5-2014

With Whom to Coordinate, Why and How in Ad-Hoc Social Media Communications during Crisis Response

Hemant Purohit
Wright State University - Main Campus

Shreyansh Bhatt
Wright State University - Main Campus

Andrew Hampton
Wright State University - Main Campus

Valerie L. Shalin
Wright State University - Main Campus, valerie.shalin@wright.edu

Amit P. Sheth
Wright State University - Main Campus, amit@sc.edu

Repository Citation
Purohit, H., Bhatt, S., Hampton, A., Shalin, V. L., Sheth, A. P., & Flach, J. M. (2014). With Whom to Coordinate, Why and How in Ad-Hoc Social Media Communications during Crisis Response. *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management*, 787-791.
https://corescholar.libraries.wright.edu/knoesis/582

With Whom to Coordinate, Why and How in Ad-hoc Social Media Communities during Crisis Response

Hemant Purohit, Shreyansh Bhatt, Andrew Hampton, Valerie Shalin, Amit Sheth, John Flach
Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University, USA
{hemant, shreyansh, andrew, valerie, amit, john}@knoesis.org

ABSTRACT

During crises directly affected people and observers join social media communities to discuss the event. They may share information relevant to response coordination, for example, specific resource needs. However, responders face a massive data overload and lack the time to monitor social media traffic for important and trustworthy information. To address these challenges, response teams may attempt manual filtering methods, resulting in limited coverage and quality. Hence, we propose a computational framework for extracting specific resource-related information, and an interface for identifying and engaging with influential participants in the dynamic, evolving social media community. Our approach helps to identify those virtual responders who serve both as sources and disseminators of important information to assist in coordinated emergency response.

Keywords
Ad-hoc communities, crisis response coordination, influencers, virtual responders, social media engagement.

INTRODUCTION

During crises people participate in social media as citizen sensors (Sheth, 2009), sharing their observations, and updating the situation. Their commentary spans multiple topics and reflects a range of demographic backgrounds. These ad-hoc, evolving social media discussion communities encompass both real and virtual volunteers (Reuter, Heger, and Pipek, 2013). Numerous messages of varying utility are shared and disseminated, concerning requests and offers for resources and services, and incident reports. Such information can provide timely and valuable situational awareness (Vieweg, Hughes, Starbird, and Palen, 2010), for example during the California fires in 2008 (Sutton, Palen, and Shklovski, 2008), the Haiti earthquake in 2010 and hurricane Sandy in 2012 (U.S. Department of Homeland Security, 2013). However, less useful sarcasm and jokes also appear. Moreover, the voluminous amount of information challenges the identification of relevant and significant information. For example, the first week of Sandy spawned more than 20 million tweets[i]. Here we address the daunting challenge of accessing the important, trustworthy information from reliable sources to assist in the coordination of formal response organizations. They seek the potentially useful contributions of a highly dynamic, ad hoc social media community. Certainly, previous literature has acknowledged the challenge of the information overload problem and emerging solutions, either by filtering information or users, and classifying resource needs to enhance situational awareness (Hiltz and Plotnick, 2013, Gupta, Joshi, A. and Kumaraguru, 2012). However, an interface to promote engagement with high-value users and information in the evolving social media community is a persisting challenge that requires further attention. One approach is to identify on-site informants (Starbird, Muzny, and Palen, 2012). This approach suffers from a potential lapse in coverage in the absence of informants, where virtual volunteers could help (Reuter, Heger, and Pipek, 2013).

We propose a framework to extract important resource-related content via influential users and an interface to engage with them in the evolving social media community on Twitter. These influential users can act as both sources and disseminators of important information and hence, contribute as emerging virtual responders to assist coordination. Our method exploits the network of user interactions (*who talks to whom)* to identify emerging influencers based on the content of social media exchanges (e.g., specific needs, such as medical, volunteering, clothing, etc). Our method for engagement leverages knowledge about user professions to further organize the user types of influentials, such as humanitarians, medical professionals, etc. The additional power of organized user information permits coordinators to tune engagement for specific types of users.

