9 July 2011

Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods

France Cheong
RMIT University, france.cheong@rmit.edu.au

Christopher Cheong
RMIT University, christopher.cheong@rmit.edu.au

ISBN: [978-1-86435-644-1]; Full paper

SOCIAL MEDIA DATA MINING: A SOCIAL NETWORK ANALYSIS OF TWEETS DURING THE AUSTRALIAN 2010-2011 FLOODS

France Cheong, School of Business IT and Logistics, RMIT University, Melbourne 3000, Victoria, Australia, france.cheong@rmit.edu.au

Christopher Cheong, School of Business IT and Logistics, RMIT University, Melbourne 3000, Victoria, Australia, christopher.cheong@rmit.edu.au

Abstract

Using tweets extracted from Twitter during the Australian 2010-2011 floods, social network analysis techniques were used to generate and analyse the online networks that emerged at that time. The aim was to develop an understanding of the online communities for the Queensland, New South Wales and Victorian floods in order to identify active players and their effectiveness in disseminating critical information. A secondary goal was to identify important online resources disseminated by these communities. Important and effective players during the Queensland floods were found to be: local authorities (mainly the Queensland Police Services), political personalities (Queensland Premier, Prime Minister, Opposition Leader, Member of Parliament), social media volunteers, traditional media reporters, and people from not-for-profit, humanitarian, and community associations. A range of important resources were identified during the Queensland flood; however, they appeared to be of a more general information nature rather than vital information and updates on the disaster. Unlike Queensland, there was no evidence of Twitter activity from the part of local authorities and the government in the New South Wales and Victorian floods. Furthermore, the level of Twitter activity during the NSW floods was almost nil. Most of the active players during the NSW and Victorian floods were volunteers who were active during the Queensland floods. Given the positive results obtained by the active involvement of the local authorities and government officials in Queensland, and the increasing adoption of Twitter in other parts of the world for emergency situations, it seems reasonable to push for greater adoption of Twitter from local and federal authorities Australia-wide during periods of mass emergencies.

Keywords: social network analysis, text mining, social media, mass emergencies.

1 INTRODUCTION

Australia experienced its worst flooding disasters in 2010 and 2011 with a series of floods occurring in several states between March 2010 and February 2011. These floods are considered the worst flood events in the history of the various states affected. First, there were the Queensland floods of March 2010 causing major flooding in south western and central Queensland (The Courier-Mail, 2010), followed by the Victorian floods of September 2010 damaging major regional towns such as Ballarat and Benalla (ABC News, 2010). Next, there was the flooding of the River Gascoyne in Western Australia in December 2010 damaging homes in Carnavon (Australian Bureau of Meteorology, 2010), the December 2010-January 2011 floods of Queensland causing three-quarters of the state to be declared a disaster zone (Brisbane Times, 2011), and ending with the New South Wales of January 2011 (Nine News) and the Victorian floods of February 2011 (The Age, 2011).

In times of mass emergencies, a phenomenon known as collective behaviour becomes apparent (Dynes & Quarantelli, 1968). It consists of socio-behaviours that include intensified information search and information contagion (Starbird, Palen, Hughes, & Vieweg, 2010). In these situations, people want to know where exactly their families and friends are as not being able to reach them or knowing they might not be able to contact you can be very frightening moments during these situations. Information is critical during emergencies as the availability of immediate information can save lives. People share information about approaching threats, where to evacuate, where to go for help, etc. Not only do they want to know about the destruction that has occurred, but they are also eager to help those affected by giving a helping hand and raise funds from donations. Thus, there is a need to keep abreast of the latest developments, however, this is difficult since information produced under crisis situations is usually scattered and of varying quality.

Social media is media used for social interaction. They are enabled by communication technologies such as the web and smartphones and they turn communication into an interactive dialogue (Wikipedia, 2011). Interactions on social media being highly distributed, decentralised and occurring in real time, they provide the necessary breadth and immediacy of information required in times of emergencies (Palen & Vieweg, 2008). Since social media offer a uniquely rapid and powerful way to disseminate information, accurate and inaccurate, good and bad spread equally alike as incorrect information can spread like wild fire. However, there is indication that social networks tend to favour valid information over rumours (Castillo, Mendoza, & Poblete, 2011).

Twitter and Facebook are good examples of social media useful in crisis situations since they provide vital information as they are happening. Twitter is a micro-blogging service, a form of lightweight chat allowing users to post and exchange short 140-character-long messages known as *tweets*. Although most tweets are conversation and chatter, they are also used to share relevant information and report news (Castillo, et al., 2011). Twitter is becoming a valuable tool in disaster and emergency situations as there is increasing evidence that it is not just a social network, it is also a news service (Yates & Paquette, 2011). In emergency situations, tweets provide either first-person observations or bring relevant knowledge from external sources (Vieweg, 2010). Information from official and reputable sources is regarded as valuable and hence is actively sought and propagated. Other users then elaborate and synthesize this pool of information to produce derived interpretations.

During the Mumbai terror attacks of 2008, online users voluntarily created a Twitter page (http://www.twitter.com/Mumbai) to update and share situational information on the attacks (Oh, Agrawal, & Rao, 2010). A study found that 52.6% of tweeted H1NI-related material in 2009 to be related to news and information on swine flu (Chew, Eysenbach, & Sampson, 2010). Twitter was used to provide time-critical information about tsunami alerts, missing and deceased people, availability of services, road conditions among other topics related to the catastrophe hours after the Chile earthquake of Santiago in 2010 (Mendoza, Poblete, & Castillo, 2010). During the Haiti earthquake, Twitter was used to create awareness about the disaster and mobilize people to help (Yates & Paquette, 2011).

Social Network Analysis (SNA) is a sociological approach for analysing patterns of relationships and interactions between social actors in order to discover underlying social structure such as: central nodes that act as hubs, leaders or gatekeepers; highly connected groups; and patterns of interactions between groups (Wasserman & Faust, 1994). SNA has been used to study social interaction in a wide range of domains. Examples include: collaboration networks (Newman, 2001), directors of companies (Davis & Greve, 1997; Davis, Yoo, & Baker, 2003), organisational behaviour (Borgatti & Foster, 2003), inter-organisational relations (Stuart, 1998), computer-mediated communications (Garton, Haythornthwaite, & Wellman, 1999), and many others.

In this study, we propose to use SNA to study the community of Twitter users disseminating information during the crisis caused by the Australian floods in 2010-2011 in order to reveal interesting patterns and features within this online community. With the help of SNA, we hope to develop an understanding of the online community that was active during that period by answering the following questions: What was the online social behaviour during the flood period? In particular, who were the active players in communicating information and how effective were they? What type of information was of importance? How can the information discovered be useful for the management of such situations in the future?

2 RELATED WORK

Given the usefulness of Twitter as a communication and interaction platform for disaster and emergency situations, it is not surprising to find out that a number of studies have been performed to analyse tweets posted under such conditions.

A content analysis of 1684 tweets collected during Black Saturday, Australia’s worst fire disaster found that these tweets contained actionable factual information contrasting with claims that the contents of tweets are of no value as they are mostly chatter (Sinnappan, Farrell, & Stewart, 2010).

