Received: 15 July 2020 | Accepted: 26 July 2020

DOI: 10.1111/1468-5973.12311

ORIGINAL ARTICLE

WILEY

Using deep learning and social network analysis to understand and manage extreme flooding

Andrei Romascanu | Hannah Ker | Renee Sieber | Sarah Greenidge | Sam Lumley | Drew Bush | Stefan Morgan | Rosie Zhao | Mikael Brunila

McGill University, Montreal, QC, Canada

Correspondence
Renee Sieber, McGill University, Montreal, QC, Canada.
Email: renee.sieber@mcgill.ca

Funding information
Environment Canada, Grant/Award Number: ECCC GCXE19M010

Abstract

Combining machine learning with social network analysis (SNA) can leverage vast amounts of social media data to better respond to crises. We present a case study using Twitter data from the March 2019 Nebraska floods in the United States, which caused over $1 billion in damage in the state and widespread evacuations of residents. We use a subset of machine learning, deep learning (DL), to classify text content of 11,982 tweets, and we integrate that with SNA to understand the structure of tweet interactions. Our DL approach pre-trains our model with a DL language technique, BERT, and then trains the model using the standard training dataset to sort a dataset of tweets into classes tailored to crisis events. Several performance measures demonstrate that our two-tiered trained model improves domain adaptation and generalization across different extreme weather event types. This approach identifies the role of Twitter during the damage containment stage of the flood. Our SNA identifies accounts that function as primary sources of information on Twitter. Together, these two approaches help crisis managers filter large volumes of data and overcome challenges faced by simple statistical models and other computational techniques to provide useful information during crises like flooding.

KEYWORDS

BERT, big data, CrisisNLP, floods, Nebraska, Recurrent Neural Network, supervised classification

1 | INTRODUCTION

The public's use of social media during crisis events can generate a wealth of information valuable to crisis managers and officials. The related fields of crisis informatics and crisis analytics have emerged to understand the ways in which data from sites such as Twitter can create near real-time situational awareness within the public (e.g., Cameron, Power, Robinson, & Yin, 2012; Helsloot & Groenendaal, 2013; Qadir et al., 2016). However, large volumes of data and irrelevant content create practical challenges for crisis managers (Hiltz, Kushma, & Plotnick, 2014). Social media data is often unstructured and challenging to represent in traditional databases. Analyses that leverage social media can be slow to scale during high traffic, making it difficult to generate useful findings within the timescales of crisis events (Qadir et al., 2016; Stieglitz, Mirbabaie, Ross, & Neuberger, 2018). While such analyses can inform long-term decision-making, crisis managers may struggle to keep pace with the dynamic and evolving needs that an effective crisis response demands (Hiltz et al., 2014).

Computational approaches can potentially address the practical challenges of processing and analysing social media data in dynamic crises. Machine learning (ML), for example, can be used to categorize text or classify sentiment, enabling crisis managers to more easily filter large volumes of data (e.g., Nguyen, Joty, Imran, Sajjad,

& Mitra, 2016). Social network analysis (SNA) also can provide insight into the complex networks by which information diffuses on social media platforms, allowing crisis managers to identify key information sources in their social networks (e.g., Gupta, Joshi, & Kumaraguru, 2012).

This paper extends existing analyses of social media to support crisis management. We present a case study using Twitter data from the March 2019 Nebraska floods. This flood event caused over $1 billion in damages in Nebraska and widespread evacuations of residents (Schwartz, 2019). We collect and analyse data from the damage containment stage of the crisis (Mitroff, 1994), as past surveys have found this to be the time when microblog data is most likely to be used by crisis managers (Tapia, Bajpai, Jansen, Yen, & Giles, 2011). To analyse this data, we use a form of ML called deep learning (DL; specifically Recursive Neural Networks or RNN) to classify text content. We augment the RNN with SNA to understand the structure of interactions between accounts. We specifically explore the following two research questions. First, to what extent is an analysis that employs DL and SNA useful to crisis managers who are responding to disasters? Second, what does this analysis reveal about the informational ecosystem on Twitter during the damage containment stage of the Nebraska flood?

In addressing these questions, our primary contributions are twofold. We first show that using pre-trained language models such as Bidirectional Encoder Representations from Transformers (BERT) and word-piece tokenization can address common issues of non-standard text in tweets and improve generalization from a small amount of labelled data. This finding is supported by experimental results that show our method outperforming previous state-of-the-art models in terms of prediction accuracy. We also demonstrate the ability of pre-trained language models to be used with real-world data as well as its applicability to new types or stages of a crisis. Second, we combine DL text classification with SNA for the analysis of social media posts during a crisis. Whereas the two have been performed separately, we are among the first to integrate them in a way that can be useful to crisis managers. This integration allows us to gain insight into both the content and the structure of communications on Twitter during this stage of the crisis. Throughout our discussion of results, we remain focussed on the usefulness of our methods and insights to the work of crisis managers. In doing so, we aim to address the need for a deeper consideration of how the analysis of social media data can be applied in practice to the work of crisis management (Zade et al., 2018).

2 | RELATED WORK

2.1 | Using ML to analyse social media data

ML has been framed as an effective approach for filtering out noisy and irrelevant information that accommodates the volume and speed of real-time social media data during a crisis (Nguyen et al., 2016; Rao, Plotnick, & Hiltz, 2017). Previous research efforts have developed ML models to classify unstructured text content of messages generated on social media during crisis events (e.g., Buscaldi & Hernandez-Farias, 2015; Reynard & Shirgaokar, 2019). For example, ML has been used to classify text content according to the stages of a crisis (Verma et al., 2011) and the sentiment of the public reacting to a crisis (Beigi, Hu, Maciejewski, & Liu, 2016). Researchers also have built ML models to map content based on location markers in the text (Cresci, Cimino, Dell'Orletta, & Tesconi, 2015; Ghahremanlou, Sherchan, & Thom, 2015). Reynard and Shirgaokar (2019) used ML to assist with geolocation and sentiment classification of tweets to help guide resource allocation during a natural disaster. Nguyen, Alam, Ofli, and Imran (2017) used a type of DL called a convolution neural network (CNN) to attempt to automatically detect the level of crisis damage from images.