Engaging with filtered layers of users serves two purposes. First, it acknowledges the information content that makes users influential, and that may be useful for situational updates. Second, important users serve as nodes in the network to direct crucial time-sensitive information effectively. For example, rumors can be controlled by channeling correct information via these agents. Resource donations could better reflect the priorities of responders, to avoid the second disaster of managing unsolicited resources. For instance, while clothing donations actually impeded the response to Hurricane Sandy[ii], more power batteries would have helped greatly.

ANALYSIS FRAMEWORK

We enable engagement with the social media community via a visualization interface for an underlying computational analysis. As in most approaches, we begin with collecting relevant data. However, we categorize the information content, identify the influential sources of important content, and categorize users to enable faceted engagement, as described in the following three subsections:

1. *Event-related data collection:* Fetch event related tweets from the social media (Twitter here)
2. *With whom to engage*: Identify influential users for evolving needs
3. *How to engage*: Classification of users for faceted engagement

We describe the proposed engagement interface after the three processing subsections.

1.) Event-related Data Collection

This study examines Twitter microblogging. We have implemented a Twitter Streaming API-based crawler for collecting tweets. It provides real-time tweet collection, including tweet text with metadata such as the timestamp, location, entity mentions, and author information. We provide a seed set of keywords for crawling to track relevant tweets about the event, and this set is updated with the most frequent terms in the collected tweets, after periodic manual inspection of the unambiguous set of crawl keywords for quality control. The Streaming API rate limits crawling if streaming data for desired keyword set rises above 1% of the Twitter firehose. We note that this may happen during large-scale disasters, resulting in data loss; however, such practical limitations can still provide a sufficiently representative set (Morstatter, Pfeffer, Liu, and Carley, 2013).

2.) With whom to engage: Influential User Identification

This part of the framework focuses on identifying important users in the social media community for engagement. Alternatives methods for this purpose can rely on on-ground twitterers (Starbird, Muzny, and Palen, 2012), centrality measures based community representatives (Gupta, Joshi, A. and Kumaraguru, 2012), and *whom-to-follow* set based on a user's topical affinity (Kumar, Morstatter, Zafarani, and Liu, 2012), etc. However, because we acknowledge the evolving nature of the social media community formed around a disaster event, we exploit user interactions to capture the dynamics of influence, specific to need types (e.g., clothing, food, etc.). This is similar in spirit to previous research for identifying influential users in brand-page communities (Purohit, Ajmera, Joshi, Verma, and Sheth, 2012). Our method analyzes user interactions about specific needs (e.g., food, clothing, medical, etc.) for a given time window. We create a network of users as nodes and directed edges based on the interactions, such that the edge is created from USER-A to USER-B if 'USER-A interacts with (*retweets/mentions/replies to)* USER-B'. The weight of the edge is equal to the number of interactions. We then apply popular algorithm, PageRank (Page, Brin, Motwani, and Winograd, 1999) on the user interaction network to identify key influential user nodes. The algorithm iteratively assigns a weight of importance to a user USER-B by aggregating the importance of all such users USER-A who have incoming edge to the USER-B. This way, a user accrues importance based on various factors; such as if other influential nodes interact with (e.g., retweet) her, a greater number of users interact with her, etc.

To identify the set of user interactions pertaining to specific needs, we created a bag-of-words model based lexicon sets for describing needs and used it to filter the corresponding tweet set; however, more sophisticated approaches beyond bag-of-words are possible such as a topic model (Wallach, 2006). For example, clothing need was represented by the set *'cloth, blanket, jacket'* (Interested readers can contact us for the lexicon set, which we do not provide here due to space limitation.) Certainly better methods can be utilized to find the subset of data relevant to specific needs. In any case, the required method must be independent of pre-established need types to allow response coordination to prioritize based on requirements. We also note that the subsets of tweets related to needs are not mutually exclusive, for example*, "Thanks for supporting #1000bearhugs, pls help other #reliefPH efforts too; food, clothing, & meds are most needed now http://goo.gl/MkgP8D"* will be present in both clothing and food type subsets.

