The Impact of Subscription Reciprocity on Charitable Content Creation and Sharing: Evidence from Twitter on Giving Tuesday

Xue (Jane) Tan
Department of Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, Indiana 47405
janetan@iu.edu

Yingda Lu
College of Business Administration, University of Illinois, Chicago, Illinois 60607
yingdalu@uic.edu

Yong Tan
Michael G. Foster School of Business, University of Washington, Seattle, Washington 98195
ytan@uw.edu

ABSTRACT

Social broadcasting sites have grown from an information diffusion channel to a public medium that fuels social movements such as charitable fundraising. Two mechanisms facilitate user participation in charitable social movements: content creation and content sharing. Users can create original content to express their attitude of giving and promote their most valued nonprofit organizations, enriching the depth of the conversation. They can also share others' content to expedite the diffusion of high-quality content, expanding the breadth of the discussion. This paper investigates the impact of reciprocal and non-reciprocal followees (i.e., a followee is an account other users subscribe to) on followers' decisions to create and share content. Analyzing the charitable movement of Giving Tuesday on Twitter, we find that original charitable content creation is prompted by reciprocal followees' participation but not non-reciprocal followees' participation in this movement. We also find that charitable content sharing is evoked by both reciprocal and non-reciprocal followees, with non-reciprocal having a greater impact. We discuss the theoretical and practical implications of these findings.

Keywords: Twitter, Reciprocity, Charitable Movements, Content Creation, Content Sharing

INTRODUCTION

Through a subscription relationship, Twitter captures people's dual needs of listening to others and expressing themselves. Users can subscribe to others' broadcasts without authorization and can broadcast to their own subscribers, which are often termed "followers." A network derived from subscription relationships, like Twitter, is different from other social networks derived from friendship, kinship, and professional relationships. You may not be an actual friend of Ellen DeGeneres but you can follow her on Twitter to see what she is up to. You may have never met Matthew, a movie fan, but you can read his commentary about new movies by following his Twitter account. Unlike Ellen who does not follow you back, Matthew is more likely to reciprocate your gesture and become your follower, given your mutual interest in movies. As such, Twitter manages a hybrid of reciprocal and non-reciprocal relationships, resulting in a short diameter, a higher frequency of updates and subsequently, a faster mode of information diffusion (i.e., diffusion refers to the spread of information among nodes in a social network) (Java et al. 2007). Rogers (2010) noted that "interactive communication technologies may be changing the diffusion process in certain fundamental ways." When information spreads beyond spatial distance, sparks of ideas may aggregate into collective actions (Oh et al. 2013). To reflect the network structure better, we follow Shi et al. (2014) to term subscription-based networks "social broadcasting networks."

Social broadcasting networks rely on two mechanisms to facilitate decentralized information diffusion. The first mechanism is sharing content, which allows content to be diffused beyond the original author's audience. On Twitter, content sharing is referred to as "retweeting" because a social post is considered a tweet. This action facilitates the cascade of information and brings new people into existing threads, sometimes making news spread more rapidly on Twitter than in mainstream news media. The second mechanism is creating original content that includes a hashtag, which directs views based on a search term or prefixed keyword. These hashtags become clickable links so users' opinions can be diffused to people interested in a certain topic. For example, #OccupyWallStreet is a topical mark of a protest movement against economic inequality (Gleason 2013). The first mechanism of content sharing focuses on the spread of information to broaden its breadth, and the second mechanism of content creation promotes self-expression of opinions concerning a topic to heighten its depth. With both depth and breadth, a popular hashtag can develop into a trending topic listed on the sidebar of Twitter, making it reach even more people.

Many trending topics concern societal issues such as the presidential election, environmental protection, and charitable fundraising (Gaffney 2010). This paper studies a charitable movement and its associated hashtag on Twitter. Charitable digital movements concern public consequences (Wejnert 2002) and play unique roles in encouraging citizen behavior such as giving and volunteering (Castillo et al. 2014). Such movements have become an increasingly important channel for individuals to express their social consciousness and for nonprofit organizations to generate awareness and recruit donors, supporting democratic vitality and encouraging charitable giving as a social norm (Miranda et al. 2016). These digital social movements require theoretically grounded investigation, due in part from being fundamentally different from business-oriented campaigns which are based on monetary incentives (e.g., Ellen DeGeneres endorses skincare products from Olay as a spokesperson).

We examine the effect of *social influence* on driving content creation and sharing with regard to charitable social movements on social media platforms. Users' participation in these movements are affected by the participation of their followees, and social influence is the key driver for individual behavior to aggregate into collective actions. The distinctive feature of the social broadcasting network is reciprocity - your reciprocal tie (Matthew) and non-reciprocal tie (Ellen) may impact your participation differently. Comparatively, the reciprocal relationship with Matthew is a strong tie that can either develop from online or offline interactions. The non-reciprocal relationship with Ellen is a weak tie that widely exists between users and content providers who are not necessarily celebrities. To understand charitable movements in social broadcasting networks and across the relationship maintained there, we sought to answer the research question: *How does the reciprocity of ties attenuate social influence regarding people's content creation (as reflected in a tweet) and content sharing (as reflected in a retweet)?*

The answer to this question concerns an underlying factor shaping the formation of social broadcasting networks - *social capital*. Social capital refers to resources embedded in a social structure for people's furtherance of their self-interests (Coleman 1988). When you advocate the same charitable movement as your reciprocal followee Matthew, he likely observes your behavior, which facilitates trust and mutual obligation. This translates to gaining more social capital, which can be used to attain mental support, social approval, or even information controls in the future (Putnam 1995). The story is different when the relationship is non-reciprocal. Due to the lack of observability, you cannot accumulate social capital from Ellen even if you advocate the same movement as her. There is no obligation for her to support you in the future because obtaining social capital hinges on "general reciprocity" (Cropanzano and Mitchell 2005). However, retweeting a celebrity's post may still be an attractive option for you because her content is likely of interest to a significant number of people and can help your reputation among your own followers (Granovetter 1973; Shi et al. 2014). A reputation for providing high-quality information is a strong motivator for participation in electronic networks (Bandura 2009;

Smith 1999). Despite the many network-level Twitter analyses uncovering the topological features of the networks and temporal trends of diffusion (Bakshy et al. ; Cunha et al. ; Romero et al. 2011), very little is known about how social capital can affect an individual's decision to create and share content. Studying the impacts of tie strength and social capital sets our study apart from the large existing body of work.

In this paper, we conduct a node-level analysis of a 2017 event of an annual charitable movement on social media. We analyze 2,033 random individuals' content creation and sharing decisions in this particular movement as related to the participation of their followee network. We delineate reciprocal and non-reciprocal followees' participation to disentangle the role of reciprocity. Further, since users and their followees may have similar preferences, the users' successive participation in regards to their followees may not necessarily indicate causation. We control for the potential endogeneity issue arisen from latent homophily using a latent instrumental variable (LIV) approach.

By comparing the diffusion processes of content creation and sharing, we find strong evidence for the theory of contagion complexity, which posits that strong ties have an advantage over weak ties in the diffusion of complex behavior (Centola and Macy 2007). In our context, users' content creation is considered more complex than content sharing. We find that content creation is affected by reciprocal followees but not by non-reciprocal followees, underscoring the strength of strong ties in a *complex* diffusion process. Weak ties, on the other hand, show a relatively higher propensity to drive content sharing, a *simple* diffusion process motivated by the novelty of information. This finding is consistent with the strength of weak tie theory, which posits that weak ties are advantageous in novel information transmission. By highlighting the *complexity* of behavior to be diffused, our study shows conditions for both the strong tie and the weak tie theories to take effect, reconciling the debate over the relative strength of strong and weak ties. Given the limited studies that examine how tie properties impact the capacity of social capital to facilitate different actions (Sandefur and Laumann 1998), our findings contribute to the literature on information systems, sociology, and consumer behavior in prosocial activities.

THEORETICAL DEVELOPMENT

Tie Strength

In his foundational account, Granovetter (1973) defined the strength of a tie (relationship) to be, among other features, a function of mutual confiding and reciprocal services. Three types of ties are emphasized in sociology literature: mutual-positive, mutual-nonpositive, and asymmetric. In sociometric tests in which subjects are asked to list people whom they like or trust, these three relationships correspond to mutual choices, mutual nonchoices, and unreciprocated (Davis 1970). Using 742 matrices of real-world data, Davis (1970) found that mutual positive corresponds to the strongest interpersonal relationship; mutual nonpositive holds the weakest relationship, and asymmetric pairs are intermediate. A decade after he introduced the concept of weak ties, Granovetter (1983) revisited a body of work on weak ties that had developed since. In Granovetter's 1983 reassessment of the theory, Friedkin (1980) made the most comprehensive attempt to examine the validity of his arguments regarding weak ties. Friedkin (1980) defined a weak tie between two faculty members as one reporting having discussed his work with another, while the other did not report the same. Friedkin concluded that treating an asymmetrical relationship as a weak tie and a reciprocal relationship as a strong tie is consistent with Granovetter's perspective. Shi et al. (2014) considered tie strength to be contingent on reciprocity when they study content sharing in the form of retweeting on Twitter. They argued that subscription networks facilitate electronic interactions, which occur more frequently between users with reciprocal ties. They also assessed overlapping neighbors between node pairs and found that reciprocal ties correspond to a higher overlap than non-reciprocal ties, a key distinction emphasized by Granovetter (1973) in terms of strong ties versus weak ties. Based on the above discussions, we consider a reciprocal tie to be stronger than a non-reciprocal tie. This allows us to draw on the rich literature of tie strength in diffusion.