Despite the wealth of potential applications, it is easy to overlook the amount of effort and computational expertise required to apply ML effectively, resources that might not be easily or quickly accessible to crisis managers. As part of their work on CNN for text, Nguyen et al. (2016) described the challenges of interpreting short messages using ML because messages like tweets contain less data, are often informal and unstructured, and use slang or misspellings. Derczynski, Meesters, Bontcheva, and Maynard (2018) noted the need to meticulously hand-label 70,000 tweets prior to developing the ML model, into "informativeness" and "actionable" categories, utilizing and paying crowdworkers (70,000, the authors note, is relatively small for training ML). Indeed, effective training of ML models has been described as "a black art that requires years of experience to acquire" (Smith, 2018, p. 1), due to their complexity and long training times, as well as the as-yet still limited understanding of the different interactions between elements of a crisis. Existing research also has investigated the unique challenges that crises pose when attempting to generalize ML models across multiple events (cf., Li, Caragea, Caragea, & Herndon, 2018). ML is still far from off-the-shelf for real-time or near real-time crisis management.

Whereas Nguyen et al. (2016) provide a framework to contextualize crisis applications; they do not account for how the structures of social media affect the spread of the data. In addition, by using a publicly available dataset to both train and test their model, they also do not account for the amount of work that is required to collect a dataset (although they attempt to ameliorate this for images in Nguyen et al., 2017). The dataset also leaves out potentially important Twitter information by omitting retweets, as well as leaving out the network that is built upon repeated retweets.

Given the challenges illustrated in this section, there is a need for more work in developing ML models that perform well when applied to short, non-standard text data and that can be generalized across multiple crises from small amounts of labelled data.

2.2 | Approaches using SNA

While ML approaches can be applied to analyse the text of social media posts, they may not capture the dynamics of how text content spreads among social media users. SNA has frequently been used to investigate the interactions between social media users during crises (e.g., Chatfield & Brajawidagda, 2012; Gupta et al., 2012; Hagen, Keller, Neely, DePaula, & Robert-Cooperman, 2018; Kim, Bae, & Hastak, 2018). For example, a network of retweets and replies between Twitter accounts during the 2017 Storm Cindy in the US revealed accounts that functioned as dominant information sources (Kim et al., 2018). Distinct communities of users were found on Twitter during Hurricane Irene, the England riots, and the 2011 Virginia earthquake (Gupta et al., 2012). Such analyses reveal how information is exchanged among numerous connected actors (i.e., user accounts) as a crisis unfolds (Haythornthwaite, 1996). Crisis data from Twitter is particularly well-suited to such analyses as user interactions (retweets, mentions and replies) that can be readily accessed by researchers. Even as ML and SNA are increasingly common in the crisis informatics literature, these two approaches are rarely used in combination. As each analytical approach reveals insight into different dimensions of crisis communication (content and structure, respectively), we argue that they can be meaningfully integrated to obtain a more holistic understanding of communication on social media during a crisis.

3 | METHODOLOGY

Our methodology has two primary components. We first used a DL model to classify a dataset of tweets from the damage containment stage of the Nebraska flood according to a number of predefined categories. We improve on current state-of-the-art social media text analytics methods by addressing the salient issues of handling non-standard text, learning from limited data and effectively generalizing to new events. We then used SNA to identify key sources of information on Twitter during this stage of the crisis.

We first outline the architecture of our model, which we used for supervised classification of large volumes of unstructured text into themes (Derczynski et al., 2018). We describe the datasets of labelled tweets we used to train this model, as well as the dataset we collected during the Nebraska flooding. We then detail our approach to SNA and highlight how we incorporated the results that we obtained from our ML methods. Our code is publicly available at [https://github.com/smacawi/].

1 shows the workflow of our model development and use.

