Network Overlap and Content Sharing on Social Media Platforms

Jing Peng[1], Ashish Agarwal[2], Kartik Hosanagar[3], Raghuram Iyengar[3]
[1]School of Business, University of Connecticut
[2]McCombs School of Business, University of Texas at Austin
[3]The Wharton School, University of Pennsylvania
jing.peng@uconn.edu ashish.agarwal@mccombs.utexas.edu {kartikh, riyengar}@wharton.upenn.edu

March 2017

Acknowledgments. We benefited from feedback from session participants at 2013 Symposium on Statistical Challenges in eCommerce Research, 2014 International Conference on Information Systems, 2015 Workshop on Information in Networks, 2015 INFORMS Annual Meeting, and 2015 Workshop on Information Systems and Economics. The authors would like to thank Professors Christophe Van den Bulte, Paul Shaman, and Dylan Small for helpful discussions. This project was made possible by financial support extended to the first author through Mack Institute Research Fellowship, President Gutmann's Leadership Award, and Baker Retailing Center PhD Research Grant.

Network Overlap and Content Sharing on Social Media Platforms

ABSTRACT

We study the impact of network overlap – the overlap in network connections between two users – on content sharing in directed social media platforms. We propose a hazards model that flexibly captures the impact of three different measures of network overlap (i.e., common followees, common followers and common mutual followers) on content sharing. Our results indicate a receiver is more likely to share content from a sender with whom they share more common followees, common followers or common mutual followers after accounting for other measures. Additionally, common followers have a higher effect than the common mutual followers on the sharing propensity of the receiver. Finally, the effect of common followers and common mutual followers is positive when the content is novel but decreases, and may even become negative, when many others have already adopted it. Using three datasets from two social media platforms (Twitter and Digg), we find that our findings apply to the sharing of content generated by both firms and users. Our findings are managerially relevant for targeting customers for content propagation in social networks.

Keywords: social media, content sharing, network overlap, multiple senders, hazards model

INTRODUCTION

Social media platforms are a popular medium for firms to reach out to customers (Schweidel and Moe 2014; Stephen and Toubia 2010). On these platforms, firms connect with users who in turn are connected to other users and these connections form the social network. When firms post content, users who are their direct connections can see it. These users can, in turn, choose to share the content with others. For example, on Twitter users "retweet" content they receive in order to share it with others. Similarly, on Facebook users can share content they receive and this allows their connections to see it. A primary requirement for the propagation of content in such social networks is that content receivers, in turn, share or rebroadcast the content that they obtain from their senders. By understanding the key factors influencing sharing on social media platforms, marketers can more effectively disseminate content on these platforms. This has led to an increased interest in studying the content sharing propensity of users on such platforms (e.g., Lambrecht et al. 2015; Suh et al. 2010; Luo et al. 2013; Zhang et al. 2016).

Users share content for a purpose and that determines what they share. Existing literature on word-of-mouth (WOM) has focused on how content and brand characteristics drive the aggregate WOM performance (see Berger 2014 and Lovett et al. 2013 for details). Similarly, studies pertaining to content sharing on social media platforms have focused on the role of content on the sharing propensity of users (Suh et al. 2010; Zhang et al. 2016; Lee et al. 2017). However, the social network structure of users can also influence their sharing decisions (see Van den Bulte and Wuyts 2007, p.43). These network characteristics of users can be easily observed in an online social network and provide quantifiable metrics to managers for operationalizing their social media marketing efforts (Van Den Bulte and Wuyts 2007, p.11). To this end, several studies have focused on the role of network characteristics on content propagation in online social networks. These include the role of sender network characteristics (Bakshy et al. 2011; Susarla et al. 2012; Yoganarasimhan 2012; Shriver et al. 2013; Suh et al. 2010) for spreading content and the role of receiver network characteristics (Luo et al. 2013) on her sharing propensity. However, a receiver's propensity to share content from a particular sender can also depend on their *shared* connections. These dyadic characteristics may represent underlying shared interests and redundancies for the sender-receiver dyad and can vary across dyads. Knowledge of the resulting sharing propensity of senders' extended network based on dyadic characteristics can be useful to improve the selection of influentials for spreading content (Trusov et al. 2010). The purpose of this article is to assess the impact of network overlap, a dyadic network characteristic, on the level of content sharing in social networks.

Network overlap is broadly defined as the number of common connections between two users (Easley and Kleinberg 2010). Network overlap[1] has been associated with effective knowledge transfer between individuals (Reagans and McEvily 2003), economic trust between users (Bapna et al. 2016), adoption of applications by users (Aral and Walker 2014) and diversity of information received by a user (Aral and Van Alstyne 2011). Its operationalization depends on whether the network is directed or not. In undirected networks (e.g., Facebook), network overlap simply means the number of common friends between two users. In directed networks (e.g., Twitter), by interpreting a connection as a followee (outgoing link), follower (incoming link) or mutual follower (bidirectional link), network overlap can be characterized by three different metrics: the numbers of common followees, common followers, and common mutual followers.

[1] Aral and Walker (2014) use the term "embeddedness" to represent network overlap. However, embeddedness has been used to represent network constraints associated with an actor in a network (Granovetter 1985). In order to avoid confusion, we use the term "network overlap". Reagans and McEvily (2003) use the term "social cohesion" to further incorporate the weight of each overlapping connection.

Table 1 summarizes the definitions of these terms. The distinction between followers and followees is important. In directed networks like Twitter and Weibo, one can follow a user without consent from the user. Followees of a focal user thereby represent the set of users whose activities are of interest to the focal user, whereas the followers represent the set of users who are interested in the focal user's activity. Mutual follower (a bidirectional link) cannot be established unless users have mutual interest. Social media platforms also allow a user to view information about another user's followees, followers and the common connections they share. For example, Figure 1 shows the detailed network information of a user followed by a focal user on Twitter. As the figure shows, the focal user can see how the other user is connected to her followees and followers and determine the extent of network overlap with the other user.

The type and extent of overlap in the network connections between two users can influence the sharing propensity in different ways. As mentioned earlier, higher number of common followees suggests that the sender and the receiver have similar interests and in turn, may have similar propensity to share a particular piece of content. In addition, the interest of the audience also plays a role (Berger 2014). More common followers and common mutual followers between the sender and the receiver may suggest that their followers share a similar taste or interest. In this case, a receiver may consider content to be more suitable for her audience and may have a higher propensity to share it. Note this can be the case even if sender and receiver do not have high number of common followees. Further, a receiver may respond differently to the taste of audience depending on whether she shares weak (followers) or strong (common followers) ties with her audience (Dubois et al. 2016). Moreover, a higher number of common mutual followers may represent a stronger social bond between the sender and the receiver (Alexandrov et al. 2013; Berger 2014). Thus, the need to maintain the social bond may also increase the propensity of a receiver to share the content. On the other hand, a larger common audience in terms of common followers and common mutual followers may suggest higher redundancy in the information received by the audience and deter a user from sharing the content to satisfy their desire for uniqueness (Alexandrov et al. 2013; Berger 2014; Cheema and Kaikati 2010; Ho and Dempsey 2010; Lovett et al. 2013). Thus, users may be less likely to share popular content as many others have already shared it. [2]

How different measures of network overlap will impact users'sharing propensity is an empirical question. Only a few studies thus far have evaluated the role of network overlap but have not focused on socially visible content sharing. The study most closely related to ours is that by Aral and Walker (2014) who evaluate the effect of common mutual followers on product adoption. However, in their context, users do not publicly share information about their adoption with others. In contrast, content sharing or broadcasting on social networks is a publicly visible activity. Certain factors are equally relevant in adoption in private as well as content sharing. For example, network overlap can represent similarity of interests between the sender and the receiver and therefore be associated with both a greater propensity to adopt and a greater propensity to share information by the receiver. However, there are important differences as well. Adoption in private is unlikely to be driven by goals such as taste of audience, social bonding or appearing unique, which are relevant for socially visible content sharing.[3] Thus, the common audience between the sender and the receiver, represented by their common followers and common mutual followers, may only be relevant for content sharing but not for private

[2] While making a sharing decision, a receiver can be expected to know the order of magnitude for different types of network overlap with a sender but may not know the exact value of each type of network overlap.

[3] On social media platforms, users can take actions such as adopting a digital product or view content without publicly revealing it to others. Nevertheless, this may not apply to actions that are inherently social in nature such as playing online social games.

adoption. In line with this notion, past work suggests that the factors influencing adoption in private differ from those relevant for adoption where the information is shared (Cheema and Kaikati 2010; Childers and Rao 1992). An additional difference is that Aral and Walker (2014) consider network overlap in an undirected network whereas our study is set in directed networks. This distinction is relevant because network overlap is operationalized differently in directed networks and some measures of network overlap may be more relevant in content sharing than others. Thus, how the sharing propensity varies with different types of network overlap should be examined separately.

In this paper, we evaluate the impact of network overlap for content sharing within sender-receiver dyads. Our micro-level model for sharing accounts for users' profile information and their social network. We estimate the model using a dataset, which contains sharing of tweets posted by Fortune 500 companies on Twitter. We show the robustness of our results using a second Twitter dataset that focuses on the sharing of tweets posted by regular users and a third dataset that contains the sharing of sponsored ads posted by companies on Digg. At the time of data collection, both websites Twitter and Digg maintained a directed social network, allowing users to follow others to keep themselves informed about their activities. We analyze the data using a novel proportional hazards model that allows an event to have more than one cause. The proposed model can identify the contribution of each co-sender based on her characteristics and has broader application in studies of content propagation in networks.

We emerge from the analyses with three key findings. First, we establish that network overlap plays a significant role in content sharing on online social networks. Second, the propensity of a receiver to share content depends on all three measures of network overlap (i.e., common followees, common followers and common mutual followers) suggesting that each measure independently contributes to the sharing propensity. Interestingly, sharing propensity increases more so with common followers as compared to common mutual followers. Third, the effects of common followers and common mutual followers are moderated by the novelty of content. Their effects are positive only when the content is relatively novel (i.e., not shared by many others). When many others have shared the content, the positive effects decrease and may even become negative, suggesting that users' need for uniqueness is a likely mechanism at work. This finding suggests a boundary condition for the positive impact of network overlap documented in earlier studies (Aral and Walker 2014; Bapna et al. 2016). We use a simulation study based on our model to show how one can improve the selection of influential users to spread content based on the network overlap with their followers. Further, the optimal set of influential users to target depends on the popularity of content.

The rest of the paper is organized as follows. We begin with a discussion of related literature and develop specific hypotheses about the impact of the three network overlap metrics on content sharing. Then, we describe the proposed model and the dataset from Twitter that we use in the application. Next, we discuss the results of model estimation and several robustness checks including generalizability of our results with two additional datasets. After that, we illustrate the quantitative effects of network overlap on content sharing using a simulation study. Finally, we conclude with theoretical and managerial implications of our work.

LITERATURE AND THEORETICAL FRAMEWORK

Our work relates to the broad literature on the role of network characteristics of users in a social network on their actions. These include studies based on unitary network characteristics of content senders (Bakshy et al. 2011; Shriver et al. 2013; Susarala et al. 2012; Yoganarasimhan

2012), unitary network characteristics of content receivers (Bapna and Umyarov 2015; Centola 2010; Haenlein 2013; Iyengar et al. 2011; Katona et al. 2011; Nitzan and Libai 2011) and dyadic characterisitics of sender-receiver pairs (Aral and VanAlstyne 2011; Aral and Walker 2014; Shi et al. 2014). Below, we briefly discuss research in each of the three areas.

Unitary network characteristics of senders. A few studies have investigated the role of unitary network characteristics of senders on the overall extent of content diffusion in the network. For example, Yoganarasimhan (2012) studies how the size and structure of the local network of a sender, posting videos, affect the diffusion of these videos in undirected networks on YouTube. The specific network characteristics investigated include the numbers of first- and second-degree friends, the clustering coefficient and the betweenness centrality of the user. Susarla et al. (2012) conduct a similar analysis but include both undirected (friendship) and directed (subscription) networks on YouTube. Bakshy et al. (2011) determine a user's influence on the diffusion of content based on the cascade sizes associated with the user's extended network. While these studies examine the influence of senders, they do not consider sender's propensity to share content per se. They also do not consider the role of receivers in propagating the content. One recent study that does consider the propensity to share content is Shriver et al. (2013) which shows users with more network ties are more likely to generate content and this, in turn, leads to more network ties. In the context of content sharing, Suh et al. (2010) show that the retweetability of a tweet depends on the author's followee and follower numbers. However, they do not consider an individual receiver's propensity to further share or rebroadcast the content and how it depends on the dyadic network characteristics between the sender-receiver pair.

Unitary network characteristics of receivers. Several studies have focused on establishing the role of local network characteristics of a content receiver on their subsequent actions. Such information may be about how other users have adopted (or disadopted) a product or posted content. For instance, Katona et al. (2011) show that the local network characteristics such as degree and density of existing adopters associated with a user has a positive impact on her adoption or registration at a site. Similarly, Centola (2010) shows that users are more likely to adopt when they receive social reinforcement from multiple connections. Rand and Rust (2011) evaluate the role of local network on the adoption behavior using an agent based model. Apart from the number of adopters, Iyengar et al. (2011) also demonstrate the impact of opinion leadership (captured by the number of ties and self-reported measures) on the adoption of a prescription drug. Similarly, Bapna and Umyarov (2015) consider the effect of a receiver's network size on her propensity to subscribe to a music site. Some research has also explored the role of unitary network characteristics on the churn (or disadoption) behavior. For example, Nitzan and Libai (2011) show that the customer churn behavior is more likely for a user who has a greater number of defecting social connections. Similarly, Haenlein (2013) investigates the role of the social contacts' churn behavior on the retention behavior of an individual in a directed network and find that the likelihood of a user to churn increases with the number of defecting users with whom she has outgoing calling relationships. In the context of content sharing, Luo et al. (2013) show that propensity of a user to retweet a tweet depends on variables representing her social status, such as number of followees and followers. However, none of the studies considers the role of *shared* network characteristics between the sender-receiver pair on the receiver's propensity to take an action or share content.

Dyadic network characteristics. More recent studies consider the role of dyadic network characteristics on a focal user's actions. At a dyadic level, there are different types of network characteristics and these can play a role on user's actions. For instance, Shi et al. (2014) study content sharing propensity of receivers and primarily focus on the role of reciprocity between senders and receivers. However, they do not consider the impact of network overlap on the sharing propensity. Using an email network, Aral and Van Alstyne (2011) show that the novelty of information a user receives is positively associated with her network diversity. Network diversity in their setup captures the lack of redundancy in the network connections and can be related to the extent of network overlap of a user with her network connections.[4] However, they do not evaluate the receiver's propensity to share the content with her followers and how it varies with her network overlap with the sender. Particularly, they do not explicitly evaluate the effect of audience overlap, represented by network overlap measures such as common followers and common mutual followers, in rebroadcasting or content sharing decisions. More recently, Aral and Walker (2014) examine the effect of common friends (common mutual followers) between a sender and a receiver and find that it has a positive effect on the adoption of an application on Facebook. However, they cannot comment on users' actions as a function of overlaps observed in directed networks such as common followees and common followers. Additionally, the adoption of the application is a private decision as users do not share this information with others.[5] The effect of network overlap on adoption in private may not apply to sharing as factors driving adoption in private can differ from those driving publicly visible content sharing. For

[4] While Aral and Van Alstyne (2011) do not conduct an actual dyadic level analysis, their measure captures the effect of average overlap of a user with her network neighbors.

[5] Online social networks allow users to share content, as well as their actions, to their connections. Aral and Walker (2014) randomize the sharing of app usage information across users in their experiment. However, the app adoption decision is private per se. Further, the visibility of the app usage information (to a small subset of friends) is passively manipulated by the app rather than actively enabled by users.

example, identity communication is more relevant for publicly consumed products as compared to privately consumed products (Cheema and Kaikati 2010; Childers and Rao 1992). In this sense, users are likely to have identity-based considerations while sharing content in the presence of common audience with their senders, which may not be a concern while adopting a product in private.

In sum, there is much interest in understanding how users' network characteristics affect content sharing in networks. While the literature has focused on unitary network characteristics of users on content sharing, an emerging stream of work has started to highlight the role of shared dyadic attributes such as network overlap. This literature, to the best of our knowledge, has not considered the role of different types of network overlap on content sharing. In this paper, we fill the gap and evaluate how different types of network overlaps affect content sharing in social networks. Table 2 provides a summary of existing literature.

