Computers and Electrical Engineering 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Emotional community detection in social networks☆

Andreas Kanavos, Isidoros Perikos, Ioannis Hatzilygeroudis*, Athanasios Tsakalidis

Computer Engineering and Informatics Department, University of Patras, Hellas, Greece

A R T I C L E I N F O

Article history:
Received 14 November 2016
Revised 7 September 2017
Accepted 7 September 2017
Available online xxx

Keywords:
Community analysis
Graph mining
Influential community detection
Sentiment analysis
Tweet emotion recognition
User influence

A B S T R A C T

It is vastly acknowledged that analyzing social networks is a very challenging research area. Take as a striking example the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. This comprises a fundamental aspect, which concerns the detection of user communities. In certain fields such as sociology and computer science where interactions and associations are often represented in the form of graphs, detecting communities is of vital importance. This paper addresses the need for an efficient and innovative methodology for community detection that will also leverage users' behavior on emotional level. Ekman emotional scale is the key point with which the methodology analyzes user's tweets in order to determine their emotional behavior. Consequently, the derived communities are estimated with the use of three different metrics, while the weighted version of a modularity community detection algorithm is utilized. There is substantial evidence indicating that our proposed methodology creates influential enough communities.

1. Introduction

Over the past few years, Twitter alongside other social media has had an increasing popularity. This in return, has generated an enormous research concernment as well as new opportunities for studying the reciprocal actions of different groups of people. Community detection and sentiment analysis are two such instances, comprising a popular topic in the investigation and better understanding of social networks. As for community detection on the one hand, it tries to analyze a social network with the capital objective of detecting clusters of associated and related users in it, while on the other hand sentiment analysis endeavors to settle upon the users' behavior on emotional level and consequently specify their attitude on a diverse number of topics, such as to recognize how individuals feel.

Being of the utmost fundamental aspects of the social network analysis, the determination of user behavior in each one of the arising communities as well as in the whole network is a vibrant concept of analyzing the exact way that users are associated for creating social communities. In order to explain social dynamics of interaction among groups of individuals, it is imperative that we study the community structure of a network. Based on this principal, there are research efforts in the literature that point to this direction [5]. A topic of extremely high research interest with wide range applications is the efficient analysis and the accurate specification of communities.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. S. Sioutas.
* Corresponding author.
E-mail address: ihatz@ceid.upatras.gr (I. Hatzilygeroudis).

http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

Community detection approaches could greatly assist in the economical and marketing domain. The accurate discovery of concrete communities and the analysis of each one could ameliorate the performance of advertising initiatives of the marketing industry by specifying and addressing the appropriate groups of users in a specific network. In this line, an effective way to offer to users recommendations adapted to their behavior and interests would be to utilize aspects of the structure of users' behavior and use the affiliations of the communities of the users [8].

People's decisions and attitudes as well as their social relationships can be affected and shaped by public stance in social networks, and since public opinion and emotions are essential to all aspects of human lives, their recognition is quite important [28]. Recognizing emotional status in user generated content in social networks is a very challenging and interesting topic in social networks and in microblogging area [3]. The analysis is necessary for understanding people status and for providing a number of indicative factors regarding public attitude towards different events. In this line, the accurate specification of the emotional content of users can potentially describe the emotional status of a community, a town or even a whole country [24].

Although users' emotional behavior as an important parameter for understanding public behavior most of the existing research efforts and methods for specifying structured communities in social networks do not utilize it. However, emotional aspects of people behavior and their opinions could provide indicative factors and assist in detecting more concrete and structured communities in terms of density.

In the present work, the main contributions concern the following aspects: Initially, a method for the analysis of user generated content and the specification of users' emotional behavior on Ekman's psychometric scale [6] is presented. Furthermore, a method for examining users' actions and their posts in social networks while calculating their influence based on their behavior and modeling a conversation as an emotional graph, is introduced. Finally, we present a method for specifying the most influential communities of users that are formulated based on each user's emotional behavior and their analytics profile in the network. Specifically, each user profile can be considered as the union of the two aforementioned characteristics, i.e. emotional profile and analytics profile.

The rest of this manuscript is structured as follows: Section 2 presents background topics in emotional detection and sentiment analysis as well as community detection. Section 3 presents our proposed methodology while in Section 4, details of the implementation are introduced. Furthermore, in Section 5, the evaluation study conducted and the results gathered regarding the community detection proposed schema are presented. Finally, Section 6 concludes the paper and provides directions for future research.

2. Related work

Community analysis in social networks has gained the interest of scientists in multiple areas such as graph theory, social network analysis, graph clustering algorithms, as well as web searching algorithms [9,15,16,20]. A community can be considered as a group of network nodes where links connecting these nodes are dense [30]. Moreover, another characteristic of a community is that it corresponds to a group of nodes on a graph or even a network that share common properties. Community detection is the problem of identifying structures of grouping of nodes which demonstrates high coupling and low cohesion. For a complete overview of wide used approaches and techniques, one can consider works presented in [21,23].

Concerning communities, the well known problem of graph partitioning draws a vast amount of attention. Concretely, in [10], an algorithm that is presented for identifying the edges lying between communities and their successive removal can be considered as a breakthrough in this area. In this algorithm, a procedure leads to the isolation of the communities after some iterations. One should also mention techniques that use the modularity metric, which designates the density of links inside communities against the density of links outside communities as proposed in [9,15,20]. The most popular approach regarding the modularity metric is the algorithm introduced in [1].

During the last years, sentiment analysis methods that can accurate recognize emotional status of users in social networks are of extreme interest [18] and many studies indicate the important role they can play in the analysis of users behavior and also the analysis of public attitude towards events and topics.

