F. Miro-Llinares et al., Int. J. of Design & Nature and Ecodynamics. Vol. 11, No. 3 (2016) 406–415

CYBER HATE SPEECH ON TWITTER: ANALYZING DISRUPTIVE EVENTS FROM SOCIAL MEDIA TO BUILD A VIOLENT COMMUNICATION AND HATE SPEECH TAXONOMY

F. MIRO-LLINARES[1] & J.J. RODRIGUEZ-SALA[2]
[1]Crimina Centre, Miguel Hernandez University, Spain.
[2]Center of Operation Research, Miguel Hernandez University, Spain.

ABSTRACT

The attack against the Charlie Hebdo weekly in Paris, in the year 2015, was a disruptive event that generated an important public reaction in social networks, creating the opportunity to study the phenomenon of violent communication and hate messages on Twitter. In the days after the attack (between January 7 and January 12), a sample of more than 255,000 tweets with the hashtags #CharlieHebdo, #JeSuisCharlie and #StopIslam was collected. An analysis was made using qualitative and quantitative approaches to contrast the level of agreement between the different methods used. In the first place, messages were classified as tweets that contained violent and hate speech or general messages, following the inclusion criteria that based on experience and the scientific literature were defined by the Principal Investigator. Then, three pairs of judges classified the sample using the excluding criteria previously defined, according to which ten types of violent speech communication were identified, which were reduced to five essential categories. After the qualitative analysis, the methods of Data Mining were used with the purpose of extracting systems of rules for the classification of the type of speech, beginning with 18 variables derived from each tweet, including date, favorites or the type of software used for the tweet, among others. The results show that disruptive events are followed by communications that show spatial temporal and textual patterns clearly identifiable; this allows the authors to propose a methodology to classify in a very precise way, those messages that contain hate or violent speech.
Keywords: cyberhate speech, data mining, social media, violent talk.

1 INTRODUCTION

Social networks have become an important source of data for scientists that study human behavior. Beginning some time ago to the present, more and more research studies avail themselves of the huge amount of data available at websites like Twitter, Facebook, and Instagram [1] instead of the traditional surveys and interviews. This also has allowed a deeper understanding of the working dynamics of the networks themselves [2], through the analysis of very large numbers of the features of the messages [3], including the study of the mode in which users behave during an emergency, as an earthquake [4]. It is precisely events like natural disasters or attacks, like those against the French weekly Charlie Hebdo, in Paris between January 7 and January 12 of 2015, or the murder of Drummer Lee Rigby in Woolwich, United Kingdom, that have facilitated a set of studies whose interest centers in the identification of what has come to be known as the trigger event. In this sense, scientific literature has widely described the features that would permit the identification of the trigger events in environments as Twitter, for example. Thus, the analysis of the combination of temporal, environmental, and textual factors of the messages can provide information relevant for improving

WIT CONFERENCES
This paper is part of the Proceedings of the International Conference on Big Data (Big Data 2016)
www.witconferences.com

ISSN: 1755-7437 (paper format), ISSN: 1755-7445 (online), http://www.witpress.com/journals
DOI: 10.2495/DNE-V11-N3-406-415

situational awareness and predicting the behavior of the event in social networks thereby providing greater support to the decision making process [5].

Within this context, one of the events that has been given greater attention has been that of hate speech and violent communication on the Internet [6]; its analysis has centered particularly on the study of two dimensions: feeling [7] and tension [8]. The first permits the classification of the opinions and emotions in a text, using a scale, built on the basis of words within a text as predictive features, that measures the degree of positive or negative feeling [9]; the second identifies, by means of the Analysis and Membership Categorization Analysis (MCA), rules for the classification of the content of the messages [8]. Other approaches to the study of hate speech on social networks are based on the development of neural language models [10]. In any case, the development of these algorithms, even though they make it possible to classify the message, do not allow for an understanding of the different nature of violent communication, thus making it absolutely necessary to arrive at a deeper study of the categorization of the different expressions that are part of this phenomenon.

2 OBJECTIVES

The research that is now being presented seeks, in the first place, to categorize the different expressions of violent communication and hate speech in order to move on to identify the patterns that permit the establishment systems of rules using variables derived from the analysis of each tweet.

A secondary objective consists in the validation of the categories created using notions that are fundamental to the juridical and social sciences, using analysis tools, after observing and understanding a sample of the tweets tweeted during the attack against Charlie Hebdo.

