

Towards credibility of micro-blogs: characterising witness accounts

Marie Truelove · Maria Vasardani ·
Stephan Winter

Published online: 27 April 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Information about events can be opportunistically harvested from social media, however, a major challenge is assessing the credibility of the information derived, and the credibility of the micro-bloggers who are the source of the information. Witnesses to events are intrinsically linked with credibility for many disciplines including journalism and the criminal justice system. This research seeks to determine whether likely witness accounts of an event can be differentiated from social media feeds. A conceptual model of a *witness account*, and related *impact accounts* and *relayed accounts* is developed. Additionally, *influence regions* defining a relationship between witnesses and events are inferred, from different categories of witness accounts. This model is explored and tested using a bushfire event as a case study. In depth manual analysis of Twitter data related to this event and its effects, confirms the expected revelations of characteristics of direct observations of a bushfire that witnesses report, and the impacts and actions potential witnesses report. A visualisation of influence regions for smoke and traffic congestion observations is provided. Additionally, for the case study event, it is observed that witness accounts contain fewer place name references, but more personal place descriptions such as ‘my home’. These findings suggest implications for automatic data

mining from place descriptions that will enable an assessment of the credibility of extracted event information.

Keywords Credibility · Witness · Place descriptions · Data mining · Social media

Introduction

Research has recognised the benefits of opportunistically harvesting information from micro-blogging and social media services such as Twitter. In the time critical phases of a crisis, contributions can be made to situational awareness (MacEachren et al. 2011), which has been described as ‘...an individually as well as socially cognitive state of understanding “the big picture” ...’ (Vieweg et al. 2010, p. 1079). Event types which have served as case studies include forest fires (De Longueville et al. 2009; Vieweg et al. 2010), floods (Poser and Dransch 2010), earthquakes (Stollberg and de Groeve 2012; Mendoza et al. 2010), and political protests from a geographically localised (Starbird et al. 2012) to country-wide scale (Cheong et al. 2012; Kumar et al. 2013; Starbird and Palen 2012). However, identifying relevant information pertaining to a particular event and assessing the credibility of users and information derived is challenging. In addition, researchers have identified false

M. Truelove (✉) · M. Vasardani · S. Winter
The University of Melbourne, Parkville, VIC, Australia
e-mail: truelove@student.unimelb.edu.au

rumours spread via social media during crisis events (Mendoza et al. 2010; Goodchild and Glennon 2010), with false reports that the New York Stock Exchange was flooded during Hurricane Sandy in 2012 misleading even mainstream news networks such as CNN (Mirkinson 2012; Wortham 2012).

MacEachren et al. (2011) found that maps that identify both the location of the event and the micro-blogger to be the most useful for crisis management personnel. These maps would enable the visual assessment of how close a micro-blogger is to the event they are contributing information about. Researchers with focus on social media have identified that ‘People who are on the ground are uniquely positioned to share information that may not yet be available elsewhere in the information space’ (Starbird et al. 2012, p. 2). Diakopoulos et al. (2012) describe the journalistic focus and importance of eyewitnesses for breaking news, where being in close proximity to an event and being able to make a report is preferred to expert knowledge on a topic. Providing eyewitness accounts to a breaking news story provides credibility to the story and is one of the ‘... quintessential acts of journalism’ (Diakopoulos et al. 2012, p. 2452). Witnesses are a fundamental component of the criminal justice system and criteria to assess witness credibility include ensuring the opportunity to view the crime (Wells and Olson 2003).

For Twitter, researchers have identified numerous sources which can be analysed to identify the geographic location of micro-bloggers and their topics (e.g. Ostermann and Spinsanti 2010). These include, but are not limited to, the optional micro-blogger’s account metadata location (metadata location), the optional mobile device determined GPS location, and spatial descriptions within the content of the micro-blogs (content locations). Despite there being numerous sources, it is a complex problem to determine the location of micro-bloggers and events, especially to a granularity of less than a city. Hecht et al. (2011) found 34 % of metadata locations contained non-geographic or blank entries, and for valid geographic entries, less than 10 % were to a granularity of city. Micro-blog content relating to crisis events has been identified to contain location references in up to 40 % of cases dependent on the type of event (Vieweg et al. 2010), but deriving unambiguous locations from such content continues to be challenging. Recent reports indicate just 3 % of tweets have linked GPS coordinates

(Leetaru et al. 2013). This complimentary research contributes to identifying locations of the micro-blogger and/or the event they are discussing, but does not focus further on the relationship between the user and the event.

This research explores whether witness accounts of an event can be differentiated from social media micro-blogs. The definition of a witness account will be further explored later but for now a witness can be understood as a person who has directly observed the event and posted a micro-blog about their observation, called the *witness account* (WA). It is envisaged, a range of characteristics might differentiate likely WAs from those which are not. These characteristics include descriptions of sensing such as ‘I see’, ‘hear’, or ‘smell’, linked content such as photos, and explicit acknowledgement of being impacted by the event. The hypothesis for this work is that likely WAs of an event can be differentiated from unlikely WAs, based on a categorisation of characteristics of micro-blogs.

To be a witness, the micro-blogger must be in the region affected by the event, which will depend on the type of event and effect that is reported. If the relationship between the event and the micro-blogger can be inferred it reduces the reliance on social media location sources to explicitly locate the micro-blogger for credibility assessment, and creates additional spatial intelligence. Previous research has identified that places may not be referred to by their name but as general place categories as they become assumed knowledge within the social network, and further used as reference points (Vieweg et al. 2010). This research also explores content that contains informal place categories related to personal action spaces (Goodchild 2009) such as ‘home’ and ‘work’ rather than place names, which may additionally identify these micro-bloggers as to have local knowledge. Fewer place names in micro-blogs from people in close proximity to an event may have implications for research seeking to identify actionable content related to events. Specific assumptions to be tested include WAs provide influence regions between micro-bloggers and events, and contain fewer named place references than non-witness accounts.

Additional goals of this research are motivated by questions raised from previous social media research. The use of metadata locations for research has limitations (Hecht et al. 2011), leading some researchers to reject them as a location source (e.g. Starbird

et al. 2012), and others to accept them (e.g. Cheong et al. 2012). This research will explore metadata locations to determine whether they are appropriate for corroborating likely WAs.

The approach adopted is an in-depth manual analysis of a single case study. The case study event was a bushfire on the northern urban boundary of metropolitan Melbourne. This was a significant local event causing widespread disruption in the surrounding suburbs. Data collected included 1711 micro-blogs using the keyword ‘bushfire’, micro-blogger account metadata, and publicly available linked content. Using manual methods, 461 on-topic, individual and original (OIO) micro-blogs were identified, and of these 198 were identified as potential WAs which were further coded for identified categories. The results of this research include models of witnesses and related concepts and indicate that likely WAs can be differentiated and categorised by the reported effects. Influence regions between witnesses and the event, based on reported effects, are visualised and further characteristics related to place name usage by witnesses are also established.

Section “[Background](#)” presents background research, in Sect. “[Theory](#)” concepts are developed, and Sect. “[Methods](#)” outlines the method for collection and preparing a corpus case study. Results are presented in the Sect. “[Results](#)” with specific interpretations, and Section “[Summary](#)” provides a high level discussion and evaluation with conclusions and potential future research in Sect. “[Conclusion and future work](#)”.

Background

When an emergency is occurring authorities and journalists alike will seek out witnesses who can provide accounts of the unfolding events. Journalists will seek witnesses because they can provide differentiation and credibility to a news story (Diakopoulos et al. 2012), and witnesses themselves will be assessed by authorities to determine suitability for providing evidence. The seminal paper of Fogg and Tseng (1999) defines the credibility of information as *believability*. Their synthesis of the academic research concludes credibility as a perceived value with a number of dimensions including *trustworthiness* and *expertise*. Research has tested the perceived credibility of topics evolving in Twitter (Castillo et al. 2011), and

which Twitter features people use to perceive the relative credibility of their fellow micro-bloggers (Yang et al. 2013; Ringel Morris et al. 2012). These examples of credibility and Twitter research do not consider the location information that can be generated from metadata and content. In comparison Thomson et al. (2012) conclude that micro-bloggers who do not publish a metadata location can be correlated with sharing less credible information, but find this correlation is mitigated when micro-bloggers are from the same country as the event.

With an excess of 400 million micro-blogs daily (Twitter 2013), there appears to be universal agreement that the micro-blogging service Twitter, is a noisy information stream. In this stream, research focuses on identifying the most relevant information, which differs from, but is related to identifying the most credible information. In research with this focus, location information plays a larger role. Relevance theory states that an input is relevant to a user when its processing in context produces a positive cognitive affect (Sperber and Wilson 2004). Kumar et al. (2013) seek to identify the most relevant information about events, and to this goal distinguish between ‘... local users who witness the unfolding event and remote users who are connected via social media’ (Kumar et al. 2013, p. 139). They describe automated methodologies employed to assign topic affinity and geo-relevancy scores. Spatial analysis to derive the geo-relevancy scores do not attempt to determine if micro-bloggers are witnesses to any of the events. Rather, GPS or metadata locations are used to determine whether the micro-blogger is within the country where the events are occurring. So though micro-bloggers with a higher-than average geo-relevancy and topic affinity score are labelled as ‘eyewitness users’, they are not necessarily a witness according to the definitions of this research.