Diffusion and Contagion

A diffusion process is defined as the dynamic by which "contagions" (i.e., ideas, actions, products, and tastes) spread through a network (Baumgarten 1975). Contagions have different levels of risk, cost, and controversy, leading to varying thresholds for adoption (Centola and Macy 2007). Research shows the diffusion of behaviors generally has a higher adoption threshold than the diffusion of information.[1] Varying scenarios—when a farmer adopts newly invented hybrid seed corn (Ryan and Gross 1943), when a doctor uses a new prescription drug (Coleman et al. 1966), and when individuals participate in social movements (Marwell and Oliver 1993)—show affirmation needs to be received from multiple sources. What determines the complexity of contagion is the number of sources needed to trigger adoption.

Our study looks separately at two diffusion actions, content creation and content sharing, to disseminate charitable information proactively and voluntarily. Content creation has a higher complexity level for a number of reasons. First, content creation is riskier because content creators are held accountable for the content (e.g., the credibility of the endorsed charity or

[1] Diffusion of information has the lowest threshold because an individual only needs to hear from one source to become aware of the information. If you are told the release date of a movie from Matthew, you do not need to acquire this information from others. Multiple sources carrying the same information result in information redundancy, sometimes inhibiting diffusion (Granovetter 1973). We study diffusion behaviors but not the diffusion of ideas (e.g., the political philosophy of Marxism or a news event) because the diffusion of information is hard to observe.

movement). Second, creating content exerts a higher cost because time and effort are needed to compose content. Third, charitable content creation is likely more controversial because others may suspect the motive behind such behavior to be reputation-driven instead of altruistic. As a result, one may need more adopted neighbors (i.e., participating followees) for content creation than content sharing. The abovementioned aspects are defining features of contagion complexity (Centola and Macy 2007) that guide us to consider content creation to be a complex contagion and content sharing to be a relatively simple contagion.

Past research on Twitter has vaguely defined diffusion, with some studies equating diffusion with sharing behaviors (Bhattacharya and Ram 2012; Taxidou and Fischer 2014) and others with original content creation (Romero et al. 2011). As Rogers (2010) notes, differentiating these two behaviors is important because they lead to different diffusion trajectories. Moreover, understanding individuals' diffusion actions (e.g., content creation and content sharing) is key to uncover the process of seeking and processing information to reduce uncertainty about the advantages and disadvantages of an action to be diffused (Rogers 2010).

Hypothesis

The concept of contagion complexity was brought up by Centola and Macy (2007) to resolve the continuing debate between two competing theories that assess the comparative strength of weak and strong ties for diffusion processes. Returning to Granovetter (1973), his strength of weak tie theory suggests that a weak tie is likely a bridge joining two sparsely connected network segments. Thus, weak ties likely possess novel information and promote diffusion that depends critically on the value of such information. Rodan and Galunic (2004) found that accessing heterogeneous knowledge from weak ties has a positive influence on managerial and innovation performance. More relevant to our study, Shi et al. (2014) developed a consumption-share model to reveal that users are more likely to share content posted by weak ties in pursuit of reputation.

Whereas weak ties can facilitate the acquisition of novel information, strong ties are considered more effective for conveying trust, applying peer pressure, and fostering cooperation (Coleman 1988). This is especially true when strong ties are defined as reciprocal ties since social capital is more likely to be accumulated when one's acts are observable. Regarding strong ties, Bond et al. (2012) found that people's voting behaviors are influenced to a greater degree by Facebook friends who are also friends offline. Aral and Van Alstyne (2011) used social network and email content from an executive recruiting firm to show that strong ties are more effective than weak ties for information diffusion in a turbulent and high-dimensional information environment. Centola (2010) conducted an experiment to show that the diffusion of health behaviors with a high adoption threshold better benefits from strong ties than weak ones. Related to our study, strong ties on Flickr were found to encourage content creation (Zeng and Wei 2013).

To reconcile the competing theories and findings, Centola and Macy (2007) used theoretical network analysis to show that weak ties diminish in strength when the diffusion is complex and requires social reinforcement from multiple sources. Hansen's (1999) network study of an electric company's new product development found that weak ties helped with information searches across groups while simultaneously hampering the transfer of complex knowledge. A recent survey found that Facebook promotes costly and time-consuming political actions like participation in street demonstrations while Twitter encourages the injection of political news and mobilization of information (Valenzuela et al. 2018).

We follow the theory of contagion complexity to propose the following hypothesis:

Hypothesis: *Strong ties have a greater impact on content creation than on content sharing, while weak ties have a greater impact on content sharing than on content creation, all else being equal.*

The essence of this hypothesis is the diminishing effect of weak ties and the increasing importance of strong ties when contagions become more complex. The risk associated with a complex contagion is more likely mitigated by affirmation from strong ties. Empirical validation of this hypothesis is very limited because it requires comparison across contagions with different complexity, a protocol we follow in our empirical investigation.

RESEARCH CONTEXT

In this section, we briefly describe the social broadcasting network of Twitter to contextualize our study. We introduce the charitable movement of Giving Tuesday as well as discuss our reasons for focusing on this movement.

Tweet and Retweet

Twitter is a microblogging site that allows users to create and share content. When users log on to Twitter, their landing page is a feed of tweets posted by those they follow in reverse chronological order. Users can "like" others' tweets as recognition of the tweet and reply to others' tweets to express their opinions. Users may follow other users (without authorization) to receive content updates and repost their tweets for their own followers' consumption, with the original author being highlighted as the owner of the content[2]. During the course of our study,

[2] There are two ways of retweeting. When users click the *retweet* button, a dialogue box opens to allow them to comment on the original post. When a comment is added, the post is published as the user's own content for which the user receives comment, likes, and further reposting. Without added comments, the retweet is published with the original content generator as author. The reposting user, in this case, does not receive further notifications concerning this post. In our study, we follow past literature in considering a repost with an additional comment to be original content (Petrovic et al. 2011).

users may also create original content for up to 140 characters long. The character count limit was expanded to 280 in November 2017. In addition, it is a convention to amplify their content with hashtags to mark the tweets topically. Specifically, a keyword is prefixed by a # symbol and included in the tweet.

Giving Tuesday

The hashtag #GivingTuesday was launched in 2012 by 92nd Street Y[3] and the United Nations Foundation. It is a movement where users are encouraged to advocate for their valued nonprofit organizations and give back to the community on the Tuesday after the U.S. holiday of Thanksgiving, Black Friday and Cyber Monday. The movement harnesses social media technology and circulates mainly on Twitter, Facebook, and Instagram.[4] Our study examines Giving Tuesday 2017, which took place on November 28, 2017. In Figure 1, we show the volume of tweets and donations reported by Blackbaud, the biggest payment processor for Giving Tuesday. We can see a higher volume of tweets containing the hashtag #GivingTuesday correlated with more total donations. Although actual donations are not itemized, it is generally believed that a higher volume of tweets related to #GivingTuesday leads to better fundraising performance.

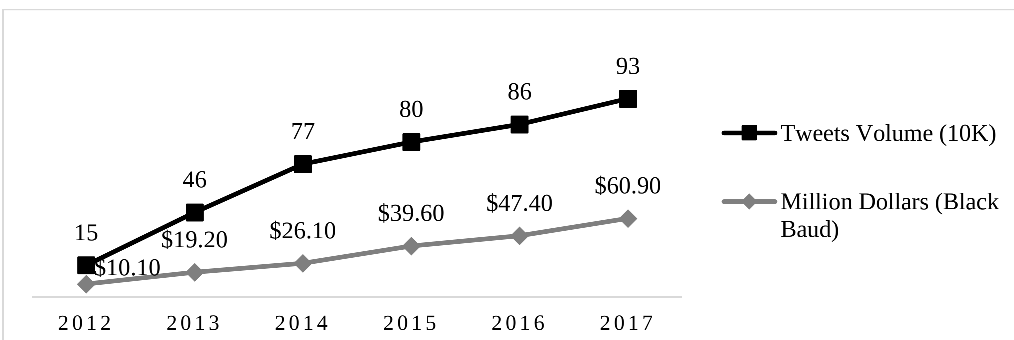
Our study prioritizes two behaviors that fuel this charitable movement: content creation and content sharing of other's tweets concerning Giving Tuesday. The behaviors of content creation and sharing show significantly different levels of engagement - content creation can involve a first-person and content sharing is based on a third-person account. Original tweets are

[3] 92nd Street Y is a cultural center in New York City that brings people together toward the value of giving back.

[4] The hashtag #GivingTuesday is a charitable campaign in the Twitterverse, and Giving Tuesday is the real-world campaign. In our paper, we use these two terms interchangeably, and they both refer to the online campaign.

usually emotionally charged with users' own experiences about the nonprofits that they endorse. For example, one original tweet in our study is "I supported Fuller Center for Housing of Greater Kansas City on #GivingTuesday because they gave Calvin a fighting chance at the good life through their Greater Blessing program, which made over $20k of needed repairs." Shared content, on the other hand, indicates an affirmation of the charity and is aimed to broaden its exposure. For example, one user of our study retweeted a message originating from the reproductive rights organization NARAL: "RT @NARAL We're in the fight of our lives. This #GivingTuesday, we need you to help us protect reproductive rights." In terms of diffusion, retweeting only expands the reach of an existing post but an original tweet reflects users' self-presentation as an activist.