The automatic recognition of emotional presence in tweets can assist in creating more sophisticated social and personal applications as well as in the study of social relations [25]. In the work presented in [17], authors study emotion recognition in tweets; they developed a corpus of 5.553 tweets manually annotated with 28 emotion categories and in following they study the performance of machine learning algorithms in fine grained classification. The results indicate that classifiers can detect 28 emotion categories in text without a huge drop in performance compared to coarser - grained classification schemes. In addition, in [2], a knowledge based approach for recognizing emotions in text is examined and a tool is presented. This tool utilizes versions of SentiWordNet lexical resource, which is a subset of WordNet-Affect with some manual added words. Specifically, this corresponding tool detects contrasts between positive and negative words that shift emotion valence. Ferrada and Camarinha-Matos [7] present an approach for developing an emotions-oriented system for assisting in monitoring and managing the interactions among users in collaborative networks. The supervision of the emotional interactions within collaborative networks provides multi-modal emotional input for achieving awareness of the participants and also offers mechanism to promote the emotional health of the network.

Xu et al. [29] present two methods for detecting communities in social networks. In the first method, sentiment can be positive or negative, while the second method assumes that the range of sentiment is divided into intervals and then users are categorized into groups based on the differences in the ranges of sentiment values. Moreover, Deitrick et al. [5] utilize

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

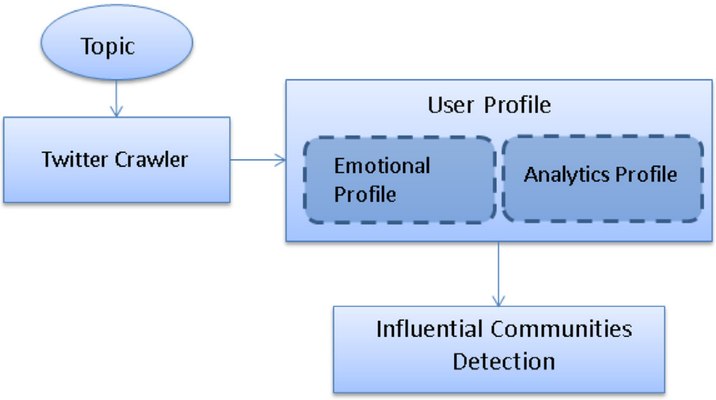

Fig. 1. System architecture.

Naive Bayes classifiers that are trained on Sanders dataset in order to improve the modularity that may yield benefits in the community detection processes as well as user recommendation. In addition, another work that help in community detection processes and user recommendation is the one proposed in [11], where authors study the contribution of user sentiments; their method aims towards the suggestion of similar users that belong to the same communities.

Although social network methods are quite important in the analyses of social networks, most of the techniques rely on node connectivity and neglect special characteristics of the nodes. In this paper, a different perspective is taken as it is believed that in social networks users' personality such as their emotional behavior are of great importance and could provide indicative information regarding users/nodes concerning the graph network. To the best of our knowledge, there are no other efforts that utilize users' emotional behavior to enhance the performance of community detection methods in contrast to our work that utilizes users' emotional behavior as specified on Ekman's scale.

3. Proposed method

In this section, we present the introduced methodology for analyzing and modeling the conversations on specific topics in Twitter network. The methodology analyzes user's tweets and specifies their emotional content on Ekman scale [6] which specifies six basic human emotions: "anger, disgust, fear, happiness, sadness, surprise". After that, user's influence in the network is determined in order to identify the more influential communities in the specific network. The extracted influential communities are utilized based on the posts' emotional content as well as on the user's influence and can be thereafter considered as the representation of the emotional interactions in the corresponding network. The architecture of the proposed system is depicted in following Fig. 1.

3.1. User emotional profile

The emotional content of user's tweets is specified by the knowledge-based tool presented in [22] is utilized. The workflow of the tool is illustrated in Fig. 2. The tool recognizes the existence or the absence of the six basic emotions that were proposed by Ekman [6], in natural language sentences.

Initially it analyzes the structure of the sentences using Part-of-Speech tagging and parsing procedures with the aim to extract meaningful information regarding the sentence structure and the grammatical roles of the words. So, the first level of analysis concerns the morphosyntactic analysis conducted by the tree tagger which is used in order to specify each word's grammatical role as well as to determine its base form (lemma). Then, a deeper analysis of the structure of the sentence is performed by Stanford Parser and the exact relationships between the words of the sentences are specified and further analyzed.

Stanford Parser analyzes natural language sentences and annotates the sentence words based on their interaction in the sentence [4,19]. The sentence dependency tree that it is created represents the exact relationships between the words of the sentence. In the tree, each node represents a word and edges are labeled by the relationship. These relationships are represented as triplets where each one consists of the relationship's name, the governor and the dependent respectively. After the formulation of the dependency tree, special parts of the sentence structure and specific words are further analyzed based on the knowledge base of the tool.

The knowledge base utilizes lexical resources and stores information regarding emotional words that convey specific emotional content. The emotional words is performed are spotted with the utilization of the WordNet Affect source [26], which has been integrated into the tool. WordNet-Affect extends WordNet, by providing a subset of synsets suitable to represent affective concepts correlated with affective words [27]. In general, WordNet synsets are assigned with one or

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

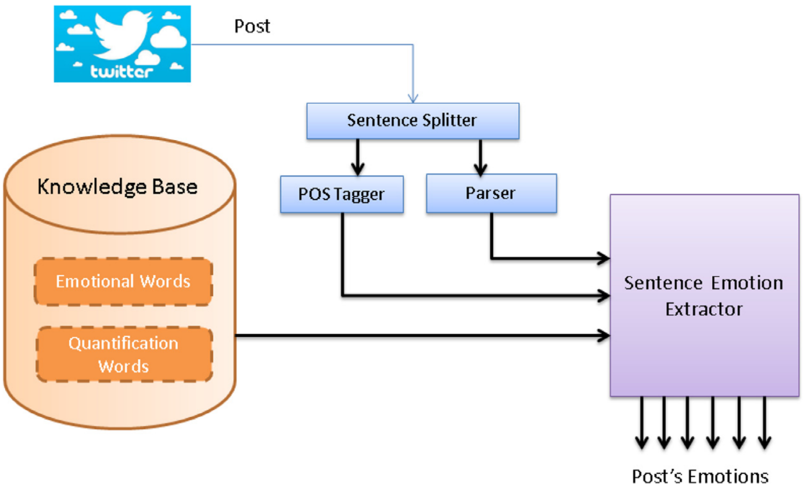

Fig. 2. Architecture of emotion recognition tool.

more affective labels. So, the affective concepts representing emotional state are individuated by synsets marked with the emotion labels.