Kumar et al. (2013) also refer to a concept of micro-bloggers being ‘on-the-ground’ (OTG). Other research classifies micro-bloggers into two categories being ‘... those who were on the ground and tweeting information from the ground, and ... those who were not on the ground or were not tweeting information about the protests from the ground’ (Starbird et al. 2012, p. 6). Manual content analysis was utilised to achieve this categorisation for the purposes of analysing crowd recommender behaviour, with earlier work looking specifically at retweet behaviour of local micro-

bloggers in comparison to other micro-bloggers (Starbird and Palen 2010). For this body of research metadata locations were not used, as it was identified these may be purposely falsified by people tweeting in political protests. They also identified micro-bloggers who were ‘... tweeting real-time information from the ground without being physically present at the event’ (Starbird et al. 2012, p.6) by utilising a live broadcast of the event. This highlights challenges of identifying micro-bloggers whose information is gained from direct observation from those observing news broadcasts and other sources.

In contrast to the researchers already described, who primarily utilise spatial information to identify those micro-bloggers which are OTG, Diakopoulos et al. (2012) develop a dictionary-based technique to classify potential witnesses based on 741 words from numerous Linguistic Inquiry and Word Count (LIWC)¹ categories including *percept*, *see*, *hear*, and *feel*. They compare their classifier with human analysis using three subjects sourced from Amazon Mechanical Turk, and conclude their method suitable for their purposes of prototyping a tool for journalists, however, acknowledge their approach to be ‘... only a first step toward the challenge of classifying eyewitnesses ...’ (Diakopoulos et al. 2012 p. 2455). Today, the main way for the news media to source early witness reports and images is from everyday citizens (Wigley and Fontenot 2010). Witness reports in social media have been identified to be two and half times more likely to contain linked content (Harlow 2011). But journalists need to be increasingly cognisant of the definition of credibility as believability when considering these sources, with investigation of the most retweeted images related to Hurricane Sandy identifying many fakes (Burgess et al. 2012).

Vieweg et al. (2010) present an in-depth analysis of what location information was contained in social media content for two emergency events, a flood and a grassfire with a view towards extracting useful information to contribute to situation awareness (SA). They found 40 % of micro-blogs related to the fires contained a clearly identifiable address or place, while 18 % did so related to the flood. Though not explored, this research identifies the presence of what is named ‘markedness’ and ‘relative references to location’.

Markedness of places is described as places no longer referred to by their name but by their general category as an emergency event unfolds, which impacts the ability to extract relevant micro-blogs. Relative references to locations are described as ambiguous reference points used in spatial descriptions. The authors identify for the flood and fire case studies that this occurs for 6 and 8 % of micro-blogs respectively.

Researchers have used Natural Language Processing (NLP) and Machine Learning (ML) techniques in efforts to automate the classification of micro-blogs that contain actionable information contributing to SA (Verma et al. 2011). The authors found that training the classifiers on spectrums of ‘...subjectivity, personal or impersonal style, and linguistic register (formal or informal style)’ enhanced the results (Verma et al. 2011, p. 386). This is based on the authors positing that micro-blogs ‘...that contribute to situational awareness are likely to be written in a style that is objective, impersonal, and formal...’ (Verma et al. 2011, p. 386). NLP has also been employed to extract place descriptions from micro-blogs (e.g. Gelernter and Balaji 2013). Numerous approaches to identifying, extracting and disambiguating place names that can be found within the content of micro-blogs and or in combination with the metadata locations, can be referred to (e.g. Sankaranarayanan et al. 2009; Cheng et al. 2010). Identifying words common to geographic regions is another approach (e.g. Eisenstein et al. 2010). Additionally, approaches for identifying and disambiguating place descriptions in other content forms such as streaming news are relevant (e.g. Lieberman and Samet 2012). This significant body of complementary research generally seeks to determine the location of the micro-blogger or event (or may not distinguish between the two), rather than further distinguish the relationship between the micro-blogger and the event. These methodologies may be used to extract the on-topic micro-blogs, which in addition to being individual and from an original source, are pre-requisites for identifying witnesses. An exception is research which implements regression analysis to establish demographic indicators between micro-bloggers who contribute actionable content that can be geocoded (a dependency for the analysis), and a wildfire event (Kent and Capello 2013).

Although it has been clarified that geospatial information opportunistically harvested from social

¹ <http://www.liwc.net/>.

media differs conceptually from Volunteered Geographic Information (VGI) (Winter and Richter 2011; Harvey 2013), research into quality of VGI can prove useful. In comparison to traditional authoritative geospatial datasets, it has been argued the quality of VGI be considered in terms of credibility rather than accuracy (Flanagin and Metzger 2008). Approaches typically focus on deriving characteristics of the contributor, including their familiarity with the environment, experience, and value of past contributions (e.g. Keßler et al. 2009; Goodchild 2007). In VGI research, many of the computational approaches leverage the availability of the contributors spatial metadata, for example home locations (Bishr and Mantelas 2008) or more sophisticated mobility models (Mashhadi and Capra 2011), which are typically not available for micro-bloggers using services such as Twitter (Leetaru et al. 2013). However, evidence of a micro-blogger's familiarity with the environment might be present in social media such as Twitter. Goodchild (2009) defines ones action space as the locations in which daily life is played out, including places of home, work, school and leisure. The major places within an action space have been defined as where people spend long periods of time '...and people usually have an explicit name for them (home, work place, etc.)' (Schmid 2007, p. 656). Using social media data, the question of whether place name references reflect a micro-blogger's action space centred on their home location has been explored (Xu et al. 2013). This collective research underpins the idea that witnesses to many event types are likely to be within their action space (Gonzalez et al. 2008), and additionally, in the context of a social network of friends, may refer to informal places such as home, work, school rather than specific place names or addresses. Using these informal place names and categories suggests familiarity with the environment.

There is a significant history of research from numerous disciplines defining an event, including models relevant to geographic phenomena (e.g. Galton 2000). An event can be defined as 'something that happens at a given place or time' (Miller 1995), or as an occurrence (Worboys 2005). However, social media specific research into events does not always differentiate *real world events* from *virtual space events*. Boettcher and Lee (2012) disambiguate these concepts beginning with the definition of an event '...as a significant occurrence or happening that is

restricted in time' (Boettcher and Lee 2012, p. 358), and then differentiating virtual space events defined as those '...that are only relevant in the Twitter user community' (Boettcher and Lee 2012, p. 358) from real world events. Real world events are described as being further categorised into those which are global events, not restricted to a specific location, in contrast to local events. In geographic information science, a more specific scale of space would be defined, for example the definition of a perceptual scale, where vista space is the space which can be '...apprehended from a single place without appreciable locomotion' (Montello 1993, p. 315), which is complementary to the direct observation of a witness.

Theory

This section presents the development of concepts to identify (likely) witnesses.

Definitions

Critical to this research is the definition of witness. Common dictionary meanings of the word include a 'person who sees an event happening, especially a crime or an accident'.² From a journalism perspective, witnesses may be defined as 'people who see, hear, or know by personal experience and perception' (Diakopoulos et al. 2012 p. 2455). 'WordNet' defines witness to be 'someone who sees an event and reports what happens' (Miller 1995), which suggests expansion from being able to perceive an event to being able to provide a report. The use of the word 'seeing' could be interpreted literally; however, this research will not restrict observation to the visual, but expand the meaning to include direct observation of the event by any of a person's senses. This expansion is supported by the criminal justice system with 'earwitnesses' providing accounts of conversations overheard, and Australians' living in bushfire prone areas report the smell of smoke as it may be detectable long before a visual verification. In this research events are generally recognised as occurrences (Worboys 2005), and a local event in social media (Boettcher and Lee 2012), with scale defined from the vista space to the

² http://dictionary.cambridge.org/dictionary/british/witness_1?q=witness.

Table 1 Witness and related categories

Content categories for on-topic, individual, and original (OIO) micro-blogs			Terminology adopted	
Topic	Location	Time	Micro-blogger	Micro-blog
Direct observation	On-the-ground (OTG)	–	Witness	Witness account (WA)
Direct observation	OTG	Delayed	Witness	Delayed WA
Direct impact	OTG	–	Potential witness	Impact account (IA) <i>Other micro-blogs may be WAs</i>
Direct impact	OTG	Delayed	Potential witness	Delayed IA <i>Other micro-blogs may be WAs</i>
Relayed	Unknown	–	Relay of witness or potential witness <i>Connected micro-bloggers may be witnesses</i>	Relay account (RA) <i>Connected micro-blogger's micro-blogs, may be WAs</i>
Relayed	Unknown	Delayed	Relay of witness or potential witness <i>Connected micro-bloggers may be witnesses</i>	Delayed (RA) <i>Connected micro-blogger's micro-blogs, may be WAs</i>
Not witness impact or relay (NWIR)	Unknown	–	NWIR micro-blogger	NWIR Account (NWIRA)

geographic space (Montello 1993). The consequences of an event that can be witnessed by people are labelled ‘effects’. It is recognised that these effects may additionally be defined as events, that is a ‘phenomenon that follows and is caused by some previous phenomenon’ (Miller 1995). Witnesses are defined as people who directly observe the event or its effects and provide a report of these observations, their WA.

