Sentiment Analysis in Social Networks through Topic Modeling

Debashis Naskar [1], Sidahmed Mokaddem [1,2], Miguel Rebollo [1], Eva Onaindia [1]

[1] Universitat Politecnica de Valencia, Camino de Vera s/n Valencia Spain

[2] University of Oran, BP 1524 El Menouar street 31000 Oran Algeria (DZ)

debashis@drtc.isibang.ac.in, sid_wise@hotmail.com, mrebollo@dsic.upv.es, onaindia@dsic.upv.es

Abstract

In this paper, we analyze the sentiments derived from the conversations that occur in social networks. Our goal is to identify the sentiments of the users in the social network through their conversations. We conduct a study to determine whether users of social networks (twitter in particular) tend to gather together according to the likeness of their sentiments. In our proposed framework, (1) we use ANEW, a lexical dictionary to identify affective emotional feelings associated to a message according to the Russell's model of affection; (2) we design a topic modeling mechanism called Sent_LDA, based on the Latent Dirichlet Allocation (LDA) generative model, which allows us to find the topic distribution in a general conversation and we associate topics with emotions; (3) we detect communities in the network according to the density and frequency of the messages among the users; and (4) we compare the sentiments of the communities by using the Russell's model of affect versus polarity and we measure the extent to which topic distribution strengthen likeness in the sentiments of the users of a community. This works contributes with a topic modeling methodology to analyze the sentiments in conversations that take place in social networks.

Keywords: sentiment analysis, topic model, emotion identification, multilayer network

1. Introduction

Despite the amount of research done in sentiment analysis in social networks, the study of dissemination patterns of the emotions is limited. It is well known that social networks exhibit some kind of positive correlation in the polarity of the sentiments associated to sequential pairs of messages (Hillmann and Trier, 2012), and that the analysis of diffusion mechanisms is clue to determine the dynamic of the evolution of a particular emotion.

Ziegler and Lausen analyzed propagation of trust and distrust on social networks, what can be considered the first paper in which sentiment propagation was studied (Ziegler and Lausen, 2005). Interesting conclusions, like that positive and negative sentiments follow a different propagation pattern (Hillmann and Trier, 2012), have been drawn from the various investigations on sentiments in social networks. Other works studied the correlation between emotions and information diffusion, finding that those messages emotionally charged were re-tweeted more often (Stieglitz and Dang-Xuan, 2013), or investigated if the topic and the opinion of the user's contacts affect the own user's opinion (Tang and Fong, 2013).

The ultimate objective of analyzing sentiments in social networks is to be able to predict the attitude of people and infer behaviour patterns like, for example, reactions against negative opinions. In this line, Nguyen et al. studied changes in collective sentiment and predicted the dynamics with statistical models that contemplate the complete network (Nguyen et al., 2012), achieving a 85% of accuracy in the direction (polarity) of the sentiment. Sentiments can also be used to predict future connections in social networks by finding similar sentiments (Leskovec et al., 2010; Yuan et al., 2014).

The focus of this work is on sentiment analysis in tweets and, particularly, on the identification of the sentiments that users show when they talk about different issues within a same conversation. We apply the Latent-Dirichlet Alloca-tion (LDA) algorithm (Blei et al., 2003) to find the most similar words that uncover the hidden thematic structure (topics) in a *single-hashtag* tweet collection and then we extract meaningful emotions from each different topic based on the ANEW dictionary (Bradley and Lang, 1999). Unlike the majority of works that only study polarity of sentiments (coarse-grained classification in positive, negative and neutral sentiments), we propose a more refined classification of sentiments based on the Russell's model (Russell, 2003). The topic sentiment analysis provides a more precise snapshot of the sentiment distribution in a social network, thus allowing the identification of communities or sub-units of users within the network. Moreover, by combining this analysis with communities detection methods, we can determine if belonging to a determined group affects the user's sentiments.

This paper is organized as follows. Section 2. presents Sent_LDA, our approach for extracting and analyzing sentiments as well as a brief note on the Russell's model. Next section presents our implementation of the Latent-Dirichlet Allocation (LDA) algorithm for topic modeling in tweet messages. Section 4. explains the process for assigning sentiment scores to tweets, topics and users. In the last section, we present the experimental evaluation, analyzing the representative social graph for the network, the communities formation and the spread of sentiments across the network. The last section concludes and presents our future work.

2. Sent_LDA: emotion identification in tweets and topics

This section presents the overview of our approach Sent_LDA but before we introduce the Russell's model of affect. Using this model enables us to define a more accurate sentiment of the messages and discriminate between cases in which the general sentiment is just positive or negative (for example, when a catastrophe occurs).

