# CPS Model Based Online Opinion Governance Modeling and Evaluation of Emergency Accidents

Xiaolong Deng
Beijing University of Posts and Telecommunications
Beijing, China
shannondeng@bupt.edu.cn

Yingtong Dou
Beijing University of Posts and Telecommunications
Beijing, China
douyingtong@bupt.edu.cn

Yihua Huang
4399 Network Co., Ltd,
Xiamen, China
fatree@4399inc.com

## ABSTRACT

In the last decades, there have been much more public health crises in the world such as H1N1, H7N9 and Ebola outbreak. In the same time, it has been proved that our world has come into the time when public crisis accidents number was growing fast. Sometimes, crisis response to these public emergency accidents is involved in a complex system consisting of cyber, physics and society domains (CPS Model). In order to collect and analyze these accidents with higher efficiency, we need to design and adopt some new tools and models. In this paper, we used CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection and decision making in the back end. Based on the online opinion data we collected, we also proposed the graded risk classification. By the risk classification method, we have built an efficient CPS Model based simulated emergency accident replying and handling system. It has been proved useful in some real accidents in China in recent years.

## CCS Concepts

**•Special-Purpose and Application-based System → Miscellaneous;**

## Keywords

APP Data Collection; CPS Model; Online Opinion Analysis; Emergency Disaster; Situation Evaluation

## 1. INTRODUCTION

The probability of public crisis and accidents around the world is increasing quickly in the past several years i.e. H1N1, Ebola and Typhoon Haiyan. Emergencies with severe effect like the Explosion of Tianjin Port also lead a huge fluctuation of online opinion in China. The increment and spread of web user makes the web becomes a major role in the information propagation. The evolution of online opinion about emergencies has become a vital factor in the management of crisis for our government [1, 2].

We have promoted CPS model based online opinion governance modeling method and the evaluation of emergency accidents. At first, the system collected the online opinion from three-dimension space (cyber, physical and society) [3, 4, 5]. In the physical space, we collected some physical information like the data about damages and deaths. For the cyber space, with the aid of distributed mining system, we collected online opinion from Sina Micro blog, Baidu News, Tianya BBS and Twitter. In the society space, we analyzed the trend of online opinion based on social network. According to the information collected from several spaces, the calculation and analysis will be applied. At last, the decision will be made based on different situations to deal with various kinds of emergencies according to our simulation computing.

In detail, our contributions are as follows:

1. We designed and implemented a CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection and decision make in the back end and it has some very good performance;

2. We proposed a detailed risk classification method with five layered risk. Also we proposed the accident types classification. It will be useful for emergency accident risk assessment;

3. We proposed a detailed analysis of Tianjin explosion on August 12, 2015 which combing the cyber information, physical information and society information.

The outline of the paper is as follows: In Section 1, the research background is introduced. In Section 3, we offered CPS model based analysis system and our APP based system framework. Furthermore, we offered the opinion mining result of 2015 Tianjin Explosion Accident in Section 4. And Section 5 gives a conclusion.

## 2. THE CPS MODEL BASED ONLINE OPINION GOVERNANCE

The CPS model has three main steps. The first step is the collection of online opinion. The second step is the calculation and analysis of collected data and the last step is decision making and feedback. Here we will introduce the CPS model.