In the case of the Mumbai terrorist attack, Oh et al. (2010) performed a qualitative analysis to argue that the terrorists monitored tweets posted by networked citizens to their advantage as they utilised situational information to mount attacks against civilians.

Over 2 million tweets were collected during a period of about 8 months during the 2009 H1N1 outbreak and manual content analysis was performed on a random sample of 5,395 tweets (Chew, et al., 2010). Content analysis revealed that 52.6% of the posts were resources-related and the most popular resources were news websites followed by web pages of government and health agencies.

The credibility of information contained in tweets was the subject of research during the Chile 2010 earthquake, manual (Mendoza, et al., 2010) and automatic credibility analysis (Castillo, et al., 2011) were performed. Manual examination of the veracity of information disseminated on a small number of tweets during that critical event showed that false rumours tend to be questioned much more than confirmed truths. An automatic classifier was built using features extracted from “trending topics” and was found to be able to classify tweets as credible or not credible with precision and recall in the range of 70% to 80%.

Tweets posted during the 2009 Red River floods were analysed to identify the mechanisms for information production, distribution and organization (Starbird, et al., 2010). It was found that the production of new information on Twitter is by means of derivative activities such as directing, relaying, synthesizing, and redistribution and is additionally complemented by socio-technical innovation. Twitter activity during the Red River floods and Oklahoma grassfires were analysed to identify information that may contribute to situational awareness (Vieweg, 2010). The objective of the study was to identify the features of the information generated for the development of framework to inform the design and implementation of software systems employing information extraction strategies. A prescriptive tweet-based syntax was proposed to increase the utility of information generated during emergencies (Starbird & Stamberger, 2010).

Apart from content analysis of tweets during emergencies, another line of research on Twitter-generated information involved the use of social network analysis techniques, but not in the context of emergencies though.

Tweets were collected in April-May 2007 by fetching them every 30 seconds from the public timeline to study the topological and geographical properties of Twitter's social network in order to determine the intentions of users and the communities they form when they share similar interests (Java, Song, Finin, & Tseng, 2007). Users were found to tweet to talk about their daily activities as well as to seek or share information. User intention was determined using the HITS algorithm (Kleinberg, 1999) (similar to Google's PageRank algorithm) to find hubs and authorities in the network while communities were identified using the Clique Percolation Method (Palla, Derényi, Farkas, & Vicsek, 2005) and their intention determined.

PageRank, Google's method for measuring the relative importance of a URL was applied to Twitter's social graph of users and their followers to determine users of importance (Noordhuis, Heijkoop, & Lazovik, 2010). This is a difficult task as it involves mining 75 million users and 1.8 billion edges and computing 50 iterations of the PageRank algorithm on this data set. However, this task was successfully completed using Amazon Cloud Services at a minimal cost ($275). Another difficult task of computing network metrics for a massive follower graph of 61.6 million users and 1.47 billion edges was successfully computed on a 128-bit Cray computer (Ediger, Jiang, Riedy, Bader, & Corley, 2010).

Dynamic graph analysis was applied to the communication graph of Twitter repliers to identify people of growing influence in social networks based purely on the structure and dynamics of their conversations (Khrabrov & Cybenko, 2010). The data used in this study consisted of 100 million tweets collected using the Twitter Streaming API and the "gardenhose" level of subscription in October-November 2009. The social graph was built from the dataset in an incremental way on a daily basis and PageRank computed at the end of the day.

SNA was used to evaluate the use of Twitter as part of a foreign language learning course by analyzing the interaction of learners and teachers over a period of 56 days (Stepanyan, Borau, & Ullrich, 2010; Ullrich, Borau, & Stepanyan, 2010). The findings indicate that there is greater interaction among students of similar levels and more attention is paid to higher achieving students. As far as gender is concerned, it was observed that there is a preference for study participants to interact with peers of the same gender, and that gender does not determine popularity.

Thus, from the foregoing survey of related research, it is evident that using SNA to analyse the main Twitter users during crisis situations does not seem to have been undertaken previously.

3 METHODOLOGY

Social network analysis (SNA) has emerged as a key technique in the social and behavioural sciences as well as in other major disciplines (Wasserman & Faust, 1994). The main focus of SNA is on the relationships among social entities (e.g. communications among members of a group) and it makes use of a variety of statistical and visual analyses to achieve this. Although, social networks were initially studied in the social sciences, such studies were restricted to rather small systems viewing these networks as static graphs consisting of nodes representing individuals and links representing various quantifiable social interactions. In contrast, recent approaches rooted in statistical physics focus more on large networks searching for universalities both in the topology of the network and in the dynamics governing its evolution (Barabási et al., 2002).

Afterwards, SNA has been increasingly used as a structured way to analyse the extent of informal relationships (among people, teams, departments, or even organisations) within various formally defined groups (Cross, Parker, Prusak, & Borgatti, 2001). SNA makes visible these otherwise invisible patterns of interaction, to identify important groups in order to facilitate effective collaboration (Cross, Borgatti, & Parker, 2003). Thus, SNA helps to identify and assess the health of strategically important networks in an organisation.

Recently, there has been an explosion of the use of SNA in a wide range of domains. A keyword search on Scopus conducted using "social network analysis" revealed 2,173 papers with SNA used in such areas as: health and medicine (Rosenquist, 2011), supply chains (Kim, Choi, Yan, & Dooley, 2011), movies (Park, Oh, & Jo, 2011), cattle movements (Aznar, Stevenson, Zarich, & León, 2011), fraud detection (Šubelj, Furlan, & Bajec, 2011), spam detection (DeBarr & Wechsler, 2010), etc. Our previous work in this area include the analysis of co-authorship networks of the Australasian Conference on Information Systems (Cheong & Corbitt, 2009a), Pacific Asia Conference on Information Systems (Cheong & Corbitt, 2009b), and evaluating student participation in virtual classrooms (Cheong & Corbitt, 2009c).

In the context of this study, we use SNA to gain an understanding of two types of networks. First, when a Twitter user (or *twitterer*) responds to a tweet, a network of twitterers is created with nodes (or vertices) representing twitterers and edges (or links) representing responses to particular tweets of particular twitterers. Since a response flows from a responder to a recipient, the links in this network are directed links. A second type of network that can be constructed from tweets is an "online resources" network as very often tweets contain links to web pages due to 140-word limit of tweets preventing more detailed description of "what is happening". Such a network contains two types of nodes, namely twitterers and resources and these networks are known as bimodal or bipartite networks in the SNA and graph theory literature (Borgatti, 2010; Borgatti & Everett, 1997).

There are many different metrics to analyse social networks. However, since our aim is to identify the most popular and influential twitterers and important resources in the digital social community that emerged during the flood period, we use various centrality measures to identify these twitterers and resources. Furthermore, we do not restrict ourselves to local centrality measures as various global centrality measures are also available to provide a better assessment of the overall community. More details on these measures are provided in the analysis section.

4 DATA COLLECTION AND PROCESSING

Tweets were harvested from Twitter using an in-house script created for the purpose and using a combination of hashtags (e.g. *#qldfloods*, *#nswfloods*, *#vicfloods*). We downloaded tweets (and their meta-data) for the period 3 to 20 February 2011. The data was stored in three separate tables of a database: one for Queensland, one for New South Wales and one for Victoria.

