© ACM, 2017. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in:
He, X., Lu, D., Margolin, D., Wang, M., Idrissi, S., Lin, Y.-R. (2017). "The Signals and Noise: Actionable Information in Improvised Social Media Channels During a Disaster." In Proceedings of Web Science 2017 (WebSci 2017)
http://dx.doi.org/10.1145/3091478.3091501

The Signals and Noise: Actionable Information in Improvised Social Media Channels During a Disaster

Xingsheng He[1,3], Di Lu[1], Drew Margolin[2], Mengdi Wang[1], Salma El Idrissi[2], Yu-Ru Lin[1]

[1]School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260, USA
[2]Department of Communication, Cornell University, Ithaca, NY 14850, USA
[3]Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, China

ABSTRACT

Web-based social and communication technologies enable citizens to self-organize relief efforts in response to crises. This work focuses on a question fundamental to the concept of collective intelligence: how effective are such self-organized channels, ungoverned by any central authority, in conforming to their intended function? In this study we examine the hashtag #PorteOuverte ("#OpenDoor") introduced during the 2015 Paris terrorist attacks, as an "improvised logistical channel" (ILC) to help individuals to find a safe shelter near the attack sites. We analyze the dynamics and effectiveness of #PorteOuverte by comparing its proportion of relevant logistical messages – individuals requesting or offering shelter – to other messages such as those offering emotional consolation or commenting on the hashtag itself. Our results reveal that the vast majority of messages are not relevant, however the crowd senses and spreads relevant messages more than others. We further demonstrate that relevant messages can be automatically detected and thus algorithmic promotion may be possible.

KEYWORDS

collective intelligence, disaster response, social media, improvised logistical channel, crowd behaviors, self-organized systems

1 INTRODUCTION

Social media empowers individuals to spontaneously organize and perform actions traditionally performed by more formal institutions. One area where these citizen-based efforts have played a particularly important role is in disaster response [1, 12, 24, 25, 36]. Of particular importance are citizen efforts to offer or ask for concrete logistical assistance [9, 13, 21, 27, 28, 34, 37, 41, 49]. To date, however, research on the citizen efforts using technology to coordinate logistical support has focused on improving responses rather than understanding community dynamics. For example, researchers have provided techniques to improve the ability of formal relief organizations to respond to needs posted to social media [28, 41]) or for citizens to find needed information [27, 37]. Little work has examined the dynamics of citizen efforts on their own with the aim of understanding how they operate and achieve their goals.

The #PorteOuverte hashtag, created in response to the November, 2015 terrorist attacks in Paris, provides an opportunity to observe such an effort. The hashtag, which means "open door," was used by individuals to offer shelter to strangers stranded by the attacks (see Fig. 1) and by individuals in need of shelter to request help and post their whereabouts. Such citizen-to-citizen coordination over social media during a disaster is not new [11, 12, 25, 27]. However, the crowd's creation of a specific, designated channel through which to accomplish this goal, similar to that afforded the study of political protests [30, 38], deserves attention. Via #PorteOuverte individuals not only communicated about logistical needs, they did so in an intentionally coordinated manner using a designated public channel. The channel bypassed formal organizations, however. It was comprised only of a simple, linguistic symbol. It thus placed the responsibility of coordinating vital resources on a self-organized conversation.

We refer to this spontaneous, crowd-based coordination of concrete resources as the use of an *improvised logistical channel* or *ILC*. We define ILCs as any **specified** means of **disseminating information** about who, where, or how individuals can take **concrete actions** to address a **novel or sudden** collective need that operates without the coordination of a central authority [19]. ILCs operate at the intersection of traditional public service and new media. Like traditional logistical infrastructure, an ILC has a designated use: to provide concrete, *actionable* information in times of distress. But unlike its more formal counterparts, an ILC can be created by anyone and disseminated in a cacophonous social space in which there is no formal mechanism of control.

In many ways hashtags are ideal candidates for ILCs. Like formal channels of communication such as radio or official websites, hashtags can rapidly draw attention to a particular flow of information [15]. Yet unlike these more formal channels, hashtags can be created cheaply by anyone through their use of natural language [5, 17, 23]. They are also easy to identify via searches [46] and spread easily through social networks [17, 32]. Thus when confronting unforeseen disasters hashtags offer communities the ability quickly self-organize to share information and coordinate responses for which there may be little pre-existing infrastructure [25].

However, like their natural language cousins, the fluid and uncontrolled nature of hashtags also makes them subject to an ongoing evolution that can undermine their logistical goals [5]. Hashtags can be re-purposed for a variety of social ends [45], lose their coherence [26], and even be "colonized" by messages designed to alter

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WebSci'17, June 25-28, 2017, Troy, NY, USA.
© 2017 ACM. 978-1-4503-4896-6/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3091478.3091501

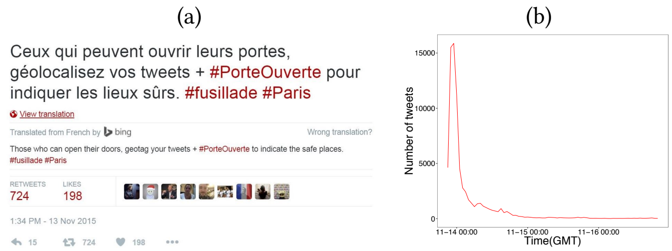

Figure 1: (a) The first tweet using the hashtag #PorteOuverte after the Paris attacks. (b) The hourly numbers of the original tweets containing the hashtag #PorteOuverte.

their meaning [10]. The effectiveness of a hashtag as an ILC thus depends on the interplay of two competing dynamics: the need to spread *broadly* enough to be effective and the need to remain *specific* enough to be useful.

In this paper, we investigate #PorteOuverte hashtag as an ILC. We address three fundamental questions: (1) What portion of the messages sent to the channel addresses its intended purpose? (2) Of the messages sent to the channel, which types draw the most public attention? (3) Can relevant, logistical messages be automatically detected and selected for promotion?

This work thus initiates the theoretical study of hashtags as self-organized channels for addressing a communal need in the wake of a disaster. Our findings indicate that the channel is heavily polluted with irrelevant information, though the community makes efforts to distinguish and promote appropriate uses. Most interestingly, we find that promotional and reflexive talk about the hashtag – celebrating it as an achievement – plays a substantial role in undermining its effectiveness.

2 BACKGROUND AND RELATED WORK

Social media has become increasingly important as a backchannel communication medium during emergencies and disasters [12, 22]. Social media enables individuals immediately proximate to or directly familiar with the disaster to communicate directly with one another and the public, improving both the speed and accuracy of information reported compared with messages from official accounts [24, 40]. Twitter in particular has been utilized for providing rapidly changing situational information about crises as they unfold [12, 39, 42].

At the same time, hashtags have emerged as an important technology for self-organized coordination [46]. Any user can create a novel hashtag to address a new situation or need [5]. Other users can disseminate the tag [32]. The crowd can learn the tag's meaning from the manner in which others use it [26]. Hashtags also democratize discourse, allowing users to coordinate conversation without relying on formal leadership [23, 30]. Hashtags can thus support logistical crisis communication by allowing individuals to create specific channels to serve particular needs as an event unfolds [19].