In the Russell's circumflex model of affect, emotions are understood as a combination of varying degrees of two main dimensions, *valence* (pleasure dimension) and *arousal* (activation dimension), which are distributed in a 2D circular space (Russell, 1980). According to the Russell's model, every affective experience is the consequence of a linear combination of valence and arousal dimensions (the so-called *core affect space*), which is then interpreted as representing a particular emotion. A numerical value for valence ranges from 1 (*unpleasant*) to 9 (*pleasant*) and arousal values range from 1 (*sleepy*) to 9 (*awake*) (see Figure 1). In the core affect map, we identify four regions (R1, R2, R3 and R4) along with 16 sentiment words that lie on the perimeter of a circle. Words are labeled with a particular name: *excited*, *sad*, *unhappy*, *bored*, etc., each having their polar coordinates on the circle (Russell, 1980).

Assuming that affect can be modified by degree of valence and arousal, it seems reasonable to assume that emotions have the potential to lie across all positions in the two-dimensions rather than just on a perimeter (Russell, 2003). The core affect map in Figure 1 will be used to identify the emotional state or sentiment label of a given entity (message, topic or user) according to its valence (X-axis) and arousal (Y-axis) values. Particularly, given an entity e with valence and arousal values (v_e, a_e), we used the Euclidean distance for identifying the closest distance sentiment to e. For instance, an entity e with valence and arousal values $(v_e, a_e) = (7.26, 3.56)$ falls within region R2 and its associated sentiment label would be $S_e = \{serene\}$.
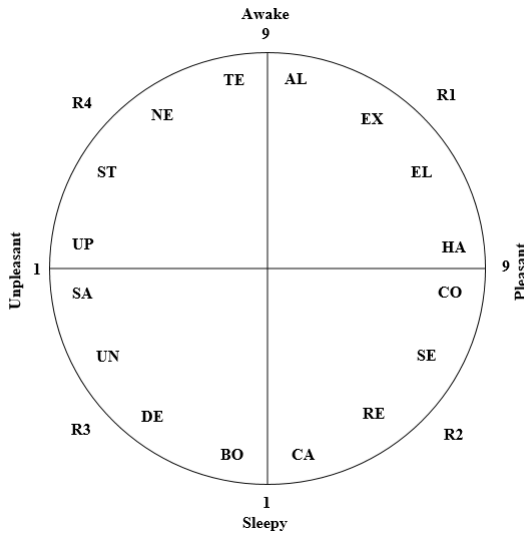

Figure 2: Global Overview of Sent_LDA

Figure 1: Representation of Russell's model (R_1= AL: alert, EX: excited, EL: elated, HA: happy; R_2= CO: contented, SE: serene, RE: relaxed, CA: calm; R_3= BO: bored, DE: depressed, UN: unhappy, SA: sad; R_4= UP: upset, ST: stressed, NE: nervous, TE: tense)

2.1. Overview of Sent_LDA

Several approaches for extracting emotion from tweet messages exist in the literature. In (Mohammad and Kiritchenko, 201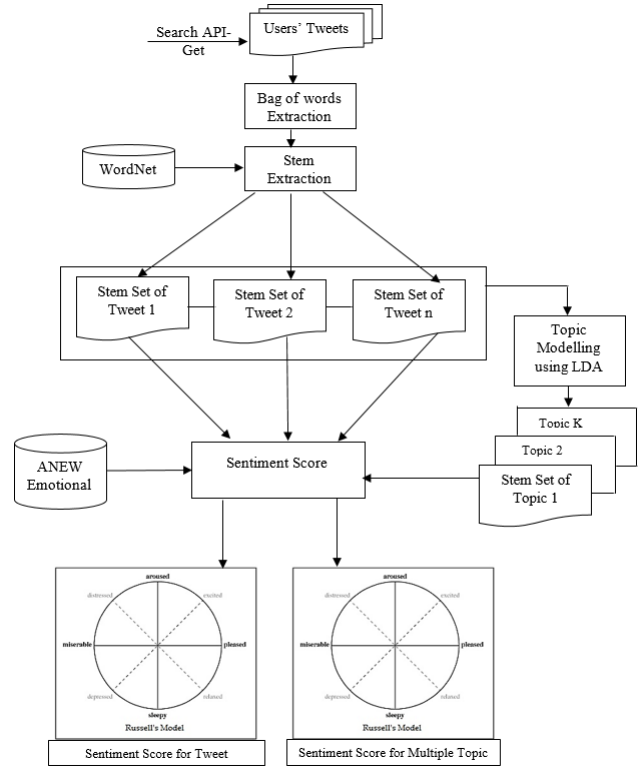5), authors create a large lexicon from tweets with sentiment-word hashtags added by tweeters corresponding to the six basic Ekman emotions (Ekman, 1992). Using emotion-related hashtags to identify the topic of the message has been the predominant choice to create emotion-labeled datasets from tweet messages (Choudhury et al., 2012), (Purver and Battersby, 2012), (Qadir and Riloff, 2014). Unlike these works, our aim is to find the most similar words that uncover the hidden thematic structure (topics) in a *single-hashtag* tweet collection and extract meaningful emotions from each different topic.

