ISSN 0354-9852 (print)
eISSN 2566-3496 (online)
Originalni naučni rad
Original scientific paper

UDK 005.511:[004.78:316.77
DOI 10.7251/AKT1737009S
COBISS.RS-ID 6970392

RESEARCH OF THE POTENTIALS OF TWITTER SOCIAL NETWORK IN CRISIS SITUATIONS

Slaviša Sovilj[1], Vojkan Vasković[2], Jelena Bačević[3], Nenad Orlić[4]

ABSTRACT

This paper describes an approach for harnessing Twitter social network in crisis situations. Main goal was to investigate possibilities of using Twitter as a platform for collecting and dissemination of information during the natural and environmental disasters. Research context is focused on the period of floods in the Western Balkans during May 2014, particularly Bosnia and Herzegovina. A method for collecting, storing, analyzing and presenting data, i.e. tweets and retweets from Twitter, is described. Further, an application for gathering and visualization of data from Twitter API was developed. Research sample included 45480 Twitter messages from users in the region of Bosnia and Herzegovina, Serbia and Croatia. Data were investigated using Social network analysis techniques with respect to location, content and user characteristics. Study results pointed out that Twitter is a valuable source of important information in crisis situations, particularly as a provider of prompt information from the endangered territories.

Keywords: *social networks, microblogging, Twitter, floods, West Balkan.*

JEL classification: *L86*

AIMS AND BACKGROUND

The research area of this study is traffic generated in the microblogging system Twitter during floods that occurred from 16. to 28. May 2014. within territories of Serbia, Croatia, Bosnia and Herzegovina. Twitter as a social media tool can play a significant role in crisis situations both by sharing

1 Ministry of Internal Affairs of the Republic of Srpska, Banja Luka, Bosnia and Herzegovina
2 Faculty of Business Economy and Entrepreneurship, Belgrade, Republic of Serbia
3 GB Solution, Belgrade, Republic of Serbia
4 MadWeb d.o.o, Pančevo, Republic of Serbia

information (Chatfield & Brajawidagda, 2012), (Aljohani et al., 2011) and collecting information valuable for responsible institutions (Terpstra et .al., 2012). Twitter is a great platform for sharing real-time information during emergencies situations because of the possibility for easy sharing short and exact content with linking external resources (Bruns & Burgess, 2012), (Huges &Palen, 2009). Twitter geolocation feature has significant importance in monitoring and analysis of communication in the scope of this social network (Li &Goodchild, 2010). This study investigates Twitter as a source of information valuable for future disaster management system. The initial assumption is that the quality and relevance of information on the situation in the field is significant if it comes from users who are in the endangered areas.

Users who are in the region hit by floods and who are in direct contact with the disaster can provide information about the real situation in the field. In this way, the relevant institutions can respond to emergency situation such as natural and environmental disasters depending on the speed of gathering information from vulnerable locations (Latonero &Shiklovski, 2011).

Main aim of this study is to determine the extent to which information from Twitter are qualitative and reliable. On that basis, the following hypotheses can be defined:

H1. Twitter microblogging platform is used in spreading information of local character.
H2. Twitter microblogging platform is used in high risk areas.
H3. Users who are located in highly affected areas have central/important role in spread of information through social networks.

Additional way to evaluate the relevance of information is monitoring the traffic according to the scope of occurrence of keywords (Stollberg & De Groeve, 2012). According to this, two additional hypotheses were defined:

H4. Leading nodes in the network are common people.
H5. Frequency of occurrence of keywords in the tweets is associated with activities in the field.

Every natural disaster is characterized by two phases: 1) stroke and emergency and 2) repair after disasters. In this study, for these two phases, the sets of keywords are designed. According to these sets, a pattern is formed in order to perform comparisons and determine the extent in which Twitter is used in various stages of disasters by keywords in the published messages. Phase of stroke is related to saving lives, evacuation of people and goods while the phase of rehabilitation is linked to damage repair and environmen-

tal aspects in terms of cleaning, pest control and prevention of infection in the flooded area.