Yet the open access and real-time adaptivity of hashtags can also be their downfall. Like natural language, hashtags are often used inconsistently [26]. This tendency toward broad but inconsistent use can be problematic in the context of disasters where people tend to post a wide variety of message types [1, 4, 14, 29, 42]. While many messages are *informative* [22], meaning they provide situational information about the disaster, including updates regarding affected individuals, infrastructure, donations and volunteer activities, cautions, and other useful emergency information, many others are *non-informative* [22], such as those that provide emotional support or express general disaster related opinions [16, 29, 43].

A variety of recent research has attempted to distinguish useful disaster-related information in social media streams [44]. These studies have primarily focused on external interventions to address specific problems, such as techniques that improve the effectiveness of relief organizations or match help requests to help offers [21, 27, 28, 37, 41]. Little work has focused on understanding the effectiveness and dynamics of the crowd's own self-organized efforts, however.

2.1 Research Questions

Our study focuses on the ability of the #PorteOuverte hashtag to draw attention to messages relevant to matching individuals in need of shelter with other individuals offering shelter nearby. As described above, studies indicate that there are multiple, competing motivations for messages on social media during disasters such as sharing information, providing sympathy and solidarity, and giving general opinions [12, 22, 29]. This leads to our first question:

RQ1: *What types of messages are sent to #PorteOuverte, and what is the proportion of messages that provide logistic information (relevant to the channel's intended purpose)?*

The proportion of relevant messages is also likely to be unstable over time. Speakers use "least effort" when expressing themselves [48]. They prefer terms that are specific but also widely recognized. The more broadly a symbol (e.g. word, hashtag) is used, the greater the incentive to tolerate its poor specificity to take advantage of its broad recognizability. This suggests that #PorteOuverte will draw more messages when users can observe that it is used as specifically intended, but as it becomes more widely diffused and recognized it will increasingly attract deviant messages designed to exploit the collective attention it receives. Hashtags in particular carry the risk of "colonization" by those who seek to co-opt or undermine their function once they are widely known [2, 10]. Thus, #PorteOuverte might begin as an effective ILC and then lose its effectiveness over time. We thus ask:

RQ2: *How does the number and proportion of logistical messages sent to #PorteOuverte change over time?*

If the audience's reaction to the channel was entirely passive then the inflow of messages sent to #PorteOuverte would define the channel's effectiveness. However, social media users are not passive recipients but actively intervene in information flows, particularly in disaster contexts [12, 37]. For example, while disasters often prompt the spread of false rumors, users often react by posting messages discouraging them [7, 35]. Similarly, users may overcome noise in the #PorteOuverte channel by rebroadcasting, via retweets, the relevant logistical messages over other kinds of messages[39]. We thus ask:

RQ3a: *What types of messages sent to #PorteOuverte are most retweeted by users?*

A related question is whether users are attending sufficiently to all logistical messages or only a select few.

RQ3b: *What portion of logistical messages sent to #PorteOuverte are retweeted by users?*

Our final question is whether simple tools can be developed to improve the integrity of the channel. Specifically, if logistically relevant messages can be detected algorithmically, it may be possible to implement automated filters to maintain the proportion of relevant messages that pass through the channel at acceptable levels.

RQ4: *To what extent can logistically relevant messages be detected automatically?*

3 DATA COLLECTION

To collect all the original tweets using the hashtag #PorteOuverte, we used Twitter's advanced search function[1] to gather retrospectively tweets containing such a hashtag.

The data were collected in two steps: First, about two weeks after the Paris attacks, we requested through the Twitter advanced search service by specifying the hashtag as #PorteOuverte, the date as from 2015-11-12 to 2015-11-16 and the rest parameters as default in the web page interface. Then we extracted all the tweet ids from the iteratively returned webpages. Second, after obtaining these tweet ids, we leveraged Twitter REST API to get more detailed information of each tweet. Fig. 1(a) shows the first tweet using #PorteOuverte. The first tweet posted on Friday, 13 Nov 2015 21:34:06 GMT created an online channel and advocated users who can offer help to use such a hashtag to provide offline shelters. Fig. 1(b) shows the hourly volume of original tweets, indicating that the majority of tweets appeared within the first 24 hours after the attacks.

With the 75,547 original tweets returned from the Twitter advanced search service, we filtered out the tweets that only contain hashtags and the *conversational* tweets (i.e., tweets that replied to other tweets). The final data set contains 66,430 tweets in a period of 74 hours from the time the hashtag was created.

4 CATEGORIZING HASHTAG FUNCTIONS

Coding scheme derived from prior research. We seek to identify messages that relate specifically to the designated logistical purpose of #PorteOuverte – identifying safe shelters and individuals in need of a safe shelter – and distinguish these from other kinds of messages. Our coding procedure began with the baseline scheme provided by [29]. We examined the applicability of the scheme to #PorteOuverte tweets and then trained coders to apply the scheme and tested their reliability. We then made adjustments, in particular, by adding two new categories not observed in the prior work.

Qu et al. (2011) provided a six-category scheme for categorizing messages sent on Sina-Weibo related in the aftermath of the 2010 Yushu earthquake: *Situation updates, General disaster-related message, Action-related message, Emotional/Social-related message, Opinion-related message,* and *Off-topic*. Of specific interest is their distinction of action-related messages from both situational updates and general disaster information messages as those which aim to coordinate tangible resources in the affected region during and after disasters.

Open coding, pilot coding and code book. In our first step, one co-author labeled randomly selected 200 tweets from our dataset, according to the definitions from [29]. The results indicated that the majority of messages fit the existing coding scheme. Nonetheless, we observed a significant number of tweets for a novel use – referencing the hashtag itself. This use was not identified by the study [29] as they examined all messages rather than those under a specific hashtag. In particular, many messages contained factual information about the hashtag with the intention of promoting its use, such as instructions and advocacy for its use. We categorized the messages as "containing *Information about the hashtag*".

Next, we conducted a pilot study in order to test whether our coding scheme is clear and sufficient to deliver reliable outcomes. We recruited two external coders along with four of our co-authors and conducted four sessions of pilot coding with 300 randomly selected tweets from our dataset. Each coder coded one batch of 100 tweets independently and therefore each batch was coded by at least two coders. We computed the inter-coder reliability for every two coders coded the same batch and the pairwise Cohen's kappa was relatively low ($\kappa = 0.35$) for some categories. (We computed the Cohen's kappa separately for each category as each tweet can have multiple categories.)

We found that the most disagreements among coders were caused by differing perceptions of *emotional/social-related* and *opinion-related* messages. In particular, many messages were highly emotional irrespective of their intention to provide support. Since social support has been considered as influential in buffering the effects of stress from disasters [3], we distinguished social support messages from those conveying emotion in a non-supportive way, such as anger at the government or terrorists.