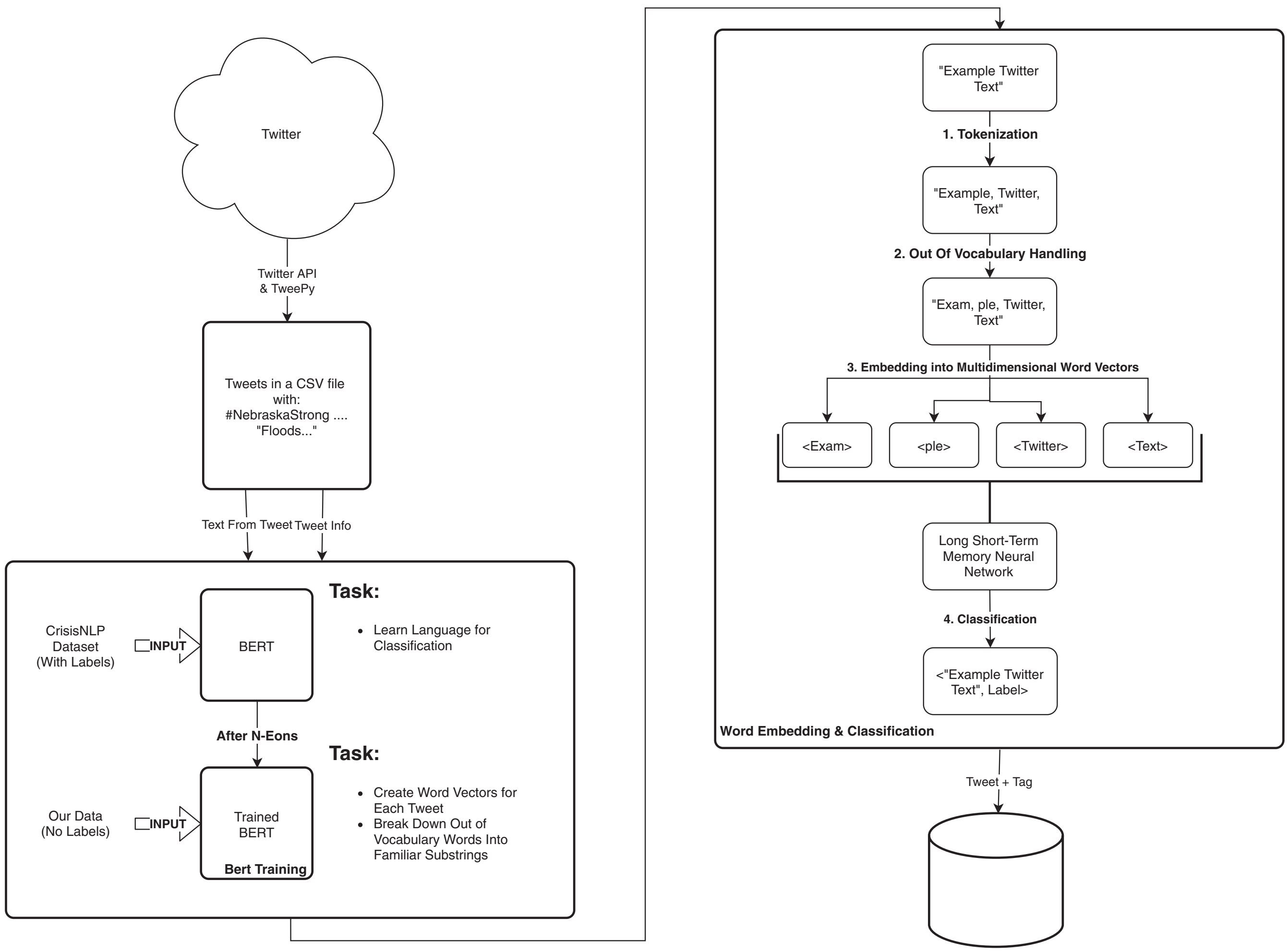

FIGURE 1 Deep learning workflow diagram

3.1 | Training the DL model

We chose to build our supervised classification model atop BERT, a DL model pre-trained on large amounts of text (Devlin, Chang, Lee, & Toutanova, 2019). This semi-supervised pre-training allows BERT to learn generalizable structural information about language (Jawahar, Sagot, & Seddah, 2019) so it can be fine-tuned on a specific task with less labelled data and better generalize to new datasets (e.g., fine-tuning for tweet classification on a previous crisis event and generalizing to a new event). Another advantage of BERT over other models is that it uses word-piece tokenization, which can further break down words into "word pieces." This solves the issue of out-of-vocabulary words commonly encountered with non-standard text in social media and which had previously been solved with less generalized text preprocessing and rule-based approaches (e.g., Imran, Mitra, & Castillo, 2016).

We trained models with a dataset of tweets from crises labelled and published by CrisisNLP (Imran et al., 2016). We used a subset of the CrisisNLP dataset with approximately 11,038 crowdflower-annotated tweets in English from 19 different crises that took place between 2013 and 2015. Tweets were classified according to the categories identified in Table 1. This essentially results in a two-tiered training approach to crisis modelling.

Kumar, Singh, and Saumya (2019) performed a comparative analysis of models to find state-of-the-art crisis-related tweet classifiers. To evaluate the performance of our model, we compared our approach with the performance measures of precision, recall, F1 scores and accuracy on a withheld test set of the 2014 Pakistan floods. We also performed hyperparameter tuning on our model to find the optimal batch size, optimizer and architecture. We used Adam, a standard adaptive momentum optimizer (Devlin et al., 2019). Adam is useful because it performs well without tuning. The following architectures were tested: BERT with a final dense layer applied to the first hidden state, BERT with an LSTM layer applied to every hidden state, as well as baseline models (LSTM + Crisis, LSTM + Glove) from Kumar et al. (2019). These analyses were then performed on three variations of training data to evaluate transferability to new crises. The first variation trains on data from the same event; the second variation trains on data from a similar event (the 2014 India floods); and the last variation trains on data from all events in the CrisisNLP dataset.

Table 2 reports validation for each model tested with optimal hyperparameters (the values of BERT/Dense are bolded to highlight that this configuration performed the best). Both BERT models were found to perform significantly better than the baseline models across all three training sets. In particular, the implementation of BERT with a final dense layer reported the highest accuracy of 77% when training on all events. A drop in performance was noted when we trained on a similar event as opposed to the same event. When we trained on data from all events, the scores improved past those reported for models trained on the same event. This speaks to the potential of the model in generalizing extreme weather event data across different crises and the utility of increasing the volume of data to improve classification accuracy.

3.2 | Classifying tweets from the Nebraska flood

Once our model was trained on the labelled CrisisNLP dataset, we collected an unlabelled dataset of tweets by using the Twitter API and filtering according to hashtags deemed relevant to the event:

- #nebraskaflood
- #flood2019
- #nebraskastrong
- #missouririver
- #nebraskaflood2019
- #prayfornebraska

TABLE 1 Categories for crisis-related tweets from CrisisNLP (Imran et al., 2016, p. 3)

Category	Description
Injured or dead people	Reports of casualties and/or injured people due to the crisis
Missing, trapped, or found people	Reports and/or questions about missing or found people
Displaced people and evacuations	People who have relocated due to the crisis, even for a short time (includes evacuations)
Infrastructure and utilities damage	Reports of damaged buildings, roads, bridges, or utilities/services interrupted or restored
Donation needs or offers or volunteering services	Reports of urgent needs or donations of shelter and/or supplies such as food, water, clothing, money, medical supplies or blood; and volunteering services
Caution and advice	Reports of warnings issued or lifted, guidance and tips
Sympathy and emotional support	Prayers, thoughts, and emotional support
Other useful information	Other useful information that helps one understand the situation
Not related or irrelevant	Unrelated to the situation or irrelevant

TABLE 2 Precision, Recall, F1 score, and validation accuracy results for our various deep learning models