So, the tool, given a sentence, it analyzes it and gives as output the emotional status of the sentence on Ekman's scale.

3.2. User analytics profile and influence metric

In this subsection, the methodology for estimating the influence (in terms of importance) of a user in a Twitter Graph is described. Concretely, as already stated in our previous works [12–14], our proposed schema could be utilized as a graph and as a result Twitter users could be represented by nodes. As a matter of fact, the edges which connect these nodes (Twitter users) are the relations entitled "Follower to Following", commonly introduced by Twitter.

As it is already shown in previous works, the influence metric should not only be based on the number of "Followers" that each Twitter user has, even in cases where this metric is big enough and thus user's tweets can be potentially received by a large number of user's "Followers". Concretely, this ratio is not sufficient and therefore, the recently introduced ratio of "Followers to Following" (FtF ratio) can be considered as a much more suitable factor. As another important metric, one can consider the number of user's Tweets. To gain a better insight into this aspect, users with higher number of posts have obviously more impact on the corresponding network. Moreover, we have also used the Frequency of user's Tweets, which depicts whether a user enjoys posting more frequent than other users.

Furthermore, we take into consideration some additional features that measure the interaction among different users in Twitter. Precisely, the number of Retweets and Replies shows whether a user enjoys republishing other users' posts or commenting on them. What is more, Retweets can be very helpful in identifying content and even web trends that interests users' followers or other users. A great amount of impact can be presented when using other three factors, namely Clicks, Favorites and Mentions; these features indicate that some users' posts may have more interest compared to other users' posts. In detail, the number of Clicks shows the number of times that users have selected a specific post from a specific user, whereas with the use of Mentions, one user can reach other users who addressed a post and do not have a "Follow" relationship amongst each other.

All the aforementioned metrics are combined so as to introduce the following Eq. (1), entitled PostImpact, which deals with posts' characteristics:

$$Post\_Impact = \frac{(Clicks+1)*(Favorites+1)*(Mentions+1)}{Tweets} * \frac{(Replies+1)*(Retweets+1)}{Tweets} \tag{1}$$

In the present work, the latest k tweets of the user for calculating the above rates according to the Twitter API, are processed; for our experiments, please see the following subsection for values of k. The proposed Influence Metric depends on all of the aforementioned features and metrics of the examined user and is presented in the following Eq. (2). Thus, the Influence of a Twitter user, based on the above parameters, is calculated as follows:

$$Influence\_Metric = Post\_Impact * Frequency * \log(FtF+1) \tag{2}$$

The above Influence Metric depends on all of the abovementioned characteristics of each user. One important aspect of Eq. (2) is that the FtF ratio is placed inside a base-10 log for avoiding outlier values. In addition, the ratio is added by 1 so as to avoid the metric being equal to 0 in cases that the value of "Followers" is equal to "Following"; if the value of "Followers" is equal to "Following" then $log(FtF)$ is equal to 0. What is more, we have added the ratio of Clicks, Favorites, Mentions, Replies and Retweets divided by the absolute number of Tweets. The 5 proposed ratios are also added by 1 so as to avoid the metric being equal to 0 in cases that Clicks, Favorites, Mentions, Replies or Retweets are equal to 0.

3.3. User emotional behavior

The accurate specification of the emotional behavior of the users necessitates two piece of information to be determined. The first deals with the quantity and consequently the frequency of user's tweets in order to further analyze and determine the emotional status of each specific user and the second concerns the specification of the user's overall emotional status based on the content of each emotional post.

Regarding the first one, the methodology analyzes user's tweets fetched the last 3 weeks in order to determine their corresponding emotional status regarding the specific period. The time window of 3 weeks has been set based on evaluation results as well as on the principle that user's emotional status can dynamically change during the passage of time. By setting a too narrow time window, one can state that a decent and balanced amount of the user's post activity would fail to be provided. On the other hand, a wider time window would fail to specify accurately as well as meaningfully represent the alternation of each user's emotional status.

In addition, user's tweets in the last 3 weeks are retrieved by the crawler and are emotionally annotated by the aforementioned tool. As long as user's tweets are analyzed and in following emotionally annotated, then user's recent emotional status can be determined. For each tweet, the tool can measure whether it conveys one out of the 6 basic emotions defined by Ekman.

Furthermore, the overall user's emotional status is calculated based on the emotional annotation of each tweet in the specific time window. Being determined from the specification whether users have a vivid emotional pulse or are emotionally neutral, they are characterized to carry a sensitivity status if at least 10% of their posts are recognized as emotional and convey one or more feelings; otherwise, they are set to be neutral.

The value of threshold equal to 10% is set based on experiments employed on different Twitter datasets that are presented in Section 4. In general, emotions in Twitter posts can vary and thus show a highly skewed distribution. In most cases, $10-15\%$ of the posts were recognized so as to convey emotions. Moreover, the analysis of the emotional tweets revealed that some emotions are more popular than others; e.g. joy (happiness) can be present in up to 50% of the emotional tweets while disgust or surprise can be present in less than 10%. For these reasons, the threshold of 10% seems to be appropriate, thus giving a balanced ratio between emotional and neutral users' posts.