In order to determine key users and their roles in dissemination of information on Twitter, current communication will be observed through mathematical form-graphs where the users are presented as nodes and their interactions as branches. In this environment it is possible to apply the methodology of Social Network Analysis - SNA, where the specific algorithms are used for determining node significance in the already formed social network (Wasserman & Faust, 2012).

SNA provides a visual and mathematical analysis of human relationships groups, organizations, computers, web pages and other information/knowledge processing entities. Furthermore, it enables management of the collected information and knowledge from social networks' users gathered around the same interest (Ehrlich & Carboni, 2005), (Haythornthwaite, 1996). An application of social network analysis related to floods in Australia in the period of 2010-2011 can be found in literature (Cheong & Cheong, 2011).

EXPERIMENTAL

Using the Twitter API - online collection of tweets

For this study, a sample was created by using the Twitter API, which allows for collection of various types of data from Twitter as well as messages-tweets, user information, trends, etc.(Satrbird & Palen, 2010). Methodology of collecting, processing, analyzing and visualization of data from Twitter is shown in Figure 1.

In this case, tweets collecting is conducted according to following keywords: #poplava (flood), #sanacija (sanation), #ciscenje (cleaning), #pomoc (help), #deratizacija (deration), #opasnost (emergency), #evakuacija (evacuation), #steta (damage).

PHP application was developed with the aim to send a specially formatted HTTP GET requests to the Twitter servers, and to receive JSON object with a tweet or retweet that contained the keyword as response (step 1). This JSON object, contains information about the message and metadata about users who posted a tweet or retweet. PHP application parses content of the JSON object (step 2), extracts data related to users and the message's content and finally stores data into a database (step 3).

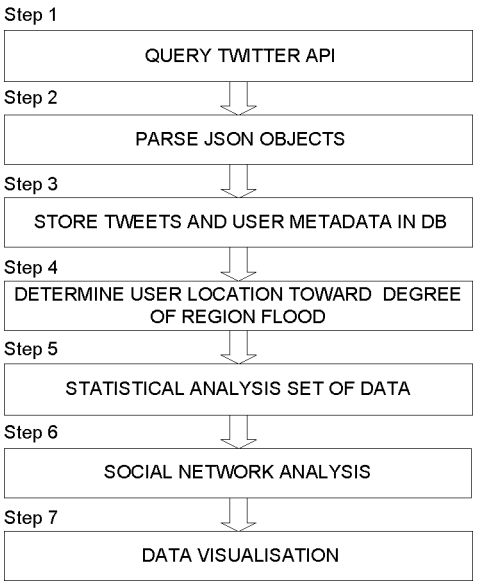

Figure. 1. Method of collecting, processing, analyzing and visualization of data

Offline preparation, analysis and visualization of collected data

After collecting data, the next step is determining users' location, the state they belong, and the degree of flood endangerment within a territory (step 4). Information about the user's location is located on their profile or in the form of geolocation and coordinates that have been provided by a mobile device. The states and degree of flood endangerment are determined by the list created for all the cities in Serbia, Bosnia, Croatia, and Montenegro. Further, a coefficient of endangerment is assigned to each city, on a scale of: high risk (HR), low risk (LR), and without risk. Furthermore, an application was made with the aim to read the information about location of the user, check if the location information is on the list of municipalities. If a user belongs to a particular municipality and state, the application stores information in the database.

In addition to location information, information about the state and flooded area for each user and their activities on Twitter are calculated and stored in the database (step 5). Metadata about users include: date of registration of a profile, number of published tweets, followers, followed, favorites, etc. Statistics for a user are related to: popularity of the user determined by the number of posted retweets, average number of posts in the analyzed period, average growth of followers for analyzed period, and other statistics of interest for the study which will be described below.

