded from the environment, and  $N_m$  is the number of messages.

#### B. TIME SIGNALS, M-MODE, AND AUTOCORRELATION

The presented simple notation and signal model aims to make easy to follow the statistical and temporal process dynamics by using well-known Stochastic-Process Principles [37]. While convenient as notation, this makes evident two limitations at this point. First, point process  $s(t, f)$  is hard to analyze in this form, as most of Stochastic-Process Theory works on temporal evolution according to a sample period  $T_s$  providing a regularly-sampled discrete-time series from the continuous time original series [38]. The second limitation, directly due to the sampling process, is that populational information is aggregated and heavily condensed, hence mixing different basic properties of the underlying multidimensional sentiment process in a single measurement.

Aiming to partially overcome these limitations, the following transformations can be done on the original set of messages  $m_j$  collected in a time interval  $(t_1, t_{N_m})$ . By choosing a time integration window  $T_w$ , it will play the role of effective time sampling to transform the point process into a discrete-time process. The total number of hits per window can straightforwardly obtained, given by

$$s_n^f[n] = \sum_{t_j \in \mathcal{T}_n} 1, \quad (3)$$

where  $\mathcal{T}_n \equiv ((n-1)T_w, nT_w)$ . Whereas this does not represent a sentiment measurement itself, it can be seen as a measure of the sampling intensity from the messages when used as an instrument. Also, an average sentiment measurement for each time window can be obtained as follows,

$$s_m^f[n] = E_{t_j \in \mathcal{T}_n}(s(t, f)), \quad (4)$$

and also the standard deviation of the sentiment measurement can be similarly obtained, and it is denoted by  $s_s^f[n]$ . Hence,  $s_n^f[n]$  and  $s_m^f[n]$  measure indirectly both the most usual sentiment on the population about topic  $\phi$  and the controversy for this topic in that same population.

Throughout the experiments, we are going to need some tool for representing in parallel and on an intuitive way the statistical and temporal dynamics of several topics. For this purpose, we use the previously proposed M-mode representation [39]. As a notation example for a multidimensional signal  $x(t)$ , we can define an M-mode using the dimension number  $f$ , so that  $\Gamma(t, f)$  denotes this representation in general terms. For instance, if the second dimension is the index corresponding to the  $j^{\text{th}}$  topic, with  $j = 1, \dots, N_t$  and  $N_t$  the

number of scrutinized topics, we can define the M-mode of the sampled process as a index-temporal and simultaneous representation of the realizations of the stochastic process as follows,

$$X(j, t) = x_j(t), \tag{5}$$

where the convention of capital notation is used for the M-mode of the represented signal.

If  $x_j(t)$  denotes in general the non-sampled sentiment measurement signals used in this work, then  $x_j(t + \tau)$  represents its replica displaced by delay  $\tau$ , and its autocorrelation function [40] is calculated with

$$r_{x_j x_j}(\tau) = \int_{-\infty}^{\infty} x_j(t + \tau)x_j(t)dt. \tag{6}$$

The M-mode of the autocorrelation functions of the  $j^{th}$  topic is then just given by

$$R_{x_j}(j, \tau) = r_{x_j x_j}(\tau), \tag{7}$$

The described elements allow us to give a set of descriptions of well-known principles in Stochastic Process Theory for the sentiment analysis based on message measurements.

**C. ENTROPY AND MUTUAL INFORMATION**

The study of the probability density function (*pdf*) of a discrete-time random process  $s[n]$  (mathematically mentioned here as  $p_f$ ) is relevant because it quantifies the certainty and the uncertainty of the results obtained in each realization. In our case, this function is defined as:

$$p_f(\rho) = pdf(\hat{s}_j(f)), \tag{8}$$

where  $\hat{s}_j(f)$  denotes the full set of scores for each message  $m_j$  of entity  $f$ , and then,  $\rho$  is the resulting independent variable of all possible score values. The value of this function is positive throughout the domain, and it represents the overall distribution of its statistical density, aggregated through time and therefore independently from it. Here again we use the convention of capital to jointly represent the estimated *pdf* of a related set of entities, as denoted by:

$$P(f, \rho) = p_f(\rho). \tag{9}$$

After the introduction of probability density notation, the analysis of the amount of information attached to an entity, based on Entropy ( $H$ ), and the Mutual Information (MI) among pairs of them, is proposed in this section. A relevant number of  $H$  measurements have been described in literature since the work published by Pincus in 1991 [41], and after Shannon’s publication in 1948 [42]. In this paper, we evaluate the Entropy ( $H(j)$ ), that is understood as the weighted average value of the amount of information of a certain variable. In our case we can rephrase it, as the amount of information of all the sentiment scores of the messages related to an entity in a period of time. Mathematically it

can then be analytically expressed in the following terms according to notation developed in this paper:

$$H(j) = - \sum_{\rho} P(f, \rho) \log_2(P(f, \rho)), \tag{10}$$

where  $\rho$  stands for the all unique possible sentiment scores,  $f$  is the entity, and  $P$  is the discrete probability density function of the score  $\rho$  for entity  $f$ . As a consequence,  $H(f)$  is represented by a one-dimensional array of the Differential  $H$  evaluated in the time slot under study for every considered entity.

The second information driver that we evaluate in this paper is the MI, which is a measurement of the cross-dependence of two paired random variables, and it specifies the amount of information that can be obtained from one random variable from knowing the other [43]. Whereas MI can be defined for continuous and discrete random variables, we work here with the discrete version. In this case,  $S_u$  and  $S_v$  denote the two random variables corresponding to two entities measured  $\hat{s}_u[j]$  and  $\hat{s}_v[i]$  for a given lexicon and sentiment analysis method, which can be divided into  $N$  and  $M$  states, respectively. Then, the MI of these random variables is given by

$$MI_{u,v} = \sum_{i,j}^{N,M} p(\hat{s}_u[j], \hat{s}_v[i]) \log \frac{p(\hat{s}_u[j], \hat{s}_v[i])}{p(\hat{s}_u[j])p(\hat{s}_v[i])}, \tag{11}$$

where  $S_u = \{\hat{s}_u[j], j = 1, 2, \dots, N\}$  and the corresponding  $S_v = \{\hat{s}_v[i], i = 1, 2, \dots, M\}$  are said states of the observed time signals.

It can be seen that this expression is symmetric in  $S_u$  and  $S_v$  and always positive, and is equal to zero if and only if  $S_u$  and  $S_v$  are independent [44]. Its units are bits, as far as we use the base-2 logarithm for  $H(j)$  calculations. Note that if  $S_u$  and  $S_v$  are independent, then the knowledge about  $S_u$  does not provide any information about  $S_v$ , and hence  $MI_{u,v} = 0$ . On the other hand, if  $S_u$  is a deterministic function of  $S_v$  and  $S_v$  is a deterministic function of  $S_u$ , then all the information conveyed by  $S_u$  is provided by  $S_v$ , and vice versa, and in this case, the MI will equal to the  $H$  of  $S_u$ , which is also equal to the  $H$  of  $S_v$ .

**D. SELECTED LEXICONS**

Several well-known lexicons have been used to perform the experiments described in Section IV. As shown in the experiments, different lexicons can result in slightly different characterizations of the sentiment distributions under study. For example, some lexicons are capable of a richer dynamic range than others. At the same time, we analyze the basic statistical properties of the distributions to determine their dependence with respect to the specific lexicon used.

For the purpose of data acquisition and processing, we developed a tweet ingestion tool that is capable of downloading tweets, up to a specified number of messages or to a specific date limit, from various user accounts or reacting to them via mentions. After tweets have been downloaded,

sentiment analysis scoring is performed using the desired lexicon. The sentiment analysis score distribution is then statistically characterized. The sentiment analysis performed by the tool is based on a simple bag-of-words model, which only takes into account individual words and their associated scores. There is no provision for n-grams, full phrases, or any other type of multi-word combinations. In addition, scoring is based on a discrete rating in a range associated to each word, and terms usually associated with a positive (negative) sentiment are associated to a positive (negative) score.

Most of the lexicons used had to be preprocessed in order to comply with the sentiment scoring model we just described. The following is a description of the used lexicons, as well as the transformations we needed to apply to be able to use them with our model.

First, lexicon AFINN-165-EN, (hereinafter referred to as *AFINN*) [45], consists of 3,382 English words to which a numeric sentiment score was manually assigned. Its score range is [-5,5]. Second, the EmoLex NRC emotion lexicon list (*NRC*) [46] is a word-based lexicon that was created using a crowdsourcing approach via Mechanical Turk. On it, 14,182 English words were annotated according to ten non-exclusive binary categories, including the eight emotions from Plutchik's wheel of emotions [47]. For our study, we selected classes named "positive" and "negative", and ignored words that were classified as neither. There is no strength or intensity gradation in this lexicon, so we arbitrarily assigned score value -3 to all negative samples, and +3 to all positive ones. This resulted in 5,636 words that could be used in our model. Third, the Sentiment Composition Lexicon for Negators, Modals, and Degree Adverbs (*SCL*) was published in [48]. Sentiment associations were also obtained manually through crowdsourcing using the Best-Worst Scaling annotation technique [27]. It consists of single words and multi-word phrases, which were created by combining single words with modifiers such as negators, auxiliary verbs, degree adverbs or a combination of those. Since our model requires single words, we discarded all multi-word phrases. We also converted the score scale (a real number between -1 and 1) to the discrete range we use. Fourth, the SemEval-2015 English Twitter Lexicon (*SemEval*) [49] is another crowd-sourced lexicon consisting of 1,515 terms (including neutral ones), with the particularity that terms are drawn from English Twitter and include general English words, misspellings, hashtags, and other categories frequently used in Twitter. It includes negated expressions that were excluded from this study. Finally, the *SentiStrength* lexicon [50], [51] was included here. Like *SemEval*, *SentiStrength* also uses terms and expressions drawn from social media, including emoticons. It assigns sentiment scores to word prefixes, like "abhor\*", in the belief that any words that use the same prefix, such as "abhorrence" or "abhorrent" in this example, will share the same sentiment score. Since our corpora were extracted from Twitter we kept all emoticons, but transformed prefixes into single words using an English dictionary. The result was a list of 7,126 non-neutral terms,

with the particularity that it includes more negative terms than positive ones. This bias is possibly a consequence of the dictionary expansion we performed on prefixes.

Though, it can then be inferred from these brief descriptions, lexicons and their structure can be very much intertwined with language. There are efforts to produce sentiment analysis lexicons in a vast amount of languages including Arabic [27] or Chinese [52]. Our study focuses solely on English-language lexicons.

#### IV. EXPERIMENTS AND RESULTS

In this section we present the results of the implemented statistical and temporal dynamics obtained with a custom-designed sentiment analysis tool, which was created to support the tasks of retrieving the tweets for a set of users, and the search result. In Fig. 2, an example of the summary results for a selected username in the front-end of the application can be seen. The application supports in the same framework the score calculation according to a set of selected lexicons for each tweet, as well as the analysis of the statistical and temporal dynamics of the sentiment scores per user, and the information-theory analysis within and among groups.

A total of 4 different sets of experiments were carried out. In Experiment I, we show an overview of statistical and dynamic results for a specific group of usernames related to universities. In Experiment II, we scrutinized the effect of not considering retweets in the same statistics evaluated in the previous one. In Experiment III, we evaluated the effect of using different dictionaries and their impact on the statistical representations. In Experiment IV, we analyzed the behaviour of user groups with different nature, namely, universities, singers, media, political leaders, technological companies, and financial companies (see Table 1 for details).