3 HYPOTHESIS

In this manner, the hypothesis that serve as the foundation of the study are the following:

H1: There occur manifestations of violence and hate on the Internet that is possible to differentiate from the 'neutral' messages and can be identified through observation under an expert criterion.

H2: Messages of violence and hate may be distinguished from one another.

H3: The quantitative variables that accompany each message present environmental patterns that are related to other rules that have been determined through observational analysis under an expert criterion, and they allow to identify to distinguish messages of violence and hate from neutral messages.

H4: Quantitative analysis allows, by means of patterns that can be objectified, the identification of the different categories of violent communication and hate elaborated after a qualitative analysis.

4 METHOD

Through a deep understanding of the phenomenon, violent and hate speech is observed and analyzed in a sample of tweets that make up a set of data about 'social conversation' generated in Twitter Because of the attack to Charlie Hebdo. A subsample of messages of violence and hate was extracted from this set of data. Then, by observing the subsample, a taxonomy was elaborated that included each of the violent and hate forms that were observed (see Table 1). After categorization, each one of the tweets was read and screened by whether they spoke of violence and hate and then the prevalence of each of them in each category was determined. On the other hand, without considering the qualitative variables, such as the set of words used in the message and their implied meaning, other quantitative variables of each message in the sample were measured (followers_count, friends_count, listed_count, text_Length, among others), with tools prepared for that purpose, in order to find patterns that related closely with the human classification.

5 DATA COLLECTION

In order to gather a sample of messages that would have all types of violent and hate manifestations, we elected the terrorist attack against the offices of the satirical weekly Charlie Hebdo. For that purpose, three hashtags were chosen because they were Trending Topic about the event in Spain; this would ensure a great representation of the communication that occurred about the event. The hashtags chosen were #CharlieHebdo, #JeSuisCharlie and/or #StopIslam; they were the most mentioned during the days after the event beginning from the day it happened, January 7 until January 12. In this manner, a sample was obtained with a total of 282.397 tweets. This data set was composed of several variables that along with those added for the analysis totaled 26 in the database.

6 PROCEDURE

6.1 Qualitative analysis

In order to make an observation that discriminates between messages of violence and hate and neutral messages in the sample, three pairs of judges were chosen. The judges observed and evaluated each message with five criteria of alternative inclusion:

Criterion One. *Serious insults, degrading expressions, of unquestionable character, directed toward unspecified persons or certain, determinate or indeterminate, groups.*

Criterion Two. *Convey a positive approach to violence against people, determinate or indeterminate, either as defense, or as glorification, justification, trivialization, incitement, induction, understanding, joy, etc.*

Criterion Three. *The attribution to specific individuals of insulting expressions, public humiliation and serious vexations, or the imputation of criminal acts or serious offenses.*

Criterion Four. *Expressions of hate or contempt directed towards certain groups, especially those who have somehow been seen, or can be seen, as deprived of their rights, and suffer intolerance, particularly those expressions that use derogatory terms against them and they ask or justify the restriction of rights of such groups.*

Criterion Five. *Nasty expressions and bad taste regarding the event, which cause severe pain to some people, particularly those expressions that show hatred towards persons, or totally dehumanizes them, including jokes and black humor particularly serious and in relation to events that are not violent (natural or accidental death), and cause much pain to indirect victims.*

In order to ascertain a valid and reliable screening by the judges, four pilot tests were conducted randomly selecting a group of 200 tweets, and the judges then individually decided if it each was a violence and hate message or not. After each of the tests, the criteria were revised in order to improve validity in the application of the criteria to the messages. Through a Kappa Test [11] the index of concordance between the judges was determined; the last pilot test resulted in a high reliability index (Kappa = 0,91). Once the concordance among the judges was established, the first screening was conducted by the three pairs of judges; they evaluated three sets of tweets, that were proportional and were distributed randomly. Each pair of judges analyzed the same messages. After the analysis of each set of tweets the inter-rater reliability was tested to ensure the concordance by the two members of each pair. The concordance was high in the three pairs of judges (couple 1 = 0.97 *k*; couple 2 = 0.86 *k*; couple 3 = 0.93 *k*).