Hypotheses

Consumers typically share content to satisfy multiple goals (Berger 2014; Lovett et al. 2013). In the case of content sharing, we posit that different types of network overlaps between the receiver and the sender are an important contextual feature that can moderate how likely a receiver will satisfy one or more of the sharing goals and, thereby, influence the receiver's propensity to share content with others. Next, we discuss our hypotheses on how the three types of network overlap satisfy user goals and their potential impact on content sharing.

Common Followees

Users share content to shape others' impression about them (Berger 2014; Chung and Darke 2006). On social media platforms, users' activities are publicly visible to others. Such visibility of individual activities makes social media platforms an ideal place to create an impression. Users may try to impress others by communicating specific identities (Berger 2014). For instance, people share topics or ideas that signal that they have certain knowledge in a particular domain (Berger 2014; Chung and Darke 2006; Packard and Wooten 2013).

In a directed social network, people follow others to keep themselves informed about their activities and posted content. Thus, the composition of one's followees largely reflects her topical interest or taste. A user is likely to share content she is knowledgeable about to create an impression. The more common followees two users have, the more likely they have similar interests. In that case, a receiver with more common followees with the sender is more likely to be knowledgeable about the content and more likely to share it with others as compared to a receiver with fewer common followees with the sender. We posit the following:

H1: The propensity of a receiver to share a piece of content from a sender is positively associated with the number of common followees between the sender and the receiver.

Common Followers

The composition of one's followers represents the taste of her audience. To establish a good impression, the taste of the audience is a factor that users are likely to consider while sharing content (Berger 2014; McQuarrie et al. 2013). The more common followers two users have, the more similar audience they have, and the more likely they will make similar decisions on whether or not to share a piece of content to their followers to create an impression. In addition, similarity in the interests of the audiences of two users also represents similarity in their own expertise or knowledge. As users tend to signal their identity by sharing their knowledge, this would further increase the propensity of the receiver to share the same content with the sender.

On the other hand, an alternative driver that may lower the propensity of a receiver to share the content obtained from a sender with whom the receiver has a lot of common followers is the need for uniqueness. Users are intrinsically motivated to achieve uniqueness (Cheema and Kaikati 2010; Tian et al. 2001) and being overly similar to others induces negative emotions (Snyder and Fromkin 1980). This desire to express uniqueness is stronger for publicly consumed products than privately consumed products (Cheema and Kaikati 2010). Moreover, the need for uniqueness is stronger in online interactions than offline interactions and leads to higher WOM for differentiated brands (Lovett et al. 2013). Need for uniqueness has been observed for other user generated content such as reviews (Ludford et al. 2004) and photographs (Zeng and Wei 2013). Past work suggests that users can satisfy their need for uniqueness by sharing novel online content (Ho and Dempsey 2010). Thus, in order to establish a unique identity on social media platforms, a user may resist sharing content that have already been shared by many others.

Following these arguments, we propose the following hypotheses.

H2: The propensity of a receiver to share a piece of content from a sender is positively associated with the number of common followers between the sender and the receiver.

H3: The positive effect of common followers on the receiver’s propensity to share content from sender decreases with the popularity of the content.

Common Mutual Followers

Due to the bidirectional nature of the links with mutual followers, the number of such links characterizes the mutual accessibility of two users through third-parties. According to the bandwidth hypothesis (Aral and Van Alstyne 2011; Burt 2001), the existence of common mutual connections expands the bandwidth of communication among users and makes their evaluation of each other more accurate. In this case, a receiver may find information received from a sender with high mutual common followers to be more useful and is more likely to share it for creating an impression (Berger 2014). In addition, as the bidirectional link represents a strong tie (Shi et al. 2014), the more common mutual followers two users have, the more likely they belong to the same social group. Thus, a user may have a higher need to interact with the sender with common mutual followers to meet the need for social bonding (Baumeister and Leary 1995).

On social media platforms, as the user actions are visible, one way to interact with the sender is to propagate the content received from the sender. The closer two users are, the stronger obligation they may have in sharing content shared by each other. Thus, the higher the number of common mutual followers two users have, the more likely they feel obligated to propagate content shared by each other to maintain a strong social bond. Finally, more common mutual followers may also suggest more similar audience with the sender, as well as a higher similarity in taste with the sender due to homophily, even after accounting for the effect of other network overlap metrics. This would further increase the receiver's propensity to share content from the sender.

However, a user's need for uniqueness can lower her propensity to share content from a sender with whom she shares mutual common followers. Similar to our earlier reasoning for the effect of common followers on content sharing, when the content to be shared is popular, a receiver with a large number of common mutual followers with the sender may resist doing so to avoid excessive similarity with the sender, as well as with other members in the same social group. However, when the content is relatively novel, the need for uniqueness is satisfied and the receiver would have a high propensity to share content due to high bandwidth and strong social bonding. We summarize the expected effects of common mutual followers in H4 and H5.

H4: The propensity of a receiver to share a piece of content from a sender is positively associated with the number of common mutual followers between the sender and the receiver.

H5: The effect of common mutual followers on the receiver's propensity to share content from sender decreases with the popularity of the content.

Common Mutual Followers vs. Common Followers

Due to higher bandwidth and stronger social bonding needs associated with a sender with common mutual followers, a receiver's propensity to share content received from such a sender should be higher as compared to her propensity to share content from a sender with common followers. However, network overlap represented by common followers and common mutual followers also captures the taste of the audience. In a broadcasting context, users focus on their need to create an impression while sharing content (Barasch and Berger, 2014). As a user on a social network typically has many followers, she can achieve this need for creating a good impression by being responsive to the taste of the followers. Thus, the taste of audience is likely to be dominant driver for sharing content as compared to other shared attributes with the sender such as bandwidth and social bond. This is especially true for directed networks where users can establish identity and create an impression in the presence of massive audience (McQuarrie et al. 2013).

Additionally, users' responsiveness to the taste of the audience may vary with audience type. For example, as users already know people with strong ties, they may only feel the need to impress others with whom they share weak ties (Berger 2014). Similarly, Dubois et al. (2016) show that need for self-enhancement is higher with weaker ties (strangers) as compared to stronger ties (friends). In our context, followers and mutual followers represent audience with weak and strong ties, respectively. Thus, a user is more likely to be responsive to the taste of audience represented by common followers as compared to that represented by common mutual followers.

Therefore, we posit:

H6: The propensity of a receiver to share a piece of content from a sender increases more with the number of common followers than with the number of common mutual followers.

Table 3 summarizes the drivers associated with the three network overlap metrics in directed networks. Note that the need for uniqueness as a driver should only come into play when there is an audience. Thus, the need for uniqueness is unlikely to moderate the effect of common followees, as they represent sources rather than the audience of a focal user. That different drivers are associated with the different three metrics illustrates the nuanced role of different types of network overlaps on content sharing in directed networks.

MODEL

Our objective is to evaluate the impact of network overlap on the propensity of a receiver to share content obtained from sender(s). We use a Cox proportional hazards model (Cox 1972) to estimate the hazard rate of sharing. In social networks, one challenge for a researcher is that a user may receive multiple feeds from different senders sharing the same content (or an aggregated feed from multiple senders) and the contribution of each co-sender on the decision to share is unclear.

At the consumer (receiver) level, a number of models have been proposed to deal with the impact of multiple senders (Toubia et al. 2014; Trusov et al. 2010) or multiple ad exposures (Braun and Moe 2013). A key difference between the present study and these studies is that our unit of analysis is a dyad rather than an individual. Individual level analysis often comes with some sort of aggregation on the sender side. For example, Aral et al. (2009) consider the overall effect of the number of shared friends on a user's likelihood to adopt a Facebook app, but the effect of individual friends' characteristics are not studied. Katona et al. (2011) accommodate multiple senders by considering the average characteristics of senders, which compromises model precision. While Trusov et al. (2010) do consider the effect of each individual sender on a user (restricted to be either 0 or 1), their model does not allow statistical inference on the effects of dyadic characteristics such as network overlap. Sharara et al. (2011) focus on an adaptive diffusion model with the objective of establishing the effect of network dynamics on content sharing. They learn the "confidence values" between sender-receiver pairs based on past sharing for the purpose of making predictions. However, they do not deal with the estimation of the effects of dyadic characteristics on the propensity to share content.

Experimental studies (Aral and Walker 2012; Aral and Walker 2014) which conduct dyadic level analyses, avoid this problem by eliminating receivers getting notifications from multiple senders. While it eliminates the statistical challenge of dealing with multiple senders, it creates a controlled (and at times artificial) setting where the experiment inadvertently also controls for drivers of sharing that can be important in a natural setting of content sharing. For example, the need for uniqueness is more likely to be a concern if multiple individuals in a user's social network have shared the content as compared to a single individual sharing the content. We address this challenge by proposing a novel proportional hazards model that allows us to estimate the contribution of individual senders when multiple co-senders collectively cause a decision to share content.

Dyadic Hazard

To ease model exposition, we present it in the context of sharing content generated or shared by a company over the social media platform, Twitter (as it is the context of our primary dataset). On Twitter, when a user (sender) retweets (shares) a piece of content (a tweet), her followers (receivers) are immediately notified about her sharing activity in the form of a feed. A receiver can have multiple senders (co-senders) if more than one of her followees retweets the same content.

Let i, j, and k index senders, receivers, and tweets, respectively. Let t be the time elapsed since the creation of a particular tweet. Let $X_i(t)$ and $X_j(t)$ represent the unitary attributes of sender i and receiver j, respectively (e.g., gender and activity level of a user on Twitter). Let X_{ij} represent the dyadic attributes concerning sender i and receiver j (e.g., network overlap measures), X_{ik} represent sender i's attributes that are specific to tweet k (e.g., the time sender i

retweets tweet k), and X_{jk} represent receiver j's attributes that are specific to tweet k (e.g., number of receiver j's followees that have shared tweet k). Let $\lambda_{ijk}(t)$ represents the dyadic level hazard of sender i causing receiver j to adopt tweet k at time t. Let $\lambda_{k0}(t)$ represents the baseline hazard for tweet k. The dyadic level hazard, stratified on tweets, is given by

$$\lambda_{ijk}(t) = \lambda_{k0}(t)\, exp\left(\beta_1 X_i(t) + \beta_2 X_j(t) + \beta_3 X_{ij}(t) + \beta_4 X_{ik}(t) + \beta_5 X_{jk}(t)\right), \qquad (1)$$

Note that the above semi-parametric formulation allows $\lambda_{k0}(t)$ to change arbitrarily over time and across tweets, allowing us to capture static content-specific effects such as the text of a tweet and time-varying effects such as the overall declining tendency to share a specific tweet over time. For example, $\lambda_{k0}(t) = 0$ represents a case when a tweet stops diffusing in the network. This formulation of dyadic hazard is similar to the formulations given in (Aral and Walker 2012; Aral and Walker 2014; Lu et al. 2013), but we allow one receiver to be exposed to the same tweet from multiple senders.

Note that X_{ik} and X_{jk} include variables representing when a sender shares and the number of co-senders of a receiver, respectively. Due to users' need for uniqueness in online communities, we hypothesize that the effects of common followers and common mutual followers are negatively moderated by the popularity of content in H3 and H5. To test these effects, we consider interaction of the popularity of tweets with common followers and common mutual followers and include these as dyadic attributes.

Spontaneous Sharing

The basic specification of dyadic hazard ignores the possibility of users to share spontaneously. For example, a user may share a brand-authored tweet received from another user in the social network, or after browsing the brand's homepage on Twitter. The latter type of sharing is termed as a spontaneous sharing and occurs via a non-social source (e.g., the brand homepage or an external site). In the context of Twitter, the sharing is spontaneous if a user shares a tweet before any of her followees do. Otherwise, the sharing is considered as potentially influenced by others. In order to incorporate the impact of non-social sources in our study, we treat them as a special sender and use a dummy variable to capture their effect on the hazard rate:

$$\lambda_{ijk}(t) = \lambda_{k0}(t)\, exp\big(\beta_0 s_i + \beta_1 X_i(t) + \beta_2 X_j(t) + \beta_3 X_{ij}(t) + \beta_4 X_{ik}(t) + \beta_5 X_{jk}(t)\big), \quad (2)$$

where the dummy variable s_i is 1 if the sender is the special sender and 0 otherwise. For the special sender, all undefined unitary and dyadic attributes are coded as missing and set to zero (or any other default value as the selection of default only affects parameter β_0). The parameter β_0 captures the combined effect of all non-social sources, as compared to a sender whose attributes are zero, on the sharing of the receiver. Since all users can adopt spontaneously, the special sender is a co-sender for every potential sharing user. Our dummy variable formulation enables us to seamlessly incorporate the effect of non-social sources.

Model Estimation

Let the parameter vector $\theta = \{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ represent the entire set of parameters of our model. Let $R_k(t)$ represent the set of receivers who have not shared tweet k before time t (excluding), which is often referred to as the risk set. Let $C_{jk}(t)$ represent the set of co-senders that have sent a feed regarding tweet k to receiver j before time t. Let E represent the set of sharing events observed in the data and let E_{jk} represents the event of receiver j sharing tweet k.

The key assumption of the proposed proportional hazard model is that the sharing of a receiver is collectively caused by all her co-senders, which is a common assumption in previous non-dyadic models to deal with multiple senders (Toubia et al. 2014; Trusov et al. 2010) or multiple ad exposures (Braun and Moe 2013). In a hazard model, this means that the time it takes the receiver to share is determined by the overall hazard of the receiver. Assume that the hazards of the receiver to be influenced by each co-sender are independent conditional on the control variables, the overall hazard of receiver j to share tweet k at time t is given by

$$\lambda_{jk}(t) = \sum_{i \in C_{jk}(t)} \lambda_{ijk}(t),$$

where $\lambda_{ijk}(t)$ represents the dyadic level hazard of sender i causing receiver j to share tweet k at time t. The additive form of the overall hazard results from the conditional independence assumption, which is a standard assumption for proportional hazards model.

Suppose event E_{jk} occurred at time τ_{jk}, the partial log likelihood of this event can be written as

$$l(E_{jk}|\theta) = ln\, P(E_{jk}|\theta) = ln\left(\frac{\lambda_{jk}(\tau_{jk})}{\sum_{j' \in R_k(\tau_{jk})} \lambda_{j'k}(\tau_{jk})}\right) = ln\left(\frac{\sum_{i \in C_{jk}(\tau_{jk})} \lambda_{ijk}(\tau_{jk})}{\sum_{j' \in R_k(\tau_{jk})} \sum_{i' \in C_{j'k}(\tau_{jk})} \lambda_{i'j'k}(\tau_{jk})}\right) \quad (3)$$

Note that the baseline hazard cancels out. The overall partial log likelihood of the entire dataset can then be written as

$$l(E|\theta) = \sum_{E_{jk} \in E} l(E_{jk}|\theta) = \sum_{E_{jk} \in E} ln\left(\frac{\sum_{i \in C_{jk}(\tau_{jk})} \lambda_{ijk}(\tau_{jk})}{\sum_{j' \in R_k(\tau_{jk})} \sum_{i' \in C_{j'k}(\tau_{jk})} \lambda_{i'j'k}(\tau_{jk})}\right) \quad (4)$$

The parameters in our model can be estimated by maximizing the partial log likelihood given in Equation (4) using the Newton-Raphson method or other numerical optimization methods. In this paper, we use an enhanced Newton-Raphson algorithm to search for the optimal parameters of the partial log likelihood. Specifically, when the parameters reaches a non-concave region, we add a small positive number to the diagonal elements of the information matrix (typically slightly larger than the smallest eigenvalue of the information matrix in absolute value), as suggested by Schnabel and Eskow (1999), to make the information matrix positive definite. The effectiveness of the enhanced Newton-Raphson algorithm has been validated through extensive simulations. The above model collapses to the standard proportional hazards model when there is only one sender for each receiver.

