Dynamics of Information Distribution on Social Media Platforms during Disasters

by

Eunae Yoo

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Elliot Rabinovich, Co-Chair
Bin Gu, Co-Chair
William Rand
John Fowler

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

When preparing for and responding to disasters, humanitarian organizations must run effective and efficient supply chains to deliver the resources needed by the affected population. The management of humanitarian supply chains include coordinating the flows of goods, finances, and information. This dissertation examines how humanitarian organizations can improve the distribution of information, which is critical for the planning and coordination of the other two flows. Specifically, I study the diffusion of information on social media platforms since such platforms have emerged as useful communication tools for humanitarian organizations during times of crisis.

In the first chapter, I identify several factors that affect how quickly information spreads on social media platforms. I utilized Twitter data from Hurricane Sandy, and the results indicate that the timing of information release and the influence of the content's author determine information diffusion speed. The second chapter of this dissertation builds directly on the first study by also evaluating the rate at which social media content diffuses. A piece of content does not diffuse in isolation but, rather, coexists with other content on the same social media platform. After analyzing Twitter data from four distinct crises, the results indicate that other content's diffusion often dampens a specific post's diffusion speed. This is important for humanitarian organizations to recognize and carries implications for how they can coordinate with other organizations to avoid inhibiting the propagation of each other's social media content. Finally, a user's followers on social media platforms represent the user's direct audience. The larger the user's follower base, the more easily the same user can extensively broadcast information. Therefore, I study what drives the growth of humanitarian organizations' follower bases during times of normalcy and emergency using Twitter data from one week before and one week after the 2016 Ecuador earthquake.

For my family and in loving memory of Dr. Janet Bell

ACKNOWLEDGMENTS

First and foremost, I would like to thank my main advisor and co-chair of my dissertation committee, Dr. Elliot Rabinovich. I am so grateful for all of the hours that he has dedicated to me over the past nine years to meet me, counsel me, and mold me into the researcher I am today. He taught me to always challenge myself and to relentlessly pursue what I am passionate about. These lessons I will carry with me for the rest of my life, not only in academia but also in my personal life.

I would also like to acknowledge and express my gratitude towards the rest of my committee members. Although Dr. Bin Gu is from another department, he embraced me and agreed to co-chair my dissertation. I am very thankful to him for making himself available for me to provide his perspectives on the dissertation and to teach me new methods. In addition, thank you to Dr. William Rand for being a part of my dissertation committee. Dr. Rand was instrumental in helping me publish the first part of my dissertation and has always been willing to discuss my research with me and give advice. Last but not least, I would like to thank Dr. John Fowler, who has been a part of my educational journey since serving on my undergraduate honors thesis committee. I am indebted to him also for paving the way for me to join the doctoral program.

I must also acknowledge the Department of Supply Chain Management for accepting and supporting me. I am sincerely grateful to Dr. Michele Pfund for helping me make my dream of becoming a doctoral student a reality. Thank you to Dr. Choi, Dr. Kull, Dr. Webster, and Dr. Yin for teaching the doctoral seminars and providing me with foundational knowledge of the field. I would also like to thank Eddie Davila for showing me how to be a teacher and for always having my back. Thank you to Dr. Mahyar Eftekhar for always being generous with his time and especially for his contributions to the first part of the dissertation. I also appreciate Dr. G, the chair of our department, for greeting me with a smile and for his care. To my cohort – Sining Song and Sina Golara – thank you for going through these past five years with me and allowing me to lean on you for support and friendship. I would also like to thank my other cohort – Yousef Abdulsalam, Sangho Chae, Zhongzhi Liu, and Zac Rogers, for adopting me and letting me tag along with them. I am so blessed to have Rachel Balven in my life as my sister from another department and fellow nerd. Thank you for always being available for me for tea time and lunch dates.

From the bottom of my heart, I thank my family for being with me through all the ups and downs during the time I was a doctoral student. I am so grateful for my husband for vicariously living through the doctoral program with me and for teaching me to program. I would also like to thank my mom and dad for immigrating to America to expand my opportunities as well as for all the food, ice cream, laundry, and hugs they have given me. To my sister, I am thankful for distracting me and making me laugh. I would like to also give thanks to my TFC family for praying for me and for being my extended family. Thank you to Hari Mohanraj for the countless chats about food, random subreddits, and data science. Jen Chen, thanks for all of the work + affogato sessions and for graduating with me again. To Debbie Jang, thank you for checking on me and giving me things to think about and learn from outside of my research.

