Datawarehouse-based approach for the analysis of terrorism-related activities in social networks

Kamel AHSENE DJABALLAH, Kamel BOUKHALFA
Laboratory of Computer Systems, University Science and Technology Houari Boumediene
adjab.kamel@gmail.com
boukhalk@gmail.com

Omar BOUSSAID
ERIC laboratory, ERIC Laboratory, Institute of Communication, University of Lyon, Lyon 2
omar.boussaid@univ-lyon2.fr

Abstract— Social network analysis techniques for activities related to terrorism are mainly based on data mining techniques. These techniques do not take into account the various axes of analysis allowing a study according to several facets. In this article we propose a comparison study of these techniques. We present our approach of analyzing these activities, based on Data warehouse and OLAP analysis. We aim to improve the analysis of these cyber threats. OLAP analysis allows us to explore social networks to detect dangerous content in the direction of targeted cyber threats. Our approach is based on five-tier architecture: (1) data sources; (2) ETL; (3) Data warehouse; (4) Analysis; (5) Presentation. In our experimentations, we used Twitter to detect and analyze the incitement to terrorism and determine the users supported the terrorism. We proposed a datamart with a metric named score, calculated using a data mining technique. Also, we used OLAP analysis techniques, based on the history of positive scores, to determine the users inciting terrorism, their locations and their retweets.

Keywords— Social Neworks; Data Warehouses; Terrorism; Sentiment Analysis; OLAP; Twitter

I. Introduction

In recent years, social networks have become an essential tool for communicating, sharing and exchanging information and joining interest groups. Users of these networks can create information, share it, and express their opinions on any subject. The latter can be political, economic, safe, social [9]. Currently, social networks concern millions of people, for example, in 2017, Facebook has 3 billion subscribers and Twitter has 500 million [22].

Nevertheless, there are several cyber threats in these social networks, such as social engineering, phishing, propaganda and activities related to terrorism. In [18] ten (10) main cyber threats in social networks have been described.

In addition, the analysis of social networks has attracted the interest of several researchers, who for the detection of cyber threats or for other purposes.

To analyse social networks, many techniques have been used such as graph analysis, data mining, analysis based on the text clustering [1], and other based on warehousing [15].

In this article, we present and study some analysis methods and more especially those allowing the activities related to terrorism, which are based mainly, on data mining. However, we can note that these analysis methods do not take into account the analysis according to detect the different faces of these cyber threats, like activities related to terrorism. Also, these analyzes were performed over periods well determined, without taking into account the historical analysis. The approach we propose based on OLAP storage allowing analysis according to several axes and taking into account the historization of the data.

The rest of this article is organized as follows. Section 2 draws a state of the art on approaches analysis of social networks and more particularly the work relating to the terrorism activities analysis. The proposal of our data warehousing approach for the detection of cyber threats is detailed in section 3. Section 4 shows the experiment of this approach. We finish this article with a conclusion and future work.

II. Related work

Analysis and understanding of social networks, like Facebook, twitter, LinkedIn, etc. arouse strong interests in several scientific communities. Many approaches have been proposed, among them those based on data warehousing and data mining.

A. Warehousing approaches

Warehousing approaches propose to store data in a data warehouse for multidimensional analysis. They can be subdivided, according to [15], into two families:

Behavior based, which focus on the analysis of the activities of the Internet users on social networks [24], [25] such as friendship creation, group creation or browsing profiles, and how people interact with an event.

Those based on the sentiment analysis, which integrate the sentiment analysis in the schema of the data warehouse [26], [27]. In these approaches we analyze the social, psychological, philosophical, behavioral aspects and the perception of a person or a group people about a product, policy, services and specific situations.

B. Data mining approaches

Concerning data mining approaches, there are several works, the most used are web methods mining, clustering and classification [19]. Other work based on data mining techniques varies from unsupervised learning, to semi-supervised and supervised learning [3].

In this article, we will focus on work related to the sentiment analysis that has aroused the interest of several researchers, such as [21], [13] and [2]. Also, sentiment analysis is used in detecting the activities related to terrorism, in social networks [20]. It should be noted that we study these cyber threats since they are the most common and widespread in the social networks.

