

SOCIAL MEDIA ANALYSIS OF PEOPLE'S HIGH-RISK RESPONSES TO FLOOD OCCURANCE

SATOSHI ANZAI & SO KAZAMA

Tohoku University, Department of Civil and Environmental Engineering, Japan

ABSTRACT

During floods, some people approach flooded riversides out of curiosity, elevating their risk of being swept away by floodwaters. Social media was used to investigate actual patterns of this high-risk behaviour and to understand its temporal and spatial distribution. Among various social media, Instagram was chosen for its real-time nature and low-noise characteristics. The selected study areas included Tama river, Kinu river, Edo river, Tone river, Sagami river, and Hirose river in Japan, areas that were all affected by heavy rain from the 7th to the 12th of September 2015. Data was collected by Instagram API detecting hash-tag search of the river names, such as “#Tama river”. The collected data was assessed by time series analysis, text analysis, and image analysis. The results indicated that usage of the relevant hashtags increased following a flood warning being issued. Additionally, the number of posts following the flood warning was influenced by population characteristics along rivers and warning issue time. Thirdly, results revealed that many people continued to remain on the riverbanks despite their awareness of the increasing risk.

Keywords: big data, Instagram, text analysis.

1 INTRODUCTION

In recent years, flood risk by heavy intensive rainfall has been increasing along with the intensification of global warming. The effect has been especially strong in Japan, because the rivers generally have a steeper slope than the rivers in continent countries [1]. Accordingly, during heavy rainfall the water level in Japanese rivers rises immediately. This characteristic has caused enormous human casualties in Japan. Ushiyama [2] surveyed characteristics of incidence of death or missing persons cases by heavy rainfall from 2005 to 2007. The study indicated that 23% of casualties resulted from people approaching flooded rivers spontaneously. This result underlines the importance of analysing people's level of interest in the river during heavy rainfall periods and the spatiotemporal distribution of this interest.

In past studies, Miceli et al. [3] surveyed disaster preparedness and perception of flood risk in Italy by questionnaire survey. Kellens et al. [4] surveyed the perception of flood risks of residents and tourists along the Belgian coast. These studies mainly collected and analysed questionnaire-based surveys, however real-time data during ongoing disasters was not used. Questionnaire surveys are able to clarify people's overall attitudes towards disasters during normal life, but they are unable to assess perceptions and attitudes in real time during an ongoing disaster on the other hand, content posted on social media includes real-time data on user reactions and behaviour. Analysis of social media content has been actively used in recent years for assessment of public perceptions and has found application in disaster management. Actual behaviour of people during a disaster can be also be assessed by this method. A better understanding of actual behaviour can contribute to improving evacuation plans. Bird et al. [5] discussed the results of the study of Facebook data generated during the floods, and the application of social media to the emergency services. Gao et al. [6] described the advantages and disadvantages of social media applied to disaster relief coordination. Yates and Paquette [7] discussed the problems of using social media in emergency management. Cheong and Cheong [8] analysed the structure of online communities to improve the efficiency of spreading information during a disaster. However, the risk-taking behaviour



of people during a disaster was not included in previous analyses. Therefore, the current study aims to fill this gap by employing social media data for analysing the behaviour of people around rivers during heavy rainfall.

2 DATA AND METHODS

2.1 Data

As stated previously, social media data has certain advantages over previously employed forms of data for grasping temporal and spatial distribution of a social phenomenon. Among various social media, Instagram data was judged by the author to be the most optimal for this research owing to its real-time nature with low noise generation. Instagram is a social media platform for sharing photos. In comparison with Instagram, Facebook does not offer the same volume of accessible data, because many users utilise privacy protection measures, while Twitter has high noise due to the employment of post-duplicating “bots” by some users to promote their online presence. Data were collected by Instagram Application Program Interface (API) detecting hash-tag search with name of rivers such as “#Tama river”. The data offered by the Instagram hash-tag search function includes posting time, texts, images and other details.

We selected six rivers Tama river, Kinu river, Edo river, Tone river, Sagami river and Hirose river in Japan for analysis (Fig. 1), which were struck by a heavy rainfall from the 7th to the 12th September 2015. Kinu river in particular was one of the most heavily damaged areas. Floods caused by this rainfall destroyed or damaged over 7000 houses and consequently about 4300 people needed rescue.