One of the issues encountered in tweet scraping is the rate-limiting restriction imposed by Twitter i.e. 350 requests per hour for registered users and 150 for anonymous users. Thus, we can only hope to collect a sub-set of the tweets and since Twitter only makes the most current tweets available, we cannot guarantee that the sample is a good representation of the population of tweets for the flood period.

Another issue is the data collected for the Queensland floods concern only the post-Queensland flood as the tools for collecting the tweets were implemented too late and since Twitter only makes the last 6-day tweets available to the public, we could not go back in time to capture them. On the other hand, the tweets collected for New South Wales and Victoria cover the entire events while they were happening. Thus, we managed to collect 6,014 tweets for the Queensland floods, 384 tweets for the New South Wales floods and 1,122 for the Victorian floods.

Once the tweets were stored in a database they were available for further analysis. In terms of data cleaning, not much was required since most of the tweets could be traced back to their owners (except for one or two cases when the user closed down his/her Twitter account after the flood) and the web pages mentioned in the tweets easily located.

In order to obtain the nodes and links (edgelist) required to generate the "*users*" and "*users-resources*" networks, individual tweets (and their meta-data) were parsed to extract the (Twitter) identity of the owner of the current tweet, the (Twitter) identity of the owner of the tweet that triggered a response, and the (compressed) URL of the link contained in the tweet. Tweets that were not a response to another tweet or did not contain a URL were ignored for the purpose of the analysis.

The edgelists extracted were processed using R and Python scripts to generate appropriately formatted network data for importation in UCINet (Borgatti, Everett & Freeman 2002), and Pajek (Batagelj & Mrvar 1998), the software used for the analysis. UCINet was used for most of the social network analysis in this study while Pajek was used mainly for visualisation purposes, as it is quite useful for very large networks in that respect.

5 VISUAL NETWORK ANALYSIS

For each flood-affected state (i.e. Queensland, NSW, Victoria), we generated and analysed two types of networks: a "*users*" network and a "*users-resources*" network. The users network during the Queensland floods is shown in Figure 1. The nodes represent users and are shown as circles while the directed links represent responses from one user to another for a particular tweet. Although it can be seen from the sociogram that there are individuals with high number of links (a couple of them is shown circled with thick lines), the diagram is too dense to clearly identify them. Thus, as a next step in the analysis, we extracted the main component (the largest sub-network in which there is a path from a user to any other user) for further visual analysis.

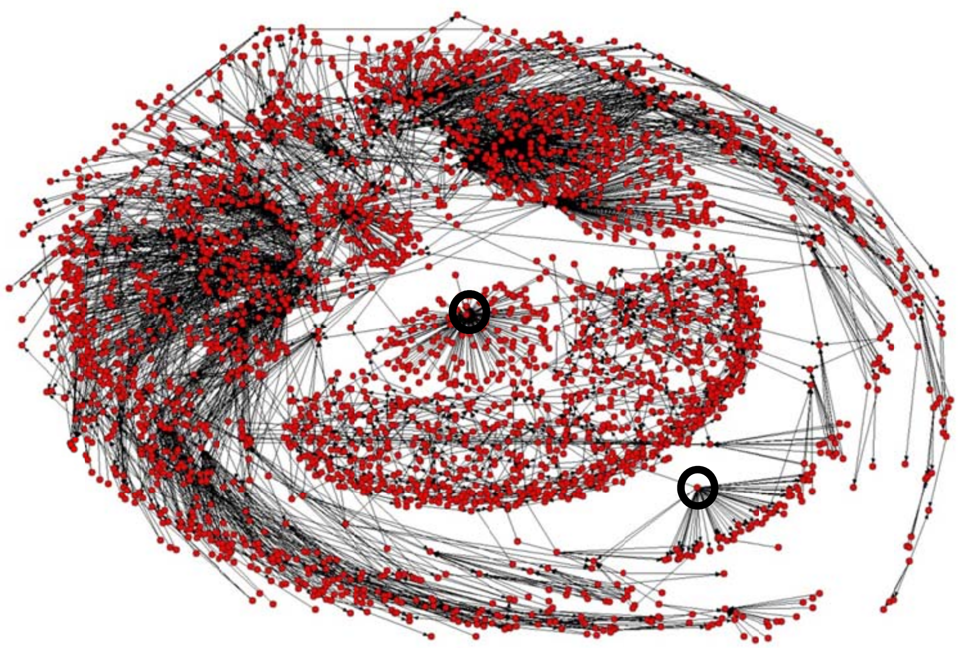

Figure 1. *Users network during Queensland floods*

The main compone nt is shown in Figure 2. This network still has too many nodes to be able to identify the major players (again a few popular players are shown by means of thick circles). Thus, the next step is to resort to ego analysis, that is, quantitative analysis of the individual actors or "egos" that make up the nodes of the network, for identification of influential and popular twitterers. This is discussed in the next section.

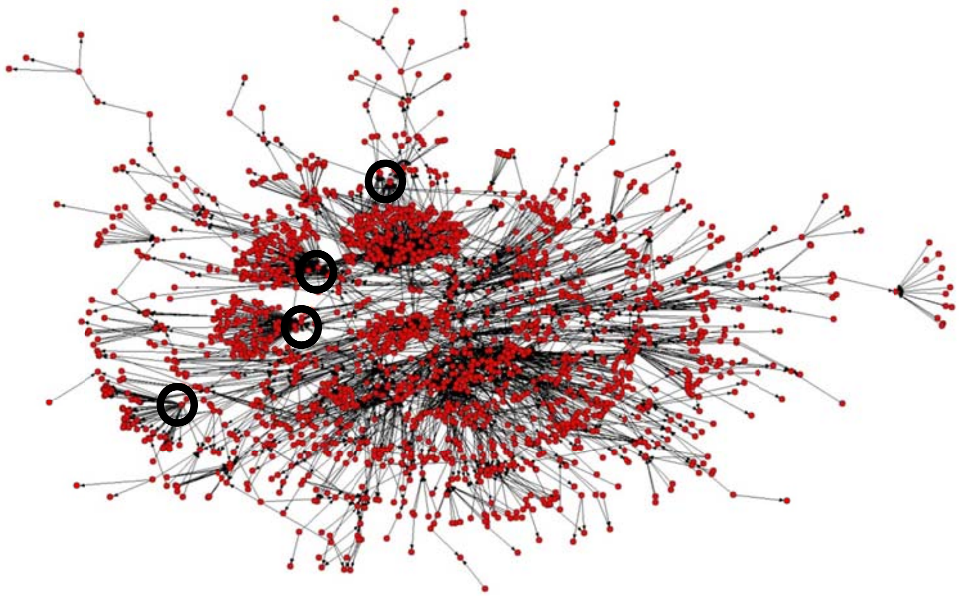

Figure 2. *Main component of users network during Queensland floods*

The users-resources network that emerged during the Queensland floods is shown in Figure 3. There are two types of nodes involved in this type of network. User nodes are represented as circles and online resources as squares. Since the tweets of the users are referring to these resources, the direction of the link is from the user node to the resource node. It can be seen from the figure that some of the resources were quite popular as many users were referring to them and at the same time some users were quite prolific in terms of the number of resources they were suggesting to others. Similarly to the users network, ego analysis is required to objectively determine the important nodes as this cannot be done visually in a network of this size.