##### A. SENTIMENT INFORMATION AND DYNAMICS

This first experiment was performed for the selected universities shown in Table 1. Only user counting on an appropriate volume of messages over a single week were considered for the study. We used the application program interface provided for public use by Twitter<sup>TM</sup> which allows compiling the messages for the past 7 days, and only English messages were considered. The dictionary *AFFIN* - 165 was used for this first analysis, as it was specifically developed for microblogs. The messages were acquired during the same week in all cases to avoid week-dependent deviations.

Figure 3 illustrates the results for the selected group of universities. Panel (a) depicts time evolution of scores for all users. Note that the tweets have been retrieved for each username during seven consecutive days, from Monday 14<sup>th</sup> to Sunday 20<sup>th</sup> of October. Accordingly, prior to time-slot consolidation, in this first panel we represented a non-uniform sampled signal. A strong variability of the time series can be observed, as well as some visual symmetry around the zero score. A different view, now using time slot consolidated and discretized score terms, is given by the score histograms

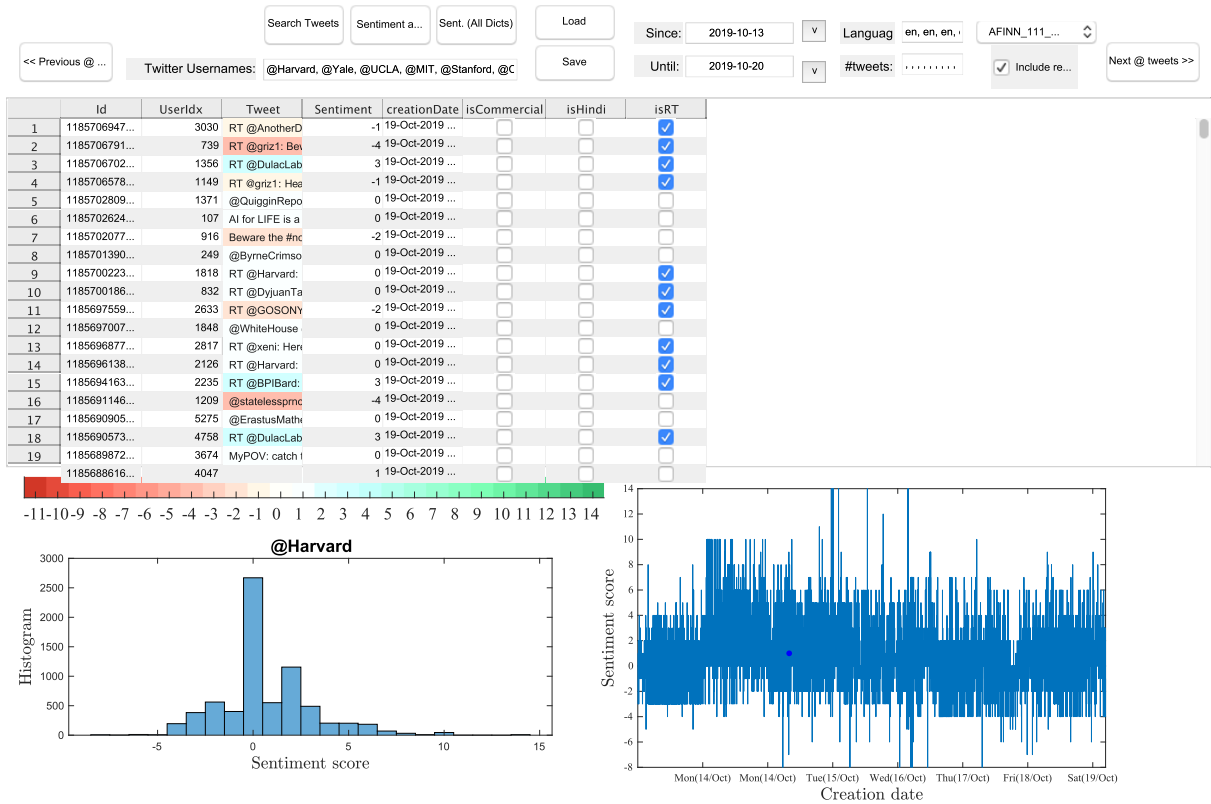


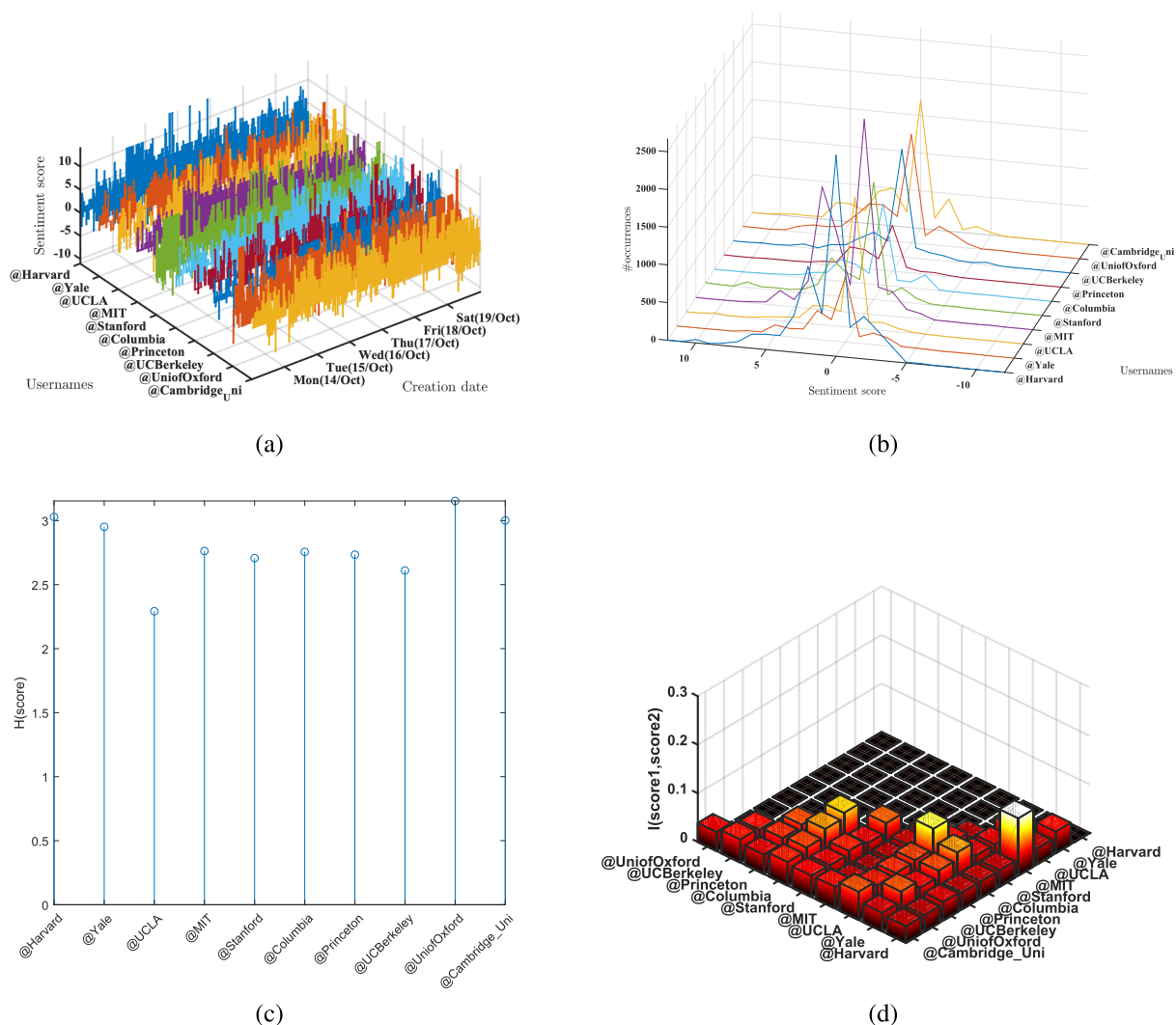
FIGURE 2. Front end of the tool developed for the message compiling, storing, quality control, and topic analysis.

TABLE 1. Usernames for different nature groups, namely, technological, universities, singers, media, political leaders, and financial companies.

Universities	Media	Singers	Technological	Politicians	Financial companies
@Harvard	@BBCBreaking	@ladygaga	@facebook	@theresa_may	@BankofAmerica
@Yale	@cnnbrk	@edsheeran	@apple	@realDonaldTrump	@WellsFargo
@UCLA	@WSJ	@eltonofficial	@amazon	@EmmanuelMacron	@Citi
@MIT	@washingtonpost	@taylorswift13	@SamsungMobile	@eucopresident	@AmericanExpress
@Stanford	@business	@shakira	@Huawei	@jeremycorbyn	@DeutscheBank
@Columbia	@thesun	@justinbieber	@Google	@netanyahu	@RBS
@Princeton	@daily_mirrow	@rihanna	@Tesla	@David_Cameron	@Chase
@UCBerkeley	@itvnews	@jtimberlake	@Dell	@GeorgeHWBush	@RBC
@UniofOxford	@skynews	@katyperry	@Microsoft	@mbachelet	@MandT_Bank
@Cambridge_Uni	@telegraph	@britneyspears	@IBM	@FedericaMog	@StanChart

represented in Panel (b) for each user. The convergence of the histograms was checked for the number of tweets retrieved in the seven day period. Multimodal distributions can be generally observed for all users, with a commonly shared high level mode at zero score, which represents a common baseline. On the other hand, additional positive and negative local maxima are easily found in almost all cases, exhibiting particular magnitudes for individual users. Positive and negative modes are quite close to the zero level. In cases where no positive or negative modes are found, a rather continuous decaying evolution of the number of messages sharing the

magnitude is exemplified. Panels (c) and (d) offer a different perspective as they show respectively,  $H(j)$ , understood as the complementary information of the each user, and the MI among users. In terms of information, users in panel (c) share a common profile as values are generally ranging a short span from 2.7 to 3, whereas one single exception is found in @UCLA. In terms of MI, only significance could be found among @Harvard vs @MIT, and @UCLA vs @Columbia. First coupled universities relation could be probably found on the geographic proximity, but that is not the case for @UCLA and @Columbia. The existence of MI