C. Analysis of activities related to terrorism in Social Networks

There is panoply of work based on the sentiment analysis for analysis of activities related to terrorism in social networks. A state of the art on these works shows that there are two approaches, the first one based on machine learning [17], [20] and the second based on the lexicon [4], [16].

- Machine learning approaches

The approaches based on machine learning are classified in supervised learning and unsupervised learning.

Concerning supervised learning approaches, [20] proposed a terrorism detection method based on sentiment analysis using supervised learning. In this work, the authors took into account the history of Tweets of a user. That is to say, after categorizing the sentence into positive, negative, and neutral categories, all of these categories were compared to a user's previous sentence of a particular user, based on the sentiment score for the last and precedent sentences.

Also, [12] dealt with the detection of terrorist activities using sentiment analysis in a distributed system. The main aim of this work was to propose architecture for the detection of terrorist activities in streaming, valid for several social networks and this, in a BIG DATA environment. For this, they proposed architecture based on a web API for collecting data from social networks, a data collector (Apache Haddop) and a processing language based on artificial intelligence for the extraction of meaning from text.

Similarly, [17] discussed the use of machine learning for the identification of jihadist messages on the social network Twitter. In this work the author used supervised learning for classifying tweets as radical and not radical. We summarize his method as follows: after cleaning the tweets and determining the characteristics vectors, which are the writing style, the time and the sentiment analysis, he used three classifiers to experiment and validate his learning which are: SVM, AdaBoost and Naive Bayes.

Equally, [7] present a supervised machine learning framework that exploits metadata, network statistics and temporal functionalities on user activities, to detect users' supporter's extremists and predicting the adoption of content. So, they exploited a dataset containing millions tweets that were manually identified, then reported, and suspended by Twitter, due to their involvement in extremist campaigns.

With regard to unsupervised learning, [8] used data mining tools in Twitter to organize the terrorist vocabulary and identify, through tweeting metadata analysis, the most likely geographic location and identities connected behind the users accounts. To achieve this goal, the author used a list of words and terms relevant to search tweets via the language R. Then he determined, using the frequency of the accounts and the word cloud, the number of "k -means clusters "for the purpose of obtaining user accounts and their associations. These user accounts are then studied with network graphs built using the R, NodeXL and Gephi languages.

- Lexicon approaches

Concerning the approaches based on lexicon, we can cite the work in [4] who used SentiWordNet, to analyze the sentiment for detecting radical content on web forums. For this, the authors proposed a model based on SentiWordNet, Word Net and NLTK to analyze web forums with radical content. This model measures and identifies the polarity of the sentiment and affects the intensity of the one that appears in the web forum. The main objective of this work was to test the effectiveness of SentiWordNet for detection opinions and emotions on the internet.

Furthermore, [16] treated the sentiment analysis in real time for the detection of terrorist activities, using the wordlist dictionary. The purpose of this research was to develop a model that could help establish crime patterns associated with terrorist activities, using information from the sentiment analysis derived from Twitter data. As such, data collected from Twitter was analyzed, in order to bring out rules for the naive bayes classifier.

D. Comparison of work of activities related to terrorism in Social Networks

A study of social network analysis methods for the detection and the analysis concerning the incitement to terrorism show that these methods use different data mining tools, from text mining, machine learning, to taking into account the history of the information contained in these networks, as well as the use of other methods and tools, for some cases. Which shows that detection and the analysis of cyber threats in social networks is a task requiring a lot

efforts and tools. We present in "Tab. 1," a comparative report of the works studied.

TABLE 1. Comparison of methods of analysis relating to activities related to terrorism

Works	Axes of analysis						
					Sentiment analysis		
	User	*Location*	*Time*	*Historical*	*Lexicon*	*M. Learning*	*Other*
Sofea et Izzatdin 2017 [20]				X		X	X
Kocharekar&Jadhav 2017 [12]						X	X
Omer 2015 [17]			X			X	X
Ferrara et al. 2016 [7]	X		X				X
Govand 2016 [8]	X					X	
Chalothorn&Ellman 2012 [4]					X		
Ngoge 2016 [16]	X	X			X		

The analysis of the works above, based on the axes on which is carried out the analysis highlights the following:

- There are works that have taken into consideration the axis of time [17], others the localization [16].
- The sentiment analysis is a technique used in all works. In this analysis, it has been used the approaches based on machine learning and the lexicon-based approaches.
- The works cited do not take into account the history of the content published by the same user, except the work [20], which showed that this account can improve the accuracy of the analysis;
- Whenever there is a consideration of other axes, it improves and enriches the analysis.
- The diversity of the analysis axes and tools used shows that the analysis of incitement to terrorism in social networks is a complex task requiring a multi-faceted analysis.