We also discovered a novel kind of message related to the hashtag itself. A number of tweets used #PorteOuverte to refer to the phenomenon of the hashtag, that is, the fact that people self-organized through Twitter to offer one another assistance. These messages often included emotional statements of awe or pride in humanity, as well as occasional statements of fear or scorn, and thus contained elements of social support and general opinion. However, they are also unique to the phenomenon of #PorteOuverte and its evolution as an improvised logistical channel. We thus distinguished these *Reflexive comments* from general messages of social support and opinion.

In addition, we found that categories in our code book are not mutually exclusive as a single tweet may fit more than one. We thus permitted each tweet in our dataset to be classified into more than one category as needed.

Based on these observations we refined our code book to clarify category boundaries, decomposing those two categories and reconstructing three new categories as *Social support*, *Reflexive comments* and *General opinion*. Table 1 lists the coding scheme and detailed selection criteria in our final code book with examples of tweets for each category. By conducting another round of pilot coding with three coders independently coding one batch of 100 tweets, we confirmed the results based on the final coding scheme reached an acceptable inter-rater reliability (pariwise $\kappa > 0.75$ for all categories).

Coding process. We started the coding process with the final coding book (Table 1). We randomly sampled 14% tweets from our dataset with a uniform distribution across time. That is, we grouped all the tweets into different sessions using a fixed time (15-minute) window and randomly sampled tweets from each session.

[1]https://twitter.com/search-advanced

Category	Abbr.	Selection Criteria	Example
Logistical Information indicated by #PorteOuverte hashtag	*Log*	Offering or requesting help about accommodations aligning with the goal of #PorteOuverte activity	Hey tourists in #Paris if you need help or some place to sleep around Jussieu 5th arrondissement, my door is open. #PorteOuverte #OpenDoor
Information about #PorteOuverte hashtag	*Use*	Providing any type of information about the #PorteOuverte activity including the fact, news, advocation or instruction of the usage of #PorteOuverte	For people in Paris looking for a shelter use hashtag #PorteOuverte. Parisians will let you in their home.
Situational Updates of the attack	*Situ*	Providing factual information about what was happening in the attack area, such as information about including victim, injury, suspects, infrastructure, utilities, situational warning and cautions. Such information improves situational awareness of readers.	French army just stormed stadium but the terrorists had killed a lot of the hostage already! #ParisUnderAttack #porteouverte
Other General Information for the attack	*Gen*	Providing or requesting any other Paris attack related relief information, such as donation, mourning activities, embassy information or directions to other relief information resources	@**** American Embassy in #Paris +33143122222 #PorteOuverte
Social/Emotional Support	*S/E*	Providing social and emotional support, including mourning, blessing/praying, comforting, encouraging, and expressing concerns for or solidarity with victims.	Our thoughts go out to those suffering from the terrible #ParisAttacks. #porteouverte #PeaceForParis
Reflexive commentary on hashtag phenomenon	*Ref*	Referring to the #PorteOuverte hashtag as a social phenomenon or meme, including commenting on whether it is good or bad, how it makes the person feel etc.	The hashtag #PorteOuverte shows love and courage in the face of hatred and fear
General expression of opinion	*Exp*	Expressing opinions or feelings related to the attacks	We could make the world a better place if we all put our racial, #religious socio-political differences aside. #PorteOuverte #PrayForParis
Off-topic/Non Codeable	*off-topic*	The tweet does NOT contain any relevant information to Paris Attack or the coder cannot understand the content or infer the purpose	It was a day full of little brains trying to find their place in -@*** 's majors. -#porteouverte

Table 1: Coding scheme for categorizing different types of tweets with hashtag #PorteOuverte.

We conducted two-round coding process. For the first round we coded 10% and the second round we coded additional 4%.

The four co-authors who helped develop the code book as described above participated in this coding process as experienced coders. We provided coders with URL of each tweet instead of the plain text, so that through the browser all the contextual information of each tweet including the embedded tweets, pictures, videos, and replies are accessible to coders to help them precisely classify the type of information conveyed by the tweets. Each of the coders annotated a bulk of tweets independently and 9240 tweets were coded in total. To validate the accuracy of our coding process, we randomly selected three batches, each containing 100 tweets, to be coded by two coders. The pairwise agreement for every two coders who coded the same batch was relatively high (agreement=0.83). In particular, the pairwise inter-rater reliability for *Logistical information* tweets (Category *Log*), the focus of our study, reached 0.90.

5 ANALYSIS: FUNCTIONAL SIGNALS AND IMPACT

In this section we address RQ1, RQ2, RQ3a and RQ3b by examining the types and timing of messages users tweet and retweet through #PorteOuverte.

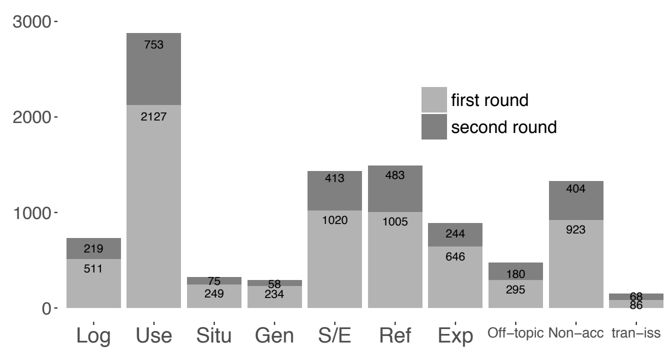

Figure 2: Category distribution of the #PorteOuverte tweets. "Non-acc" (abbr. for Non-accessible) are the tweets that can not be accessed from the webpages and "tran-iss" (abbr. for translation-issue) are the tweets that the coders have difficulty in understanding the translated content.

Content distribution (RQ1). We applied the classification scheme to the uniformly sampled tweets in a two-round coding process: We first coded 10% and then compared the coding results with the rest 4%. Overall, the coding results are consistent across different categories. For example, for the coded tweets, 7.74% and 8.42%

tweets were labeled as *Logistical information*, 15.45% and 15.88% tweets were labeled as *Social support* in the first and second round, respectively. Fig. 2 gives the results of each round and shows the category distribution. Besides 475 tweets coded as *off-topic* and 154 tweets have the translation issue, most tweets fall into the seven categories: there are 324 *Situational updates* tweets, 292 *General information* tweets, 730 *Logistical information* tweets, 2880 *Information about the hashtag* tweets, 1433 *Social support* tweets, 1488 *Reflexive comments* tweets , and 890 *General opinion* tweets.

As shown in Fig. 2, the proportion of *Logistical information* tweets, those which are in alignment with the goal of the hashtag, occupied only a small portion of the channel. The modal category is *Information about the hashtag*, messages that are of uncertain usefulness. These tweets that promote the hashtag may be of value in getting the word out about the existence of the channel, but too many *Information about the hashtag* tweets can also overwhelm the useful information. Furthermore, several other categories typically found in response to disasters (*Social support*, *General opinion*, and also *Situational updates*) also dominate relative to *Logistical information.* We also observed that, ironically, *Reflexive comments*, in which individuals celebrate the channel for its usefulness, also appears to drown out relevant information.

