Journal of Operations Management 45 (2016) 123–133

Contents lists available at ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises

Eunae Yoo [a], William Rand [b], Mahyar Eftekhar [a], Elliot Rabinovich [a,*]

[a] Department of Supply Chain Management, W. P. Carey School of Business, Arizona State University, P.O. Box 874706, Tempe, AZ 85287-4706, USA
[b] Department of Business Management, Poole College of Management, North Carolina State University, 2801 Founders Drive, Raleigh, NC 27695, USA

ARTICLE INFO

Article history:
Accepted 4 May 2016
Available online 9 July 2016
Accepted by: Mikko Ketokivi

Keywords:
Humanitarian operations
Social media
Information diffusion theory

ABSTRACT

The rapid diffusion of information is critical to combat the extreme levels of uncertainty and complexity that surround disaster relief operations. As a means of gathering and sharing information, humanitarian organizations are becoming increasingly reliant on social media platforms based on the Internet. In this paper, we present a field study that examines how effectively information diffuses through social media networks embedded in these platforms. Using a large dataset from Twitter during Hurricane Sandy, we first applied Information Diffusion Theory to characterize diffusion rates. Then, we empirically examined the impact of key elements on information propagation rates on social media. Our results revealed that internal diffusion through social media networks advances at a significantly higher speed than information in these networks coming from external sources. This finding is important because it suggests that social media networks are effective at passing information along during humanitarian crises that require urgent information diffusion. Our results also indicate that dissemination rates depend on the influence of those who originate the information. Moreover, they suggest that information posted earlier during a disaster exhibits a significantly higher speed of diffusion than information that is introduced later during more eventful stages in the disaster. This is because, over time, participation in the diffusion of information declines as more and more communications compete for attention among users.

1. Introduction

The management of humanitarian operations during disasters is often highly complex due to the extreme uncertainty and diversity of stakeholders involved in these crises (Van Wassenhove, 2006). In such instances, gathering and sharing timely information regarding infrastructure, supply of resources, and needs is critical to develop an understanding of existing conditions and coordinate an effective response (Pettit and Beresford, 2009). To that end, researchers have stressed the importance of rapid information diffusion for humanitarian organizations (HOs) to gather intelligence about conditions in affected communities (Oloruntoba and Gray, 2006) and for HOs to distribute information among stakeholders in order to foster collaboration (Altay and Pal, 2014).

Internet-based social media hosted on platforms like Twitter or Facebook may help facilitate information diffusion because they provide the means through which stakeholders can upload and share information with others in real-time and at virtually no cost. Many HOs have recognized the value of social media platforms and have started using them to access and share information from various sources. This includes data from informants with first-hand knowledge of what is occurring in affected areas (Gao et al., 2011), and recently, HOs have aggregated these data to create crisis maps showing landmarks like damaged infrastructure and shelters (Meier, 2012). HOs have also used social media to share capacity levels and resource availability to enhance coordination among stakeholders (Sarcevic et al., 2012).

Despite these experiences, and calls by experts for additional research on the use of social media for humanitarian operations (e.g., Holguín-Veras et al., 2012; Kumar and Havey, 2013), the literature on this subject is still at an embryonic stage. Most of this work has focused on descriptions and characterizations of social media responses to humanitarian crises (e.g., Kaigo, 2012; Kogan et al., 2015) and has yet to rigorously consider the dynamics of information dissemination during these events and their influence on humanitarian operations.

* Corresponding author.
E-mail addresses: Eunae.Yoo@asu.edu (E. Yoo), billrand@gmail.com (W. Rand), Eftekhar@asu.edu (M. Eftekhar), Elliot.Rabinovich@asu.edu (E. Rabinovich).

http://dx.doi.org/10.1016/j.jom.2016.05.007
0272-6963/

Our paper addresses this deficiency by analyzing diffusion dynamics of information in social media from a disaster case. To that end, we follow Ellison et al. (2007) and focus on a network representation of social media platforms on the Internet in which users can forge connections and share information directly with each other, as well as indirectly through other users. These connections will form social media networks in which information produced by a user (i.e., an originator) will create cascades when those connected directly to her receive it and, in turn, share it with those with whom they are connected. These information cascades will continue to spread as long as more users join these cascades by sharing the information they receive with those connected to them.

To address this objective, we develop and test a set of theoretical propositions regarding the role played by three key determinants of information diffusion dynamics in social media networks. Although past work has discussed the importance of these determinants in the crisis informatics literature (e.g., Ringel Morris et al., 2012; Starbird and Palen, 2010; Vieweg et al., 2010), their impact on information diffusion across social networks remains undetermined. The first determinant focuses on the *influence* that information cascade originators have in these networks as a function of their social connections. The second one focuses on the *type of content* being shared in these networks and whether it contributes to improving situational awareness during a crisis. The third determinant corresponds to the *timing* in the introduction of information in these networks with respect to the progression of disaster events. Since the propositions focus on characteristics of cascades, the unit of analysis in our study is a cascade.

Our results show that information can spread faster when it originates from users that are influential in these networks. They also indicate that the timing when information is initially posted by an originator relative to a disaster's development of events will impact the information's rate of diffusion across social media networks. Information that is originally posted later, as a disaster intensifies, will spread at a lower rate than information that is posted at earlier stages of the disaster because, over time, participation in the diffusion of information cascades declines as more cascades compete for attention among users. This phenomenon underscores a paradox in which as a disaster's effects build up, there will be more cascades contributed by originators, but the information in those cascades will spread more slowly.

In the next section, we expand on theoretical explanations for the diffusion of information on social media networks and develop the propositions that guided our study. In Section 3, we detail how we collected the data and operationalized the variables to test the propositions. We then present the empirical model and the results pertaining to the evaluation of the propositions in Section 4, followed by a discussion of the results, implications, and conclusions in Section 5.

2. Information diffusion on social media networks: background, theory, and propositions

Research based on Information Diffusion Theory has relied on different types of models of adoption to explain the dynamics of information cascades' diffusion in network settings. Two of the seminal models are the *Independent Cascade* (IC) model developed by Goldenberg et al. (2001) and Kempe et al. (2003) and the *Linear Threshold* (LT) model developed by Granovetter (1978). These models assume each member contributes monotonically to the diffusion of information (i.e., there is no dis-adoption or forgetting of the information). In these models, information diffusion proceeds iteratively over time starting from a set of members that contribute information to be subsequently distributed by other members across the network (Guille et al., 2013). IC and LT models also account for information diffusion due to a member receiving information from sources external to the network or internally from those informed participants that are adjacent to her in the network (Myers et al., 2012).

IC and LT models, however, differ from each other in several aspects. IC models assume that an informed member has one chance at a time of independently sharing information with one uninformed member adjacent to her in the network (Kempe et al., 2003). Thus, at any point in time, an uninformed member has a likelihood, q, of becoming aware of the information when at least one of her neighbors in the network has already become aware of the information. But, in many versions of the IC model (Goldenberg et al., 2001), there is also a probability, p, that the individual will become aware of this information from external sources. High values for q and p will denote a high information diffusion rate throughout the network due to the internal influence of network connections or influence of sources external to the network, respectively (Guille et al., 2013).