Table 1 presents the terminology adopted by this research for micro-bloggers and their micro-blogs, distinguished by the primary content categories of topic, location and time, for which Table 2 provides more details. Pre-requisite for establishing witnesses are the content is OIO. In addition to a direct observation, the content categories of direct impact and relay have been coined. A micro-blog is classified as an *impact account* (IA) when the content does not include a direct observation, but indicates the user is personally impacted by the event or is undertaking an action because they are impacted by the event. For example, micro-blogs can provide a status of how close a bushfire is or indicate the activation of evacuation plans. Users who send IAs are defined as potential witnesses. Potentially, other micro-blogs in their social media timeline might be WAs. The micro-blog is classified as *relayed* when the account is about a direct observation or impact of a person who is not

the micro-blogger. In the context of a social network, the micro-blogger might be relaying the observations of friends or family. An alternative to micro-blogging ‘I can see smoke’ is taking a picture with a phone and sharing with a social network. This micro-blogger fits the definition of a witness providing a WA. As with the textural content of an account, the linked content needs to be classified as being original and depicting an effect, to be categorised as a WA.

This research seeks to differentiate more probable WA from those which are less probable, and characterise those from which the status of a micro-blogger as a witness or potential witness can only be inferred. It assumes that accounts have not been maliciously fabricated. It should also be noted that witnesses can only be identified from their micro-blogs, and therefore, micro-bloggers might be witnessing the event and even micro-blogging on the topic, but will only be identified if their content contains witness characteristics that can be observed through human analysis.

Influence regions

As outlined previously, researchers have used a variety of mechanisms to identify micro-bloggers who are OTG, including manual analysis of a micro-blogger’s social media feed (e.g. Starbird et al. 2012), and automated analysis of metadata locations (e.g.

Table 2 Explanations of primary content categories

Topic	<p><i>Direct observation:</i> A subset of the OIO category, which includes a direct observation of the event or its effects</p> <p><i>Direct impact:</i> A subset of the OIO category, which includes a description indicating being impacted by the event and/or taking an action because of the event or its effects</p> <p><i>Relayed:</i> A subset of the OIO category, which includes a direct observation or indication of being impacted by the event, which is being relayed by the micro-blogger from knowledge of another person. The relay micro-blogger may or may not be OTG</p> <p><i>NWIRA:</i> A subset of the OIO category that does not include a direct observation, a direct impact or relay. The micro-blogger may or may not be OTG</p> <p><i>Overlapping categories:</i> Overlap between the categories <i>direct observation</i> and <i>direct impact</i> is expected. For example, a micro-blogger might indicate they can smell smoke and are evacuating</p>
Location	<p><i>OTG:</i> It is probable the micro-blogger is in the area affected by the event, and therefore, it is possible for the micro-blogger to witness the event by direct observation. For example, in a bushfire event a micro-blogger may need to look out the window to see smoke, or open the window to smell smoke. Being OTG doesn't automatically qualify a micro-blogger as posting a WA</p> <p><i>Unknown:</i> OTG or not OTG</p>
Time	<p><i>Delayed:</i> There are observables in the account which make it apparent the account was made some time after the observation, or the micro-blogger has moved since posting the account</p>

Kumar et al. 2013). This research complements the existing by inferring a micro-blogger is OTG, if it is determined they have provided a WA or IA. This research seeks to further refine OTG by the identification of an influence region, which is the region in which the micro-blogger can be inferred to be located, by the effect or impact they have reported.

Various events have different characteristics of their influence regions. For example, in a bushfire event a micro-blogger who reports seeing flames will need to be in a region in relatively close proximity to the event compared to a micro-blogger who reports seeing smoke. The boundaries of an influence region are considered to be vague in two ways. Spatially, in that the boundaries are indeterminate, and in actuality, in that it is uncertain which effects and impacts are considered to be caused by the event and which are

not. This research assumes the micro-blogger links their account to an event by keywords. An assumption is that the effect or impact that the micro-blogger is reporting reflects their proximity. For example, a micro-blogger seeing flames reports this, rather than seeing smoke because it is a more pertinent observation to share. If the micro-blogger seeing flames and smoke reports smoke, the micro-blogger will be placed in the larger influence region.

Typically spatio-temporal characteristics of many effects are defined to support prediction modelling for emergency response efforts, and in a real-life scenario these predictions would be adapted to accommodate evolving conditions. In conjunction with external data sources, a validity test of the observations being reported might be possible (e.g. Bishr and Mantelas 2008; Mashhadi and Capra 2011; Yanenko and Schlieder 2012). For example, as it is possible to infer that a micro-blogger is in close proximity to the event because they have reported seeing flames, it might also be possible to discredit such an account. In a bushfire, very few individuals would be in proximity to the fire, and if they were, they are likely to be defending their property and have limited opportunity to micro-blog about it. The number of individuals with the opportunity to observe and report smoke could be vast in comparison. Similar considerations can be established for other categories of events. A traffic accident can be in vista space, or the noise of the crash can be heard, or the queues of congesting cars behind the accident can be observed. An open air concert can be attended, or heard from a distance.

Place descriptions

This section outlines the concepts related to the characteristics of place name and place category references within the content of WAs and IAs.

Place names

Work in this paper will characterise how place names are used in WAs and IAs. In addition to place names, the use of formal and informal place categories will be explored, as to whether these are used more frequently by micro-bloggers who are OTG. For this purpose Table 3 provides descriptions of categories and terminology. Informal place categories are separated, as these may reflect more personal places such as neighbourhood rather than a broader environment.

Table 3 Explanation of place description categories

Terminology	Description
Place names	In addition to formal place names that might be found in the local official gazetteer ^a , abbreviated and vernacular forms are included
Place categories	Place name categories or types which might be found in the local official gazetteer. Abbreviated and vernacular versions of category names are included
Personalised place categories	When place categories are personalised using ownership words, or through context (e.g. ‘my suburb’)
Informal place categories	Informal place categories are distinguished from place categories in that these categories (or types) will not appear in a gazetteer (e.g. ‘home’)
Personalised informal place categories	When informal place categories are personalised using ownership words (e.g. ‘my home’)
Egocentric	In the absence of a place name or category, the user might use themselves as a reference object. This might be explicitly stated or implicitly

^a <http://services.land.vic.gov.au/vicnames/>

Place names are used to name an event

For many events, the names of the places in which they occur, naturally become the names which are used to identify the event (e.g., the ‘Kingslake Bushfire’, or the ‘Queensland Floods’), or vice versa (Chan 2014). In emergency response scenarios, these *event place names* will be broadcast widely by news and emergency services to provide updates to citizens. For OIO micro-blogs which use place names, an exploration will be made of the use of those which were widely broadcast and those which were not, to establish if any characteristics can be identified. Though many witnesses are expected to use these event place names, an exploration of those which do not, may provide unique observations of the event.

Corroboration

Existing methods to identify if micro-bloggers are OTG are used to corroborate WAs and IAs. Depending

on the end-use of harvested information and the extents of the effects of the event, different granularities and needs for redundancies might be considered. This will have particular implications for accounts that do not contain place names in their content. To enable comparison with previous research, and enable observations of appropriateness, each account will be compared with the metadata location and GPS locations when available, to form a matrix of corroboration. This exploration will provide an indication of what proportion of accounts cause issues due to inconsistent locations, and what scenarios might cause these. Additionally, an analysis will also be completed for those micro-bloggers who micro-blogged multiple times, after their individual micro-blogs have been categorised. It is expected a mix of account categories will be present.

Method

This section uses a case study to apply the concepts presented previously, and test them against the research hypothesis.

Event description

The event used for this case study was a bushfire which commenced at approximately 1 pm on Monday 18th February 2013, which was not a school or public holiday. The ignition point was identified on a rural road on the northern urban boundary of metropolitan Melbourne and progressed southwards towards more densely populated suburbs, where residents were advised to evacuate. The Hume Freeway, the main arterial road connecting Melbourne and Sydney was closed. The fire was attended by 175 fire-fighting personnel including volunteers and considered under control during Tuesday 19th February 2013. The fire burnt approximately 2040 ha, a number of buildings were lost but casualties were not reported. The event is described as a significant event,³ leading evening news bulletins in the state of Victoria.

³ As identified by the Fires Services Commissioner Victoria <http://www.firecommissioner.vic.gov.au/our-work/review/community-response-to-bushfires-during-201213-fire-season/>.