Although some studies have taken into account some areas in their analyzes, the problem of the presence of irony and sarcasm in the content related to terrorist activities remains unresolved and the methods of sentiment analysis deserve to be improved , including aspects related to the behavior of the user. For example analyze the history of a user.

Also, these approaches are much more focused on detection, for example content analysis, while we also need to take into account other areas of analysis to understand this phenomenon, in particular for decision makers. These approaches will provide an overview of a user's opinion of a period through their history, the opinions of a set of users, their locations or the interactions between these users. This will assist in a broader understanding of this phenomenon and take the necessary decisions to combat certain activities related to terrorism and take appropriate measures, including by government institutions.

As such, as the state of the art has shown that the analysis of terrorist activities in social networks is a complex task that requires a multi-faceted analysis and that the microscopic vision does not respond that we left to this problem, then it is useful to analyze the behavior of users and take into account the various axes, on the one hand, and to have a macroscopic vision to analyze the interactions between users, on the other hand.

So, we can say that it is useful to propose an approach to detect and analyze these activities along several axes. In addition to taking into account the historization. This would help to understand the different aspects to this phenomenon. For example, the evolution of terrorist recruitment networks and better understand some aspects related to the process of radicalization.

While emphasizing that during the operation of the work of storage of social networks, and to the best of our knowledge, no works have used the storage of social networks for the analysis of activities related to terrorism, the storage and analysis OLAP turn out, therefore, a very interesting track to improve the analysis of these cyber threats.

This choice stems from the characteristics of the warehousing which allow the logging and the navigation according to different granularity and several dimensions. These features will allow us a richer analysis and a better correlation of data stored over time. Of course, this OLAP storage and analysis work needs to be combined with other data mining methods.

III. Our approach

The state of the art has shown that detecting the activities related to terrorism in social networks is a complex task that requires multi-faceted analysis, hence, multi-dimensional analysis can respond to the complexity of the analysis process of these activities, and OLAP storage and analysis could improve the analysis of this phenomenon.

This choice stems from the characteristics of storage that allow historization and navigation according to different granularities and several dimensions. These characteristics enable us a richer analysis and offer a better view on data stored over time. Obviously, OLAP storage and analysis job needs to be combined with data mining techniques which are widely used in the literature for to study terrorism.

In our opinion, the analysis of social networks, according to several axes, for example subject, user, location, time, etc. allows to analyze and detect the different aspects related to this cyber threat. Our approach resembles traditional data warehouse approaches in other areas of application, with some specifications in the data sources layer and the use of a Data mining techniques in the ETL layer "Fig.1,".

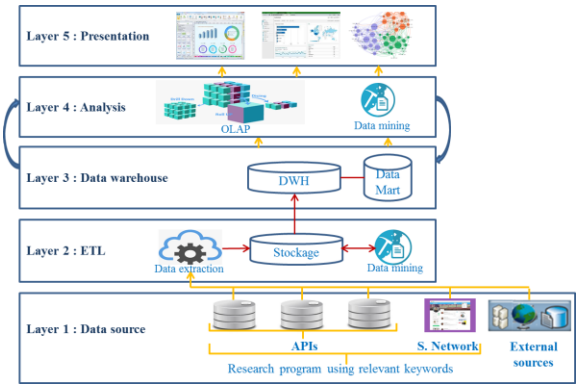
A. *The Data Sources Layer*

This layer is represented by the APIs available for searching social networks data and their metadata, using the connection keys provided by the providers of these networks, through interfaces. In order to start the process of search, libraries (APIs) are used, while introducing the search keywords relevant, to collect data related to a subject. In addition, this layer includes external sources of data that will be exploited when needed, such as location data, for example.