In LT models, it is assumed that a participant will share information with her uninformed neighbors in the network if, over time, the number of informed members adjacent to her in the network exceeds her own influence threshold (Granovetter, 1978). The lower this threshold across the network, the faster the participant will share information with her uninformed neighbors and the faster information will diffuse internally throughout the network. In prior work, this threshold is denoted by ϕ (Watts and Dodds, 2007). In our paper, we operationalize this threshold by setting $\phi = 1 - q$. This allows us to maintain a relationship consistency with the IC model where high values of q indicate faster diffusion, and low values of q indicate slower diffusion. In some prior work, the q parameter is fixed for all individuals, while in other contexts it is chosen from a distribution for each individual (Watts, 2002). Traditionally, the LT model has not incorporated a p parameter, instead relying on the initial seeds of the network to propagate the information (Kempe et al., 2003; Watts and Dodds, 2007), but a p parameter playing the same role that it does in the IC model can be added to this model instead of an initial seed (Dodds and Watts, 2005).

Though previous work has created a generalized model that incorporates both the IC and LT models (Dodds and Watts, 2005), we developed a framework that allows for versions of both the IC and LT models to be described using the same two parameters of p and q. To that end, we modeled the user decision process in the following sequential steps:

(1) **Effect of *p***: Independent of the adoption model (LT or IC), each agent who has not yet adopted the information adopts the information with probability p due to discovering the information from a source of information diffusion outside the network structure.
(2) **Effect of *q***: Depending on the adoption model, users take different actions.
   a. ***q* in the LT model:** Each user who has not adopted observes the number of neighbors who have adopted divided by the total number of neighbors they have. If that ratio exceeds ϕ, the focal user adopts the information (Watts and Dodds, 2007).
   b. ***q* in the IC model:** Each user who adopted information in the most recent previous time step has q probability of transmitting the information to any neighbor who has not adopted the information (Goldenberg et al., 2001).

Though each of these models has found success in analyzing diffusion processes (e.g., Goldenberg et al., 2001; Guille et al., 2013; Rand et al., 2015; Watts and Dodds, 2007), it is not obvious whether both models can be used jointly in studying information diffusion on social media networks in the same context. As part of our contribution to the literature, we will first examine how IC and LT models explain these cascades' diffusion dynamics within the same context. Then, we will use this analysis to focus our line of inquiry on the effects of the three diffusion determinants we introduced in Section 1. We will expand on these determinants' effects below.

2.1. The effect of influential originators on the diffusion of cascades

The diffusion of an information cascade will depend on the level of *influence* that the cascade's originator carries in the social network. An originator's influence is particularly relevant to the context of cascades in social media networks during humanitarian crises since users previously reported having significant concerns about the credibility of disaster information they received through social media (Ringel Morris et al., 2012). While influence can be assessed in a number of different ways, prior results from information diffusion models concentrate on influence measured by a user's number of social connections and suggest that users with large network audiences are perceived to have superior credibility (Bhattacharya and Ram, 2012). These perceptions will allay concerns about trustworthiness and induce individuals to conform to cascades launched by influential originators (Goldenberg et al., 2009). Based on this evidence, we expect that users will be inclined to join cascades originated by network members with extensive influence, and as a result, these cascades will exhibit greater rates of internal diffusion.

Moreover, research has relied on the principle that influential cascade originators usually have numerous social connections that will expose large audiences to their cascades soon after they are launched (Kempe et al., 2003). This implies that if a cascade's originator is well-connected, the cascade will diffuse rapidly because a wider audience will be exposed early on to the cascade. We anticipate that this principle will also apply in the context of information diffusion in social media networks during a disaster. Hence, we conjecture that an information cascade's diffusion may experience a surge soon after a highly influential user exposes the cascade's information to her network links. This will contribute to the cascade's overall rate of diffusion throughout the social media network. Proposition 1 summarizes this argument for our setting.

Proposition 1. In the context of cascades carrying disaster-related information throughout social media networks, the influence of a cascade's originator contributes positively to the cascade's speed of diffusion.

2.2. The effect of content promoting situational awareness on the diffusion of cascades

Research shows that diffusion rates will increase if network members perceive that cascades' contents are informational and that sharing these contents will be helpful to others (Rogers-Pettite and Herrmann, 2015). Based on this evidence, we argue that, during humanitarian crises, network members are more inclined to participate in cascades carrying informational content that is seen as useful to disaster relief operations. For many of these members, the decision to join cascades conveying informational content related to disaster relief will follow altruistic and emotional motivations to help victims. In joining these cascades, these members anticipate no material gains. Instead, they look to obtain rewards resulting from their cooperation with other cascade participants and from offering support to others in need (Fowler and Christakis, 2010).

In a humanitarian context, these information cascades will convey content that will heighten *situational awareness*. Situational awareness, in itself, is defined as a complete and coherent understanding of what is going on during emergencies, and it is gained from information that helps to assess the situation at hand (Sarter and Woods, 1991; Vieweg et al., 2010). In humanitarian operations, information supporting situational awareness is vital because decision parameters are highly dynamic (Holguín-Veras et al., 2012). Hence, situational awareness is required to make decisions that are well-informed and reflective of current events.

Given the value of situational awareness, we expect that network members will have a greater disposition to join cascades that carry information that could improve situational awareness. Our expectation follows evidence showing that cascades with information that improves situational awareness exhibit greater participation among social media users (Vieweg et al., 2010). Thus, messages meant to improve situational awareness during a crisis are likely to strengthen the diffusion of information cascades across social networks. Proposition 2 formalizes this argument.

Proposition 2. In the context of cascades carrying disaster-related information throughout social media networks, speed of diffusion will be higher for cascades carrying information that heightens situational awareness than for cascades carrying other types of information.

2.3. The effect of timing in the launch of cascades on the diffusion of cascades

Past work on information diffusion has underscored the role played by temporal patterns in the dissemination of information across networks. As part of this body of work, Boyd et al. (2010) identified a preference by participants in social media networks to share time sensitive information with others. This is particularly relevant in a humanitarian context, in which participants will be motivated to share urgent information that will help address directly their own needs and those of others in the network.

Leskovec et al. (2009) argued that the level of motivation among network participants to share time-sensitive information will contribute to the likelihood of certain topics gaining initial traction among network participants and eventually forming a cascade. These topics, for example, may comprise the development of urgent news events during a humanitarian crisis. At an early stage during a disaster, cascades addressing such topics will spread quickly as more participants imitate one another in sharing information. But over time, the rate of participation in the diffusion of cascades will decline as newer topics compete with older ones for attention. As a result, the diffusion of new cascades is likely to become increasingly difficult, regardless of the urgency embedded in an information cascade. Cascades that are launched at later stages during the course of a crisis are therefore expected to diffuse at a lower rate than cascades launched at earlier stages. That is, the diffusion of information cascades on social media networks will decline as a disaster unfolds. Proposition 3 formalizes this argument.

Proposition 3. In the context of cascades carrying disaster-related information throughout social media networks, the speed of diffusion will be lower for cascades that are launched later than for cascades launched earlier during the progression of a disaster event.

3. Research methodology

3.1. Context: Twitter and Hurricane Sandy

We focused on Twitter to test our propositions. Social networks on Twitter are based on directional links between users. On Twitter, a user can follow, or track, the messages (or "tweets") of another user or be followed by other users (called "followers"). Users can receive the tweets of those they follow and broadcast all of their own tweets to their followers. Twitter also gives a user the ability to "retweet" original tweets or other retweets posted by users that she follows in order to share these messages with her own followers. A user's retweets preserve the contents of the original message, and these retweets may be shared in turn by the user's own followers, who may or may not be a part of the network of the user who uploaded the original tweet.