In the scope of network analysis, communication pairs were formed, where user A posts original message-tweet and user B gives retweet on this post (step 6). These communication pairs are imported in the form of graph G=(V,E) where V is a set of all nodes while E represents branches, i.e. relations. Since the direction of communication in the sample is known, this graph represents oriented, direct graph. In order to identify influential nodes in the social network, general methods of social network analysis are applied (Sovilj &Vaskovic, 2013):

a) density of graph that represents ratio of number of existing branches and the maximum number of branches and direct graph can be expressed as $|E| / (|V| * (|V| - 1))$,
b) In degree - number of input massages for every node is a number of retweets posted on original massage - user's tweet.
c) Out degree - number of output massages is a number of posted retweets or tweets.
d) Betweenes centrality - denotes the number of shortest paths from all nodes to all other nodes that pass through node analyzed. These are bridges in the network, users who connect different clusters. It is based on a calculation which demonstrates the role of a particular node in terms of connecting groups and formed clusters within the network.
e) Clossenees centrality - closeness centrality of node describes how the observed node near other nodes in the network. The highest value will be the node from which it is possible to access other nodes in the network passing through the smallest number of nodes.

Practical realization of the social network analysis was performed by using the Java libraries - JUNG framework, which is a set of software algorithms to calculate the metric graph. For the visualization and presentation of established social networks the application add-in for MS Excel-NodeXL was used (step 7). A significant problem was related to importing the data from a database in NodeXL, which detects the data in GraphML standard, accordingly, a web service that transfer data from database in format of GraphML was written.

RESULTS AND DISCUSSION

Analyses of users' locations

After data cleaning, the sample contained 45480 messages, i.e. tweets. Approximately 76% of messages (34804) contained information about user's location. Distribution of the tweets by countries is provided in table 1, whe-

re global character of the communications via Twitter platform can be noticed. Column "in" provides number of all retweets that are answers on tweets made by users from each country. Column "from" provides number of retweets made exclusively by users from each country. Column "original" presents a percentage of original tweets made by users from each country, while column "retweet" provides percentage of retweets on tweets from each country.

Table 1. Distribution of the tweets by countries

Country		Total	On Country	in	from	original	retweet
1.	Bosnia and Herzegovina	13397	4868	5797	2732	13.10%	86.90%
2.	Croatia	7390	2555	2940	1895	16.10%	83.90%
3.	Serbia	35294	18484	2709	14101	15.99%	84.01%
4.	Montenegro	994	180	466	348	15.29%	84.71%
Total:		57075	26087	11912	19076		

For instance, it can be noticed from table 1 that significant portion of communication was done with users that are outside Bosnia and Herzegovina (BIH). Accordingly, an analysis was performed in order to find out with which countries users from BIH communicated and the scope of the communication as well.

Table 2. Directions of the Twitter communication from BIH

Direction	**counter**	**Direction**	**counter**
From BIH to Croatia	10.58%	In BIH from Montenegro	8.24%
From BIH to Montenegro	5.32%	In BIH from Croatia	17.35%
From BIH to Other locations	26.42%	In BIH from Other locations	29.58%
From BIH to Serbia	57.68%	In BIH from Serbia	44.83%

Term "Other locations" indicates the communications with the users for whom information about location was not available or their locations were not included with this research. It can be noticed that most of the communication from Bosnia was performed with neighboring Serbia. According to the fact that more than 40% of total communications were done among users within one country, the hypothesis H1 (Twitter is used for dissemination of information among users that are near to each other) was confirmed. The communications are mostly local.

Figure 2 confirms assertion that micro blogging is used in territories that are at high risk of floods. More than 53% of total messages on Twitter were from those territories.