Training set	Model	Precision	Recall	F1	Accuracy
All events	BERT/Dense	**0.77**	**0.77**	**0.76**	**0.77**
	BERT/LSTM	0.73	0.72	0.70	0.72
	LSTM + Glove	0.63	0.59	0.59	0.59
	LSTM + Crisis	0.63	0.59	0.57	0.59
Similar event	BERT/Dense	**0.60**	**0.62**	**0.59**	**0.62**
	BERT/LSTM	0.42	0.48	0.44	0.48
	LSTM + Glove	0.22	0.37	0.27	0.37
	LSTM + Crisis	0.47	0.43	0.39	0.43
Same event	BERT/Dense	**0.73**	**0.74**	**0.72**	**0.74**
	BERT/LSTM	0.65	0.65	0.61	0.65
	LSTM + GloVe	0.57	0.61	0.57	0.61
	LSTM + Crisis	0.61	0.62	0.60	0.62

Tweets from this dataset were input to the model to be automatically classified along the categories from Table 1. To validate our model, we sampled a representative subset (class-balanced and with no retweets) of the classified tweets and manually labelled them, showing the model had an accuracy of 76% on this data, comparable to our tests on the CrisisNLP dataset.

3.3 | Conducting SNA

Once our dataset of tweets from the Nebraska flood event was labelled according to the categories identified in Table 1, we conducted a SNA to identify the Twitter accounts within our dataset that functioned as key sources of information on Twitter during our period of data collection.

We converted our Twitter dataset into a network structure of nodes and edges to model flows of information between Twitter users during the Nebraska flood. One user account corresponds to one node in the network. The edges connecting these nodes are informational exchanges between accounts, as represented by tweets that are mentions, retweets and replies. This approach to modelling a social network of information flow using Twitter data follows past work on crises (Chatfield & Brajawidagda, 2012; Gupta et al., 2012; Kim et al., 2018; Willis, Fisher, & Lvov, 2015). In creating this network, we removed all tweets that were classified by our ML model as “Not related or irrelevant.”

We identified influential Twitter accounts in this network using several measures of node centrality: degree centrality, betweenness centrality and PageRank score. Degree centrality measures the number of “one-hop” connections (edges) between nodes in a network and betweenness centrality measures the number of times that a node lies on the shortest path between two other nodes. Nodes with high betweenness centrality may be considered “bridges” between communities in a network. PageRank is used by the Google Search Engine to rank the importance of websites (Page, Brin, Motwani, & Winograd, 1999) and has been applied to Twitter networks to identify central nodes (Willis et al., 2015). Past research efforts do not use a consistent centrality measure when identifying the most central nodes in the network. We applied all three measures and looked for nodes that ranked highly according to multiple measures. The use of multiple measures of node centrality to identify influential Twitter accounts follows past work in this domain (Hagen et al., 2018).

We identified the top ten nodes with the highest values for each centrality measure. Nodes that appeared in at least two out of the three top ten rankings were considered to be among the most central nodes in the network.

4 | RESULTS

From March 22 to March 26, 2019, we collected a total of 11,982 tweets relating to the damage containment stage (Mitroff, 1994) of the Nebraska flood. Of these tweets, 1,055 are regular tweets, 9,658 are retweets, 1,239 are mentions, and 30 are replies.

4.1 | Classifying the tweets

Figure 2 shows the number of tweets that were classified by our model into nine categories, as specified in Table 1. We found tweets categorized as “Donations” occurred 5.75 times as much as the first four categories combined. The volume of tweets in the “Sympathy” category also exceeded the first four combined by 1.2 times. The shift in the stage from the immediacy of the event to its aftermath likely explains the popularity of these categories. Even when we consider the DL model's ability to categorize into nine categories, the existing categories provide insufficient guidance. The “Other useful information” category is the largest category of all (40% greater than “Donations”).

The vast majority of tweets in our dataset are retweets, supporting the findings of Starbird, Palen, Hughes, and Vieweg (2010), who characterize retweet activity during crises as a means for those affected to spread information that they believe to be valuable and trustworthy. When retweets are removed from the dataset, the “Donations” category supersedes the “Other useful information” category and outflanks

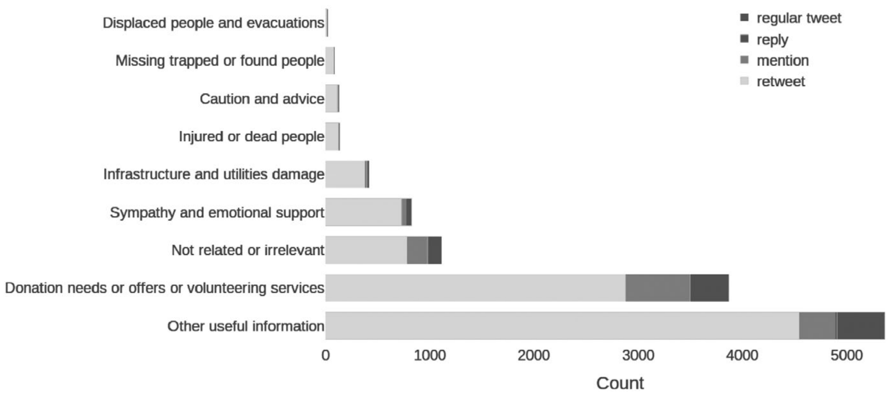

FIGURE 2 The number of tweets assigned to each category

the first four categories by more than 14 times. "Other useful information" comprises almost half (41%) of the unique tweets.

Table 3 shows examples of tweets that were classified into each category. A tweet with the text, "How to help those impacted by the floods?", exemplifies how Twitter has become a medium for crisis communication. "Moderate flooding going on here," classified as "Caution and advice," exemplifies the success of our model's classification. An example from the "Sympathy" category shows how classification models need to include emojis and be able to capture emergent folk taxonomies or "folksonomies" (e.g., #fbproud, meaning Nebraska Farm Bureau proud, NEFB).

Due to biases in our training data, certain classes are under-represented, leading to measurable variances in our model's confidence for each class. These are shown in Figure 3 and can play a valuable role in a crisis manager's interpretation of our model's outputs, helping address the black-box issue of ML. To the best of our knowledge, we are the first to report such metrics in this context.