The overview of Sent_LDA, our sentiment analysis model, is shown in Figure 2. It includes tweet processing mechanisms, topic modeling method through LDA and sentiment extraction tool for classification of sentiments. We consider three types of entities: tweets or messages, topics and users, all associated to an opinion orientation expressed either with an individual sentiment or a set of sentiments represented through valence and arousal values.

For our purpose of analyzing real-time events, we chronologically retrieved as many tweets through the Twitter *Search API*. We used *hashtags* for collecting tweets which are posted by several users on the basis of a particular domain (conversation) and we built a corpus of D tweets. The first step is to have a word representation for each message in order to facilitate machine manipulation as well as elimination of noisy words. Processing each individual tweet message involves extracting bags of words, filtering all stop words and extracting stems by using the stemmer tool of the WordNet dictionary (Figure 2, top). As a result, we obtain the vocabulary for our particular dataset composed of V stem words. Afterwards, the topic model LDA algo-

rithm is applied over the tweets' stem sets (vocabulary) and it returns a stem set for each found topic (Figure 2, middle; details are presented in Section 3.). Finally, the emotional role of the stem sets of tweets and topics is obtained through the emotional base ANEW dictionary (Figure 2, bottom) and words are annotated by their valence and arousal values (see section 4.). Words are then classified according to the Russell's emotion model.

Formally, let's consider a given tweet $d \in D$, an entity which is tokenized and filtered by eliminating stop words to a bag of words w_d. We transform w_d to a stem set, s_d, through the WordNet stemmer tool, which allows us to get the appropriate stem for each token. The sentiment words associated to s_d are spotted based on the ANEW dictionary, where emotional words are annotated by their valence and arousal values (examples of such emotional words are agreement, love, sad, quite, disagree, etc.). Finally, the overall opinion orientation or emotional tweet status conveyed by the user is determined by combining the content of each emotional word identified in the tweets sent by the user and classified according to the Russell's model. Additionally, each topic identified by the LDA algorithm constitutes a new entity, t, and the emotional value of each t is similarly obtained. The final outcome of the sentiment analysis process is a pair of lists: $\langle oo_{d_1}, \ldots, oo_{d_D} \rangle$ for the D tweet messages and $\langle oo_{t_1}, \ldots, oo_{t_K} \rangle$ for the K topics elicited by the LDA algorithm .

3. Topic modeling using LDA

Our assumption is that multiple scattered sentiments are typically found in a *single-hashtag* general conversation and that other more useful and suitable sentiment groupings can be extracted by identifying the topics involved in the conversation. The aim of this section is to elicit a topic distribution over the corpus of D tweets and a sentiment distribution of topics.

The Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003) is one of the most successful topic models to infer the topics discussed in a collection of documents. The probabilistic generative model LDA models the D documents in the corpus as mixtures of K latent topics where each topic is a discrete distribution over the V words of the collection's vocabulary. The LDA generative process results in the joint distribution:

$$p(\mathbf{w}, \mathbf{z}, \theta, \phi | \alpha, \beta) = p(\phi|\beta)p(\theta|\alpha)p(\mathbf{z}|\theta)p(\mathbf{w}|\phi_z) \quad (1)$$

where $\mathbf{w}$ is the observed words, the unobserved latent variables are ϕ (the $K \times V$ per-topic word distribution matrix), θ (the $D \times K$ per-document topic distribution matrix) and $\mathbf{z}$ (the $D \times V$ matrix that represents the topic index assignment for each word w_i in document $j \in D$). Given θ and ϕ, drawn from the hyperparameters α and β for the symmetric Dirichlet distribution, the aim is to learn the latent variables; that is, the words associated with each topic (expressed through the topic distributions ϕ). LDA therefore sees each document as a set of topic occurrences that appear in an arbitrary order, which is similar to the 'bag-of-words' model. Choosing a topic is performed independently for each word of the document under the constraint of overall compliance of the fixed distribution over topics θ_d.

Rather than learning θ and ϕ directly, we applied the reverse generative process and learnt the posterior distributions of the latent variables by means of collapsed Gibbs sampling, a Markov Chain Monte Carlo algorithm that samples one random variable at a time (Porteous et al., 2008; Griffiths and Steyvers, 2004). This process learns the probability of a topic z_{ij} in document j being assigned a word w_i, given all other topic assignments to all other words. That is, learning the assignment of words to topics $\mathbf{z}$ given the observed words $\mathbf{w}$ by repeatedly sampling the latent variables z_{ij} regarding the other values of $\mathbf{z}$. As a result of using the collapsed Gibbs sampler for LDA, we can draw the values of θ and ϕ.

Two key aspects affect the performance of this process: the *number of iterations* for Gibbs sampling to converge and the selection of the *optimal number of topics*. Regarding the first issue, we know that convergence is theoretically guarantee with LDA Gibbs sampling but there is no way of knowing how many iterations are required to reach the stationary distribution. In practice, a visual inspection of the log-likelihood can give us an acceptable estimation of convergence.