### 2.1 CPS Model



*EM-GIS 16, October 31-November 03 2016, Burlingame, CA, USA*



DOI: http://dx.doi.org/10.1145/3017611.3017619

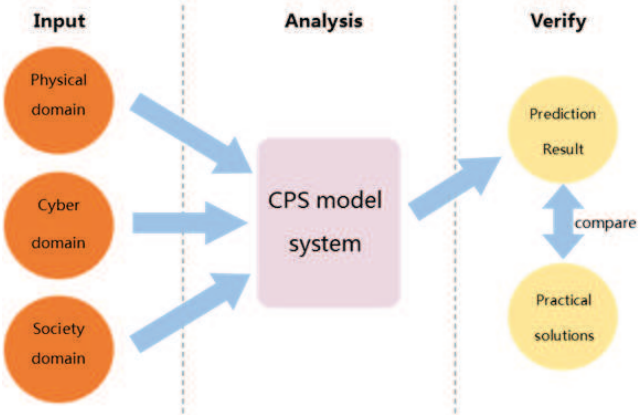


**Figure 1: CPS Model Based Risk Reply**

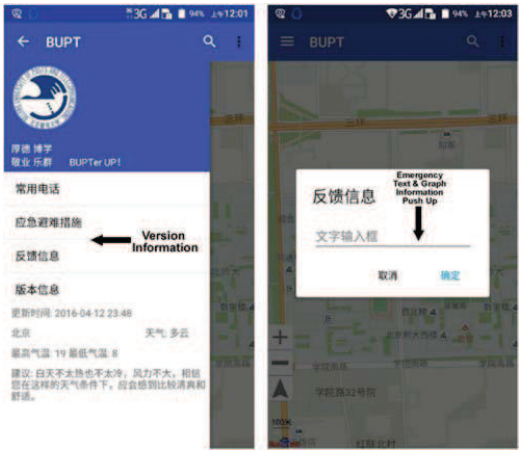


**Figure 2: APP GUI**

The basic way to evaluate the model is analyzing the event that have occurred is input to the model. Compare the conclusion of event with the actual process and verify whether they are similar or not.

Crisis response is involved in a complex system consisting of cyber, physics and society domains. Different domains are closely coupled and dynamically evolved. Cross domain analysis is applied for better response of natural disasters. A data-model integration method is proposed to support situation awareness and emergency response planning. Models are used for the prediction of crisis scales and its impact on infrastructure system and local society. Data mining is used to sense society readiness and mentality of affected citizen. The integration of cross-domain data mining and emergency modeling is essential for supporting scientific decision-making (planning and response) [3, 4, 5].

## 2.2 Opinion Collection APP

To collect the online opinion, we combine the way using APPs and web crawlers together. The mobile APPs collect the information of events happened in the real world (i.e. photos, locations, time). The web crawler gets the online opinion in the web. The APP-based information collection is a user-driven information pushing mechanism, i.e. the client-server structure. Users send information (including longitude and latitude information) of emergency accidents via specific APPs like what is shown in Figure 2 to the server where store the storing and analyzing the data.

## 2.3 CPS Model

It can be found from Figure 2 that user can send out the accident information in emergency and this is a demo GUI for emergency report for our university campus (www.bupt.edu.cn). In our system, user can send out text information, photo information and location information (including latitude and longitude information). After the information was sent to back end server, it can be extracted and helpful for further data mining in Figure 3.

The web crawler is a server-based online information collection process. The web crawler running 7*24 hours on the server collects the online information about the emergencies from BBS, micro-blogging and others websites. Besides, the crawler could also collect information from news website and other online libraries according to the specific emergency ac-

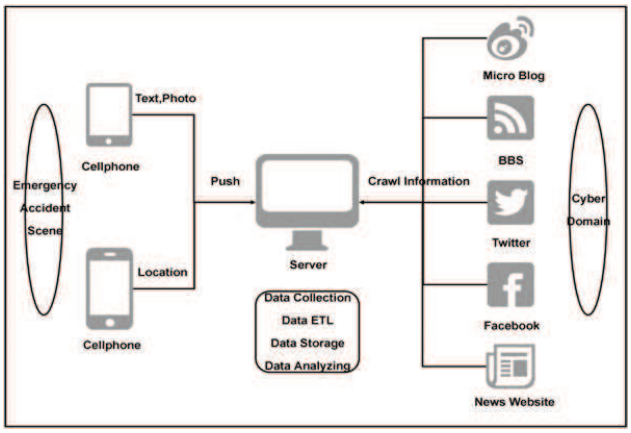


**Figure 3: Opinion Collection System Framework**

cident such as Micro Blog, BBS, Twitter, and Facebook. As a result, we offered some suggestions to the government to handle opinion situation in these emergency accidents [6, 7, 8, 9].

# 3. GRADED EMERGENCY ACCIDENT RISK

## 3.1 Risk Classification

According to the accumulated emergency accidents by our research team in recent year, we have proposed a detailed risk classification method with five layered risk which can be found in Table 1.

Risk grade judgment is based on the content analysis. An important step in the content analysis is to extract and quantify text. The content builds an identifiable vector relating to the key words of risk grade in the content of Micro Blog and WeChat at the initialization of pre-judgment model. There is a method combining the word frequency method, the mutual information method and the information entropy method to extract, identify with cleaning and finally transfer the eigenvector [7, 8, 9].