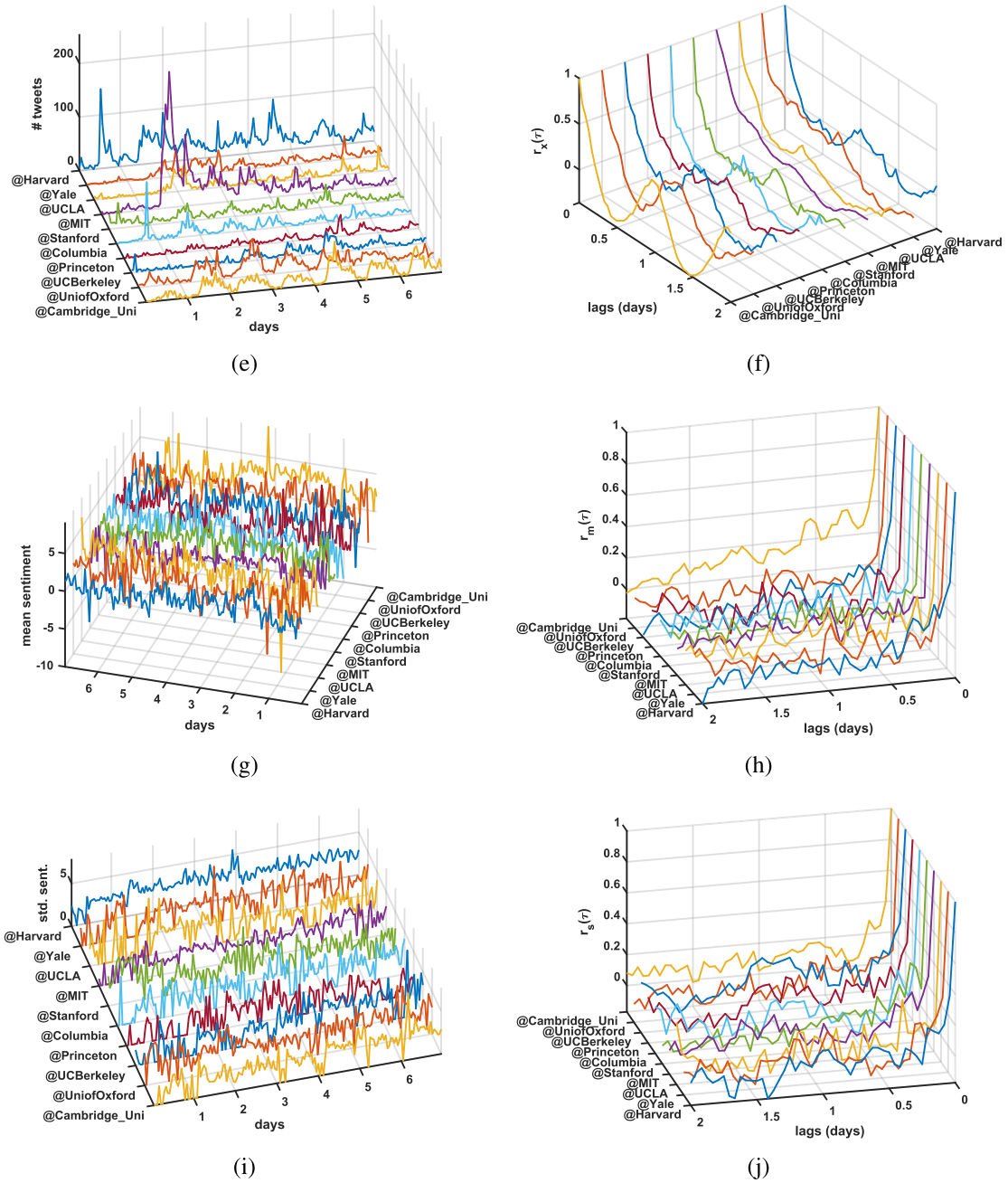
**FIGURE 3.** Experiment I: statistical and temporal dynamics of sentiment scores in the group of universities, for the 10 selected users and 7 days: (a) Time evolution of sentiment score for the retrieved tweets; (b) Histogram of the sentiment score for each user; (c,d)  $H$  per user and MI among users in the group.

when retweets are considered showed isolated specific couples ( $@MIT$  and  $@Harvard$ , and  $@UCLA$  and  $@Columbia$ ). The relation of the first set could be related to the geographical proximity but the existence of this second group require of a deeper analysis of specifics. This effect is present both, with or without retweets, but special evidence is visible when excluding retweets and  $@Yale$  showed a relevant relation with a number of entities. This might be related to the tractor effect of Yale as one of the top universities in USA. Further analysis would required for a better understanding of these trends.

Figure 4 shows a graphical representation of the key statistics analyzed in this paper. For reader convenience, panels in this figure are consecutively named after the previous figure, to emphasize the figures related to the same reality or experiment. In particular, the first panel of Figure 4 is named (e) following the last panel (d) of the previous figure. This structure

is repeated later on with some of the subsequent experiments, for an easier comparison among them. Panel (e) incorporates the aggregated number of tweets over hourly slots. This representation offers the intensity perspective in the communication. It can be generally appreciated a daily circadian rhythm with effective minimums over the night and relative lower intensity over the weekend. An example of this behavior can be found in  $@Cambridge\_Uni$ . Clear exceptions to this daily cycle are certain days in  $@MIT$  or  $@Harvard$ . In  $@MIT$ , there is a relevant increase in the number of tweets after day 1, which starts to decrease after one day. Periodicity of this intensity is presented in panel (f) through the autocorrelation function. As expected from previous results, a maximum is visible in one- and two-days delays. In this panel, we find not such a strong seasonality for  $@Harvard$  and  $@MIT$ . For this last user, the autocorrelation decays progressively, showing that it is statistically a non-persistent time processes. For the





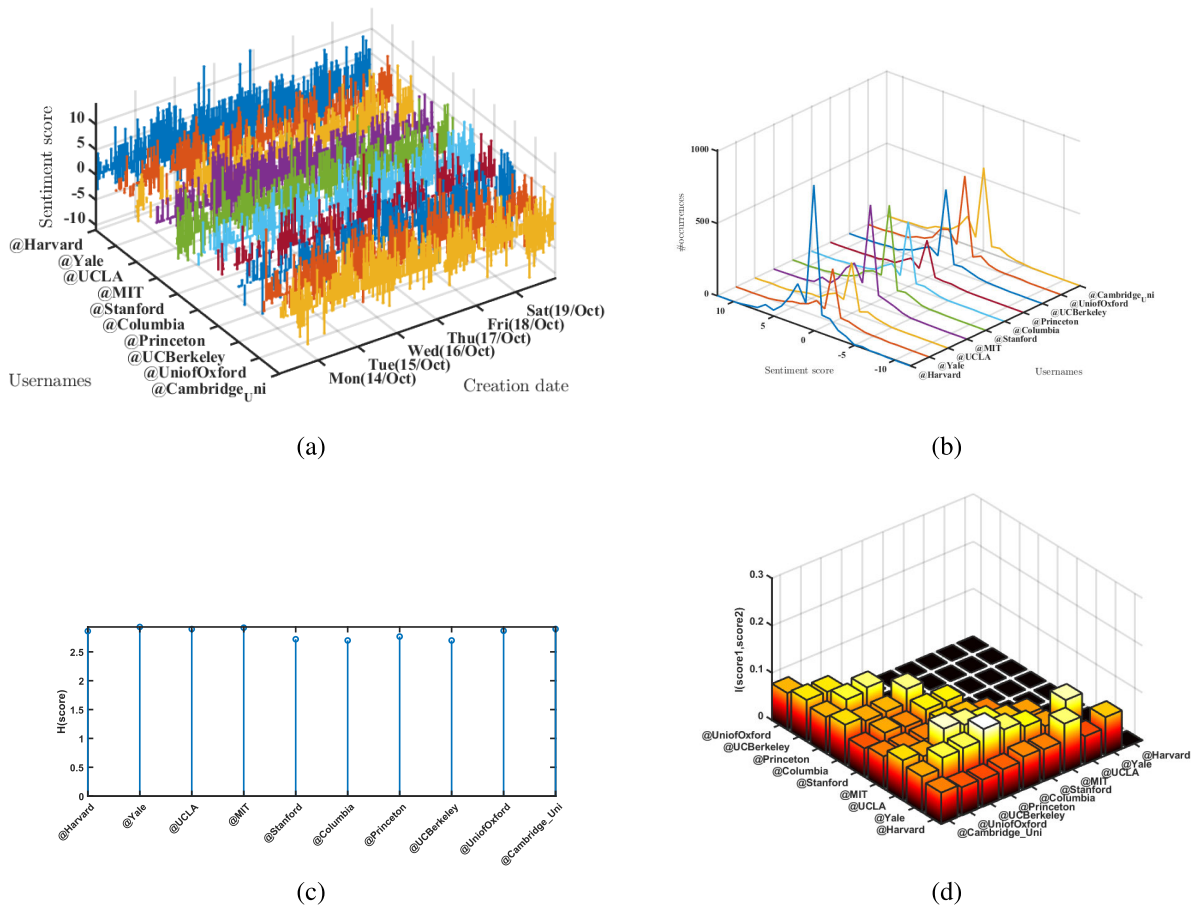
**FIGURE 4.** Experiment I (cont). statistical and temporal dynamics of sentiment scores in the group of Universities, for the 10 selected users and 7 days: (e,f) # of tweets when considering one-hour time windows and their autocorrelation functions; (g,h) Mean and autocorrelation function of the sentiment score; (i,j) Standard deviation and autocorrelation function of the sentiment score.

rest of the users, the peaks are present in day one and, to a lesser extent, in the second day, showing somehow a lack of short-time memory regardless of the rationale behind these effects.

Panels (g) and (i), included in Fig. 4, show the mean and the standard deviation of the sentiment score for all users during the week under study. Panels (h) and (j) represent the

autocorrelation of the mean and standard deviation, respectively. In this four panels no significant visual pattern can be appreciated. No clearly visible stationary in the autocorrelation, nor definite shapes in the absolute values.

The exclusion of retweets, the use of different lexicons, and the different types of user groups, will allow us further comparative analysis in the next experiments.



**FIGURE 5.** Experiment II: statistical and temporal dynamics of sentiment scores in the group of universities without retweets, for the 10 selected users and 7 days: (a) Time evolution of sentiment score for the retrieved tweets; (b) Histogram of the sentiment score for each user; (c,d)  $H$  per user and MI among users in the group.