As for the number of topics, some approaches like Hierarchical Dirichlet Process can be used to estimate the best topic number and this is line we intend to explore in the near future. Nevertheless, a relatively simple way to find the optimal number of topics is by iterating through models with different numbers of topics and select the model with the maximum log-likelihood, given the data. In general, since we aim for discovering topics in tweet messages which are all labeled under the same hashtag, the number of topics in the general conversation will be relatively low. Thus, the parameters of the Dirichlet distribution used in LDA can be assigned small values in order to render only a few topics allocated to each document.

Additionally, LDA presents another weakness related to the overlapping topics composition. It is desirable to have independent topics and thus avoiding the appearance of the same word in multiple topics. Consequently, a better fit will be a model with fewer overlapping words. In practice, the best fitting model can also be discovered through analysis of the log-likelihood.

Topic Modeling in Twitter. Based on other works on topic modeling in twitter, we justify here the use of LDA for topic modeling and sentiment identification. First we must note that, unlike supervised statistical approaches that require manual annotation of tweets with emotions (Mohammad and Kiritchenko, 2015) or constrain the topic model to use only those topics that correspond to a document's (observed) label set (Ramage et al., 2009), we focus exclusively on unsupervised learning algorithms.

The short and sparse texts of tweets messages pose a serious challenge to the efficacy of topic modeling. Common techniques to overcome this limitation rely upon aggregation strategies as a data preprocessing step for LDA. In (Hong and Davidson, 2010) authors train two models, LDA and an author-topic model (Rosen-Zvi et al., 2010), in short text environments (Twitter). They introduce several aggregation techniques to obtain topics associated with messages and their authors, and they conclude that a standard LDA

model on user aggregated profiles yields better results than the author-topic model. In (Mehrotra et al., 2013), authors apply various tweet pooling schemes in a standard LDA to finally conclude that *hashtag-based pooling*, creation of pooled documents for each hashtag, outperforms all other pooling strategies and the unpooled scheme.

In general, LDA has proven to work well on tweets (Weng et al., 2010), particularly when messages are very focused and very few topics are discussed in the composition of the entire tweet (Naveed et al., 2011). This is precisely the case of our corpus where tweets are all labeled under the same single hashtag.

4. Sentiment extraction with ANEW dictionary

In this section, we explain the extraction of sentiments from the stem sets of tweets and topics and how to associate sentiment scores to these entities based on the ANEW dictionary. First, we justify the use of Affective Norms for English Words (ANEW) dictionary versus other dictionaries like SentiWordnet and then we explain the sentiment extraction and score assignation.

4.1. ANEW dictionary

The aim of sentiment extraction is to compile sentiment words. One of the most efficient approaches for this purpose is the dictionary-based approach. Dictionary-based approaches use dictionaries of emotional words which are associated to a sentiment score. There exist several affect dictionaries in the literature like ANEW dictionary of affect (Bradley and Lang, 1999; Nielsen, 2011) or SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010).

The performance of dictionary-based approach can be evaluated according to two aspects: 1) the number of emotional words covered by the dictionary and 2) the nature of sentiment score provided by the dictionary. ANEW, for instance, computes this score with the valence and arousal values of the word, which range from 1 to 9; SentiWordNet, instead, uses polarity. Hence, ANEW allows us to calculate a more accurate sentiment value which fits better our aim of having a bi-dimensional representation of sentiments as well as to measure the intensity of expressed sentiments. Additionally, the new version of the ANEW dictionary (Nielsen, 2011) provides the mean and standard deviation of normative emotional ratings (valence v and arousal a) for 2477 unique words in English (see next section).

Corpus	ANEW	SentiWordnet	Vocabulary
Tweet 1	377	471	1138
Tweet 2	1117	1262	2419
Review	1097	1012	7202

Table 1: #words covered by different dictionaries of affect

We compared the number of emotional words covered by ANEW and SentiWordNet in three of our experiments (results are shown in Table 1). The first two corpus shown in Table 1 correspond to microblog tweets (tweets do not contain more than ~ 12 words) and the third corpus corresponds to a set of beer reviews [1], where reviews contain more than 30 words in average. As we can observe in Table 1, the much larger size of SentiWordNet (155287 words) versus ANEW (2477 words) only implies a very insignificant higher coverage or even slightly lower as in the case of the beer reviews.

4.2. Calculating Sentiment scores

Following, we show an example of a tweet message that comprises three emotional words that exist in the ANEW dictionary along with their valences and arousal values.