B. *The ETL layer*

In the ETL process we extract data from heterogeneous sources (Extraction). Then, they are transformed to perform necessary treatments and cleanings, using text mining techniques (Transformation). Finally, we load this data in the warehouse (Loading) [5]. In our approach, we also propose to integrate data mining techniques in the ETL process, in order to extract our measurements.

FIGURE 1. Proposed approche for analysis the activities related to terrorism



C. *The data warehouse layer*

This layer is represented by a multidimensional data model, dedicated to the analysis of business processes. It is in this model that fact tables and dimensions as well as their hierarchies are defined. In [6], the authors define the dimensions and indicators of a multidimensional model as the elements to answer the subject of a business process to questions such as the following : Who? What? When? Where? Why? How? How many? They call these seven questions the 7Ws.

The first six questions make it possible to identify the dimensions while the seventh one allows to define the measures in the table of fact. The dimensions associated with a fact represent the context of occurrence of the fact. In our case, the dimensions could be: the user having published content on a social network, its location and the time of publication of this content. Regarding the table of fact, it represents an event or a situation and an indicator constitutes the measure of a fact. In our case a fact is the publication of content in a social network and the extent of the nature of this content deduced by data mining techniques, for example a positive sentiment towards a content related to terrorism.

D. *The analysis layer*

The analysis layer includes an OLAP server, its role is decisive insofar as it translates users' queries into requests on the Data Warehouse and provides the results to decision support tools. It has OLAP cubes, where data is sorted by dimension. In our case the OLAP cubes will manipulate the data of social networks, which can be aggregated according to different dimensions (axes), by using aggregation functions: min, max, count, sum, avg. We can use specific functions. It should be noted that in the architecture that we proposed results from the analysis layer, including data mining can be stored, a second time, in the data warehouse as a measure or dimension, which is represented by a loop between the data warehouse layer and the analysis layer.

E. *The presentation layer*

The presentation layer is the different reporting tools that enable the different visualizations of the analysis layer as diagrams, maps, graphs, charts etc.

IV. Experiment

A. *Realization of a Datamart*

We chose to implement our approach on the social network "Twitter", although our approach is applicable to other social networks.

We made a datamart, which we can extend, and we took as subject (subject) "terrorism". With our datamart we can study other topics, so to detect other cyber threats by adding other measures and dimensions, if necessary.

To realize our datamart, we followed these steps:

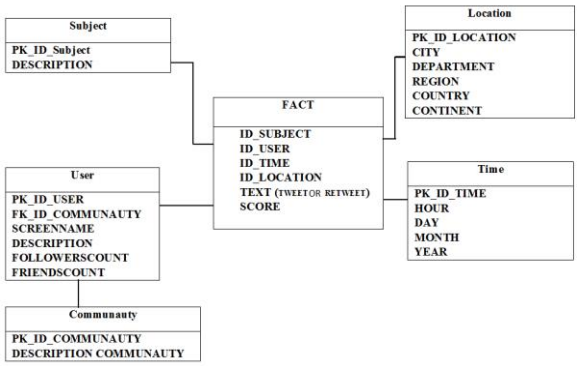
- Search, then extraction of tweets and metadata, using R language. For this, the data needed for our multidimensional analysis (extraction) were extracted using the bag-of-words (BOW) approach. For this extraction, we chose the most representative keywords of the "terrorism" theme (in French), for that we used two semi-automatic methods, one is based on a BOW representation of the text, and the other is based on a Word2Vec representation.
- Perform a job of cleaning, processing and structuring this data (transformation). Then, this transformed data was loaded into a data warehouse (fact table and dimensions), in the Postgres RDBMS (loading). In this process we have integrated text mining techniques to determine the score of a tweet (sentiment analysis).
- The sentiment analysis, in our case, was based on the dictionary FEEL (a French Expanded Emotion Lexicon) [1] which contains, in addition to terms related to the polarity (positive or negative), others concerning emotions such as joy, anger, surprise, sadness, etc. The choice of the FEEL dictionary is motivated by the fact that wealth of this dictionary

allows a better classification of our tweets. To classify our tweets, we considered the text content of these tweets belong to three classes: positive (if Score> 0), negative (if Score <0), neutral (if Score = 0).