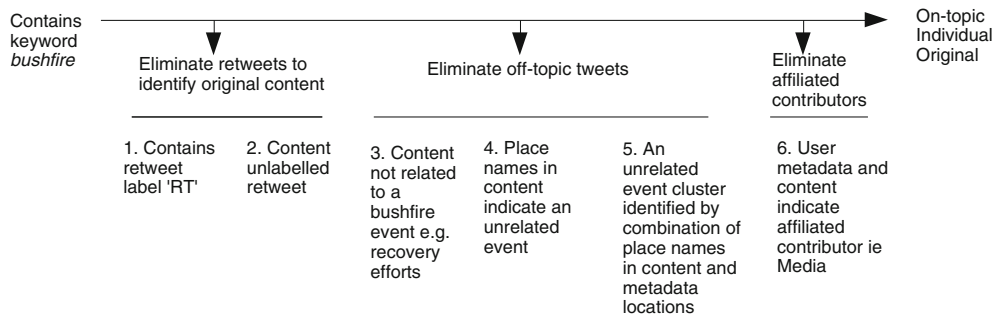


Fig. 1 Overview of the process to establish the OIO corpus

Data collection

Twitter is used as the source social media for this case study. Micro-blogs or ‘tweets’ were collected using the keyword ‘bushfire’, which is considered appropriate for the event of the case study. ‘Bushfire’ is used ubiquitously and predominantly in Australia in reference to uncontrolled fires outside urban areas, which enables the exclusion of uncontrolled fires outside of Australia, and urban building fires. As previous research indicates, only a small proportion of tweets come with GPS coordinates (e.g. Leetaru et al. 2013), and therefore, the more geographically specific keyword contributed to the ability to ensure the integrity of on-topic tweets.

The package of software tools described by Bruns and Liang (2012) and Bruns and Burgess (2011) was utilised. This includes a tweet retrieval and storage environment, which predominately uses the Twitter streaming API to collect tweets containing user configured keywords. This environment was running for the majority of the bushfire season for south-eastern Australia, and as such it collected tweets from the beginning of the event and as the event unfolded in real-time. Micro-blogger profile metadata was also collected for each micro-blogger that appeared in the tweet archive, with a script checking for new micro-bloggers every two hours.

Pre-processing to establish the OIO corpus

Each tweet in the archive was linked with the micro-blogger metadata. Manual processing on a number of passes was undertaken to establish a corpus of data with OIO tweet content. Figure 1 provides an

overview of the processing of each tweet. This was a process of elimination and does not guarantee that the tweets left are OIO. However, during this event there were no other significant bushfire events bordering suburban regions of Australia. This contributes significantly to the confidence that tweets without place names or metadata locations are likely related to the event of the case study.

Linked content

Once the OIO corpus was established, the linked content was processed, which involved reconstructing URLs, manual inspection and collection of any original content. Of the 461 OIO tweets, 102 had linked content, 95 of which could be stored for analysis. The content that could not be stored was either privacy protected or had been removed since the event. The content was then manually inspected to establish if it was original. In this case study it was common for tweets with original text content to link to mainstream news and emergency service websites, which is relatively simple to identify as an unoriginal source once the URLs have been reconstructed. However, there were instances where very compelling and professional photographs were linked on personal websites but not credited, thus requiring further investigation.

Coding of characteristics

The coding of the witness and related categories was through manual inspection of each tweet via a number of passes, by the author. The priority for each pass through the OIO corpus was the category

of interest in the following order, direct observation, a direct impact or action, and relays. Each tweet for each category was then further inspected and sub-categorised developed and applied for observable effects and impacts. Linked content identified as original, was further categorised as to whether it constituted a direct observation or not, for example smoky sky versus a screen grab of a mobile phone application. The resulting dataset is referred to as the reference dataset.

Coding process evaluation

Experiments to evaluate the coding process, by testing whether agreement could be stated for the reference dataset were completed. Procedures involved the participants completing three experiments, with 10 % of the OIO corpus extracted to support training tasks. However, it must be noted that the training exercises could not replicate the familiarity the author has with the corpus. The first experiment required the participant to read each tweet in the OIO corpus and code WA and IA that they identified in separate passes. Testing of the experimental procedures revealed that fatigue with the categorisation tasks could be reduced if the participants focused on a single category at a time, and completed experiments over a number of days rather than a single sitting. Additionally, the limited number of RA proved insufficient to enable adequate training or justify a third pass through the corpus. The second and third experiments were to read each WA and IA from the reference dataset respectively, and code the sub-categories defined. Testing of the experimental procedures revealed difficulties in training participants to identify the sub-categories *other*, however, these were included and their influence is discussed further in the results. Two participants completed the experiments, and can be described as native English speakers and Australian residents since childhood.

Results

In this results section, where complete or near complete tweet text or images from linked content are presented, the full URL is provided in footnotes. However, where ‘snippets’ of tweet text are presented the full URL of the source is not provided, and the

Table 4 Summary results for each primary category

Category	# Accounts	% of OIO	# Delayed
NWIRA	159	34	–
WA	198	43	11
IA	94	21	9
RA	10	2	1

authors can provide further details on request. The root source of snippets is <https://twitter.com>, access date Monday or Tuesday 18th and 19th February 2013.

Summary

461 tweets representing OIO were differentiated. Table 4 presents a summary of results for the primary categories defined in Table 1. Accounts classified as delayed only include those that could be identified from the text content alone. With only ten tweets, RAs were not a large category, but distinguishable with careful consideration. References to personally known people (e.g. ‘son’, ‘mum’, ‘dad’, ‘cousin’, ‘family’), and the impacts these people are experiencing dominated the category with eight of the ten accounts. As this category is small, absolute numbers rather than percentages are presented throughout this paper.

Witness accounts

As outlined in Table 5, in this case study WAs are dominated by observations of *smoke* (77 %) which is not unexpected due to the potential spatial extent of this effect. A total of 234 effects were identified, with 34 WAs coded with two effects and one with three effects. Examples of explicit and implicit sensing were present in the majority of sub-categories. Explicit observations of seeing smoke dominated during the day, with smelling smoke at night. Nine accounts report observation of a *bushfire moon*, which is assumed from descriptions to be an observation of what the moon looks like when smoke is in the sky at night. A number of WAs were additionally identified to include impact descriptions. For example, three WAs also refer to evacuation and numerous provide spatial descriptions indicating the bushfire is near to them or indications of being fearful.

Table 5 Summary of bushfire effect categories

Effect	# Effects	% WAs	Example tweet content including snippets
Smoke	153	77	<i>haze, bushfire forming clouds, orange sunshine, sky was black, look of bushfire, orange tinge, orange hue, grey/brown clouds, dark cloud, thick plume, ominous smell, strong burning smell</i>
Ash	5	3	<i>Wow. I was just communing with my giant lime tree in inner urban Melb; it's covered in bushfire ash. Scary^a</i>
Emergency vehicles	6	3	<i>@Asher_Wolf @jennynorton Bushfire? I can hear the whirlybirds right now^b</i>
Road closures	1	<1	<i>#dannybrook fire seen from closed Hume Freeway #thomastown #epping #bushfire http://t.co/eeVvJ9Ka^c</i>
Traffic conditions	5	2	<i>30 min to get out of the #Epping Plaza car park. It's a car yard #bushfire^d (only example not on a road)</i>
Bushfire moon	9	5	<i>Very red bushfire moon outside, beautiful terrible bushfire moon tonight, #moon...devoured by smoke</i>
Linked content	48	24	<i>#bushfire #Donnybrook http://t.co/c1hMwiJW^e #smoky #bushfire in nearby Epping http://t.co/FbjzBUf^f</i>
Other	7	4	<i>Drove past the bushfire in donnybrook/epping on the way back from Benalla. Looks bad^g</i>

^a https://twitter.com/Kristen_Boschma/status/303780135373512704. Access date 18 February 2013

^b <https://twitter.com/InterzoneRebels/status/303390158005342208>. Access date 18 February 2013

^c <https://twitter.com/chriscorneschi/status/303411867542507520>. Access date 18 February 2013

^d https://twitter.com/Craigsta90_10/status/303390521127211008. Access date 18 February 2013

^e http://twitter.com/JayBull_95/status/303386818030751744. Access date 18 February 2013

^f <https://twitter.com/om4james/status/303368575769522176>. Access date 18 February 2013

^g https://twitter.com/Reynolds_R/303413868825296897. Access date 18 February 2013

Linked content

Of the 95 OIO tweets with linked content that could be collected, 59 were categorised as linking to original content, and of these 48 were categorised as direct observations of the event, and therefore considered WAs. 34 of these WAs also had textual content which stand-alone would have been categorised as a WA. Almost all are photographs of smoke in the sky, with a number of examples of recognisable buildings in the skyline (see Fig. 2). The only images reporting traffic related effects are also shown in Fig. 2. No linked content was differentiated in the corpus of actual fire-fighting, though uncredited examples were identified through comparison of images published by the media, shown in Fig. 3. There were a number of daytime images posted at night (determined via timestamp) which were not previously coded as delayed. This suggests that more delayed accounts exist than can be identified from the text content alone.