In addition, in cases that a user is determined to have an emotional status in the specific time window, their emotional profile on Ekman's emotional scale is utilized. This is implemented as all different emotions, which the specific user has expressed through their posts, are aggregated and as a result, these emotions compose user's emotional behavior (profile) in terms of existence or absence. So, the user's emotional status aims to provide a deeper and more complete insight of user's behavior and characteristics as it expresses the entirety of their posts' activity.

3.4. Community detection

Since our motivation stems from the fact that we are interested in identifying the most influential communities in the produced users' graph, we utilize one of the most common, modularity based, community detection algorithm [1], with the difference of adding an additional conversion/transformation as a pre-processing step.

More to the point, the proposed influential community detection approach combines the modularity optimization of network community structure with the emotional state of each user's retrieved tweets in the graph. This corresponding information is taken into account as the transformation of the retrieved graph to its dual graph, which is known as line graph, is utilized. In following, we have used the weighted version of community detection algorithm [1] in order to extract the influential communities in terms of a ranked list. As a final step, the transformation of the line graph to its dual is introduced so as to effectively match the extracted communities based on the initial retrieved social graph.

Comprehensively, the methodology is modulated in the following steps:

1. *Conversion to ine graph*, where line graph is the dual of an initial graph; a dual is the inverted nodes-edges graph. An example of the corresponding conversion is depicted in the following Fig. 3. More to the point, the users (e.g. nodes) are represented by the vector of their tweets' emotional scale based on Ekman model, where a value equals to 1 means that the corresponding emotion is present and a value equals to 0 means that it is absent. On the other hand, the edges between these nodes represent the "Following" relationship where we can observe that they have different labels as they are connected to different nodes.
2. *Utilization of weighted community detection algorithm*, which is a modularity optimization method described in [1] so as to identify communities within the Twitter network.
3. *Conversion to initial dual node graph*, which is the reverse of the aforementioned step 1.

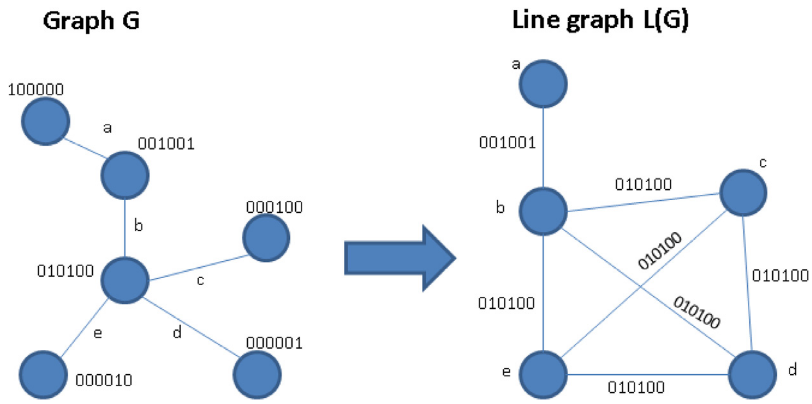

Fig. 3. Conversion to line graph.

3.5. Emotional graph representation

In this subsection, the weight of the emotional edge interactions between two nodes is calculated. Namely, each one of the 6 Ekman emotional scales is represented by a node while the emotional interactions between two nodes are represented by the corresponding edge.

Initially, each topic is characterized by the way that the corresponding information is expressed in tweets. So, the proposed emotional graph represents the emotional interaction between the users that took part in the topic with the use of the according tweets. Moreover, the emotional graph is based on each post's emotional status as well as the user's influence that made the post.

Concretely, the emotional graph is a weighted directed graph $G = (V, E)$; where the nodes represent the basic human emotions as defined on Ekman's scale and the edges represent the emotional alternations. More specifically, a directed edge starting for node V_1 and ending in node V_2 represents that the emotion of node V_1 is annotated in users' comments on the emotion of the node V_2. We also assume that there is a weight in each edge in a way that represents the intensity of these two emotion interactions.

The weight is calculated as follows:

$$W_{emotion1 \rightarrow emotion2} = \frac{\sum_{i=1}^{m} Inf\_Metric_i * Post_i}{\sum_{j=1}^{k} Inf\_Metric_j * Post_j} \tag{3}$$

where k represents the number of posts made on a certain topic which are emotionally annotated with the emotion "emotion1". Accordingly, the parameter m represents the number of posts with emotion "emotion1" in response to a post with emotion "emotion2". Post i represents the emotional post i made and the parameter Inf_Metric represents the influence metric of the user who made the post i. The value of the weight takes values in the range [0, 100]. In addition, this weight indicates the weighted portion of users' tweets containing the emotion "emotion1" in order to reply or to comment to a specific tweet expressing emotion "emotion2".

So, a value close to 0 denotes that the posts annotated with "emotion2" are not commented by users' posts which are annotated with emotion "emotion1". In contrast, a value close to 100 denotes that the initial post annotated with emotion "emotion2" is emotionally commented on by users with emotion "emotion1".

4. Implementation

4.1. Twitter datasets

We based our experiments on Twitter and used Twitter API in order to collect tweets. The Twitter4J[1] is used for collecting tweets that have been addressed during various periods of time and also for a diversity of topics with use of corresponding keywords.

For the development of the corpus from different topic conversations, we have collected the tweets made by popular news portals regarding specific events and topics where each topic consists of at least 10.000 posts. Furthermore, another very important issue is that Twitter users who made one or more comments to the initial post as well to other people's tweets, were analyzed in order for their influence to be estimated with use of Eq. (2). In following, all the corresponding tweets (either posts or comments etc.) made on the topic were emotionally annotated based on Ekman's model. In Table 1,

[1] Twitter4J: http://twitter4j.org/en/index.html.