Figure 1. Social Mention and Donated Amount on Giving Tuesday[5]



Giving Tuesday provides us with a unique opportunity to answer our research question for several reasons. First, since Giving Tuesday is an annual event that takes place on a specific date—always the Tuesday following Thanksgiving—our pre-selected random users are not likely to participate in the movement beforehand or afterward. This reduces the risk of truncation. Second, given the nature of charitable fundraising, posts relating to this topic have a homogenous

[5] The social mentions reported here are different from some statistics reported online because we do not consider social posts generated prior to or after the launch of Giving Tuesday. Those tweets are usually not generated by individual users.

sentiment, eliminating the potential confoundedness deriving from diverse content (Shore et al. 2016). Third, the reputation motivation inherent in prosocial behavior augments the reputational gain from content creation, making it easier for us to observe how content sharing and creation are impacted differently by reciprocity. Lastly, the pervasive engagement in Giving Tuesday provides us with ample variation in the degree to which individual users are exposed to it, allowing us to identify the effect of social influence within a single theme.

DATA

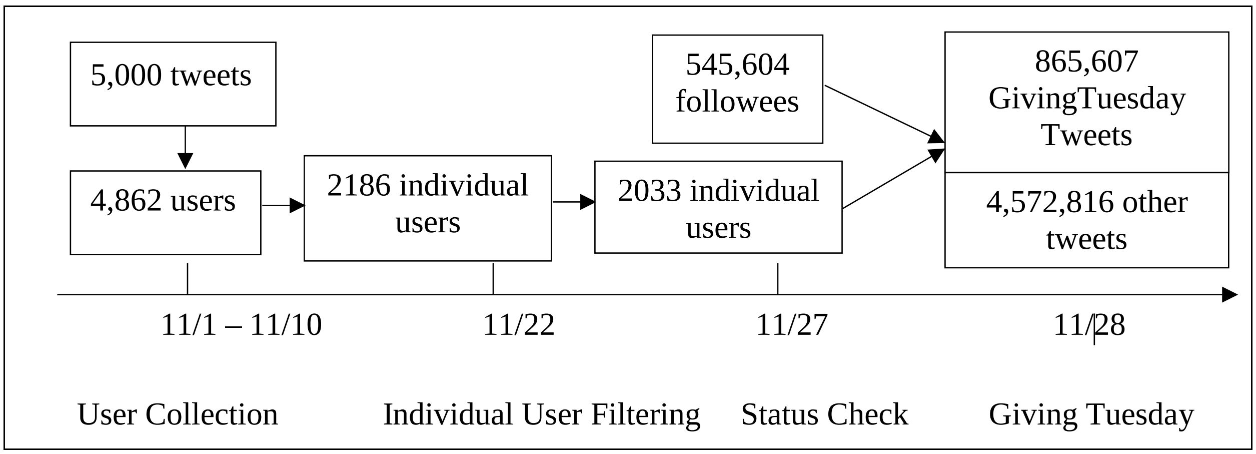
Data Collection

At the user level, we select random individual users to evaluate their diffusion behaviors. We randomly sampled 5,000 Twitter statuses published in the first ten days of November 2017 using Twitter API statues/sample, with five hundred random tweets collected daily. We chose this data collection period to avoid inactive or silent users as it was close to Giving Tuesday. These tweets were created by 4,862 unique vocal Twitter users located in different local network structures of the Twittersphere. We excluded protected users—those whose tweets are "locked" and thus only available to approved followers but not for general consumption—and evaluated each remaining user to keep only individual parties. We do not include Twitter accounts administered by organizations as they likely have scheduled content to post and are less prone to social influence. Each user was kept in our sample if consensus among three Amazon Mechanical Turk workers was reached that the user was unambiguously individual and not an organization, news media, or other types of account. All Turk workers were given instructions to examine the username, profile picture, bio, and associated tweets when drawing conclusions[6].

[6] Our study was approved with IRB exempt status as we have benigne interaction with human subjects - the Amazon Mechanical Turkers (#51968 at University of Washington and #1510 at Rensselaer Polytechnic Institute).

After this process, we had 2,186 individual users for the analysis. Prior to Giving Tuesday, we collected these users' basic information, including their number of followers, followees, and statuses as well as, the date they joined Twitter. We also collected the list of their followees to construct their followee network. These users have 545,604 total followees. Right before Giving Tuesday, we conducted a user status check and found 153 users were either deleted or suspended, leaving us with 2033 focal users (See Figure 2).

Figure 2. Data Collection Process



At the content level, we obtained all tweets including the keyword "GivingTuesday" or "Giving Tuesday" on November 28, 2017.[7] There were 865,607 tweets that fit these criteria, consisting of 319,145 original tweets, 530,685 retweets, and 15,777 replies, and our study focuses only on original tweets and retweets. We matched this tweet set with our focal users and their followee networks to identify user participation. It turns out that 39,665 tweets associated with Giving Tuesday were generated by 4,632 reciprocal followees and 12,414 non-reciprocal followees, which means our relatively small group of focal users received 4.7% of all Giving

[7] This data was purchased through an official partner of Twitter. Out of all the tweets, 800,326 contain "GivingTuesday", 64,963 contain "giving Tuesday", and 797,775 contain "#GivingTuesday".

Tuesday tweets. This demonstrates that the Twittersphere has a small effective diameter and is advantageous for information dissemination. Finally, we collected an additional 4,572,816 tweets not associated with Giving Tuesday, generated by the followee network on November 28, 2017, to control for focal users' newsfeed intensity.

Variables

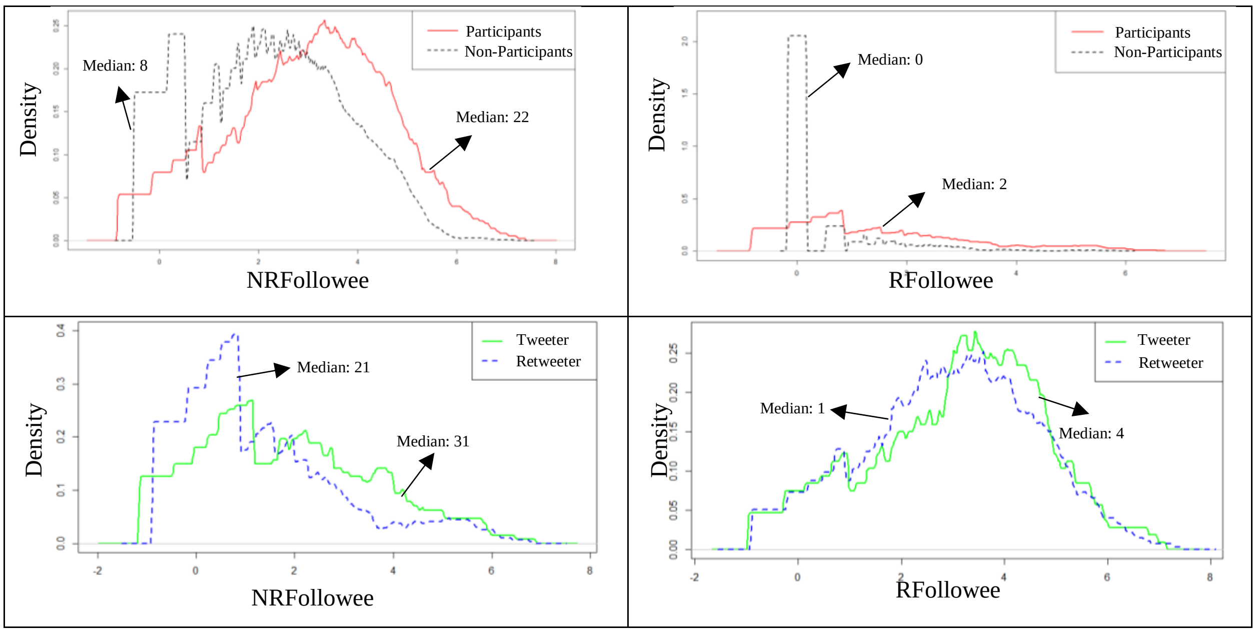
We define two binary dependent variables to represent users' participation in Giving Tuesday. $Tweet_i$ takes the value of 1 if user *i* creates at least one original post containing the keyword "Giving Tuesday" or "GivingTuesday." $Retweet_i$ takes the value of 1 if user *i* retweeted at least one tweet about Giving Tuesday. Otherwise, both variables take the value of 0. Among the 2,033 focal users, 56 generated original content[8] and 157 retweeted others' content.

Our independent variables are the number of reciprocal followees ($RFollowees_i$) and non-reciprocal followees ($NRFollowees_i$) who participated in Giving Tuesday 2017. If a focal user participated in Giving Tuesday, only her followees who participated before her first tweet or retweet will be counted in these two measures. To provide some model-free intuition, we divide the 2,033 focal users into participants (207 users) and nonparticipants (1,826 users) of this movement. Non-participants are represented in black dashed lines in the top panel of Figure 3. This segment of users has a median of 8 non-reciprocal followees and 0 reciprocal followees participated in the movement. It is notable that a considerable number of non-participants do not have any reciprocal followees involved in this movement; this pattern is not seen for participants. Participants that either created original tweets or retweeted other's content are represented by red solid lines in the top panel of Figure 3. These participants have a median of 22 non-reciprocal

[8] One original tweet is a retweet with added comment "Support." No content generated by our focal users expressed negative opinion about Giving Tuesday.

followees and 2 reciprocal followees who participated in this movement. From the distribution at the top panel of Figure 3, the red solid curves that represent participants significantly shift to the right of the black dashed curves that represent non-participants, suggesting that followees' participation plays an important role in driving focal users' participation.