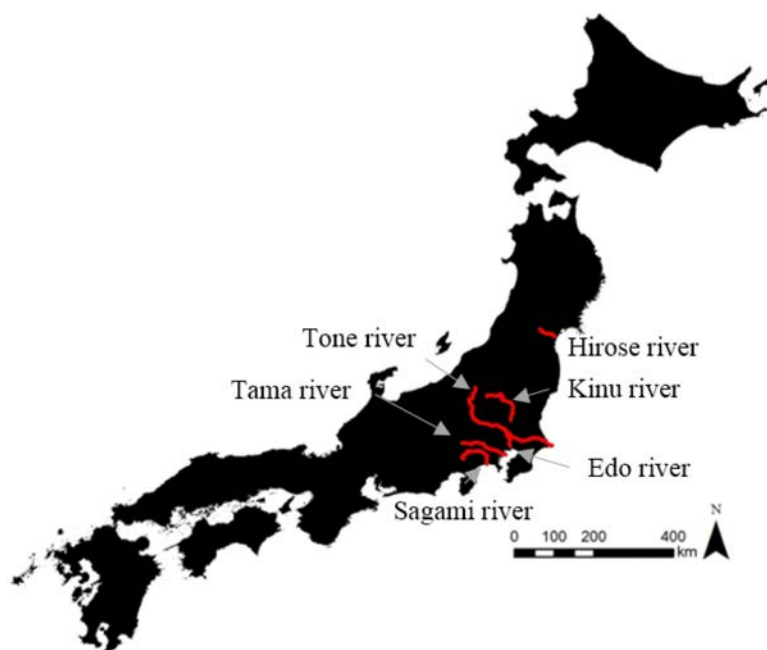


Figure 1: The target rivers in Japan.

2.2 Methods

Data was analysed by time series analysis, text analysis and image analysis. The time series analysis was used to understand fluctuation in public interest concerning the rivers. Time series changes in the number of posts were compared with the time a flood warning was given and recalled and with the average number of posts. Japanese government sets five stages in disaster warning, as follows: level 1 promotes people's awareness of a disaster risk; level 2 calls for preparation for an evacuation; level 3 encourages the public to begin evacuation procedures; level 4 indicates that evacuation procedures should be completed by this time; finally, level 5 informs of the arrival of floodwaters. In this study, only warnings that are level 2 or above are defined as "flood warning" because level 1 is assumed to be less urgent. To simplify the analysis, the posting time of each post and the warning issue time were rounded down by 30 minute-intervals (e.g. 23:52 to 23:30). The rounding down by 10-minute, 20 minute and 30 minute intervals was tried and no significant effect was found on the results. Accordingly, 30 minute intervals were used for simplicity.

The average Av number of posts in each 30-minute interval was calculated by excluding the number of posts during flood warning from the total number of posts in September (eqn (1))

$$Av = \frac{np_A - np_w}{t_A - t_w}, \quad (1)$$

- np_A is the number of posts in the September.
- np_w is the number of posts during flood warning.
- t_A is the number of 30-minute time steps in September.
- t_w is the number of 30-minute time steps during flood warning.

The text analysis carried out to assess user perceptions of river conditions. Mecab, the open source Japanese language analysis system, was used for this purpose. It is useful to separate one sentence to each word and obtain frequency for each word. Tables of frequency rankings of words were generated by Mecab. In order to investigate the number of posts recognizing dangerous river conditions, certain words indicating unusual conditions of a river were extracted from the posts. To give a simple example, words "flood", "muddy stream" and "high flow" were judged to indicate perception of unusual conditions among users. The posts including words indicating unusual condition of a river were counted. The ratios R of the number of such posts to the number of posts during flood warning were calculated as expressed in eqn (2).

$$R = \frac{np_r}{np_w} \times 100, \quad (2)$$

- np_r is the number of posts recognizing unusual condition of a river.

As Instagram is primarily an image-sharing platform, most posts include photographs along with textual commentary; posts regarding rivers include photographs of said rivers. Image analysis was used to investigate the distance from the photographer to the river. The location of the poster taking the photo was manually determined by observing the photographs. The recorded locations were grouped into eight distance categories from "d0" to "d7". "d1" refers to images taken by the waterside, on the surface of the water or under the water. "d2" describes images taken on the riverbed. "d3" images were taken on the bank or



on the bridge. “d4” images were taken from the landside, for example from an apartment building or park some distance away from the river. “d5” images were taken from transport such as cars or buses. The “d6” category refers to images that only included a screenshot of a flood warning, while “d7” refers to images that either were unrelated to the river, or those where the distance could not be determined. “d0” was used in cases where there were no images included with the post.