6.2 Quantitative analysis

To conduct the quantitative analysis, the set of classified tweets was used in an attempt to define the rules and patterns that allow to distinguish the features that differentiate a tweet that expresses hate

or violence from a tweet that uses a neutral or positive speech. It was desired to determine the features that define tweets with negative content as a function of the message they transmit (hate, violence, discrimination, etc.). The objective of this procedure was to build a model, based on the patterns, that allows for predicting when a tweet is potentially dangerous (for example: it expresses hate or incites violence), using data-mining techniques. Data mining is the fusion of statistical modeling, the storage of data bases, and artificial intelligence techniques [12]. There are many research projects that have used data mining to study problems related to delinquency: one such project is the study by Estivill-Castro and Lee [13] who used clustering techniques and association rules to detect spatial-temporal patterns in the registries of criminal events; other, more general studies, as the one by Chen *et al*., [14] show how to approach an investigation of different types of crimes with several data mining techniques. As currently, social networks are also used for criminal ends, the importance of analyzing the contents of a social network, likeTwitter, are demonstrated by investigations like that of Bendler *et al*. [15]. In this investigation, we have used techniques based on the generation of association rules. Given a data set D with N records or rows and A attributes or columns, in which each attribute may be assigned a certain value from a finite set of values, one rule r_i is one tuple of the pairs <attribute, value> with one antecedent and one consequent expressed as follows:

$$r_i = \{\text{antecedent}\} \rightarrow \{\text{consequent}\},$$

where both {antecedent} and {consequent} are a set of one of several pairs <attribute, value>. The probability that the antecedent of the said rule, that is the combination of values that it represents, can be found in the original set of data is called the 'support' of the rule. On the other hand, the probability that the consequent of the rule can be found in the subset of rows in which the antecedent is found is called the 'trustworthiness' of the rule. Formally,

$$\text{Support } (r_i\text{: ant}\rightarrow\text{con}) = N_{ant}/N \quad \text{Trustworthiness (ri: ant}\rightarrow\text{con)} = N_{ant\rightarrow con}/N_{ant},$$

where N_{ant} that represents the antecedent is found in the data set D, and $N_{ant\rightarrow con}$ represents the times that the entire combination of values of the rule, antecedent and consequent, is found in D. Association rules are characterized by being patterns that seek the combinations of probable values within a database, that is, with high support and trustworthiness values. In an association rule, both, the antecedent and the consequent will have a variable set of pairs <attribute, value>; for example, for one set of data with four attributes (A = 4) we could have association rules like these:

ar_1: $\{<a_1, v_{1,1}>, <a_2, v_{2,1}>, <a_3, v_{3,1}>\} \rightarrow \{<a_4, v_{4,1}>\}$

ar_2: $\{<a_1, v_{1,2}>, <a_2, v_{2,2}>\} \rightarrow \{<a_3, v_{3,2}>\}$

ar_3: $\{<a_4, v_{4,3}>\} \rightarrow \{<a_1, v_{1,3}>, <a_2, v_{2,3}>\}$

Algorithms for the search of association rules are frequently used in data mining, where the algorithm '*a priori*', presented in the study of Agarwal and Srikant [16], is the reference method to extract this type of rules from large volumes of data.

7 RESULTS

A subset consisting only of messages of violence and hate, very different from each other, was culled. It was made up of 2,304 original tweets culled from 282,397 tweets in the sample; this represents 0.8% of the messages of violence and hate in the total. This subsample was observed first, to discover the phenomenon as a whole. The taxonomy was created after an observational analysis that

Table 1: Hate speech & violent communication taxonomy. Own elaboration.

Regarding Causation To		Category
Physical violence	Physical damage	Violent incitement
No physical violence	Personal moral damage	Personal offence
		Discrimination incitement
	Colective moral damages	Collective offence

Table 2: Categories' prevalence (n = 2.304).

Category detail	Quantity	Percentage
Violence incitement	135	5,87%
Personal offence	114	4,96%
Discrimination incitement	984	42,61%
Collective offence	1,071	46,57%
	2.304	100

Table 3: Day speech prevalence (n = 2.304).

		Quantity			Percentage	
DayStretch	**Hours**	**Total**	**Neutral**	**Hate/Violence**	**Neutral (%)**	**Hate/ Violence (%)**
Dawning	00:00 → 08:00	16,899	16,818	81	6,00	3,52
Morning	08:00 → 13:00	60,068	59,634	434	21,29	18,87
Midday	13:00 → 16:00	70,259	69,630	629	24,86	**27,35**
Afternoon	16:00 → 20:00	69,271	68,581	690	24,48	**30,00**
Night	20:00 → 24:00	65,900	65,434	466	23,36	20,26

completed and improved the categorization as more and more observations were made on the sample of interest. The next step of this method consisted in reading and classifying each of the 2,304 tweets in the categories of the taxonomy, each of them being classified as a function of the interests that were at play in the message.