One possibility is that the channel is more accurately comprised of two channels: a useful, low-noise channel for individuals communicating logistical information in French and a more generic, high noise channel for those using other languages to talk about other things. As shown in Table 2, we do see differences in the channel's use based on languages; however, the results are not consistent with this characterization. First, we observed that the French-only tweets in the channel have improved, but still low, precision, as the proportion of *Logistical information* is still only (26.7%), with *Logistical information* plus *Information about the hashtag* still only combining to be just over half (53.6%). Moreover, French-only tweets are not particularly strong in recall, as only 71% of *Logistical information* tweets are in French. In other words, almost 1/3 of relevant messages would be missed if users of the channel focused only on French tweets. Furthermore, the channel's performance among these non-French tweets is very poor, as only 3.4% of non-French tweets to #PorteOuverte are relevant, *Logistical information* tweets. Thus, while there are important differences in the channel based on language use, it is not a case of two distinct channels with different applications, and so our subsequent analysis continues to focus on the channel as a whole.

Category trends (RQ2). Assessing the effectiveness of #PorteOuverte requires not only determining the overall mix of messages sent to it but also the timing of these different kinds of messages relative to one another. We summarize the temporal change of categories in Fig. 3 to help understand how the content and the ratios change as the events on the ground unfold.

As shown in Fig. 3 panel A, the hashtag, created at 11-13 21:34:06 GMT, began with an immediate spike triggered by the violent attacks. For roughly the next four hours a stream of *Logistical information* messages are sent through the channel. As indicated in panel B, over time, other types of information entered the channel but behaved differently. The most important information *Logistical information*, was, appropriately, the dominant form of information for the first 30 minutes after the hashtag's creation. It was

Table 2: The number of French and non-French tweets in each category. The sum of all the categories may exceed the number of "on-topic tweets" as a tweet can be coded with multiple categories.

	# of tweets		# of retweets	
	Fr	Non-Fr	Fr	Non-Fr
on-topic tweets	1839	5405	38339	49817
Log	506	224	17464	13824
Use	549	2331	16322	23764
Situ	72	252	3527	3159
Gen	51	241	1020	2886
S/E	201	1232	2935	4190
Ref	451	1037	2234	6703
Exp	137	753	564	2366

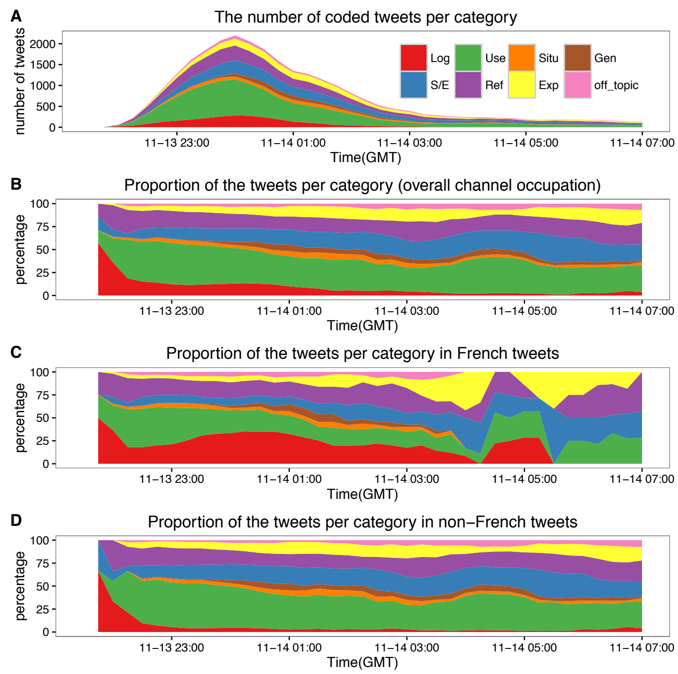

Figure 3: The number of coded tweets and the channel occupation of each category overtime.

quickly replaced by *Information about the hashtag*, however, which continued to dominate.

Similar to the previous content analysis, after its emergence the largest category remained *Information about the hashtag* while the channel was active. In addition, *Social support* and *General opinion* gradually became more dominant hours after the attacks as the logistical *Logistical information* messages began to wane. Nonetheless, our analysis indicates that noisy (*Social support*, *Reflexive comments*, *General opinion*, *off-topic*) information was prominent from the beginning (34% at 11-13 23:00), and increasingly squeezed the bandwidth as the evening wore on (45% at 11-14 01:00 and 54% at 11-14 03:00 GMT). As with the analysis of the overall proportions, the French-only tweets (panel C) were less noisy; nonetheless, *Logistical information* tweets were at no time more than 35% of the channel after the first 30 minutes.

Comparing the raw number of *Logistical information* messages to their proportion provides further insight into whether these messages are interfering with the logistical coordination intended by #PorteOuverte. The number of *Logistical information* messages roughly indicates the number of individuals trying to use the channel for its intended purpose, while the proportion of noise indicates how hard it would be to find those messages. The problem is aggravated when the peaks of useful information (*Logistical information*) and proportion of noise coincide.

Results show that the proportion of noise is quite high even at the peak volume of actionable messages. Even excluding promotional use of the hashtag (*Information about the hashtag*), the peak of logistical messages, in terms of raw count, occurs between 11-13 23:34 and 11-13 23:49 GMT, when these messages are only a 13.67% share of the total, compared with 13.72% for *Social support*, 17.05% for *Reflexive comments*, and 7.17% for *General opinion*.

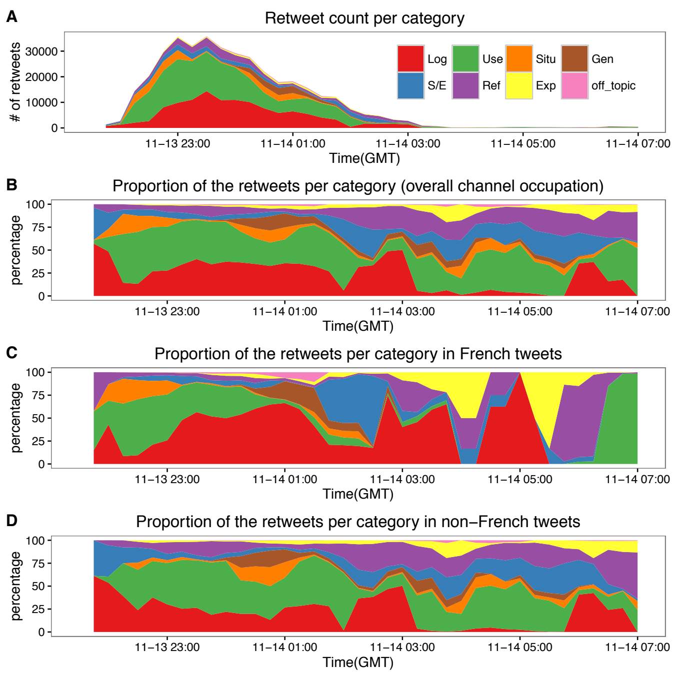

Figure 4: The retweet count and the channel occupation of each category overtime.