- For the determination of the dimensions, we propose a datamart conceptual model contains five (05) dimensions and one hierarchy, which are:
  - Subject: the subject of the tweet.
  - User: the author of the tweet.
  - Community, which constitutes the hierarchy of a set of user having the same class.
  - Time: the time and date of creation of the tweet.
  - Location: The geographical location of the author of the tweet.
- The table of fact is the publication of a tweet.
- Concerning measurements, we have taken S (Score> 0): an integer that represents the positive sentiment of the user of tweet concerning the incitement to terrorism.

For this purpose, the conceptual model, which has been implemented in a physical model of data under a relational DBMS, is presented in "Fig.2," below.

FIGURE 2. Datamart for the analysis incitement to terrorism



B. *Analysis and presentation*

To determine the users inciting terrorism, the analysis in our approach is carried out as follows: It is known that one aspect of behavioral analysis is the tracking of the variation of a user's behavior over a period of time. Our idea is to follow the sentiment expressed by a user during a period of time. This will result in the calculation of the average score of a user's tweets that can classify this user by "inciting terrorism" or not. Since, if we take the sentiment for a single tweet of a user, it can't refer to a good ranking, since it can be a parody or joke. But if we calculate the average of the sentiments of this user for several tweets, according to a threshold that we will determine, tweets for a period of time in our case, we can be more sure of the orientation of this user.

Thus, we can propose the following modelization:

- Let $U = \{u_1; u_2; \ldots; u_n\}$: all users of Twitter.
- Let S_i be an integer that represents the sentiment expressed in a tweet tweet by a user u_j (S_i = nbrPositifsTerms - nbrNegativesTerms) .
- Let Avg be a real representative of the average of the scores of a set of tweets $\{tw_1; tw_2; \ldots; tw_n\}$ published by a given user u_j in a given period. we thus have Avg = average (S_1; S_2;…; S_n), n is the total number of tweets expressed by a user u_j. This average reassures us to classify more accurately users into two (02) classes, C1: "Incitement to terrorism" and C2: "No incitement to terrorism".

Our decision rule is:

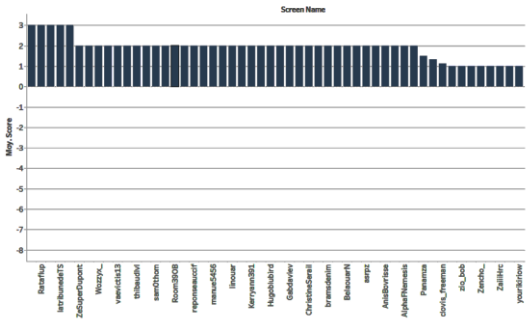
- if Avg > 0 then assign the user to the class C1.
- if Avg ≤ 0 then assign the user to the class C2.

For the representation we used the platform "Tableau Software". This software enables efficient analysis and publishing in a dynamic way, from a selection of data, while allowing to find the best possible representation.

To show the validity of our approach, here are the results of some queries, which we realized:

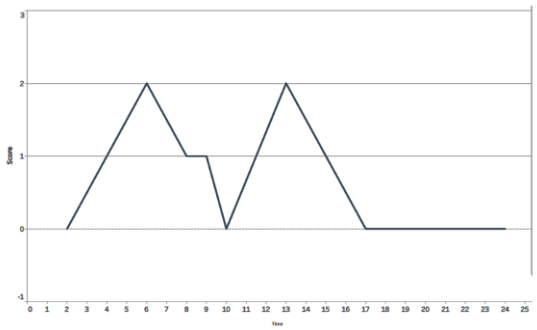
- Who are the users who incite to terrorism in the last two months? This question is equivalent to say who are the users that have a positive average score (Avg>0), over a period of the last two months? The response offred by our plateforme is represented by "Fig. 3,".

FIGURE 3. Users with a positive average score (inciting terrorism)



- What is the variation of the positive score of a user, over a period of time? "Fig. 4," shows the positive variation of the user's scores "Ratafloup" over a period of 25 days.