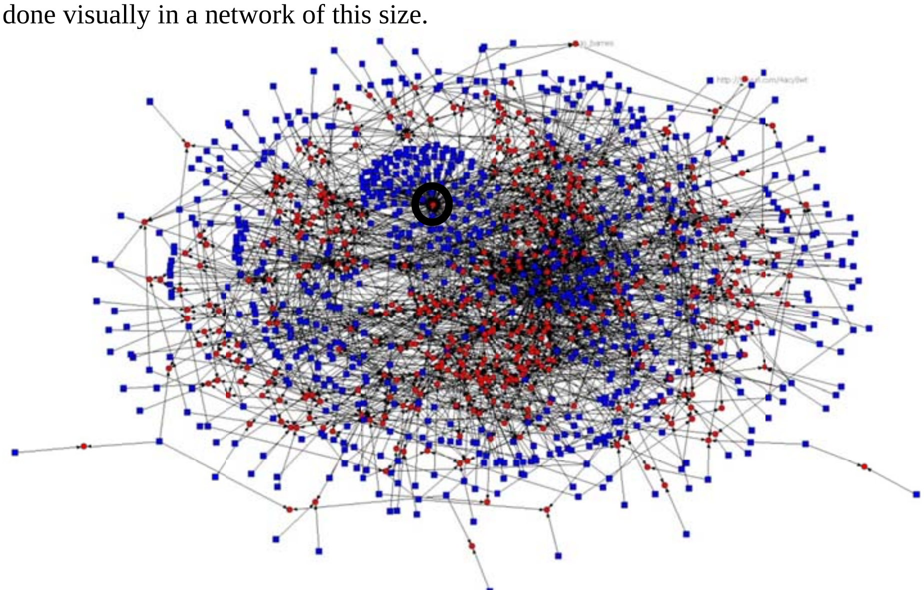

Figure 3. *Users-resources network during Queensland floods*

Figure 4 shows the users network for NSW while Figure 5 shows the users-resources for NSW. These networks are quite small as compared to the Queensland ones, the main reason being the lesser magnitude of the flood disaster in NSW as compared to Queensland and hence less activity on Twitter.

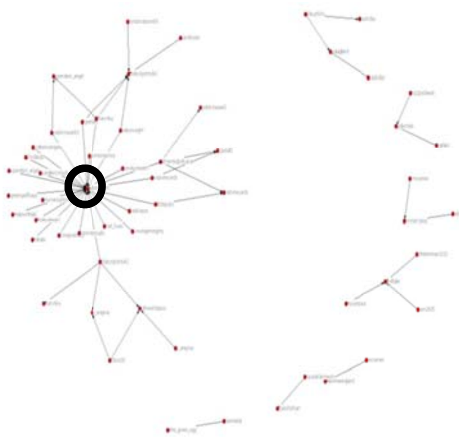

Figure 4. Users network for NSW floods

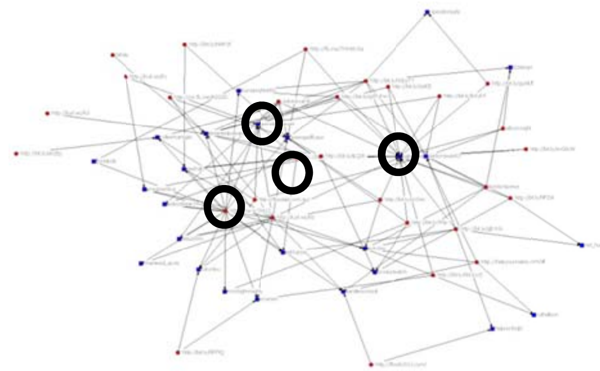

Figure 5. Users-resources network during NSW floods

Although the extent of the flood damage in Victoria was more significant than in NSW, it was still much less disastrous than Queensland. This is reflected in the user and user-resources networks of Figure 6 and Figure 7. The important nodes are slightly more apparent than in the Queensland networks but we will still use ego analysis to rank them in order of importance.

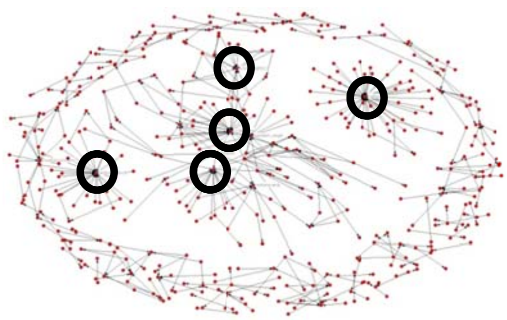

Figure 6. Users network during Victorian floods

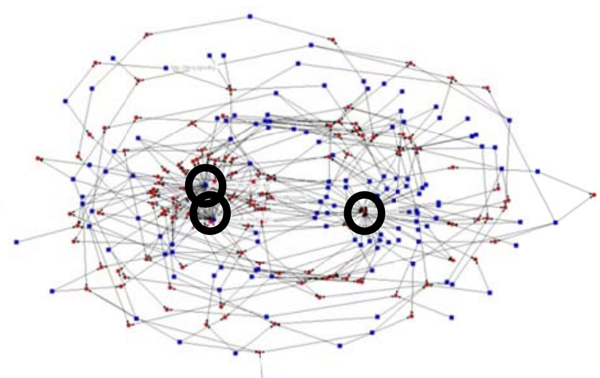

Figure 7. Users-resources network during Victorian floods

6 EGO ANALYSIS OF *USERS* NETWORKS

In this section, we analyse the nodes in terms of their *centrality* in the online network. The idea of centrality of nodes was one of the earliest used by social network analysts and the origins of this idea can be found in the sociometric concept of the *star* i.e. the most popular person or the person at the centre of attention (Scott 2007). Thus, a central node is one at the centre of a number of connections i.e. a node with a large number of direct links with other nodes.

Centrality is measured by the *degree* of the various nodes in the network, with degree representing the number of other nodes to which a node is adjacent. This measure of centrality is known as *local* *centrality* since indirect connections to the particular node are ignored. Thus, the notion of centrality has been extended to *global centrality* (Freeman 1979) to include the distant connections of the nodes. This is measured by the *closeness* of the nodes to other nodes expressed in terms of the distances among the various nodes. *Betweenness* (Freeman 1979) is another centrality measure which measures the extent to which a particular node lies between the various other nodes of the network. A node of relatively low degree may play an important intermediary node (e.g. broker, gatekeeper, etc) and hence be a central node in the network. *Eigenvector* (Bonacich 1972) is another measure of centrality proposed based on the belief that the centrality of a particular node cannot be assessed in isolation from the centrality of all the other nodes to which it is connected. Centrality scores are assigned to nodes based on the principle that connections to high-score nodes contribute more to the score of the particular node than connections to low-score nodes.

The term *structural hole* was coined by Burt (1992) to refer to some important aspects of positional advantage (or disadvantage) of nodes in a network. He developed a number of measures to explain how and why the ways nodes are connected affect their constraints and opportunities and hence their behaviour.