Figure 3. Density Plot of NRFollowee and RFollowee by Diffusion Choices



Note: Rectangular kernel was employed when constructing the density plots. The variables are plotted with log-transformed values, and the median is reported at the original scale.

We further divide the 207 participants into original content creators (or "tweeters") and content sharing users (or "retweeters") in the bottom panel of Figure 3. Tweeters are represented with green solid lines and retweeters are represented with blue dashed lines. Retweeters have a median value of 21 for $NRFollowees_i$ and 1 for $RFollowees_i$, and tweeters have a median of 31 for $NRFollowees_i$ and 4 for $RFollowees_i$. From the comparison of distributions, we see that non-reciprocal followees better differentiate tweeters from retweeters than reciprocal followees do.

We plot the histogram of log-transformed number of non-reciprocal followees ($NRFollowees_i$) and reciprocal followees ($RFollowees_i$) in Figure 4. The distributions are significantly different from a normal distribution, assuring the validity of the LIV method which we detail later. Other variables are defined in Table 1 and the statistics are reported in Table 2.

Figure 4. Histogram of Participating Followees by Reciprocity

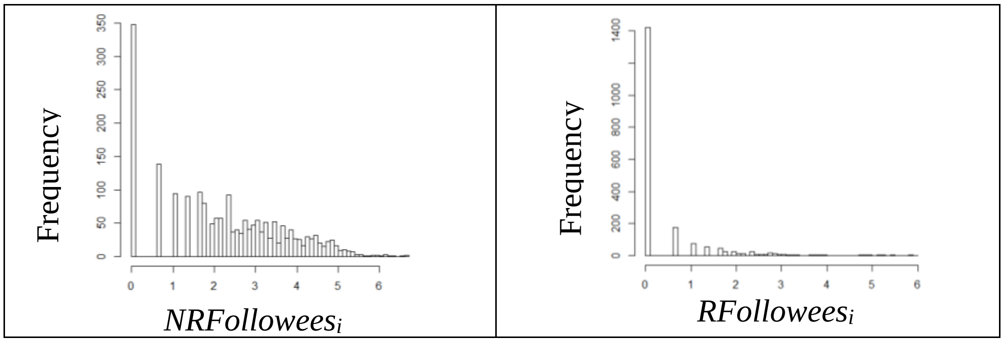


Table 1. Variable Definitions

Variables	Description
Dependent Variables	
Tweet$_i$	Indicator variable for individual *i* who generated original post(s) about Giving Tuesday
Retweet$_i$	Indicator variable for individual *i* who generated retweet post(s) about Giving Tuesday
Independent Variables	
RFollowees$_i$	The number of reciprocal followees who participated in Giving Tuesday (before *i*'s first tweet or retweet of the same topic, if any)
NRFollowees$_i$	The number of non-reciprocal followees who participated in Giving Tuesday (before *i*'s first tweet or retweet of the same topic, if any)
Control Variables	
Statuses$_i$	The number of all Twitter statuses generated by *i* before Giving Tuesday 2017.
Followers$_i$	The number of Twitter users following *i* before Giving Tuesday 2017.
Followees$_i$	The number of Twitter users followed by *i* before Giving Tuesday 2017.
OtherPosts$_i$	The number of posts not associated with Giving Tuesday on Giving Tuesday 2017 (before *i*'s first tweet or retweet of the same topic, if any)
Tenure$_i$	The number of days since *i* joined the Twitter platform.

Note: Log-transformations are conducted for all variables except for *Tenure$_i$* due to the high skewness. *Tenure$_i$* was normalized to 1 for estimation convenience.

Table 2. Summary Statistics (N=2,033)

	Transformed Scales				Original Scales			
	Mean	Sd	Min	Max	Mean	Sd	Min	Max

Tweet$_i$	-	-	-	-	0.03	0.16	0	1
Retweet$_i$	-	-	-	-	0.07	0.26	0	1
RFollowees$_i$	0.55	1.06	0.00	5.99	5.30	27.46	0	398
NRFollowees$_i$	2.24	1.53	0.00	6.67	27.24	55.59	0	789
Followees$_i$	5.87	1.32	0.00	10.58	730.36	1359.23	0	39161
Followers$_i$	5.70	1.60	0.00	10.49	815.39	1749.82	0	36114
Statuses$_i$	8.89	1.97	0.69	13.64	26118.26	49716.31	1	839573
OtherPosts$_i$	6.39	2.05	0.00	9.94	678.16	826.46	0	9421
Tenure$_i$	0.45	0.25	0.00	1.00	1561.24	968.45	0	3705

MODEL

Baseline Model

Our baseline model is a Probit model with a binary outcome $Tweet_i$ (or $Reweet_i$). We present the model with the dependent variable being $Tweet_i$, and estimate both models when the outcome is $Tweet_i$ or $Reweet_i$.

$$Tweet_i = \begin{cases} 1, & \text{if } Tweet_i^* \geq 0; \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

A latent continuous variable $Tweet_i^*$ determines whether individuals make the decision to participate:

$$Tweet_i^* = \gamma_1 RFollowees_i + \gamma_2 NRFollowees_i + \mathbf{\beta x}_i + u_i, \quad (2)$$

where $\mathbf{\beta x}_i = \beta_0 + \beta_1 Tenure_i + \beta_2 OtherPosts_i + \beta_3 Statuses_i + \beta_4 Followers_i + \beta_5 Followees_i$.

In this baseline model, u_i is assumed to be independent and identically distributed. However, the independence assumption of u_i is likely to be violated due to its potential correlation with the independent variable $RFollowees_i$, or even $NRFollowees_i$. If users who actively advocate for charities are more likely to hold a tie with others who share the same concerns, the correlation between users' number of participating followees and their own participation may be more an expression of latent homophily rather than social contagion (McPherson et al. 2001). Since marketers can only control the choice of influencers but not the formation of the network structure, researchers are most interested in the effect of social contagion. To test and account for the endogeneity bias resulting from homophily, a common solution is to use observed instrumental variables (IVs) that meet the "exclusion restriction"—variables need to be correlated with the endogenous variables and uncorrelated with the error term.

Despite the popularity of this method, questions have been raised regarding its untestable nature and weak instrument problem. The IVs are not testable because the exogeneity test (e.g., Hausman test) is based on the availability of valid IVs. This is a circular problem that leads to the overuse of IVs (Hueter 2016). The other common issue is the weak instrument problem that occurs when the IVs do not sufficiently explain the variation of the endogenous variable. Under such cases, using IVs will lead to more bias than not using the IVs at all (Bound et al. 1995). A further complication is related to the context of the social broadcasting network. When estimating social influence, a commonly used IV is the characteristics of focal users' friends' friends. These characteristics will affect users' friends but not the focal users directly. However, this approach is prohibited by the size of the followee network. Given the small diameter of a social broadcasting network, collecting focal users' second-degree followees would require obtaining a huge partial network. Given the above reasons, we adopt an "instrument-free" approach to test and account for the endogeneity issues.

Latent Instrumental Variable Method

A family of frugal IV estimations was developed to account for endogeneity issues without the use of observed instruments (Ebbes et al. 2009; Park and Gupta 2012). We use a parametric approach in the main analysis and a semi-parametric approach in the robustness check.[9] The LIV approach with linear regression was introduced by Ebbes et al. (2005), and the nonlinear adaptation is discussed in Ebbes (2004). LIV assumes the existence of a latent categorical instrument variable that satisfies the exclusion restriction required for an observable instrumental variable and estimates the distribution of the latent IV together with other coefficients of interest. It not only solves the circular problem of testing exogeneity but also alleviates the weak instrument issue because the latent instrument is, by construction, the "best" IV. It has been applied in marketing studies on the effect of visual attention to advertisement (Zhang et al. 2009), the position of text ads in search campaigns (Rutz et al. 2012), and has been extended to time series data to understand the role of online communications in sales (Sonnier et al. 2011). Most relevant to our setting, LIV has been used to account for latent homophily in the estimation of social influence (Ma et al. 2014). In our study, we follow Ebbes (2004) to devise a Probit-LIV estimator to fit the binary outcome.

In the following model specification, we assume the number of participating reciprocal followees ($RFollowees_i$) to be endogenous and the number of participating non-reciprocal

[9] Three parametric approaches were discussed in Ebbes et al. (2009), and one semi-parametric approach was discussed in Park and Gupta (2012). The three parametric approaches include the latent instrumental variable (LIV) approach, the higher moments approach (HM), and the identification through heteroscedasticity (IH) approach. In this paper, we use the LIV approach as our main analysis. We do not use HM because it requires that the second moments of the mean-centered dependent variable and a mean-centered endogenous regressor be uncorrelated. We do not use IH for it requires an observable grouping variable to determine the heteroscedastic error structure. Park and Gupta (2012) proposed the copula method that nonparametrically estimates the marginal distribution of the endogenous regressors and uses a copula function to build the joint distribution of the endogenous regressors and the structural error term. We use this copula method in our robustness check where we assume both $RFollowees_i$ and $NRFollowees_i$ to be endogenous.

followees ($NRFollowees_i$) to be exogenous. Theoretically, the stronger a tie connecting two people, the more similar the two people are (Laumann 1968). Therefore, homophily occurs within strong ties which manifest in a reciprocal following relationship. Also, given the success of Giving Tuesday, many organizations and news media accounts that are one-way connected by individual accounts have participated in the event regardless of whether they possess a charitable nature or not. Further, a separate model in which we assumed $NRFollowees_i$ to be endogenous was estimated and did not report significant dependency between the error term and the endogenous variable. Lastly, we account for the endogeneity of both $RFollowees_i$ and $NRFollowees_i$ in a robustness check, where our results hold.