4.2 | Twitter accounts functioning as key information sources

We constructed a social network of retweets, replies and mentions that had a total of 7,583 nodes and 8,832 edges. The accounts that ranked highly according to our three measures of node centrality (degree, betweenness and PageRank) primarily correspond to official information sources, such as government officials and institutions and news figures. For example, the official Twitter account for Nebraska's Governor Pete Ricketts ranks highly in two of the three centrality measures, as well as accounts for the Nebraska Emergency Management Agency (NEMA) and the Nebraska State Patrol. One influential account is an official account associated with an Omaha reporter.

4.3 | Characterizing tweet content from key information sources

The results of the previous two sections can be combined to offer insight into the characteristics of the tweets published by accounts that functioned as key information sources during the Nebraska flood. We selected two accounts that ranked highly in the measures of node centrality and can thus be considered official information sources: State Governor Pete Ricketts and the Nebraska Emergency Management Agency (NEMA). Figures 4 and 5 show the distribution in categories of the tweets from each account. As shown in Figure 4, tweets from Governor Ricketts primarily related to donations and volunteering services. A large number of his tweets were classified as "Other useful information." Figure 5 shows how many of the tweets from NEMA related to donations and other useful issues. Notably, many of the tweets from NEMA were classified according to the "Sympathy and emotional support" categories, which was present to a much lesser extent in the tweets from Governor Ricketts. The inclusion of retweets in Figures 4 and 5 shows categories of the tweets that gained more visibility through higher retweet volumes. Overall, the SNA pointed to the most influential Twitter accounts during the time period in question. The results revealed subtle distinctions between the type of content generated (through unique posts or retweets) from highly influential posters, suggesting different roles that Twitter accounts played in spreading information during the flood.

5 | DISCUSSION

We developed a DL model, not only for situational awareness of injuries and infrastructure damage, but also for mobilizing volunteering, aid and donation efforts. For any stage of a crisis, crisis managers can use the overall distribution of categorized tweets (Figure 2) for an initial sense of the affected population's needs during that stage. Crisis managers also can use the tweet classifications to search for specific updates more effectively. For example, should a crisis manager need up-to-date information about the status of a road network within a crisis-affected area, they could target their search by manually examining tweets that were assigned to the "Infrastructure and utilities damage" category. In our case study of the Nebraska flood, the relatively large number of tweets relating to requests for donations and volunteering services suggests a primary need to rebuild and recover.

Our combination of DL with SNA allows for a more targeted understanding of tweet content during the flooding. Surveys with crisis

TABLE 3 Examples of tweets classified in each category

Injured or dead people		
@[REDACTED] @[REDACTED] It's so bad, it's unbelievable. Three people have died, so many livestock, wild animals, countless people have lost everything & are misplaced, & roads & water systems have been devastated #NebraskaFloods #NebraskaStrong #PrayForNebraska	Worst flooding damage in our state's history—As many as a million calves dead in Nebraska, AT LEAST SOMEBODY IS A HAPPY CAMPER #AOC #GreenNewDeal https://t.co/cDTUDPWLH2 #NebraskaFlood #Nebraskaflooding https://t.co/OnM7x1Zwmc	Flooding in the midwest has already caused three deaths and $3 billion in damage. These events, once unthinkable, are now commonplace We face a simple choice: decarbonise now, or watch the suffering increase every day. #ClimateCrisis #NebraskaStrong https://t.co/iFeNGS4AEa
Missing, trapped, or found people		
Search turns up no signs of missing Norfolk man #NebraskaFlood2019 https://t.co/sgvhTe7Nvk	A needed smile: It's #NationalPuppyDay, so here is a dog rescued with his family last week by a #Blackhawk helicopter crew. The #NEGuard helped rescue 111 people &; 13 pets during the #NebraskaFlood #NebraskaStrong https://t.co/tdyJkgSv9x	Went to pick up what was left at grandmas yesterday, this poor girl also managed to get stuck in the river behind her house. Saved her tho! #nebraskastrong #nebraskaflood2019 @ [REDACTED] https://t.co/n5kDdKMfBG
Displaced people and evacuations		
Chaotic evacuations with emergency sirens blaring as the #MissouriRiver rises to the top of the three-story-high levee wall in St. Joseph, Missouri! https://t.co/BCUcxXrF05 via @[REDACTED]	It made the hair on my neck stand up going by #Elwood where cops are at the exits to keep people from getting into town after it's been evacuated. Helicopters are flying around. It's insane. The town is just empty. #Flood2019 #missourifloods https://t.co/xJnJu3n71k	They've sounded the alarm. Mandatory evacuation for Elwood Kansas. Please pray for them #Flood2019 https://t.co/gDLvJIwBeo
Infrastructure and utilities damage		
Surveyed flood damage in Plattsmouth and many other southeast Nebraska communities along the Missouri River which is so high it's almost impossible to see across to the Iowa/Missouri banks in some places #NebraskaFlood | #NebraskaStrong https://t.co/XPsTBM37or	There hasn't been much national coverage but, Nebraska is flooded & lives are completely devastated & even lost. #NebraskaStrong	Busy weekend for @DouglasCountyNE road crews as they rebuild washed out roads and shoulders. Please go slow and give them room. #NebraskaSTRONG #NebraskaFlood https://t.co/kPGx1PTgsA
Donation needs or offers or volunteering services		
Thank you to @[REDACTED] and all the people from across the country providing hay for flood relief! #NebraskaFlood | #NebraskaStrong https://t.co/kjGtG3a6hR	@[REDACTED] Thank you to all the volunteers in all the communities that are taking time to give back A truly amazing Volunteer is: Selfless Generous Helpful Thoughtful Valuable Patient Kind Giving #volunteers #NebraskaStrong #loveinaction #k9comfortdogs #Flooding2019 https://t.co/nYKJB0y0VL	How to help those impacted by the floods? Drop off supplies and donations at any @[REDACTED] location. Hang in there, everyone. We'll get through this together. #nebraskastrong https://t.co/Sx1w7vccVV

(Continues)

TABLE 3 (Continued)