**B. DYNAMIC ANALYSIS EXCLUDING RETWEETS**

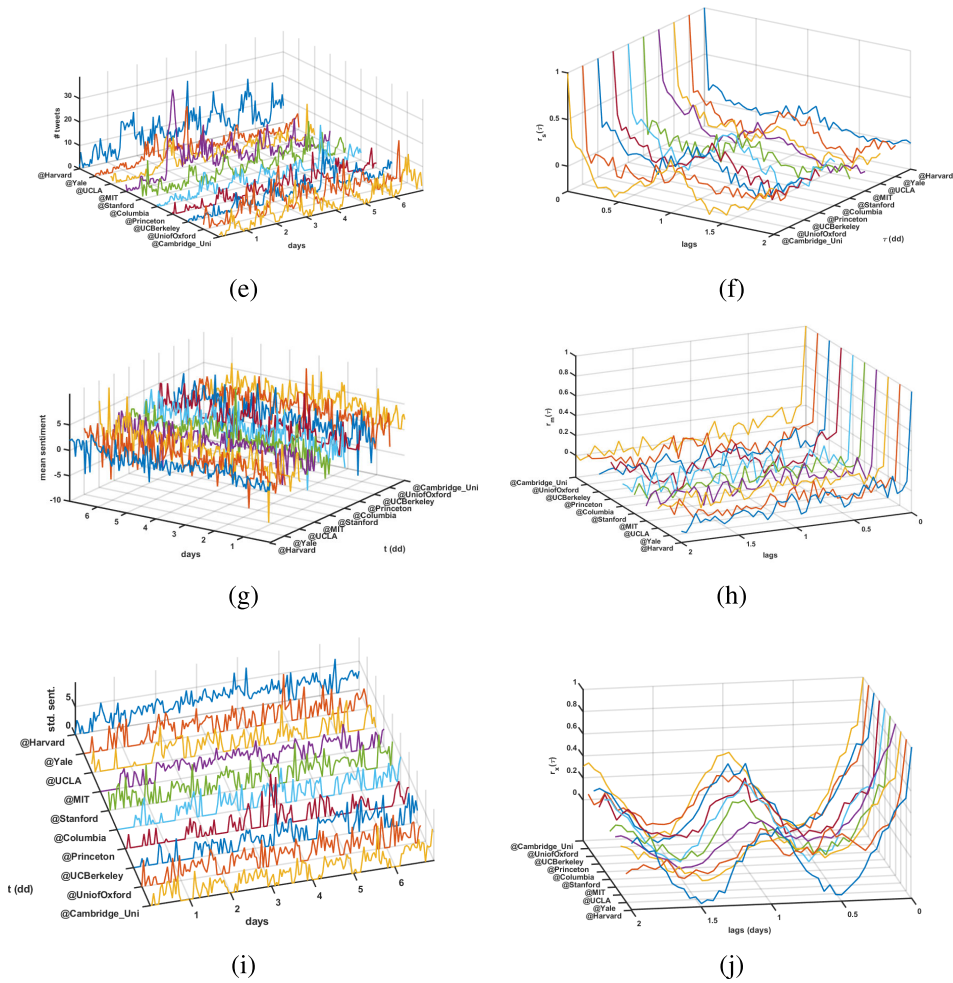
This second experiment has been conducted for the same group of university usernames, but without considering retweets in the message acquisition. Several conclusions can be drawn when the statistics and the dynamics are compared with and without retweets, which emerge from the time evolution of the scores.

Figure 5 (b) of Experiment II shows positive monomodality for @Harvard (it has reduced negative peaks). In all the universities, the positive peak was found to be patent, and exceptionally no clear peak was found on the positive side. There were a negligible number of peaks on the negative side, and when they did appear, they did not reached a comparable level versus their positive counterparts. There were also a reduced number of bimodalities present in the same side, and when found, they were very mild on the negative side. A comparison of the histograms of the sentiment score for each user with and without retweets, in Figures 3 (b) and 5 (b), shows that the number of occurrences is much lower when retweets are not considered (see Panel (b)), as trivially expected, but the convergence and stabilization of histograms still occurs.

In this second experiment, the positive or negative local modes are present but to a lesser extent visible in relation to zero modes growing. An exception to this behavior is @Harvard, where positive modes appear not to be so affected by the exclusion of the retweets.

As far as  $H$  is concerned, Figure 5 (c) of Experiment II shows much more homogeneous  $H$  across the users, exhibiting higher values in those users that exhibited lower values when retweets were considered. Visible previous drops of entropies, especially sharp for the @UCLA in Figure 3 (c), are not that perceptible. A close comparison of these figures shows, for example, relevant reduction of information for @UCLA and @UCBerkeley. On the contrary @Oxford, @Harvard and @Cambridge, did not change much in terms of  $H$  when the retweets were not considered. Therefore, from a statistical perspective, incorporating retweets can have an even impact in terms of  $H$  for different users, but in some cases the reduction of  $H$  in academic institutions is noticeable.

The MI in Figure 3 (d) for Experiment I and Figure 5(d) for Experiment II is very regular. Especial interest can be



**FIGURE 6.** Experiment II (cont). Statistical and Temporal Dynamics of sentiment scores in the Group of Universities without retweets, for the 10 selected users and 7 days: (e,f) # of tweets when considering one-hour time windows and their autocorrelation functions; (g,h) Mean and its autocorrelation function of the sentiment score; (i,j) Standard deviation and its autocorrelation function of the sentiment score.

noticed in Figure 3 (d), where over the general reduced values of MI, two couples of users, @Harvard vs @MIT, and @UCLA vs @Columbia, manifest a singular MI. This appears to be isolated effects, and not generalizable for the other user combinations. By the same token, in panel (d) of Figure 5 we can appreciate higher values in general and much wider cross relation as far as MI is concerned. Although the previous relation among the two paired users is still present, the case of Yale requires specific attention, as it now arises with relevant MI relationship with a number of universities.

In Experiment II, Figure 6 on panel (f) depicts the autocorrelation of the intensity, showing the same seasonal daily effect described when retweets were considered, although in this second case it appears to incorporate a higher amount of variability or noise. On the contrary, when it comes to analyze the autocorrelation of the standard deviation, once the retweets are removed on panel (j), it shows a clear seasonal-daily effect, which was previously hidden. Note that, in both autocorrelation analysis when retweets are not

considered, a second peak is materialized on the second day. These results show not only a circadian pattern (daily) in the behavior of the number of tweets (intensity), but also and even more clearly in the standard deviation of the score that stands. It is remarkable that this effect is not visible at all in the mean of the sentiment, but it is in the standard deviation in this case.

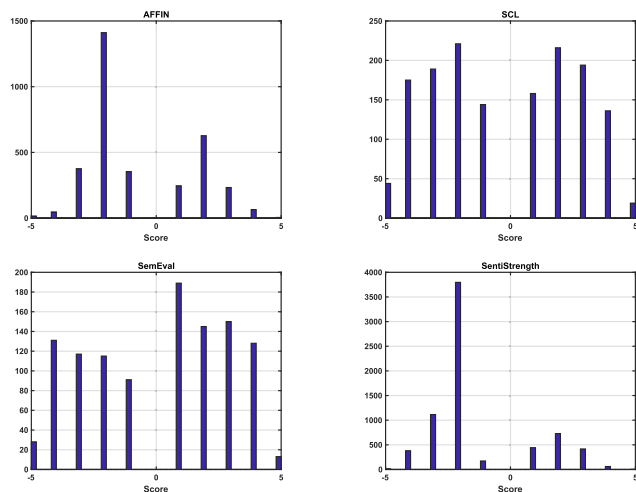
Figure 6 (i) of Experiment II shows replicable patterns throughout the days, including lack of signal during the night, reflecting also the inactivity also visible in panels (e) and (g), which is coherent with the time when people sleep and they are inactive. For example, in @Cambridge we can find a flat effect caused by the night. This clearly seasonable profile solely does not justify the results of panel (j), as this night-day reality was also visible in Figure 4, but the corresponding panel (j) does not suggest it.

### C. IMPACT OF DICTIONARIES AND SCORES

We further aimed to scrutinize the impact of the different dictionaries and scores used as the basis for the sentiment

**TABLE 2.** Experiment III: Number of terms (diagonal) and number of shared terms (super-diagonal) for each considered dictionary.

	<i>AFINN</i>	<i>NRC</i>	<i>SCL</i>	<i>SemE.</i>	<i>SentiS.</i>
<i>AFINN</i>	3382	1306	680	359	1775
<i>NRC</i>		5636	936	317	2002
<i>SCL</i>			1496	204	946
<i>SemE.</i>				1107	361
<i>SentiS.</i>					7124



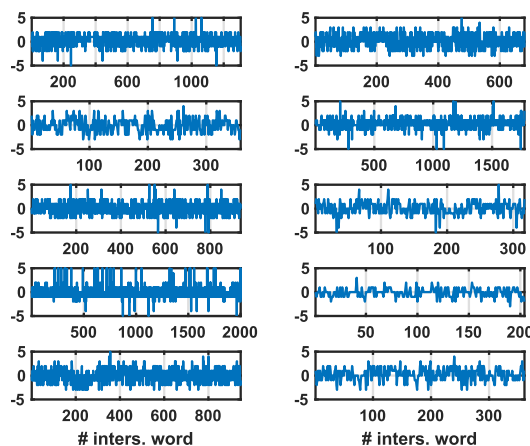
**FIGURE 7.** Experiment III: Histograms for scores in the terms for the dictionaries in our study.

retrieval from short messages. For this purpose, we continued using the same set of universities, without taking into account resent messages.

It is relevant to emphasize not only the different number of terms in the selected dictionaries, but also their common intersections, all of whose countings are shown in Table 2. The largest and the smallest dictionaries are *SentiStrength* and *SemEval*, respectively, in such a way that we could consider as large dictionaries also *AFINN* and *NRC*, as well as reduced dictionary also *SCL*. We can also see that *SentiStrength* is to some extent the dictionary with more shared terms with the others in general, but not with *SemEval*, which turns to be the dictionary with lower number of shared terms with all the others.

We can also analyze the discrete distribution of the scores in each dictionary, as seen in Fig. 7. Both *AFINN* and *SentiStrength* show strong bimodality in the positive and negative selected terms, whereas *SCL* and *SemEval* tend to work with more equalized histograms for positive and negative terms. In all cases, the tails (corresponding to score values of  $\pm 5$ ) are much more under-represented, though in the equalized histograms they are notably more present. It is also interesting to note the preponderance of negative terms in the non-equalized dictionaries.

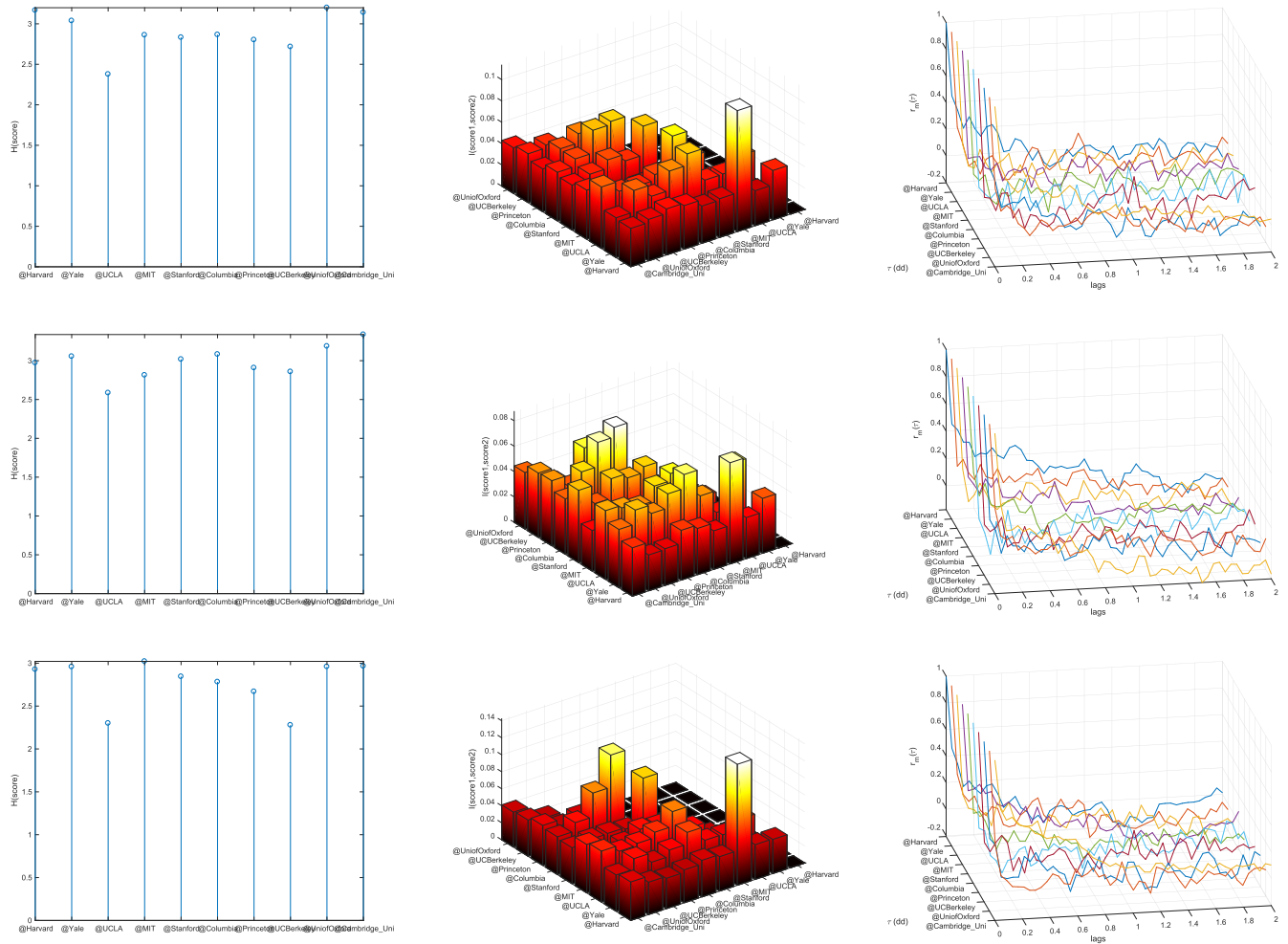
Additional information can be seen when representing the difference of the scores in those shared terms by pairs of dictionaries, as exhibited in Fig. 8, where each panel shows