3 RESULTS

3.1 Results of the time series analysis

Fig. 2 shows the time series change in the number of post with the hashtag of each river. Fig. 2(a) shows a result in the case of Tama river. The warning was issued on 9th September from 17:50 until 21:23 in Tama river. A_v was approximately 1.2 posts in September. After the warning, the number of posts increased rapidly. Tama river is located near a big city, therefore many people are able to observe the river regularly and in particular in times of changing river conditions. Fig. 2(b) shows result of Edo river. The A_v for September was approximately 0.28 posts. It rained heavily around the rivers, but a flood warning was not issued. Despite the fact that Edo river is located near Tama river, there was no parallel increase in the number of posts. People living along Edo river did not pay attention to the river due to the lack of a flood warning. This demonstrates the impact of flood warnings in stimulating elevated levels of attention towards rivers from the population. Fig. 2(c) shows the results for Sagami river. The warning was issued from 9th September from 16:03 until 20:03 in Sagami river. A_v was approximately 0.08 posts in September. The number of posts did not increase even after the flood warning. The typical A_v of Sagami river is much lower than that of other rivers; this indicates general low public interest in Sagami river which does not change much even under flooding conditions. Fig. 2(d) shows the results for Hirose river. A flood warning for Hirose river was issued from 10th September at 23:21 to 8:01 on 11th September. A_v was approximately 0.11 posts in September. After the warning, the number of posts increased rapidly. Like Tama river, Hirose river is located near a big city, hence it is observed by large numbers of people. Fig. 2(e) shows the results for Kinu river. The warning was issued from 9th September at 20:40 until 6:40 on 11th September. A_v was approximately 0.69 posts in September. After the warning, did not rapidly increase. Kinu river is located near small cities, thus, few people pay attention to the river. However, after the floodwaters overtopped the local dikes, causing partial collapse, the number of posts did significantly increase. Because overtopping of dikes was reported by local mass media, many people began to pay attention to the river. Fig. 2(f) shows the results for Tone river. The warning was issued from 9th September at 23:39 until 9:50 on 12th September. A_v was approximately 0.17 posts in September. After the warning was issued, the number of posts did not rapidly increase. Because Tone river is located in rural areas and the warning started at midnight, there were few posts when it was first issued. In the morning, the number of posts increased. These results suggest that the level of interest in river conditions even following a flood warning highly depends on the population characteristics of the area along the river and warning issue time.

3.2 Results of text analysis and image analysis

Textual content of posts was analysed to evaluate the public's perception of the river conditions. Table 1 shows the ranking of frequency of words in the posts during flood