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

Table 1
Categories and topics used.

Category	Topic
Accident	Malaysia Airlines Flight 370 disappear
Accident	Philippine Typhoon Kills Thousands
Arts	Spectre
Economy	Stock market
Economy	Greek Bailout
Science	Space Expedition 38
Social	European Parliament Election 2014
Social	Obamacare
Social	SyrianRefugees
Sports	Maria Sharapova wins Grand Slam
Sports	Brazil vs. Croatia (World Cup 2014)

Table 2
Distribution of tweets.

Topic	Emotional	Neutral
Malaysia Airlines Flight 370 disappear	67%	33%
Philippine Typhoon Kills Thousands	75%	25%
Spectre	27%	73%
Stock market	35%	65%
Greek Bailout	61%	39%
Space Expedition 38	55%	45%
European Parliament Election 2014	37%	63%
Obamacare	31%	67%
SyrianRefugees	43%	57%
Maria Sharapova wins Grand Slam	21%	79%
Brazil vs. Croatia (World Cup 2014)	25%	75%

Table 3
Percentages for Ekman emotional scales.

Topic	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Malaysia Airlines Flight 370 disappear	9%	2%	9%	5%	65%	10%
Philippine Typhoon Kills Thousands	9%	2%	13%	0%	68%	8%
Spectre	3%	4%	6%	43%	32%	12%
Stock market	5%	9%	42%	22%	19%	3%
Greek Bailout	6%	4%	22%	6%	51%	11%
Space Expedition 38	1%	2%	5%	75%	8%	9%
European Parliament Election 2014	6%	3%	5%	39%	34%	13%
Obamacare	6%	8%	15%	49%	9%	13%
SyrianRefugees	23%	4%	19%	2%	47%	5%
Maria Sharapova wins Grand Slam	3%	5%	2%	55%	20%	15%
Brazil vs. Croatia (World Cup 2014)	3%	1%	1%	65%	23%	7%

the topics studied as well as their related category are presented. News portals contain BBC, CNN, Huffington Post, New York Times, Reuters, etc.

In order to get an insight regarding users' emotional attitude, the number of Tweets that express specific emotional dimensions versus the number of Tweets that do not express any emotional dimension, is calculated and is depicted in Table 2. We can observe that these percentages depend mainly on the topic of the discussion; e.g. the emotional tweets are 75% regarding the rather emotional topic "Philippine Typhoon Kills Thousands", while the percentage of the emotional posts is 27% when considering the more neutral topic regarding the movie "Spectre".

What is more, the aforementioned table shows that topics that do not seem to contain tweets with emotional dimensions, such as "Maria Sharapova wins Grand Slam" and "Brazil vs. Croatia (game of World Cup 2014)", they still have a considerable percentage of emotional tweets, i.e. about 20% to 25%. On the other hand, the topic entitled "Space Expedition 38" contains more than half emotional posts, and as will be shown in the next subsection, the majority of these emotional posts are categorized in the dimension of Happiness.

Furthermore, in Table 3, the topics studied along with their six corresponding Ekman emotional scales are presented. In the context of this study, we tried to select and in following to examine topics based on the principle that they possess diversity in their emotional content. Thus, the eleven topics are emotionally rich and as a result they demonstrate a diversification in their distribution. The emotional analysis of the tweets indicates that the emotion of Happiness is the predominating emotion in five (or six considering "European Parliament Election 2014") out of the eleven topics, whereas the emotion of Fear achieves higher percentage in one topic and the emotion of Sadness in the remaining four. Indeed,

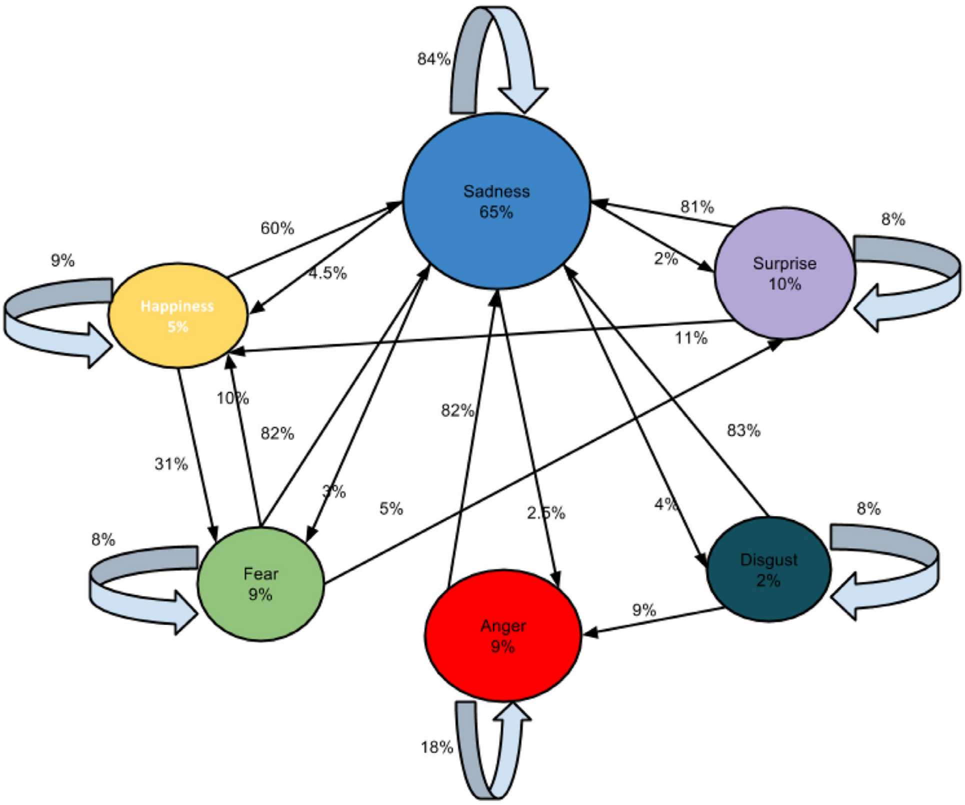

Fig. 4. Emotional graph for topic “Malaysia Airlines Flight 370 disappearance”.

regarding the topic entitled “Malaysia Airlines Flight 370 disappearance”, the emotion of Sadness is expressed in about 65% out of the emotional tweets, while the emotion of Happiness achieves approximately the half percentage in “Spectre” and “Obamacare”.