**FIGURE 8.** Experiment III: differences between scores in the shared terms in pairs of dictionaries, in the same order that super-diagonal elements in Table 2. See text for details.

the non-ordered shared words by pairs of dictionaries in the horizontal axis, and the score difference for each term in the vertical axis. The order of pairs of dictionaries follows the super-diagonal in Table 2 being read by rows. It can be noted that many shared terms do not share a strongly similar score in the rank. Moreover, the visible spikes in these plots (both in the positive and negative directions) often indicate shared terms which even score with opposite sign in different dictionaries.

With all these descriptive statistics of the dictionaries, their scores, and their shared and different information, we now analyze some of the statistical and temporal dynamics as captured upon said dictionaries. Figure 9 shows several panels for  $H(j)$ , the MI, and the mean autocorrelation of representative dictionaries. We checked that the  $H$  profiles were strongly similar for all dictionaries, with some more patent increase in the range for *SentiStrength*, so that this increase in the range could be explained by adding much more words than in the other, rather than by equalizing the positive and negative scores. With respect to MI, it can be observed that there is a different behavior in *SemEval* when compared with all the others that turn out to share a similar profile, and that it consists on an increase of the information shared by different used being more clearly retrieved with the former. Also, *SemEval* is the only one with a trend to yield some more persistence in the mean-score autocorrelation functions for several of the users. In all cases, the autocorrelations of the process itself were strongly similar among all, and mostly dominated by the seasonal components, as seen in the previous experiments. Probably the criteria for selecting the terms in *SemEval* could be the key for the smallest dictionary yielding some more information in terms of MI and of temporal dynamics. However, it seem clear that the tasks performed in the different dictionaries (such as increasing the number of terms or histogram equalization in scores) do not seem to impact much on the statistical and temporal dynamics, and resolution and sensitivity, respectively, could be further pursuit in future dictionaries.



**FIGURE 9.** Experiment III: obtained  $H$  (left),  $MI$  (center), and mean autocorrelation (right), for dictionaries *AFINN* (up), *SemEval* (middle), and *SentiStrength* (down).

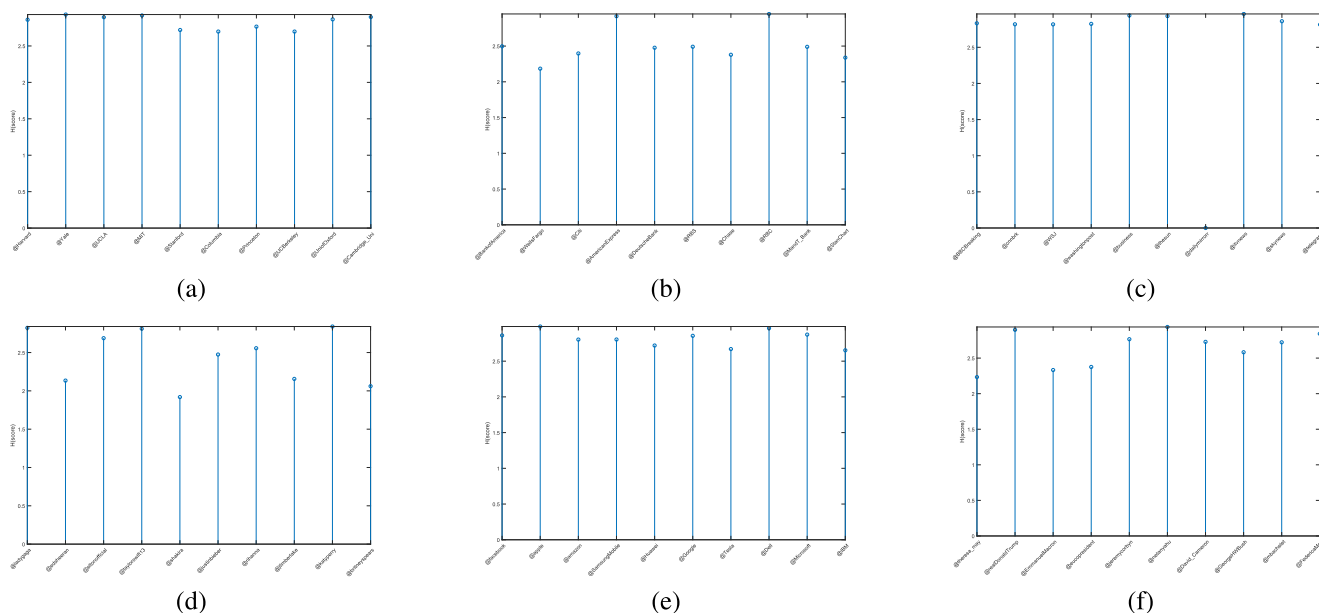
**D. ANALYSIS ON DIFFERENT NATURE GROUPS**

Experiment IV compares users from different environments, allowing us to compare the previously defined statistics and features across different nature groups. For that propose we incorporated over the previously evaluated set, the academic entities, five others industries, namely: (i) Financial Institutions, (ii) Communication Media Companies, (iii) Singers, (iv) Technological and Internet Firms, and (v) Politicians. Each group comprises ten of the most recognizable user-names, to guarantee the necessary minimum amount of messages for statistical significance. The *AFIN* dictionary was considered, and analysis was performed not including retweets.

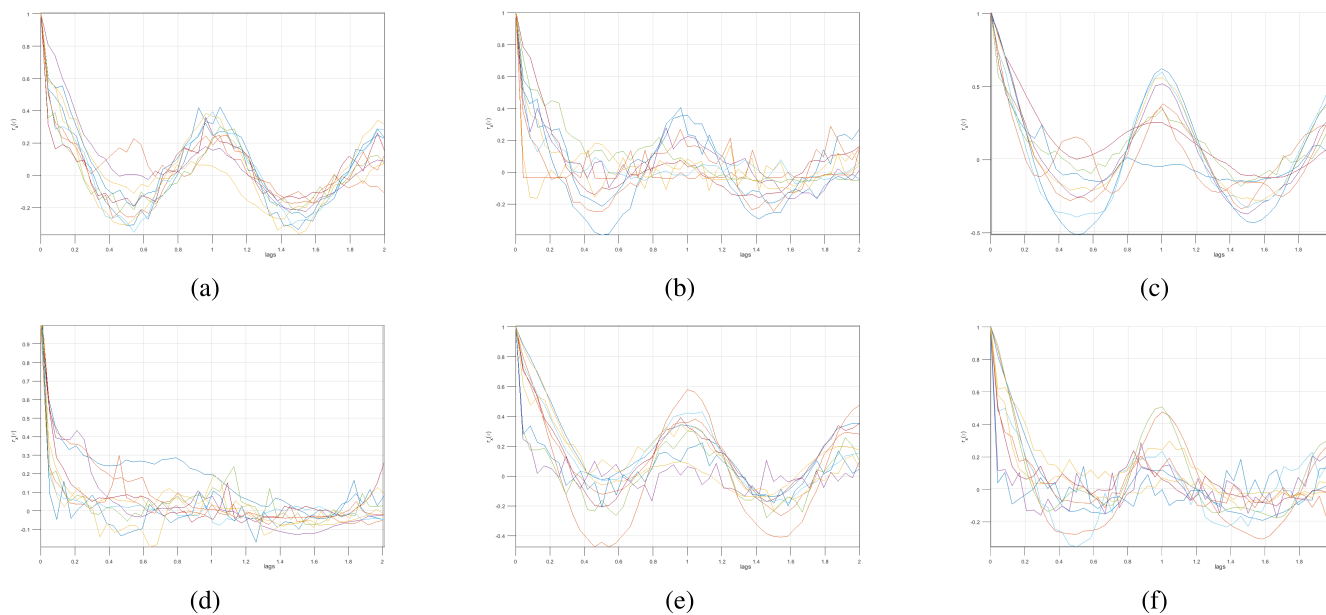
Figures 10, 11, 12, 13 and 14, show a summary of all the key results from this experiment. As the main objective of this experiment is to show the industry focus, figures are represented in a consolidated perspective overlapping smashing the three dimensional M-Modes into, two dimensional figures where curves are overlapped showing the consolidated

and evocated view. From an industry perspective, as far as  $H$  is concerned, results depicted in Fig. 10 show extensive high  $H$  for all users in Academy, Communication Media, and Technological. A different outlook can be observed in the case of Singers such as EdSheeran, Sakira, Tiberlake and Britney Spears, and with the inancial Institutions counting on a visible lower level of  $H$ , with the exception of American Express and RBC.

As seen previously for the academy industry, the auto-correlation of the sentiment score intensity shows a clear repetitive daily profile (see Fig. 11, and this profile is again shared across industries. The only exception to this pattern is the singers industry, where this outline is not visible at all, either in a consolidated approach, or at an individual user level. This pattern seems to be very singular to this particular industry, and it will require special attention in the discussion. A different reality is represented in the autocorrelation of the standard deviation of the sentiment score, where the periodic effect detected visible in the case of universities, seems to be



**FIGURE 10.** Experiment IV: industry comparison of  $H$ , being read by rows: (a) academic entities, (b) financial institutions, (c) media communication companies, (d) singers, (e) technologicals, and (f) politics.