Feb 24, 8:02pm: @Suvi_90 @Bloody_Mary0812 yeah who **hate** him? It's so strange and **sad** but we **love** him so much :) #SPNFamily #GetJensenToOneMillion

- **hate**, $v = [\mu : 1.79, \sigma : 1.32], a = [\mu : 6.79, \sigma : 2.66]$
- **sad**, $v = [\mu : 1.61, \sigma : 0.95], a = [\mu : 4.13, \sigma : 2.38]$
- **love**, $v = [\mu : 8.72, \sigma : 0.7], a = [\mu : 6.44, \sigma : 3.35]$

The aim of this phase is to associate each entity e with a tuple (v_e, a_e). The average sentiment score of a tweet message d is calculated with the valence and arousal of the stem words of d that appear in the ANEW dictionary (emotional words of d). Then, the sentiment score of a user u is calculated with the score of his/her tweet messages and we associate the corresponding sentiment label S_u as explained in Section 2.

In order to combine the mean values of the valence and arousal of the emotional words, we have to assume that individual mean values reported for each stem form a normal distribution. Supposedly, if a stem has a high σ of valence (equivalently for arousal) then the valence ratings of the word are distributed over a wider range of values; and lower values of σ imply that ratings are closer to μ. Thus, we used a probability weight based on the probability density function of each word in ANEW to estimate that the stem's valence (arousal) falls exactly at the mean.

$$X = \frac{\sum_{i=1}^{N} \frac{\phi_{i,t}\mu_i}{\sigma_i}}{\sum_{i=1}^{N} \frac{\phi_{i,t}}{\sigma_i}} \qquad (X, \phi, \mu, \sigma) \qquad (2)$$

X	mean value of valence (Y, mean value of arousal)
N	total number of emotional words within the message
$\phi_{i,t}$	topic distribution for word i in the tth topic
μ	word's mean value of valence (equivalently for arousal)
σ	word's standard deviation of valence (equivalently for arousal)

Table 2: Notation of different attributes

Formula (2) calculates the sentiment score of a message by estimating the overall mean value of all emotional words within the message (see Table 2). Particularly, X (respectively, Y) is the overall mean value of the valence (respectively, arousal) considering the N emotional words within the message. When calculating the sentiment score of a

message without considering the topic distribution ϕ returned by LDA, we set $\phi_{i,t}$ to 1. In this case, X and Y (valence and arousal) denote the *primary sentiment* of the conversation; that is, the sentiment score of the general conversation without considering the word distribution per topic. Then, the sentiment score of a user is calculated as the average emotional value of all the tweets sent by the user. For example, if we amalgamate the three words **hate**, **sad** and **love** of the above message d, the result of the weighted average formula (2) for the valence and arousal is $X_d = 4.78$ and $Y_d = 5.69$, respectively.

On the other hand, if topics are taken into account, the sentiment scores of the messages and users will be subject to such particular topic distribution ϕ. The same formula (2) shows how to calculate the mean value of valence and arousal of a message when the word w_i of the message appears in topic t with a probability $\phi_{i,t}$.
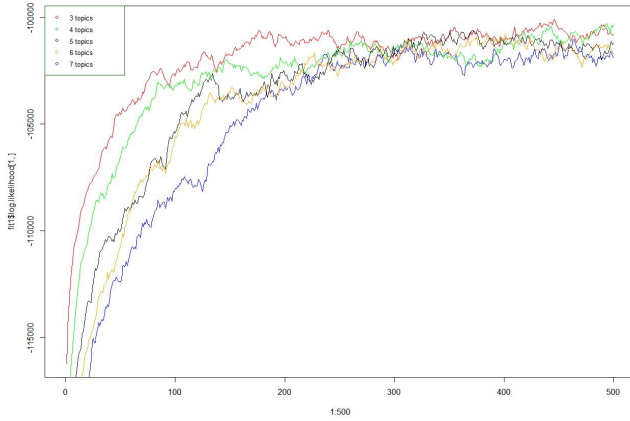

Figure 3: Log-likelihood representation for all Topics

5. Experimental Evaluation

The purpose of the experimental evaluation is to identify the behaviour of users' sentiments over a conversation. Particularly, we are interested in analyzing the impact of topic distribution of a general conversation in the users' sentiments. Our hypothesis is that the sentiments of users talking about a particular topic are more alike than the sentiments of the same users in the general conversation.

To lead our study over the proposed framework, we collected over 10,600 tweets with the *Search-API* related to one of the most important and recent sport events, based on the hashtag (#elclasico). Once the tweets were collected, we kept those involved in a direct conversation (replies, mentions and re-tweets). We will refer to this tweet collection as the *primary conversation*, where a total of 3,600 users were identified.

Network representation. A network representation is required to study the emotional behaviour of users through a primary conversation and to evaluate the users sentiment tendencies on specific topic discussion. A multilayer network is formally defined by (Boccaletti et al., 2014) as a pair $M = (G, C)$ where $G = \{G^1, \dots, G^K\}$ is a family of graphs, $G^\alpha = (X^\alpha, E^\alpha)$ is a layer, and $C = \{E^{\alpha\beta} \subseteq X^\alpha \times X^\beta\}$ is the set of connections between two different layers G^α and G^β. Elements of E^α are the *intralayer* connections and the elements of C are the *interlayer* or *crossed layers*. The characteristic of the multiplex network is that all the layers have the same set of nodes $X^1 = \dots = X^K = X$ and the cross layers are defined between equivalent nodes $E^{\alpha\beta} = \{(x,x), x \in X\}$. In our case, a topic corresponds to a layer of the multiplex network and the nodes represent the users (Boccaletti et al., 2014).