Our study focused on Twitter data associated with Hurricane Sandy, a disaster for which Twitter usage has received some research attention (e.g., Rand et al., 2015). Hurricane Sandy is considered to be the largest Atlantic hurricane on record in the United States (U.S.). It began as a tropical storm in the Caribbean in October of 2012, grew into a Category 3 hurricane at its peak, and impacted the Eastern U.S. We determined Hurricane Sandy to be an appropriate disaster case for our study for two reasons. First, the hurricane's major effects were felt in a densely populated, highly developed area. Because of the hurricane's magnitude and Twitter's popularity in this area, a large volume of tweets were posted in relation to this event, creating a rich dataset for empirical analyses. Second, as the main effects of Hurricane Sandy were felt in the U.S., tweets were mostly sent in English. This eliminated the need for translation to address our research objectives.

3.2. Data collection

Our data contain original tweets and retweets posted from October 26 until October 30, 2012. These dates correspond to the periods before, during, and after Hurricane Sandy effects were experienced in the U.S and overlap with the stages when preparation and response activities to the hurricane occurred. Preparation and response stages are usually the most relevant for humanitarian operations in many disasters as high levels of uncertainty and volatility in conditions on the ground are pervasive at these times (Van Wassenhove, 2006).

The collected data include the actual contents of the tweets and retweets, information about the users responsible for these posts, and the date and time, to the second, when each of the posts appeared on Twitter. The data were gathered in real-time using Twitter's Search API, an interface through which one can program queries to collect tweets and retweets posted within the past seven days. Twitter limits the amount of data that can be downloaded per IP address using the Search API. To overcome this limit, a script using the Search API was run constantly on ten different machines with a rule that would pull tweets and retweets containing the keywords "Sandy," "hurricane," "storm," and/or "superstorm". Based on the volume of data downloaded, we were confident that the Search API extracted a high percentage of the tweets and retweets that contained our search keywords during our data collection period. Nevertheless, we decided to evaluate the completeness of the data gathered through the Search API by comparing it against a sample we acquired from Gnip, a Twitter subsidiary with access to the entire Twitter firehose (i.e., all activity ever posted on Twitter). To draw the Gnip sample, we used identical keywords and date ranges to those specified for the Search API sample. Our comparison demonstrated that the Search API only missed 7.81% of the messages in the Gnip dataset. This suggests that our sample contains a vast majority of the tweets and retweets posted during Hurricane Sandy and with the selected keywords.

Subsequently, we used a program to separate the original tweets from the retweets that the Search API extracted. We manually reviewed all of the original tweets and filtered out those that we deemed irrelevant along with their retweets. Although they contained the chosen keywords, irrelevant tweets included jokes, song lyrics, emotional responses, and discussions of topics unrelated to Hurricane Sandy. Please refer to Table 1 for more detail on irrelevant tweets. After removing the irrelevant messages, we were left with 18.27% of the original tweets in the sample along with their retweets.[1] In total, these tweets and retweets corresponded to 333,968 messages.

Because our propositions dealt with information cascade effects, the unit of analysis for our study is the cascade. In view of this, we organized the tweets and retweets in the dataset into cascades. We followed the lead of authors who have previously conceptualized information cascades in Twitter as retweet chains (e.g., Galuba et al., 2010; Lerman and Ghosh, 2010). Each original tweet represented the start of a cascade, and retweets by additional users signaled participation in a cascade. In Twitter, the text in all retweets is usually identical to the text in the original tweet that launched the cascade since Twitter makes retweets possible through the push of a single button. Retweets are also marked at the beginning by "RT@username," followed by the original tweet's text. The username following "RT@" identifies the user that posted the original tweet and launched the cascade.

Based on these attributes, we compiled cascades in our data by identifying and grouping retweets that shared the same text and embedded originator usernames. Then, to ensure that each group of matching retweets constituted an actual cascade and not background conversations among select users, we confirmed that each cascade consisted of at least ten retweets issued at varying intervals. This process generated 5683 cascades. We chose a threshold of ten retweets because cascades on Twitter usually do not require many retweets to develop (Lerman and Ghosh, 2010).

We then developed a program to examine in detail the original tweets that began each cascade. Through this program, we isolated the username embedded in the beginning of each retweet's "RT@username" and separated the original tweet's text that followed. Then, the program searched through the dataset and pulled each original tweet with the matching username and content. In this process, we found that 249 cascades (comprised of 19,558 retweets) could not be matched to their original tweet because they had missing information about the originating users.[2] This prevented us from identifying the time when each of these cascades started, and therefore, we were unable to examine their diffusion. Although this left us with no option but to drop these cascades from our sample, the removal of these cascades had a negligible impact on our results since they constituted only 4.3% of our observations. After we filtered these cascades, we were left with a final sample of 311,429 retweets forming 5434 cascades to evaluate our propositions. Table 2 shows the distribution of the cascades across six content categories.

3.3. Operational measures

In this section, we expand on the operationalization of the variables introduced in the propositions. Moreover, we introduce a set of control variables to be used as part of the empirical testing of these propositions. Table 3 lists the variables in the propositions and the control variables along with their operationalization.

[1] The process of cleaning and categorizing the cascades took approximately 45 h to complete.

[2] This information may be missing from the data because privacy settings chosen by the originators did not allow the Search API to access this information or because the original tweet was posted before the start of the data collection.

Table 1
Irrelevant tweets.

Irrelevant category	Example tweet
Emotional response	"actually really scared of the hurricane coming:("
Joke	"Hurricane Sandy sounds like a delicious mixed drink."
Not related to Sandy	"Yay!!(: hanging out with my bestfriend @strong_sandy"
Opinion	"I get the feeling this hurricane in gonna be just like irene and barley [sic] hit us.."
Song lyric	"The voice that calmed the sea would call out through the rain and calm the storm in me … -Casting Crowns. I love this song! #whoami"
Vague forecast	"Sandy is coming"

Table 2
Breakdown of cascade categories[a].

Category	Count of cascades	Description	Sample retweet
Advisories	2024	Transportation shutdowns, evacuation warnings, survival/safety tips, and updates on hurricane intensity/trajectory	RT @Timcast: Reports that all NYC bridges will be closing at 7pmEST via @NYScanner #Sandy #Frankenstorm
Business	445	Reports of business-related shutdowns and forecasts of economic impacts	RT @Reuters_Biz: Stock bond markets shut on Tuesday may reopen Wednesday http://t.co/JL6fEHea
Declarations	141	Declarations of emergencies by states	RT @USNationalGuard: So far governors in MD VA NY DC PA CT NC NJ DE MA and VT have declared states of emergency ahead of #Hurricane #Sandy.
Forecasts	640	Forecasts of weather and hurricane effects	RT @twc_hurricane: BREAKING: TWCs experts now expect localized wind gusts of 90 + mph near the coast of NJ NYC and Long Island later today. #Sandy
Humanitarian	246	Information related to shelters, relief efforts, and deployment of aid	RT @femaregion2: #Sandy Search for open shelters by texting: SHELTER + a zip code to 43362 (4FEMA). Ex: Shelter 01234 (std rates apply)
Reports	1938	Status updates of weather, damage, outages, etc.	RT @News12LI: As of 10:32 a.m. LIPA is reporting 15695 outages across Long Island. #Sandy

[a] Adapted from Olteanu et al. (2014) and Vieweg et al. (2010).