Table 1 shows the top 24 users (due to space limitations) in the main component of the Queensland flood online community ranked by order of importance on the centrality measures previously discussed, namely: (1) *degree*, (2) *betweeness*, (3) *closeness*, (4) *eigenvector* (5) and *structural holes*. In order to have a better idea who these users are, we have visited their Twitter page and extracted their real names and biographies whenever these details were available.

Since we chose to represent our users network as a directed network, a centrality degree analysis yielded two scores: *out degree* (number of tweets sent out by a particular user when responding to tweets of other users) and *in degree* (number of tweets received by a particular user as response from other users). The first section of Table 1 shows the ranking of the top 24 individuals on the *out degree* score while the second section of the table ranks individuals by the *in degree* score. The top scorers in terms of out degree (number of tweets sent out) are users having high influence in the network and just to name a few of them, they are: Sean Robertson (Australian Extreme Weather Event & Disaster Updates), Alexandra Worlson (Not for profit organisation), Isagold Button (virtual name), Wilson Voight (virtual name), Cathy Border (Ten News), George Hall (social media volunteer), etc. Although these users and others from news channels and humanitarian organisations were busy posting tweets, online shopping organisations were also doing the same. The time of tweet collection was during the aftermath of the flood and these organisations were busy promoting their products,

Users with high *in degree* scores are regarded as prestigious or popular individuals and some of them are: Wilson Voight (virtual name), Tony Abbott (Leader of the Opposition), Queensland Police Service (QPS) Media Unit, Andrew Bartlett (ex-Senator), Rove McManus (media personality), Anna Bligh (Queensland Premier), ABC News, Julia Gillard (Prime Minister), Michael Bubble (Canadian singer), etc. In regards to *betweeness* centrality, the top individuals are: QPS Media Unit, Tony Abbott, Anna Bligh, Andrew Bartlett, Wilson Voight, Rove MacManus, ABC News, Sean Robertson, Animal Welfare League, Julia Gillard, etc. Thus, these users can be viewed as leaders in the online network since being on the shortest paths between other users they are able to control the flow of information in the network.

In terms of *closeness* centrality, the leading users were: QPS Media Unit, George Hall (social media volunteer), Liz Baillie (anonymous user), Alexandra Worlson, Wilson Voight, Sean Robertson, Tony Abbott, Operation Angel (Humanitarian Org), Anna Bligh (Queensland Premier), etc. Since closeness centrality measures the distance of a node to all others in the network, the closer a node is to others, the more favoured that node is. Nodes with high closeness scores are likely to receive information more quickly than others as there are fewer intermediaries between them. It is well known that the Queensland Police Service played a very active role in the network and is thus acknowledged as the leader in terms of closeness centrality. Tony Abbott was the leading user when the *eigenvector* centrality criterion is used. This means that he is connected to many other users who are well connected and thus is most likely to receive new ideas. This fits well with his role as the leader of the opposition.

1a. Out Degree								
1	seldomsean63	Sean Robertson	9	dmentedpollyana	Kath Cantarella	17	spencerhow son	Spencer How son
2	babysgotstyle2	Alexandra Worlson	10	visitvineyards	Visit Vineyards	18	tomtomprince	Tommy Prince
3	isagold	IsaGold Button	11	operation_angel	Humanitarian Org	19	annfinster	Online shopping
4	w ilsonvoight	Wilson Voight	12	emmaw right13	Emma Wright	20	digellabakes	Danielle Crismani
5	tencb	Cathy Border	13	612brisbane	ABC New s	21	ecrameri	Emma Crameri
6	geehall1	George Hall	14	qldonline	New s and informatior	22	jayne13	Jayne
7	liz_baillie	Liz Baillie	15	molkstvtalk	Media Commentator	23	stgusface	Gusface
8	askkazza	Karen s	16	karaleecomm	Community Ass.	24	tennew sqld	TEN New s Qld
1b. In Degree								
1	w ilsonvoight	Wilson Voight	9	benmacqueen	Benjamin MacQueen	17	babysgotstyle2	Alexandra Worlson
2	tonyabbottmhr	Tony Abbott	10	clembastow	Clem Bastow	18	feebsquared	Fee Bamford-Bracher
3	qpsmedia	QPS Media Unit	11	benpaddlejones	Ben Jones	19	qldonline	New s and information
4	andrew bartlett	Andrew Bartlett	12	fandoms4floods	FandomsFightTheFloc	20	emmaw right13	Emma Wright
5	rove1974	Rove MacManus	13	612brisbane	ABC New s	21	firstdogonmoon	Mr Onthemoon
6	theqldpremier	Premier Anna Bligh	14	stevejdixon	Steve Dixon	22	pollytics	Possum Comitatus
7	operation_angel	Humanitarian Org	15	juliagillard	Julia Gillard	23	tencb	Cathy Border
8	drew _bow ie	Drew Bow ie	16	michaelbuble	Michael Bubble	24	guttertw its	reb of Gutter Trash
2. Betweenness								
1	qpsmedia	QPS Media Unit	9	liz_baillie	Liz Baillie	17	stevejdixon	Steve Dixon
2	tonyabbottmhr	Tony Abbott	10	operation_angel	Humanitarian Org	18	isagold	IsaGold Button
3	theqldpremier	Premier Anna Bligh	11	612brisbane	ABC New s	19	unclechilliman	I w as Indica Man
4	andrew bartlett	Andrew Bartlett	12	seldomsean63	Sean Robertson	20	jools18	Julie Jones
5	w ilsonvoight	Wilson Voight	13	w inecountrydog		21	molkstvtalk	Media Commentator
6	rove1974	Rove MacManus	14	net_hues	Annette	22	juliagillard	Julia Gillard
7	babysgotstyle2	Alexandra Worlson	15	aw lq	Animal Welfare Leagu	23	leemareegallo	Lee-Maree Gallo
8	geehall1	George Hall	16	qldonline	New s and informatior	24	w olfie_rankin	Wolfie Rankin
3. Closeness								
1	qpsmedia	QPS Media Unit	9	tonyabbottmhr	Tony Abbott	17	w olfie_rankin	Wolfie Rankin
2	geehall1	George Hall	10	jools18	Julie Jones	18	eireaus	
3	liz_baillie	Liz Baillie	11	qldonline	New s and informatior	19	_buyqld	buyQLD.org
4	babysgotstyle2	Alexandra Worlson	12	tdeb007	Tania de Bruin	20	johnalchin	John Alchin
5	w ilsonvoight	Wilson Voight	13	operation_angel	Humanitarian Org	21	can_do_campbell	Fake Lord Mayor of Bris
6	seldomsean63	Sean Robertson	14	theqldpremier	Premier Anna Bligh	22	onegreenbus	
7	unclechilliman	I w as Indica Man	15	tomtomprince	Tommy Prince	23	kate_eltham	
8	mjw ill90		16	visitvineyards	Visit Vineyards	24	ecrameri	Emma Crameri
4. Eigenvector								
1	tonyabbottmhr	Tony Abbott	9	andrew bartlett	Andrew Bartlett	17	domslashryan	
2	drew _bow ie	Drew Bow ie	10	racergirl86		18	nicholosophy	Nicholas Perkins
3	benmacqueen	Benjamin MacQueer	11	melissarghj		19	leemareegallo	Lee-Maree Gallo
4	clembastow	Clem Bastow	12	davidbew art		20	alliejoys	Allie Joy
5	benpaddlejones	Ben Jones	13	mikichoo		21	alw eber	Alick Weber
6	stgusface	Gusface	14	unclechilliman	I w as Indica Man	22	auspoltragic	
7	feebsquared	Fee Bamford-Brach	15	harriettibet	Harriet Tibet	23	billie_mae	
8	urbancreature	Aaron Hew ett	16	nadinelambert	Nadine lambert	24	bspargo7	Brent Spargo
5. Structural Holes								
1	tonyabbottmhr	Tony Abbott	9	w ilsonvoight	Wilson Voight	17	geehall1	George Hall
2	qpsmedia	QPS Media Unit	10	clembastow	Clem Bastow	18	fandoms4floods	FandomsFightTheFlood
3	andrew bartlett	Andrew Bartlett	11	benpaddlejones	Ben Jones	19	juliagillard	Julia Gillard
4	makeuseof		12	isagold	IsaGold Button	20	michaelbuble	Michael Bubble
5	rove1974	Rove MacManus	13	612brisbane	ABC New s	21	feebsquared	Fee Bamford-Bracher
6	drew _bow ie	Drew Bow ie	14	stevejdixon	Steve Dixon	22	qldonline	New s and information
7	benmacqueen	Benjamin MacQueer	15	operation_angel	Humanitarian Org	23	firstdogonmoon	Mr Onthemoon
8	theqldpremier	Premier Anna Bligh	16	babysgotstyle2	Alexandra Worlson	24	pollytics	Possum Comitatus