Topic distribution with LDA. To evaluate the optimal number of topics, LDA was run with different values of K, from $K = 3$ to $K = 7$ and we selected the model with the maximum log-likelihood. Figure 3 shows the log-likelihood for each model with respect to the number of iterations represented in the X-axis. As we can observe, the five models tend to convergence at about 500 iterations and $K = 3$ shows the best log-likelihood. This was also confirmed by checking that $K = 3$ was the model with fewer overlapping words across topics. Therefore, we selected the 3-topic model identified by LDA, so the multiplex graph is composed of $K = 3$ layers, one per topic. Table3 shows the top-10 words for the resulting model $K = 3$. For instance, Topic 3 contains different words with high probability (Bale (0.10), Car (0.06) and Attack (0.04)) which means that the subject of this topic is about Bale's car attack after the *'el clasico'* game. Interestingly, Topic 1 reveals a conversation about the Most Valuable Player (MVP (0.04), Vote (0.04) and Player (0.03)).

Community detection. For evaluating the effectiveness of our topic-sentiment approach, we propose a community detection method to extract the community structure in large networks[2]. A community is defined as a group or cluster of users who are densely connected with intra-community edges and scattered by inter-community connection (Kauffman et al., 2014). The modularity measure of a partition is a scalar value $[-1, 1]$ that measures the density of links inside communities as compared to links between communities (Newman, 2006). We used the detection algorithm based on modularity maximization proposed in (Blondel et al., 2008) to identify the existing communities in the conversation through the density and the frequency of the messages among the members of each group. A total of 19 communities were identified in the primary conversation (Table 4). Our final objective is to analyze the sentiments of the communities identified in the primary conversation and then check if the sentiments of users who talk about the same topic within each community have a higher degree of similarity.

5.1. User's Sentiment Identification and Analysis

Users are depicted in the extracted graph M, which models the conversation about the analyzed hashtag. We calculate the sentiment score of every user u in the primary conversation; that is, we calculate the mean value of valence (X_u) and arousal (Y_u) on the basis of the tweets sent by u as explained in section 4.2., and assess the sentiment of users according to two different measures:

---

[2]Large networks : https://sites.google.com/site/findcommunities/

Topic	Word (Frequency)									
Topic 1	Win (0.05)	**MVP (0.04)**	**Vote (0.04)**	**Player (0.03)**	Complete (0.03)	Score (0.02)	Fight (0.02)	Pass (0.02)	Chance (0.02)	Second (0.02)
Topic 2	CL (0.10)	Headline (0.07)	Make (0.07)	World (0.07)	Victory (0.06)	Goal (0.06)	Camp (0.03)	Barcelona (0.02)	Look (0.02)	like (0.01)
Topic 3	Madrid (0.12)	Real (0.12)	**Bale (0.10)**	Fan (0.07)	**Car (0.06)**	Barcelona (0.06)	**Attack (0.04)**	Play (0.04)	Number (0.03)	Defence (0.03)

Table 3: Different word frequency of each topic

Comm.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Polarity	0.28 (83)	0.00 (519)	0.00 (29)	0.37 (57)	0.00 (74)	0.00 (159)	0.00 (69)	0.00 (5)	x (1)	0.92 (3)	0.81 (4)	0.00 (11)	0.77 (40)	0.22 (83)	0.00 (49)	0.09 (92)	0.00 (132)	0.00 (4)	0.00 (27)
Primary Sentiment	**0.50** (83)	1.07 (519)	**0.36** (29)	1.18 (57)	1.43 (74)	**1.42** (159)	**0.19** (69)	0.72 (5)	x (1)	**0.92** (3)	1.50 (4)	**0.00** (11)	1.30 (40)	1.85 (83)	0.25 (49)	0.60 (92)	0.80 (132)	0.81 (4)	0.82 (27)
Topic 1	1.78 (130)	0.99 (147)	**0.36** (29)	1.06 (24)	1.45 (61)	1.80 (140)	0.30 (69)	**0.00** (5)	x (1)	-	x (1)	**0.00** (11)	0.95 (14)	**1.69** (58)	0.57 (20)	**1.00** (2)	**0.00** (18)	**0.00** (4)	**0.00** (4)
Topic 2	1.70 (108)	0.71 (272)	x (1)	**1.04** (29)	**0.00** (5)	2.07 (14)	x (1)	-	-	-	0.92 (3)	-	1.50 (4)	1.84 (18)	**0.23** (27)	1.58 (3)	**0.00** (5)	x (1)	-
Topic 3	1.55 (96)	**0.63** (145)	-	1.15 (7)	1.21 (18)	2.20 (9)	-	x (1)	-	0.92 (3)	1.00 (2)	-	**0.00** (25)	1.92 (5)	0.81 (4)	**0.09** (89)	0.07 (112)	-	0.26 (23)

Table 4: Shannon Entropy associated with the different communities identified and its size

- **Polarity**. A sentiment is evaluated as positive if $X_u > 5$ and negative otherwise. In general, polarity gives more general emotional overview which is explained by its 2-dimensional discriminants.