To account for the dependency between $RFollowees_i$ and u_i , the LIV approach decomposes the endogenous variable $RFollowees_i$ into an exogenous part $\boldsymbol{\alpha}_\mathbf{i}'\boldsymbol{\pi}$ and an endogenous part v_i ,

$$RFollowees_i = \boldsymbol{\alpha}_\mathbf{i}'\boldsymbol{\pi} + v_i . \qquad (3)$$

The exogenous part $\boldsymbol{\alpha}_\mathbf{i}'\boldsymbol{\pi}$ is a function of the latent categorical instrument $\boldsymbol{\alpha}_\mathbf{i}$ and the categorical means $\boldsymbol{\pi}$. The latent instrument $\boldsymbol{\alpha}_\mathbf{i}$ is assumed to follow an M-dimensional multinomial distribution with probability λ_j for each dimension *j*, where $j = 1, \cdots, M$. The other component $\boldsymbol{\pi}$ is a vector of categorical means π_j , where $j = 1, \cdots, M$. When the value of the latent instrument for a specific dimension *j* is 1 ($\alpha_{ij} = 1$), individual *i* belongs to category *j* and the exogenous part takes the value of π_j . For example, suppose we have $M = 2$ and $\pi = (\pi_1 \quad \pi_2)' = (0.5 \quad 1.5)'$. We draw $\boldsymbol{\alpha}_\mathbf{i}$ for individual *i* from the binomial distribution to find that $\boldsymbol{\alpha}_\mathbf{i} = \begin{pmatrix}1 & 0\end{pmatrix}'$. We can then calculate the exogenous part for *i* to be

$\boldsymbol{\alpha}_i'\pi = \begin{pmatrix}1 & 0\end{pmatrix}\cdot(0.5 \quad 1.5)' = 0.5$. If $\lambda_1 = 0.2$ and $\lambda_2 = 0.8$, we know that $\boldsymbol{\alpha}_\mathbf{i} = \begin{pmatrix}1 & 0\end{pmatrix}'$ with a probability of 20% and $\boldsymbol{\alpha}_\mathbf{i} = \begin{pmatrix}0 & 1\end{pmatrix}'$ with a probability of 80%. According to this setting, $\boldsymbol{\alpha}_i'\pi$ is uncorrelated with u_i, and is essentially a latent discrete variable with M categorical levels.

From the above equations, the endogeneity issue arises if the error term u_i of equation (2) is correlated with the error term v_i of equation (3). This leads to the correlation between $RFollowees_i$ and u_i. By modeling the dependency explicitly, we control for this endogeneity problem. We assume that the error terms (u_i, v_i) follow a bivariate normal distribution with mean zero and variance matrix $\begin{pmatrix}\sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2\end{pmatrix}$. In Appendix A, we derive the log-likelihood function as below:

$$LL = \sum_i \ln\left(\sum_{j=1}^{M}\lambda_j \cdot \left(\begin{array}{l}\left(1 - Tweet_i\right) + \left(2\cdot Tweet_i - 1\right) \\ \times\Phi\left(\dfrac{\gamma RFollowee_i + \boldsymbol{\beta}\mathbf{x}_i + \rho\left(\sigma_u / \sigma_v\right)\left(RFollowee_i - \pi_j\right)}{\sqrt{1-\rho^2}\sigma_u}\right)\end{array}\right)\cdot\phi\left(\frac{RFollowee_i - \pi_j}{\sigma_v}\right)\right)$$

$$-\sum_i \ln\sigma_v, \qquad (4)$$

where $\Phi(\cdot)$ is the cumulative density function of a standard normal distribution, and $\phi(\cdot)$ is the probability density function of a standard normal distribution. In our estimation, we cannot identify both σ_u and σ_v. Therefore, we estimate the ratio of them as $\sigma = \sigma_u / \sigma_v$. Given the likelihood function, a maximum likelihood estimator was employed for estimation.

The identifying assumption of the LIV model is that the distribution of the endogenous variable is significantly different from a normal distribution. This stems from the normality assumption of v_i. If the latent variable does not differ significantly from a normal distribution, the algorithm will fail to distinguish it from the error term. To validate the identification of our estimation, we first show in Figure 4 that *RFollowees* and *NRFollowees* are very skewed. We then conduct a comprehensive simulation study in Appendix A.2 where we simulate an endogenous variable that appears similar to the one we have. The simulation shows the good performance of our estimation method in alleviating endogeneity.

RESULTS

Model Selection

For the LIV model, we determined the number of levels (M) for the latent discrete variable following Ebbes et al. (2005). We calculated the integrated classification likelihood (ICL), Akaike information criterion (AIC), and Bayesian information criterion (BIC) when M equals 2, 3, and 4. The ICL is a modified version of BIC which adds a penalty for the entropy of the categorization and is more suitable for selecting the component number in mixture models. The fitness statistics are reported in Table 3, where a three-class model outperforms both when the outcome is tweeting original content and when the outcome is retweeting.

Table 3. Information Criteria for LIV Model Selection

	ICL		AIC		BIC	
	Original Tweet	Retweet	Original Tweet	Retweet	Original Tweet	Retweet
$M = 2$	5684.363	6185.222	5606.812	6107.659	5679.837	6180.683
$M = 3$	4817.755	5318.947	4737.607	5238.787	4806.866	5308.046
$M = 4$	4829.707	5330.862	4760.607	5261.787	4822.101	5323.281

Main Results

The estimated coefficients for the native Probit model and the LIV model are reported in Table 4. We see that content creation is positively affected by reciprocal followees but not non-reciprocal followees. Conversely, retweeting behavior is positively affected by both reciprocal and non-reciprocal followees, with the latter having a greater impact, as shown in the LIV model. The coefficient for the effect of the reciprocal followees on retweeting is reduced from $\gamma_1^{Baseline-Retweet}$ =0.172 (p-value<0.01) in the baseline model to $\gamma_1^{LIV\text{-}Retweet}$ =0.112 (p-value<0.05) in the LIV model, showing that 34.9% of the estimated effect in the baseline model is likely spurious.

Table 4. Estimation Results

	Original Tweet		Retweet	
	Probit	LIV	Probit	LIV
RFollowees_i	0.337***	0.293***	0.172***	0.112**
	(0.0590)	(0.056)	(0.0460)	(0.044)
NRFollowees_i	0.0213	0.075	0.163***	0.190***
	(0.0619)	(0.059)	(0.0470)	(0.044)
Followers_i	-0.0185	-0.006	-0.152**	-0.139**
	(0.0797)	(0.080)	(0.0633)	(0.063)
Statuses_i	-0.124***	-0.121***	0.245***	0.243***
	(0.0463)	(0.046)	(0.0396)	(0.040)
Followees_i	0.0621	0.039	0.155***	0.142**
	(0.0787)	(0.079)	(0.0550)	(0.055)
Tenure_i	0.611**	0.592**	-0.506***	-0.515***
	(0.253)	(0.253)	(0.186)	(0.187)
OtherPosts_i	-0.0750*	-0.104**	-0.0970***	-0.114***
	(0.0422)	(0.043)	(0.0278)	(0.028)
ρ		0.03		0.102*
		(0.0716)		(0.0523)
σ		0.407***		0.408***
		(0.00705)		(0.00719)
Likelihood	-217.339	-2346.304	-469.787	-2596.894

Note: *p<0.1; **p<0.05; ***p<0.01.

To test our hypothesis, we vertically compare the effects of reciprocal and non-reciprocal followees on both processes. Such a comparison is meaningful because we use the same sample in both processes with standard binary outcomes. The reciprocal followees have a larger impact on content creation ($\gamma_1^{LIV-Tweet}$ =0.293, p-value<0.01) than on retweeting ($\gamma_1^{LIV-Reweet}$ =0.112, p-value<0.05). The difference between coefficients is significant according to the Z-test with a p-value less than 0.01 (Clogg et al. 1995; Paternoster et al. 1998). As noted earlier, non-reciprocal followees have a null effect on content creation and a positive effect on content sharing ($\gamma_2^{LIV-Reweet}$=0.19, p-value<0.01). Such a difference in the statistical significance confirms that weak ties play a more salient role in content sharing than content creation. Thus, our hypothesis is fully supported.

Other findings involving the control variables reveal interesting patterns. Neither the number of users' followers ($Followers_i$) nor followees ($Followees_i$) have a significant correlation with users' content creation during this event. However, on retweeting, the number of users' followers ($Followers_i$) has a negative impact while the number of users' followees ($Followees_i$) has a positive impact. This shows that retweeting has a higher dependency on the network in which users are embedded and with richer sources of information, users are more likely to share. However, users' concerns regarding sharing charitable content seem to grow with audience size. The number of followees' other tweets ($OtherPosts_i$) is generally negatively related to users' participation in the charitable movement, possibly because other posts dilute the message of Giving Tuesday. In addition, with more past statuses ($Statuses_i$), users are more likely to share content but less likely to create original content. Finally, from the estimated coefficient of $Tenure_i$, early Twitter adopters are more likely to create original content whereas new users are more inclined to share other's content.