Table 3
Variable operationalization.

Construct	Variable label	Operationalization
Information cascade's diffusion speed	*DIFFUSION*	Ratio of q/p values obtained from the *IC* model
Cascade originator's influence	*INFLUENCE*	Number of users following the cascade originator (at the time of cascade launch)
Cascade content's contribution to situational awareness	*AWARENESS*	Dummy variable coded 1 if the cascade content contributed to situational awareness; 0 otherwise
Lateness in the launch of the cascade during the disaster	*LATENESS*	Lag in the launch of the cascade relative to the start of the data collection (measured in hours)
Incidence of cascade boosts by originator	*BOOST*	Dummy variable coded 1 if originator boosted the cascade; 0 otherwise
Misleading cascade	*FALSE*	Dummy variable coded 1 if the cascade content was misleading; 0 otherwise

3.3.1. Dependent variable

To measure the cascades' diffusion speed on Twitter's network, we followed Rand's et al. (2015) approach and ran an agent-based model (ABM) to evaluate how well the IC and LT models we introduced in Section 2 represented the cascade data. This generated an overall adoption rate of information at discrete time steps. ABM offers a robust understanding of information diffusion on social networks since it represents not only the properties of the individual agents but also their communication channels via local network connections. Rand and Rust (2011) identify up to six properties of a system that make it useful to model using ABM: (1) a medium number of agents, (2) local and potentially complex interactions among agents, (3) agents' heterogeneity, (4) rich environments, (5) temporal aspects, and (6) agents' adaptability. Information diffusion on social media features all six of these properties to an extent, making ABM a suitable method for our study. Please refer to Part I of the electronic appendix that accompanies this paper for a more detailed discussion of the appropriateness of ABM. The ABM was constructed, verified, and validated following the guidelines of Rand and Rust (2011). Parts II through IV of the electronic appendix contain supplemental information of model construction, verification, and validation beyond the details given below, and Part VI shows the natural language version of the code used to create the ABM.

There were two basic entities in the ABM: (1) a Twitter user interested in receiving and transmitting information and (2) the relationship between each pair of users in a cascade, i.e., a social tie or a link. Ties between users enabled the transmission of information across each cascade. In Twitter, two users are connected to each other if one of the users follows the other and/or vice versa. Thus, the agents in the ABM possessed a set of links that corresponded to the social links of each user to other users based on their "following" relationships. We patterned these relationships against the links observed across a sample of 4076 participants in the longest cascade in our dataset. Using Twitter's RESTful API, we identified the users followed by each cascade participant at the time of Hurricane Sandy. This yielded a total of 1,322,814 links, of which 3315 served to cascade the information by being direct connections between users who were part of the cascade. Because of the rate limits on Twitter's RESTful API, it would have taken a prohibitive amount of time to pull all of the networks for each cascade. Therefore, we used the network for the longest cascade as the pattern for all of the other cascades we examined. While this decision simplified the modeling process, it is not a major limitation since Twitter exhibits scale-free properties, meaning that subnetworks are similar to their corresponding larger networks

(Kwak et al., 2010).

Since the main observation in the ABM was the overall adoption of information at each time step for each cascade, agents had a property that specified whether or not they had adopted new information. By adoption of new information, we mean the joining of a cascade by retweeting. In addition, agents had both a coefficient of external influence (p) and a coefficient of internal influence (q) that controlled the rate of adoption of a new piece of information in each cascade following external or internal stimuli, respectively. At the beginning of the ABM, all agents started in an "un-adopted" state, and a directed social network linking the agents was formed based on the empirical Twitter networks described above. Then, at each time step, any agents that still had not adopted the information decided whether to adopt the information based on p, q, and the state of their neighbors in the network. Agents followed the unified model discussed in Section 2 to make these decisions. The agents first chose whether to adopt based on external influence. To do this, they drew a random number from the uniform distribution of [0,1). If that number was less than p, they then adopted that information. This decision rule for external influence was identical regardless of whether the LT or IC models were considered.

The role of internal influence of network links was subsequently considered. In the LT model, each agent counted the number of neighbors that had adopted the information and divided this sum by the total number of neighbors. The agent then compared this number to $\phi = 1 - q$, and if the ratio was higher than ϕ they proceeded to adopt the information. In the IC model, each agent who adopted the information in the most recent time step transmitted the information to all of its neighbors who had not adopted. These uninformed agents drew a random number from the uniform distribution of [0,1), and if the number was less than q, they adopted the information. After all of the non-adopting agents had considered whether or not to adopt according to the rules described above, statistics on the number of adoptions that occurred during that time step were calculated. The model then iterated again until every agent in the network had adopted the information. We calibrated our model so that a time step was roughly one minute. This enabled a seamless comparison to the observed data, which was also set in a resolution of one-minute increments.

The ABM provided observations for each cascade on the adoption of information at each time step for the IC and the LT models. We then compared this information to the empirical data to determine for each adoption model and each cascade which values of p and q best matched the empirical data. To complete this task, we used a simulated annealing (SA) approach. This method works by generating iterative values of p and q and measuring the performance of the model between the time series of the model data and the observed data for each cascade until identifying the parameter values for p and q that optimize this performance. We chose to use SA since a full search of the parameter space was precluded by the computational cost, and SA provides a robust way to search the space quickly for a set of parameters that minimizes errors. For technical details on the number of runs and implementation of the SA algorithm, please refer to Part V of the electronic appendix.

To estimate the performance measure from each model run for each cascade network, we obtained values for $Y(t)$, the number of agents in the network who had adopted the information at each time step, t. Next, we compared $Y(t)$ to the actual number of adopters per time step from our empirical data, $Empirical(t)$, using the Mean Absolute Percentage Error (MAPE). As Equation (1) shows, the MAPE is equal to the absolute difference between the empirical value of information adoption observed at time step, t, throughout the duration of the cascade and the ABM's value at that same time step, divided by the empirical value at time step t and averaged over all values (n).

Table 4
Descriptive statistics for *MAPE* values.

Model	Median	Mean	SD	Min	Max
IC	13.36	22.91	94.58	0.61	5611.07
LT	12.97	21.89	86.26	1.08	4821.76

$$MAPE = 100 \times \frac{1}{n} \sum_{t=0}^{n} \frac{|Empirical(t) - Y(t)|}{Empirical(t)} \tag{1}$$

We then averaged the MAPE across k runs. For a sample of the cascades, we observed that the average MAPE did not change markedly with more than ten runs for a given parameter setting. Thus, we chose to use ten model runs to provide an adequate estimate of the underlying adoption patterns for a given cascade network and a given set of parameters. It was this average MAPE value over ten runs that was then used by the SA approach to optimize the parameter values.

Table 4 provides a distribution of the MAPE values across all the cascades for the IC and LT models. A comparison of the MAPE values for p and q across the cascades revealed that the MAPE values for p and q were consistently low across the IC and the LT models and similar to values identified for this metric in previous studies (Rand et al., 2015). Since both the IC and LT models performed well, we chose to focus on the IC model to operationalize information cascades' diffusion speed as our dependent variable (henceforth labeled as *DIFFUSION*). This is because the IC model allowed for a more direct measurement of *DIFFUSION* as the ratio of q/p values obtained from the model's output.[3] By operationalizing the dependent variable as q/p, we were able to account for diffusion forces due to sources internal and external to the networks underlying the cascades.