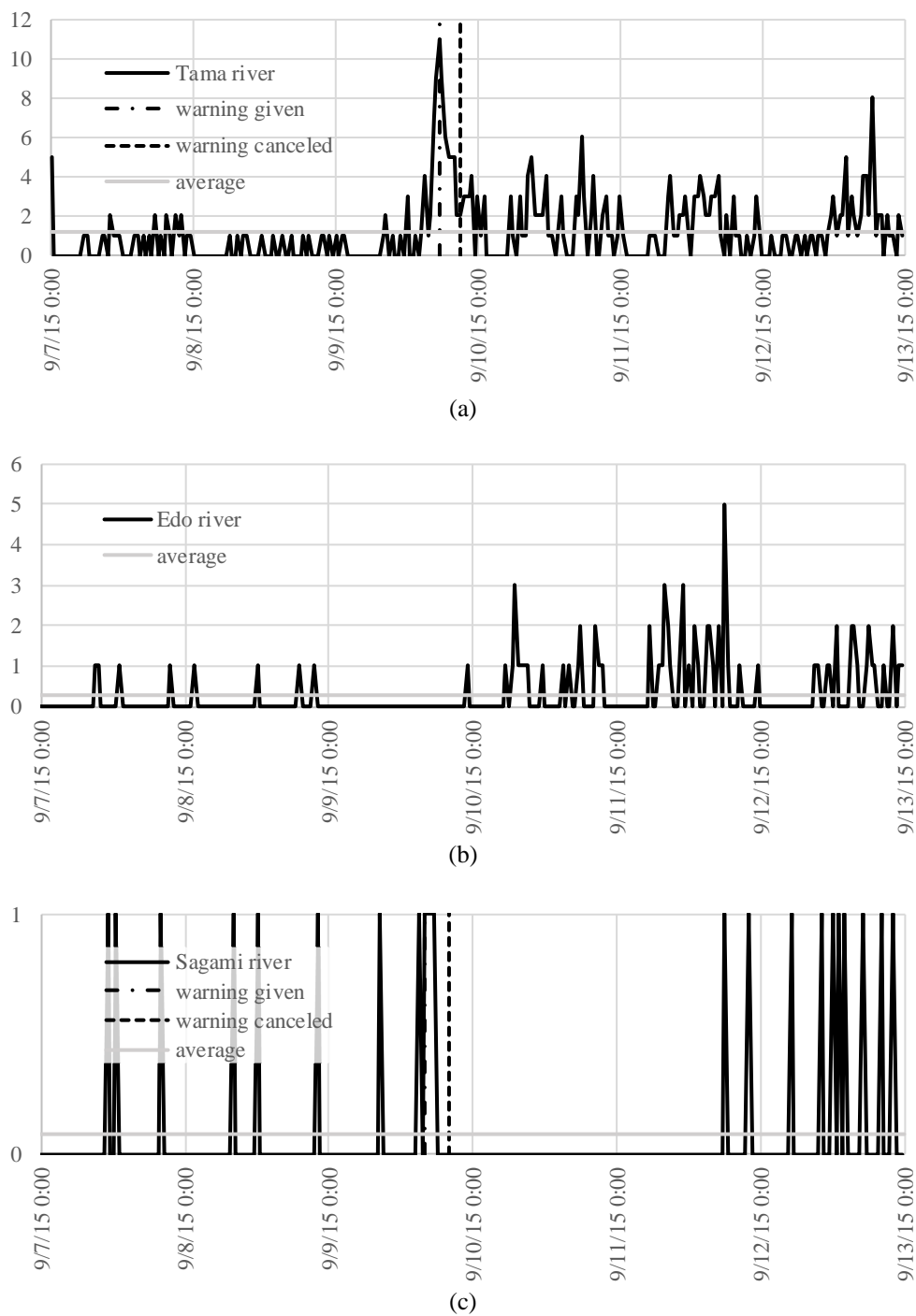
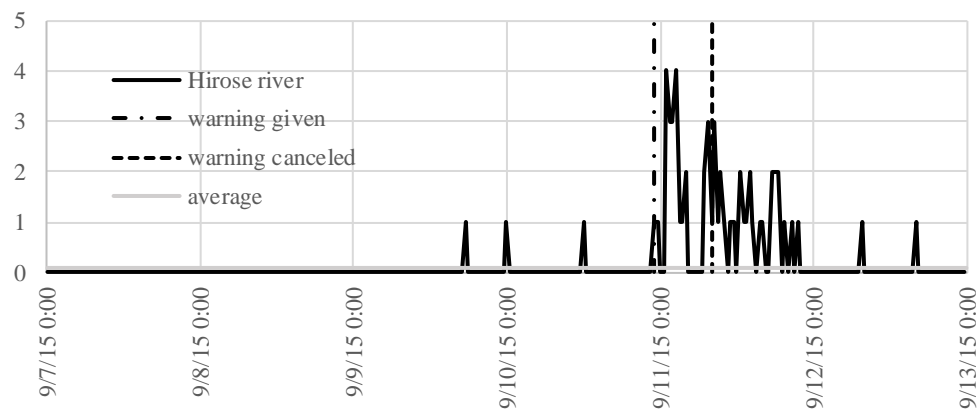
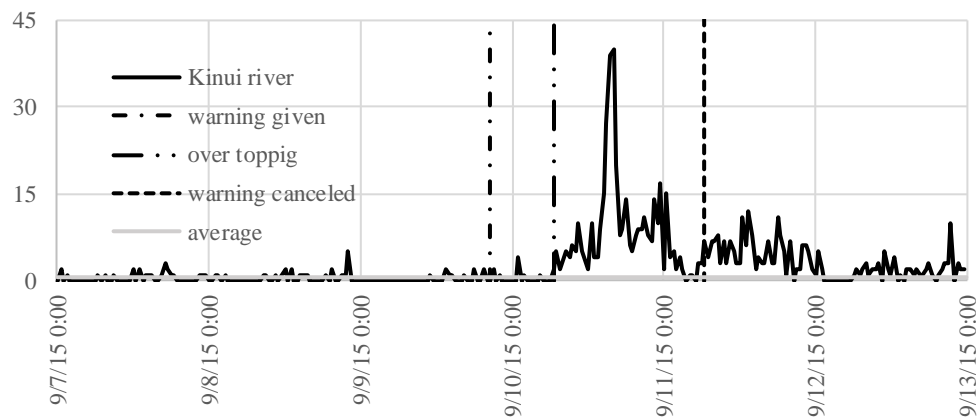