Whether this stream of irrelevant messages is problematic for the channel depends not only on its proportion but whether it is sufficient to distract from relevant information.

User attention (RQ3a). We began by analyzing the flow of vital *Logistical information* and noisy messages relative to users' attention for distinguishing them. At the peak time at which *Logistical information* messages were sent to #PorteOuverte the hashtag received approximately 4.8 tweets per second of which 13.67% were *Logistical information*. This means that a user would receive 1 logistically actionable message every 1.52 seconds, and need to discover and pick it out of an additional 6.32 messages received during this same 1.52-second interval. This rate is more than 100 times greater than the rate at which researchers have estimated information overload takes place, and more than twice the highest rate tested [31]. Thus, we conclude that even if these coarse calculations substantially underestimate users' attention or ability to navigate the channel, it was nonetheless substantially overloaded and any gain in the signal-to-noise ratio would appreciably improve the dissemination of relevant information.

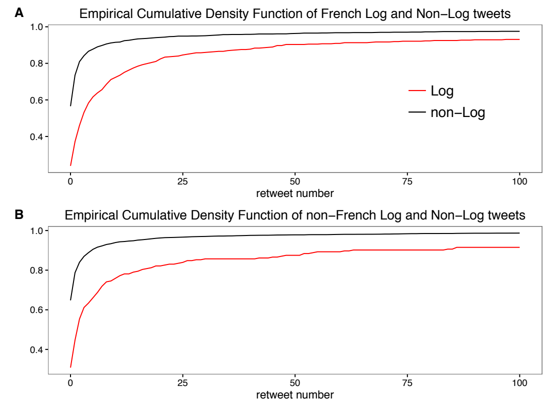

Figure 5: The ECDF of retweet numbers per tweet in the *Log* and Non-*Log* categories across different languages. Tweets with more than 100 retweets were not shown in the plots (less than 3% in Non-*Log* and 9% in *Log*).

To further examine user attention we consider the extent to which *Logistical information* tweets are identified by users for retweeting. Though Twitter does not re-post retweets directly into the channel the rate of retweets contributes to such tweets being identified as "Top" in Twitter's feed to the hashtag and also disseminates the tweet directly to the retweeting users' followers. Fig. 4 gives the attention each category gained over time and is more optimistic for the channel's effectiveness. Though the majority of tweets sent into the channel were not logistical, the audience did effectively select these relevant tweets for rebroadcasting such that, across Twitter, they ended up being a substantial portion of the #PorteOuverte messages disseminated. In particular, *Logistical information* tweets received an approximately fortyfold boost from retweets when the raw number of original *Logistical information* reached its peak. This enables this logistical information to populate about 37.42% of messages tagged with #PorteOuverte, even though they are only 13.67% of the original content tagged this way. Combined with the amplified *Information about the hashtag* tweets, this relevant information dominated until 11-14 03:00 GMT time, after the demand for *Logistical information* appeared to have subsided according to the raw count of original tweets. The effect is particularly dramatic for French-only tweets, which shows several bursts of retweets that drive the proportion of *Logistical information* to steadily grow and reaches as high as 66% when this channel was active. Thus, consistent with [24], self-organized participation in the channel is corrective against noise.

The unseen *Logistical information* (RQ3b). The preceding analyses indicate that though the channel is overloaded, the crowd has good precision in identifying relevant, logistical messages for rebroadcasting. This raises the question of whether this rebroadcasting effort is well distributed across all messages or focuses only on a few. In Fig. 5, we show the Empirical Cumulative Density Function of the retweets of *Logistical information* and

Non-*Logistical information* tweets. As in our previous analysis, *Log* tweets generally attracted more attention than non-*Logistical information* tweets. However, a considerable portion of valuable *Logistical information* tweets were not selected for re-broadcasting as 25% of *Logistical information* got no retweets at all and about half of *Logistical information* information got less that 4 retweets.

These results indicate that though the crowd attends to and selects for *Logistical information* messages, this attention is focused on only a subset of relevant messages. This phenomenon suggests that automated tools may help in identifying these messages and inspires us to develop an algorithmic approach that can automatically detect *Logistical information* tweets for better promotion under disasters.

6 MACHINE CLASSIFICATION

To answer **RQ4** – "*To what extent can logistically relevant messages be detected automatically?*" – we conducted experiments on various machine classification techniques. With the tweet labels obtained from the human coding process, the goal is to automatically detect the human-coded label(s) for a given tweet. We are interested more in the most relevant categories, *Log* and *Use*, which are central to the communication of the logistical messages, while we also tested our machine classification techniques on the other two categories *S/E* and *Ref*, for comparison.

Experiment setting. We operationalized the problem into a binary classification task with an objective to classify whether a given tweet was labeled with a particular category. For example, in the classification for category *Log*, we consider *Log*-labeled tweets as positive and otherwise negative. The binary classification setting enables us to examine each category separately and to study the most relevant categories in depth. The experiments used the first-round coded tweets (6,200 in total), and our text pre-processing step (see details below) resulted in 4,331 tweets. Table 3 lists the basic statistics in the experiment dataset.

Table 3: The number of tweets, proportion and positive-negative ratio for each category in the experiment dataset.

	# of tweets	Percentage	Ratio of +/-
Log	399	9.21%	1:10
Use	1814	41.88%	1:1.38
S/E	869	20.06%	1:3.98
Ref	922	21.29%	1:3.69

Text pre-processing. The text pre-processing step ensures the tweets used for experiments are interesting and avoids the results from being dominated by the peculiar or systematic features in the data. In the current work, we focused on how to automatically classify the event-related categories and excluded tweets annotated as *off-topic* (5.14%) or labeled as with translation-issue (1.67%) by annotators. In the real-world scenarios, both tweets with translation issue and *off-topic* tweets can be detected and filtered out by sophisticated translation systems (e.g. Google Translate[2]) or technical methods such as event-specific content segregation[18] or topical content distillation [47].

We also removed the token "#PorteOuverte" from each tweet as it appeared in all tweets. URLs and mention (@username) were replaced with standardized tokens and will be treated differently from other text tokens. We further filtered out tweets with less than two words as the semantics in such short tweets are often obscure. As tweets in our corpus were written in different languages, we translated all the tweets into English using Google Translate[2] and compared the translated data with non-translated ones. Tweets with translation issues were removed. We also applied standard text pre-processing such as removing punctuations and applied *Snowball Stemmer* for word stemming.

Feature engineering. Two types of features were extracted from each tweet as the input of the classification task.

Content features: features derived from the words appeared in the tweet text. We considered two alternatives: *N*-gram and word2vec [20]. *N*-gram is a continuous sequence of *n* words in a text. We extracted various *N*-gram features including *unigrams*, *bigrams*, and *trigrams*. The initially extracted *N*-grams resulted in a high-dimensional set of features (35,907 in total). To reduce the computational complexity and also to remove the noisy features, the dataset was pruned to ignore the infrequent terms (e.g., less than two times). We tested and compared several state-of-the-art methods to select the most discriminative features and decided to use the "Bi-Normal Separation" [8] over other commonly used metrics, such as "Information Gain," "Term Frequency" and "Chi-Squared." Finally, to further reduce the feature dimension, we applied the distributional representation method word2vec [20] to obtain the mean 100-dimensional vectors for each given tweet according to its contained words.