Caution and advice		
We decided to plot all flood warnings since March 14th. Over 35,000 square miles and 5.2 million people covered! #Flood2019 https://t.co/q7axID5fOi	#Nebraska, the Flood Warning continues for #MissouriRiver areas near #Blair affecting #Harrison and #WashingtonCo. #NebraskaFlood #Flood2019 https://t.co/fJdm0wf4kO	Moderate flooding going on here. Worried about the people in the bottoms. Hopefully it will crest soon. #flood2019
Sympathy and emotional support		
This is incredible! $325,000 and counting! So many #NebraskansHelpingNebraskans. #NebraskaStrong #NebraskaStrongDay https://t.co/qhv8g2BMLZ	This is truly devastating. A true natural emergency. Stay strong. My prayers are with you all until we find out how we can help. 💔🌧💙🌧💔 #Flooding2019 #Flood19 #Flood2019 https://t.co/v02lpvsDGT	Our thoughts and prayers are with those in Nebraska. As we get the latest updates of videos, pictures, and stories, our hearts break, but we also know Nebraska will get past this. #NebraskaStrong https://t.co/gx5RH4GfPB
Other useful information		
Farmers affected by #flooding look ahead to planting #Nebraskastrong	After a flood, what happens to local #realestate markets? #Flood2019	If you or someone you know is being affected by the flood, counseling and information services are available. #NebraskaFlood | #NebraskaStrong
Not related or irrelevant		
Lunch successfully served! NEFB and @[REDACTED] were in Verdigre today with burgers, hotdogs, water and more! We also dropped off a donation of supplies for the recovery effort. We're honored to help our communities after historic floods. #nebraskastrong #fbproud https://t.co/uBD2XThpl4	I really enjoyed my visit to the Black Hills Unity Concert in 2017. People have a sense of humor about themselves that's very Midwestern #lakota #missouririver #greatplains https://t.co/uXCAkiATSU	Happy Spring! New calf born on the island-Our cattle are on a high spot next to the river. We can go feed them everyday by boat, hopefully will be able to rescue them soon. 🙏 #Flood2019 #FarmFamily #WeCanMakeIt #PrayForAll https://t.co/V31GDbriX9

Note: Mentions of Twitter usernames have been redacted out of respect for user privacy. The usernames of verified accounts and official government accounts have been preserved.

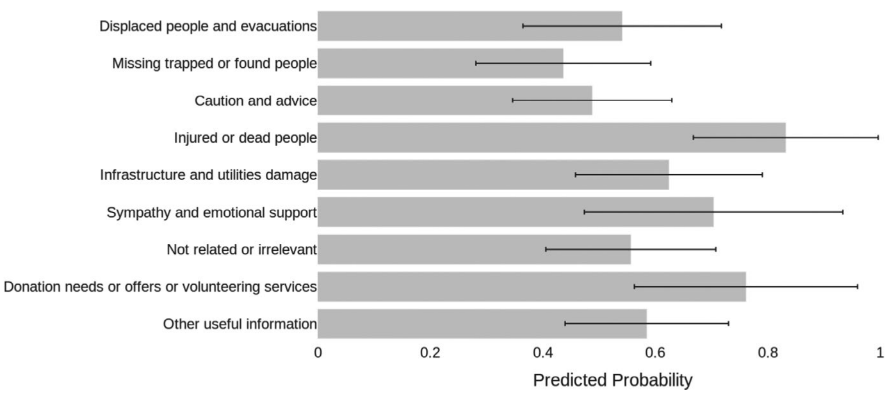

FIGURE 3 Mean probability values for categories. Lines indicate standard error values for each category

managers have found that a lack of trust in information garnered from social media can pose a significant barrier to using this information in practice (Hiltz et al., 2014; Rao et al., 2017). Combining DL with SNA could allow crisis managers to better filter through the results of categorized information to locate information coming from trusted sources. Our approach contributes to efforts to make social media more actionable and more manageable, by allowing managers to target content for manual inspection.

What the model fails to reveal is as important as what it does reveal. Limitations in the training dataset, CrisisNLP, results in some of this opacity. Unsurprisingly, CrisisNLP is not tailored to the damage containment phase that was the focus of this work. The number of training data for each category varied from 402 to 2,610 tweets. Accuracy can be impacted by the volume of tweets used to train the model to recognize a particular category. Training data also impacts occurrences unique to a particular crisis. During our collection phase, a relief concert trended at 20 to 40 retweets, references and replies per hour (Figure 6). The concert, its celebrity as well as the presence of the governor, skewed the distribution of content. These retweets suggest that there will be challenges in transferring the model to other crisis events.

The large volume of tweets classified as “Other useful information” demonstrates gaps as well as opportunities for expanding categories. In Table 2, we illustrate this with examples related to farming