3.3.2. Determinants

We are interested in investigating three determinants: (1) the cascade originator's influence, (2) the cascade content's contribution to improving situational awareness, and (3) the timing of the launching of the cascade. To measure a cascade originator's influence, we followed Cha et al. (2010), who explained that an agent is influential when it acts as an information channel to a large audience. This is consistent with opinion leadership models that support the notion that individuals are influential when they have a high number of connections with others (Bonacich, 1972). Thus, we measured a cascade originator's influence as the number of the originator's followers on Twitter at the time the cascade was launched (*INFLUENCE*).

The second explanatory variable serves to identify those cascades that spread information related to situational awareness. To identify whether a cascade included this type of content, we

[3] In the IC model, q represents a probability of internal influence, i.e., adoption due to internal influence is q multiplied by the fraction of neighbors who have adopted. Therefore, q/p describes the difference in spreads due to internal influence vs. external influence. However, in the LT model, q is a measure of how low the threshold to adoption is due to internal influence. This is different than a probability of adoption. Hence, q in the LT model is not directly comparable to p in the LT model since p is a direct measure of the probability of adoption due to external influence. This makes it difficult to make direct claims about the rate of internal vs. external adoption in the LT model based on these parameters. Nevertheless, since we developed the ABM under both IC and LT models, the ABM could serve to evaluate which rules cause the agents to adopt, and, from that, count up the number of agents that adopt due to internal influence and external influence in the LT model and compare those numbers to gauge diffusion speed indirectly.

created a dummy variable using the categorization scheme introduced in Section 3.2. This dummy (*AWARENESS*) equals 1 if the cascade belonged to advisories, humanitarian, or reports categories since, as detailed in Table 2, all dealt with information about safety, shelters, or the functional state of the affected areas. Otherwise, *AWARENESS* equals 0. We validated this operationalization by having four raters independently classify whether a randomly sampled set of 100 cascades pertained to situational awareness as defined in this study. We then checked the inter-rater agreement of our and the raters' classifications using Fleiss' kappa (Fleiss, 1971). The kappa statistic was equal to 0.68, which indicates substantial agreement (Landis and Koch, 1977).

Finally, the third explanatory variable captures the timing of each cascade's launch during the disaster. To that end, we measured the difference in hours between each cascade's launch and the time when we began our data collection. By calculating these intervals, we captured how late a cascade was launched during the disaster. We labeled the variable for this measure as *LATENESS*.

3.3.3. Control variables

As part of our empirical model, we accounted for instances in which cascade originators attempted to artificially increase the rate of diffusion of information in their cascades. Therefore, our first control variable accounts for instances when users boosted (or bumped up) those cascades that they themselves originated in order to increase the cascades' visibility on Twitter. A user may attempt to give a cascade that she originated a "boost" by reposting, at least once, the same tweet that initiated a cascade. However, in doing so, the originator may contribute to artificially distorting the cascade's growth pattern and its rate of diffusion. We controlled for this effect by using a binary indicator (*BOOST*) that specified which cascades in our sample were boosted by their originators or not. We set *BOOST* to 1 if a cascade was boosted by its originator or 0 otherwise.

Moreover, we controlled for whether the information conveyed in a cascade was misleading. Prior studies have documented the circulation of manufactured information in online social networks during disasters (e.g., Kaigo, 2012). In our sample, some cascades contained information that purposefully exaggerated the size of the hurricane while others conveyed messages designed to convey outlandish claims about damages caused by the hurricane. Because such reports can generate a sense of panic among users (Gupta et al., 2013), they may artificially increase the rate of diffusion of information in these cascades. We controlled for this effect with a dummy variable (*FALSE*) that is set to 1 if the cascade's contents were false and 0 otherwise.

4. Empirical analysis

We used regression analysis to test the propositions based on Equation (2). The use of regression analysis enabled us to specify the rate of diffusion for a cascade, *i*, as a function of the explanatory and control variables discussed in Section 3.3 in addition to an error term, u_i.

$$\begin{aligned} DIFFUSION_i = \beta_0 + \beta_1\, INFLUENCE_i + \beta_2\, AWARENESS_i \\ + \beta_3\, LATENESS_i + \beta_4\, BOOST_i + \beta_5\, FALSE_i + u_i \end{aligned} \tag{2}$$

Fig. 1 shows a cumulative distribution of the cascades' originations over time, and Table 5 lists the descriptive statistics for the variables in Equation (2). Since the mean for *DIFFUSION* (37.28) is statistically higher than 1 ($p < 0.01$), our data suggest that internal information diffusion on social media networks advances at an average rate that significantly exceeds the average speed at which information originates from external sources. Please note that we limited the range of our parameters to historically observed values (Chandrasekaran and Tellis, 2007). Thus, it might be argued that we did not explore a large enough range to observe model fits with very large p values. As a robustness check, we examined the number of cascades where the optimal p values were at the maximum range of exploration we allowed. Out of 5434 cascades, only 648 of the IC model fits had p values at their maximum value, and of those 648, only 12 had the minimal q values. This means that for at least approximately 88% of our cascades, the best model fit was one where internal influence of network connections was much higher than external influence. In fact, removing the runs where p reached its maximum value changes the mean for *DIFFUSION* to 40.25, which illustrates how strong a role internal influence plays in the vast majority of these cases.

4.1. Statistical modeling

We used a Generalized Linear Model (GLM) with a gamma distribution to model Equation (2). This approach was suitable for our model because *DIFFUSION* only took on positive values and displayed a right-skewed distribution of values. Also, after probing the relationship between *DIFFUSION* and *INFLUENCE* and *LATENESS*, we observed that the variance of *DIFFUSION* increased with the mean. This is consistent with the gamma distribution ($Var\,[Y_i] = \mu^2/\nu$). Separate plots of *DIFFUSION* versus *INFLUENCE* for each of the two categories in *AWARENESS* also revealed that there were some outlying *DIFFUSION* values at extreme *INFLUENCE* values, which is another property consistent with the gamma distribution (Dobson and Barnett, 2008).

To ensure an appropriate use of GLM, we also followed several additional steps. First, we used a Pearson Chi-Squared estimation method to estimate the GLM scale parameter (McCullagh and Nelder, 1989). Second, we examined a log link function and an identity link function as possible alternatives to transform the dependent variable to estimate the GLM. Although the GLM results were fully consistent across both link functions, the identity link function provided significantly better Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) fit measures than the log link function. Thus, the results we report in this paper correspond to those obtained using the identity link function (Hardin and Hilbe, 2007). The results obtained using the log link function are available upon request. Third, we used the Huber-White sandwich estimators to estimate standard errors that are robust to possible misspecification of the variance and link functions in the GLM. Finally, we checked for multicollinearity among the explanatory and control variables and found that almost all correlations among these variables were fairly small (Table 5).