FIGURE 4. Variation of the positive score of user with the screen name "Ratafloup"

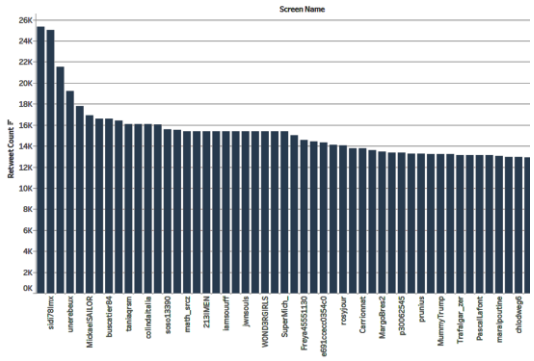


- What are the locations of users inciting to terrorism? "Fig. 5," presents the different locations of users inciting terrorism.

FIGURE 5. locations of users inciting terrorism

- What is the number of retweets for users inciting terrorism? Which is represented by "Fig. 6,".

FIGURE 6. Numbers of retweets of users inciting terrorism



It should be noted that the number of retweets, in Figure 6, is in thousands, because we took k = 100, so 4k means 4000 retweets, 6k means 6000 retweets.

C. Evaluation

The evaluation of our classification is presented in the confusion matrix "Tab.2,". As such, we took the tweets of 100 users who were automatically classified as inciting terrorism, using our approach, and then exposed these tweets to three specialists to classify these users into "inciting to terrorism" and "not inciting to terrorism ".

TABLE 2. Confusion matrix

		Classification by human expertise	
		Inciting	***Not inciting***
Automated classification (100 users)	Inciting	42	58
	Not inciting	26	74

So, out of these 100 users classified "inciting terrorism" by our approach, 42 users were judged "inciting terrorism" by these specialists. Similarly, of the 100 users "not inciting" by our approach 26 were judged "inciting terrorism," which has precision of 58%.

v. Conclusion and future work

In this work we used the techniques of data mining and data warehouse to analyze a related to terrorism in the social network twitter. With this in mind, with the help of Data mining techniques were able to determine the positive scores, that is to say the tweets inciting terrorism. Also, we realized a datamart and with the help of the determined multidimensional analysis the users inciting terrorism, their localization as well as their retweeters.

We have seen that our approach allows detecting and analyzing the users inciting terrorism and this, taking into account the history in our analysis.

For further work in this area, several avenues of research are possible:

- Enrich and improve our model, especially the ETL layer by exploring other tracks, such as those based on machine learning methods and continuous contexts "continuous-bag-of-words"[10]. From methods that can be used in this approach is Word2vec is defined in [14]. This exploration will allow us to improve our analysis process and stand out better results.
- Introduce the notion of graph cube [23].
- Experiment our cyber threat analysis model in a BIG DATA context using the MapReduce programming paradigm.
- Experimenting with our datamart on other political, security and social issues and enrich it with other measurements and other dimensions to detect and analyze other cyber-threats.

References

[1] A. Abdaoui, J. Azé, S. Bringay et P. Poncelet, *"Feel : a french expanded emotion lexicon,"* Language Resources and Evaluation, vol. 51, no. 3, pages 833–855, 2017.

[2] J. Acosta, Lamaute, N. Luo, M. Finkelstein, E. Cotoranu, "*Sentiment Analysis of Twitter Messages Using Word2Vec*",

Proceedings of Student-Faculty Research Day, CSIS, Pace University, May 5th, 2017

[3] M. Adedoyin-Olowe, M. M. Gaber, & F. Stahl, "*A survey of data mining techniques for social media analysis*", arXiv preprint arXiv:1312.4617, 2013.

[4] T. Chalothorn and E. Jeremy, "*Using SentiWordNet and Sentiment Analysis for Detecting Radical Content on Web Forums*", In: 6th Conference on Software, Knowledge, Information Management and Applications (SKIMA 2012), Chengdu University, 2012.

[5] S. Chaudhuri and Umeshwar Dayal, "*An overview of data warehousing and OLAP technology*", ACM Sigmod record 26.1: 65-74, 1997.

[6] L.Corr et J. Stagnitto, "*Agile data warehouse design: Collaborative dimensional modeling, from whiteboard to star schema*". DecisionOne Consulting, 2011.