Contextual features: features describing the contextual information associated with the tweets. We extracted four types of contextual features to augment the content features: (1) ContainURL, (2) ContainMention, (3) ContainParisLocation, and (4) InParis. The first three features are binary features indicating whether the tweet contains a URL, an @-mention, or any identified location names in Paris, e.g., the name of a street or a subway station name. The street and subway station names were obtained from public website[3,4]. The last feature provides geographic proximity with respect to the event locus, Paris, with three categories: in Paris, not in Paris or unknown. Due to the sparseness of geo-codable tweets (only 0.3% tweets were posted with geotags), we extracted the geo-location from the users' profiles (available from 75% of users).

It is worth noting that the information about language and posting time of a tweet was excluded intentionally. While the goal of the classification is to distinguish the logistically relevant messages, our ultimate goal is to enhance the use of the ILCs during the course of the event and across communities. Thus, the machine classification should be built without the knowledge about the language and time of the tweets.

Training the classifiers. From the feature engineering, we obtained four combinations of features as the input of the classification: (1) *N*-gram alone, (2) word2vec alone, (3) *N*-gram with contextual features (*N*-gram-C), and (4) word2vec with contextual features (word2vec-C). We employed both traditional binary

[2]https://translate.google.com/

[3]http://www-groups.dcs.st-and.ac.uk/history/Societies/ParisNames.html

[4]https://en.wikipedia.org/wiki/List_of_stations_of_the_Paris_Metro

classifiers such as Naive Bayes, AdaBoost and Support Vector Machines (SVM), and a tree boosting classifier XGBoost [6], which has been shown to work better than other classifiers like Deep Neural Network in a moderate-size dataset.

Due to the imbalanced positive-negative class distributions in all categories (see Table 3), we have tested classifiers trained with data under various sampling strategies – ranging from a 1:1 under-sampling condition (where the negative instances were under-sampled to create a balanced positive-negative ratio) to a no-sampling condition to retain the original class ratios. Our comparison showed that the classification training with any under-sampling strategy did not have a significant effect (with only less than 1% or no improvement). Hence, we report and discuss the results based on the training data that retained the original class ratios.

Evaluation of the classifiers. The classification performance was evaluated using a 10-fold cross-validation. We report the standard evaluation metrics, including Precision, Recall, F-score and AUC (Area Under ROC curve). The AUC metric is particularly informative in the case of imbalanced class distribution.

Table 4: Classification performance (%), with 10-fold cross-validation, for different classifiers and across categories.

	Precision	Recall	F-Score	AUC	Method
Log	**93.67**	60.40	73.44	96.31	SVM
	93.60	58.64	72.11	95.96	AdaBoost
	64.40	**82.21**	72.24	94.90	NB
	91.06	68.92	**78.46**	**96.32**	XGBoost
Use	82.22	85.17	83.67	**93.59**	SVM
	82.47	**87.93**	85.11	93.03	AdaBoost
	80.61	84.34	82.43	91.22	NB
	82.82	87.85	**85.26**	93.34	XGBoost
S/E	**71.80**	51.20	59.81	**86.52**	SVM
	70.71	53.62	**60.99**	85.17	AdaBoost
	40.09	**73.30**	51.83	80.35	NB
	69.57	52.35	59.75	85.84	XGBoost
Ref	**72.15**	51.40	60.03	**86.60**	SVM
	71.80	53.58	61.36	86.20	AdaBoost
	49.26	**69.19**	57.55	82.93	NB
	71.48	54.12	**61.60**	86.50	XGBoost

Table 5: Classification performance in AUC (%), with 10-fold cross-validation, using different features with SVM.

	N-gram	N-gram-C	word2vec	word2vec-C
Log	94.60	95.61	95.89	**96.31**
Use	93.11	93.13	93.10	**93.14**
S/E	84.46	84.72	84.75	**86.52**
Ref	86.11	86.42	86.15	**86.60**

Table 6: Classification performance in AUC (%), testing with future data, using different features with SVM.

	N-gram	N-gram-C	word2vec	word2vec-C
Log	92.16	92.27	92.79	**93.10**
Use	90.02	91.94	90.39	**91.09**
S/E	82.02	83.10	83.28	**83.28**
Ref	81.55	84.85	85.05	**85.56**

6.1 Classification results

Table 4 shows the overall performance for different classifiers across different categories, with the best feature combination, word2vec-C, which captures both syntactic and semantic word relationships and performs the best across all different classifiers. All reported results are based on the translated texts, which have 1.12%–4.61% performance gains over the non-translated ones. As shown in Table 4, all the classifiers have reached to an AUC above 80%, indicating that the features that we selected have a good discriminating ability. In particular, we were able to obtain high precision in some cases (e.g., 93.67% when using SVM on *Log*), indicating the ability of a machine to correctly identify a specific kind of messages (e.g. logistical messages).

Classifier comparison. Based on the AUC metric (in Table 4), SVM and XGBoost perform the best across all categories. XGBoost achieves the best in *Log* and has similar performance with SVM in other categories. NB performs the worst in terms of AUC and F-score, but performs the best on Recall metric except for *Use*. Most of the classifiers (except for NB) have a Precision value higher than the Recall value, which indicates that the classifiers produced more false negative cases than false positive ones. Considering the F-score metric, XGBoost is the best except for *Use*, in which AdaBoost performs the best.

Feature comparison. As SVM performs the best overall, we use SVM to compare the strength of different types of features in the classification task. Table 5 provides a comparison of different features in terms of the AUC metric. Overall, the contextual features only bring 0.02–2.08% performance gain. For example, we found that in the *Log* classification, the additional gain for the contextual features is 1.06%, with InParis feature performs the best (0.6%). SVM with word2vec-C performs the best and the contextual features help improve the performance but not significantly.

Category comparison. From Table 4, we observe that the two most relevant categories, *Log* and *Use*, are more distinguishable in the classification task as they both have higher AUC than other two categories. We further look into some of the salient characteristics in each category.

The category *Log* has high AUC and Precision but low Recall, indicating the classifiers produce more false negative cases. Our further examination of the misclassified cases reveals several limitations including the location identification and label ambiguity. For example, two tweets are classified as negative (i.e., not *Log*) incorrectly: *"Shooting in Paris. If you can not go home and are on Paris-West, DM #PorteOuverte"* and *"#PorteOuverte #fusillade #Paris #RT a room available at Paris 15 for tourists or families in the..."*. These tweets contain the location names ("Paris-West" and "Paris 15") in Paris; however, our feature extraction failed to recognize the location names of these tweets, in part due to our limited location lexicons and informal names (e.g, "Paris 15") used in the tweets. Some false negative cases were produced because of the ambivalence in their labels. For example, the tweet (*"#PorteOuverte to Brochant do not hesitate. Tonight we are the France."*) was coded with two labels (*Log* and *S/E*) with each sentence signaling one category. The tweet is positive in the *Log* classification, but the second sentence has a considerable portion. Therefore, the machine failed to classify this tweet correctly.