4.2. Results

Table 6 presents the results from the GLM. To generate these results, we used a hierarchical approach. We first considered a restricted model in which we regressed the dependent variable only upon the control variables (GLM 1). Then, we regressed the dependent variable on the control variables as well as the explanatory variables in the propositions (i.e., unrestricted model or GLM 2). The results from likelihood ratio chi-squared test of GLM 2 indicate that the group of explanatory variables is statistically significant. Significant reductions of the AIC, BIC, and Deviance measures for GLM 2 also confirm that the addition of the predictors in GLM 2 makes a statistically significant contribution in explaining our dependent variable's variance, above and beyond the contribution made by the control variables (Coxe et al., 2013; Hardin and Hilbe, 2007).

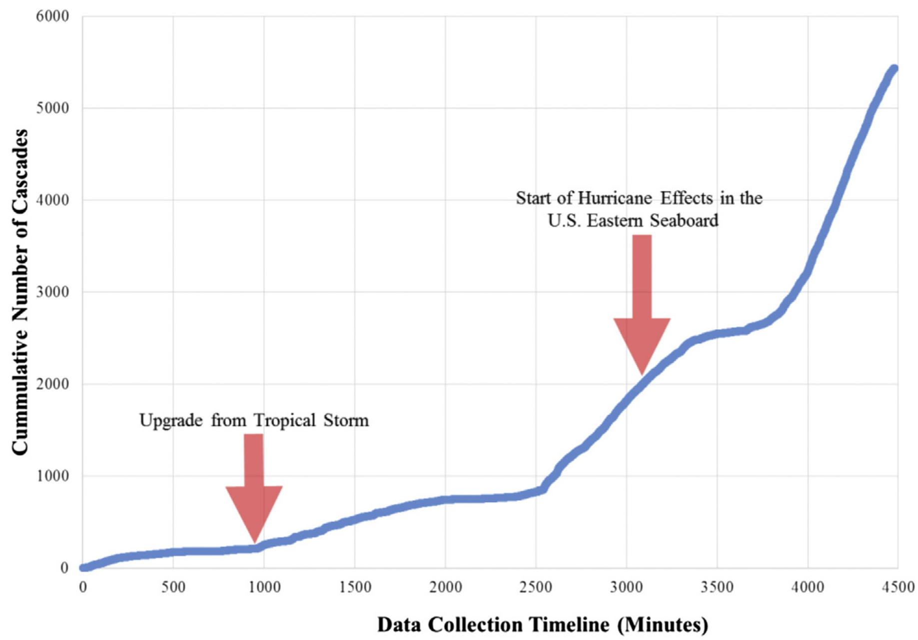

Fig. 1. Cumulative distribution of cascades over time.

Table 5
Correlations and descriptive statistics.

	1	2	3	4	5	6
1. DIFFUSION	1					
2. INFLUENCE	0.03*	1				
3. AWARENESS	−0.01	−0.04**	1			
4. LATENESS	−0.25**	0.01	0.16**	1		
5. BOOST	0.28**	0.01	0.01	−0.09**	1	
6. FALSE	−0.01	−0.04**	0.11**	0.16**	−0.03	1
Mean	37.28	234,447.66	0.78	55.34	0.02	0.04
Std. Deviation	55.63	848,551.55	0.42	18.04	0.13	0.19
Minimum	12.67	0	0	0.00	0	0
Maximum	737.80	9,133,950	1	74.65	1	1

*$p < 0.05$, **$p < 0.01$.

From Table 6, the coefficient for *INFLUENCE* was positive and statistically different from zero ($p < 0.05$). Therefore, Proposition 1 is confirmed: as a cascade originator's influence rises, the speed of information diffusion in the cascade increases. Moreover, the effect by *LATENESS* on the dependent variable was negative and statistically different from zero ($p < 0.01$). This means that, during a disaster event, the rate of information cascades' diffusion decreases over time as cascades are launched later during the disaster event. Proposition 3, therefore, is also confirmed. Proposition 2, however, received no support since the coefficient for *AWARENESS* was not significantly different from zero. Thus, we have no evidence to conclude that cascades carrying information that heightens situational awareness during a crisis will experience faster diffusion than cascades carrying other types of information. The lack of support for Proposition 2 is surprising based on theory and previous findings (Vieweg et al., 2010) but raises an important point that social media networks like Twitter may be limited in effectively spreading certain types of content. This is vital for HOs to understand as they create policies and strategies for managing information in a crisis.

Table 6
GLM results.

	GLM 1 Coefficients (Std. errors)	GLM 2 Coefficients (Std. errors)
INFLUENCE		1.05E-6 (4.92E-7)*
AWARENESS		−0.27 (1.25)
LATENESS		−0.75 (0.05)**
BOOST	124.31 (20.37)**	100.65 (18.12)**
FALSE	−1.76 (1.64)	9.13 (1.61)**
Intercept	35.36 (0.68)**	77.15 (3.36)**
Scale factor	1.84	1.45
Likelihood Ratio Chi-Square	181.51**	612.32**
AIC	49864.94	49320.01
BIC	49884.74	49359.61
Deviance	3170.49	2619.56
Obs.	5434	5434

*$p < 0.05$, **$p < 0.01$.

Among the control variables, we observed that the coefficient for *BOOST* was positive and significant ($p < 0.01$). Hence, boosting a cascade's original message is associated with an increase in the cascade's diffusion rate. Another result is that cascades that contain false information circulate at a faster rate than cascades that do not. This is evident from the positive and statistically significant coefficient for the control variable *FALSE* ($p < 0.01$).

5. Discussion of results and conclusions

The planning and execution of humanitarian operations depends on a variety of resources that have very short shelf lives. Our research builds on the fact that information constitutes one of those resources. During times of crisis, it is critical to gather and share information quickly, but accomplishing this goal has been difficult for reasons that include a restricted diffusion of information relevant to humanitarian operations during the course of disasters (Day et al., 2012). While it has been theorized that social media networks built on open Internet platforms can contribute to address these restrictions (Meier, 2015), there is limited work in the humanitarian operations literature that examines whether and how this can be accomplished. Moreover, while extant research in this field has focused on the development of analytical models to manage information (Özdamar and Ertem, 2015), it is only recently that empirical research has begun to study these phenomena, particularly in social media settings (e.g., Korolov et al., 2015).

Our study addresses this deficit in the literature by applying Information Diffusion Theory to the context of humanitarian disasters. Our findings show that, in this context, cascades on social media networks can advance at a rate that significantly exceeds the speed at which information originates from external sources. This finding is important because, during humanitarian crises, speed is key in the diffusion of information among HOs and other stakeholders in order to plan and respond effectively to rapid changes that occur during this type of events. Establishing that social media networks can diffuse information via connections among its users at a rate above that in which external sources of information permeate these networks during a crisis constitutes an important contribution to assessing these networks' effectiveness.

Another contribution from our results is that they show that this speed of diffusion is contingent upon the type of users that originally publish this information. When information is issued by users with high levels of influence, as measured by their number of followers, it will diffuse quickly. However, this will not be the case if the originators' influence is limited. For HOs, this implies that the development of social connections in these networks will be a valuable strategy to pursue in order to ensure fast communication with stakeholders like public donors and beneficiaries during times of crisis. Still, a question that deserves further investigation is whether information diffusion speed will experience different rates of growth as a function of the originator's number of followers once that number reaches certain thresholds. An examination of our data revealed that the rate of growth in diffusion speed as a function of the number of followers seems to increase as that number reaches a threshold of approximately 600,000 followers. Originators with an amount of followers above this threshold appear to have a significant leverage on the diffusion of information. A reason for this is that observations above this threshold sit at the head of a power law distribution across users in our dataset and, thus, can exert a significant pull on diffusion. This is in line with past research that has identified the presence of power law distributions underlying properties of social media networks like Twitter (e.g., Hodas et al., 2013).