Table 1. Centrality measures of users in main component of Queensland floods users

Structural holes was measured in terms of *Effective size of the network*, i.e. the number of connections a user has, minus the average number of connections that each individual has to other users. Tony Abbott, QPS Media Unit, Andrew Bartlett, Rove MacManus, Anna Bligh, etc. led on this criterion suggesting that they have more opportunities to act as brokers or coordinators.

Due to space limitations, we are unable to present a detailed analysis of the NSW and Victorian networks, instead we will discuss the main findings. The level of Twitter activity regarding the NSW floods was too low to draw any meaning conclusion. Although the local authorities were criticised for not taking an active role on Twitter (StreetCorner, 2011) like the Queensland Police Service did, an army of individuals would have stepped in to provide information about relief centres and traffic conditions should the floods have had a more damaging impact. A number of Twitterers actively dissemination information during the Victorian floods were those involved in the Queensland floods. Notable examples are: Sean Robertson (Australian Extreme Weather Event & Disaster Updates), Alexandra Worlson (Not for profit organisation), Karalee Community Association, ABC News, Wilson Voight (virtual name), and Operation Angel (humanitarian organisation). We also noted the presence of the State Library of Queensland and the Australian Librarian and Information Association who were actively providing a range of material and information about floods and natural disasters.

7 EGO ANALYSIS OF *USERS-RESOURCES* NETWORKS

We analysed the *users-resources* network as a 2-node network (Borgatti, 2010; Borgatti & Everett, 1997) using centrality measures for online resources as well as users. The top 25 resources as measured by the degree centrality measure during the Queensland floods are shown in Table 2. At the top of the list, were online resources discussing the proposed flood levy (as this was the end of the Queensland flood period) followed by web pages related to the following: people with disability, animal rescue, donations, legal help, damaged produce, fraud investigations, fund raising, counselling, temporary homes, volunteering, and making fun of the situation on YouTube.

1. Resources centrality			
	Compressed URL	**URL**	**Title**
1	http://cot.ag/guICOS	http://w w w .email.sw ordcdc.com/t/View Email/r/E0C4	We can do better than Labor's flood tax
2	http://tw itpic.com/3voi8i	http://tw itpic.com/3voi8i	tw itpic
3	http://bit.ly/gCg8y1	Not found	
4	http://bit.ly/fIl6nD	http://w w w .youtube.com/w atch?v=oexWpfZ_2-w	A Message For QLD
5	http://bit.ly/fdacpZ	Microsoft Word document	People w ith disability media release
6	http://bit.ly/htxBDD	http://lostfound.rspcaqld.org.au/search/?s=lost&new	RSPCA home page
7	http://t.co/zDdznef	http://guttertrash.w ordpress.com/2011/02/03/tony-ak	Tony Abbott begs for donations to block financial support for QLD
8	http://bit.ly/fqdzmU	http://w w w .flickr.com/photos/telstra-corp/sets/7215	SAT COW on Palm Island - Cyclone Yasi
9	http://w w w .qld.gov.au/floods/donate.html		Donate to the disaster relief appeal
10	http://bit.ly/f0RuyR	http://w w w .legalaid.qld.gov.au/floods/Pages/default.	Flood and cyclone legal help for Queenslanders
11	http://bit.ly/hGEaTY	http://w w w .brisbanetimes.com.au/environment/w eat	Buying damaged produce show s same spirit that fought disaster
12	http://bit.ly/hNLpZa	http://w w w .rspcaqld.org.au/	RSPCA Queensland Animal rescue
13	http://bit.ly/iaPlSh	http://arterystore.com/index.php?option=com_conten	The Give and Take of the Flood of 2011
14	http://helpyourmates.com/all		Mates affected by the floods around Australia need your help.
15	http://bit.ly/edQOCl	http://w w w .brisbanetimes.com.au/environment/w eat	Fraud investigations into 1400 relief claims
16	http://bit.ly/hecb1y	http://insidecuisine.com/2011/02/04/floodlight-dinner-	Floodlight dinner Sydney raises $70K
17	http://on.fb.me/f6mXAI	http://w w w .facebook.com/pages/Adopt-a-cyclone-Y	Adopt a cyclone Yasi affected tow n
18	http://bit.ly/aXKEeb	?	
19	http://bit.ly/dIRAJB	http://w w w .blazeaid.com/index.html	BlazeAid - volunteering
20	http://bit.ly/e1hr6F	http://new s.ninemsn.com.au/national/floods/8209229	Temporary homes arrive in Grantham
21	http://bit.ly/ectnJh	http://w w w .readfearn.com/2011/02/roots-of-resilien	Roots of resilience in community gardens
22	http://bit.ly/fMLpl1	http://tw itter.com/Operation_Angel#	Not for profit humanitarian organisation
23	http://bit.ly/i0Cl74	https://w w w .eheadspace.org.au/	Headspace - online counselling service
24	http://bit.ly/f0evBM	http://floodelectricianbuildersqld.com/	Flood relief tradies - helping each other
25	http://bit.ly/f86BYa	https://salvos.org.au/donate/secure-online-donations	Salvos - Secure online donations □

Table 2. Centrality of resources during the Queensland floods

A degree centrality analysis was also performed on the users on this users-resources network and details are shown in Table 3. Some of the users actively disseminating information on these resources include: Wilson Voight, Sean Robertson, Alexandra Worlson, etc.