- **Primary Sentiment**. We associate one of the 16 primary sentiments of the Russell's model to each user. We identify the appropriate region of the model through X_u and Y_u and then we use Euclidean distance to identify the closest sentiment(s).

In order to evaluate the similarity of users' sentiments within each community according to both measures, we calculated the Shannon entropy, a measure of the uncertainty in the probability distribution of sentiments (mixture ratio of sentiments present in each community). The first two rows of Table 4 show the entropy results for Polarity and Primary Sentiments, respectively. Below the entropy value in each community, the number of users found in such community is shown between parenthesis. When the community has a single user, it makes no sense to calculate the entropy value (this is indicated with a (x) sign).

An entropy value close to zero means the community has a unique predominant sentiment. Higher entropy values are associated to communities in which more than one sentiment appears. Unsurprisingly, the best values were obtained from the polarity level, where only two sentiments - positive and negative - are handled, and a predominant positive sentiment was detected in the network. The Primary Sentiment level shows less uniformity among the users since 16 different sentiments are considered at this level. Nevertheless, if we consider that the max entropy value is 4 ($log_2 16$) and that no value in Primary Sentiment exceeds 1.85, we can say there is a fairly low uncertainty in the predominant Russell's sentiment in each community.

	#tweets	#users	Sentiment assortativity
Primary conversation	10,600	3,600	0.2047
Topic 1	4,152	1,482	**0.446**
Topic 2	2,500	1,008	**0.341**
Topic 3	3,979	1,235	**0.391**

Table 5: Structural properties of the complete network and each one of the layers

5.2. Topics impact on communities

We now study the impact of topic distribution in the sentiments of the communities. The three topic conversations are modeled in three different layers G^1, G^2, G^3 of the multiplex network, each associated to the corresponding topic. A layer G^i was obtained by retrieving the tweet messages sent by the users about topic i. Thus, each layer contains the set of users participating in the topic conversation, where the same user can participate in more than one topic distribution; that is, the same user can be found in different layers of the network (inter-layer connections). Table 5 shows the structural properties of the whole network and each layer, where the first two columns show the number of tweets and users, respectively. The last column is the value of the sentiment assortativity, that is, the tend of users to be connected with users of similar sentiments. Assortativity is a correlation coefficient between +1 and -1; negative values indicate a negative relationship that connects people with different sentiments; a positive value, however, shows a positive correlation connecting users with similar sentiments. The interesting thing here is that the assortativity in the primary conversation has a lower positive value than the assortativity in the topic conversations, thus indicating that users in the layers have more connections with sentimentally alike users than they have in the primary conversa-

tion. This can be interpreted as a first indication that topics strengthen likeness in sentiments.

Subsequently, we analyzed the entropy values in the topic conversations per community according to the primary sentiments of the Russell's model. We calculated X_u^i and Y_u^i for every user u and topic i in each community and we obtained the entropy values shown in the last three rows of table 4 along with the number of users. The (-) sign means there are no users associated to that community. In our analysis, we will compare the entropy values of the topic conversations with the results of the Primary Sentiment level (the best value is shown in bold).

The first observation is that 11 communities show an entropy reduction when the topic is considered, meaning that topic distribution strengthens the likeness of sentiments in these communities. For instance, in community 2, the biggest community among all (519 users), topic distribution gathers users together in three topic sub-communities that show a much higher sentiment similarity, notably in topic 2 and topic 3. Another interesting observation is community 8, where the same five users happen to be have the same sentiment, what was discovered when we detected the five users talking about the same topic (topic 1). Notice that the entropy value of these five users in the Primary Sentiment is 0.72, an indication that the users show different sentiments when the primary conversation is considered. In community 13, for instance, we observe the existence of 25 users perfectly aligned around the same sentiment in topic 3; and, almost all of the users in community 16 show the same sentiment when conversation about topic 3 is analyzed (note the drop of the entropy from 0.60 in the Primary Sentiment level to 0.09 in topic 3). A similar happening takes place in community 19 between the Primary Sentiment level and topic 3 level although in this case it was found 4 users with exactly the same sentiment in topic 1. On the other hand, there are communities that are only present in some topics, such as communities 3, 7, 8, 11, 12 or 19. This also shows that the topic layer tends to give more optimized communities based on sentiments.