The speed of information diffusion on social media networks during a disaster is also contingent upon the time when information is introduced in these networks. Information that is posted earlier during a disaster exhibits a significantly higher speed of diffusion than information that is introduced later during the disaster. This is because, over time, participation in the diffusion of information cascades declines as more cascades compete for attention among users. Such a phenomenon is particularly acute in the context of a hurricane like the one in our study in which the number of new cascades increases sharply over time after hurricane effects materialize in large population areas (see Fig. 1). This phenomenon also underscores a paradox in which, as a disaster progresses, there are increasingly more cascades contributed by originators, but the information in those cascades diffuses more slowly. As a result, a major challenge emerges for HOs trying to introduce urgent information and promoting its diffusion among an increasingly larger volume of new messages posted by other users. How can HOs increase the rate of diffusion of information among all this chatter? Addressing this information directly to followers or requesting explicitly that they retweet the information can augment diffusion (Huberman et al., 2008), particularly if those followers are themselves influential. Including hashtags and links in messages can influence the rate in which users spread information as well (Galuba et al., 2010).

We also observed that cascade originators may be able to increase the speed of diffusion by posting the same information repeatedly in order to raise its visibility. This practice can be justified among HOs in particular situations where information is of urgent nature, particularly during times of excessive chatter like those described above. However, it remains to be seen whether this practice carries with it diminishing marginal returns in increasing the rates of information diffusion. Moreover, we observed that cascades with fabricated information infect the network at a faster pace. Although our data demonstrate that cascades transmitting misleading information transpire rarely (only 4% of the cascades were false), this finding does raise troublesome questions about the ability by HOs and other participants in social media networks to detect and correct this type of cascade. For instance, what attributes do cascades carrying misleading information share that could be used to identify them before they spread too far? What mechanisms can be instated in order to alert the public about these cascades and reverse their diffusion? The design of policies that address these questions and their joint implementation by a wide variety of HOs will help improve the effectiveness of social media in diffusing reliable information to other stakeholders.

It is also important to note that our research found no evidence to suggest that cascades carrying content that enhances situational awareness exhibit significantly higher diffusion rates relative to other cascades. This is surprising given that authors have previously noted that user participation is greater for cascades with information related to situational awareness (e.g., Vieweg et al., 2010). It is possible that the effects of other content-related factors, such as the use of Twitter hashtags or directional operators, on cascades' diffusion rates supersede the effect of situational awareness content. It is also possible that high diffusion rates may be observable but only for those cascades contributing new situational awareness content. That is, content that offers the most up-to-date information of how a disaster event is unfolding.

Another limitation in our research is that it does not assess the geographical implications of information flows in social media networks. We do know from our data that as information diffused on the networks, it reached a substantial amount of local individuals affected by our study's focal disaster. We found that almost 35% of all users in our data were located in geographical areas affected by the disaster. In total, users located in the areas affected by Hurricane Sandy participated in almost all (96.87%) of the cascades. In addition, in 80% of the cascades in our data, 20% or more participants were located in areas affected by the disaster. Thus, a large amount of information in these cascades did manage to reach people located in areas of need.

Prior evidence suggests that local individuals who are geographically vulnerable during a disaster share information differently in social media networks than individuals located in areas unaffected by the disaster (Starbird and Palen, 2010). In particular, local individuals are more likely to contribute information during a humanitarian crisis than other individuals. Those local to a disaster are also more likely to propagate information received from other local individuals during a disaster (Kogan et al., 2015). Given this evidence, we expect that an increase in local users' participation in information cascades will improve the cascades' rate of diffusion in social media networks. Future research in the context of cascades carrying disaster-related information in social media networks could assess empirically whether local users' participation in these cascades will contribute positively to the cascades' rate of diffusion.

Lastly, this research empirically tests theoretical propositions using data from a disaster that was not completely unexpected or unpredictable. However, some disasters that HOs must respond to occur without warning (e.g., earthquakes, terrorist attacks). Future research can analyze whether the theoretical propositions presented in this paper hold in the context of sudden-onset emergencies and whether additional factors specific to this setting impact the diffusion rate of information on social media platforms.

Inspired by the emergence of social media usage during disasters, our study examined the effectiveness of information propagation on social media platforms and identified factors that affected the rate of information diffusion. Beyond this context, commercial firms have also started to leverage social media to catalyze word-of-mouth marketing and enhance brand awareness and engagement (Hoffman and Fodor, 2010). However, key differences exist regarding information cascades on social media in humanitarian versus commercial settings. For instance, in an anticipated event, such as the release of a new product, firms often initiate cascades and engage with consumers to generate buzz. HOs and other stakeholders can also use social media platforms to share preparation information as forecasted disasters draw closer and intensify. However, commercial firms are better able to control and manipulate cascade formation and diffusion in these events since information typically originates from the firm and does not involve as many stakeholders as in humanitarian settings.

Firms also utilize social media as an information tool during unexpected events involving product and service failures. For example, firms in the electronics industry frequently monitor social media to identify information about hardware and software defects reported by consumers while firms in the transportation industry routinely use social media to trace information about unexpected service failure events. Cascades with this information are more likely to originate from dispersed geographical areas unlike cascades with information from victims of unexpected, sudden-onset disaster events (e.g., earthquakes, terrorist attacks), which can largely be traced to more limited geographical areas. While these characteristics help differentiate cascades on social media in commercial and humanitarian contexts, we encourage researchers to continue investigating cascade behavior to increase our understanding of how information disseminates on social media.

Acknowledgements

We are grateful to the special issue editors, the associate editor, and two anonymous reviewers for their insightful comments and guidance. We also would like to thank Hyun-Woo (Anthony) Kim and Tommy Benning from Gnip for their invaluable help with data collection. We would also like to acknowledge Bin Gu, Fred Morstatter, and Shamanth Kumar from Arizona State University for improving our understanding of information diffusion theories and the different properties of Twitter data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jom.2016.05.007.