3.) How to engage: User Profession Categorization

Domain familiarity influences both the crafting and sharing of a message. Thus slicing and dicing access paths to information potentially helps coordination. In this step, we categorize the influential users based on profession, such as those related to humanitarian, journalism, or medical. We first created ten popular user profession domains and then created a lexicon using identifiers from Wikipedia and the U.S. Department of Labor statistics, borrowing from the occupation-based method of previous research on user interest presentation (Purohit, Dow, Olonso, Duan, and Haas, 2012). The set was then expanded manually to capture general terms. For example, the lexicon for the humanitarian profession domain contains words such as 'humanitarian', 'emergency', 'disaster', etc. The final step is to perform entity spotting of the lexicon terms in the user's description metadata of his or her Twitter profile. We acknowledge that alternative methods can be employed. Our objective is to support the key functionality at the interface to enable faceted engagement for coordination.

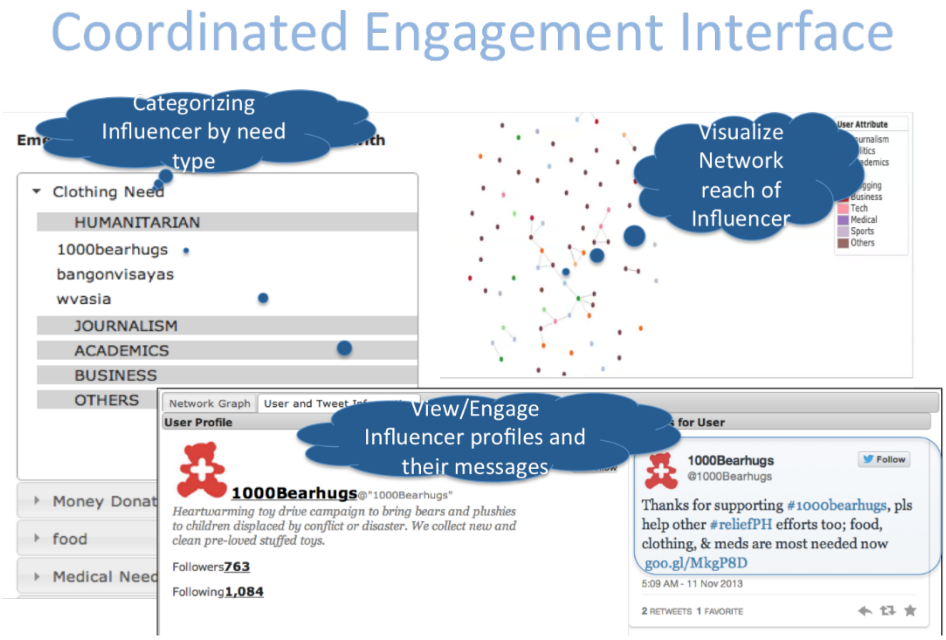

Figure 1. Engagement interface components to show three major functions to assist coordination. Online at http://twitris.knoesis.org/yolandastorm2013/networkTest/ **for Philippines typhoon event, Nov. 2013.**

ENGAGEMENT INTERFACE AND PRELIMINARY EXPERIMENTATION

We designed a prototype to create a faceted engagement interface as shown in Figure 1. It uses the outputs of processing steps 2 and 3 from our framework. It has three major parts: i.) an interaction network of the top 100 influential users with respect to a need (e.g., clothing) as shown in the right; ii.) the top users with respect to chosen need, categorized by user profession domains (e.g., humanitarian) as shown in the left; and iii.) a snippet of the user profile and the tweets that contributed to defining that user as influential as shown in the bottom for a user [*Clothing -> Humanitarian -> 100Bearhugs*]. The user profile and tweets appear after clicking on a user in the list on the left under a profession domain, e.g., humanitarian.