Meanwhile, the word frequency method is the simplest method for extracting the eigenvector, taking the frequency scale of key words as a criterion to choose the feature words

**Table 1: Graded Emergency Accident Risk**

| Risk grade and characteristics | Events types | Events examples |
|---|---|---|
| **Primary risk**<br>Characteristics: events that cannot be prejudged and controlled by humankind and that are unexpected and greatly harmful to the safety of life and property as well as extremely bed effects on the stability of country and society. | Natural disaster, Geologic and meteorological disasters, Ecological disaster, Violence and terror incidents | HaiYan typhoon, Wenchuan earthquake, SARS, Violence and terror incidents at Kunming Station |
| **Secondary risk**<br>Characteristics: unexpected mega-events due to human factor, resulting in a certain damage to the safety of life and property and to the economy and society. | Safety accidents<br>Traffic accidents<br>Environmental pollution in a large scale and so on | Fire in Harbin 2016, Jan, 2nd.<br>Explosion at Tianjin Port |
| **Tertiary risk**<br>Characteristics: events that has no threat to safety of life and property, but a great influence on the social stability and public opinion guidance | Issue of law about livelihood<br>Implementation of policy and so on<br>National and even global events caused by technological economy | Delay retirement under the pension dual system<br>Anticorruption within party<br>The crash of stock market, circuit breakers and share trading stops<br>Global financial crisis |
| **Quaternary risk**<br>Characteristics: events that has no threat to safety of life and property, but a certain influence on the social stability and public opinion guidance | Domestic and foreign policy;<br>Diplomatic incidents | American presidential election, the TPP accident,<br>Agreement to Xi Jinping visit<br>Parade on Victory over Japan Day |
| **Others**<br>Characteristics: informational events that has no threat to safety of life and property and no significant influence on the social stability and public opinion guidance | Public figures<br>Events like typical phenomenon of society and so on | Li Shishi vs. AlphaGo<br>Attack on a girl at H hotel |

and considering that a word with increasing frequency of occurrence in text is more representative.

The mutual information method is to measure the relevance between the feature item and a particular category at the time of extraction of feature words. If the calculating value of mutual information is greater, the relevance between the corresponding feature item and the category is higher; instead, the relevance between the corresponding feature item and the category is lower, which should be dropped out.

Given the joint distribution of two random variables $(X, Y)$ is $p(x, y)$, the marginal distribution is $p(x)p(y)$, and the mutual information $I(X;Y)$ is the relative entropy of the joint distribution $p(x, y)$ and the product distribution $p(x)p(y)$:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)} \qquad (1)$$

The information gain method is mainly according to the information quantity (i.e. the information entropy) of feature item in the entire category system to measure the importance of the feature, and to determine its selection with the contribution degree feature item to the category system. To calculate the contributed information quantity is to compare the change of systematical information quantity under the circumstance of the selection or nor. In information theory, the entropy is to measure the expected value of a random variable. The entropy value H of a random variable X beneath the range of $\{\chi_1, ...\chi_n\}$ is defined as $H(x) = E(I(X))$. Meanwhile $I(X)$ is the self- information of random variable $X$. At the same time, another expression form of entropy $H$, which is expanded according to the definition of expectation and the formula of self- information, is:

$$H(X) = \sum_{i=1}^{n} p(\chi_i)I(\chi_i) = -\sum_{i=1}^{n} p(\chi_i)logp(\chi_i) \qquad (2)$$

The information entropy is to measure the average expected value of the random variable whose sentence appears in the form of some word combination mode. And it can be used to calculate the sentence information entropy according to the value selection of the random variable in the mentioned range.

Using the above methods, it can carry out the extraction and identification to some particular content key words when searching, and then calculate the oriented risk grade. If there is a basic conformity, a primary early- warning can be launched; if not, other operations could be continued. For example, in the primary risk judgment procedure, key words, such as “great”, “bad” “injuries and deaths”, can be marked when building the identifiable vector of key words of Micro Blog content, using the word frequency method to decrease the data dimension, to abolish the noisy data and to reduce the scope. To confirm the relevance between the Micro Blog content and “the primary risk category”, the mutual

information method can be used to calculate the mutual information value (relative entropy) of feature quantity of key words, while finally to confirm the contribution degree of the Micro Blog content to the entire "primary risk category" for determining the selection situation of the feature quantity [10, 11, 12].