The estimated coefficient for ρ is an indicator for the existence of endogeneity. Our results show an insignificant ρ for the process of content creation and a significant ρ for the process of retweeting (p-value<0.1). Since endogeneity occurs when omitted variables impact both the error terms and the endogenous variables,[10] we interpret the result of ρ as follows: Unobservable factors (e.g., mutual interest in advocating for charities) drive both users' retweeting behaviors and their reciprocal followees' participation in Giving Tuesday; such factors do not drive people's original content creation concerning the charitable movement. We also note that when we assume $NRFollowees_i$ to be the endogenous variable, the estimated ρ is insignificant for both processes. The advantage of LIV to test exogeneity helps us to accurately specify our model, for which we only assume $RFollowees_i$ to be endogenous.

Other Coefficients

We report the other parameters for the LIV model in Table 5. The point estimates for the categorical means ($\boldsymbol{\pi}$) show good separation and the separation of the category probabilities ($\boldsymbol{\lambda}$) also seems to be sufficient. Ebbes (2004) considers the parameters ($\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$) that govern the exogenous part of the endogenous variable to be "nuisance" parameters—that is, of no theoretical interest. In rare cases, the predicted LIV instrument can be profiled using observed data (Ebbes 2004). However, most applications of LIV do not correspond to interpretable predicted LIV, and our study falls in this category.

Table 5. LIV Parameters of the Latent Variable

	Original Tweet		Retweet	
π_1	0.126***	(0.011)	0.127***	(0.011)
π_2	2.261***	(0.037)	2.265***	(0.037)
π_3	4.815***	(0.071)	4.818***	(0.069)
λ_1	0.822***	(0.0108)	0.823***	(0.0107)
λ_2	0.141***	(0.00784)	0.140***	(0.00781)
λ_3	0.0377***	(0.00551)	0.0376***	(0.00541)

[10] Theoretically, omitted variables, measurement error, self-selection, and simultaneity all lead to endogeneity. Technically, all these issues may be equated with the omitted variable problem because they are all concerned with an instrument that is correlated with the omitted relevant variable (Hueter 2016).

Note: *p<0.1; **p<0.05; ***p<0.01

ROBUSTNESS

We conducted a series of checks to examine the robustness of our findings. The first robustness check used a different and smaller charitable movement. The second and third robustness checks estimate the content creation and content sharing processes simultaneously. We incorporate users' past participation in previous Giving Tuesdays in the fourth robustness check. We use another "instrument-free" approach to handle the scenario with two endogenous variables as the last robustness check. In Appendix B1, we construct a panel data structure to control for the effect of time. In Appendix B2, we incorporate the length of content as an additional control variable.

Alternative Social Movement- Red Nose Day

In this robustness check, we examine users' participation in Red Nose Day, a small-scale charitable movement carried out on May 24, 2018. We obtained all tweets associated with this event through an official partner of Twitter. In total, 12,924 original tweets, 26,445 retweets, and 1,446 reply messages were procured. We matched these tweets' authors with the 2,033 Twitter users in the main analysis and found that 13 users retweeted posts containing the keyword "RedNoseDay" or "red nose day." Due to the relatively low participation rate, we did not find anyone creating original tweets related to Red Nose Day. We also obtained focal users' followees participating in Red Nose Day and their other tweets on the same day. We re-did the analysis and report the results in Table 6, for which the dependent variable is the retweeting decision. From this small-scale campaign, positive effects were found from non-reciprocal followees but not from reciprocal followees, which is consistent with our main finding.

Table 6. Robustness - Red Nose Day (N=2,033)

	Retweet	
$RFollowees_i$	0.0337	(0.183)
$NRFollowees_i$	0.265*	(0.16)

Note: Standard errors in parentheses,* p < 0.10, ** p < 0.05, *** p < 0.01, controls not reported.

Alternative Model - Biprobit Model

Our analyses to this point have estimated the process of content creation and content sharing separately. However, these two processes may be interrelated and estimating them simultaneously may lead to an efficiency improvement.[11] In this section, we propose a Biprobit model to estimate the process of tweeting and retweeting simultaneously (Roodman 2009). This model assumes that the error terms of the tweeting equation and the retweeting equation follow a bivariate normal distribution. The correlation between these two error terms is assumed to be ρ. The means of both error terms are assumed to be 0 and the variance of both is assumed to be 1. We report the results in Table 7. For the content creation process, only the effects of reciprocal followees are significantly positive. For the content sharing process, effects from both types of followees are significantly positive. Our hypothesis still holds that reciprocal followees have strength in promoting original content creation and non-reciprocal followees play a more significant role in content sharing. We note that the estimation of ρ is significant, indicating a positive association between the error terms of these two processes.

Table 7. Robustness - Biprobit Model (N=2033)

Original Tweet		
$RFollowees_i$	0.33***	(0.054)
$NRFollowees_i$	0.0435	(0.061)
Retweet		
$RFollowees_i$	0.202***	(0.04)
$NRFollowees_i$	0.159***	(0.0425)

[11] Theoretically, treating the two processes separately does not lead to biasedness but may result in larger standard errors with small samples.

ρ	0.442***	(0.077)

Note: Standard errors in parentheses,* p < 0.10, ** p < 0.05, *** p < 0.01, controls not reported.

Alternative Model - Multinomial Probit Model

In this robustness check, we classify users who both tweeted and retweeted as original content creators, thus users' options are mutually exclusive. We then estimate both diffusion processes at the same time by allowing users to choose from three options: tweet, retweet, and not to participate. A multinomial Probit model is employed with both tweeting and retweeting decisions compared to the base choice of not participating (Butler and Moffitt 1982). The results are presented in Table 8, and our hypothesis still holds.

Table 8. Robustness - Multinomial Probit (N=2,033)

Base		
Original Tweet		
$RFollowees_i$	0.472***	(0.078)
$NRFollowees_i$	0.0815	(0.0844)
Retweet		
$RFollowees_i$	0.255***	(0.0614)
$NRFollowees_i$	0.234***	(0.0652)

Note: Standard errors in parentheses,* p < 0.10, ** p < 0.05, *** p < 0.01, controls not reported.

Additional Control - Past Participation

In our main analysis, we did not account for users' past participation in Giving Tuesday because our users are a random sample from the Twittersphere. In this section, we collected all focal users' 62,208 tweets generated on Giving Tuesdays for the previous five years (11/27/2012, 12/3/2013, 12/2/2014, 12/1/2015, 11/29/2016). We found that among our focal users, four participated in the 2013 event, six participated in the 2014 event, eight participated in the 2015 event, and twelve participated in the 2016 event. We constructed a count variable $PastEvents_i$ to denote the number of past participation among users and included it in our analysis to examine the robustness of our result. As shown in Table 9, our hypothesis is still supported. From the estimated coefficient for $PastEvents_i$, we see that past participation in Giving Tuesday promotes users' content creation about this charitable movement but not content sharing.

Table 9. Robustness Check – Past Events (N=2,033)

	Original Tweet		Retweet	
$RFollowees_i$	0.306***	(0.0620)	0.166***	(0.0461)
$NRFollowees_i$	0.0311	(0.0645)	0.163***	(0.0470)
$PastEvents_i$	1.336***	(0.267)	0.315	(0.201)

Note: Standard errors in parentheses,* p < 0.10, ** p < 0.05, *** p < 0.01, controls not reported.

Alternative Model - Two Endogenous Variables

In this robustness check, we assume both $RFollowees_i$ and $NRFollowees_i$ to be endogenous. We use another statistical method, the copula approach, to account for the endogeneity issue. The copula approach nonparametrically estimates the marginal distribution of the endogenous regressors and uses a copula function to build the joint distribution of the endogenous regressors and the structural error term. Suppose that $H_1(\cdot)$ and $H_2(\cdot)$ are marginal distribution functions of $RFollowees_i$ and $NRFollowees_i$ that we obtained nonparametrically, and $G(u_i)$ is the marginal distribution of u_i where u_i is assumed to be normally distributed.

Let $F(RFollowees_i, NRFollowees_i, u_i)$ be the joint distribution function, and Sklar's Theorem states the existence of a copula function C such that

$$\begin{aligned} F(RFollowees_i, NRFollowees_i, u_i) &= C(H_1(RFollowees_i), H_2(NRFollowees_i), G(u_i)) \\ &= C(RFollowees_i^*, NRFollowees_i^*, u_i^*) \end{aligned} \quad , \quad (5)$$