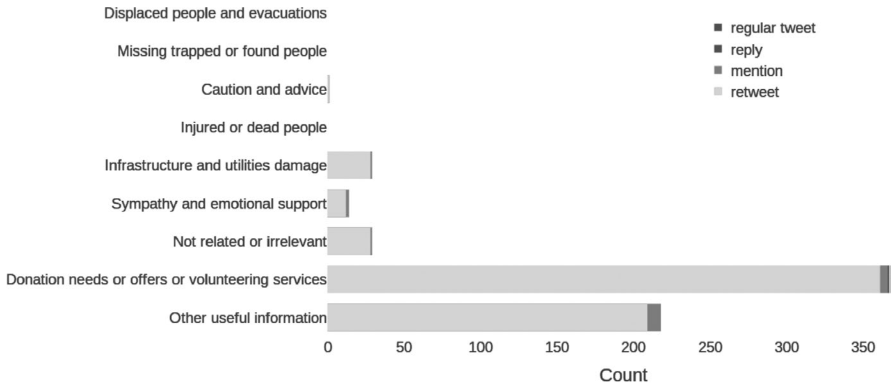

FIGURE 4 The number of tweets from Governor Ricketts assigned to each category

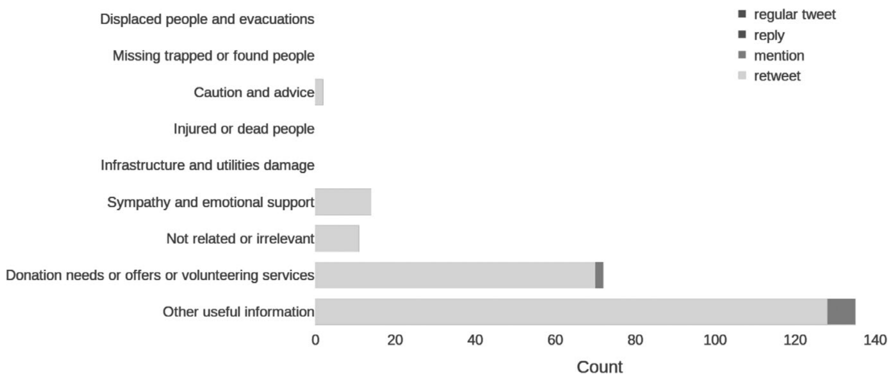

FIGURE 5 The number of tweets from NEMA assigned to each category, including retweets

FIGURE 6 A tweet that trended during the 2019 Nebraska flood (Source: https://twitter.com/jtimberlake/status/1109857242118086656) [Colour figure can be viewed at wileyonlinelibrary.com]

and real estate, which suggest a relevant additional category related to planning and recovery. Consider the following example in Table 2: "We face a simple choice: decarbonise now, or watch the suffering increase every day. #ClimateCrisis," which suggests a useful addition could be the connection of acute events to broader causations. Work remains on creating a robust standard for training data that supports the various stages of a crisis as well as explicates the relevant content not captured by the nine categories. That being said, robust training data is costly and time-consuming to develop; it requires a large volume of input data and a certain degree of expertise (e.g., to train workers to accurately label content).

Finally, our model exposes the trade-off between conducting real-time analysis and recognizing emergent topics critical to a particular crisis. Topic modelling is an alternate method that allows categories to emerge unsupervised from the data. Topic modelling works by detecting co-occurring groups of words—latent structure—in large collections of text. The insights afforded by this unsupervised learning demands substantial data cleaning and human interpretation to make sense of those categories. Additionally one needs to wait until a large corpus of crisis content is collected. Predetermined categories provide a degree of automation and speed at the cost of what might be unique to the crisis. Semi-supervised topic modelling might offer a solution that leverages the benefits of unsupervised and supervised classification (Kuhn, 2018).

6 | CONCLUSION

We developed a combined DL/SNA approach, based on free and open-source analytical tools for analysing Tweets from the 2019 Nebraska flood to aid in crisis management. This approach extends the existing analytical toolbox in the field of crisis informatics by augmenting DL-based text analysis with a SNA, providing deeper insight into the network structure of communication between accounts on Twitter. Furthermore, leveraging pre-trained BERT allows our model to better generalize to new crises with fewer training examples, automatically handle out-of-vocabulary words, and more easily predict multiple categories with one model. This contrasts with other approaches that require complex methods necessary to handle limited training data or out-of-vocabulary words, as well as approaches that rely on one binary classifier per category (Imran et al., 2016).

DL is a method that is still in its very early stages of usability by individuals who are not computer scientists or software engineers.

We argue that a model that reveals its benefits and costs assists in trust of ML, its refinements like RNN and the output of those models.

Our approach can help crisis managers reduce the information overload associated with social media, enabling the timely extraction of critical information during crises. In future, we will test our model on other floods and investigate its transfer capacity to other extreme events (e.g., snowstorms, freezing rain). We hope that tests of transferability will continue discussions on the challenges of ML and the opportunities for its combination with other methods like SNA for crisis management.

ACKNOWLEDGEMENTS

This research was funded by a grant and contract from Environment and Climate Change Canada, Social Media Crowdsourcing Assessment of Weather Impacts.

REFERENCES

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. In W. Pedrycz, & S.-M. Chen (Eds.), *Sentiment analysis and ontology engineering: An environment of computational intelligence* (pp. 313–340). Springer. doi: https://doi.org/10.1007/978-3-319-30319-2_13

Buscaldi, D., & Hernandez-Farias, I. (2015). Sentiment analysis on microblogs for natural disasters management: A study on the 2014 Genoa floodings. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1185–1188). doi: https://doi.org/10.1145/2740908.2741727

Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). Emergency situation awareness from Twitter for crisis management. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 695–698). doi: https://doi.org/10.1145/2187980.2188183

Chatfield, A., & Brajawidagda, U. (2012). Twitter tsunami early warning network: A social network analysis of Twitter information flows. In J. W. Lamp (Ed.), *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian conference on information systems* (pp. 1–10). Geelong, Australia: Deakin University.

Cresci, S., Cimino, A., Dell'Orletta, F., & Tesconi, M. (2015). Crisis mapping during natural disasters via text analysis of social media messages. In J. Wang, W. Cellary, D. Wang, H. Wang, S.-C. Chen, T. Li, & Y. Zhang (Eds.), *Web information systems engineering* (pp. 250–258). Cham, Switzerland: Springer.

Derczynski, L., Meesters, K., Bontcheva, K., & Maynard, D. (2018). Helping crisis responders find the informative needle in the tweet haystack. In K. Boersma, & B. M. Tomaszewski (Eds.), *Proceedings of the 15th international conference on information systems for crisis response and management* (pp. 649–662). Rochester, NY: Rochester Institute of Technology.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of North American chapter of the association for computational linguistics: Human language technologies 2019* (pp. 4171–4186). Minneapolis, USA: Association for Computational Linguistics.