Table 1 lists the basic characteristics of the event-related data used in the present experimentation. We used 24 hours of data collected on Nov. 11, 2013 for the Philippines typhoon event. We analyzed the data considering the entire day as the window for our preliminary study, generated networks and identified influential users. Here we focus on three types of needs to show the added value of influential users based filtering for clothing, medical and volunteering services. We generated two sets of tweets of approximately 100 messages for qualitative study; one randomly picked from the keyword based set specific to need, and another from the filtered set of influential users. We present the following examples of tweets from both sets in Table 2 to demonstrate the value of faceted information access and engagement for coordination. In the right cell, the example messages are from the filtered set by users relevant to particular profession types and discussing specific needs during the typhoon. These are accessible via our prototype of the interface at: http://twitris.knoesis.org/yolandastorm2013/networkTest/

		Keyword based Filtering	Influential users based Filtering
Total tweets	379888		
Clothing related tweets		939	92
Medical related tweets		2139	131
Volunteer related tweets		3087	181

Table 1. Dataset characteristics for the Philippines (Yolanda) typhoon event. Keyword based filtering suffers from the information quality and still leaves considerable volume through which responders must sort.

KEYWORD-BASED FILTERING	INFLUENTIAL USER-BASED FILTERING
K1. [Clothing] "*Donated clothes for the victims of Yolanda. I hope it helps. #ReliefPH*" K2. [Clothing] "*I won't believe it's a true disaster until Anderson Cooper heads to the Philippines wearing his typhoon flak jacket and poncho. Oh, wait!* ??" K3. [Medical] "*Typhoon #Haiyan: Doctors of the World sends medical teams to worst-affected areas | DOTW http://t.co/2X4c0Csjva via* @DOW_UK" K4. [Volunteer] *"RT @_iamcLaR_: @Realtaeyang RT please Not in the Philippines but want to help for relief efforts? Details: #PrayForThePhilippines"*	I1. Clothing -> Humanitarian -> @1000BearHugs : *"Thanks for supporting #1000bearhugs, pls help other #reliefPH efforts too; food, clothing, & meds are most needed now http://goo.gl/MkgP8D"* I2. Clothing -> Journalism -> @David : "*#Ormoc urgently needs food, water, medicines, blankets. Barge is headed from Cebu to Ormoc tomorrow , please spread @MovePH #ReliefPH"* I3. Medical -> Humanitarian -> @HumanityRoad : *"#ReliefPH* @TeamRubicon *sent team of 15 to #Tacloban with medical kits supplied by @DirectRelief http://bit.ly/19SWygc #hmrd ^ct"* I4. Volunteer -> Humanitarian -> @AllHands: "*Interested in volunteering with our #SuperTyhpoon* #Haiyan *response? Let us know here: http://bit.ly/19W3k4X #volunteer* #YolandaPH*"*

Table 2. Examples of tweets randomly selected from the keyword based filtering on the left, and the proposed influential users-based filtering by faceted browsing [NEED-> PROFESSION->USER] on the right. Example K2 shows the limitation of keyword-based approach due to lack of semantics of relevance.

Tables 1 and 2 show how filtering based on influential users reduces the information overload by identifying tweets with unique and useful information of greater relevance to situational awareness, for example, I1 and I2 in table 1 indicates the prioritized needs. Keyword based filtering, on the other hand, does not address this issue due to mere syntactic approach of filtering without context. Influential users become so via attribution from community members and therefore, are likely to be sources of important information. We anticipate that coordinators will be able to locate useful sources more easily with this indication of reliability.

Additionally, users with different professional backgrounds may share different information. For example, the following two cases come from a journalist, *"Toyota to donate P10M to aid in relief efforts #TyphoonHaiyan #ReliefPH #YolandaPH"*, *"Royal Caribbean Cruises pledges support amounting to $1m. CEO says said they employ 12,800 Filipinos #reliefPH #givingback #YolandaPH"*. Though such newsworthy content is retweeted often, it does little to inform situational awareness.