6. Conclusion

In this paper, we have presented a sentiment detection approach to extract sentiments from single-hashtag twitter conversations to demonstrate that user-level sentiment analysis can be significantly improved when incorporating topic modeling. Following the topic distribution, our Sent_LDA approach generates several layers (topics) from the primary conversation, each representing a network of sentiments associated to different messages and users. Experimentation showed that topic modeling is very helpful for sentiment classification of twitter messages since different contextual views of sentiments are obtained. We used various levels of sentiments and we observed that the primary sentiment identification is an appropriate level to analyze users' emotional tendencies. Finally, the detection of communities through the topic distribution analysis highlights a more precise picture of users' sentiments.

8. Bibliographical References

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *International Conference on Language Resources and Evaluation,LREC,2010*, volume 10, pages 2200–2204.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C., Gómez-Gardeñes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.

Bradley, M. M. and Lang, P. J. (1999). Affective norms for english words (anew): Technical manual and affective ratings. *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*.

Choudhury, M. D., Gamon, M., and Counts, S. (2012). Happy, nervous or surprised? classification of human affective states in social media. In *AAAI International Conference on Weblogs and Social Media, ICWSM, 2012*.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *International Conference on Language Resources and Evaluation,LREC,2006*, volume 6, pages 417–422.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *National academy of Sciences of the United States of America*, 101(suppl 1):5228–5235.

Hillmann, R. and Trier, M. (2012). Dissemination patterns and associated network effects of sentiments in social networks. In *IEEE International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 2012*, pages 511–516.

Hong, L. and Davidson, B. D. (2010). Empirical study of topic modeling in twitter. In *ACM Workshop on Social Media Analytics,SOMA 2010*, pages 80–88.

Kauffman, J., Kittas, A., Bennett, L., and Tsoka, S. (2014). Dyconet: a gephi plugin for community detection in dynamic complex networks. *PloS one*, 9(7):0101357.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In *ACM International conference on World wide web, WWW, 2010*, pages 641–650.

Mehrotra, R., Sanner, S., Buntine, W. L., and Xie, L. (2013). Improving LDA topic models for microblogs via

tweet pooling and automatic labeling. In *ACM Conference on research and development in Information Retrieval, SIGIR 2013*, pages 889–892.

Mohammad, S. M. and Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, pages 301–326.

Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A. C. (2011). Searching microblogs: coping with sparsity and document quality. In *ACM Conference on Information and Knowledge Management, CIKM 2011*, pages 183–188.

Newman, M. E. (2006). Modularity and community structure in networks. *National Academy of Sciences*, 103(23):8577–8582.

Nguyen, L. T., Wu, P., Chan, W., Peng, W., and Zhang, Y. (2012). Predicting collective sentiment dynamics from time-series social media. In *ACM International workshop on issues of sentiment discovery and opinion mining, WISDOM, 2012*, page 6.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. In *ESWC Workshop on 'Making Sense of Microposts': Big things come in small packages, MSM, 2011*, pages 93–98.

Porteous, I., Asuncion, A., Newman, D., Smyth, P., Ihler, A., and Welling, M. (2008). Fast collapsed gibbs sampling for latent dirichlet allocation. In *ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD, 2008*, pages 569–577.

Purver, M. and Battersby, S. (2012). Experimenting with distant supervision for emotion classification. In *ACL Conference of the European Chapter of the Association for Computational Linguistics,EACL, 2012*, pages 482–491.

Qadir, A. and Riloff, E. (2014). Learning emotion indicators from tweets: Hashtags, hashtag patterns, and phrases. In *Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014*, pages 1203–1209.

Ramage, D., Hall, D. L. W., Nallapati, R., and Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Conference on Empirical Methods in Natural Language Processing, EMNLP 2009*, volume 1, pages 248–256.

Rosen-Zvi, M., Chemudugunta, C., Griffiths, T. L., Smyth, P., and Steyvers, M. (2010). Learning author-topic models from text corpora. *ACM Transactions on Information Systems, TOIS, 2010*, 28(1):4:1–4:38.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145–172.

Stieglitz, S. and Dang-Xuan, L. (2013). Emotions and information diffusion in social media - sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4):217–248.

Tang, J. and Fong, A. (2013). Sentiment diffusion in large scale social networks. In *IEEE International Conference on Consumer Electronics, ICCE, 2013*, pages 244–245.

Weng, J., Lim, E.-P., Jiang, J., and He, Q. (2010). Twitterrank: finding topic-sensitive influential twitterers. In *ACM International conference on Web search and data mining, WSDM 2010*, pages 261–270.

Yuan, G., Murukannaiah, P. K., Zhang, Z., and Singh, M. P. (2014). Exploiting sentiment homophily for link prediction. In *ACM Conference on Recommender systems, RecSys, 2014*, pages 17–24.

Ziegler, C.-N. and Lausen, G. (2005). Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, 7(4-5):337–358.