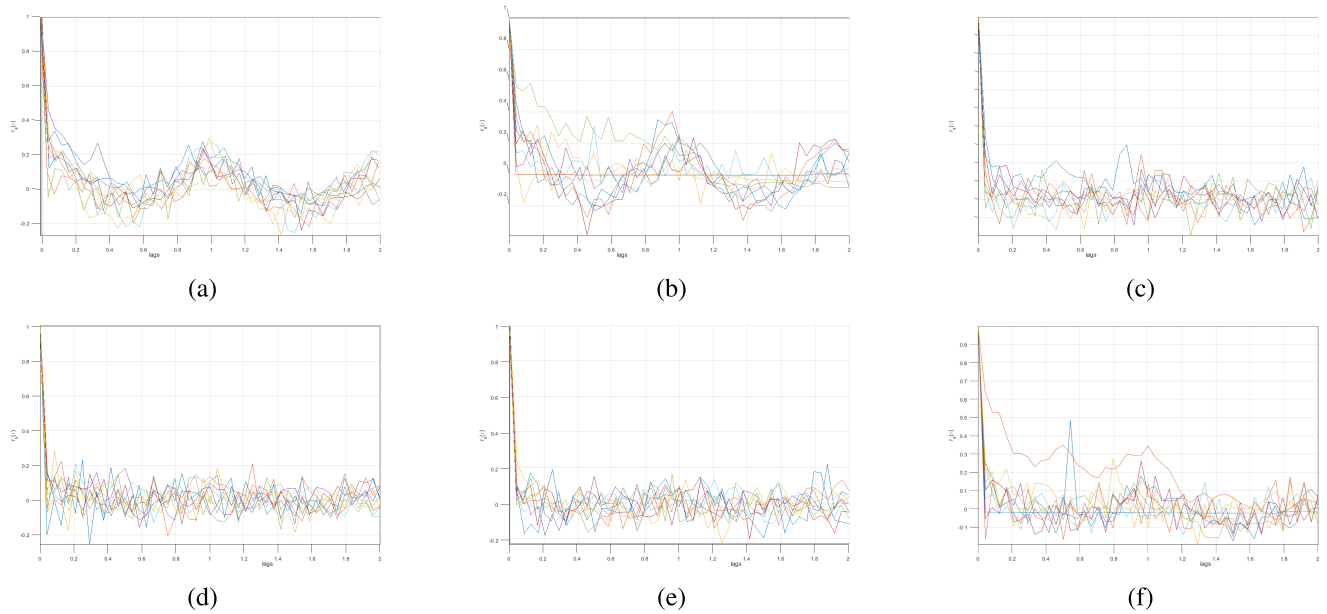


**FIGURE 11.** Experiment IV: industry comparison of Auto-correlation of Message Intensity, being read by rows: (a) Academic entities, (b) financial institutions, (c) media communication companies, (d) singers, (e) technologicals, and (f) Politics.

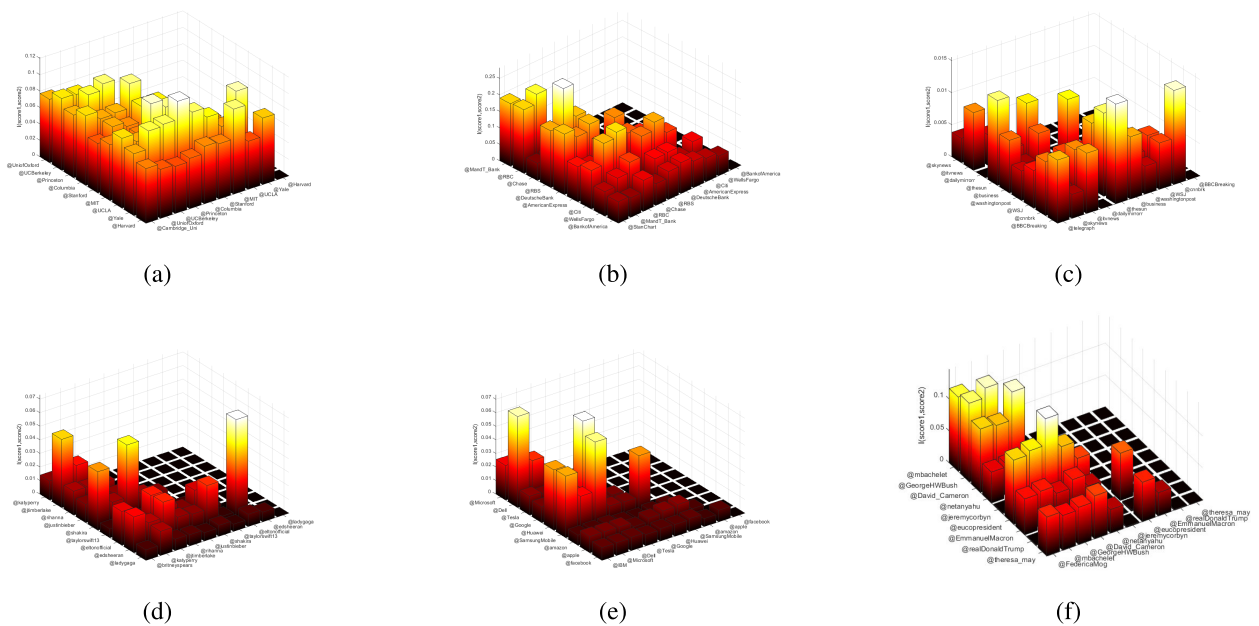
replicable only in the case of financial institutions, but not in the rest of the analyzed industries (See Fig. 12. No reference or representation is devoted to the autocorrelation of the mean of the scores, as no different pattern was found in this representation in any industry compared with the expressed in Experiments I and II.

On the other hand, MI comparative analysis is shown in Fig. 13, which did not offer relevant differences from the previous results for the case of universities. In general terms,

it appears to be industry specific, or more precisely user specific, as no replicable pattern can be observed in this comparison among industries. But special attention requires the case of *@IBM @Dell, and @Huawei* in the Technology sector as they showed relevant MI among peers. A similar situation appears in Politics with four relevant users leading the MI, namely, *@mbachelet, @GeorgeHWBush, @DavidCameron* and *@eucopresident*. Additional discussion and further analysis is required for an adequate interpretation and feature



**FIGURE 12.** Experiment IV: industry comparison of auto-correlation of standard deviation of the score, being read by rows: (a) Academic entities, (b) financial institutions, (c) media communication companies, (d) singers, (e) technologicals, and (f) politics.

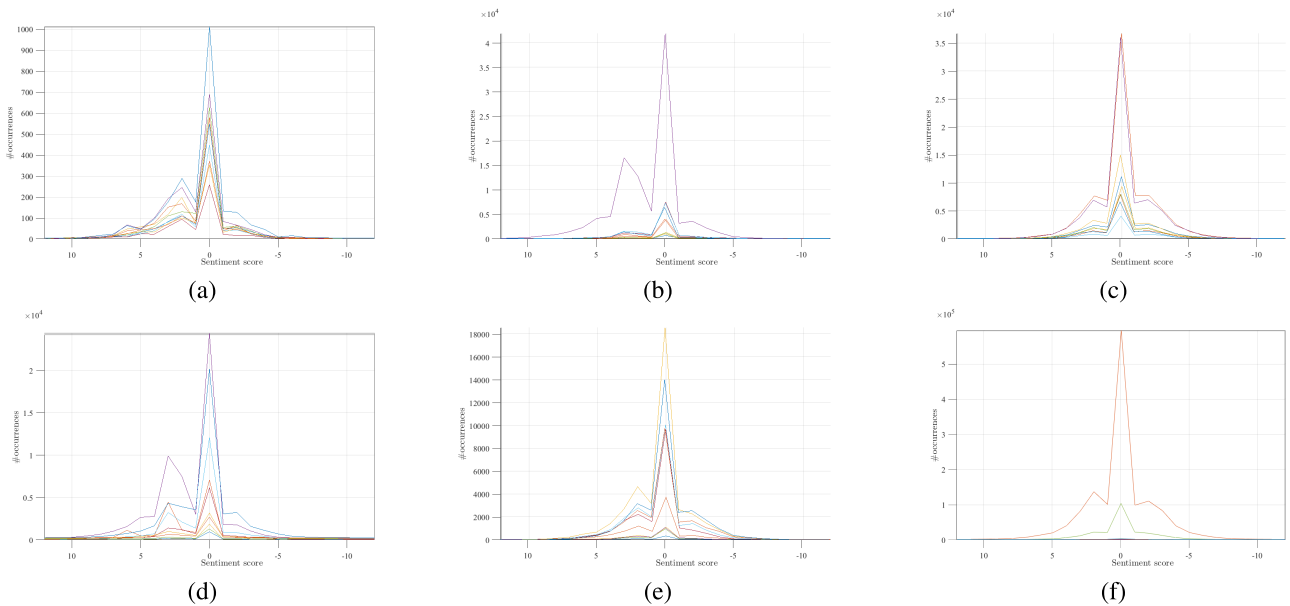


**FIGURE 13.** Experiment IV: industry comparison of MI, being read by rows: (a) Academic entities, (b) financial institutions, (c) media communication companies, (d) singers, (e) technologicals, and (f) politics.

extraction, as in certain industries high values among users could imply significance in terms of cross-relation and sentiment contagion among those users.

Special attention requires the score histograms shown on the first column of panels in Figure 14. This consolidated evocated view shows the presence of the neutral mode on zero that is extremely prominent in all industries and users. Separate behavior can be found in industries when it comes

to lateral modes, although apparently all industries present at least a slightly higher weigh of the positive mode, and in a deeper analysis it turned out not to be so in certain cases. Visual inspection of consolidated histograms offers two separate patterns. The first pattern, especially represented by Academic Institutions and Singers, shows a clear positive mode that happens to appear higher than their negative counterpart, which is in some cases almost non-existent.



**FIGURE 14.** Experiment IV: industry comparison of score histograms, being read by rows: (a) Academic entities, (b) financial institutions, (c) media communication companies, (d) singers, (e) technologicals, and (f) politics.

And a second pattern can be observed in which both sides are more parallel-like, with some individual exceptions.

It is also relevant to mention that, in specific industries, wider lobes in the positive branches are clearly present. These situations are less visible in a first visual inspection, but it might be of much higher significance in a mathematical evaluation. In an attempt to make it visible in this experiment, we evaluated the relation among the weighted positive and negative legs of the histogram for each industry and use, excluding the baseline mode in zero. To do so, we created a new index by dividing the compounded positive side of the histogram over the compounded negative side, excluding zero. For better representation, ten times the decimal logarithm of these values are plotted, converting this non-dimensional into logarithmic units (dB) and highlighting especially the values close to origin. From a mathematical perspective, if we define the histogram of the sentiment score as the probability density function, but this time discretized attending to the different possible values of the sentiment according to our model, the histogram function will be mathematically as follows:

$$h_f(\rho) = p_f[\rho], \tag{12}$$

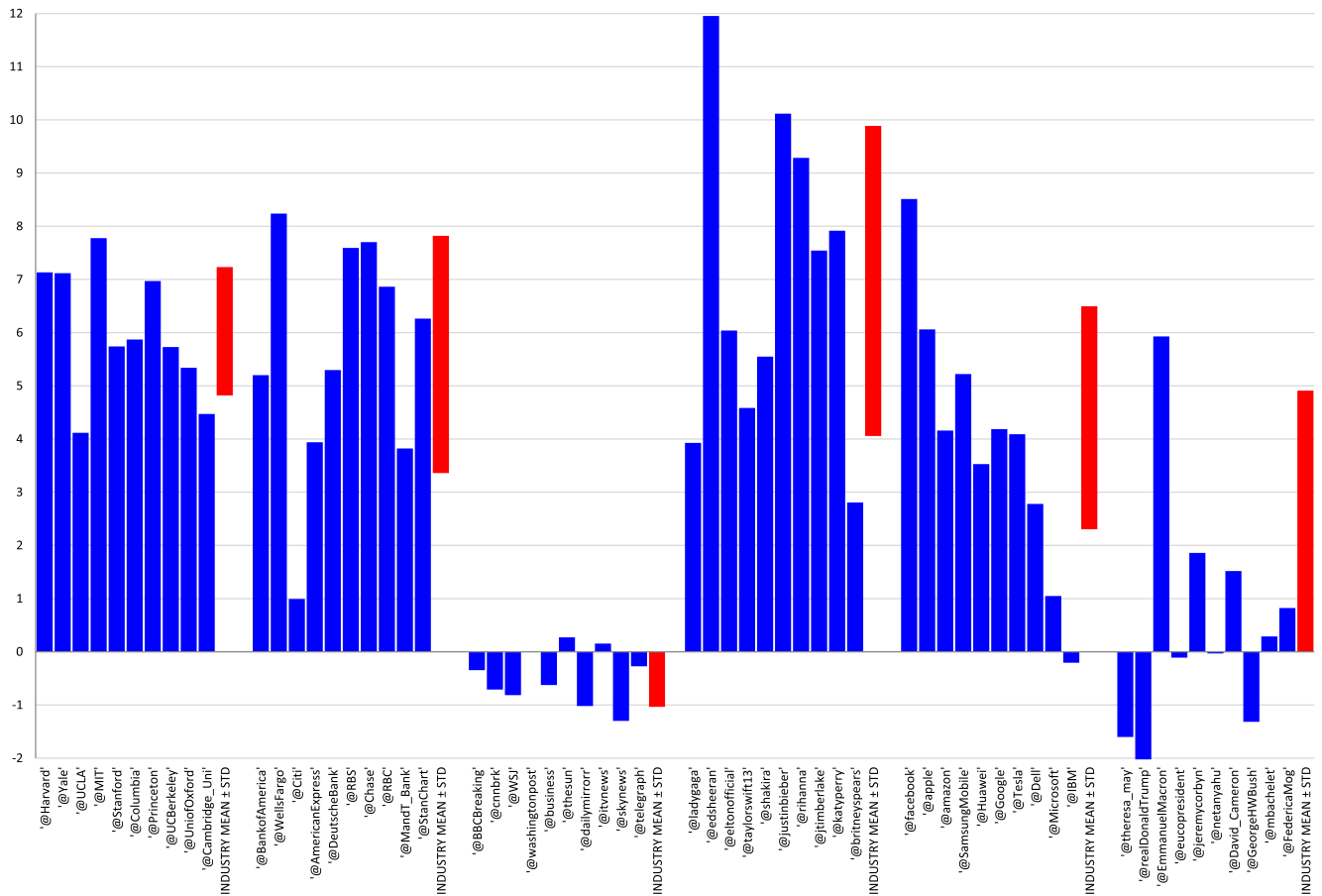
where  $\rho$  are all possible score values, and  $f$  stands for the entity or user under study. Now we can define the *Compounded Aggregated Positivity Index* or *CAPi* as ten times the logarithm of relation among the compounded aggregated positive and negative sentiment, and mathematically could be computed as

$$CAPi_f = 10 \cdot \log \left( \frac{\sum_{\rho>0} (\rho \cdot p_f[\rho])}{\sum_{\rho<0} (\rho \cdot p_f[\rho])} \right) (dB), \tag{13}$$

where  $CAPi_f$  stands for the user  $f$  (in dB units),  $\rho$  are all possible score values, and  $p(\rho)$  is the value of the histogram of the score of the sentiment for  $\rho$ , and in other words the number of messages of the user  $f$  where sentiment score was evaluated as  $\rho$ .

Figure 15 shows the computation results of *CAPi* for all users and industries. According to results, *CAPi* values are very much industries dependent. Academic users tend to keep similarity among all users in the sector, that is expressed with a lower standard deviation of the values obtained. It is generally found an average positive valuation of this index (close to 6 dB). The second set, although with quite relevant positive values, presented a much wider variability of this index. This characteristic of Financial Institutions, is shared by Singers, and Technological Entities. The industry with higher *CAPi* is the Singers, with an average close to 7 dB. In absolute terms the three users with larger index belong also to this industry, almost doubling the average of all sectors. We can mention Lady Gaga close to 12 dB, Justin Bieber with about 10 dB, and Riana about 9 dB. It should be noticed at this point the existence of two sectors with relevant negative values of *CAPi*. This is the case of the Media and Politicians. The first of these two groups has mostly negative values over the sample, where only two of them (The Sun and ITV News) have a slightly greater weight of positive sentiments against the negative ones. The second groups, The Politicians, present a more balanced model between positive and negative, where only in three out of ten cases, the negative weight exceeds the positive. The largest negative values of the index in absolute terms are achieved in this industry. We should highlight the existence of a very wide dynamic range of *CAPi* values in this industry, where one of the users exceeds almost three times





**FIGURE 15.** Experiment IV: CAPI (dB) calculated for all industries and users. In blue the index calculated for the each user in the industry. Industry range, in red, calculated as two times the standard deviation over the mean of all users included in the corresponding industry.

his next counterpart, Emmanuel Macron. Special analysis and discussion is required to evaluate this situation.

According to the experiments carried out in this section, we can summarize hereafter the key findings. On the one hand, the comparison between the analysis with and without retweets has shown that the existence of retweets in some cases slightly emphasizes existing modes, but on the contrary, it hinders the vision of some dynamic patterns of interest. On the other hand, comparing diverse dictionaries, the results have not shown significant differences that justifies intensively enough the use of dictionaries with a greater or lesser number of words. On the other hand, in the comparative analysis of the different industries, wide differences in their statistical characterization with different tools have been observed with intensity. Next section elaborates and discusses in detail these results, arguing on possible justification for them

**V. DISCUSSION**

The main objective of this work has been to evaluate the possibility of characterizing and modeling the sentiment aroused on social media, in an example of Twitter community. The detailed analysis of this reality can offer institutions,

companies, entities, or users themselves, a valuable tool to know the effectiveness of their communication, and not only through digital media but also by any other means, since the sentiment collected could be considered as a consolidation of all aggregated inputs. The aim is then to evaluate the subjective and subconscious consolidated emotional feeling attitude awoken by a user, brand, entity, company, or institution, as expressed directly or indirectly by the interactions on the social media under evaluation throughout the words they used.

We first analyzed ten selected Academic Entities, including direct tweets and with and without retweets. In the benchmarking analysis including and not considering retweets, no relevant information according to our analysis was missing when retweets were not included. However, this second scenario showed that new patterns arose that were not visible when the retweets were present. Hence, we argue that retweets do not add statistical relevant information (apart from the volume of total messages), and they even can statistically hide relevant patterns, which were visible once retweets were removed from the sample-base.

A more detailed view showed, for the histogram of scores excluding the zero-modes, that the presence of retweets

strengthened the peak levels of the existing modes, either they were positive or negative scored. This could be interpreted as if the presence of retweets reinforced the sentiment bias of non-baseline modes, but they do not add additional information as far visual patterns are concerned. In this very same analysis, a reduction in entropy was found for some individual users when retweets were considered. A possible explanation for this reduction, taking into account that this is not a general behavior, could be found to our knowledge, such as the existence of actuator automatic bots, or the relevant number of users just replicating the original tweet but not adding additional personal bias, among others. As a consequence and prior to further analysis of this reality, we could argue that this effect could not a-priori be considered as a value-building element for the model.

Additionally, and due to the fact that the signal extracted as the consolidation of an eventual infinite number of fully independent sentiments are expected to behave as a clean aleatory stochastic process, large values of entropy should not be judged necessarily negative. In summary, and to our consideration, the joint effect of a negligible increment in amplitude of the sentiment modes in the histograms, the hinder of certain periodicity dynamics, and the possibility of a not positive sentiment bias, will suggest the consideration of not including the retweets in the sentiment analysis when scrutinizing the dynamics.

A relevant number of different lexicons, in terms of size, methodology, or even sentiment structuring scores, are present in the literature. In this paper we evaluated and benchmarked only five of them over the Academic Institutions. The tasks performed on the different dictionaries (such as increasing the number of terms or histogram equalization in scores) do not seem to impact much on the statistical and temporal dynamics, and resolution and sensitivity should be further pursuit in future dictionaries. In this setting, we can argue that the use of small databases in terms of number of words, although convenient from a computational stand point, might not be recommendable given that their generalization capabilities could also be strongly limited. So, for analysis across industries we proposed AFIN, a dictionary keeping some balance between computational efficiency and generalization expectations.

Relevant findings were obtained in our cross-industry evaluation. Specifically, daily-circadian behavior was found not only in terms of intensity (number of tweets), but also and even more clearly visible in the Standard Deviation of score, suggesting that there exists a pattern during the day when the tweets are being released, and that the sentiment evoked, although not following this daily behavior in terms of the mean, it clearly does in terms of variability. This might suggest that, although there is no relation over what is the sentiment people show over the day, it is in terms of how disperse it is. General intensity circadian cycles, could be very much explained in a double fold way: First, due to the normal circadian cycle of the people (not tweeting during night and doing it essentially at certain moments in the day);

Second, by the professional activity from official community managers and communications departments. A different and more interesting result is the existence of the circadian cycle in the standard deviation of the score, which requires of much deeper analysis to justify if any sociological perspective could be underling this effect showing that people are more predisposed at certain times of the day to maximize their subjective assessments (feelings). Another possible argument could be that this effect highlights the fact that the communication departments of the academic entities generate, of course positive tweets, and always at the same time of day. Although this would be consistent with the pattern in the standard deviation, it would be not with the fact that this pattern is not reflected in the autocorrelation of the mean.

Considering the relevant multimodal effects presented in the score sentiment histograms, and given the particularity of each one of them, both in industries and among users in the same group, in this paper we propose a new indicator that allows a compounded aggregated sentient index (referred in the paper as *CAPI*) to summarize all effects expressed by the different observers over certain users. This index revealed interesting results that were not visible in visual inspection of histograms, that turned out to be characteristic in some industries. In particular, stands out the negative values and generalized results of this index for Media Communications and Politicians, whereas on the other hand, Universities and Singers stick out in the positive side. Also, at individual user level, manifested index out of the pack compared to their peers.

## VI. CONCLUSION

In conclusion, we consider that it is possible to create indices that allow evaluating the sentiment generated by a brand, entity or individual person, making use of the systems like Twitter and their provided info, through sentiment statistics and dynamics as presented here. On the other hand, considering the difficulties found to justify some findings, we understand that there is much room to continue improving the analytical capacity of this type of techniques by applying new and more sophisticated processing and semantical analysis, as well as by expanding the sample base in terms of the number users incorporated in the study, and of course by widening the temporal scope.