Electronic copy available at: https://ssrn.com/abstract=2850301

where $H_1(RFollowees_i)$, $H_2(NRFollowees_i)$, and $G(u_i)$ are probability integral transformations represented by $RFollowees_i^*$, $NRFollowees_i^*$, and u_i^*. We operationalize C as a Gaussian copula such that

$$C(RFollowees_i^*, NRFollowees_i^*, u_i^*) = \Psi(\Phi^{-1}(RFollowees_i^*), \Phi^{-1}(NRFollowees_i^*), \Phi^{-1}(u_i^*)), \quad (6)$$

where Ψ denotes a three-dimensional multivariate normal distribution. We then have

$$\begin{pmatrix} RFollowees_i^* \\ NRFollowees_i^* \\ u_i^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{u1} \\ \rho_{12} & 1 & \rho_{u2} \\ \rho_{u1} & \rho_{u2} & 1 \end{bmatrix} \right), \quad (7)$$

where the variance for $RFollowees_i^*$ and $NRFollowees_i^*$ are assumed to be one, and that for the structural error term is assumed to be σ_u. Using the marginal distribution of multivariate normal distribution, we have

$$\begin{aligned} Tweet_{it}^* = {} & \gamma_1 RFollowees_i + \gamma_2 NRFollowees_i + \sigma_u \cdot \frac{\rho_{u1} - \rho_{12}\rho_{u2}}{1-\rho_{12}^2} \cdot RFollowee_i^* \\ & + \sigma_u \cdot \frac{\rho_{u2} - \rho_{12}\rho_{u2}}{1-\rho_{12}^2} \cdot NRFollowee_i^* + \boldsymbol{\beta}\mathbf{x}_i + \sigma_u \cdot \sqrt{1 - \rho_{u1}^2 - \frac{(\rho_{u2} - \rho_{12}\rho_{u1})^2}{1-\rho_{12}^2}} \cdot \omega \end{aligned}, \quad (8)$$

where $\boldsymbol{\beta}\mathbf{x}_i = \beta_0 + \beta_1 Tenure_i + \beta_2 OtherPosts_{it} + \beta_3 Status_i + \beta_4 Followers_i + \beta_5 Followees_i$. ω is a random variable not related to any other variable in equation (8), and we can unbiasedly estimate the coefficients of γ_1 and γ_2 with variables $RFollowee_i^*$ and $NRFollowee_i^*$ included in the estimation. From the estimation results in Table 10, we find that tweeting is only affected by reciprocal followees and retweeting behavior is only affected by non-reciprocal followees, which is consistent with our hypothesis. The diminished significance from reciprocal followees on retweeting decisions may be due to specifying an exogenous variable *NRFollowees* as endogenous.

Table 10. Robustness Check – Copula Approach (N=2,033)

	Original Tweet		Retweet	
$RFollowees_i$	0.538**	(0.268)	0.142	(0.215)
$NRFollowees_i$	0.491	(0.32)	0.849***	(0.257)
$\sigma_u \cdot (\rho_{u2} - \rho_{12}\rho_{u2}) / (1 - \rho_{12}^2)$	-0.807	(0.531)	-1.16***	(0.428)
$\sigma_u \cdot (\rho_{u1} - \rho_{12}\rho_{u1}) / (1 - \rho_{12}^2)$	-0.419	(0.659)	0.196	(0.526)

Note: Standard errors in parentheses,* p < 0.10, ** p < 0.05, *** p < 0.01, controls not reported.

DISCUSSIONS

Key Findings

In this study, we found that charitable content creation is prompted by reciprocal followees but not non-reciprocal followees. However, charitable content sharing is invoked by both reciprocal and non-reciprocal followees, with non-reciprocal followees having a greater impact. Below, we leverage the rich tweet content to exemplify theoretical arguments in support of the hypothesis and further interpret the results.

Qualitative Observations & Future Research

Our hypothesis posits that weak ties promote content sharing because they bring high-quality content. Indeed, we observe that tweets from non-reciprocal ties that prompt content sharing are often novel, such as this one: "*Google adds a donate button to search results to encourage giving during the holidays.*" In contrast, reciprocal followees' tweets that were not retweeted typically did not contain novel information, such as "*When communities need us the most, we're there. Please, remember the #RedCross today."*

To explain why weak ties have no significant impact on content creation, we argue that content creation is usually associated with high-cost actions such as charitable giving. As evidence, all tweets in our sample that report giving are originally created content because self-announcement is based on the first-person account. For example, one wrote, "*I gave to @kcpetproject on #GivingTuesday because I had to show some love to the creatures, too!*" Besides, content creation is of high risk as content creators may be held accountable when they endorse a charitable organization, especially a less reputable or less popular one. For example, one focal user followed multiple reciprocal followees to support @A4A_org, a non-profit organization with only four thousand followers. This is less likely to happen if this user does not observe sufficient affirmation from strong ties.

Lastly, we argue that a non-reciprocal relationship prohibits the accumulation of social capital (Rogers and Kincaid 1981). A significant form of social capital is an effective norm, or more intuitively, social pressure. If your participation is driven by social pressure from your reciprocal followee Matthew, you may want to leverage visible channels to ensure that your participation is observed. When retweeting an existing post, Matthew may not notice your effort because the original tweeter is highlighted as creating the content. As evidence of people's desire to be observed, one focal user tweeted "*On #GivingTuesday; I warmly giveBackToALLofU!*" with a few mentions of this users' reciprocal followees who had participated in Giving Tuesday. The story is different when users mention their non-reciprocal Twitter tie such as Ellen DeGeneres because she is unlikely to notice such mentions.

Other than the above observations specific to our results, we have some other observations that may pave the way for future study. The tweet that was the most often seen includes simply the hashtag #GivingTuesday and no other content. Such content was exposed to ten unique users in our data, yet none of these users participated in Giving Tuesday. Apparently, this tweet was neither informative nor emotionally charged. The tweet "*#GivingTuesday is powered by #netneutrality. Tell Congress to stop the FCC's plan to end the open internet now*." was exposed to three unique users in our sample and was unsuccessful in inducing further participation. This tweet is more politically oriented, possibly leading to reluctance in response from the audience. One tweet that successfully induced action reads "*Today #GivingTuesday, I wrote about why I give to the Internet Archive: this year we scanned a long-lost reel of film from the Japanese American incarceration camp at Jerome, AK. It hasn't been seen in 72 years!*" This tweet conveys a detailed personal story. Another successful tweet in invoking participation reads "*Thanks, Jessica! We hope more of our followers will be moved to support the reporting too.*" This tweet acknowledged a previous giver before calling for similar behaviors and could be a good social media marketing strategy. Such observations make us believe that it is promising to examine the lexical features of content in persuasiveness – the ability to induce desirable economic or social behaviors. Given the increasing call for socially responsible digital platforms, such studies are urgently needed to understand digital social movements.

Theoretical Implications

Our study makes a distinct theoretical contribution to the literature. Information technology has been considered a new vehicle for building social networks (Bandura 2009). It has been well documented that electronic social networks transcend the barriers of time and space unlike traditional social networks (Garton et al. 1997). Wellman (1997) proposed the importance of studying online interaction using a social network approach and specifically listed strength of tie as a topic of investigation. Our study followed this research agenda and discovered that reciprocal and non-reciprocal followees facilitate distinct yet complimentary paths for information diffusion. This new perspective allows us to tap into the rich literature of sociology and social network to understand social broadcasting networks. More importantly, we provide empirical evidence for Centola and Macy's (2007) contagion complexity theory in the social media context and expand the explanory power of the theory in the digital world.

We expanded past works by differentiating between content creation and content sharing. In contrast to the previous finding that shows a similar topical pattern between original and shared content (Geva et al. 2019), we discover the different motivations behind those two diffusion processes. The delineation between content creation and sharing also enhances the understanding of social media as a way to process citizen-driven information collectively (Oh et al. 2013). Specifically, content creation drives content diversity (i.e., emancipatory) and content sharing promotes content concentration (i.e., hegemonic). To this end, we answer the call of Miranda et al. (2016) to "compare different digitalized processes" to develop a more comprehensive theory of the differentiating effects of these processes.

Practical Implications

Our findings have direct implications for social media marketing. To promote content sharing and raise more awareness, charitable campaign managers should encourage celebrity or media accounts (i.e., weak ties) to broadcast novel information like corporate matching programs and new ways of fundraising. To deepen the conversation, managers should encourage users to share their own stories and invite their friends to do the same. Further, when the contagion is complex, like dumping a bucket of icy water over one's head, campaigns making use of social pressure, like friend tagging, may work better (Townsend 2014).

The diffusion patterns we found can potentially be generalized into social movements concerning public welfare or even social media marketing in general. Individual behaviors can only be integrated into collective actions when the behaviors meet the expectations and values of the community. Therefore, it is critical to form social norms in order to launch social movements or marketing campaigns successfully. For example, to encourage Twitter users to announce their voting behavior in a presidential election, marketing managers need to prioritize users who are deeply embedded in the network. Users with many reciprocal relationships are good choices to spend more marketing budget on, as they may effectively drive the chain of adoption actions.

LIMITATIONS AND CONCLUSIONS

There are a few limitations to our study. First, our focus on charitable movements means that the diffusion complexity of tweeting and retweeting is significantly different than other contexts. For example, corporations sometimes launch social responsibility programs to pledge a fixed monetary amount to charities for every social media mention that includes a specific keyword. Our findings may not hold in this case as the original content creation merely necessitates copying and pasting the keyword. Second, we focus on the tweets posted on the exact date of Giving Tuesday. Different findings may be generated by looking at the dynamics before and after the event. Third, our work is conducted at a node level but not a dyad level because we do not know which tweet trigged users to post on a particular topic. Therefore, we lack insight into how different tweet content persuades readers. Moreover, social broadcasting sites are rich in ties that feature the need for content consumption. In other networking environments with different types of ties, like Facebook, where online friends often maintain offline relationships, we may see different dynamics of diffusion. In addition, our data sample represents users who actively generate or share content but not those who only use Twitter to acquire information. Finally, due to data limitations, we did not examine social influence measures in conjunction with fundraising outcomes. Future research can fill this void by linking social media topic richness (results of content creation) and topic mention volume (results of content sharing) to fundraising outcomes.