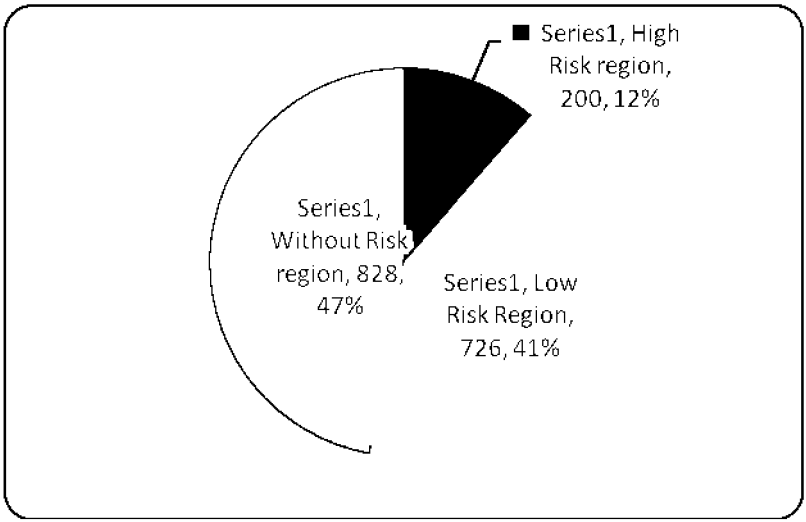

Figure. 2. Tweets distribution by level of risk of floods

Next issue was to find out in which extent Twitter is used for information distribution with respect to level of endergement of the teritory. In this context, distribution means that an user A have put an information, i.e. tweet and an user B made retweet in order to inform their followers and disseminate the information via network. A matrix of tweets was created and shown in table 3. Users that are from the regions with same level of endergment of floods were paired. It can be noticed that the least number of communication s were performed among users from the floodplains, whereas the biggest number of the communications was among users from teritories that were out of floods.

Table 3. Matrix of the communication with respect to level of endangerment of the territory

		User A: tweeter		
		Region:HighRisk	**Region:LowRisk**	**Region:WithoutRisk**
User B: retweeter	**Region:HighRisk**	2.56%	8.99%	8.76%
	Region:LowRisk	3.86%	19.39%	11.52%
	Region:WithoutRisk	3.11%	13.08%	28.74%

Analysis of locations showed that Twitter is used both locally and globally in dissemination of information related to natural and environmental disasters. Local dissemination implies that users who made original messages and retweets were from the teritory of the same country. In BIH this ratio is 37% for local and 63% for global. Share of users from highly endan-

gered teritories in local communication is 12%. According to the above results Twitter can be used as resource of information in emergency situation. On the other hand, it should be noticed that the number of tweets from highly endangered teritories is significantly lower in comparison to safe teritories.

Analysis of users' positions in social network

Analysis of users that are important for information dissemination and sharing was performed by social network analisys methods. Main task was to identify liders in social network and according to that find out if "important nodes" in social networks are located in highly endengered teritories (H3.). JUNG java library that enabled computation of values, input degree, output degree, betweenees and clossenees centrality for all nodes in the network (Yan et al., 2013). Top ten ranked users were presented in table 4. Ranking according to in_degree parameters enable to find out the most popular users in network, while out_degree parameter shows the most active users within the network, i.e. users that have posted the biggest number of mesagges-tweets. Betweenees centrality criteria enable identification of the users whose tweets connect formed group of users. Clossenees centrality aims to identify users that are in the highest extent connected with other users in network. Following types of users were found: common user - user, government institutions - e-gov, nongovernmental organizations - NGO and media - media.

Table 4 shows ranking of users, their location and types. It is obvious that users from highly endangered territories are placed only in clossenes centrality ranking, while in terms of popularity, activity and connectivity of users from those territories are not present in the list. Accordingly, it can be noticed that users form highly endangered areas don't have important roles in creating and dissemination of data from social network. Thus, H3 is refused. Majority of users are in group of common users. Further, government users have significant roles via their official profiles. Accordingly, H4 that claims that leader nodes are common users, is confirmed.