Again due to space restrictions, we shall summarise our findings regarding the identification of important online resources during the NSW and Victorian floods. During the NSW floods, new (in terms of what we have already identified during the Queensland flood) important online resources were related to: accounting implications of natural disasters, donor information, tips for supporting children in disasters, Australian Red Cross, rights and obligation of employers and employees during natural disasters, and Australian Taxation Office. No new important resources appeared during the Victorian floods as mostly resources that have already been identified were referred to.

2. Users centrality								
1	w ilsonvoight	Wilson Voight	9	ljloch	LJ Loch	17	tomtomprince	Tommy Prince
2	seldomsean63	Sean Robertson	10	lyndsayfarlow	Lyndsay Farlow	18	net_hues	Annette
3	babysgotstyle2	Alexandra Worlson	11	greengadflyaus	DianeO'Donovan	19	w inecountrydog	
4	askkazza	Karen S	12	annfinster	Online shopping	20	servicecentralz	Service Central
5	visitvineyards	Visit Vineyards	13	ecrameri	Emma Crameri	21	findatoow oomba	Finda Toow oomba
6	qldonline	New s and informati	14	minxyferret		22	collectiveact	Rosie Williams
7	dmentedpollyana	Kath Cantarella	15	liz_baillie	Liz Baillie	23	autoday	
8	geehall1	George Hall	16	digellabakes	Danielle Crismani	24	alertnetclimate	Alertnet Climate

Table 3. Centrality of users in the users-resources network of the Queensland floods

8 DISCUSSION

Based on Powell and Rayner's widely used taxonomy (1952), several stages can be identified in a disaster, namely: (1) warning, (2) threat, (3) impact, (4) inventory, (5) rescue, (6) remedy and (7) recovery. Most studies focus on the impact, inventory and rescue stages since traditional communications are less effective during these stages (Mendoza, et al., 2010). In this aspect, Twitter has proved itself to be a valuable platform for disseminating vital information.

Although the tweets for the Queensland floods were collected after the impact stage, we believe they are still valuable since social network analysis revealed a number of users who were known to be active in that online community (StreetCorner, 2011). Additionally, using several local and global centrality measures, SNA helped to identify the effectiveness of these users. Active and effective users identified were: local authorities (mainly the Queensland Police Services), political personalities (Queensland Premier, Prime Minister, Opposition Leader, Member of Parliament), social media volunteers, traditional media reporters, and people from not-for-profit, humanitarian, and community associations.

It is well known that Queensland Police took a very active role on Twitter, providing the public with regular updates on the situation every few minutes as well as dealing with the spread of misinformation on Twitter (StreetCorner, 2011). Queensland Police was also very active on Facebook, providing more detailed updates than is possible with 140-character tweets.

Although the analysis of the users-resources network during the Queensland floods identified a wide range of important resources, these resources were mostly web pages and blogs providing information of a more general nature rather than vital information and updates on the disaster. Since it is more effective to disseminate critical information on Facebook (because of high penetration) and mining Facebook was not part of the study, we would have missed such information. If this was the case, it makes more sense to conclude that the resources identified supplement the resources posted on Facebook.

Unlike the Queensland floods, data collection for the NSW and Victorian floods was during the threat and impact stages, and unlike Queensland there was no evidence of Twitter activity from the part of the local authorities and the government in NSW and Victoria. Furthermore, the level of Twitter activity during the NSW floods was almost nil. Most of the active players during the NSW and Victorian floods were volunteers who were active during the Queensland floods.

The Federal Emergency Management Agency (FEMA) of the US Department of Homeland Security recognises the usefulness of Twitter (and other social media) during emergencies and uses Twitter during all stages of a disaster, before the event strikes, during the event and after.

Given the positive results obtained by the involvement of the local authorities and government officials in Queensland, and the increasing adoption of Twitter in other parts of the world for emergency situations, it seems reasonable to push for greater adoption of Twitter by local and federal authorities Australia-wide during periods of mass emergencies. This will help to ensure that vital information of an official and reliable nature is quickly propagated throughout the network and false rumours dealt with as they emerge.

9 CONCLUSION AND FUTURE WORK

We used SNA to study interactions between Twitter users during the Australian 2010-2011 floods, one of the worst Australian flooding disasters occurring across several states. SNA provides techniques to analyse the structure of a network as an entity as well as with techniques to analyse individual nodes and their place in the network. Using SNA metrics and techniques we were able to identify influential members of the online communities that emerged during the Queensland, NSW and Victorian floods as well as identify important resources being referred to. The most active community was in Queensland, possibly induced by the fact that the floods were orders of magnitude greater than in NSW and Victoria. The analysis confirmed the active part taken by local authorities, namely Queensland Police, government officials and volunteers. On the other hand, there was not much activity from local authorities in the NSW and Victorian floods prompting for the greater use of social media by the authorities concerned. As far as the online resources suggested by users are concerned, no sensible conclusion can be drawn as important ones identified were more of a general nature rather than critical information. This might be comprehensible as it was past the impact stage in the Queensland floods and participation was at much lower levels in the NSW and Victorian floods.

One of the acknowledged limitations of the research is the insufficient number of tweets collected to be able to generalise the findings. Thus, a first future work is to improve the tweet collection rate in order to obtain a much larger sample as close as possible to the whole population of tweets for the event. This is a difficult task as Twitter limits the number of tweets that can be downloaded and special permission is required for greater volumes of download.

Although SNA provides an objective method for identifying the main actors in the online community, it mainly measures patterns and frequencies of communications; other techniques are required to gauge the nature and the quality of communications. Text mining and analytics would be helpful in that regard as it can be used for analysing the content of tweets and the content of online resources being suggested by users. Thus, another future work we are contemplating is the combination of SNA with text mining and analytics for improved assessment of the effectiveness of Twitter during mass emergencies.

10 ACKNOWLEDGEMENTS

We would like to acknowledge the contribution of Nicholas Potesta, who was working in the School of Business IT and Logistics as part of an internship scheme. He assisted in the development of the initial prototype of the social media analytics system.

References

ABC News. (2010, 5 September). Army to evacuate Victorian flood victims. Retrieved 8 March, 2011, from http://www.abc.net.au/news/stories/2010/09/05/3002955.htm

Australian Bureau of Meteorology. (2010). Gascoyne River Flood, 16 - 20 December 2010. Retrieved 8 March, 2011

Aznar, M. N., Stevenson, M. A., Zarich, L., & León, E. A. (2011). Analysis of cattle movements in Argentina, 2005. *Preventive Veterinary Medicine, 98*(2-3), 119-127.

Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations *Physica A: Statistical Mechanics and its Applications, 311*(3-4), 590-614

Borgatti, S. P. (2010). 2-Mode Concepts in Social Network Analysis (to appear in Encyclopedia of Complexity and System Science). Retrieved 9 May 2011, from http://www.steveborgatti.com/papers/2modeconcepts.pdf

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks, 19*, 243-269.

Borgatti, S. P., & Foster, P. C. (2003). The Network Paradigm in Organizational Research: A Review and Typology *Journal of Management*, *29*(6), 991-1013.