4.2. Emotional graph instance

We consider the emotional graph regarding the topic “Malaysia Airlines Flight 370 disappearance”, which is illustrated in Fig. 4 and contains nodes and edges. More specifically, the nodes correspond to the six emotional dimensions and the bigger a node is, the higher the percentage of the corresponding emotion in users conversation. What is more, the value of an edge indicates the weight portion of users’ tweets, which contain the emotional dimension “emotion1”, so as to comment or reply to a specific tweet, which contain the emotional dimension “emotion2”.

Regarding the abovementioned topic “Malaysia Airlines Flight 370 disappearance”, the total numbers of tweets analyzed are 11.000, out of which the 7.370 were annotated with emotions while 3.630 were neutral. As expected, the emotional dimension of Sadness was the prevalent emotion in most of users’ posts. Concretely, 4.790 posts were annotated with the corresponding emotion, thus revealing the disappointment as well as the users’ frustration about the accident.

In contrast, 368 posts were annotated with the emotional dimension of Happiness, thus showing the users’ expectation and hope for a happy outcome, mainly after the new founding from the rescue team. In addition, Anger was found in 663 users’ posts showing the frustration about the rescue actions, while Fear was found in 660 users’ posts stating the unpleasant emotions regarding the corresponding accident. Surprise was found in 739 posts revealing users’ thoughts about potential extraordinary scenarios, such as alien kidnapping, etc. Finally, Disgust achieves the smallest percentage with 150 posts.

5. Evaluation

In a case study of the methodology, the graph for the topic of the movie #Spectre is used and the corresponding graph consists of 10.000 nodes. In addition, each node represents a user that addressed a post in the corresponding topic. The performance of our methodology is illustrated in the following Figs. 5–8 and is compared versus the modularity optimization method in determining the influential communities (regarding a topic or an event for a specific time period). Namely,

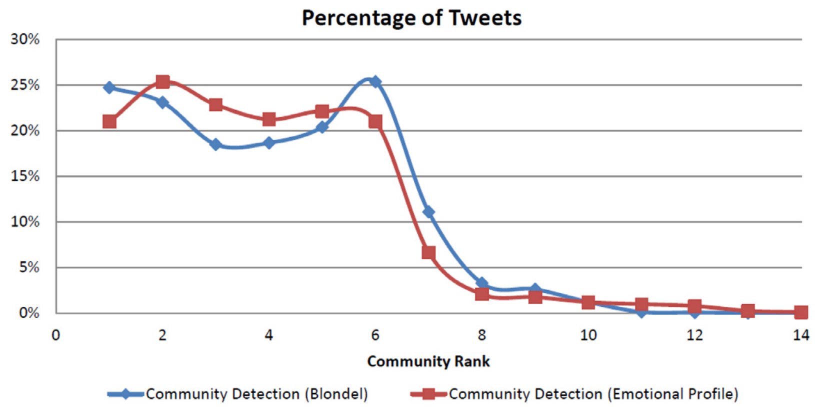

Fig. 5. Influential community detection methods based on the percentage of tweets.

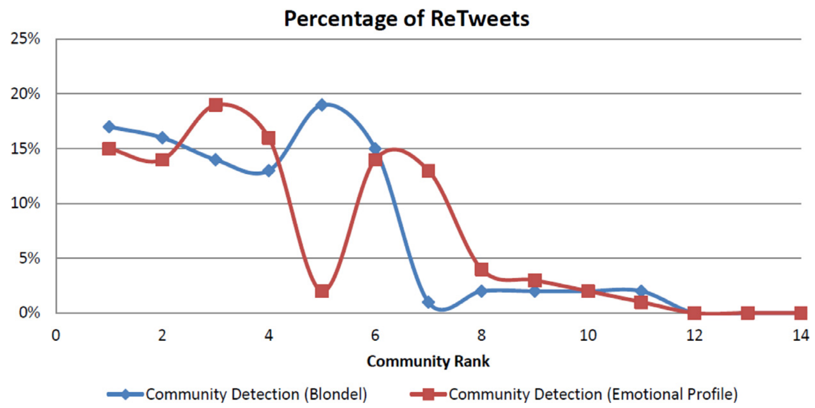

Fig. 6. Influential community detection methods based on the percentage of ReTweets.

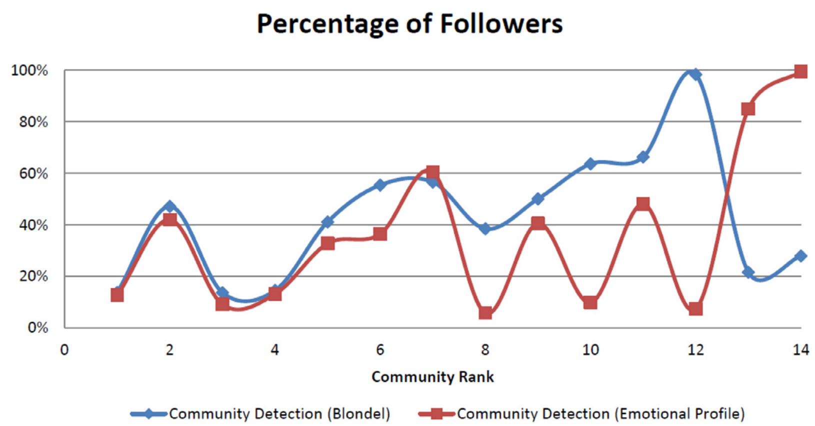

Fig. 7. Influential community detection methods based on the percentage of followers.

the influence of the biggest, in terms of users, communities using different metrics for different application scenarios, is depicted.