Ghahremanlou, L., Sherchan, W., & Thom, J. A. (2015). Geotagging Twitter messages in crisis management. *The Computer Journal*, *58*(9), 1937–1954. https://doi.org/10.1093/comjnl/bxu034

Gupta, A., Joshi, A., & Kumaraguru, P. (2012). Identifying and characterizing user communities on Twitter during crisis events. In *Proceedings of the 2012 workshop on data-driven user behavioral modelling and mining from social media* (pp. 23–26). doi: https://doi.org/10.1145/2390131.2390142

Hagen, L., Keller, T., Neely, S., DePaula, N., & Robert-Cooperman, C. (2018). Crisis communications in the age of social media: A network analysis of Zika-related tweets. *Social Science Computer Review*, *36*(5), 523–541. https://doi.org/10.1177/0894439317721985

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research*, *18*(4), 323–342. https://doi.org/10.1016/S0740-8188(96)90003-1

Helsloot, I., & Groenendaal, J. (2013). Twitter: An underutilized potential during sudden crises? *Journal of Contingencies and Crisis Management*, *21*(3), 178–183. https://doi.org/10.1111/1468-5973.12023

Hiltz, S. R., Kushma, J., & Plotnick, L. (2014). Use of Social Media by U.S. Public Sector Emergency Managers: Barriers and wish lists. In S. R. Hiltz, L. Plotnick, M. Pfaf, & P. C. Shih (Eds.), *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management* (pp. 602–611). University Park, Pennsylvania: The Pennsylvania State University. doi: https://doi.org/10.13140/2.1.3122.4005

Imran, M., Mitra, P., & Castillo, C. (2016). Twitter as a lifeline: Human-annotated Twitter Corpora for NLP of crisis-related messages. In N. Calzolari, K. Choukri, H. Mazo, A. Moreno, T. Declerck, S. Goggi, M. Grobelnik, J. Odijk, S. Piperidis, B. Maegaard, & J. Mariani (Eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation* (pp. 1638–1643). European Language Resources Association.

Jawahar, G., Sagot, B., & Seddah, D. (2019). What does BERT learn about the structure of language? In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3651–3657). Association for Computational Linguistics. doi: https://doi.org/10.18653/v1/P19-1356

Kim, J., Bae, J., & Hastak, M. (2018). Emergency information diffusion on online social media during storm Cindy in U.S. *International Journal of Information Management*, *40*, 153–165. https://doi.org/10.1016/j.ijinfomgt.2018.02.003

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, *87*, 105–122. https://doi.org/10.1016/j.trc.2017.12.018

Kumar, A., Singh, J., & Saumya, S. (2019). A comparative analysis of machine learning techniques for disaster-related tweet classification. In *Proceedings of Humanitarian Technology Conference, IEEE Region 10* (pp. 222–227). IEEE. doi: https://doi.org/10.1109/R10-HTC47129.2019.9042443

Li, H., Caragea, D., Caragea, C., & Herndon, N. (2018). Disaster response aided by tweet classification with a domain adaptation approach. *Journal of Contingencies and Crisis Management*, *26*(1), 16–27. https://doi.org/10.1111/1468-5973.12194

Mitroff, I. I. (1994). Crisis management and environmentalism: A natural fit. *California Management Review*, *36*(2), 101–113. https://doi.org/10.2307/41165747

Nguyen, D., Alam, F., Ofli, F., & Imran, M. (2017). Automatic image filtering on social networks using deep learning and perceptual hashing during crises. In T. Comes, F. Bénaben, C. Hanachi, M. Lauras, & A. Montarnal (Eds.), *Proceedings of the 14th International conference on information systems for crisis response and management* (pp. 499–511). ISCRAM.

Nguyen, D. T., Joty, S., Imran, M., Sajjad, H., & Mitra, P. (2016). *Applications of online deep learning for crisis response using social media information*. [Conference presentation and paper]. Fourth International Workshop on Social Web for Disaster management Indianapolis, IN, USA

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford, CA: Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/

Qadir, J., Ali, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: Big data-driven crisis response. *Journal of International Humanitarian Action*, *1*(1), Article 12. https://doi.org/10.1186/s41018-016-0013-9

Rao, R., Plotnick, L., & Hiltz, R. (2017). Supporting the use of social media by emergency managers: Software tools to overcome information overload. In T. Bui, & T. R. Sprague Jr. (Eds.), *Proceedings of the 50th Hawaii international conference on system sciences* (pp. 304–312). Manoa, Honolulu, Hawaii: University of Hawaii.

Reynard, D., & Shirgaokar, M. (2019). Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? *Transportation Research Part D: Transport and Environment*, *77*, 449–463. https://doi.org/10.1016/j.trd.2019.03.002

Schwartz, M. (2019). *Nebraska faces over $1.3 billion in flood losses*. NPR [radio Broadcast]. National Public Radio. https://www.npr.org/2019/03/21/705408364/nebraska-faces-over-1-3-billion-in-flood-losses

Smith, L. N. (2018). *A disciplined approach to neural network hyper-parameters: Part 1—learning rate, batch size, momentum, and weight decay (NRL Technical Report 5510-026)*. Washington, D.C.: US Naval Research Laboratory.

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on the red: What hazards threat reveals about the social life of microblogged information. In S. Whittaker, & E. F. Churchill (Eds.), *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 241–250). ACM. https://doi.org/10.1145/1718918.1718965

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics—Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Tapia, A. H., Bajpai, K., Jansen, J., Yen, J., & Giles, L. (2011). Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations. In M. A. Santos, L. Sousa, & E. Portela (Eds.), *8th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2011* (pp. 1–10). Lisbon: ISCRAM.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., … Anderson, K. M. (2011). Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 385–391). Barcelona: Association for the Advancement of Artificial Intelligence.

Willis, A., Fisher, A., & Lvov, I. (2015). Mapping networks of influence: Tracking Twitter conversations through time and space. *Participations: Journal of Audience & Reception Studies*, *12*(1), 494–530.

Zade, H., Shah, K., Rangarajan, V., Kshirsagar, P., Imran, M., & Starbird, K. (2018). From situational awareness to actionability: Towards improving the utility of social media data for crisis response. *Proceedings of the ACM on Human-Computer Interactions*, *2*, 1–18. https://doi.org/10.1145/3274464

How to cite this article: Romascanu A, Ker H, Sieber R, et al. Using deep learning and social network analysis to understand and manage extreme flooding. *J Contingencies and Crisis Management*. 2020;28:251–261. https://doi.org/10.1111/1468-5973.12311