## 3.2 Accident Types

Depending on the type of event emergency network public opinion, it can be divided into five areas, political, society, the people's livelihood, natural or man-made disasters and other.

**Table 2: Accident Types**

| The type of event | Instructions |
|---|---|
| Political | Diplomatic or political incidents happened at home and abroad policy implementation. |
| People's livelihood | The events involving the people's livelihood, such as laws enacted and implementation of policy. |
| Society | The events involving the public safety, social celebrities and social phenomenon etc. |
| Natural/Man-made disasters | Natural disasters, accidents, transportation accidents and environment pollution etc. |
| Other | The events triggered by the science, technology and economic factors. |

# 4. CPS MODEL ANALYSIS OF TIANJIN EXPLODES ION

The 2015 Tianjin explosions were a series of explosions that killed over one hundred people which occurred at a container storage station at the Port of Tianjin on Wednesday, 12 August 2015. The first two explosions occurred within 30 seconds at the factory, which located in the Binhai New Area of Tianjin, China. The second explosion was much larger and caused by about 800 tons of ammonium nitrate. Fires caused by the initial explosions continued to burn out of control throughout the weekend, repeatedly causing some secondary explosion which eight additional explosions occurring on Saturday, 15 August.

The cause of the explosions was not immediately known at first, but an investigation was concluded in February 2016 that an overheated container of dry nitrocellulose was the cause of the initial explosion.

Poor coverage of the event and the emergency response to it received criticism. By 12 September 2015, the official casualty report was 173 deaths, 8 missing, and 797 non-fatal injuries.

## 4.1 Cyber Domain

Figure 4 is the overall trend line which based on the result of Baidu index when "Tianjin explosion" as the search keyword. As it can be seen from the chart, the day of August 13 incident search index is at its peak, the next two days to August 15 search index reached a valley, before August 17, the search index more moderate ups and downs, 8 after May 24, the event has been concern about completely


**Figure 4: Online Public Opinion Data**

normal situation.

We have found Baidu search results page more than 3,210,000 points and Baidu news over 3,050,000 articles by keywords of "Tianjin explosion" by June 9, 2015.And the news has been broadcasted by SinaWeibo over 66,962,864time and Tencent Micro Blog over 519,465 times. We collected the data and analyzed the case by some useful methods [13, 14, 15, 16].

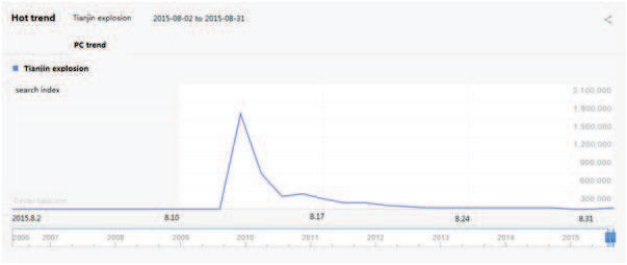


**Figure 5: Online Public Opinion Data Trend of PC End**

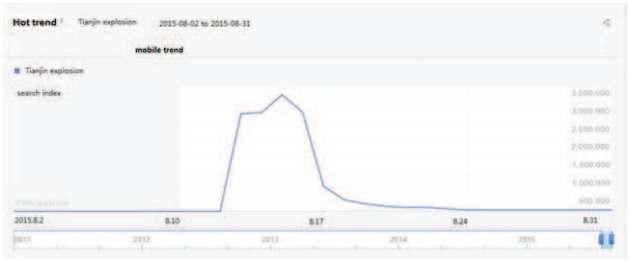


**Figure 6: Online Public Opinion Data of Mobile End**

Figure 5 and Figure 6 above represented the "Tianjin explosion" search results in 24 hours form the incident happened. The mobile terminal searches reached its peak on the third day of the event. Overall, the trend of attention from the mobile terminals is much higher than that from the PC side, to some extent, showing timeliness and convenience of mobile terminal.