Thank you to the Departments of Supply Chain Management and Information Systems at the W. P. Carey School of Business for providing me with research funding. I would also like to thank the Center for Services Leadership for awarding me a dissertation grant and Arizona State University's Office of Knowledge Enterprise Development, the GPSA, and Graduate Education for awarding me the Graduate Research and Support Program award. Lastly, I would like to acknowledge the AWS Cloud Credits for Research program and Lorena Costanzo for supporting my work.

Above all, I thank God for providing for me and for giving me strength.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

APPENDIX

LIST OF TABLES

LIST OF FIGURES

PREFACE

Humanitarian operations management is concerned with the coordination and delivery of resources that can alleviate the suffering of those affected by a disaster. Like the commercial sector, humanitarian operations must run efficient supply chains to be successful (Van Wassenhove 2006), but key differences distinguish humanitarian from commercial operations. First, the mission of humanitarian operations is not necessarily to minimize operational costs but, rather, to minimize human suffering (Holguín-Veras et al. 2013). Humanitarian organizations (HOs) face extreme levels of variability from the demand side since disasters cannot always be predicted as well as from the supply side since HOs are dependent on the availability of a highly uncertain resource amounts under varying lead times. Moreover, the operating environment is turbulent due to destabilized infrastructure and the convergence of many stakeholders (e.g., local government, military, and other HOs) with goals that may not be aligned toward a common objective (Kovács and Spens 2007, Van Wassenhove and Pedraza Martinez 2012).

Despite these challenges, humanitarian operations must fulfill their objective of distributing all required resources and services to beneficiaries. Examples of commonly demanded resources and services include food, water, and medical services. Another vital resource is information, especially since information facilitates the sourcing and delivery of other resources and services to beneficiaries and other stakeholders. In fact, the effective management of information is one of the most critical factors in determining the success of humanitarian operations (Long and Wood 1995). With accurate information about beneficiaries' needs, for instance, HOs can allocate resources such that the right products can reach the right population at the right time. HOs also issue donor appeals and exchange information with collaborating HOs to enhance coordination and avoid redundant efforts.

However, the management of information has been reported as a major struggle for humanitarian operations. As noted previously, the operational environment during a disaster is volatile due to factors like a damaged physical landscape, population migration, and disrupted economic and political states (Holguín-Veras et al. 2012). This means that decision parameters related to the operational environment are changing constantly, and what may have been relevant or accurate information yesterday is no longer so today. For example, the number of beneficiaries that a HO expects to serve may change suddenly due to notices of mandatory evacuations. Because information is highly perishable in the humanitarian context (Meier 2015), HOs require a robust information network that can quickly diffuse information among the appropriate stakeholders.

Social media platforms have emerged as a useful tool to address this need, and many HOs maintain an active presence on these platforms. HOs have found social media platforms to be valuable because information is shared in real-time and propagates rapidly through platforms' sharing functions. Using these platforms, HOs broadcast information about their available services and share updates about their projects. Furthermore, HOs employ social media platforms to collect information from beneficiaries that post first-hand knowledge of conditions at disaster sites (Gao et al. 2011). The purpose of this dissertation is to develop insights into how social media platforms can disseminate information during times of crisis by answering the following three research questions:

1. What user-related and content-related factors affect the diffusion speed of information on social media platforms in a disaster?
2. How is the dissemination rate of social media content affected by the concurrent diffusion of other content?
3. What mechanisms drive the growth of HOs' social media networks in periods with and without a disaster?

CHAPTER 1

Evaluating Information Diffusion Speed and its Determinants in Social Media networks during Humanitarian Crises[1]

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] This paper was previously published. The citation is as follows: Yoo, E., Rand, W., Eftekhar, M. and Rabinovich, E., 2016. Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management*, *45*, pp.123-133.

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 (e.g., 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.

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).