The classification of category *Use* also has high AUC, Precision and Recall. We observe that in this category, users tended to use similar words to promote the channel, e.g., *"If you're in Paris and you need a safe place to stay use the hashtag #PorteOuverte"* and *"If in Paris and need somewhere safe to go use the hashtag #PorteOuverte*). As a result, the content features (N-gram or word2vec) may capture the information of the category *Use* more easily than those in other categories.

Compared with category *Log* and *Use*, the classifications for category *S/E* and *Ref* are relatively hard. We observe that the tweets in the two categories often contain more diverse words. This can be further confirmed by the Bi-Normal Separation (BNS) scores of the content features in these categories. The mean BNS scores of the top 1000 N-grams in *S/E* and *Ref* are 0.52 and 0.58, respectively, much lower than the mean BNS scores for *Log* (1.02) and *Use* (0.77). This suggests that the content features in categories *S/E* and *Ref* may not be sufficiently discriminative for classification.

Testing with future data. In this section, we test the performance of the classifiers in a more realistic situation, which classify new tweets using previous ones. The whole dataset was sorted in chronological order and divided into two sets: the first 80% as training and the remaining 20% as testing. We trained the same classifiers using the same feature combinations described earlier. The results are shown in Table 6. Overall, the results are similar to the cross-validation results shown previously (see Table 5): F-score and AUC decrease only slightly by less than 4%, indicating the effectiveness of the classifiers to be applied to the unseen, future data.

Overall, our experiment results indicate the capacity of using machine classification to distinguish the logistically relevant messages in the improvised #PorteOuverte channel, with possibility and insights to further improve the classification performance.

7 DISCUSSION AND FUTURE WORK

In this paper we investigated the potential effectiveness of an improvised logistical channel, #PorteOuverte. Moving beyond prior work on social media use after disasters [22, 29], we reported the first empirical assessment of the distribution of message types as they relate to the specific, logistical purpose of an improvised channel.

Channel content & effectiveness. Overall, our results indicate that #PorteOuverte was significantly hampered by noise. From early in the channel's history, and through its peak use, individual sent many more irrelevant messages to the channel than relevant ones. This problem was substantially worse for non-French tweets, a set of tweets that could not be ignored as they contained almost 1/3 of the relevant messages. In addition, the overall rate of tweets to the channel indicated that it was overloaded, and thus noise had a direct, unmitigated impact.

More encouragingly, the crowd was able to make substantial improvements to the raw message flow. Crowd members "voted up" relevant tweets at a substantially higher rate than they did other message types. This was particularly effective among French-only tweets, where relevant retweets were modal in the channel and dominated on occasion. Unfortunately, the distribution of attention within the retweeted population is skewed, with some relevant offers and requests for logistical information receiving substantial attention while others received little.

Channel improvement. Our results also suggest ways that the channel can be improved through intervention. First, the crowd's attempt to intervene via retweets could be more systematically supported. As noted above, retweets do not re-post a tweet to the live feed from a hashtag, but rather "vote" it into the "Top" tweets. Promoters of ILCs may want to make such procedural nuances salient in their explanations of how to use a channel. Specifically, our results indicate that the "Top" feed is likely to contain substantially better precision and recall on relevant, logistical information than the "Live" feed. Secondly, our machine classification experiments indicate that there is an opportunity for substantial improvement through algorithmic intervention. In particular, logistical information can be identified automatically with good precision and thus promoted algorithmically. Promotional uses of the channel, which may or may not be of benefit depending on the state of the channel at a particular time, can also be identified.

Limitation and Future work. While the present work provides the first attempt to understand hashtags as self-organized channels for addressing a logistical need in the wake of a disaster, our investigation is not conclusive.

First, our results indicate the capacity of supervised machine classification to distinguish the logistically relevant messages in the improvised #PorteOuverte channel; however, the labeled training data and the testing data are both from the same disaster. A practical system must detect logistical information in real time without waiting for relevant data to be labeled after a disaster strikes. Future research might consider testing the performance of machine classifiers in a practical cross-domain scenario [33] where the classifiers is trained on tweets from previous disasters and classify tweets of a new disaster event.

Secondly, the endogenous dynamics of the channel's growth also merit attention. The dominant message-type sent to #PorteOuverte, *Information about the hashtag*, is of ambiguous value. In particular, early in the channel's development it was likely that these messages helped to diffuse the idea of #PorteOuverte, indicating a "learning curve" for the audience regarding the improvised channel. At some point, however, it appears that these messages competed for and drew attention away from the relevant, logistical messages. Future work should consider modeling the dynamics of this natural tension between the need to promote the channel and the need to use it effectively. Results of our machine classification indicate that if these time points can be identified relevant messages could be promoted or suppressed automatically.

Further, our analysis indicates that individuals select some logistical tweets for rebroadcast but ignore others. Similar to research examining which notices from official authorities get the most reshares [39], future research might consider which kinds of logistical offers or requests draw the most user attention. In addition, our research uncovered a new kind of message, *Reflexive comments*, that emerged endogenously and created noise in the channel. Modeling how and when comments about the phenomenon begin to grow could be useful in anticipating the dynamics of future ILCs.

Nevertheless, the current work makes an important initial effort to understand the dynamic and effectiveness of hashtags as self-organized ILCs during and after disasters. It sheds a light towards a more effective mechanism to detect and disseminate logistical needs in self-organized channels on social media in the wake of disasters.

ACKNOWLEDGEMENTS

This work is part of the research supported by NSF Grants #1423697, #1634944 and #1634702, and the CRDF & CIS at the University of Pittsburgh. Xingsheng He also thanks the support of China Scholarship Council. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

REFERENCES

[1] Adam Acar and Yuya Muraki. 2011. Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities* 7, 3 (2011), 392–402.

[2] Jisun An, Haewoon Kwak, Yelena Mejova, De Oger, Sonia Alonso Saenz, and Braulio Gomez Fortes. 2016. Are you Charlie or Ahmed? Cultural pluralism in Charlie Hebdo response on Twitter. *arXiv preprint arXiv:1603.00646* (2016).

[3] Leonard Bickman and Jerome D Cook. 1986. Social support following a disaster. (1986).

[4] Axel Bruns, Jean E Burgess, Kate Crawford, and Frances Shaw. 2012. #qldfloods and @QPSMedia: Crisis communication on Twitter in the 2011 south east Queensland floods. (2012).

[5] Paola-Maria Caleffi. 2015. The "hashtag": A new word or a new rule? 12, 2 (2015), 46–70.

[6] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. *arXiv preprint arXiv:1603.02754* (2016).

[7] Dharma Dailey and Kate Starbird. 2014. Visible Skepticism: Community Vetting after Hurricane Irene. In *ISCRAM 2014*. 777–781.