Our proposed model has two advantages over prior specifications. First, it does not speculate on the contribution of each co-sender apriori, but allows the data to automatically determine the contribution of individual co-senders based on their characteristics. Second, it is applicable even if only some of the co-senders have a significant impact on the sharing, as the likelihood in Equation (3) essentially captures the probability of the true cause belonging to the set of co-senders. Lacking information on which subset of co-senders have real effects will increase the standard errors of the parameter estimates, but will not bias the point estimates. In Web Appendix A, we show using simulations that the proposed model can recover the true parameters with negligible errors, regardless of whether the sharing events are caused by all co-senders collectively or only one of the senders. In contrast, we find that models that make assumptions on the contributions of co-senders apriori can result in substantial bias (see Table A1 in Web Appendix A).

Note that our model does not make any assumption about the direction of the effect of co-senders on the sharing propensity. This allows us to flexibly capture the saturation effect (a negative coefficient on co-senders) or the reinforcement effect (a positive coefficient on co-senders), as the number of co-senders increases.

Identification

A primary challenge for determining the impact of the network characteristics on user actions is that the results could be biased due to unobservable characteristics. For example, a sender with high popularity offline might be more influential than other senders with similar online characteristics. While such offline information might be observable to the receiver, it is often unknown to the researcher. Similarly, a receiver with stronger interest in brand-related content might be more likely to share their tweets in general, and such topical interest of individual receivers is often not available to the researcher. Missing information on either senders or receivers can bias model estimates. To address this concern, we allow for random effects at the sender-level and the receiver-level, which allow each sender and receiver to have a random intercept that captures the main effect of unobserved characteristics. Given that the special sender representing the effects of non-social sources is intrinsically different from other senders, we allow the variance of the frailty term for the special sender to be different from other senders. We also consider random effects at the dyadic level to account for dyad-specific unobservables, following previous studies in network contexts (Hoff 2005; Lu et al. 2013; Narayan and Yang 2007). Note that it is possible that the unobserved characteristics are correlated with observed characteristics. For example, a sender with high unobservable popularity may also have lot of connections and, as a result, a larger overlap with the receiver's connections as compared to a less popular sender. As random effects cannot accommodate such correlations, we also estimate models with fixed effects at the sender level (fixed effects allow unobserved characteristics to be correlated with observed characteristics).

In addition to unobserved characteristics, two additional concerns for identification are spontaneous shares and endogenous communication patterns (Aral and Walker 2014). For the former, we explicitly control for the possibility of spontaneous shares, by treating all non-social sources as a special sender. Such a control not only teases out the effect of non-social sources, but also alleviates, to some extent, the concern that a receiver is sharing due to her inherent propensity to share. For the latter, in our application, the platform sends a notification to all followers of a sender. Thus, there is no selection bias on who can see the content (i.e., no endogenous communication patterns).

A fourth problem with identifying content sharing across a dyad is that a receiver often sees the same content from multiple senders before sharing, and the quantitative contribution of each co-sender may be unclear. We address this challenge statistically by proposing a novel proportional hazards model that determines the contribution of each co-sender based on her characteristics.

DATA

We seek to understand how different types of network overlaps between a sender and a receiver connected in a social network impact the sharing behavior of the receiver. A dyadic level study imposes stringent requirements on the data. First, we need a sample of content generated or shared on a social media platform by firms. This is important as we can establish the implications of our results for firms utilizing social media to reach out to consumers. Next, for each piece of content, we need complete information regarding how the content propagates through the network from activated users (senders) to their followers (receivers). Such information includes the profile and social graph information of all involved users (both senders and receivers), as well as time-stamped sharing information at the individual user level. The sample of involved users can be identified by traversing the audience of activated users progressively. Specifically, we can start from a set of seeds (e.g., the author or users who spontaneously share the content) and then treat the followers of these seeds as receivers. This process iterates when a receiver become activated, i.e., she shares the content, until the end of the observation time window. This progressive user sampling approach based on ego's network allows us to focus on users who are relevant to our analysis. A similar approach has been employed by other researchers interested in the effects of dyadic network characteristics (Aral and Walker 2014; Shi et al. 2014). The set of users chosen by the progressive sampling approach are all the activated users (senders) and their followers (receivers). Finally, the profile and social graph information on these users can be collected retrospectively from historical data on social media platforms. Note that if there are users with regular exposures to non-social sources (e.g., a brand's homepage), we can also consider them as receivers.

We collected one dataset with the desired information from Twitter. To improve the managerial relevance of our study, we focus on the sharing of brand-authored tweets. In the context of Twitter, the act of sharing is retweeting. As noted earlier, we assume that sharing is spontaneous if a user shares a tweet before any of her followees do. Otherwise, the sharing is considered as potentially influenced by others. We focus on nine brands listed by Fortune magazine as the top fortune 500 companies using social media.[6] In response to the API restrictions, we used 40 Twitter accounts to collect our data in parallel. [7] We first collect the tweets authored (or retweeted in some cases) by each brand in a 30-day time window around April 2016.[8] Then for each tweet, we collect the social graph information needed for our analysis retrospectively in two steps. As the first step, we collected the social graph information of the author and retweeters of each tweet. These users represent the set of senders for the focal tweets. Next, we spend about 50 days to collect the social graph information for the followers (receivers) of the senders. Since the density and network size of Twitter users is very high (a Twitter user in our dataset has more than 8000 followers on average and the median number of followers is 741), collecting social graph information for all followers of every sender is extremely time-consuming.[9] In order to reduce the size of the data, we consider a smaller set of followers. For every sender, we consider all followers who retweet. From the remaining followers who do not retweet, we randomly sample 100 followers using the risk set sampling approach (Langholz and BORGAN 1995; Langholz and Goldstein 1996), if there are more than 100 non-activated followers. The risk set sampling approach can produce unbiased estimates. Finally, we collect the profile information for all the identified users.

For each potential receiver, we generate one dyadic observation for her if one of her followees shares the tweet. Since everyone can retweet a tweet spontaneously without the influence of its followees, we add an additional dyadic observation for each receiver, in which the sender is a special sender that captures the effect of the non-social source (as discussed in the Model

[6] http://fortune.com/2014/06/02/500-social-media/
[7] https://dev.twitter.com/rest/public/rate-limits
[8] Since we could not collect data on all brands concurrently in a short time interval (the network structure among users may change if they are not collected in a short time interval), we collected data on the 9 brands in three different time windows. All tweets in our sample were posted during March 14 ~ May 4, 2016.
[9] It would take us 6.2 years to collect information for all receivers using our setup due to API restrictions.

section). The act of retweeting allows the user to share the tweet with her followers. One converts from a receiver to a sender immediately after the sharing activity. Table 4 shows the summary information of the dataset. The table shows that 6.4% of shares have more than one co-sender (excluding the special sender), and the average number of co-senders is 2.12, including the special sender. This validates the need for a model that accounts for multiple co-senders.

We use several control variables pertaining to the sender, the receiver, and the sender-receiver dyad. These variables, summarized in Table 5, include the unitary network attributes of the sender/receiver, the engagement level of the sender/receiver, the demographics of the sender/receiver, the timing of the sender's share, the number of co-senders in the receiver's network, and so forth. Table 6 summarizes the summary statistics for the main unitary and dyadic network attributes and key control variables.

Table 7 outlines the correlation among dyadic network characteristics. In order to clearly identify the effects of different overlapping connections, we exclude common mutual followers when counting the number of common followees and common followers. The correlations among the three network overlap metrics are not particularly high and suggest that these metrics are capturing different drivers. Further, the estimates of the correlated variables were stable with changes in model specifications and data samples, suggesting that multicollinearity is unlikely to be an issue.

In order to understand how tweets were shared over time, we plot the Kaplan-Meier survival curve for a random subsample of tweets (see Figure B1 in the Web Appendix B). Note that the sharing activities on most tweets basically ceased in about a week. In Figure B3 (Web Appendix B), the hierarchical visualization of the network among activated users shows how content spread across users. This figure demonstrates that path length is short (around 2 on average) for content as they propagate through the user network, in agreement with the observation made by Goel et al. (2016) about short path lengths for diffusion in online social networks.

Preliminary Evidence

To evaluate the potential relationship between network overlap and propensity to share content, we compute the average network overlap of activated and non-activated sender-receiver dyads, respectively. A dyad is considered as activated if the receiver shares the content. Figure 2 shows the difference in the magnitude of each type of network overlap between activated and non-activated dyads. Activated dyads have higher network overlap than non-activated dyads. This suggests that higher network overlaps in terms of common followees, common followers and common mutual followers can be associated with higher propensity to share. In order to determine the effect of popularity of content, we further divide activated dyads into two groups based on whether they were activated when the popularity of the tweet is above or below the average popularity of all tweets. Figure 3 shows the number of average common followers associated with dyads activated at high popularity is lower as compared to that associated with dyads activated at low popularity. This figure suggests that relatively fewer dyads with high number of common followers get activated when the tweet popularity is high, implying that popularity may negatively moderate the effect of common followers on the propensity to share. Common mutuals show a similar pattern except that the difference in the average number of common mutuals for activated dyads at high and low tweet popularity is relatively small. To infer the true effects of network overlap, we also have to control for confounding factors that may affect both network overlap and propensity to share. We achieve this by estimating our proportional hazards model discussed earlier. Next, we discuss our model results on the role of network overlap on the sharing propensity of the receiver.

RESULTS

Main Results

Table 8 summarizes the results of four model specifications. Our main model of interest is model 4 that includes interaction terms representing the moderating effect of tweet popularity on common followers and common mutual followers. We also estimate models with no interaction terms or including only one of the two interaction terms (Models 1-3, respectively). Likelihood ratio tests suggest that model 4 is preferred over models 2 and 3 (p < 1e-3). The following discussion is based on the estimates from model 4 unless otherwise specified.

Common followees. The number of common followees has a positive effect on the sharing propensity of the receiver. This result validates H1. Higher the number of common followees between a sender and a receiver, higher is the similarity in their interests and knowledge. Thus, the more common followees the receiver has with the sender, the more likely the receiver is knowledgeable about or interested in the sender's content and more likely she will share the content from the sender in order to impress others. Note that we obtain this result after controlling for the effect of common mutual followers, which represent close friends. Thus, our result suggests that common followees can also be used to capture similarity or homophily between users (McPherson et al. 2001).

Common followers. The simple effect of common followers (when the logarithm of the content popularity is zero) is positive, suggesting that the number of common followers has a positive effect on dyadic influence when the popularity of a tweet is low. This finding validates H2. As more common followers reflects higher similarity between the audiences of the sender and the receiver, the receiver is likely to make the same decision as the sender (i.e., to share), especially when the content is relatively novel and the concern around uniqueness is not strong. The negative interaction of common followers with content popularity confirms H3: the effect of common followers decreases with content popularity, validating users' need for uniqueness in content sharing (Ho and Dempsey 2010). This is similar to extant finding that consumers with a high need for uniqueness may decrease the consumption of a product if it becomes commonplace, also known as the reverse-bandwagon effect (Cheema and Kaikati 2010; Granovetter and Soong 1986).

Common mutual followers. The simple effect of common mutual followers (when the logarithm of content popularity is zero) is positive and demonstrates that, when the content is relatively novel, common mutual followers has a positive impact on sharing. This finding validates H4. A high number of common mutual followers between a sender and a receiver represents higher similarity in their interests and the taste of their audiences. In addition, such overlap suggests stronger social bonding, as well as higher bandwidth to better evaluate each other's content (Aral and Van Alstyne 2011; Burt 2001). Thus, a receiver is more likely to find content from a sender with high common mutual followers useful and is more likely to share it.

The negative interaction of common mutual followers with popularity confirms H5. This finding shows a boundary condition for the positive effect of common friends (common mutual followers) previously reported in undirected networks (Aral and Walker 2014; Bapna et al. 2016). Specifically, the effect of common friends is positive only when the information to be communicated is relatively novel (or not as popular).

Our results show that the effect of network overlap in directed networks varies across different types of “connections”. Moreover, the impact of common followers and common mutual followers are negatively moderated by content novelty. The negative interaction effects suggest that users are eventually going to cease sharing due to concerns around uniqueness. As a result, the content is likely to diffuse for short distances within a network. This may explain the short information cascades reported in literature (Goel et al. 2016) and also observed in our dataset (Figures B1 and B3). The negative interaction effects also confirm that common followers and common mutual followers do represent characteristics such as the similarities in audiences with weak and strong ties and are not just redundant measures representing homophily.

Common mutual followers vs. Common followers. The coefficient of common followers is higher than the coefficient of common mutual followers (Table 8). Wald test suggests that the difference is significant (see Table C2 in Web Appendix C). Note that model 4 shows the simple effect of common followers and common mutual followers (when the logarithm of the content popularity is zero). The difference between the two coefficients is also significant in Model 1, which captures the effects of common followers and common mutual followers averaged across all levels of content popularity. In addition, we employ an alternate model where we constrain the coefficients of common followers and common mutual followers to be the same. We find that our current model specification provides a much better fit than this alternate model (the difference in BIC is larger than 100). These results confirm that the difference between the coefficients of common followers and common mutual followers is positive and significant. Thus, the propensity to share increases more with common followers than that with common mutual followers. In sum, H6 is supported.

A possible explanation is that users pay much more attention to the taste of audience with weaker ties (followers) than that for audience with stronger ties (mutual followers). As a result, they tend to share content from the sender with whom they share more common followers. That different types of followers have differential effects confirms the importance of considering the *directionality* of connections in social networks. Our results show that in targeting users for content propagation, it is better to select users who share common followers with their audience as compared to users who share common mutual followers with their audience.

In addition to the findings on the three network overlap measures, it is worthwhile highlighting the estimates on two additional variables (i.e., co-senders and shareTime), which help us understand how each co-sender contributes to a receiver's propensity to share. First, the effect of co-senders is negative, showing that the marginal effect of a co-sender on content sharing decreases with the number of co-senders (though the overall effect of all senders may increase). This echoes a previous finding on how multiple friends affect the sharing of URLs on Facebook (Bakshy et al. 2012). Second, the effect of shareTime is positive[10], suggesting that the later a co-sender shared, the stronger effect the co-sender has on the receiver. This pattern documents a recency effect for co-senders, consistent with previous findings that social effects decay over time (Haenlein 2013; Nitzan and Libai 2011; Trusov et al. 2009).

Robustness Checks

Unobserved Characteristics. A potential concern with our analysis is that the sharing of content could be driven by unobserved characteristics at the sender, the receiver, and even the

[10] It can be easily shown that, in a proportional hazards model, using shareTime (i.e., how long did it take for a sender to adopt) is equivalent to using recency (i.e., how long ago did the sender adopt), because the sum of the two variables equals the time elapsed since the creation of the ad. The only difference is that the estimates on both variables will have opposite signs. We use shareTime as it does not vary over time, which facilitates the estimation.

dyad level. The dyadic observations with the same sender, receiver or dyad may not be independent because of common unobserved characteristics. In our main analysis, we consider sender-specific, receiver-specific and dyad-specific random effects. As a robustness check, we also account for the effects of unobserved characteristics with a fixed effects approach as it allows for unobserved characteristics to be correlated with observed characteristics. While the fixed effects approach appears to be more flexible than the random effects model in terms of its assumptions, it is more sensitive to the issue of insufficient reoccurrence. Specifically, in the proportional hazards modeling framework, a random effects approach tends to provide more reliable estimates than the fixed effects approach as the former penalizes large individual effects (Therneau 2000) and prevents the model from over-fitting. With that being said, we still estimate fixed effects on the sender level but not on the receiver-level as the low reoccurrence frequency of receivers in our data may result in substantial incidental parameter bias in the estimates (Allison 2002; Lancaster 2000). Fixed effects on the dyadic level are not a viable alternative as well, as then the effects of dyadic network characteristics are not identified. Note that the random/fixed effects allow us to account for unobserved factors such as the fact that some users might be bots on Twitter.

Table 9 presents the results from different models with random and fixed effects at sender, receiver and dyad levels. Overall, the estimates on the dyadic network characteristics are qualitatively similar across different model specifications.

Validation with Additional Datasets

To test whether our findings generalize to other types of content and other directed networks, we collect two additional datasets. In the first additional dataset, we focus on tweets posted by regular Twitter users (instead of brands) which are related to three different topics: Apple (Technology), NBA (Sports), and Election (Politics). We chose these three topics randomly from a set of trending topics on Twitter. We collected 2,081 such tweets using the Streaming API of Twitter during May 21~30, 2016. These tweets are retweeted 18,873 times in total. Then, following the same approach to collecting the main dataset, we collect the profile and social graph information of all involved senders and receivers. To collect the dataset within a reasonable time frame, we only sampled 10 non-activated followers for each sender. Web Appendix E shows detailed statistics for this dataset. Table 10 shows the main results for this additional Twitter dataset.[11] The results on the two Twitter datasets are highly consistent, demonstrating that our findings apply not only to brand-authored tweets, but also to general tweets.