By looking separately at content creation and content sharing, our findings contribute to the debate on the comparative strength of strong and weak ties. This unique angle distinguishes our study from existing research on information systems and online communities. Given the need to engage individuals in collective actions of various contexts (e.g., political movements, product adoption, professional collaboration, and cultural interactions), close scrutiny of diffusion complexity is compelling. Along with contributing to the theoretical understanding of the diffusion process in social broadcasting networks, we hope our study helps practitioners better strategize for their social media campaigns.

REFERENCE

Aral, S., and Van Alstyne, M. 2011. "The Diversity-Bandwidth Trade-Off," *American Journal of Sociology* (117:1), pp. 90-171.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. "Everyone's an Influencer: Quantifying Influence on Twitter," *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65-74.

Bandura, A. 2009. "Social Cognitive Theory of Mass Communication," in *Media Effects*. Routledge, pp. 110-140.

Baumgarten, S. A. 1975. "The Innovative Communicator in the Diffusion Process," *Journal of Marketing Research* (12:1), pp. 12-18.

Bhattacharya, D., and Ram, S. 2012. "Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter," *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*: IEEE, pp. 966-971.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature* (489:7415), p. 295.

Bound, J., Jaeger, D. A., and Baker, R. M. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American statistical association* (90:430), pp. 443-450.

Butler, J. S., and Moffitt, R. 1982. "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica: Journal of the Econometric Society* (50:3), pp. 761-764.

Castillo, C., El-Haddad, M., Pfeffer, J., and Stempeck, M. 2014. "Characterizing the Life Cycle of Online News Stories Using Social Media Reactions," *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*: ACM, pp. 211-223.

Centola, D. 2010. "The Spread of Behavior in an Online Social Network Experiment," *science* (329:5996), pp. 1194-1197.

Centola, D., and Macy, M. 2007. "Complex Contagions and the Weakness of Long Ties," *American journal of Sociology* (113:3), pp. 702-734.

Clogg, C. C., Petkova, E., and Haritou, A. 1995. "Statistical Methods for Comparing Regression Coefficients between Models," *American Journal of Sociology* (100:5), pp. 1261-1293.

Coleman, J. S. 1988. "Free Riders and Zealots: The Role of Social Networks," *Sociological Theory* (6:1), pp. 52-57.

Coleman, J. S., Katz, E., and Menzel, H. 1966. *Medical Innovation: A Diffusion Study*. Bobbs-Merrill Co.

Cropanzano, R., and Mitchell, M. S. 2005. "Social Exchange Theory: An Interdisciplinary Review," *Journal of management* (31:6), pp. 874-900.

Cunha, E., Magno, G., Comarela, G., Almeida, V., Gonçalves, M. A., and Benevenuto, F. c. "Analyzing the Dynamic Evolution of Hashtags on Twitter: A Language-Based Approach," *Proceedings of the Workshop on Languages in Social Media*, pp. 58-65.

Davis, J. A. 1970. "Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociomatrices," *American Sociological Review*), pp. 843-851.

Ebbes, P. 2004. "Latent Instrumental Variables," *Unpublished Ph. D. thesis, University of Groningen, Groningen*).

Ebbes, P., Wedel, M., and Böckenholt, U. 2009. "Frugal Iv Alternatives to Identify the Parameter for an Endogenous Regressor," *Journal of Applied Econometrics* (24:3), pp. 446-468.

Ebbes, P., Wedel, M., Böckenholt, U., and Steerneman, T. 2005. "Solving and Testing for Regressor-Error (in) Dependence When No Instrumental Variables Are Available: With New Evidence for the Effect of Education on Income," *Quantitative Marketing and Economics* (3:4), pp. 365-392.

Friedkin, N. 1980. "A Test of Structural Features of Granovetter's Strength of Weak Ties Theory," *Social networks* (2:4), pp. 411-422.

Gaffney, D. 2010. "Iranelection: Quantifying Online Activism," *In Proceedings of the Web Science Conference (WebSci10*: Citeseer.

Garton, L., Haythornthwaite, C., and Wellman, B. 1997. "Studying Online Social Networks," *Journal of computer-mediated communication* (3:1), p. JCMC313.

Geva, H., Oestreicher-Singer, G., and Saar-Tsechansky, M. 2019. "Using Retweets When Shaping Our Online Persona: Topic Modeling Approach," *MIS Quarterly* (43:2).

Gleason, B. 2013. "# Occupy Wall Street: Exploring Informal Learning About a Social Movement on Twitter," *American Behavioral Scientist* (57:7), pp. 966-982.

Granovetter, M. 1983. "The Strength of Weak Ties: A Network Theory Revisited," *Sociological theory* (1), pp. 201-233.

Granovetter, M. S. 1973. "The Strength of Weak Ties," *American journal of sociology* (78:6), pp. 1360-1380.

Hansen, M. T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Administrative science quarterly* (44:1), pp. 82-111.

Hueter, I. 2016. "Latent Instrumental Variables: A Critical Review," *Institute for New Economic Thinking Working Paper Series*:46).

Java, A., Song, X., Finin, T., and Tseng, B. 2007. "Why We Twitter: Understanding Microblogging Usage and Communities," *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*: ACM, pp. 56-65.

Laumann, E. 1968. "Interlocking and Radial Friendship Networks: A Crosssectional Analysis," *Mimeographed. Ann Arbor: University of Michigan*).

Ma, L., Krishnan, R., and Montgomery, A. L. 2014. "Latent Homophily or Social Influence? An Empirical Analysis of Purchase within a Social Network," *Management Science* (61:2), pp. 454-473.

Marwell, G., and Oliver, P. 1993. *The Critical Mass in Collective Action*. Cambridge University Press.

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a Feather: Homophily in Social Networks," *Annual review of sociology* (27:1), pp. 415-444.

Miranda, S. M., Young, A., and Yetgin, E. 2016. "Are Social Media Emancipatory or Hegemonic? Societal Effects of Mass Media Digitization," *MIS quarterly* (40:2), pp. 303-329.

Oh, O., Agrawal, M., and Rao, H. R. 2013. "Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises," *Mis Quarterly*), pp. 407-426.

Park, S., and Gupta, S. 2012. "Handling Endogenous Regressors by Joint Estimation Using Copulas," *Marketing Science* (31:4), pp. 567-586.

Paternoster, R., Brame, R., Mazerolle, P., and Piquero, A. 1998. "Using the Correct Statistical Test for the Equality of Regression Coefficients," *Criminology* (36:4), pp. 859-866.

Petrovic, S., Osborne, M., and Lavrenko, V. 2011. "Rt to Win! Predicting Message Propagation in Twitter," *ICWSM* (11), pp. 586-589.

Putnam, R. D. 1995. "Tuning in, Tuning Out: The Strange Disappearance of Social Capital in America," *PS: Political science & politics* (28:4), pp. 664-683.

Rodan, S., and Galunic, C. 2004. "More Than Network Structure: How Knowledge Heterogeneity Influences Managerial Performance and Innovativeness," *Strategic management journal* (25:6), pp. 541-562.

Rogers, E. M. 2010. *Diffusion of Innovations*. Simon and Schuster.

Rogers, E. M., and Kincaid, D. L. 1981. "Communication Networks: Toward a New Paradigm for Research,").

Romero, D. M., Meeder, B., and Kleinberg, J. 2011. "Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter," *Proceedings of the 20th international conference on World wide web*: ACM, pp. 695-704.

Roodman, D. 2009. "Estimating Fully Observed Recursive Mixed-Process Models with Cmp,").

Rutz, O. J., Bucklin, R. E., and Sonnier, G. P. 2012. "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research* (49:3), pp. 306-319.

Ryan, B., and Gross, N. C. 1943. "The Diffusion of Hybrid Seed Corn in Two Iowa Communities," *Rural sociology* (8:1), p. 15.

Sandefur, R. L., and Laumann, E. O. 1998. "A Paradigm for Social Capital," *Rationality and society* (10:4), pp. 481-501.

Shi, Z., Rui, H., and Whinston, A. B. 2014. "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter," *MIS Quarterly* (38:1), pp. 123-142.
Shore, J., Baek, J., and Dellarocas, C. 2016. "Network Structure and Patterns of Information Diversity on Twitter," *Boston University Questrom School of Business Research Paper*:2813342).
Smith, M. A. 1999. "Communities in Cyberspace,").
Sonnier, G. P., McAlister, L., and Rutz, O. J. 2011. "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Marketing Science* (30:4), pp. 702-716.
Taxidou, I., and Fischer, P. M. 2014. "Online Analysis of Information Diffusion in Twitter," *Proceedings of the 23rd International Conference on World Wide Web*: ACM, pp. 1313-1318.
Townsend, L. 2014. "How Much Has the Ice Bucket Challenge Achieved?."
Valenzuela, S., Correa, T., and Gil de Zúñiga, H. 2018. "Ties, Likes, and Tweets: Using Strong and Weak Ties to Explain Differences in Protest Participation across Facebook and Twitter Use," *Political Communication* (35:1), pp. 117-134.
Wejnert, B. 2002. "Integrating Models of Diffusion of Innovations: A Conceptual Framework," *Annual review of sociology* (28:1), pp. 297-326.
Wellman, B. 1997. "An Electronic Group Is Virtually a Social Network," *Culture of the Internet* (4), pp. 179-205.
Zeng, X., and Wei, L. 2013. "Social Ties and User Content Generation: Evidence from Flickr," *Information Systems Research* (24:1), pp. 71-87.
Zhang, J., Wedel, M., and Pieters, R. 2009. "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis," *Journal of Marketing Research* (46:5), pp. 669-681.