[8] George Forman. 2003. An extensive empirical study of feature selection metrics for text classification. *JMLR* 3, Mar (2003), 1289–1305.

[9] Huiji Gao, Xufei Wang, Geoffrey Barbier, and Huan Liu. 2011. Promoting coordination for disaster relief–from crowdsourcing to coordination. In *SBP 2011*. Springer, 197–204.

[10] Philip N. Howard and Samuel C. Woolley. 2016. Bots and Automation over Twitter during the US Election. (2016). http://politicalbots.org/wp-content/uploads/2016/11/Data-Memo-US-Election.pdf

[11] Qunying Huang and Yu Xiao. 2015. Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information* 4, 3 (2015), 1549–1568.

[12] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. *Computing Surveys (CSUR)* 47, 4 (2015), 67.

[13] Peter M Landwehr and Kathleen M Carley. 2014. Social media in disaster relief. In *Data mining and knowledge discovery for big data*. Springer, 225–257.

[14] John Boaz Lee, Michael Ybañez, Marlene M De Leon, and Ma Regina E Estuar. 2013. Understanding the behavior of filipino twitter users during disaster. *JoC* 3, 2 (2013), 37.

[15] Janette Lehmann, Bruno Gonalves, Jos J. Ramasco, and Ciro Cattuto. 2012. Dynamical classes of collective attention in twitter. In *WWW 2012*. ACM, 251–260.

[16] Yu-Ru Lin and Drew Margolin. 2014. The ripple of fear, sympathy and solidarity during the Boston bombings. *EPJ Data Science* 3, 1 (2014), 1–28.

[17] Yu-Ru Lin, Drew Margolin, Brian Keegan, Andrea Baronchelli, and David Lazer. 2013. #Bigbirds Never Die: Understanding Social Dynamics of Emergent Hashtags. In *ICWSM 2013*.

[18] Debanjan Mahata, John R Talburt, and Vivek Kumar Singh. 2015. From Chirps to Whistles: Discovering Event-specific Informative Content from Twitter. In *WebSci 2015*. ACM, 17.

[19] David J. Mendonca and William Al Wallace. 2007. A Cognitive Model of Improvisation in Emergency Management. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 37, 4 (July 2007), 547–561.

[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS 2013*. 3111–3119.

[21] Tahora H Nazer, Fred Morstatter, Harsh Dani, and Huan Liu. 2016. Finding requests in social media for disaster relief. In *ASONAM 2016*. IEEE, 1410–1413.

[22] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *CSCW 2015*. ACM, 994–1009.

[23] Ruth Page. 2012. The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. 6, 2 (2012), 181–201.

[24] Leysia Palen, Kate Starbird, Sarah Vieweg, and Amanda Hughes. 2010. Twitter-based information distribution during the 2009 Red River Valley flood threat. *ASIS&T* 36, 5 (2010), 13–17.

[25] Leysia Palen, Sarah Vieweg, Sophia B. Liu, and Amanda Lee Hughes. 2009. Crisis in a networked world features of computer-mediated communication in the April 16, 2007, Virginia Tech Event. 27, 4 (2009), 467–480.

[26] Lisa Posch, Claudia Wagner, Philipp Singer, and Markus Strohmaier. 2013. Meaning as collective use: predicting semantic hashtag categories on twitter. In *WWW 2013*. 621–628.

[27] Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. 2013. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* 19, 1 (2013).

[28] Hemant Purohit, Andrew Hampton, Shreyansh Bhatt, Valerie L Shalin, Amit P Sheth, and John M Flach. 2014. Identifying seekers and suppliers in social media communities to support crisis coordination. *CSCW 2014* 23, 4-6 (2014), 513–545.

[29] Yan Qu, Chen Huang, Pengyi Zhang, and Jun Zhang. 2011. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In *CSCW 2011*. ACM, 25–34.

[30] Raquel Recuero, Gabriela Zago, Marco T Bastos, and Ricardo Arafljo. 2015. Hashtags Functions in the Protests Across Brazil. *SAGE Open* 5, 2 (2015).

[31] Manuel Gomez Rodriguez, Krishna Gummadi, and Bernhard Schoelkopf. 2014. Quantifying information overload in social media and its impact on social contagions. In *ICWSM 2014*.

[32] Daniel M. Romero, Brendan Meeder, and Jon Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *WWW 2011*. ACM, 695–704.

[33] Koustav Rudra, Subham Ghosh, Niloy Ganguly, Pawan Goyal, and Saptarshi Ghosh. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In *CIKM 2015*. ACM.

[34] Christopher Slagh. 2010. Managing chaos, 140 characters at a time: How the usage of social media in the 2010 Haiti crisis enhanced disaster relief. (2010).

[35] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M. Mason. 2014. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In *iConference 2014*. iSchools.

[36] Kate Starbird, Leysia Palen, Amanda L Hughes, and Sarah Vieweg. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *CSCW 2010*. ACM, 241–250.

[37] Kate Starbird and Jeannie Stamberger. 2010. Tweak the tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. (2010).

[38] Zachary C Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani, and James Fowler. 2015. Online social networks and offline protest. *EPJ Data Science* 4, 1 (2015), 19.

[39] Jeannette Sutton, C. Ben Gibson, Nolan Edward Phillips, Emma S. Spiro, Cedar League, Britta Johnson, Sean M. Fitzhugh, and Carter T. Butts. 2015. A cross-hazard analysis of terse message retransmission on Twitter. *PNAS* 112, 48 (2015), 14793–14798.

[40] Jeannette Sutton, Leysia Palen, and Irina Shklovski. 2008. Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires. In *ISCRAM 2008*. 624–632.

[41] István Varga, Motoki Sano, Kentaro Torisawa, Chikara Hashimoto, Kiyonori Ohtake, Takao Kawai, Jong-Hoon Oh, and Stijn De Saeger. Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster.

[42] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *SIGCHI 2010*. ACM, 1079–1088.

[43] Xidao Wen and Yu-Ru Lin. 2016. Sensing Distress Following A Terrorist Event. In *SBP-BRiMS 2016*.

[44] Xidao Wen, Yu-Ru Lin, and Konstantinos Pelechrinis. 2016. PairFac: Event Analytics through Discriminant Tensor Factorization. In *CIKM 2016*.

[45] Lei Yang, Tao Sun, Ming Zhang, and Qiaozhu Mei. 2012. We know what@ you# tag: does the dual role affect hashtag adoption?. In *WWW 2012*. ACM, 261–270.

[46] Michele Zappavigna. 2015. Searchable talk: the linguistic functions of hashtags. 25, 3 (2015), 274–291. http://www.hca.westernsydney.edu.au/gmjau/?p=1762

[47] Hang Zhang and Vinay Setty. 2016. Finding diverse needles in a haystack of comments: social media exploration for news. In *WebSci 2016*. ACM, 286–290.

[48] G.K. Zipf. 1999. *The psycho-biology of language: An introduction to dynamic philology*. Routledge, London.

[49] Matthew Zook, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy* 2, 2 (2010), 7–33.