The second additional dataset is collected from Digg.com, a large online social news aggregation website. At the time we collected the data, Digg maintained an internal Twitter like social network structure. Users can highlight ("digg") their favorite content and the activity is visible to all of their followers. Digg introduced a native advertising model, called diggable ads, in 2009. Both the internal social network and native advertising model remained on the website until Digg's acquisition in August 2012. The feature allowed an advertiser to promote sponsored content in the feeds of Digg users. We focus on the digging (sharing) activities of 31 ads in a month-long period. Since, Digg network is much smaller as compared to Twitter, we were able to collect social graph information for all involved senders and receivers. Web Appendix F provides more details regarding the collection of this dataset. Web Appendix G shows detailed

[11] Due to small number of sampled receivers, we cannot reliably estimate our model with random/fixed effects for this dataset. However, analyses on our main dataset show that results without such effects are directionally similar.

statistics for this dataset. We estimate a model with all three levels of random effects, identical to what we do for the main Twitter dataset, and find a similar pattern of results (see Table 11 and Table C2 in Web Appendix C).

MANAGERIAL RELEVANCE OF THE RESULTS

Marketers can use network overlap to better target users for content propagation. For example, Twitter allows advertisers to choose audience based on gender, device, interests, and even any list of users provided by advertisers (see Figures D1 and D2 in Web Appendix D). Thus, it is possible to target individual users based on their network overlap with their followers.

To quantify how network overlap can facilitate the sharing of content, we develop a simulation study that uses the social network structure in the main Twitter dataset together with the estimated model parameters. We assume all senders (excluding brands) are activated and then simulate how long it takes the senders to activate 1% of all receivers, both with and without considering network overlap. Since we did not estimate the baseline hazard function in our main model, we have to make a parametric assumption for the baseline hazard function to simulate survival times for receivers. Without loss of generality, we assume the baseline hazard function follows a Weibull distribution with shape parameter k. Note that if $k = 1$, the baseline hazard function remains constant over time. The baseline hazard decreases over time if $k < 1$ and increases over time if $k > 1$. Based on the survival curves shown in Figure B1, the latter case is unlikely, but we still consider it for the purpose of completeness. Figure 4 summarizes the contribution of network overlap for accelerating content sharing under different scenarios. As compared to not considering network overlap (equivalent to assuming all coefficients related to network overlap variables are zero), accounting for network overlap saves about 35~70% of time to activate 1% of receivers across a wide range of the shape parameter (k). The amount of time saved is similar for activating 5% and 10% of receivers. We also tried to parametrize the baseline hazard function with a Gompertz distribution and the results are similar.

To better describe the marginal effect of network overlap, we artificially increase the network overlap between a dyad by a certain percentage and determine how much does it lowers the time for activation. Figure 5 shows how the percentage of time saved increases with the increase in network overlap. For instance, a 20% increase in network overlap can reduce the activation time of a dyad by about 13%. These simulation results demonstrate the value of network overlap in increasing the speed of content propagation in social networks.

Our model results also indicate that the effect of network overlap varies with the popularity of content. To illustrate how this finding can influence the selection of seeders, we choose two sets of seeders having high and low network overlap with their followers, respectively, and then compare the time it takes to activate 1% of all their receivers. To obtain seeders with high and low network overlap, we first select the top 200 senders with the largest number of followers in the main Twitter dataset. Next, we calculate the average number of common followers each user shares with her followers. Using the median of the average numbers for the 200 users as a threshold, we assign these users to high and low network overlap groups. The 100 users in the high network overlap (i.e., common followers) set have 40,367 followers on average, whereas the 100 users in the low network overlap set have 118,324 followers on average. See Table C3 in Web Appendix C for more details.

To facilitate the comparison, we compute the ratio of the time taken to activate 1% receivers by the high and low network overlap seeders. We repeat the computation at different levels of content popularity. Figure 6 summarizes the results. When content popularity is low, the ratio is less than one, indicating that the high overlap seeders activate 1% receivers faster than the low network overlap seeders, even though the high overlap seeders have much less followers on average. However, when the content popularity is high, the low network overlap seeders activate 1% receivers faster. These results illustrate that seeders with high network overlap should be selected only when the content is not popular.

In practice, the popularity of content often varies across brands. For example, a tweet posted by Starbucks is usually much more popular than one posted by Allstate. Our finding suggests that different brands may want to target different sets of seeders. Specifically, it should be more effective for popular brands (e.g., Starbucks) to target users with lower network overlap (with their followers). In contrast, less popular brands (e.g., Allstate) struggling for engagements may want to choose high network overlap users as seeders.

DISCUSSION & CONCLUSION

Social media platforms hold the potential to reshape the manner in which consumers generate, spread and consume content. Understanding what leads to effective content sharing at the dyadic level lies at the core of cost-effective content propagation on these platforms. While the effects of unitary network attributes have been well-studied in the literature, studies on the effects of dyadic network attributes on content sharing are nascent.

In this paper, we study the effect of a dyad's network overlap on content sharing in directed networks. More specifically, we quantify the effects of common followees, common followers, and common mutual followers between a sender and a receiver on the propensity of sharing by the receiver. Substantively, our results show that the effect of network overlap in directed networks varies across different types of "connections". The number of common followees is positively associated with receiver's propensity of sharing. Other network overlap measures such as number of common followers and common mutual followers also have positive effect on this propensity. However, the latter positive effect decreases with the popularity of the shared content. Thus, our study provides insight into consumer behavior in online content sharing. Previous work has focused on role of unitary characteristics of senders and receivers and dyadic characteristics such as reciprocity in content sharing. We extend previous work by demonstrating how shared network characteristics such as network overlap can also inform the sharing propensity at a dyadic level. Further, we document the moderating role of content popularity on the effect of network overlap on sharing. In doing so, we add to the existing literature by highlighting the role of uniqueness in social consumption (Cheema and Kaikati 2010; Zeng and Wei 2013). More specifically, we show how the uniqueness concern is revealed through differential responses to network overlap measures with the increase of content popularity. Finally, we demonstrate the importance of directionality of connections by documenting the difference in the sharing propensity depending on the type of network overlap. Specifically, we show that sharing propensity is more likely to increase with common followers as compared to common mutual followers. To the best of our knowledge, ours is the first study to document differential responses of users on social networks based on the directionality of their shared connections or network overlap.

Our paper makes a methodological contribution as well by proposing a new hazard rate modeling approach to more accurately determine the contribution of individual senders on influencing a receiver when multiple senders are involved. Quite often, consumers may respond only after the content is seeded by multiple senders (Centola and Macy 2007). Even if detailed tracking information is available for each user, it would be difficult to determine the exact contribution of each sender in the content sharing process.[12] Previous work either makes strong assumptions about how the contribution should be attributed to different senders (Aral et al. 2009; Braun and Moe 2013; Katona et al. 2011; Toubia et al. 2014) or does not focus on the identification of the effect of shared characteristics (Sharara et al. 2011; Trusov et al. 2010). Our approach makes no such assumptions and, as a consequence, can help better tease apart the effects of shared network attributes.

For marketing managers, we provide insights on how to target customers in a directed network at a micro level. Many platforms support micro level targeting to improve the efficacy of targeting (e.g., display of promoted tweets on Twitter) and prevent information overload for their members (e.g., filtering of feeds on Weibo). Our results show that platforms such as Twitter or Weibo can improve their targeting or filtering by focusing on dyads embedded in different types of connections (i.e., followees, followers, mutual followers). As a concrete example, when deciding whether or not to show a promoted tweet to a given user[13], Twitter may want to consider how many common connections this user shares with the author, as well as the overall popularity of the tweet. Specifically, targeting users who have more common followees with the

[12] While a platform can track the actual time when a receiver sees content from one or more senders and the sequence in which the content is received, it cannot determine how consumer is weighing these different feeds in her decision to adopt the content and in turn send it to her followers.
[13] Once a tweet is promoted, Twitter can display the tweet to any user on the platform, even though this user doesn’t follow the author of the tweet. However, in practice, to avoid spamming users, Twitter only displays promoted tweets to selective users deemed relevant. Note that an advertiser can promote a tweet authored by a random user.

author can be more effective. Targeting users who have large numbers of common followers and common mutual followers can also be effective when the tweet is not that popular, but might be counterproductive when the tweet is already sufficiently popular. Finally, as compared to most previous studies that primarily focus on the sharing of organic content in social networks, the analysis of this paper is based on the sharing of brand-authored tweets, which makes our findings of direct relevance to marketers.

Our work can be extended in several ways. First, it is likely that characteristics of the content can influence how much it is shared within dyads (Berger and Milkman 2012). Our modeling framework allows us to account for the heterogeneity of content but it would be useful to understand if the magnitude or direction of our results is sensitive to the type of content being shared. Second, from a modeling standpoint, we did not have information on whether or not a user actually saw the feed. Without the impression information, we are essentially modeling the overall hazard of a user to read and adopt an ad. This coarse modeling structure may increase the standard errors of our estimates. However, the impression information is typically only known to social media platforms. Future research should explore alternative approaches to address the lack of impressions such as conducting experiments where such information can be obtained from users (De Bruyn and Lilien 2008) or developing a multi-stage model to capture the effect of impressions (Shi et al. 2014). Finally, the conditional independence of co-senders' hazards assumes that the existence of one co-sender does not cannibalize or reinforce the effects of other co-senders. In our analysis, we relax this assumption by allowing the hazard of a co-sender to change with the number of co-senders. The negative coefficient on shared followees suggests that the marginal effect of a co-sender decreases with the number of co-senders (i.e., the cannibalization effect exists). However, this remedy strategy may not be satisfactory if the hazards of individual co-senders change by different multiplicative scales as the number of co-senders increase. Future studies should explore the non-linear effect of the number of co-senders on the outcome.

REFERENCES

Alexandrov, A., Lilly, B., and Babakus, E. 2013. "The Effects of Social- and Self-Motives on the Intentions to Share Positive and Negative Word of Mouth," *Journal of the Academy of Marketing Science* (41:5), pp. 531-546.

Allison, P. 2002. "Bias in Fixed-Effects Cox Regression with Dummy Variables." Available at: http://www.statisticalhorizons.com/wp-content/uploads/2012/01/BiasFE-Cox.pdf.

Aral, S., Muchnik, L., and Sundararajan, A. 2009. "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences* (106:51), pp. 21544-21549.

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

Aral, S., and Walker, D. 2012. "Identifying Influential and Susceptible Members of Social Networks," *Science* (337:6092), pp. 337-341.

Aral, S., and Walker, D. 2014. "Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment," *Management Science* (60:6), pp. 1352 - 1370.

Bakshy, E., Hofman, J.M., Mason, W.A., and Watts, D.J. 2011. "Everyone's an Influencer: Quantifying Influence on Twitter," *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 65-74.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. 2012. "The Role of Social Networks in Information Diffusion," *Proceedings of the 21st International Conference on World Wide Web*, pp. 519-528.

Bapna, R., Gupta, A., Rice, S., and Sundararajan, A. 2016. "Trust and the Strength of Ties in Online Social Networks: An Exploratory Field Experiment," *MIS Quarterly* (forthcoming).

Bapna, R., and Umyarov, A. 2015. "Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks," *Management Science* (61:8), pp. 1902–1920.

Barasch A., and Berger, J. 2014. " Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research* (51:3), pp. 286-299.

Baumeister, R.F., and Leary, M.R. 1995. "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation," *Psychological Bulletin* (117:3), p. 497.

Berger, J. 2014. "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology* (24:4), pp. 586-607.

Berger, J., and Milkman, K.L. 2012. "What Makes Online Content Viral?," *Journal of Marketing Research* (49:2), pp. 192-205.

Braun, M., and Moe, W.W. 2013. "Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories," *Marketing Science* (32:5), pp. 753-767.
Burt, D.R. 2001. "Bandwidth and Echo: Trust, Information, and Gossip in Social Networks," in *Networks and Markets: Contributions from Economics and Sociology*. Russell Sage Foundation.
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[1]," *American Journal of Sociology* (113:3), pp. 702-734.
Cheema, A., and Kaikati, A.M. 2010. "The Effect of Need for Uniqueness on Word of Mouth," *Journal of Marketing Research* (47:3), pp. 553-563.
Childers, T.L., and Rao, A.R. 1992. "The Influence of Familial and Peer-Based Reference Groups on Consumer Decisions," *Journal of Consumer Research* (19:2), pp. 198-211.
Chung, C.M., and Darke, P.R. 2006. "The Consumer as Advocate: Self-Relevance, Culture, and Word-of-Mouth," *Marketing Letters* (17:4), pp. 269-279.
Cox, D.R. 1972. "Regression Models and Life Tables," *Journal of the Royal Statistical Society. Series B* (34:2), pp. 187-220.
De Bruyn, A., and Lilien, G.L. 2008. "A Multi-Stage Model of Word-of-Mouth Influence through Viral Marketing," *International Journal of Research in Marketing* (25:3), pp. 151-163.
Dubois, D., Bonezzi, A., and De Angelis, M. 2016. "Sharing with Friends Versus Strangers: How Interpersonal Closeness Influences Word-of-Mouth Valence," *Journal of Marketing Research* (53:5), pp. 712-727.
Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets*. Cambridge University Press.
Goel, S., Anderson, A., Hofman, J., and Watts, D.J. 2016. "The Structural Virality of Online Diffusion," *Management Science* (62:1), pp. 180-196.
Granovetter, M., and Soong, R. 1986. "Threshold Models of Interpersonal Effects in Consumer Demand," *Journal of Economic Behavior & Organization* (7:1), pp. 83-99.
Haenlein, M. 2013. "Social Interactions in Customer Churn Decisions: The Impact of Relationship Directionality," *International Journal of Research in Marketing* (30:3), pp. 236-248.
Ho, J.Y., and Dempsey, M. 2010. "Viral Marketing: Motivations to Forward Online Content," *Journal of Business Research* (63:9), pp. 1000-1006.
Hoff, P.D. 2005. "Bilinear Mixed-Effects Models for Dyadic Data," *Journal of the American Statistical Association* (100:469), pp. 286-295.
Iyengar, R., Van den Bulte, C., and Valente, T.W. 2011. "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science* (30:2), pp. 195-212.
Katona, Z., Zubcsek, P.P., and Sarvary, M. 2011. "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research* (48:3), pp. 425-443.
Lambrecht, A., Tucker, C.E., and Wiertz, C. 2015. "Advertising to Early Trend Propagators? Evidence from Twitter." Available at SSRN: https://ssrn.com/abstract=2419743.
Lancaster, T. 2000. "The Incidental Parameter Problem since 1948," *Journal of Econometrics* (95:2), pp. 391-413.