## REFERENCES

- [1] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *8th Int. AAAI Conf. weblogs social media*, May 2014, pp. 216–225.
- [2] W. Fan and M. D. Gordon, "The power of social media analytics," *Commun. ACM*, vol. 57, no. 6, pp. 74–81, 2014.
- [3] M. P. Anto, M. Antony, K. M. Muhsina, N. Johny, V. James, and A. Wilson, "Product rating using sentiment analysis," in *Proc. Int. Conf. Electr., Electron., Optim. Techn. (ICEEOT)*, Mar. 2016, pp. 3458–3462.
- [4] Y. Sun, M. Fang, X. Wang, and S. Diao, "GubaLex: Guba-oriented sentiment lexicon for big texts in finance," in *Proc. 13th Int. Conf. Semantics, Knowl. Grids (SKG)*, Aug. 2017, pp. 25–32.
- [5] T. Mankar, T. Hotchandani, M. Madhwani, A. Chidrawar, and C. S. Lifna, "Stock market prediction based on social sentiments using machine learning," in *Proc. Int. Conf. Smart City Emerg. Technol. (ICSCET)*, Jan. 2018, pp. 1–3.

- [6] Z. Madhoushi, A. R. Hamdan, and S. Zainudin, "Sentiment analysis techniques in recent works," in *Proc. Sci. Inf. Conf. (SAI)*, Jul. 2015, pp. 288–291.
- [7] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Statistical approaches to concept-level sentiment analysis," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 6–9, May 2013.
- [8] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7653–7670, Nov. 2014.
- [9] G. Fung, J. Yu, and H. Lu, "The predicting power of textual information on financial markets," *IEEE Intell. Inform. Bull.*, vol. 1, pp. 1–10, May 2005.
- [10] T. Karppi and K. Crawford, "Social media, financial algorithms and the hack crash," *Theory, Culture Soc.*, vol. 33, no. 1, pp. 73–92, Jan. 2016.
- [11] B. Bradbury and I. Kim, "What if 140 characters can impact your company's market value? Tweeting corporate social responsibility and stock prices in the financial sector," in *Proc. Surfing IT Innov. Wave Amer. Conf. Inf. Syst.*, 2016.
- [12] G. Ranco, D. Aleksovski, G. Caldarelli, M. Gräar, and I. Mozetiä, "The effects of Twitter sentiment on stock price returns," *PLoS ONE*, vol. 10, no. 9, 2015, Art. no. e0138441.
- [13] S. Y. Yang, S. Y. K. Mo, and A. Liu, "Twitter financial community sentiment and its predictive relationship to stock market movement," *Quant. Finance*, vol. 15, no. 10, pp. 1637–1656, Oct. 2015.
- [14] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth, and Shubham, "Reviewer credibility and sentiment analysis based user profile modelling for online product recommendation," *IEEE Access*, vol. 8, pp. 26172–26189, 2020.
- [15] J. Hornikx and B. Hendriks, "Consumer tweets about brands: A content analysis of sentiment tweets about goods and services," *J. Creative Commun.*, vol. 10, no. 2, pp. 176–185, Jul. 2015.
- [16] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, Nov. 2013.
- [17] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated Twitter sentiment," in *Proc. 28th Int. Conf. Hum. Factors Comput. Syst. (CHI)*, 2010, pp. 1195–1198.
- [18] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, "Analyzing political sentiment using Twitter data," in *Proc. Inf. Commun. Technol. Intell. Syst.*, vol. 107, 2019, pp. 427–436.
- [19] Q. Ji, "Exploring the motivations for live posting during entertainment television viewing," *Atlantic J. Commun.*, vol. 27, no. 3, pp. 169–182, May 2019.
- [20] R. Baldania, "Sentiment analysis approaches for movie reviews forecasting: A survey," in *Proc. Int. Conf. Innov. Inf., Embedded Commun. Syst. (ICIECS)*, Mar. 2017, pp. 1–6.
- [21] S. M. Vohra and J. B. Teraiya, "A comparative Study of Sentiment Analysis Techniques," *J. Inf. Knowl. Res. Comput. Eng.*, vol. vol., no. 2, pp. 313–317, 2012.
- [22] R. Besançon, F. M. S. Spaccapietra, and M. Rajman, "Text mining: Natural language techniques and text mining applications," in *Proc. Data Mining Reverse Eng.*, 1998, pp. 50–64.
- [23] Y. Du, W. Wen, F. Cao, and M. Ji, "A case-based reasoning approach for land use change prediction," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5745–5750, Aug. 2010.
- [24] S. Paliwal, S. Kumar Khatri, and M. Sharma, "Sentiment analysis and prediction using neural networks," in *Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2018, pp. 1035–1042.
- [25] Y. Fei, "Simultaneous support vector selection and parameter optimization using support vector machines for sentiment classification," in *Proc. 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Aug. 2016, pp. 59–62.
- [26] W. Zheng and Q. Ye, "Sentiment classification of chinese traveler reviews by support vector machine algorithm," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Appl.*, 2009, pp. 335–338.
- [27] S. Kiritchenko and S. M. Mohammad, "Capturing reliable fine-grained sentiment associations by crowdsourcing and best–worst scaling," in *Proc. Conf. North Amer. Chapter Assoc. for Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 811–817.
- [28] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese text sentiment analysis based on extended sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019.
- [29] J. Wu, K. Lu, S. S. Su, and S. Wang, "Chinese micro-blog sentiment analysis based on multiple sentiment dictionaries and semantic rule sets," *IEEE Access*, vol. 7, pp. 183924–183939, 2019.
- [30] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model," *IEEE Access*, vol. 8, pp. 14630–14641, 2020.
- [31] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, "Sentiment embeddings with applications to sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 496–509, Feb. 2016.
- [32] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuska, "Natural language processing (Almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Aug. 2011.
- [33] A. Go, L. Huang, and R. Bhayani, "Twitter sentiment analysis," in *Proc. Int. AAAI Conf. weblogs Social Media* vol. 17, 2009, p. 252.
- [34] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. 7th Int. Conf. Lang. Resour. Eval.*, vol. 75013, N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, Eds. Paris, France: Rue Brillat-Savarin, 2010.
- [35] A. Go, L. Huang, and R. Bhayani, "Twitter sentiment analysis," The Stanford Natural Language Processing Group, Final Projects CS224N 2008/2009, 2009.
- [36] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.
- [37] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.
- [38] S. R. Laxpati and V. Goncharoff, *Practical Signal Processing and Its Applications: With Solved Homework Problems*. Singapore: World Scientific, 2017.
- [39] M. Rodríguez-Ibanez, S. Munoz-Romero, C. Soguero-Ruiz, F.-J. Gimeno-Blanes, and J. L. Rojo-Alvarez, "Towards organization management using exploratory screening and big data tests: A case study of the spanish red cross," *IEEE Access*, vol. 7, pp. 80661–80674, 2019.
- [40] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Cham, Switzerland: Springer, 2016.
- [41] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 2297–2301, Mar. 1991.
- [42] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
- [43] L. Zheng, W. Pan, Y. Li, D. Luo, Q. Wang, and G. Liu, "Use of mutual information and transfer entropy to assess interaction between parasymphathetic and sympathetic activities of nervous system from HRV," *Entropy*, vol. 19, no. 9, p. 489, 2017.
- [44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
- [45] F. Koto and G. Y. Rahmangtyas, "Inset lexicon: Evaluation of a word list for indonesian sentiment analysis in microblogs," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Dec. 2017, pp. 93–98.
- [46] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013, doi: 10.1111/j.1467-8640.2012.00460.x.
- [47] R. Plutchik, "The nature of emotions," *Amer. Sci.*, vol. 89, no. 4, pp. 344–350, 2001.
- [48] S. Kiritchenko and S. Mohammad, "The effect of negators, modals, and degree adverbs on sentiment composition," in *Proc. 7th Workshop Comput. Approaches to Subjectivity, Sentiment Social Media Anal.*, 2016, pp. 43–52.
- [49] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal text," *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, May 2014.
- [50] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 12, pp. 2544–2558, Dec. 2010.
- [51] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social Web," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 1, pp. 163–173, Jan. 2012.
- [52] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 english lemmas," *Behav. Res. Methods*, vol. 45, no. 4, pp. 1191–1207, Dec. 2013.



**MARGARITA RODRÍGUEZ-IBÁÑEZ** received the Ph.D. degree in communication and marketing from the Universidad Rey Juan Carlos, Madrid, Spain, in 2008. Since 2008, she has been a Full Professor of the Department of Business with different Spanish universities in the main teaching areas of marketing and communication. She is currently a Postdoctoral Researcher with the Biomedical Engineering and Data Science Group, Universidad Rey Juan Carlos. Her main research interests

include new technologies and their impact on organizations and society, data science, marketing, and machine learning.



**SERGIO MUÑOZ-ROMERO** received the B.Sc. degree in telecommunication engineering and the Ph.D. degree in machine learning from the Universidad Carlos III de Madrid. He has led pioneering projects, where the machine learning knowledge was successfully used to solve real Big Data problems. Since 2015, he has been the Head of Data Science and Big Data with Persei vivarium. He is currently a Researcher with the Universidad Rey Juan Carlos. His current research interests are centered in explainable machine learning algorithms, and statistical learning theory and their applications to big data.

tered in explainable machine learning algorithms, and statistical learning theory and their applications to big data.



**FRANCISCO-JAVIER GIMENO-BLANES** received the Diploma degree in advanced studies in taxes and business administration and the Ph.D. degree in communications technologies. He is a Telecommunications Engineer. With 25 years of professional experience, he has devoted almost half in the industry, both in Spain and internationally, holding various management positions at the industry and academic side. He has devoted over 15 years in the academy as a Teacher and

Researcher. He is currently the National Vice President of the Spanish Red Cross and a University Professor with Universidad Miguel Hernández. He has published 29 conference papers and 17 articles. He was supported by 14 research projects. He also founded a technology-based spin-off company and issued two intellectual property registers.



**CRISTINA SOGUERO-RUIZ** received the Ph.D. degree in machine learning with applications in healthcare through the Joint Doctoral Program in Multimedia and Communications with Universidad Rey Juan Carlos and University Carlos III, in 2015. She has published several articles in JCR journals and international conference communications. She was supported by the FPU Spanish Research and Teaching Fellowship (granted in 2012). She has participated in several research

projects (with public and private funding) related to healthcare data-driven machine learning systems. Her main research interests include machine learning, data science, and statistical learning theory. She won the Orange Foundation Best Ph.D. Thesis Award by the Spanish Official College of Telecommunication Engineering.



**PEDRO MANUEL CUENCA-JIMÉNEZ** received the B.Sc. degree in telecommunications engineering from the Universidad Politécnica de Madrid. He is an experienced professional who has devoted his career to the development of software-based projects and services enjoyed by millions. He was a member of the teams that created the first massive online education portal in Spain, the first online bookstore and the first free Internet service provider in the same country. After spending several

years at Spanish branches of the France Télécom Group, he co-founded LateNiteSoft, in 2008, to create mobile applications for iOS. Their Camera+ app brings advances in digital image processing and computational photography to the public and has been downloaded more than 12 million times. He has been an Assistant Professor with Universidad Rey Juan Carlos, since 2017.



**JOSÉ LUIS ROJO-ÁLVAREZ** (Senior Member, IEEE) received the B.Sc. degree in telecommunication engineering from the University of Vigo, in 1996, and the Ph.D. degree in telecommunication engineering from the Universidad Politécnica de Madrid, in 2000. He is a Professor with the Department of Signal Theory and Communications, Universidad Rey Juan Carlos, Spain. He has coauthored more than 130 international articles and has contributed to more than 180 conference

proceedings. His research interests focus on statistical learning methods for signal and image processing, arrhythmia mechanisms, robust signal processing methods for cardiac repolarization, and Doppler image post-processing.

...