Impact accounts

IAs are dominated by a category which has been named *event near me* because as the example in Table 6 indicates, this is what the content communicates. This content category appears similar to the relative location referencing described by Vieweg et al. (2010), but with approximately 12 % of OIO in this case study a greater proportion. For the *evacuation* category, potential witnesses considering their evacuation or 'Bushfire Survival Plans' were identified, as were those who had already evacuated. Similarly, content indicating plans might be affected, or had been affected were categorised under *plans change*, mostly travel plans. With a larger corpus it might be possible to further differentiate categories of anticipated impacts, from impacts which have already occurred. Ten of the *other* impact accounts were distinguished because of an apparent heightened emotional state alone. However, this can be observed in accounts across all categories in a variety of ways. Two



Fig. 2 Linked content reporting: **a** traffic congestion on the Hume Highway (https://twitter.com/holly_yeatman/status/303394166686224384/photo/1. Access date 6 March 2013), **b** closure of the Hume Freeway (<https://twitter.com/chriscoreschi/status/303411867542507520/photo/1>. Access date 6

March 2013), **c** smoke from a suburban backyard (<https://twitter.com/nmg75/status/303408200122769408>. Access date 6 March 2013), **d** smoke from Melbourne's central business district (CBD) (<https://twitter.com/taitems/status/303351685848379392>. Access date 6 March 2013)



Fig. 3 Example of linked content identified as uncredited media source, and therefore not a WA (<https://twitter.com/SukhSandhu/status/303398640167288832/photo/1>. Access date 6 March 2013)

tweets imply that the potential witnesses are involved with *emergency response* activities. On further investigation, it can be confirmed that one of the users is a volunteer fire fighter, and the other cannot be discredited.

Coding Process Evaluation

Table 7 presents the results of the first experiment to evaluate the coding process for primary WA and IA categorisation. These results indicate agreement, in particular for WA, and therefore, the validity of the methodology employed. Difficulty in training participants in *other* sub-category tweets (refer to Table 5 and Table 6 for WA and IA respectively) was reported, leading to the investigation of results with

their inclusion and without. Their influence is more pronounced for IA, due in part to their larger proportion of this smaller category.

Participant A and B achieved 97 % and 98 % agreement on average with the WA sub-categories (refer to Table 5) in experiment two, and 80 % agreement on average with the IA sub-categories (refer to Table 6) in experiment three, or 84 % agreement when the *other* sub-categories are excluded. Of note, Participant A commented they had not coded all references to a 'Bushfire Survival Plan' as belonging to the *evacuation* sub-category because the plan might be to 'stay and defend'⁴ rather than evacuate, reducing their agreement for this sub-category to 67 % and highlighting the importance of robust definitions. For the sub-category of IA with a significant number of examples to support training, *event near me*, Participant A and B achieved 88 and 81 % agreement respectively. This acceptable but lower figure than the WA sub-categories may reflect a reliance on expertise in spatial science, whereas WA sub-categories did not require such expertise.

Place descriptions

This section presents results for place descriptions contained within the different categories of accounts.

⁴ <http://www.cfa.vic.gov.au/plan-prepare/stay-and-actively-defend/defending-your-property.pdf>.

Table 6 Summary of bushfire impact or action categories

Impact	# Impacts	% IAs	Examples
Event near me	57	61	<i>#Bushfire that is only a 15 min drive away from my house is scary. Lucky we are East of it^a</i>
Evacuation	16	17	<i>hmm...i wonder if a bushfire near my house and consequent evacuation is a good enough excuse to not do homework^b</i>
Plans change	6	6	<i>Had to cancel my last home visit of the day due to a bushfire. http://t.co/xUe5et82^c</i>
Emergency response	2	2	<i>Currently on the back of the fire truck on the way to the #Epping #bushfire as a relief crew! It should be a day of just blacking out^d</i>
Other–emotion	10	11	<i>OMFG BUSHFIRE #craigieburn help:(http://t.co/DYw2wGrJ)^e</i>
Other	10	11	<i>ThankU to those that were thinking of me & my fellow Melbournians thru yesterdays bushFire! The CFA worked thru the night keeping us safe!^f</i>

^a https://twitter.com/lovejess_xo/status/303403252777644032. Access date 18 February 2013

^b <https://twitter.com/BavleenKaurKK/303428975932362752>. Access date 18 February 2013

^c https://twitter.com/curran_joanna/status/303444057533849601. Access date 18 February 2013

^d https://twitter.com/jimmy_reade/status/303578955561185280. Access date 18 February 2013

^e <https://twitter.com/LukeDamnBrooks/status/30339346357408563>. Access date 18 February 2013

^f <https://twitter.com/LoyalLonz82/303627888853409792>. Access date 19 February 2013

Table 7 Coding process evaluation results for primary categorisation

Participant	WA (<i>other excluded</i>)		WA		IA (<i>other excluded</i>)		IA	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
A	1	0.92	0.94	0.90	1	0.86	0.75	0.76
B	1	0.84	0.96	0.81	1	0.85	0.77	0.77

Place names

Place names and categories described in Table 3 were coded and results presented in Table 8. The categories defined as *personalised place categories*, *informal place categories* and *personalised informal place categories* described in Table 3 have been combined in Table 8 under the heading *personal place category*. This decision was made primarily due to the small sample size. The WAs and IAs in combination contain fewer place names, 47 % compared to 69 %, and more place categories or personal place categories, in combination 34 % compared to 5 %. These differences are more pronounced for certain categories such as *event near me*. Additionally, it can be noted that egocentric spatial descriptions are not present in NWIRA, which reflects that micro-bloggers who tweet from a personal perspective appear to be credible witnesses. The figures in Table 8 are the number of accounts that include place categories, not a count of instances within the tweet. There are very few examples observed where more than one instance of a

non-place name was within a single tweet. However, a list of multiple place names within a single tweet is more common. Table 9 provides example tweet content and further observations.

Place names used to name the event

An event such as the case study is widely broadcast via the news media. Below are two of the first WAs identified from Twitter for the case study event:

1. *yo does anyone know what's up with the haze coming from the north of melbourne? smells like bushfire. can't find any news online.*⁵
2. *Hmm what looks to be a sizeable bushfire off in the north-east? <http://t.co/jRIk5upX6>*⁶

⁵ <https://twitter.com/tetzlol/status/303350066230460416> Access date 18 February 2013.

⁶ <https://twitter.com/taitems/status/303351685848379392> Access date 18 February 2013.

Table 8 Place name and category summary

Account category	# accounts	Place name (%)	Place category (%)	Personal place category (%)	Ego-centric (%)	None (%)
NWIRA	159	69 %	4 %	<1 %	–	27 %
IA and WA	292	47 %	11 %	23 %	11 %	23 %
WA	198	52 %	12 %	16 %	8 %	25 %
IA	94	36 %	9 %	38 %	18 %	18 %
RA	10	3	0	6	0	1
E. near me	57	30 %	11 %	58 %	30 %	0 %

Table 9 Place name and category comments and example snippets

Category	Example content snippets and related interpretations
Place names	Vernacular, abbreviations and hashtag creations were all coded as place names. Examples include: <i>straya, melbs, #meanwhileinAustralia</i>
Place categories	Examples include: <i>the highway, the hospital, the road, the city</i>
Personal place categories	Place categories were personalised explicitly or implicit in context. Examples include: <i>my suburb, my area, school, campus, my town, suburb near mine</i> Informal place categories were personalised explicitly or implicit in context. Examples include: <i>home, office, bedroom, work, my house, my room, my place, where I live, classroom, #backyard</i> Did not code <i>home</i> when referenced in the collective, for example <i>residents can now return to their homes</i>
Egocentric	Both explicit and implicit egocentric references coded. Examples include: <i>near me, near us, nearby, close by, from here</i>

These accounts provide evidence that it might be difficult to determine from observing smoke in the sky which places are on fire. For the general majority, the names of the places affected might only become known from observing mainstream news. Identifying which accounts do not contain widely broadcast place names might provide further evidence supporting WAs. Table 10 lists which place names were identified as being widely broadcast and therefore eliminated and Table 11 lists lesser used place names which remained for WAs and IAs versus NWIRA. Lesser used place names for IAs and WAs are mostly suburb

names whereas NWIRAs appear to report more precise information. But when these accounts are investigated, it appears that each of these can be attributed to another source such as the media rather than the individual (and therefore, why they were eliminated as WAs at processing time). Five of the seven road names presented in Table 11 can be attributed to a single micro-blogger who is identified as the most prolific in the corpus. This micro-blogger is discussed further in Sect. “Corroboration”.

Influence regions

Two methods were used to explore influence regions between the event and witnesses. Figure 4 maps the place names presented in Table 11, corresponding to WAs that reported smoke or traffic conditions. Google Maps Engine was used to geocode place names and create the visualisation, with point icons representing each place name as Google Maps suggests. In this small corpus, smoke and traffic conditions were the only categories with enough content to be suitable for this visualisation. The number of reports, and the potential spatial extents and proximity to the event which can be derived from this content, appear to fit the model described in Sect. “Influence regions”. Figure 4 also shows what is interpreted to be a boundary of ‘metropolitan Melbourne’ as presented by Google Maps in the search result for ‘Melbourne, VIC, Australia’.⁷ The fire extents are as depicted by Emergency Services⁸ during the event.