Figure 2: The time series change of the number of posts. (a) Tama river; (b) Edo river; (c) Sagami river; (d) Hirose river; (e) Kinu river; (f) Tone river.

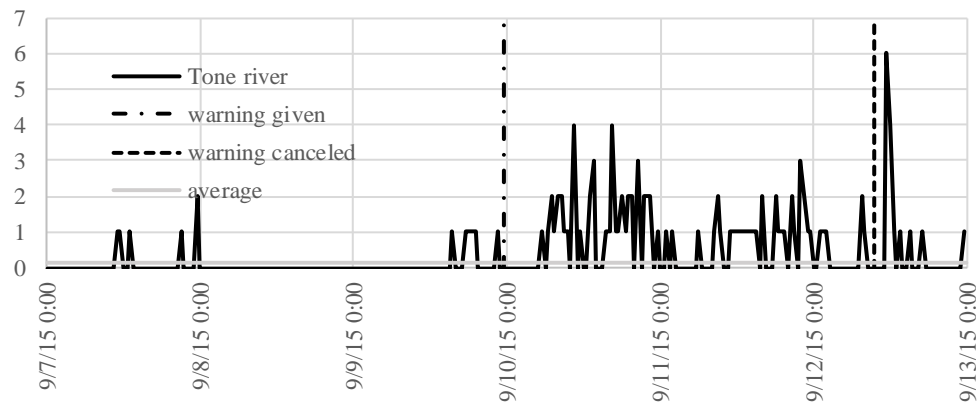




(d)



(e)



(f)

Figure 2: Continued.



Table 1: The frequency rankings of words from 1st to 10th during flood warning in each river. (a) Tama river; (b) Hirose river; (c) Kinu river; (d) Tone river.

(a)			(b)		
rank	word	posts	rank	words	posts
1	Tama-river	39	1	Hirose-river	24
2	typhoon	16	2	Sendai	11
3	rain	9	2	flood	11
4	rainbow	8	2	refuge	11
5	flood	7	5	advise	9
6	Futakotamagawa	6	6	heavy rain	7
6	sky	6	7	bridge	5
6	heavy rain	6	8	city	4
9	river	5	8	water level	4
10	riverbed	4	8	warning	4
10	rainbow	4	8	river	4
10	overflow	4	8	Hirose-	4
10	bank	4	8	typhoon	4
10	water	4			
10	water level	4			
10	sun	4			
(c)			(d)		
rank	word	posts	rank	word	posts
1	Kinu-river	410	1	Tone-river	85
2	Ibaraki	144	2	Ibaraki	17
3	rescue	124	2	flood	17
4	city	116	4	high flow	15
5	collapsed	115	4	rain	15
6	Jouso	114	6	typhoon	14
7	flood	110	7	heavy rain	13
8	safety	89	7	water level	13
9	prefecture	86	9	river	11
10	Self - Defense Forces	84	10	prefecture	10

warning in Tama river, Hirose river, Kinu river, and Tone river. Posts related to Edo river and Sagami river were not analysed by text analysis because there were few posts during flood warning. Japanese particles, punctuation marks, and non-textual symbols were excluded from the rankings.

Table 1(a) shows the results for Tama river. The words “flood” and “over flow” were the 5th and 10th most frequently used. On the other hand, “rainbow” was the fourth most frequently used, because the weather cleared before the flood warning was cancelled and people’s interest was drawn to the rainbow that appeared at this time. R of eqn (2) was 31%.

Table 1(b) shows the results for Hirose river. “Flood” was the 3rd most frequently used word and other unusual words were also used. This is because flood warnings have rarely been issued for Hirose river. R for Hirose river was 58%. Table 1(c) shows results for Kinu river. The words “collapsed” and “flood” were the 5th and 7th most frequently used, respectively. Additionally, words indicating an emergency situation were also used frequently. Because the dikes on Kinu river partially collapsed, the people along Kinu river were more likely to perceive themselves to be in danger. R for Kinu river was 57%. Table 1(d) shows the results for Tone river. The words “flood” and “high flow” were the 2nd and 4th most frequently used. R is 40%. Hence, words illustrating the change of river conditions such as flooding and high flow were frequently used, and about half of posts display recognition of dangerous conditions.