Langholz, B., and BORGAN, Ø.R. 1995. "Counter-Matching: A Stratified Nested Case-Control Sampling Method," *Biometrika* (82:1), pp. 69-79.
Langholz, B., and Goldstein, L. 1996. "Risk Set Sampling in Epidemiologic Cohort Studies," *Statistical Science* (11:1), pp. 35-53.
Lee, D., Hosanagar, K., and Nair, H. 2017. "Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook." Available at SSRN: https://ssrn.com/abstract=2290802.
Lovett, M.J., Peres, R., and Shachar, R. 2013. "On Brands and Word of Mouth," *Journal of Marketing Research* (50:4), pp. 427-444.
Lu, Y., Jerath, K., and Singh, P.V. 2013. "The Emergence of Opinion Leaders in a Networked Online Community: A Dyadic Model with Time Dynamics and a Heuristic for Fast Estimation," *Management Science* (59:8), pp. 1783-1799.
Ludford, P.J., Cosley, D., Frankowski, D., and Terveen, L. 2004. "Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 631-638.
Luo, Z., Osborne, M., Tang, J., and Wang, T. 2013. "Who Will Retweet Me?: Finding Retweeters in Twitter," *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 869-872.
McPherson, M., Smith-Lovin, L., and Cook, J.M. 2001. "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* (27), pp. 415-444.
McQuarrie, E.F., Miller, J., and Phillips, B.J. 2013. "The Megaphone Effect: Taste and Audience in Fashion Blogging," *Journal of Consumer Research* (40:1), pp. 136-158.
Narayan, V., and Yang, S. 2007. "Modeling the Formation of Dyadic Relationships between Consumers in Online Communities." Available at SSRN: http://ssrn.com/abstract=1027982.
Nitzan, I., and Libai, B. 2011. "Social Effects on Customer Retention," *Journal of Marketing* (75:6), pp. 24-38.
Packard, G., and Wooten, D.B. 2013. "Compensatory Knowledge Signaling in Consumer Word-of-Mouth," *Journal of Consumer Psychology* (23:4), pp. 434–450.
Rand, W., and Rust, R.T. 2011. "Agent-Based Modeling in Marketing: Guidelines for Rigor," *International Journal of Research in Marketing* (28:3), pp. 181-193.
Reagans, R., and McEvily, B. 2003. "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range," *Administrative Science Quarterly* (48:2), pp. 240-267.
Schnabel, R.B., and Eskow, E. 1999. "A Revised Modified Cholesky Factorization Algorithm," *SIAM Journal on Optimization* (9:4), pp. 1135-1148.
Schweidel, D.A., and Moe, W.W. 2014. "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research* (51:4), pp. 387-402.
Sharara, H., Rand, W., and Getoor, L. 2011. "Differential Adaptive Diffusion: Understanding Diversity and Learning Whom to Trust in Viral Marketing," *Proceedings of Fifth International AAAI Conference on Weblogs and Social Media*, pp. 345-352.
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.
Shriver, S.K., Nair, H.S., and Hofstetter, R. 2013. "Social Ties and User-Generated Content: Evidence from an Online Social Network," *Management Science* (59:6), pp. 1425-1443.
Snyder, C.R., and Fromkin, H.L. 1980. *Uniqueness: The Human Pursuit of Difference*. Plenum Press New York.

Stephen, A.T., and Toubia, O. 2010. "Deriving Value from Social Commerce Networks," *Journal of Marketing Research* (47:2), pp. 215-228.
Suh, B., Hong, L., Pirolli, P., and Chi, E.H. 2010. "Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," *Proceedings of 2010 IEEE Second International Conference on Social Computing*, pp. 177-184.
Susarla, A., Oh, J.-H., and Tan, Y. 2012. "Social Networks and the Diffusion of User-Generated Content: Evidence from Youtube," *Information Systems Research* (23:1), pp. 23-41.
Therneau, T.M. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer.
Tian, K.T., Bearden, W.O., and Hunter, G.L. 2001. "Consumers' Need for Uniqueness: Scale Development and Validation," *Journal of Consumer Research* (28:1), pp. 50-66.
Toubia, O., Goldenberg, J., and Garcia, R. 2014. "Improving Penetration Forecasts Using Social Interactions Data," *Management Science* (60:12), pp. 3049-3066.
Trusov, M., Bodapati, A.V., and Bucklin, R.E. 2010. "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research* (47:4), pp. 643-658.
Trusov, M., Bucklin, R.E., and Pauwels, K. 2009. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing* (73:5), pp. 90-102.
Van den Bulte, C., and Wuyts, S.H.K. 2007. *Social Networks in Marketing*. Marketing Science Institute.
Yoganarasimhan, H. 2012. "Impact of Social Network Structure on Content Propagation: A Study Using Youtube Data," *Quantitative Marketing and Economics* (10:1), pp. 111-150.
Zeng, X., and Wei, L. 2013. "Social Ties and User Content Generation: Evidence from Flickr," *Information Systems Research* (24:1), pp. 71-87.
Zhang, Y., Moe, W.W., and Schweidel, D.A. 2016. "Modeling the Role of Message Content and Influencers in Social Media Rebroadcasting," *International Journal of Research in Marketing* (forthcoming).

TABLES

Table 1. Glossary

Glossary	Description
Connections	
Friend	A user mutually connected with the focal user (undirected networks)
Followee	A user followed by the focal user (directed networks)
Follower	A user following the focal user (directed networks)
Mutual follower	A user following and followed by the focal user (directed networks)
Network overlap	
Common friend	A user mutually connected to both the sender and the receiver (undirected networks)
Common followee	A user followed by both the sender and the receiver (directed networks)
Common follower	A user following both the sender and the receiver (directed networks)
Common mutual follower	A user following and followed by both the sender and the receiver (directed networks)
Others	
Share	Retweet a tweet or digg an ad
Feed	Information notifying a user about the sharing activity of one's followees
Co-senders	The set of followees of the focal user who have already shared the tweet/ad

Table 2. Literature on the Role of Network Characteristics on User Actions

	Existing Work
Unitary Network Characteristics of Senders	Impact of sender characteristics on content propagation (Bakshy et al. 2011; Suh et al. 2010; Susarala et al. 2012; Yoganarasimhan 2012) Impact on network ties on a sender's propensity to generate content (Shriver et al. 2013)
Unitary Network Characteristics of Receivers	Impact of receiver characteristics on adoption: website registration (Katona et al. 2011); music site subscription (Bapna and Umyarov 2015); drug adoption (Iyengar et al. 2011); agent-based model (Rand and Rust 2011) Impact of receiver characteristics on churn (Hanlein 2013; Nitzan and Libai 2011) Impact of receiver characteristics on content propagation (Luo et al. 2013)
Dyadic Network Characteristics	Impact of network diversity on the novelty of information received by a user (Aral and Van Alstyne 2011) Impact of reciprocity on a receiver's propensity to share content (Shi et al. 2014) Impact of common friends on app adoption (Aral and Walker 2014) Impact of common followees, common followers and common mutual followers on a receiver's propensity to share content **(Present Study)**

Table 3. Drivers Associated with the Three Network Overlap Metrics

Network Overlap Metric	**Positive Driver**	**Negative Driver**
Common followees	Identity Signaling	
Common followers	Taste of Audience, Identity Signaling	Need for Uniqueness
Common mutual followers	Bandwidth, Social Bonding, Taste of Audience, Identity Signaling	Need for Uniqueness

Table 4. Summary Statistics

Number of tweets	397
Number of senders	12,565
Number of receivers	869,899
Number of <sender, receiver, tweet> tuples	2,972,026
Number of spontaneous tuples	1,483,729 (50%)
Number of social tuples	1,488,297 (50%)
Number of observations after accounting for time-varying variables	949,480,746
Number of shares (retweets)	18,493
Number of spontaneous shares	3,695 (20%)
Number of potential influenced shares	14,798 (80%)
Percentage with more than one co-senders (excluding special sender)	6.4%

Table 5. Descriptions of Independent Variables

Independent Variable		**Description**
$\boldsymbol{X_i}/\boldsymbol{X_j}$		**Attributes of sender *i* / receiver *j***
Network attributes	followees	Number of followees (out-degree)
	followers	Number of followers (in-degree)
	mutuals	Number of mutual followers
	lists	Number of lists subscribed
Engagement levels	statuses	Total number of tweets, including retweets
	favorites	Total number of favorites
Others	verified	Whether the Twitter account is verified
	regMon	How many months have the user been registered on Twitter
	isSocial ($\boldsymbol{s_i}$)	1 if sender $\boldsymbol{i}$ is a social source (i.e., followee), otherwise 0
	isAuthor	1 if the sender is the author of the tweet, otherwise 0
$\boldsymbol{X_{ij}}$		**Attributes of a sender-receiver dyad**
Dyadic network attributes	isMutual	Do the sender and the receiver follow each other mutually
	commonFollowees	Number of followees shared by the sender and the receiver
	commonFollowers	Number of followers shared by the sender and the receiver
	commonMutuals	Number of mutual followers shared by the sender and the receiver
$\boldsymbol{X_{ik}}$		**Sender-specific attributes of a tweet**
Sharing timing	wday	Day of a week when sender *i* retweeted tweet $\boldsymbol{k}$
	hour	Hour of a day when sender *i* retweeted tweet $\boldsymbol{k}$
	shareTime	Hours taken for sender $\boldsymbol{i}$ to retweet since the creation of tweet $\boldsymbol{k}$
$\boldsymbol{X_{jk}}$		**Receiver-specific attributes of a tweet**
	co-senders	Number of followees (co-senders) of the receiver who have already shared
$\boldsymbol{X_k}$		**Attributes of ads *k* (only interaction with other variables can be identified)**
	popularity	Number of retweets at a given time point

Table 6. Key Statistics of Main Variables

	Zeros	Mean	SD	Min	Median	Max
Unitary Network Attributes of All Users						
Number of followees	5,610	4,553.0	24,351.6	0	741	4,651,052
Number of followers	8,268	8,049.2	174,775.8	0	363	79,317,306
Number of mutuals	17,910	2,967.0	18,544.3	0	200	1,720,546
Dyadic Network Attributes of Sender-Receiver Dyads						
isMutual (1– reciprocal, 0 – non-reciprocal)	612,468	0.42	0.49	0	0	1
Number of common followees	189,767	25.7	79.1	0	6	3,947
Number of common followers	359,837	12.4	292.4	0	2	193,179
Number of common mutual followers	469,259	19.5	111.7	0	1	15,520
Popularity of Content						
Number of retweets	0	46.6	114.2	1	13	795

Table 7. Correlation among Dyadic Network Characteristics

	isMutual	logCommonFollowees	logCommonFollowers	logCommonMutuals
isMutual	1.00	0.17	0.30	0.51
logCommonFollowees		1.00	0.52	0.51
logCommonFollowers			1.00	0.74
logCommonMutuals				1.00

Table 8. Parameters Estimates of Different Model Specifications

	Model1	**Model2**	**Model3**	**Model4**
Network Overlap				
logCommonFollowees	0.072***	0.081***	0.087***	0.089***
logCommonFollowers	0.116***	0.239***	0.113***	0.230***
logCommonMutuals	0.063***	0.063***	0.224***	0.079*
Interactions with Popularity				
logCommonFollowers:logPopularity		-0.032***		-0.029***
logCommonMutuals:logPopularity			-0.052***	-0.017*
Fitness				
logLikelihood	-100246	-100123	-100136	-100113
BIC	200855	200620	200646	200610

Significance levels: p<0.001 (***), p<0.01 (**),p<0.05 (*), and p<0.1 (.). The main effect of logPopularity cannot be identified as everyone sees the same retweet number at a given time point, the effect of which is cancelled out in the likelihood. Model 4 is chosen as our main model based on fitness. We omit the coefficients on control variables for clarity. Please see Table C1 in Web Appendix C for the complete set of parameter estimates.

Table 9. Parameters Estimates from Different Random/Fixed/Mixed Effects Models

	none	rs	fs	rs-rr	fs-rr	rs-rr-rd	fs-rr-rd
Network Overlap							
logCommonFollowees	0.057***	0.057***	0.052***	0.096***	0.092***	0.089***	0.094***
logCommonFollowers	0.364***	0.369***	0.370***	0.218***	0.210***	0.230***	0.200***
logCommonMutuals	-0.072.	0.017	-0.001	0.092*	0.098**	0.079*	0.102**
Interactions with Popularity							
logCommonFollowers:logPopularity	-0.034***	-0.031***	-0.031***	-0.026***	-0.027***	-0.029***	-0.026***
logCommonMutuals:logPopularity	0.008	-0.011	-0.006	-0.023**	-0.019*	-0.017*	-0.020*
Fitness							
Log Likelihood	-112004	-110917	-110342	-100386	-99476	-100113	-99409
BIC	224351	222198	338193	201145	316471	200610	316347

In row 1, the first letter represents whether fixed (f) or random (r) effects is used. The second letter indicates the subject ("s" for sender, "r" for receiver, and "d" for dyad) on which the specified effect is applied. For example, "rs" represents a model with random effects on sender, and "fs-rr-rd" represents a model with fixed effects on sender, random effects on receiver, and random effects on dyad. "rs-rr-rd" is the main model used in this paper. The model "none" doesn't include random or fixed effects on any subject.

Table 10. Parameter Estimates on Additional Twitter Dataset

	Model1	Model2	Model3	Model4
Network Overlap				
logCommonFollowees	0.203***	0.207***	0.204***	0.206***
logCommonFollowers	0.114***	0.344***	0.116***	0.225***
logCommonMutuals	-0.013	-0.027*	0.237***	0.139***
Interactions with Popularity				
logCommonFollowers:logPopularity		-0.104***		-0.049***
logCommonMutuals:logPopularity			-0.120***	-0.076***
Fitness				
Log Likelihood	-65778	-65485	-65462	-65436
BIC	131880	131305	131258	131217

Table 11. Parameters Estimates on Digg Dataset

	Model1	Model2	Model3	Model4
Network Overlap				
logCommonFollowees	0.192.	0.212**	0.073	0.191**
logCommonFollowers	1.125***	2.287***	0.154**	1.520***
logCommonMutuals	-0.079	0.034	2.739***	0.768***
Interactions with Popularity				
logCommonFollowers:logPopularity		-0.517***		-0.339***
logCommonMutuals:logPopularity			-0.589***	-0.171***
Fitness				
Log Likelihood	-20116	-19753	-19693	-19675
BIC	40549	39832	39712	39683