⁷ <https://www.google.com/maps/place/Melbourne+VIC/@-37.8602828,145.079616,9z/data=!3m1!4b1!4m2!3m1!1s0x6ad646b5d2ba4df7:0x4045675218ccd90> Access date 17 March 2014.

⁸ <http://emergency.vic.gov.au/map#now> Access date 18 February 2013 approximately 21:00.

Table 10 Place names identified as event place names or highly publicised place names

Suburbs: *Donnybrook, Epping, Wollert, Craigieburn*

Roads: *Hume Highway, Hume Freeway (road closure), Donnybrook Road (ignition point), Cooper St (containment line)*

Other: *The Northern Hospital, Epping Plaza, Aurora Estate (places related to evacuation advice)*

Table 11 Place names other than the event place names or highly publicised place names present in the corpus

More unique place names in WAs and IAs	37 place names, representing 27 % of IA and WAs with place names <i>Suburbs: Campbellfield, Benalla, #thomastown, Doncaster east, #3073, south morang, Thomastown, Hampton Park, Hawthorn, Northcote, Templestowe, Reservoir, Braybrook, North Melbourne, #melbourneCBD, Footscray, Kensington/Flemington (not a complete list)</i> <i>Roads: sydney rd</i> <i>Other: Victoria park, The Pines Shopping Centre, Merri Creek, city of whittlesea, MCG</i>
More unique place names in NWIRA	23 place names, representing 21 % of NWIRA with place names <i>Suburbs: Geelong, Thomastown, Thomastown, #northmelbourne, mount Waverley, Mount Waverley, #campbellfield, Sydney Epping, Greensborough, Campbellfield, Campbellfield, Hawthorn, Thomastown</i> <i>Roads: High St Woodstock, McDonalds Rd South Morang, High St, Kings Drive Lalor, ohearns st, Craigieburn Rd East, Vearings Rd</i> <i>Other: BP on Hume Hwy Cooper St Exit, melb airport</i>

The *event near me* category is characterised and differentiated by content that appears to directly define an influence region using spatial descriptions, which when generalised communicate *bushfire near me*. The qualitative spatial relation *near* is specified in a variety of ways, with varying levels of precision: some simply used the words *near* or *close*, ten accounts were in units of time (e.g. *10–15 min drive*), five accounts in units of distance (e.g. *not even 10 km away*), and five accounts in units of suburbs (e.g. *neighbouring suburb*).

Corroboration

Table 12 presents a summary of the metadata locations for the OIO corpus. Note these statistics were calculated by the number of accounts, not the number of unique micro-bloggers. They compare broadly with those presented in Hecht et al. (2011), however, differences can be noted as to the granularity of the valid geographic entries, perhaps in part due to population density differences between Australia and USA. Potential witnesses creating IAs have significantly fewer valid geographic entries.

Table 13 presents corroboration results using metadata location or GPS. The metadata location is considered not suitable for corroboration if it is blank, contains a non-geographic entry, or valid geographic

entry that is not specific enough. For this event, the granularities considered suitable were from the most precise GPS coordinates to metropolitan area (see Fig. 4 to visualise, for example, ‘metropolitan Melbourne’). In the majority of cases, the metadata location corroborates WAs and IAs, and most usefully for 60 and 29 accounts respectively, where place names were not present in the content. Seven WAs and IAs were identified which were not consistent, and these were explored further. Five were confirmed as WAs and IAs even though their metadata did not indicate Melbourne, and two were identified as false positives. The majority of NWIRA was also from micro-bloggers disclosing their location to be Melbourne, confirming that micro-bloggers can be micro-blogging about an event but not provide evidence that they are witnesses or are impacted. Another interpretation of these results might be that the event was primarily a significant local event resolved in less than 1 day, and therefore primarily of local interest only.

Table 14 outlines the number of micro-bloggers who sent multiple accounts related to the event, and indicates how many could be described as *mixed* categories or *single* categories. A single category is assigned if all the micro-bloggers accounts are IAs and WAs, or are NWIRAs, and a mixed category is assigned if a combination of WAs and IAs and

Fig. 4 An approximation of influence regions for smoke effects and traffic effects using WA which contain place names that are not event names or highly publicised place names, with fire extents and ‘metropolitan Melbourne’

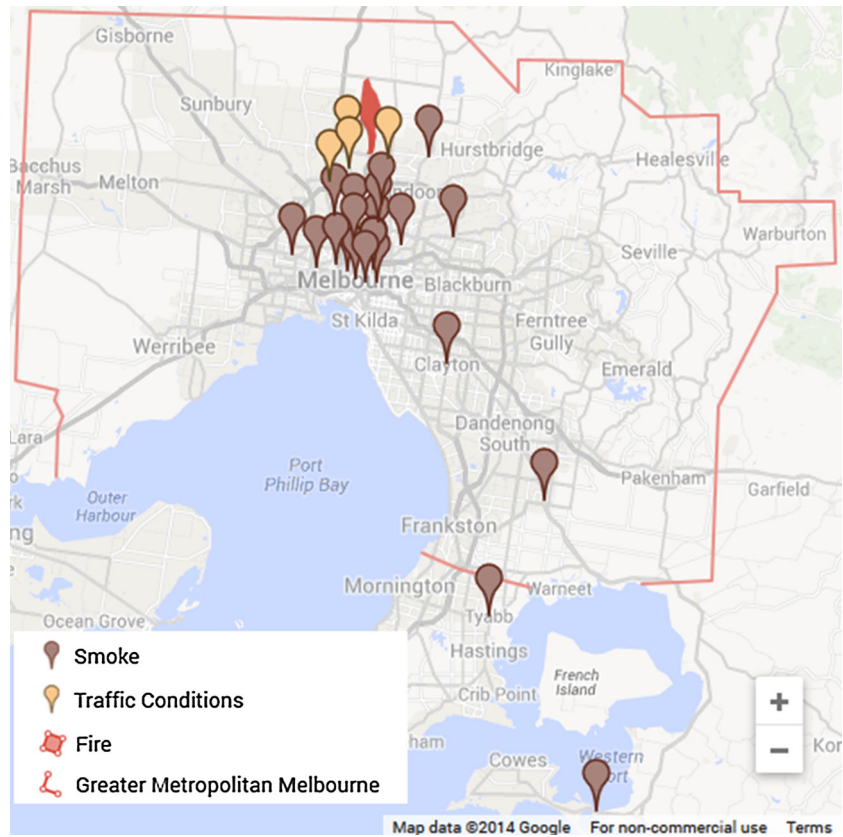


Table 12 Summary of metadata locations including breakdown by granularity for valid geographic entries

	Summary			Granularity for valid geographic entries					
	Blank	Non-Geo	Valid Geo	Country	State	Greater Metro	Suburb	Street	GPS
OIO	11 %	12 %	77 %	14 %	2 %	80 %	1 %	<1 %	<1 %
WA	8 %	9 %	83 %	15 %	1 %	81 %	2 %	<1 %	<1 %
IA	17 %	20 %	63 %	11 %	3 %	80 %	2 %	2 %	2 %
RA	0	2	8	2	1	3	2	0	0
NWIRA	14 %	8 %	78 %	15	2 %	73 %	7 %	0	2 %

NWIRAS is detected. All micro-bloggers categorised as mixed categories were further investigated and verified, with only two requiring additional comment. The first micro-blogger had five accounts all NWIRAs, except one WA which is consequently identified as a false positive. The credibility of this micro-blogger was already in question as the example in Fig. 3 indicates. For the second micro-blogger who contributed over 16 tweets, the context provided, when

considering all tweets collectively does bring into question the categories previous tweets were given. This suggests that as an event is progressing, the status of each micro-blogger might be reviewed, when a certain number of tweets are reached. Additionally, this micro-blogger provides an example that a person can legitimately be a witness to parts of the event or be personally affected, but not necessarily be a witness to other parts of the event.

Table 13 Corroboration matrix indicating whether metadata or GPS locations corroborate the account

	Metadata L. Corroborates?			GPS Corroborates?		
	Yes	No	Not suit.	Yes	No	Not suit.
Witness account	134	5	59	2	0	196
Without place names	60	0	38	–	–	96
With linked content	35	0	13	1	0	47
Impact account	48	2	44	4	0	90
Without place names	29	2	29	3	0	57
Relay account	5	0	5	–	–	10
Without place names	2	0	5	–	–	7
NWIR account	84	19	56	2	2	155
Without place names	28	4	17	–	–	49

Table 14 Summary of micro-bloggers by the number of accounts, and category of accounts contributed

# Accounts	1	2	3	4	5	6	7	8	16
# Micro-bloggers	336	26	5	1	2	1	1	1	1
# Single account category micro-bloggers	–	22	2	0	0	0	1	0	0
# Mixed account category micro-bloggers	–	4	3	1	2	1	0	1	1

Discussion

This case study was suited to test the hypothesis and models developed, despite being based on a small dataset and a single event. Manual analysis of the corpus enabled an in-depth exploration of the characteristics of the micro-blogs related to the event, and avoided the challenges of automatic interpretation which when dealing with unstructured social media data can be significant. Additional advantages of this approach and the event selected include the ability to incorporate local knowledge for the event area, the ability to relatively define temporal and spatial characteristics of the event, and the reasonable certainty of the integrity of the OIO corpus in the absence of absolute data such as GPS. Disadvantages include potential bias being introduced because of relying on human interpretation and that the model was applied to a single type of event, which might introduce characteristics not applicable to a general model. Additionally, it did not support an in-depth exploration of relayed or delayed accounts characteristics.