In this manner, the observation during the qualitative analysis established the prevalence of each category of the taxonomy in the sample of 2,304 messages of violence and hate. Big differences were observed between the two. In this sense, the referent group impacting collective sensibility (collective offense) is the group that presents the highest percentage, 46.57%, followed by the group that incites discrimination (discrimination incitement) (42.61%). It is followed by direct violence (violence incitement) with a much smaller percentage, 5,87%, and by affectation of honor or dignity (personal offence) (4,96%) (see Table 2).

On the other hand, with respect to the descriptive analyses that were conducted about the temporal values, it is observed that the percentages of violent communication and hate in relation to the total sample are greater around noon and in the afternoon, with a difference between 3% during the noon and a 6% in the afternoon. This means that the expressions of violence and hate are broadcast with a greater frequency between noon and the afternoon, especially the latter. Nevertheless, the samples of hate speech and violent talk in relation to neutral messages that occur at dawn, in the morning, as well as at night, are smaller in al cases (see Table 3).

If we do a temporal analysis of violence and hate communication from a different perspective, in this case from the moment the trigger events occur, the case is divided into three moments: the moment of the first terrorist attack, that is, the attack on Charlie Hebdo (Terrorist 1), the shootout of Coulibaly and the police (Terrorist_2) and the police assault that ended the lives of the three terrorists (Police). From this perspective, it is seen that the greater flow of messages, neutral as well as of violence and hate, is at the moment of the attack on Charlie Hebdo. Also, during this first moment, the messages of violence and hate represented the greater percentage in relation to the total. That is to say that the first moment of the three was the one that generated more hate (58.3%) eventually becoming proportionally greater than the total of neutral messages that were sent (49.6%). The second moment (Terrorist_2), represents, in a very egalitarian way, the phenomenon of violence and hate discourse and the prevalence of the last is somewhat greater. The moment that ended the event, the police assault is the event that caused less social conversation on the Internet and and also lower manifestations of violence and hate (see Table 4).

In relation to the number of followers, the variable was discretized in order to make a more efficient analysis. In this manner, those tweeters that had less than 100 followers were considered as 'noob twitter', and 'middle twitter' were those that had between 100 and 1,000 followers. 'Advance twitter was assigned to those users that had between 1,000 and 10,000 followers, while the 'referent twitter' was assigned to those that had more than 10,000 followers. What can be inferred from this

Table 4: Event stretch prevalence (n = 2.304).

	Quantity			Percentage	
EventStretch	**Total**	**Neutral**	**Hate/Violence**	**Neutral (%)**	**Hate/Violence (%)**
Terrorist_1	140,287	138,947	1,340	49,61	**58,26**
Terrorist_2	81,731	81,122	609	28,96	26,48
Police	60,379	60,028	351	21,43	15,26

Table 5: Followers prevalence (n = 2.304).

	Quantity			Percentage	
FollowersD	**Total**	**Neutral**	**Hate/Violence**	**Neutral**	**Hate/Violence**
NoobTwitter	42,296	41,934	362	14.97	15.74
MiddleTwitter	164,296	163,030	1,266	58.20	55.04
AdvancedTwitter	65,308	64,701	607	23.10	26.39
ReferentTwitter	10,497	10,432	65	3.72	2.83

analysis is that, on the one hand the noob twitter has a parallel influence on neutral messages in contrast to the messages of the violence and hate, even though we can see a small increase of 0.95% in the latter. On the other hand, the middle twitter is the more representative of the entire conversation, neutral or of violence and hate (58.20% in the neutral conversation and 55% in the violence and hate). In the case of the advanced twitter, even though a smaller prevalence than the previous case is represented, a significant upward difference is observed between the messages of hate and violence and the neutral (26.4% messages of violence and hate vis-à-vis 23.1% of neutral). Lastly, the smaller prevalence is found in the group of referent twitters, both, in the neutral conversation (3,72%) and in the conversation with violence and hate (2.83%) (see Table 5).