FIGURES

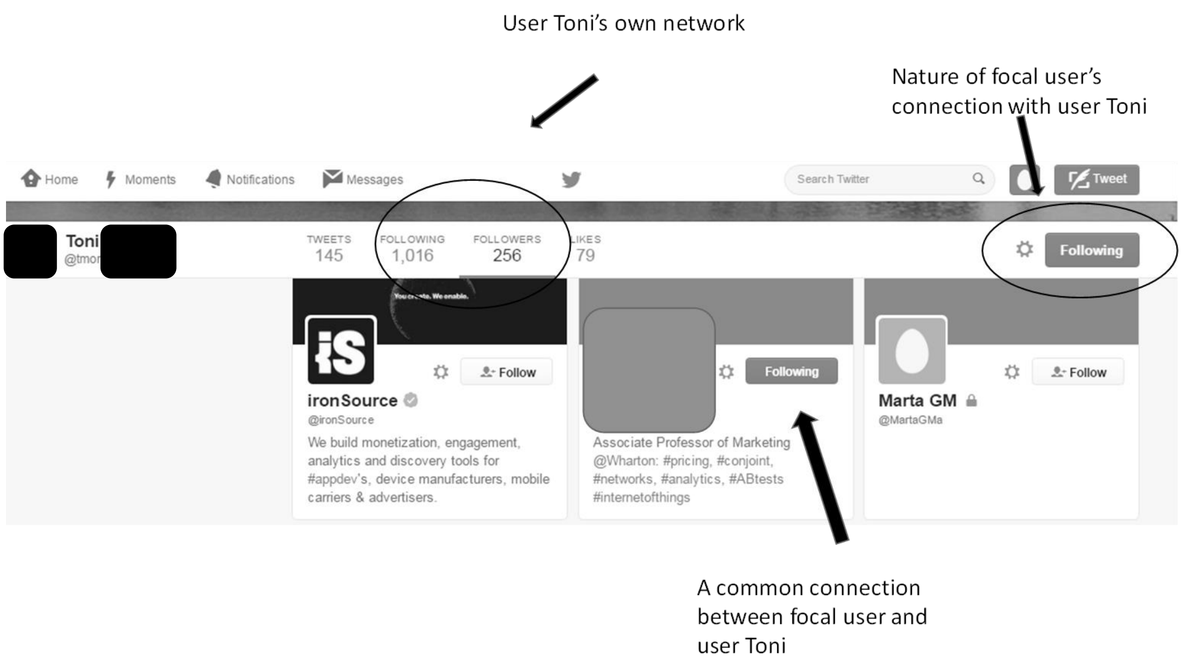


Figure 1. The Network Information A Focal User Can Observe about Her Followees

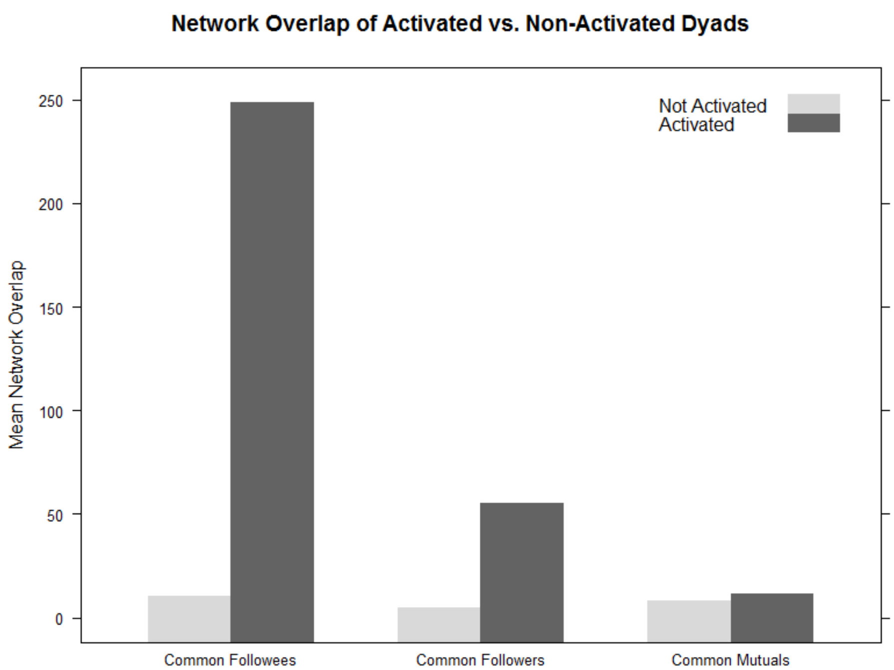


Figure 2. The Mean Network Overlap of Activated vs. Non-Activated Dyads

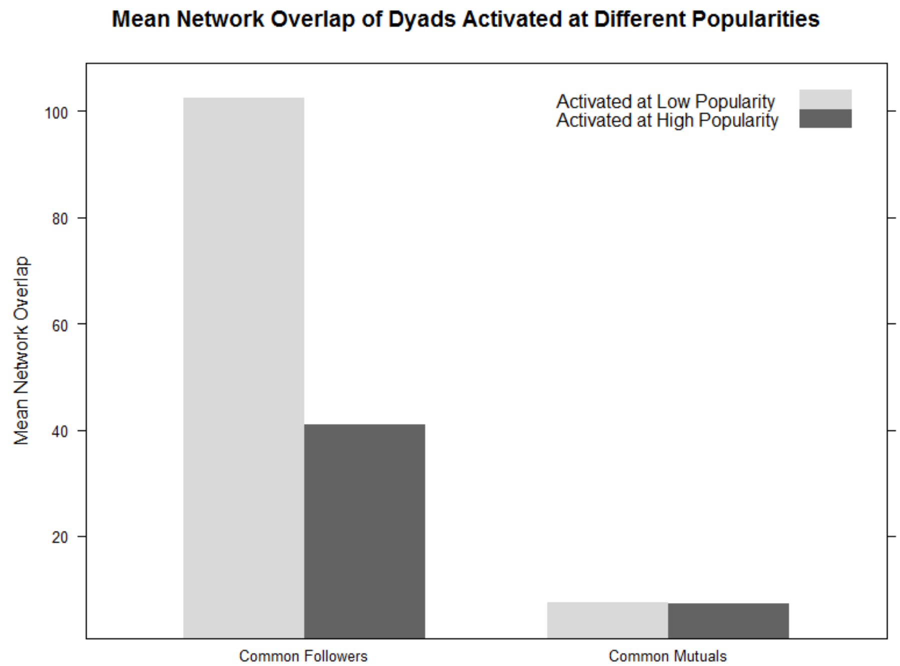


Figure 3. The Mean Common Followers and Common Mutual Followers for Dyads Activated at High vs. Low Popularity

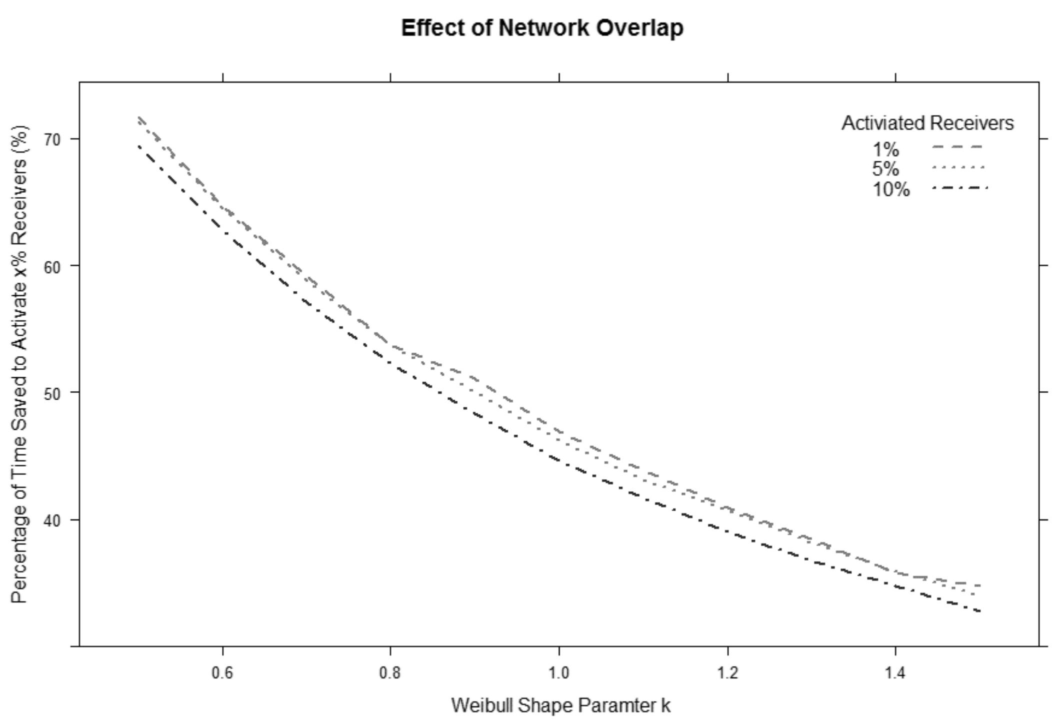


Figure 4. The Effect of Network Overlap in Speeding up the Sharing of Content

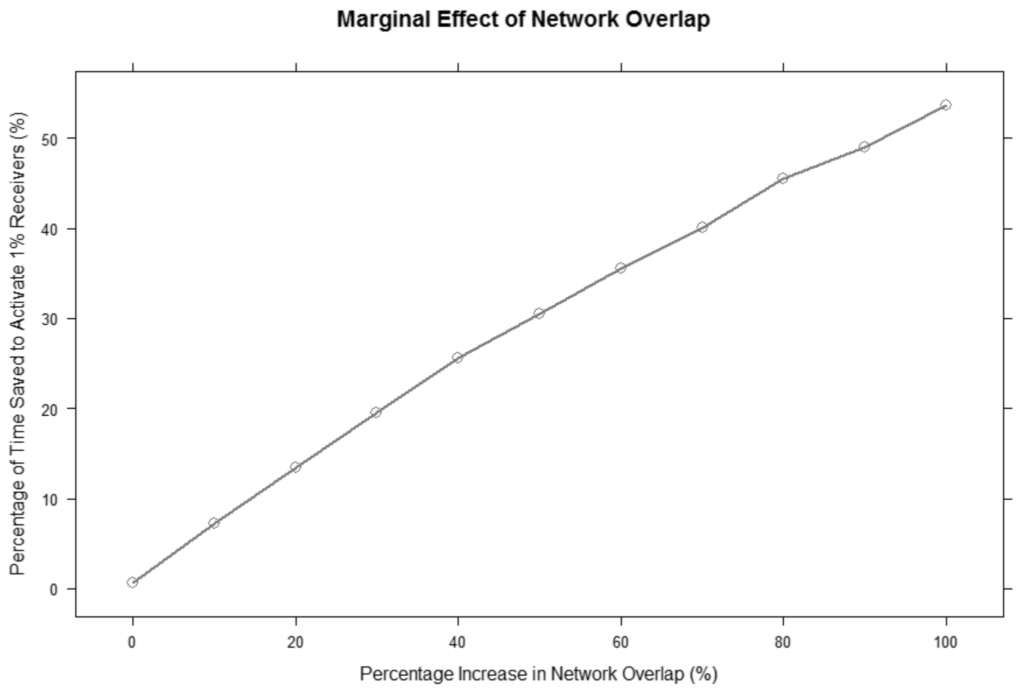


Figure 5. The Marginal Effect of Network Overlap in Activation Time

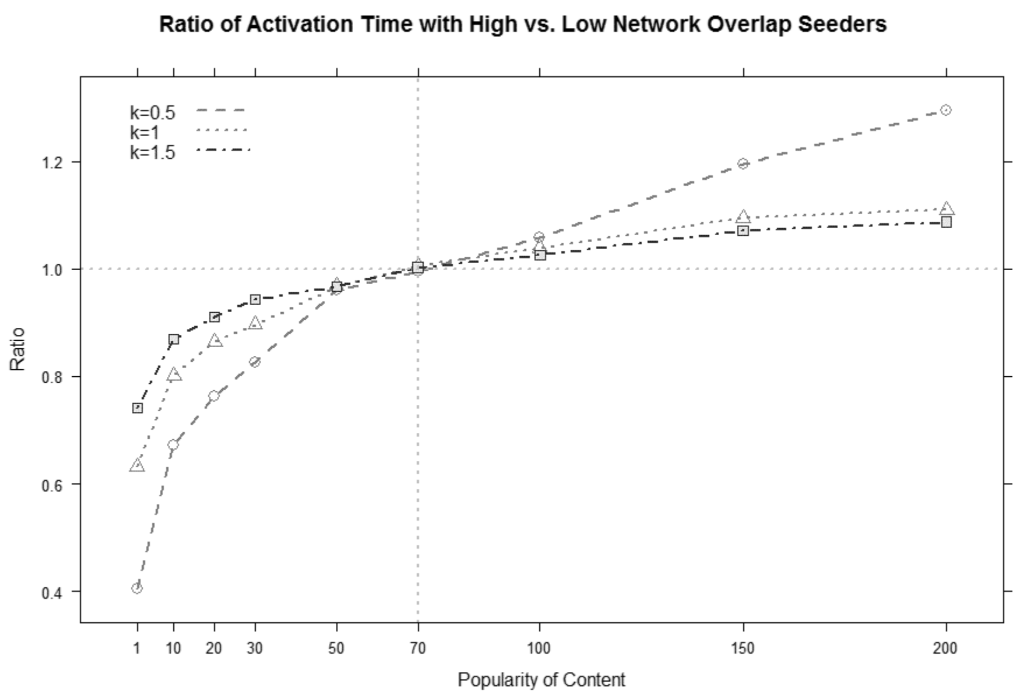


Figure 6. How the Relative Effectiveness of High Network Overlap Seeders, as compared to Low Network Overlap Seeders, Varies with the Popularity of Content

WEB APPENDIX

Web Appendix A: Simulation Studies

Our proposed model is called a collective cause model because it rests on the assumption that the event is caused by all co-senders collectively. We test the performance of the model in recovering the true parameters when the data are generated under the collective cause assumption. In practice, it is also possible that only part of the co-senders contributes to the event. To demonstrate the effectiveness of the collective cause model in dealing with such data, we focus on an extreme case in which the event is caused by one of the co-senders independently (called single-cause data). For simplicity, we assume that all the co-senders of a receiver adopt simultaneously at the beginning. This assumption has no effect on identification but greatly simplifies the data generation process. To test the robustness of the collective cause model to the distribution of independent variables, we assume that every user has three attributes drawn from three different distributions, namely, normal, binomial, and exponential. With a goal to generate a dataset with 10K events, we construct the collective-cause and single-cause datasets as follows:

1) Generate 200 senders and 5000 receivers, each has three attributes drawn from three different distributions: one normal, one binomial, and one exponential.
2) Randomly sample 10,000 senders and 10,000 receivers with replacement from the pool of 200 senders and 5,000 receivers, respectively. A one-to-one mapping between the 10K senders and 10K receivers results in 10K dyadic observations.
3) Randomly sample another 2,000 senders with replacement from the pool of 200 senders and map each of them to one of the 10K receivers in step (2) randomly. Those matched receivers in this step will therefore have multiple senders.
4) For each dyadic observation, compute the dyadic hazard, assuming the baseline hazard and all model parameters equal to 1.
5) *Collective-cause*: for each of the 10K receivers, compute her aggregated hazards by summing up the hazards from all her co-senders. Simulate a survival time for each receiver based on her aggregate hazards (Bender et al. 2005).

   Single-cause: simulate a survival time for each of the 12K dyadic observations, following the method proposed by Bender et al. (2005). If a receiver has multiple survival times associated with multiple senders, choose the minimum survival time as the survival time of the receiver.

6) To make the data more realistic, choose the lower 20% quantile of all survival times as the censoring time, such that 80% of conversion events are censored in the final data.

The data generation process of the collective-cause and single-cause data are exactly the same, except for step (5). We use a dyadic setup to ensure that the structure of the simulated dataset is similar to the structure of the dataset used in the application. Moreover, we censored 80% of events to test the effectiveness of the single cause model on incomplete observations.

To show the effectiveness of the proposed collective cause model, which doesn't speculate on the quantitative contribution of co-senders, we compared its performance with two benchmark models developed based on the idea of linear attribution in advertising.[14] The key idea of linear attribution is that each touch point contributes equally to the conversion. In the first benchmark model, we assume that every co-sender has equal probability to be the sole cause of event and maximize the expected likelihood of the event to be caused by any co-sender. We call this model the equal probability model. In contrast to the first benchmark model which assumes that only one of the co-senders is the true cause, in the second model we assume that every co-sender is part of the true cause. Specifically, we treat an event with multiple co-senders as multiple independent events caused by the co-senders each. We restrict the total case weight of each receiver to be one and evenly split the unit case weight among multiple co-senders. The second benchmark model is called the tied events model as it can be estimated by the tie handling methods of proportional hazard models (Therneau 2000).

Table A1 summarizes the relative errors (i.e., $\frac{\hat{\beta} - \beta}{\beta}$) of three models on two types of datasets, averaged over 20 runs.

[14] https://support.google.com/analytics/answer/1662518?hl=en

Table A1. Relative Errors of the Collective Cause Model

	Single-Cause Data			Collective-Cause Data		
	Tied Events	Equal Prob.	Collective Cause	Tied Events	Equal Prob.	Collective Cause
normal	-0.1883 (0.02)	-0.2102 (0.03)	0.0064 (0.02)	-0.3312 (0.02)	-0.2122 (0.03)	0.0014 (0.02)
binomial	-0.1787 (0.04)	-0.1996 (0.05)	0.0060 (0.05)	-0.3236 (0.04)	-0.2079 (0.04)	-0.0053 (0.04)
exponential	-0.1539 (0.01)	-0.1776 (0.01)	-0.0006 (0.02)	-0.2729 (0.02)	-0.1766 (0.01)	-0.0003 (0.02)
rnormal	-0.1897 (0.02)	-0.2109 (0.02)	0.0030 (0.02)	-0.3254 (0.02)	-0.2079 (0.02)	0.0066 (0.02)
rbinomial	-0.1811 (0.05)	-0.2045 (0.05)	-0.0034 (0.05)	-0.3200 (0.05)	-0.2088 (0.05)	-0.0064 (0.04)
rexponential	-0.1541 (0.01)	-0.1758 (0.01)	0.0015 (0.01)	-0.2740 (0.01)	-0.1743 (0.01)	0.0022 (0.02)

The prefix “r” indicates covariates on the receiver side. Enclosed in parentheses are the standard deviations of the relative errors.

As can be seen, the proposed collective cause model can recover the true parameters with negligible errors not only on the collective-cause data, but also on the single-cause data. This finding demonstrates that the collective cause model is a valid model even if only part of the co-senders contributes to the event. The mathematical proof regarding why the collective cause model can still recover the true parameters when only one of co-senders contributes to the event is available from the authors upon request. The intuition behind this finding is that, in the single-cause data, the overall hazard of a receiver given in the numerator of Equation (2) can be reinterpreted as the overall hazard of the receiver to be influenced by any single source she has seen. In this sense, the collective cause model is a truthful representation of the single cause data, except that it does not use the true cause information. The estimates of the tied events model and equal probability model are both substantially biased downwards, which demonstrates that arbitrary assignment of credits among co-senders may lead to misleading results. The effectiveness of the collective cause model in recovering the true parameters are robust to censoring, scaling, distribution of survival times, and the average number of co-senders on a receiver.