## 4.2 Society Domain

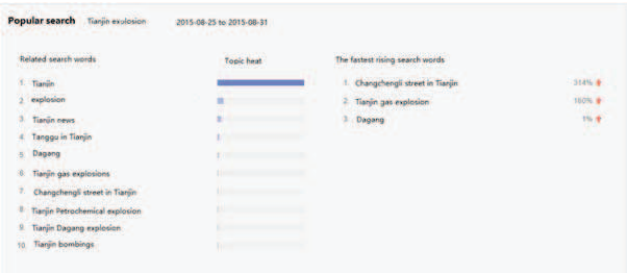


**Figure 7: Hot Search Distribution of Tianjin Event**


**Figure 8: Word Frequency of SINA Micro Blog**

## 4.3 Further Data Analysis

We can get event data area of physical domain and the initial situation. Our system (in adjustable intervals) collected can collect the online information from social network in each 5 minutes, and the user can monitor events in real-time development of the state, according to meteorological and hydrological events traffic situation around the area develop mass evacuation routes, events such as the deployment of firefighters' disposal program.

We use a method for automatic summarization based on LDA model and information entropy for Chinese document. It uses LDA model to do shallow semantic analysis work on documents and gets the distribution of topics under each document. Through analyzing the topics of document, we got the topic which has the best expression of central idea for document. Meanwhile, this paper proposed a new method to compute the sentence weight and extract the most important sentence based on measuring the information entropy for each sentence. It treats the sentence as a random variable and calculates the information entropy for every random variable. Experimental results show that this method can pick out the most important sentence in the document.

By the above table it can be found that, at the beginning of the Tianjin Explosion event, people greatly concerned about the incident itself. And lately people have expressed concern about the injured persons.

The above table shows public opinion for these accident victims when the incident occurred. Lately, people became concerned about conducting a thorough investigation and

**Table 3: Word Frequency Rank in the Early Stage**

| Words | Part of speech | Weights | The number of occurrences |
|---|---|---|---|
| explosion | vn | 5.68 | 2753 |
| occur | V | 5.82 | 874 |
| Tianjin Binhai New Area | n-newword | 6.52 | 599 |
| explosion | n-new | 10.07 | 463 |
| Clifford | V | 6.24 | 457 |
| Dangerous goods stores | n-newword | 3.81 | 338 |
| Hope | V | 6.7 | 300 |
| Tianjin Binhai | n-newword | 4.49 | 280 |
| Accident | n | 6.71 | 271 |
| Victims | vi | 4.95 | 269 |
| hospital | n | 6.7 | 260 |
| Help | vn | 6.75 | 246 |
| Know | v | 5.78 | 232 |
| store | n | 3.67 | 227 |
| Injured | vi | 4.86 | 218 |
| blessing | v | 7.06 | 191 |
| Binhai New Area | n-newword | 7.89 | 191 |

**Table 4: Word Frequency Rank in the Late Stage**

| Word | Part of speech | Weights | The number of occurrences |
|---|---|---|---|
| Explosion | vn | 6.23 | 7306 |
| Tianjin | ns | 8.78 | 4942 |
| explosion | n-newword | 8.23 | 1409 |
| occur | v | 7.36 | 1176 |
| victims | n-newword | 8.73 | 1020 |
| accidents | n | 8.63 | 973 |
| Tianjin harbor | ns | 6.76 | 881 |
| Tianjin Binhai New Area | n-newword | 8.29 | 861 |
| Ruihai company | n-newword | 9.3 | 695 |
| News | n | 7.97 | 674 |
| government | n | 9.09 | 582 |
| Hope | v | 7.69 | 528 |
| Clifford | v | 7.01 | 518 |
| Disaster | n | 7.45 | 486 |
| Marina | n | 7.09 | 470 |
| The company | n | 8.65 | 469 |
| life | n | 7.56 | 446 |
| know | v | 6.7 | 437 |
| help | vn | 8.03 | 419 |
| survey | v | 6.47 | 419 |
| warehouse | n | 6.07 | 404 |
| Binhai New Area | n- newword | 9.21 | 391 |
| Country | n | 8.52 | 386 |
| Hero | n | 6.82 | 377 |
| Jobs | vn | 7.53 | 369 |

respect to the sacrifices of the heroes of fire fighter.