Table 4. Review

In degree				**Out degree**			
username		region	type	username		region	type
	Edis Jxxxxxxx	WR	user		anel dzxxxxxxx	WR	user
	Republika Srpska	LR	e-gov		tacixxxxxxx	WR	user
	Mahpeyxxxxxxx	WR	user		Miso Dxxxxxxx	LR	user
	Al Jazeera Balkans	WR	media		Ladyxxxxxxx	WR	user
	Arduana Kxxxxxxx	WR	user		prayforBosniaSxxxxxxx	WR	user
	Vanja xxxxxxx	LR	media		Ljiljana Txxxxxxx	LR	user

	Radio Sarajevo	WR	media		Edis Jxxxxxxx	WR	user
	sinisavxxxxxxx	LR	user		Vedad Hxxxxxxx	WR	user
	Miso Dxxxxxxx	LR	user		velmanuhxxxxxxx	WR	user
	DarkoTxxxxxxx	LR	user		090 290053 BHT	LR	media
Betweenees centrality				**Clossenees centrality**			
username		region	type	username		region	type
	Edis Jxxxxxxx	WR	user		Jasminxxxxxxx	WR	user
	Republika Srpska	LR	e-gov		Mjesecevaxxxxxxx	WR	user
	Miso Dxxxxxxx	LR	user		Dzanaxxxxxxx	WR	user
	Mahpeyxxxxxxx	WR	user		SMS NA 1003	WR	user
	Vanja xxxxxxx	LR	user		Nikola Nxxxxxxx	LR	user
	090 290053 BHT	WR	media		Darjanas	LR	user
	Arduana Kxxxxxxx	WR	user		Izeta Fxxxxxxx	WR	user
	Radio Sarajevo	WR	media		Marexxxxxxx	HR	user
	Al Jazeera Balkans	WR	media		Stalkerxxxxxxx	HR	user
	090 290053 BHT	WR	media		CristinaKxxxxxxx	LR	user

In order to present disposition of network nodes in their role we have performed visualization of social network data with respect to level of endangerment of territories (figure 3).

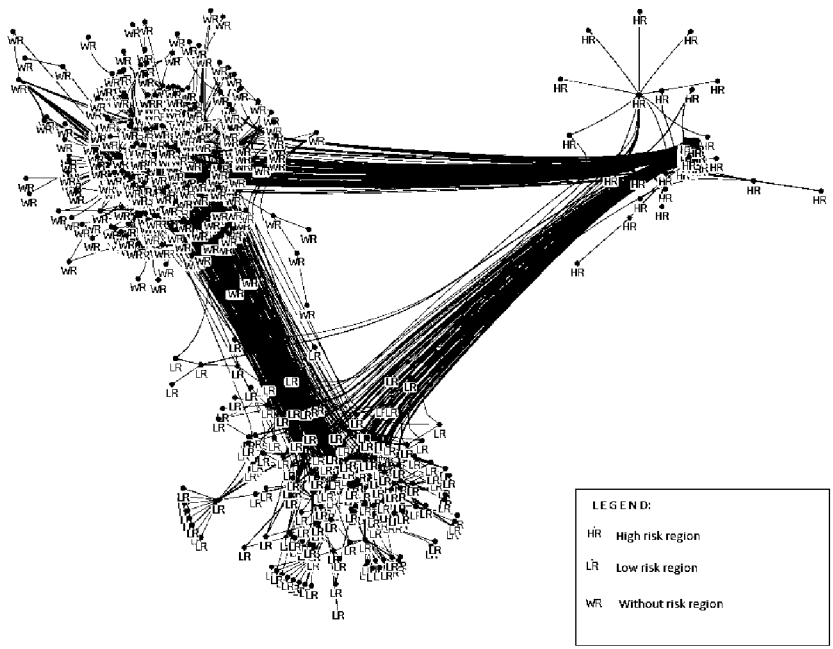

Figure. 3. Social network of users according to level of endangerment of territories

Time and location frequency analysis of key words

Analysis of keywords distribution with respect to location of users and time of messages in order to analysis of Twitter usage in dissemination of information in different phases of the catastrophe. Importance of keyword analysis within Twitter messages is explained in literature (Abilhoa & Castro, 2014). This analysis aims to test H5. Key words are classified in two groups. First group titled Impact contains keywords related to "phase of stroke", while second group titled Repair include keywords related to sanitation activities: deration, cleaning, prevent diseases, etc. First group includes following words: #poplava (flood), #hitno (emergency), #evakuacija (evacuation), #pomoc (help), while the second group contain: #sanacija (repair), #ciscenje (cleaning), #deratizacija (deration), #steta (damage).