Web Appendix B: Diffusion Plots

The Kaplan-Meier curve shows how the average probability of a user who has been exposed to a piece of content to remain inactivated (i.e., not share the content) changes over time, which is also known as the survival probability. In the plots below, each line represents a tweet or an ad. When the content is new, the probability of a user to share is relatively high, and the number of inactivated users drop relatively faster. Thus, the survival probability declines relatively quickly in the beginning. When the content become old, the probability of a user to share it diminishes. Accordingly, the survival probability stabilizes. By definition, the survival probability decreases upon each sharing event and remains flat in-between. In the Twitter dataset, we plot the survival curves for 30 randomly sampled tweets. In the Digg dataset, we plot the survival curves for all 31 ads.

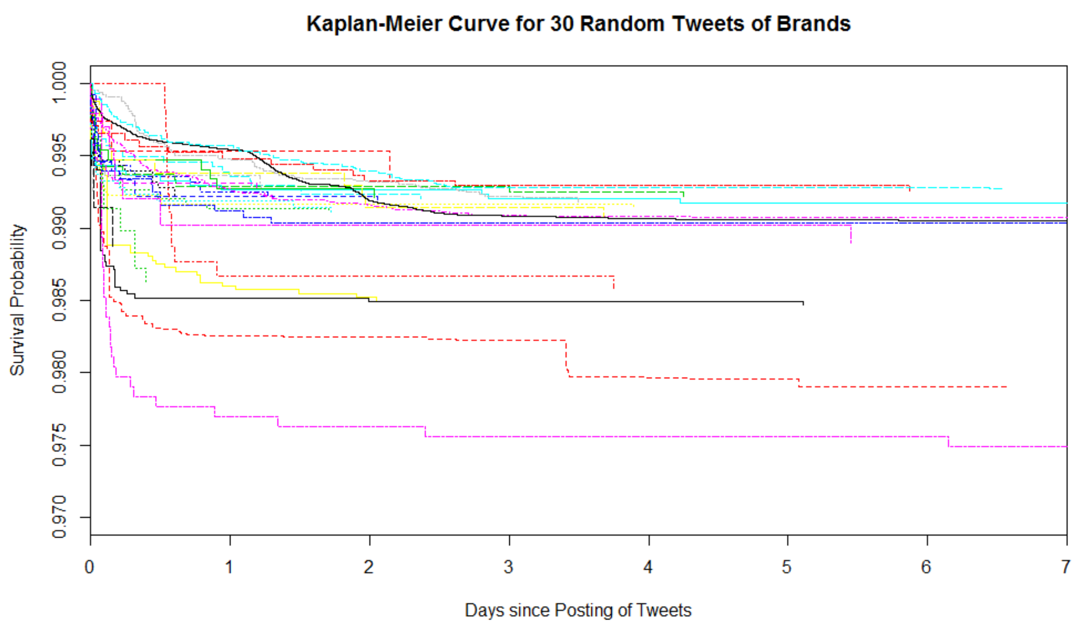


Figure B1. Kaplan-Meier Survival Curve for 30 Randomly Sampled Tweets

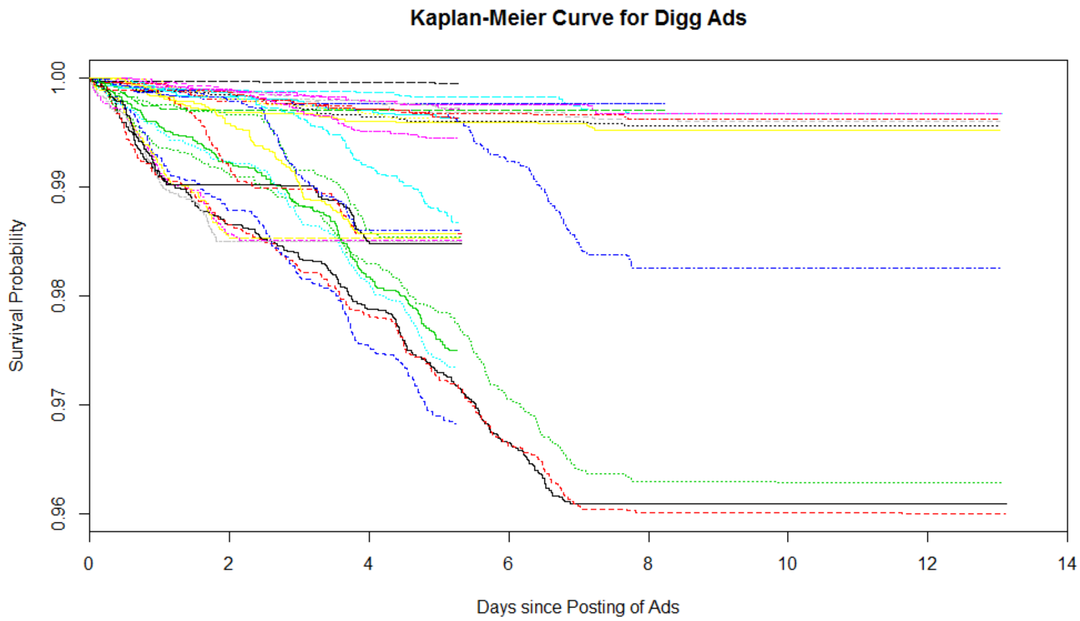
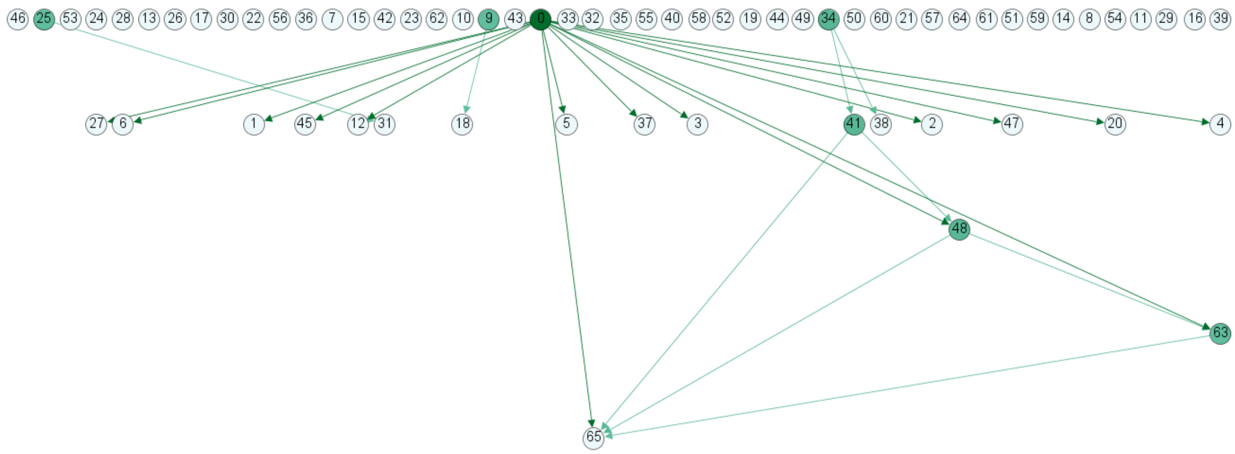


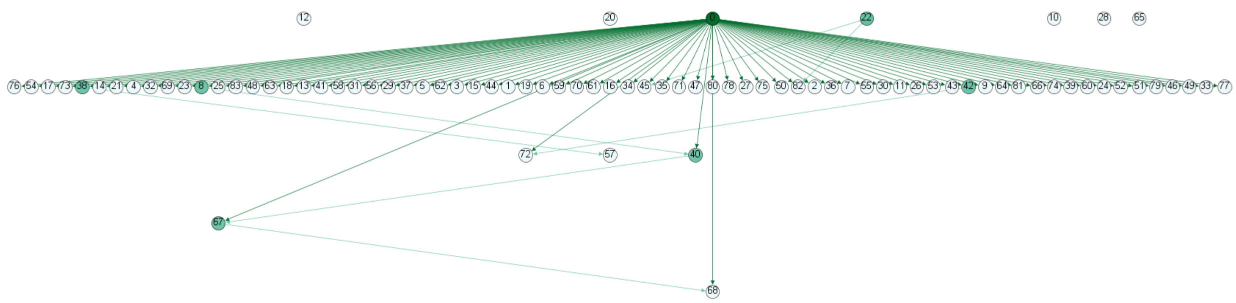
Figure B2. Kaplan-Meier Survival Curve for All Digg Ads

The plots below show how the content spread across users. These plots are made in Gephi using the DAG layout[15]. In such a layout, nodes are layered based on the longest path to activate it. The arrow represents the direction of information flow. The darkness of a node's color is proportional to its outdegree. The color of a link between two nodes is the same of the source node. If a node has no incoming link, it shares the content spontaneously. Due to the subtle difference between Twitter and Digg, we use node 0 to represent the author of a tweet whereas node 1 to represent the author of a Digg ad. On Digg, the author is automatically considered as the first digger. On Twitter, a fresh tweet has no retweet.

[15] https://marketplace.gephi.org/plugin/dag-layout/

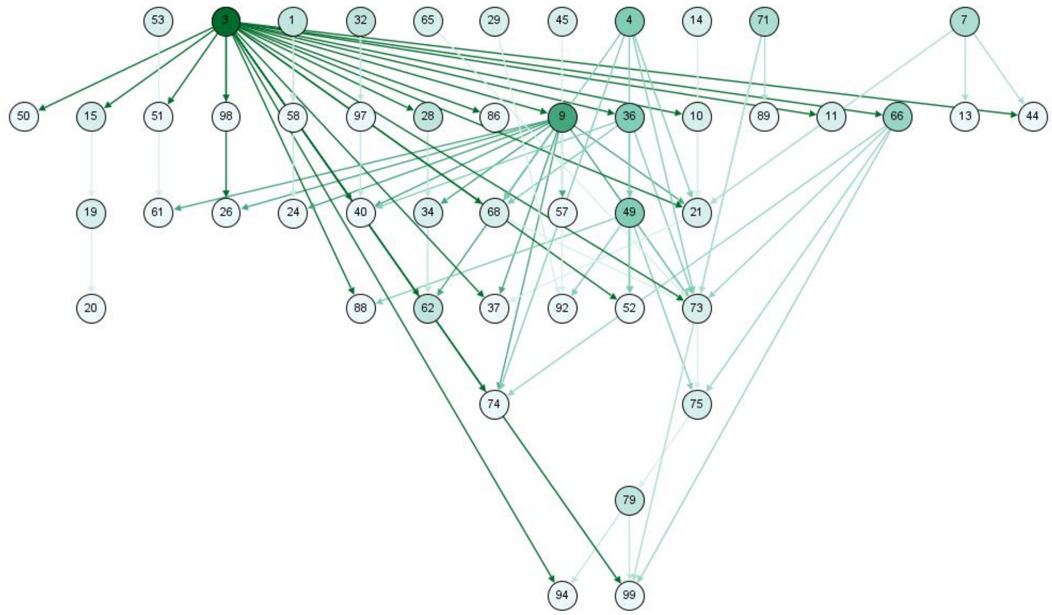


(a)

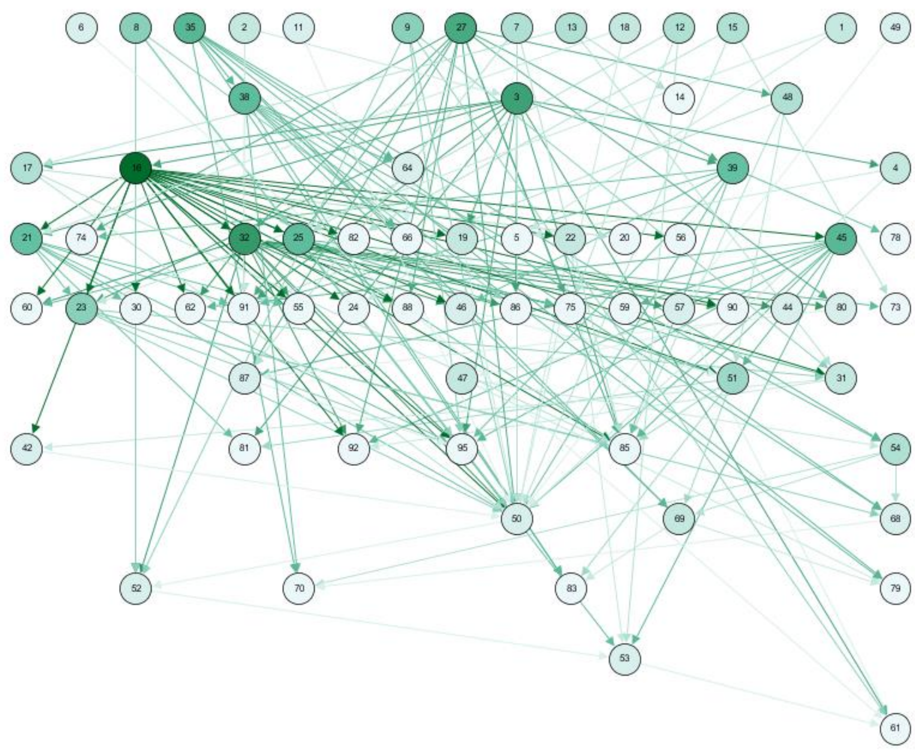


(b)

Figure B3. Sharing Graph for Two Sample Tweets



(a)



(b)

Figure B4. Sharing Graph for Two Sample Digg Ads

Web Appendix C: Main Twitter Dataset

Table C1. Complete Results on the Main Twitter Dataset

	Model1	Model2	Model3	Model4
Characteristics of Sender				
isSocialTRUE	-13.022***	-13.217***	-13.227***	-13.275***
isAuthorTRUE	-0.484**	-0.514**	-0.557**	-0.551**
logFollowees	-0.870***	-0.939***	-0.985***	-0.966***
logFollowers	1.046***	1.047***	1.037***	1.043***
logMutuals	0.540***	0.658***	0.715***	0.702***
logLists	0.339***	0.312***	0.299***	0.306***
logStatuses	-0.165***	-0.192***	-0.205***	-0.211***
logFavourites	-0.180***	-0.138***	-0.111***	-0.113***
verified	-0.323**	-0.258*	-0.149	-0.191
logRegMon	-0.024	-0.011	-0.017	-0.013
Characteristics of Receiver				
logFollowees	-0.178***	-0.177***	-0.183***	-0.180***
logFollowers	0.000	0.002	-0.005	-0.001
logMutuals	-0.125***	-0.126***	-0.119***	-0.122***
logLists	-0.021**	-0.025***	-0.018**	-0.024***
logStatuses	0.249***	0.251***	0.249***	0.250***
logFavourites	0.076***	0.077***	0.076***	0.077***
verified	-0.213.	-0.292*	-0.239.	-0.293*
logRegMon	-0.234***	-0.225***	-0.222***	-0.221***
Sharing Timing of Sender				
wday2	-0.027	-0.019	0.004	-0.006
wday3	-0.263	-0.289	-0.258	-0.257
wday4	-0.036	-0.062	-0.036	-0.037
wday5	-0.127	-0.133	-0.154	-0.152
wday6	0.174	0.176	0.163	0.161
wday7	-0.689*	-0.659*	-0.654*	-0.657*
hour(5,11]	-0.376.	-0.361.	-0.323	-0.346
hour(11,17]	-0.109	-0.080	-0.064	-0.073
hour(17,23]	-0.239	-0.249	-0.227	-0.255
shareTime	0.045*	0.041.	0.038.	0.035
Number of Co-senders				
co-senders	-2.000***	-2.062***	-2.029***	-2.065***
Dyadic Characteristics				
isMutualTRUE	-0.054	-0.054	-0.072	-0.071
logCommonFollowees	0.072***	0.081***	0.087***	0.089***
logCommonFollowers	0.116***	0.239***	0.113***	0.230***
logCommonMutuals	0.063***	0.063***	0.224***	0.079*
logCommonFollowers:logPopularity		-0.032***		-0.029***
logCommonMutuals:logPopularity			-0.052***	-0.017*
Fitness				
Log Likelihood	-100245.8	-100123.3	-100136.4	-100113.2
BIC	200855.0	200620.0	200646.1	200609.6

For wday, Monday-Sunday are coded as 1-7. Hour of a day is grouped into four bins. For dummy variables, the missing levels are the reference levels.