[7] E. Ferrara, W. Q. Wang, O. Varol, A. Flammini, & A. Galstyan, "*Predicting online extremism, content adopters, and interaction reciprocity*", In International conference on social informatics (pp. 22-39). Springer, Cham,2016.

[8] A. Govand, "*Identifying Terrorist Affiliations through Social Network Analysis Using Data Mining Techniques*", Information technology master theses, 2016.

[9] A. Kaplan et M. Haenlein, "*Users of the world, unite! The challenges and opportunities of Social Media*", Business horizons, vol. 53, no. 1, pages 59–68, 2010.

[10] H. Kim, K. Hyunjoong Kim and C. Sungzoon, "*Bag-of-concepts: Comprehending document representation through clustering words in distributed representation*". Neurocomputing 266 : 336-352, 2017.

[11] L. Kirichenko, T. Radivilova et A. Carlsson, "*Detecting cyber threats through social network analysis: short survey*", Socio-economic challenges, Volume 1, Issue1, 2017.

[12] M. Kocharekar et U. Jadhav, "*Detecting Terrorist Activities using Sentiment AnalysisIn a Distributed System*", International Journal of Scientific Research Engineering Technology (IJSRET), vol. 6, no. 3, pages 185–187, 2017.

[13] Z. Madhoushi, A. R. Hamdan and S. Zainudin, "*Sentiment Analysis Techniques in Recent Works*", Science and Information Conference (SAI), 2015. IEEE, 2015.

[14] T. Mikolov, K. Chen, G. Corrado & J. Dean, "*Efficient estimation of word representations in vector space*". arXiv preprint arXiv:1301.3781, 2013.

[15] I. Moalla, A. Nabli, L. Bouzguenda & M. Hammami, "*Data warehouse design approaches from social media: review and comparison*", Social Network Analysis and Mining, 7(1), 5, 2017.

[16] L. A. Ngoge, "*Real–time sentiment analysis for detection of terrorist activities in Kenya*", Diss. Strathmore University, 2016.

[17] E. Omer, "*Using machine learning to identify jihadist messages on Twitter*", Examensarbete 30 hp, thesis 2015.

[18] A. N. Palo, "*Top 10 social networking threats*", Retrieved from http://www.networkworld.com/article/2213704/collaboration-social/top-10-social-networking-threats.html, 2017.

[19] A. Sharma, M. K Sharma & R. K. Dwivedi, "*Literature Review and Challenges of Data Mining Techniques for Social Network Analysis*", Advances in Computational Sciences and Technology, 10(5), 1337-1354, 2017.

[20] A. V. Sofea and I. Abdul Aziz, "*Terrorism Detection Based on Sentiment Analysis Using Machine Learning*", Journal of Enginneering and Applied Science 12 (3): 691-698, 2017.

[21] G. Vinodhini and RM. Chandrashekharan, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume2, Issue 6, June 2012.

[22] J. Wang, A. Wable et O. S. Cuervo, "Sharing digital content on a social network", Mars 21 2017. US Patent 9,602,605, 2017.

[23] P. Zhao, X. Li, D. Xin & J. Han, "*Graph cube: on warehousing and OLAP multidimensional networks*", In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (pp. 853-864). ACM, 2011.

[24] M. B. Kraiem, J. Feki, K. Khrouf, F. Ravat, O.Teste, "Modeling and OLAPing social media: the case of Twitter", J Social Netw Anal Inf Syst 5(1):47, 2015.

[25] A. Cuzzocrea, C. De Maio, G. Fenza, V. Loia & M. Parente, Towards "*OLAP analysis of multidimensional tweet streams*", In Proceedings of the ACM Eighteenth International Workshop on Data Warehousing and OLAP (pp. 69-73). ACM, 2015.

[26] I. Moalla & A. Nabli, "*Towards data mart building from social network for opinion analysis*", In International Conference on Intelligent Data Engineering and Automated Learning (pp. 295-302). Springer, Cham, 2014.

[27] A. Walha, F. Ghozzi, & F.Gargouri, "*ETL design toward social network opinion analysis*", In Computer and Information Science 2015 (pp. 235-249). Springer, Cham.