Figure 4 presents comparation of two groups of keywords by teritories and degree of endangerment. It is evident that tweets with keywords from the Impact group are tweeted from highly endangered teritories. At the same time, keywords from second group are posted from safe teritories.

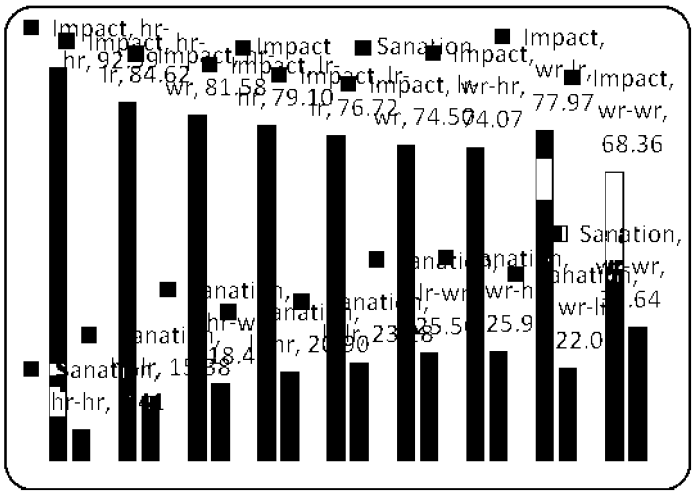

Figure. 4. Keywords in flooded territories

Figure 5 shows comparative analysis of keywords during particular period of time, where appearance and frequency of keywords is connected with situation in the field. In the begining of the floods dominant keywords were related to floods, evacuation of population and saving assets, while after stabilization topics related to echological aspects became dominant.

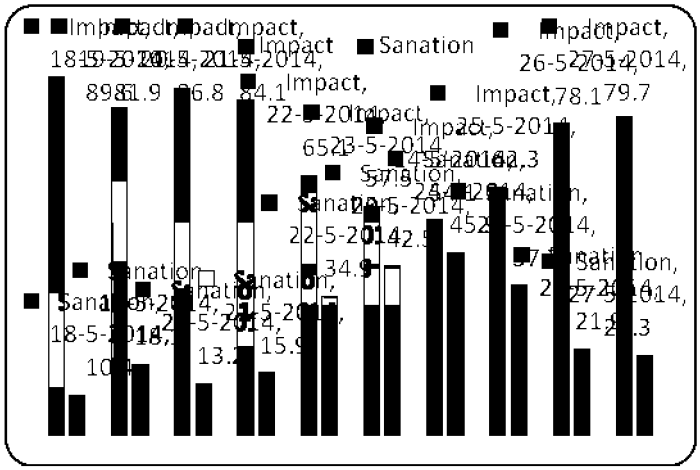

Figure. 5. View of represented keywords in time

Based on analisys of keywords it can be concluded that traffic on Twitter is in accordance with situation in the field, i.e. frequency of keyword appearance in tweets is connected with activities in the field (H5).

CONCLUSIONS

This paper presents an investigation of using Twitter as platform for collecting and dissemination of information during the floods in West Balkan during May 2014, particularly in Bosnia and Herzegovina. The initial assumption of this research was that the data that coming from users with highly endangered areas have significant informational value for the management of crisis situation. This conclusions are in accordance with results of researches where government used Twitter to early warnings, rescue and recovery (Chatfield et al., 2013).

Further, the research investigated in which extent Twitter was used in territories that were directly endangered by floods. Location analysis showed that Twitter is used in those territories. Social network analysis pointed out that users from territories that are at high risk of floods do not have important role in network considering criteria, such as: activities, popularity, connectivity and centrality. Keyword frequency analysis showed that prompt information from the field can be provided, particularly early warnings and following situation in the field during natural disaster.