In these figures, the extracted communities (for each one out of the four cases) are ranked based on the Influence Metric which has been described in Eq. (2). Since our motivation stems from the fact that we are interested in identifying the more influential communities and not just the first one, our examination is focused on the first 5 ranked communities. As a result, Figs. 5–8 present the 14 ranked communities in which the percentage of Tweets, ReTweets, Followers and Size for each Community is respectively considered.

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

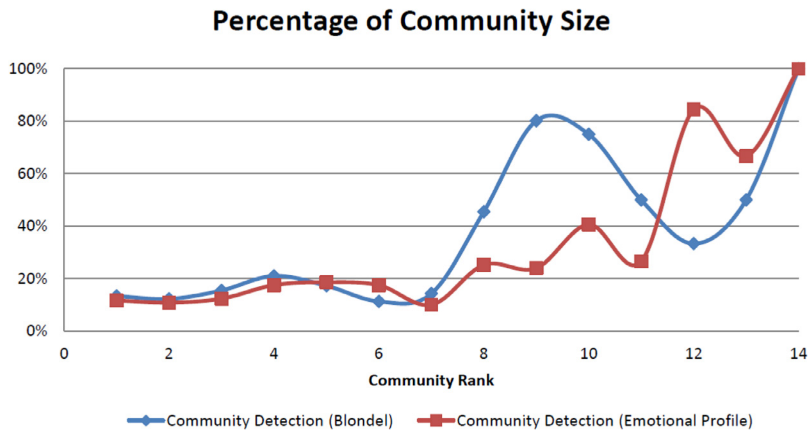

Fig. 8. Influential community detection methods based on the percentage of community sizes.

Table 4
Normalized metrics for estimating influential communities.

Influential communities detection	Tweets / Size	Followers / Size
Community detection based on modularity [1]	1326	1,640
Emotional community detection	1587	1,549

As it is obvious in Fig. 5, the proposed emotional approach produce more Tweets than influential communities detected only from modularity optimization community detection algorithm [1]. More specifically, the emotional approach performs slightly better for the first 5 communities.

In addition, Fig. 6 presents the percentage of ReTweets for each produced community computed by the two compared approaches. The two approaches seem to have identical percentages of ReTweets per community, except some specific community number; for example communities 3 and 7 have more ReTweets in the case of the emotional approach whereas communities 1 and 5 gain a bigger percentage of ReTweets for the simple community detection methodology.

We can observe in Fig. 7 that influential communities based on the emotional factors slightly decrease the percentage of Followers in the top 5 communities as compared to Blondel et al. [1]. One potential reason is the fact that the proposed Influential Metric in Eq. (2) is more generic as it deals with an overall estimation of each user's impact in the produced community; this estimation relies on the number of Clicks received, Mentions, Replies, Retweets, etc.

As it is obvious from Fig. 8, the communities of the proposed emotional methodology require fewer nodes that the modularity algorithm [1]. This occurs due to the fact that the modularity optimization community detection algorithm is affected by the inequality of the weights distribution in the connected nodes. Of course another potential reason could be the density of links inside the communities compared to the density of links between communities. In addition, this metric can be very useful, especially in cases where the cost is associated with the size of the communities and as a result, smaller communities, having larger impact, are required.

Table 4 proves the aforementioned analysis as the Tweets/Size and Followers/Size metrics are totally consistent with the results of the above Figures. Taking into considerations the meaningful metrics, we observe that the communities extracted from our proposed approach are denser than from the modularity optimization method.

Specifically, the results show that the emotional approach results in a higher number of Tweets per Community. This result denotes that the finer and more sophisticated proposed approach drives into a more structured formulation of the arising communities. On the other hand, the number of Followers per Community is slightly lower when considering the emotional approach and this can be attributed due to the fact that Influential Metric deals with an overall estimation of user's impact in the produced community. Summing up, Table 4 supports the rational that the communities detection could benefit from the meaningful data of users' emotional behavior.

6. Conclusions and future work

In this paper, we proposed a novel method for identifying influential communities in social networks with use of users' emotional behavior as well as users' influence in a specific timeframe. The proposed method is based on the emotional content of each post addressed by a user as well as on her influence metric as long as she participates in a specific topic. Thereafter, since all users are assigned a specific influence metric and in following are modeled as emotional or neutral, the proposed method identifies the most influential communities based on these two features. The results obtained support our schema for identifying influential communities as it outperforms the popular modularity optimization method used in related works.

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

Future work will focus on scalability problems that emerge when analyzing bigger graphs. In addition, it is in our keen interest to examine the performance of the methodology using several subjects and identify the parameters that influence its performance in a finer granularity level. Another key aspect of our future work will be the integration of our approach in initiatives and existing methods for viral marketing and for branches' advertising purposes. Finally, another aspect to examine is the evolution of influential communities in time with the utilization of temporal networks.