Once the tweets were collected, we extracted those whose discourse could be qualified as negative (hate, violence, etc.), in such a way that we now had two sets of data, a global set with all of the tweets that we called 'TT' (Total Tweets) and another set that we called 'NT' (Negative Tweets), the second being a subset of the first. Some of the patterns found that a certain difference was reflected between the sets TT and NT as follows:

Table 6: Patterns found in TT (n = 282.397).

Support	**Confidence**	**Even Stretch**	**Hashtag**		**Agent**
36.35%	37.28%	Terrorist_1	HT_N_N_Y	⇒	**Android**

Tabla 7. Patterns found in NT (n = 2.304)

Support	**Confidence**	**Even Stretch**	**Hashtag**		**Agent**
36.17%	40.75%	Terrorist_1	HT_N_N_Y	⇒	**WebClient**

Considering the fact that the three hashtags of the sample were subsumed under a single variable (Hashtags) that determined which of the three was found in the message, the first was #StopIslam, #JeSuisCharlie was the second and #CharlieHebdo the third, (e.g. "HT_Y_Y_Y" would mean that it has the three hashtags), and the interpretation of these patterns would be the following: after the first attack, the majority of the tweets that included exclusively the hashtag #CharlieHebdo in their content were sent from an Android device (37,28%). Nevertheless, in the set NT, the device most used was the web client (40,75%) (see Tables 6 and 7). The following example is also interesting:

Table 8: Patterns found in TT (n = 282.397).

Support	**Confidence**	**#StopIslam**	**#CharlieHebdo**	**Re-tweet**		**#JeSuisCharlie**
21.38%	77.71%	SI_N	CH_Y	RT_Y	⇒	**JSC_N**

Table 9: Patterns found in NT (n = 2.304).

Support	**Confidence**	**#StopIslam**	**#CharlieHebdo**	**Re-tweet**		**#JeSuisCharlie**
2.65%	81.97%	SI_N	CH_Y	RT_Y	⇒	**JSC_Y**

Table 10: Patterns found through the hashtags (n = 2.304).

#	Support (%)	Confidence (%)	Antecedent		Consequent
StopIslam	34,39	75,35	SI_Y	⇒	**Discrimination incitement**
JeSuisCharlie	18,48	48,71	JSC_Y	⇒	**Colective offense**
	18,48	34,59	JSC_Y	⇒	**Discrimination incitement**
CharlieHebdo	56,39	63,07	CH_Y	⇒	**Colective offense**

This means that the majority of the tweets that did not use the hashtag #StopIslam, did use #CharlieHebdo and were re-tweeted and did not use the hashtag #JeSuisCharlie (77.71%). In contrast, among the negative tweets with the same antecedent, the majority used this hashtag (81.97%) (see Tables 8 and 9).

On the other hand, when studying those classified as 'NT', that included the messages of hate and violence, etc. (N = 2,304), significant patterns were observed that were able to provide support with a high degree of trustworthiness. In this manner, when choosing each hashtag by itself and crossing its data with the established categories, it was observed that the hashtag #StopIslam occurs with a probability of 75.3% of the cases and has a support of 34.39%, inciting discrimination.

The hashtag #JeSuisCharlie, meanwhile, occurs with a probability of 48.7% of the cases vis-a-vis the rest of the categories, while collective offence was followed closely by incitation to discriminate with a probability of 34.5%. Both categories have a support of 18.48%. Finally, the hashtag #CharlieHebdo, the most used during the event, occurs with a probability of 63.1 of the cases, and the category of collective offence with a support of 56.39% (see Table 10).

8 DISCUSSION AND CONCLUSIONS

Despite the fact that the patterns show that there are indications of differences between the variables of the neutral messages that were generated in the social conversation about the attack to Charlie Hebdo and the messages of violence and hate, there are no patterns that determine significant differences in these expressions of hate vis-a-vis the rest of the social conversation. In this sense, the small quantity of information provided by the low percentage of messages of violence and hate, in relation to the total sample, makes it difficult to complete the task of finding associations that can be differentiated. Nevertheless, the fact that some of the combinations found through the quantitative analysis do not differentiate between messages of violence and hate from the rest of the neutral conversation may mean an improvement with respect to the understanding of the data that provide no relevant information for the identification of these messages. Therefore, this study and the results shown in it can help to identify objectifiable patterns using the messages on the Web.