Brisbane Times. (2011, 11 January). Three-quarters of Queensland a disaster zone Retrieved 8 March, 2011, from http://www.brisbanetimes.com.au/environment/weather/threequarters-of-queensland-a-disaster-zone-20110111-19mf8.html

Castillo, C., Mendoza, M., & Poblete, B. (2011, 28 March - 1 April). *Information Credibility on Twitter.* Paper presented at the WWW 2011, Hyderabad, India.

Cheong, F., & Corbitt, B. J. (2009a). *A social network analysis of the co-authorship network of the Australasian Conference of Information Systems from 1990 to 2006*. Paper presented at the 17th European Conference on Information Systems (ECIS 2009)

Cheong, F., & Corbitt, B. J. (2009b). *A social network analysis of the co-authorship network of the Pacific Asia Conference on Information Systems from 1993 to 2008*. Paper presented at the Pacific Asia Conference on Information Systems (PACIS 2009)

Cheong, F., & Corbitt, B. J. (2009c). *Using Social Network Analysis to Evaluate Participation in Online Discussions in a Virtual Classroom*. Paper presented at the 20th Australasian Conference on Information Systems.

Chew, C., Eysenbach, G., & Sampson, M. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One, 5*(11), 1-13.

Cross, R., Borgatti, S., & Parker, A. (2003). Making Invisible Work Visible. . In R. Cross, A. Parker & L. Sasson (Eds.), *Networks in the Knowledge Economy* (pp. 261-282). New York: Oxford Press.

Cross, R., Parker, A., Prusak, L., & Borgatti, S. P. (2001). Knowing what we know: supporting knowledge creation and sharing in social networks. *Organizational Dynamics, 30*(2), 100-120.

Davis, G. F., & Greve, H. R. (1997). Corporate Elite Networks and Governance Changes in the 1980s. *The American Journal of Sociology, 103*(1), 1-37.

Davis, G. F., Yoo, M., & Baker, W. E. (2003). The Small World of the American Corporate Elite, 1982-2001 *Strategic Organization, 1*(3), 301-326.

DeBarr, D., & Wechsler, H. (2010). Using social network analysis for spam detection (Vol. 6007 LNCS, pp. 62-69).

Dynes, R. R., & Quarantelli, E. (1968). Group behavior under stress: a requered convergence of organizational and collective behavior perspectives. *Sociology and Social Research, 52*, 416-429.

Ediger, D., Jiang, K., Riedy, J., Bader, D. A., & Corley, C. (2010). *Massive Social Network Analysis: Mining Twitter for Social Good.* Paper presented at the International Conference on Parallel Processing (ICPP '2010).

Garton, L., Haythornthwaite, C., & Wellman, B. (1999). Studying Online Social Networks. In S. Jones (Ed.), *Doing Internet Research* (pp. 75-105). Thousand Oaks, CA: Sage Publications.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, 12 August). *Why we twitter: understanding microblogging usage and communities.* Paper presented at the Joint 9th WEBKDD and 1st SNA-KDD Workshop '07, San Jose, California , USA.

Khrabrov, A., & Cybenko, G. (2010). *Discovering influence in communication networks using dynamic graph analysis*, 2nd IEEE International Conference on Social Computing (SocialCom 2010).

Kim, Y., Choi, T. Y., Yan, T., & Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management, 29*(3), 194-211.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604-632.

Mendoza, M., Poblete, B., & Castillo, C. (2010, 25 July). *Twitter Under Crisis: Can we trust what we RT?* Paper presented at the SOMA '10, Washington, DC, USA.

Newman, M. E. J. (2001). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences, 101*(Suppl. 1), 5200-5205.

Nine News. (10 January 2011). More than 1000 isolated in NSW floods.   Retrieved 8 March, 2011

Noordhuis, P., Heijkoop, M., & Lazovik, A. (2010, 5-10 July 2010). *Mining Twitter in the Cloud: A Case Study.* Paper presented at the Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on.

Oh, O., Agrawal, M., & Rao, H. R. (2010). Information control and terrorism: tracking the Mumbai terrorist attack through Twitter. *Information Systems Frontiers*, 1-11.

Palen, L., & Vieweg, S. (2008, 6-10 February). *The Emergence of Online Widescale Interaction in Unexpected Events: Assistance, Alliance & Retreat.* Paper presented at the CSCW '08.

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature, 435*(7043), 814-818.

Park, S. B., Oh, K. J., & Jo, G. S. (2011). Social network analysis in a movie using character-net. *Multimedia Tools and Applications*, 1-27.

Powell, J. W., & Rayner, J. (1952). *Progress notes: Disaster investigation*: Maryland Army Chemical Center.

Rosenquist, J. N. (2011). Lessons from social network analyses for behavioral medicine. *Current Opinion in Psychiatry, 24*(2), 139-143.

Sinnappan, S., Farrell, C., & Stewart, E. (2010, 1-3 December). *Priceless Tweets! A Study on Twitter Messages Posted During Crisis: Black Saturday.* Paper presented at the 21st Australasian Conference on Information Systems (ACIS 2010), Brisbane.

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010, 6-10 February). *Chatter on the red: what hazards threat reveals about the social life of microblogged information.* Paper presented at the CSCW '10, Savannah, Georgia, USA.

Starbird, K., & Stamberger, J. (2010, May). *Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Syntax to Support Citizen Reporting.* Paper presented at the 7th International ISCRAM Conference, Seattle, USA.

Stepanyan, K., Borau, K., & Ullrich, C. (2010, 5-7 July). *A social network analysis perspective on student interaction within the twitter microblogging environment.* Paper presented at the 10th IEEE International Conference on Advanced Learning Technologies, ICALT 2010 Sousse.

StreetCorner. (2011). Police and public turn to social media & maps in the Queensland flood crisis Retrieved 11 March, 2011, from http://www.streetcorner.com.au/news/showPost.cfm?bid=20421&mycomm=ES

Stuart, T. E. (1998). Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-technology Industry. *Administrative Science Quarterly, 43*(3), 668-698.

Šubelj, L., Furlan, S., & Bajec, M. (2011). An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications, 38*(1), 1039-1052.

The Age. (2011, 14 January). Flash flooding hits Victoria Retrieved 8 March, 2011, from http://www.theage.com.au/environment/flash-flooding-hits-victoria-20110113-19pu7.html

The Courier-Mail. (2010). Damage from southwest Queensland floods to become clearer as waters recede Retrieved 8 March, 2011, from http://www.couriermail.com.au/news/queensland/damage-from-southwest-queensland-floods-to-become-clearer-as-waters-recede/story-e6freoof-1225837963854

Ullrich, C., Borau, K., & Stepanyan, K. (2010). Who students interact with? A social network analysis perspective on the use of twitter in language learning *Lecture Notes in Computer Science* (Vol. 6383 LNCS, pp. 432-437).

Vieweg, S. (2010, February). *Microblogged contributions to the emergency arena: Discovery, interpretation and implications.* Paper presented at the Computer Supported Collaborative Work.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York City, New York, U.S.A.: Cambridge University Press.

Wikipedia. (2011). Social media. Retrieved 13 March, 2011, from http://en.wikipedia.org/wiki/Social_media

Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *International Journal of Information Management, 31*(1), 6-13.