Based on the conclusions it can be concluded that Twitter can be used as source for data about endangered regions, as there are people who use Twitter in these situations. This was proved in similar researches (Yin et al., 2012), (Gao et al., 2011). On the other hand, Twitter should not be used as

single and only reliable source of information from the field, but in combination with other information channels (Morales et al., 2014). This is supported by the fact that people from highly endangered areas do not have significant role in the social network and their tweets are not found as significant for users from other territories.

This research has a few constraints. Firstly, data used in analyses are collected via Twitter API. Thus, there is no guarantee that Twitter provided all content generated in the particular period of time. Further, locations of the users were determined in two ways: geocoordinates that were generated via mobile device or by location a user provided within their profile during registration process. At the particular moment, an user can be at location that differs from the one that was entered in the profile.

Future researches are directed toward developing system for automated collection and sentiment analysis of data related to natural and environmental disasters. Further, this implies developing a language ontology for analyzing Twitter posts (Konopoulos et al., 2013).

REFERENCES

1. Chatfield, A. T., & Brajawidagda, U. (2012). Twitter Tsunami early warning network: A social network analysis of twitter information flows. *In 23rd Australasian conference on information systems Geelong*, Australia.
2. Aljohani, N. R., Alahmari, S. A., & Aseere, A. M. (2011). An organized collaborative work using Twitter in flood disaster, *Proceedings of the ACM WebSci'11*, Koblenz, Germany, 1-2.
3. Terpstra, T., De Vries, A., Stronkman, R., & Paradies, G. L. (2012). Towards a realtime Twitter analysis during crises for operational crisis management, *Proceedings of the 9th International ISCRAM Conference* , Vancouver, Canada.
4. Bruns, A., & Burgess, J. E. (2012). Local and global responses to disaster:# eqnz and the Christchurch earthquake. In Disaster and emergency management conference, *Conference proceedings. AST Management Pty Ltd.*, Bisbaine, 86-103.
5. Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3-4), 248-260.
6. Li, L., & Goodchild, M. F. (2010). The role of social networks in emergency management: A research agenda. *International Journal of Information Systems for Crisis Response and Management*, 2 (4), 49-59.
7. Latonero, M., & Shklovski, I. (2011). Emergency management, Twitter, and social media evangelism. *International Journal of Information Systems for Crisis Response and Management*, 3(4), 1-16.
8. Stollberg, B., & De Groeve, T. (2012, April). The use of social media within the global disaster alert and coordination system (GDACS). *In Proceedings of the 21st International Conference on World Wide Web* .
9. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge university press.
10. Ehrlich, K., & Carboni, I. (2005). *Inside social network analysis*. Boston: Boston College.

11. Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & information science research*, 18(4), 323-342.
12. Cheong, F., & Cheong, C. (2011). Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. *Pacific Asia Conference on Information Systems (PACIS) 11.*
13. Starbird, K., & L. Palen. (2010) 'Pass it on? Retweeting in mass emergency', *in Proceedings of the 7th International ISCRAM Conference,* Seattle: ISCRAM.
14. Sovilj, S., & Vasković, V. (2013). Analysis of corporate social networks using the JUNG framework. *Info M*, 12(48), 15-20.
15. Yan, Q., Wu, L., & Zheng, L. (2013). Social network based microblog user behavior analysis. *Physica A: Statistical Mechanics and Its Applications*, 392(7), 1712-1723.
16. Abilhoa, W. D., & De Castro, L. N. (2014). A keyword extraction method from twitter messages represented as graphs. *Applied Mathematics and Computation*, 240, 308-325.
17. Chatfield, A. T., Scholl, H. J. J., & Brajawidagda, U. (2013). Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services. *Government information quarterly*, 30(4), 377-386.
18. Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6), 52-59.
19. Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, 26(3), 10-14.
20. Morales, A. J., Borondo, J., Losada, J. C., & Benito, R. M. (2014). Efficiency of human activity on information spreading on Twitter. *Social Networks*, 39, 1-11.
21. Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, 40(10), 4065-4074.