We analysed images in posts during flood warning to determine the distance from the rivers to the users taking the photograph. Kinu river was not analysed by image analysis, because due to extensive flooding d2 and d3 could not be judged from the images.

Table 2 shows the number of posts in each distance category in posts displaying recognition of dangerous river conditions and in posts not displaying this recognition. The

Table 2: The number of posts in each distance category (*). (a) Posts recognizing the dangerous condition of river; (b) Posts not recognizing the dangerous condition of river.

(a)							
	d1 & d2	d3	d4	d5	d6	d7 & d0	total
Tama river	2	6	3	0	1	0	12
Hirose river	1	3	3	0	4	3	14
Tone river	3	17	5	4	1	4	34
(b)							
	d1 & d2	d3	d4	d5	d6	d7 & d0	total
Tama river	4	13	3	0	1	6	27
Hirose river	0	3	3	0	1	3	10
Tone river	8	20	4	3	2	14	51
(*) Eight distance categories are defined.							
distance	define						
d1	by the waterside, on the surface of the water or under the water						
d2	on the riverbed						
d3	on the bank or on the bridge						
d4	from the landside						
d5	in the transportation						
d6	only contents of a warning						
d7	unconcerned or unable to judge						
d0	no image						

results indicate that images were taken most frequently on the river bank (“d3”) in Tama river and Tone river during the flood warning. At the same time, the social media response to Hirose river flooding was dominated by posts in the “d6” category. This is likely because the flooding was highly unusual for that area, and people were afraid to approach the river closely.

The results of text analysis and image analysis suggest that many people approached flooding rivers even though they were aware of dangerous river conditions.

4 CONCLUSIONS

Three conclusions can be drawn from the results of this study. First, the number of posts increased following the issuance of a flood warning in most cases. Additionally, the number of posts during the flood warning was influenced by the population surrounding rivers and warning issue time. Finally, many images with posts were taken along the rivers while the flood warning was in effect and the people were aware of the increased risk. Generally, the government issues flood warnings in order to make people aware of the danger of flooding and take precautionary measures. However, our results indicate that along with increasing awareness of dangerous conditions, the warning can also increase people’s interest in the river and lead them to potentially expose themselves to risk.

ACKNOWLEDGEMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in Aid for Challenging Exploratory Research, 2015–2017 (15K14036, So Kazama). We wish to thank Tohoku Regional Development Association for their generous financial assistance. This publication was supported by Masaki Sawamoto research grant.

REFERENCES

- [1] Yoshimura, C., Omura, T., Furumai, H. & Tocker, K., Present state of rivers and streams in Japan. *River Research and Applications*, **21**, pp. 93–112, 2005.
- [2] Ushiyama, M., Characteristics of human damage caused by heavy rainfall disasters in Japan from 2005 to 2007. *Geographical Reports of Tokyo Metropolitan University*, **42**, pp. 49–56, 2008.
- [3] Miceli, R., Sotgiri, L. & Settanni, M., Disaster preparedness and perception of flood risk: A study in an alpine valley in Italy. *Journal of Environmental Psychology*, **28**, pp. 164–173, 2008.
- [4] Kellens, W., Zaalberg, R., Neutens, T., Vanneuville, W. & De Mayer, P., An analysis of the public perception of flood risk on the Belgian coast. *Risk Analysis*, **31**(7), 2011.
- [5] Bird, D., Ling, M. & Haynes, K., Flooding Facebook? The use of social media during the Queensland and Victorian floods. *The Australian Journal of Emergency Management*, **27**(1), 2012.
- [6] Gao, H., Barbier, G. & Goolsby, R., Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, **26**(3), 2011.
- [7] Yates, D. & Paquette, S., Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management*, **31**(1), pp. 6–13, 2011.
- [8] Cheog, F. & Cheong, C., Social media data mining: A social network analysis of tweets during the 2010–2011 Australian floods. *Pacific Asia Conference on Information Systems, PACIS 2011: Quality Research in Pacific Asia*, Brisbane, Queensland, Australia, 7–11 Jul. 2011.