References

Altay, N., Pal, R., 2014. Information diffusion among agents: implications for humanitarian operations. Prod. Oper. Manag. 23, 1015–1027.
Bhattacharya, D., Ram, S., 2012. Sharing news articles using 140 characters: a diffusion analysis on Twitter. In: Presented at the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, pp. 966–971.
Bonacich, P., 1972. Technique for analyzing overlapping memberships. Sociol. Methodol. 4, 176–185.
Boyd, D., Golder, S., Lotan, G., 2010. Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In: Presented at the 2010 43rd Hawaii International Conference on System Sciences (HICSS). IEEE, pp. 1–10.
Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K., 2010. Measuring user influence in Twitter: the million follower fallacy. In: Presented at the ICWSM, pp. 10–17.
Chandrasekaran, D., Tellis, G.J., 2007. A critical review of marketing research on diffusion of new products. In: Malhotra, N.K. (Ed.), Review of Marketing Research, vol. 3. Emerald Group Publishing Limited, pp. 39–80.
Coxe, S., West, S.G., Aiken, L.S., 2013. Generalized linear models. In: Little, T.D. (Ed.), The Oxford Handbook of Quantitative Methods in Psychology, Statistical Analysis, vol. 2. Oxford University Press, New York, NY, pp. 26–51.
Day, J.M., Melnyk, S.A., Larson, P.D., Davis, E.W., Whybark, D.C., 2012. Humanitarian and disaster relief supply chains: a matter of life and death. J. Supply Chain Manag. 48, 21–36.
Dobson, A.J., Barnett, A.G., 2008. An introduction to generalized linear models. Chapman & Hall/CRC Texts in Statistical Science, third ed. Chapman and Hall/CRC, Boca Raton, FL.
Dodds, P.S., Watts, D.J., 2005. A generalized model of social and biological contagion. J. Theor. Biol. 232, 587–604.
Ellison, N.B., Steinfield, C., Lampe, C., 2007. The benefits of Facebook "Friends:" social capital and college students' use of online social network sites. J. Comput. Mediat. Commun. 12, 1143–1168.
Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378–382.
Fowler, J.H., Christakis, N.A., 2010. Cooperative behavior cascades in human social networks. Proc. Natl. Acad. Sci. 107, 5334–5338.
Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., Kellerer, W., 2010. Outtweeting the twitterers-predicting information cascades in microblogs. In: Presented at the Proceedings of the 3rd Conference on Online Social Networks, USENIX Association, 3–3.
Gao, H., Barbier, G., Goolsby, R., 2011. Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell. Syst. 26, 10–14.
Goldenberg, J., Han, S., Lehmann, D.R., Hong, J.W., 2009. The role of hubs in the adoption process. J. Mark. 73, 1–13.
Goldenberg, J., Libai, B., Muller, E., 2001. Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark. Lett. 12, 211–223.
Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83, 1420–1443.
Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: a survey. In: Presented at the SIGMOD Record, 2503797, pp. 17–28.
Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A., 2013. Faking Sandy: characterizing and identifying fake images on Twitter during hurricane Sandy. In: Presented at the Proceedings of the 22nd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, 2488033, pp. 729–736.
Hardin, J.W., Hilbe, J.M., 2007. Generalized Linear Models and Extensions, second ed. Stata Press, College Station, TX.
Hodas, N.O., Kooti, F., Lerman, K., 2013. Friendship paradox redux: your friends are more interesting than you. In: Presented at the 7th International AAAI Conference on Weblogs and Social Media.
Hoffman, D.L., Fodor, M., 2010. Can you measure the ROI of your social media marketing? MIT Sloan Manag. Rev. 52, 41–49.
Holguín-Veras, J., Jaller, M., Van Wassenhove, L.N., Pérez, N., Wachtendorf, T., 2012. On the unique features of post-disaster humanitarian logistics. J. Oper. Manag. 30, 494–506.
Huberman, B., Romero, D.M., Wu, F., 2008. Social Networks that Matter: Twitter under the Microscope. First Monday 14.
Kaigo, M., 2012. Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake. Keio Commun. Rev. 34, 19–35.
Kempe, D., Kleinberg, J., Tardos, E., 2003. Maximizing the spread of influence through a social network. In: Presented at the Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 956769, pp. 137–146.
Kogan, M., Palen, L., Anderson, K.M., 2015. Think local, retweet global: retweeting by the geographically-vulnerable during Hurricane Sandy. In: Presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, 2675218, pp. 981–993. http://dx.doi.org/10.1145/2675133.2675218.
Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismail, M., Wallace, W., 2015. Actions are louder than words in social media. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ASONAM '15. ACM, New York, NY, USA, pp. 292–297.
Kumar, S., Havey, T., 2013. Before and after disaster strikes: a relief supply chain decision support framework. Int. J. Prod. Econ. 145, 613–629.
Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter, a social network or a news media?. In: Proceedings of the 19th International Conference on World Wide Web, WWW '10. ACM, New York, NY, USA, pp. 591–600.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Lerman, K., Ghosh, R., 2010. Information contagion: an empirical study of the spread of news on Digg and Twitter social networks. In: ICWSM 10, pp. 90–97.
Leskovec, J., Backstrom, L., Kleinberg, J., 2009. Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09. ACM, New York, NY, USA, pp. 497–506.
McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, second ed. Chapman and Hall/CRC, London.
Meier, P., 2015. Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response. CRC Press, Boca Raton, FL.

Meier, P., 2012. How the UN Used Social Media in Response to Typhoon Pablo iRevolution.
Myers, S.A., Zhu, C., Leskovec, J., 2012. Information diffusion and external influence in networks. In: Presented at the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2339540, pp. 33–41.
Oloruntoba, R., Gray, R., 2006. Humanitarian aid: an agile supply chain? Supply Chain Manag. Int. J. 11, 115–120.
Olteanu, A., Castillo, C., Diaz, F., Vieweg, S., 2014. CrisisLex: a lexicon for collecting and filtering microblogged communications in crises. In: Presented at the Eighth International AAAI Conference on Weblogs and Social Media.
Özdamar, L., Ertem, M.A., 2015. Models, solutions and enabling technologies in humanitarian logistics. Eur. J. Oper. Res. 244, 55–65.
Pettit, S., Beresford, A., 2009. Critical success factors in the context of humanitarian aid supply chains. Int. J. Phys. Distrib. Logist. Manag. 39, 450–468.
Rand, W., Herrmann, J., Schein, B., Vodopivec, N., 2015. An agent-based model of urgent diffusion in social media. J. Artif. Soc. Soc. Simul. 18, 1–24.
Rand, W., Rust, R.T., 2011. Agent-based modeling in marketing: guidelines for rigor. Int. J. Res. Mark. 28, 181–193.
Ringel Morris, M., Counts, S., Roseway, A., Hoff, A., Schwarz, J., 2012. Tweeting is believing?: understanding microblog credibility perceptions. In: Presented at the Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, ACM, 2145274, pp. 441–450.
Rogers-Pettite, C., Herrmann, J., 2015. Information diffusion: a study of Twitter during large scale events. In: Cetinkaya, S., Ryan, J.K. (Eds.), Presented at the 2015 Industrial and Systems Engineering Research Conference.
Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., Anderson, K., 2012. “Beacons of Hope” in decentralized coordination: learning from on-the-ground medical twitterers during the 2010 Haiti earthquake. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12. ACM, New York, NY, USA, pp. 47–56.
Sarter, N.B., Woods, D.D., 1991. Situation awareness: a critical but ill-defined phenomenon. Int. J. Aviat. Psychol. 1, 45–57.
Starbird, K., Palen, L., 2010. Pass it on?: retweeting in mass emergency. In: Presented at the Proceedings of the 7th International ISCRAM Conference, pp. 1–10.
Van Wassenhove, L.N., 2006. Humanitarian aid logistics: supply chain management in high gear. J. Oper. Res. Soc. 57, 475–489.
Vieweg, S., Hughes, A.L., Starbird, K., Palen, L., 2010. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10. ACM, New York, NY, USA, pp. 1079–1088.
Watts, D.J., 2002. A simple model of global cascades on random networks. In: Presented at the Proceedings of the National Academy of Sciences of the United States of America, pp. 5766–5771.
Watts, D.J., Dodds, P.S., 2007. Influentials, networks, and public opinion formation. J. Consum. Res. 34, 441–458.