For this event, witnesses could be differentiated and categorised by the effects they reported observing. The

effects were dominated by *smoke* (77 % of WAs). As smoke can be detected over vast areas, it might be expected, that other types of events may have proportionally fewer witnesses. Micro-bloggers reported their direct observations in either explicit (e.g. ‘I see smoke’) or implicit ways (e.g. ‘thick orangey haze’). Additionally, subtle changes in language could be detected as day became night, where the sense of *smell* became more prominent. These findings provide clues to the challenges of future automatic interpretation to differentiate WAs. The categories of effects and sensing need to be identified and interpreted in both explicit and implicit forms. For two of the effect categories, smoke and traffic congestion there were sufficient WAs with unique geocodable place names to support a visualisation of the influence regions. Additionally, it was observed that a significant number of WAs and IAs contained spatial descriptions, and in particular the category *event near me*. Potentially, these may be formalised (e.g. Vasardani et al. 2013), to enable the refinement of influence regions for particular witnesses and potential witnesses.

The category of IA was created to include those micro-blogs which described how the micro-bloggers were directly impacted or what actions they were

considering. Though these micro-blogs do not include direct observations, it can be inferred that they may be in areas where effects would be experienced. Additionally, for some IA categories, it might be inferred that the potential witnesses could be in closer proximity than if they reported a direct observation, an example being IAs of evacuation compared to WAs of smoke. Impacts were not distinguished from actions in this research, nor were intentions to undertake an action distinguished from actions already completed. This categorisation is not always apparent in micro-blogs, but a larger corpus might enable further exploration of whether possible and beneficial.

Additionally, such categorisation might be potential WAs and IAs are less likely to contain place names—47 % compared to NWIRAs with 69 %. This was not because they included fewer spatial descriptions, but because they included more personal place categories, with 23 % compared to less than 1 % for NWIRAs. This finding suggests that likely witnesses and potential witnesses use personal place categories instead of place names, at least for a bushfire event, or perhaps more generally for events that occur in people's action spaces with which they have familiarity. Additionally, within the context of a social network, the environmental context may not need to be explicitly stated. An analysis of place name presence, in accounts not widely broadcast by the media or emergency services, revealed both IAs and WAs versus NWIRAs were comparable, with 27 % versus 21 % respectively. However, on initial observation, the NWIRAs appeared to contain more street names and landmarks compared to suburb names. Further investigation revealed that in general, though the textual content for these NWIRAs were original, the information they contained could be attributed to other sources. Whereas the place names in the WAs and IAs reflected unique perspectives of the event, with suburb names the preferred granularity level in an urban environment to report the effect of smoke.

Though some researchers have been identified to use, as this research did, manual analysis to determine the likely geographic location of micro-bloggers (e.g. Starbird et al. 2012), the majority rely on the metadata location (e.g. Kumar et al. 2013). 64 % of metadata locations were identified as being from a granularity of GPS coordinates to metropolitan Melbourne, which was deemed suitable for a corroboration exercise for this event. 60 WAs and 29 IAs without place names in

their content were associated with corroborating metadata locations, and only seven were not. These were investigated further and five were found to be legitimate accounts, although the metadata location suggested otherwise, and two were confirmed as false positives. Dependent on the intended end use of the information sourced from social media, future work might seek redundant location data for corroboration at the granularity of the effect or impact, applying greater scrutiny. For example, a report of a road block would not be considered corroborated unless it could be grounded with a location at that granularity. For micro-bloggers who had posted multiple accounts, additional investigation was completed to ascertain the level individual accounts corroborated each other or not. The overarching outcome of this analysis was confirmation that it is legitimate to have micro-bloggers sending accounts categorised as WAs, IAs and NWIRAs. For example, a micro-blogger might post accounts indicating they see smoke is in the sky of a neighbouring suburb, then that they might need to evacuate, then good wishes and luck to Melbournians. As events can span beyond the vista space it is valid to find that a single micro-blogger can be a witness for some effects but not others.

Although formal analysis was not undertaken on the linguistic style of the text content of the micro-blogs, it was observed that WAs and IAs seemed more credible, more believable, if they were personal and informal in style, as defined by Verma et al. (2011). Accounts which were formal and objective were often more difficult to believe as to be unique and created by an individual, because the style is similar to that which emergency services and the mainstream news media would use. This is especially so when uncredited retweets, uncredited linked content, and accounts made by users who are micro-blogging about what they see on TV were identified. Contributing to the perception that the style of WAs and IAs is more informal might be the inclusion of personal place categories.

The accounts were categorised as to the primary category of interest, firstly a WA, then an IA or RA. It was observed in a small number of cases, WAs also made reference to evacuation, though it was more common that two effects are found in a single tweet. Only one account was identified to have three effects reported. With a 140 character limit for micro-blogs in Twitter, there is probably a limit to how much can be

communicated. Considering this overlap might be useful from a corroboration perspective, because the influence regions of multiple effects, impacts or combinations in a single micro-blog could be tested to be consistent with each other or not. Emotion is an exception. It is observed that the WA and IA categories do appear to have an elevated level of emotion. Although ten IAs were distinguished as being IAs for this reason alone, it appeared that emotion was elevated for all accounts, though this was not the subject of formal analysis.

Conclusion and future work

From a range of disciplines, previous research has stated a relationship between a micro-blogger's proximity to an event and the relevance of their contributions, and a relationship between witnesses and credibility. But due to location information sparsity for social media including Twitter, it is only possible for the smallest minority of micro-bloggers to establish their proximity to an event in absolute terms. Consequently, this research sought to establish if it is possible that likely witnesses could be differentiated by assessing the content of their micro-blogs. A defining model of WAs, and related IAs and RAs were established. The hypothesis was supported by the results of this research. Likely WAs could be differentiated and categorised by the effect the micro-blogger reported as their direct observation of the event. Direct observations of numerous effects were identified from the case study bushfire event with smoke dominating. Observations of traffic congestion, road closures and emergency response vehicles were also reported by witnesses and categorised. The witnesses often reported explicitly the sense used to make their observations, especially for observations of smoke. IAs could also be differentiated and categorised, with reports of having undertaken or intending to evacuate, and volunteer fire fighters travelling to and from the event identified. However, the dominate category observed was named *event near me* as this is what the potential witnesses reported, using a variety of qualitative spatial relations and personal place categories. Influence regions for the case study event type could be visualised based on geocoding content present in two sub-categories of WAs, smoke and traffic conditions. There were fewer place names in

WAs and IAs for the case study event type, which was not because they had less spatial descriptions, but because of the increased presence of personal place categories. This may suggest that witnesses are reporting from their action spaces with which they have familiarity, and/or within the context of a social network, the environmental context may not need to be explicitly stated.

Many avenues for future research have been identified. Though this case study enabled initial models of witness categories to be defined and explored, other events in terms of a larger corpus size and differing event types is required for further model testing and refinement. Specific challenges identified for the automatic interpretation of witness and related accounts may be pursued. Particular challenges are the identification and interpretation of categories of effects and sensing in both explicit and implicit forms. Expansion of testing to different event types will also provide direction on when the use of place names and categories found in this research can be assumed for other event types. Additionally, it would be beneficial to expand analysis of witness characteristics from a single micro-blogger and micro-blog, to complete timelines, including off-topic content. A more formal exploration of the linguistic style of WAs, might provide important contributions for research focused on identifying relevant or actionable content for situational awareness. Finally, exploration on whether spatial descriptions—which are present in a significant proportion of WAs and IAs—can be formalised, may generate significant spatial intelligence in addition to the refinement of influence regions.

References

- Bishr, M., & Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal*, 72(3–4), 229–237.
- Boettcher, A., & Lee, D. (2012). EventRadar: A real-time local event detection scheme using twitter stream. In *IEEE international conference on green computing and communications, conference on internet of things, and conference on cyber, physical and social computing* (pp. 358–367).
- Bruns, A., & Burgess, J. (2011). Gawk scripts for twitter processing. Mapping Online Publics (1.0 ed.). <http://mappingonlinepublics.net/http://mappingonlinepublics.net/resources/>.

- Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4).
- Burgess, J., Vis, F., & Bruns, A. (2012). How many fake Sandy pictures were really shared on social media? Datablog Facts are scared. <http://www.theguardian.com/news/datablog/2012/nov/06/fake-sandy-pictures-social-media>. The Guardian.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *20th international world wide web conference*, Hyderabad, India (pp. 675–684). New York: ACM.
- Chan, C. (2014). Leveraging Twitter to detect event names associated with a place. *Journal of Spatial Science*, 59(1), 137–155.
- Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, Toronto, Canada (pp. 759–768). ACM.
- Cheong, M., Ray, S., & Green, D. (2012). Interpreting the 2011 London Riots from Twitter Metadata. In *12th International conference on intelligent systems design and applications (ISDA)* (pp. 915–920).
- De Longueville, B., Smith, R. S., & Luraschi, G. (2009). “OMG, from here, I can see the flames!”: A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 international workshop on location based social networks* (pp. 73–80). New York: ACM.
- Diakopoulos, N., De Choudhury, M., & Naaman, M. (2012). Finding and assessing social media information sources in the context of journalism. In *The ACM SIGCHI conference on human factors in computing systems*, Austin, Texas (pp. 2451–2460). New York: ACM.
- Eisenstein, J., O’Connor, B., Smith, N. A., & Xing, E. P. (2010). A latent variable model for geographical lexical variation. In *The 2010 conference on empirical methods in natural language processing*, MIT, Massachusetts (pp. 1277–1287).
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3–4), 137–148.
- Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Conference on human factors in computing systems* (pp. 80–87). ACM, New York.
- Galton, A. (2000). *Qualitative spatial change*. Oxford: Oxford University Press.
- Gelernter, J., & Balaji, S. (2013). An algorithm for local geoparsing of microtext. *Geoinformatica*, 17(4), 635–667.
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221.
- Goodchild, M. F. (2009). The quality of geospatial context. In K. Rothermel, D. Fritsch, W. Blochinger, & F. Durr (Eds.), *Quality of context* (pp. 15–24). Berlin: Springer.
- Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth*, 3(3), 231–241.
- Harlow, S. (2011). Social media and social movements: Facebook and online Guatemalan justice movement that moved offline. *New Media & Society*, 14(2), 225–243.
- Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowd sourced geographic information. In D. Z. Sui, S. Elwood, & M. F. Goodchild (Eds.), *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice* (pp. 31–42). Berlin: Springer.
- Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber’s Heart: The dynamics of the “Location” field in user profiles. In *The ACM CHI conference on human factors in computing systems* (pp. 237–246). ACM, Vancouver.
- Kent, J. D., & Capello, H. T., Jr. (2013). Spatial patterns and demographic indicators of effective social media content during the Horse thief Canyon fire of 2012. *Cartography and Geographic Information Science*, 40(2), 78–89.
- Keßler, C., Janowicz, K., & Bishr, M. (2009). An agenda for the next generation gazetteer: geographic information contribution and retrieval. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, Seattle, WA, USA (pp. 91–100).
- Kumar, S., Morstatter, F., Zafarani, R., & Liu, H. (2013). Whom should I follow? Identifying relevant users during crises. In *24th ACM conference on hypertext and social media*, Paris, France (pp. 139–147). New York: ACM.
- Leetaru, K. H., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18(5).
- Lieberman, M. D., & Samet, H. (2012). Adaptive context features for toponym resolution in streaming news. In *Proceedings of the 35th international ACM SIGR conference on research and development in information retrieval*, Portland, OR (pp. 731–740). New York: ACM Press.
- MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Saveliev, A., Mitra, P., et al. (2011). SensePlace2: GeoTwitter analytics support for situational awareness. In *2011 IEEE conference on visual analytics science and technology* (pp. 181–190).
- Mashhadi, A. J., & Capra, L. (2011). Quality control for real-time ubiquitous crowdsourcing. In *Proceedings of the 2nd international workshop on ubiquitous crowdsourcing Beijing* (pp. 5–8).
- Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis: can we trust what we RT? In *Proceedings of the first workshop on social media analytics* Washington, DC, USA (pp. 71–79). New York: ACM.
- Miller, G. A. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Mirkinson, J. (2012). CNN, weather channel falsely report NYSE flooding during hurricane sandy. http://www.huffingtonpost.com/2012/10/29/cnn-weather-channel-nyse-flooding-sandy_n_2042415.html.
- Montello, D. R. (1993). Scale and multiple psychologies of space. In A. U. Frank, & I. Campari (Eds.), *Spatial information theory a theoretical basis for GIS* (pp. 312–321,

- Lecture Notes in Computer Science, Vol. 716). Berlin/Heidelberg: Springer.
- Ostermann, F. O., & Spinsanti, L. (2010). A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. In *The 14th AGILE international conference on geographic information science*, Utrecht, The Netherlands.
- Poser, K., & Dransch, D. (2010). Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica*, 64(1), 89–98.
- Ringel Morris, M., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012). Tweeting is believing? Understanding microblog credibility perceptions. In *The 2012 ACM conference on computer supported cooperative work*, Seattle, Washington (pp. 441–450). New York: ACM.
- Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., & Sperling, J. (2009). TwitterStand: News in tweets. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, Seattle, WA, USA (pp. 42–51). ACM.
- Schmid, F. (2007) Formulating, identifying, and analyzing individual spatial knowledge. In *Seventh IEEE international conference on data mining workshops*, Omaha, NE (pp. 655–660).
- Sperber, D., & Wilson, D. (2004). Relevance theory. In G. Ward, & L. Horn (Eds.), *Pragmatics* (pp. 607–632).
- Starbird, K., Grace, M., & Leysia, P. (2012) Learning from the Crowd: Collaborative filtering techniques for identifying on-the-ground twitterers during mass disruptions. In L. Rothkrantz, J. Ristvej, & Z. Franco (Eds.), *9th International ISCRAM Conference, Vancouver, Canada* (pp. 1–10).
- Starbird, K., & Palen, L. (2010). Pass it on? Retweeting in mass emergency. In *7th International ISCRAM conference*, Seattle, USA (pp. 10).
- Starbird, K., & Palen, L. (2012). *(How) Will the Revolution be Retweeted? Information Diffusion and the 2011 Egyptian Uprising*. Paper presented at The ACM 2012 conference on computer supported cooperative work, Seattle, Washington.
- Stollberg, B., & de Groeve, T. (2012). The use of social media within the global disaster alert and coordination system (GDACS). In *The 21st international conference the world wide web*, Seoul, Korea (pp. 703–706).
- Thomson, R., Ito, N., Suda, H., Lin, F., Liu, Y., Hayasaka, R., et al. (2012). Trusting Tweets: The fukushima disaster and information source credibility on twitter. In *9th ISCRAM conference*, Vancouver, Canada (pp. 10).
- Twitter (2013). Celebrating #Twitter7. <https://blog.twitter.com/2013/celebrating-twitter7>. Accessed 29 August 2013.
- Vasardani, M., Timpf, S., Winter, S., & Tomko, M. (2013). From descriptions to depictions: A conceptual framework. In T. Tenbrink, J. Stell, A. Galton, & Z. Wood (Eds.), *Spatial information theory* (Vol. 8116, pp. 299–19, LNCS). Berlin: Springer.
- Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., et al. (2011). Natural language processing to the rescue? Extracting “Situational Awareness” tweets during mass emergency. In *5th International AAAI conference on weblogs and social media*, Barcelona, Spain (pp. 385–392).
- Vieweg, S., Hughs, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *ACM conference on human factors in computing systems*, Atlanta, GA, USA (pp. 1079–1088).
- Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, 54(1), 277–295.
- Wigley, S., & Fontenot, M. (2010). Crisis managers losing control of the message: A pilot study of the Virginia Tech shooting. *Public Relations Review*, 36(2), 187–189.
- Winter, S., & Richter, K.-F. (2011). Citizens as database: Conscious ubiquity in data collection. In D. Pfoser (Ed.), *Advances in spatial and temporal databases* (pp. 445–448). Berlin: Springer.
- Worboys, M. (2005). Event-orientated approaches to geographic phenomena. *International Journal of geographical Information Systems*, 19(1), 1–28.
- Worham, J. (2012). On Twitter, sifting through falsehoods in critical times. http://www.nytimes.com/2012/11/01/technology/on-twitter-sifting-through-falsehoods-in-critical-times.html?smid=pl-share&_r=0.
- Xu, C., Wong, D. W., & Yang, C. (2013). Evaluating the “geographical awareness” of individuals: An exploratory analysis of twitter data. *Cartography and Geographic Information Science*, 40(2), 103–115.
- Yanenko, O., & Schlieder, C. (2012). Enhancing the Quality of Volunteered Geographic Information: A Constraint-Based Approach. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Bridging the geographic information sciences* (pp. 429–446, Lecture Notes in Geoinformation and Cartography). Berlin, Heidelberg: Springer.
- Yang, J., Counts, S., Ringel Morris, M., & Hoff, A. (2013). Microblog credibility perceptions: Comparing the United States and China. In *ACM conference on computer supported cooperative work and social computing*, San Antonio, Texas, ACM.