CONCLUSION AND FUTURE WORK

We propose a framework to extract information and an interface to help response teams engage with influential users in the evolving social media community. This addresses information overload in two ways: 1) by identifying important information nuggets and 2) by facilitating the identification of and engagement with important users. These users can act as both sources and disseminators of important information and thereby contribute as emerging virtual responders. Our method exploits networks of user interactions (*who talks to whom)* to identify emerging influencers based on the social media discussions about specific needs, such as medical, volunteering, clothing, etc. The engagement interface presents influential users and their professions, their profiles and the messages that made them influential. Crucially, the proposed framework for faceted engagement with social media community supports analysis for the dynamic need profile of a crisis.

We plan to follow up on this work by expanding and evaluating prioritization of needs, categorization of professions and their interdependence during complex dynamics of crisis response. We shall incorporate geo-fencing based data collection, which we avoided in this preliminary study due to highly sparse distribution for messages with location-enabled metadata. We are working on user-based evaluation for the role of influence and user categorization and metrics for computing time, and will exploit parallel processing

ACKNOWLEDGEMENTS

We thank NSF for the SoCS grant IIS–1111182 'Social Media Enhanced Organizational Sensemaking in Emergency Response' and our colleagues, especially undergrad Jeremy Brunn, for their help and feedback.

REFERENCES

1. Gupta, A., Joshi, A. and Kumaraguru, P. (2012) Identifying and characterizing user communities on twitter during crisis events, *Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media* (pp. 23-26). ACM.
2. Hiltz, S. R. and Plotnick, L. (2013, May) Dealing with information overload when using social media for emergency management: emerging solutions, *Proceedings of the 10th International ISCRAM Conference*. T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T. Müller, eds.
3. Kumar, S., Morstatter, F., Zafarani, R. and Liu, H. (2013) Whom should I follow?: identifying relevant users during crises, In *HT* (pp. 139-147).
4. Page, L., Brin, S., Motwani, R. and Winograd, T. (1999) The PageRank citation ranking: bringing order to the web.
5. Purohit, H., Ajmera, J., Joshi, S., Verma, A, and Sheth, A. P. (2012) Finding influential authors in brand-page communities, In *ICWSM*.
6. Purohit, H., Dow, A., Alonso, O., Duan, L. and Haas, K. (2012) User taglines: Alternative presentations of expertise and interest in social media, *Proceedings of the first ASE International Conference on Social Informatics*. IEEE.
7. Reuter, C., Heger, O. and Pipek, V. (2013) Combining real and virtual volunteers through social media, *Proceedings of the 10th International ISCRAM Conference.* T. Comes, F. Fiedrich, S. Fortier, J. Geldermann and T. Müller, eds.
8. Sheth, A. (2009). Citizen sensing, social signals, and enriching human experience, *IEEE Internet Computing*, volume 13, number 4, pp. 87–92.
9. Starbird, K., Muzny, G. and Palen, L. (2012) Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions, *Proceedings of the 9th International ISCRAM Conference*. L. Rothkrantz, J. Ristvej, and Z. Franco, eds.
10. Sutton, J., Palen, L. and Shklovski, I. (2008) Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. *Proceedings of the 5th International ISCRAM Conference*. F. Fiedrich and B. Van de Walle, eds.
11. Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013) Is the sample good enough? Comparing data from twitter's streaming api with twitter's firehose. In *ICWSM*.
12. U.S. Department of Homeland Security, Science and Technology Directorate, Virtual Social Media Working Group (2013). Lessons learned - Social media and hurricane sandy.
13. Vieweg, S., Hughes, A. L., Starbird, K. and Palen, L. (2010) Microblogging during two natural hazards events: what twitter may contribute to situational awareness, In *CHI*.
14. Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. *Proceedings of the 23rd International Conference on Machine Learning* (pp. 977-984). ACM.

[i] http://www.huffingtonpost.com/2012/11/02/twitter-hurricane-sandy_n_2066281.html

[ii] http://www.npr.org/2013/01/09/168946170/thanks-but-no-thanks-when-post-disaster-donations-overwhelm