With reference to the generation of patterns for the validation of the elaborated taxonomy, the results show predictive patterns that can facilitate the task of classifying messages of violence and hate. In relation to this, the analysis shows that the variable that provides greater predictability about the type of message of violence and hate is the tag itself with which the user hashtags the message. This means that the variable that best predicts the type of violence and hate message is the hashtag used in the tweet, for, as it is indicated by the data, users that manifest discriminant hate will use, with a great probability the tag #StopIslam, that has on face value a heavy discrimi-

natory load, while the users that use the tag of the event, #CharlieHebdo, or the supportive #JeSuisCharlie, will express a violent discourse based on gross language to manifest intense anger after a terrorist attack.

ACKNOWLEDGEMENT

Research conducted under the Ciber Hache Project "Incitement to violence and hate speech on the Internet. Real reach of the phenomenon, typologies, environmental factors and limits to the legal intervention against it" funded by the Ministry of Economy and Competitiveness (DER2014-53449-R)

REFERENCES

[1] Burnap, P. & Williams, L., Cyber hate speech on twitter: an application of machine classification and statistical modeling for policy and decision making. *Policy and Internet,* **7**, pp. 223–242, 2015.
http://dx.doi.org/10.1002/poi3.85

[2] Sloan, L. & Morgan, J., Who tweets with their location? understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS One*, **10**(11), e0142209, 2015
http://dx.doi.org/10.1371/journal.pone.0142209

[3] Paltoglou, G., Sentiment-based event detection in Twitter. *Journal of the Association for Information Science and Technology,* 2015.
http://dx.doi.org/10.1002/asi.23465

[4] Zielinski, A., Bügel, U., Middleton, L., Middleton, S.E., Tokarchuk, L., Watson, K. & Chaves, F., Multilingual analysis of Twitter news in support of mass emergency events. In *EGU General Assembly Conference Abstracts,* 14, p. 8085, 2012.

[5] Alsaedi, N., Burnap, P. & Rana, O., Identifying disruptive events from social media to enhance situational awareness. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* pp. 934–941, 2015.
http://dx.doi.org/10.1145/2808797.2808879

[6] Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V. & Bhamidipati, N., Hate speech detection with comment embeddings. In *Proceedings of the 24th International Conference on World Wide Web Companion*, International World Wide Web Conferences Steering Committee, pp. 29–30, 2015.
http://dx.doi.org/10.1145/2740908.2742760

[7] Magdy, W., Darwish, K. & Abokhodair, N., Quantifying Public Response towards Islam on Twitter after Paris Attacks. *arXiv preprint arXiv:1512.04570*, 2015.

[8] Williams, M.L., Edwards, A., Housley, W., Burnap, P., Rana, O., Avis, N., Morgan, J. & Sloan, L., Policing cyber-neighbourhoods: tension monitoring and social media networks. *Policing and Society*, **23**(4), pp. 461–481. 2013.
http://dx.doi.org/10.1080/10439463.2013.780225

[9] Pang, B. & Lee, L., Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, **2**(1–2), pp. 1–135, 2008.
http://dx.doi.org/10.1561/1500000011

[10] Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V. & Bhamidipati, N., Hate speech detection with comment embeddings. In *Proceedings of the 24th International Conference on World Wide Web Companion*, International World Wide Web Conferences Steering Committee, pp. 29–30, 2015.

http://dx.doi.org/10.1145/2740908.2742760
[11] Viera, A.J. & Garrett, J.M., Understanding interobserver agreement: the kappa statistic. *Family Medicine*, **37**(5), pp. 360–363, 2005.
[12] Mena, J., *Investigative Data Mining for Security and Criminal Detection*, Butterworth-Heinemann, Elsevier Science (USA), p. 452, 2003, ISBN 0-7506-7613-2.
[13] Estivill-Castro, V. & Lee, I., Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. *Proceeding Sixth International Conference Geocomputation, Brisbane (Australia)*, 2001.
[14] Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y. & Chau, M., Crime data mining: a general framework and some examples. *Computer*, **37**(4), pp. 50–56, 2004. http://dx.doi.org/10.1109/MC.2004.1297301
[15] Bendler, J., Tobias Brandt, T., Wagner, S. & Neumann, D., Investigating crime to Twitter relationships in urban environments - facilitating a virtual neighborhood watch. *Proceedings of 22th European Conference on Information Systems*, Tel Aviv (Israel), 2014.
[16] Agrawal, R. & Srikant, R., Fast algorithms for mining association rules. *Proceedings International Conference Very Large Data Bases (VLDB'94)*, Santiago (Chile), 1994.