## 5. CONCLUSION

In this paper, we proposed a new CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection and decision make in the back end. And based on the online opinion data we collected, we also proposed the graded risk classification method and accident classification method. Based on the risk classification method, we have built an efficient CPS Model based simulated emergency accident replying and handling system. It has been proved that integrated data from cyber domain, physical domain, and society domain will help us to make the final management decision and we proposed the detailed analysis of 2015 Tianjin Explode accident.

## 6. ACKNOWLEDGMENTS

Thanks to National Natural Science Foundation of China (NSFC) (Grants No. 91124002), National Culture Support Foundation Project of China (2013BAH43F01), and National 973 Program Foundation Project of China (2013CB329600) in social network analysis. We appreciate direction from professor Hui Zhang and his aid form Joint-Operated project from National Natural Science Foundation of China (NSFC) (Grants No. 91224008-14) from Tsinghua University.

## 7. REFERENCES

[1] Joseph, K., Landwehr, P. M., Carley, K. M. An Approach to Selecting Keywords to Track on Twitter During a Disaster, *Proceedings of the 11th International ISCRAM Conference – University Park*, Pennsylvania, USA, May 2014.

[2] Slavkovikj, V., Verstockt, S., Hoecke, S. V., et al. Review of wild fire detection using social media, *Fire Safety Journal*, 68, 109–118.2014.

[3] Peng, Y. L., Li, J. J., Xia, H., et al., The effects of food safety issues released by we media on consumers' awareness and purchasing behavior: A case study in China, *Food Policy*, 51, 44 - 52.2015.

[4] Ma, Y. F., Deng, Q., Wang, X. Z., Liu, J. Q., Zhang, H., Keyword-based Semantic Analysis of Microblog for Public Opinion Study in Online Collective Behaviors, *Web-Age Information ManagementLecture Notes in Computer Science 2014*, pp 44-55.2014.

[5] German, N., Leonie, R., Astrid, M., et al., Psychosocial functions of social media usage in a disaster situation: A multi-methodological approach, *Computers in Human Behavior*, 34, 28-38.2014.

[6] Hadiguna, R. A., Kamil, I., Delati, S., Reed, R. Implementing a web-based decision support system for disaster logistics: A case study of an evacuation location assessment for Indonesia, *International Journal of Disaster Risk Reduction*, 9, 38 - 47.2014.

[7] Campos, V., Bandeira, R., Bandeira, A., A method for evacuation route planning in disaster situations, *Procedia - Social and Behavioral Sciences*, 54, 503 - 512.2012.

[8] Ndiaye, I. A., Neron, E., Linot, A., Monmarche, N., Goerigk, M. A new model for macroscopic pedestrian evacuation planning with safety and duration criteria, *Transportation Research Procedia*, 2, 486 - 494.2014.

[9] Onorati, T., Malizia, A., Diaz, P., Aedo, I., Modeling an ontology on accessible evacuation routes for emergencies, *Expert Systems with Applications*, 41, 7124 - 7134.2014.

[10] Wilensky, H., Twitter as a Navigator for Stranded Commuters during the Great East Japan Earthquake, *Proceedings of the 11th International ISCRAM Conference – University Park*, Pennsylvania, USA, May 2014.

[11] Sun, D., Zhang, L. H., Chen, F. X., Comparative study on simulation performances of CORSIMand VISSIM for urban street network, *Simulation Modelling Practice and Theory*, 37, 18-29.2013.

[12] Amy R. Donovan, Clive Oppenheimer, Modelling risk and risking models: The diffusive boundary between science and policy in volcanic risk management, *Geoforum*, 58, 153-165.2014.

[13] RuiGaspar, Claudia Pedroc, et al., Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events, *Computers in Human Behavior*, 56, 179-191.2016.

[14] Hannes Seppänen, Kirsi Virrantaus, Shared situational awareness and information quality in disaster management, *Safety Science*, 77, 112-122.2015.

[15] Maria M. C., Neville A. S., Ioannis M. D., The concept of risk situation awareness provision: Towards a new approach for assessing the DSA about the threats and vulnerabilities of complex socio-technical systems, *Safety Science*, 79, 126-138.2015.

[16] K. Mora, J. Chang, A. Beatson, C. Morahan, Public perceptions of building seismic safety following the Canterbury earthquakes: A qualitative analysis using Twitter and focus groups, *International Journal of Disaster Risk Reduction*, 13, 1-9.2015.