References

[1] Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of community hierarchies in large networks. J Stat Mech 2008;P10008.
[2] Chaumartin F. UPAR7: a knowledge-based system for headline sentiment tagging. In: International workshop on semantic evaluations (SemEval@ACL); 2007. p. 422–5.
[3] Choudhury MD, Gamon M, Counts S. Happy, nervous or surprised? classification of human affective states in social media. International AAAI conference on weblogs and social media (ICWSM); 2012.
[4] De Marneffe M-C, MacCartney B, Manning CD, et al. Generating typed dependency parses from phrase structure parses. In: International conference on language resources and evaluation (LREC); 2006. p. 449–54.
[5] Deitrick W, Valyou B, Jones W, Timian J, Hu W. Enhancing sentiment analysis on twitter using community detection. Communications and Network 2013;5(3):192–7.
[6] Ekman P. Basic emotions. New York: John Wiley & Sons Ltd; 1999.
[7] Ferrada F, Camarinha-Matos LM. Emotions in collaborative networks: a monitoring system. In: Doctoral conference on computing, electrical and industrial systems (DoCEIS); 2012. p. 9–20.
[8] Ferrara E. A large-scale community structure analysis in facebook. EPJ Data Sci 2012;1(9).
[9] Fortunato S. Community detection in graphs. Phys Rep 2010;486(3–5):75–174.
[10] Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci 2002;99(12):7821–6.
[11] Gurini DF, Gasparetti F, Micarelli A, Sansonetti G. iscur: nterest and sentiment-based community detection for user recommendation on twitter. In: International conference on user modeling, adaptation, and personalization (UMAP); 2014. p. 314–19.
[12] Kafeza E, Kanavos A, Makris C, Vikatos P. T-pice: twitter personality based influential communities extraction system. In: IEEE International congress on big data; 2014. p. 212–19.
[13] Kanavos A, Perikos I, Vikatos P, Hatzilygeroudis I, Makris C, Tsakalidis A. Conversation emotional modeling in social networks. In: IEEE International conference on tools with artificial intelligence (ICTAI); 2014. p. 478–84.
[14] Kanavos A, Perikos I, Vikatos P, Hatzilygeroudis I, Makris C, Tsakalidis A. Modeling retweet diffusion using emotional content. In: Artificial intelligence applications and innovations AIAI; 2014. p. 101–10.
[15] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis: Invited presentation, extended abstract. Tech. Rep. Physical Review E80, 056117; 2009.
[16] Leskovec J, Lang KJ, Mahoney MW. Empirical comparison of algorithms for network community detection. In: International conference on world wide web (WWW); 2010. p. 631–40.
[17] Liew JSY, Turtle HR. Exploring fine-grained emotion detection in tweets. In: Conference of the North American chapter of the association for computational linguistics: human language technologies; 2016. p. 73–80.
[18] Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Mining text data; 2012. p. 415–63.
[19] Manning CD, Surdeanu M, Bauer J, Finkel JR, Bethard S, McClosky D. The stanford corenlp natural language processing toolkit. In: Annual meeting of the association for computational linguistics (ACL); 2014. p. 55–60.
[20] Newman MEJ. Fast algorithm for detecting community structure in networks. Phys Rev E 2004;69(6):066133+.
[21] Papadopoulos S, Kompatsiaris Y, Vakali A, Spyridonos P. Community detection in social media. Data Min. Knowl. Discovery 2012;24(3):515–54.
[22] Perikos I, Hatzilygeroudis I. Recognizing emotion presence in natural language sentences. In: International conference on engineering applications of neural networks (EANN); 2013. p. 30–9.
[23] Plantie M, Crampes M. Survey on social community detection. In: Social media retrieval, Computer Communications and Networks. Springer; 2013. p. 65–85.
[24] Quercia D, Ellis J, Capra L, Crowcroft J. Tracking "gross community happiness" from tweets. In: ACM conference on computer supported cooperative work (CSCW); 2012. p. 965–8.
[25] Sintsova V, Musat CC, Pu P. Semi-supervised method for multi-category emotion recognition in tweets. In: IEEE International conference on data mining workshops (ICDM); 2014. p. 393–402.
[26] Strapparava C, Valitutti A. Wordnet-affect: an affective extension of wordnet. In: International conference on language resources and evaluation (LREC), 4; 2004. p. 1083–6.
[27] Valitutti A, Strapparava C, Stock O. Developing affective lexical resources. Psychnol J 2004;2(1):61–83.
[28] Wang W, Chen L, Thirunarayan K, Sheth AP. Harnessing twitter "big data" for automatic emotion identification. In: Conference on privacy, security, risk and trust (PASSAT) and confernece on social computing (SocialCom); 2012. p. 587–92.
[29] Xu K, Li J, Liao SS. Sentiment community detection in social networks. In: iConference; 2011. p. 804–5.
[30] Yang B, Liu D, Liu J. Discovering communities from social networks: methodologies and applications. In: Handbook of social network technologies and applications; 2010. p. 331–46.

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011

Andreas Kanavos is an Adjunct Lecturer at the Department of Computer Engineering and Informatics at University of Patras. His research interests span the broad areas of Data Mining, Graph Mining, Social Network Analysis, Information Retrieval, Bioinformatics and String Algorithmic. He has published over 40 papers in refereed scientific conferences and journals.

Isidoros Perikos is a Post-Doctoral researcher at the Department of Computer Engineering and Informatics at University of Patras. His main research interests include Artificial Intelligence, Data Mining and Knowledge Extraction, Human Computer Interaction, Natural Language Processing and Affective Computing. He has published over 50 papers in international conferences, journals and workshops. He is a member of IEEE and ACM.

Ioannis Hatzilygeroudis is currently an Associate Professor at the Department of Computer Engineering and Informatics at University of Patras. He is an Associate Editor-in-Chief of the International Journal of Artificial Intelligence Tools (IJAIT). He has also been member of the PCs of over 100 AI related conferences. He has published over 100 papers in international journals, edited volumes and proceedings.

Athanasios Tsakalidis is a Professor at the Department of Computer Engineering and Informatics at University of Patras. He is one of the contributors to the writing of the "Handbook of Theoretical Computer Science". He has published many scientific articles having a especial contribution to the solution of elementary problems in the area of Data Structures.

Please cite this article as: A. Kanavos et al., Emotional community detection in social networks, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.09.011