Table C2. The p-values of Wald Tests on the Difference between the Main Effects of Common Followers and Common Mutual Followers

	Twitter	Additional Twitter	Digg
Model 1	0.009	<0.001	<0.001
Model 4	0.002	0.02	0.004

Table C3. Mean Characteristics of Two Sets of Seeders

	Low Network Overlap	High Network Overlap
Followers	118,324.3	40,367.1
Common Followees	13.1	37.0
Common Followers	2.1	38.6
Common Mutuals	0.8	95.8

Web Appendix D: Targeting on Twitter

Below are two snapshots showing how targeting works on Twitter. Advertisers can ask Twitter to deliver their tweets to selective users of their own choice. Advertisers only need to pay for the engagements from the targeted users (or seeders), but not the further engagements from the followers of the targeted users.

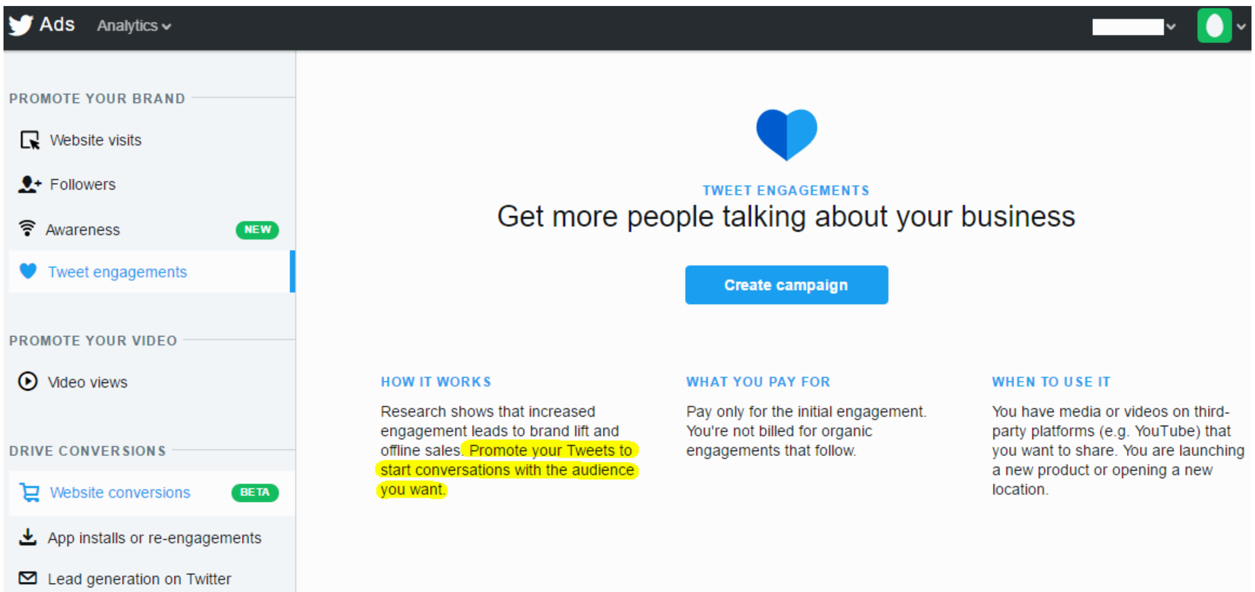


Figure D1. Creating Campaign to Increase Engagements on Twitter

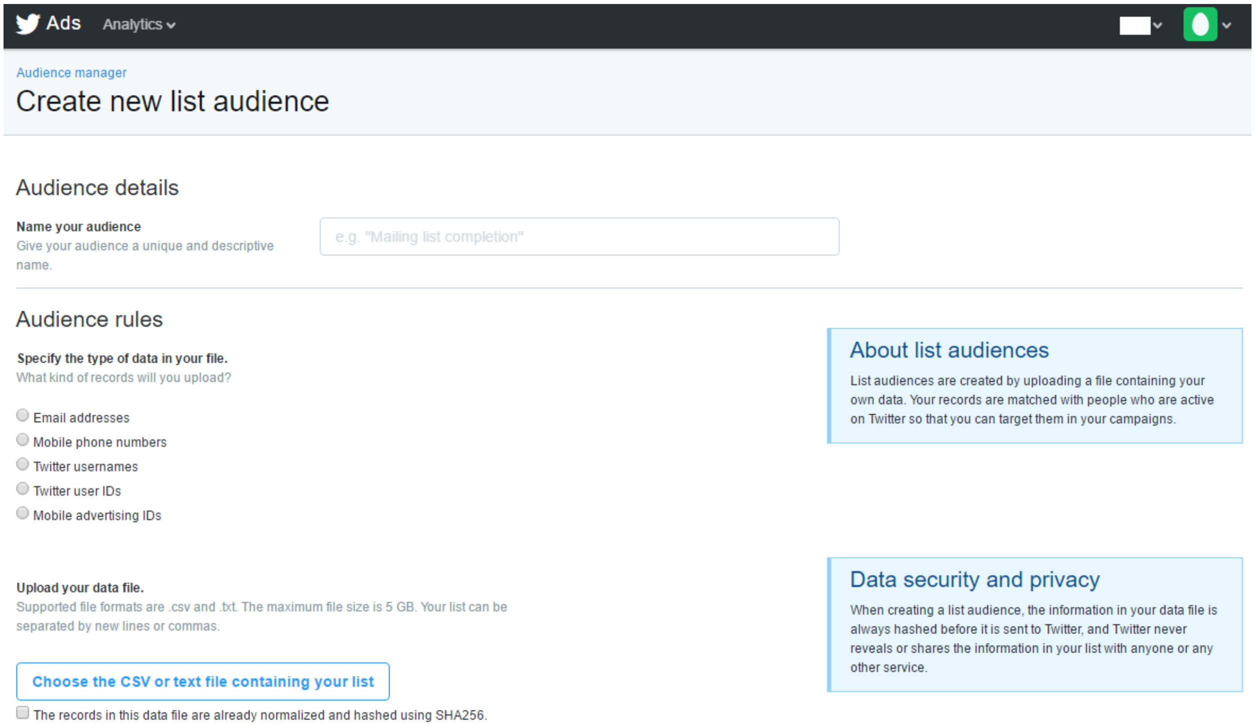


Figure D2. Customization of Twitter Campaign Audience

Web Appendix E: Additional Twitter Dataset

Table E1. Summary Statistics

Number of tweets	2,081
Number of senders	17,174
Number of receivers	162,802
Number of <sender, receiver> dyads	360,920
Number of <sender, receiver, tweet> tuples	445,156
Number of spontaneous tuples	210,457 (47%)
Number of social tuples	234,699 (53%)
Number of observations after accounting for time-varying variables	12,903,017
Number of shares (retweets)	18,873
Number of spontaneous shares	3,115 (17%)
Number of potential influenced shares	15,758 (83%)
Percentage with more than one co-senders (excluding special sender)	32.6%

Table E2. Key Statistics of Main Variables

	Zeros	Mean	SD	Min	Median	Max
Unitary Network Attributes of All Users						
Number of followees	937	6,185.9	37,657.0	0	769	4,798,543
Number of followers	1,486	10,039.5	203,555.2	0	437	75,257,863
Number of mutuals	3,555	4,116.7	27,067.9	0	228	1,714,751
Dyadic Network Attributes of Sender-Receiver Dyads						
isMutual (1– reciprocal, 0 – non-reciprocal)	99,479	0.5	0.5	0	0	1
Number of common followees	29,033	46.3	97.4	0	13	3,476
Number of common followers	48,733	59.9	414.7	0	5	33,902
Number of common mutual followers	60,449	60.3	244.3	0	5	27,132
Popularity of Content						
Number of retweets	0	9.1	13.8	2	4	100

Table E3. Correlation among Dyadic Network Characteristics

	isMutual	logCommonFollowees	logCommonFollowers	logCommonMutuals
isMutual	1.00	0.16	0.29	0.51
logCommonFollowees		1.00	0.61	0.64
logCommonFollowers			1.00	0.77
logCommonMutuals				1.00

Table E4. Complete Model Estimates on the Additional Twitter Dataset

	Model1	Model2	Model3	Model4
Characteristics of Sender				
isSocialTRUE	-5.068***	-5.544***	-5.499***	-5.570***
isAuthorTRUE	1.536***	1.601***	1.606***	1.615***
logFollowees	-0.487***	-0.523***	-0.504***	-0.516***
logFollowers	0.983***	1.033***	1.009***	1.023***
logMutuals	0.351***	0.379***	0.380***	0.384***
logLists	-0.061***	-0.076***	-0.075***	-0.076***
logStatuses	-0.141***	-0.151***	-0.150***	-0.152***
logFavourites	0.010	0.019*	0.016.	0.018*
verified	-0.397**	-0.486***	-0.536***	-0.529***
logRegMon	-0.138***	-0.143***	-0.136***	-0.140***
Characteristics of Receiver				
logFollowees	-0.164***	-0.162***	-0.161***	-0.162***
logFollowers	-0.015	-0.006	-0.007	-0.005
logMutuals	-0.346***	-0.349***	-0.351***	-0.350***
logLists	0.026**	0.017.	0.018*	0.016.
logStatuses	0.484***	0.485***	0.484***	0.484***
logFavourites	0.068***	0.069***	0.068***	0.068***
verified	-0.037	-0.182*	-0.109	-0.152.
logRegMon	-0.330***	-0.329***	-0.327***	-0.328***
Sharing Timing of Sender				
wday2	-1.274***	-1.312***	-1.277***	-1.291***
wday3	-1.925***	-1.857***	-1.831***	-1.829***
wday4	-2.694***	-2.894***	-2.800***	-2.854***
wday5	-2.357***	-2.345***	-2.332***	-2.335***
wday6	-1.845***	-1.815***	-1.808***	-1.801***
wday7	-1.004***	-0.995***	-0.969***	-0.976***
hour(5,11]	0.151.	0.114	0.100	0.103
hour(11,17]	0.331***	0.333***	0.304***	0.316***
hour(17,23]	0.379***	0.398***	0.379***	0.388***
shareTime	0.060***	0.088***	0.090***	0.093***
Number of Co-senders				
co-senders	0.481***	0.583***	0.600***	0.606***
Dyadic Characteristics				
isMutualTRUE	0.537***	0.517***	0.522***	0.518***
logCommonFollowees	0.203***	0.207***	0.204***	0.206***
logCommonFollowers	0.114***	0.344***	0.116***	0.225***
logCommonMutuals	-0.013	-0.027*	0.237***	0.139***
logCommonFollowers:logPopularity		-0.104***		-0.049***
logCommonMutuals:logPopularity			-0.120***	-0.076***
Fitness				
Log Likelihood	-65777.6	-65485.3	-65461.8	-65436.2
BIC	131880.1	131305.4	131258.3	131217

Web Appendix F: Digg Data Collection

Diggable ads were seamlessly integrated with organic stories and displayed at three fixed positions of the eighteen slots available on the front page. Initially ads are only shown on the front page. Users can digg an ad after viewing it just like digging an organic story. In that case, the ad is also included in the news feed of all their followers. Other users can explore the ad on the front page or navigate through feeds of their followees' activities in the "My News" page. All activities associated with an ad are automatically combined into a single feed for clarity. The identities of the involved followees are displayed side by side in the combined feed. Due to this feed combining feature, it is likely that each followee (co-sender) more or less has some effect on the activity of the focal user (receiver). Diggable ads were identical to organic stories except for an inconspicuous flag "sponsored by xx" below them. Diggable ads are removed from the front page when the associated advertiser runs out of budget, but users can still see them from social feeds.

We investigate the sharing of diggable ads. For the purpose of this study, we focus on all ads (31) created during a randomly chosen month-long period (May 24th, 2012 to June 25th, 2012). As mentioned earlier, we need the profile and social graph information of all involved users in the ad sharing process to study the effect of overlap associated with dyads on the sharing behavior. In the Digg setting, since all users can see the ads from the front page, they are all potential receivers. In order to control the size of our dataset, we only consider active users who can potentially digg or share these 31 ads.[16] We define a user as active if she has dugg at least

[16] Focusing on active users allows us to remove inactive users who are not at risk of sharing anymore. In practice, marketers often focus on such high-risk users in their targeting campaigns (e.g., sending coupons to customers who have purchased their products in the past or who have met some threshold on the amount spent).

one ad in the past and still maintained some activity on Digg such as posting, digging and commenting other content in the focal time period.[17]

It is important to highlight that there are a few differences in how we collect and analyze the Digg and Twitter datasets, mainly to incorporate the contextual differences between the two platforms. The first difference is that, in the Digg dataset, we treat all users as candidates for spontaneous sharing of an ad, as they all can see the ad on the front page of Digg. In the Twitter dataset, however, for each tweet, only the followers of the author (i.e., the brand) or retweeters are candidates for spontaneous sharing because there are no such non-social sources like front page that guaranteed substantial exposure for non-followers. Second, in contrast to Digg, Twitter often only shows the feed from the earliest co-sender to the receiver and does not provide any clue about the other co-senders' activity on the same tweet. Fortunately, our model can effectively handle the case even if only one of the co-senders has a significant impact (see performance on single-cause data in Table A1 of Web Appendix A). What is noteworthy is that despite these differences between Digg and Twitter, we obtain highly similar results and it further demonstrates the generalizability of our findings.

[17] We have access to profile information of all users who ever dugg one of the diggable ads between October 2010 and July 2012, including gender, location, number of diggs, number of comments, number of submissions, number of followers, and number of followees.

Web Appendix G: Digg Dataset

Table G1. Summary Statistics of Digg Dataset

Number of ads	31
Number of senders	1,058
Number of receivers	8,164
Number of <sender, receiver> dyads	95,144
Number of <sender, receiver, ad> tuples	560,044
Number of spontaneous tuples	222,846 (40%)
Number of social tuples	337,198 (60%)
Number of observations after accounting for time-varying variables	1,857,163
Number of shares (diggs)	2,810
Number of spontaneous shares	1,438 (51%)
Number of potential influenced shares	1,372 (49%)
Percentage with more than one co-senders (excluding special sender)	32.1%

Table G2. Descriptions of Independent Variables

Independent Variable		**Description**
$\boldsymbol{X_i}/\boldsymbol{X_j}$		**Attributes of sender $\boldsymbol{i}$ / receiver $\boldsymbol{j}$**
Network attributes	followees	Number of followees (out-degree)
	followers	Number of followers (in-degree)
	mutuals	Number of mutual followers
Engagement levels	diggs	Total number of diggs
	comments	Total number of comments
	submissions	Total number of submissions
Others	gender	Male, female, or missing
	regMon	How many months have the user been registered on Digg
	isSocial ($\boldsymbol{s_i}$)	1 if sender $\boldsymbol{i}$ is a social source (i.e., followee), otherwise 0
	isSubmitter	1 if the sender is the submitter of the ad, otherwise 0
$\boldsymbol{X_{ij}}$		**Attributes of a sender-receiver dyad**
Dyadic network attributes	isMutual	Do the sender and the receiver follow each other mutually
	commonFollowees	Number of followees shared by the sender and the receiver
	commonFollowers	Number of followers shared by the sender and the receiver
	commonMutuals	Number of mutual followers shared by the sender and the receiver
$\boldsymbol{X_{ik}}$		**Sender-specific attributes of an ad**
Sharing timing	wday	Day of a week when sender *i* dugg ad $\boldsymbol{k}$
	hour	Hour of a day when sender *i* dugg ad $\boldsymbol{k}$
	shareTime	Hours taken for sender $\boldsymbol{i}$ to adopt since the creation of ad $\boldsymbol{k}$
$\boldsymbol{X_{jk}}$		**Receiver-specific attributes of an ad**
	co-senders	Number of followees (co-senders) of the receiver who have already shared
$\boldsymbol{X_k}$		**Attributes of ads $\boldsymbol{k}$ (only interaction with other variables can be identified)**
	popularity	Number of diggs on an ad at a given time point

Table G3. Key Statistics of Main Variables

	Zeros	**Mean**	**SD**	**Min**	**Median**	**Max**
Unitary Network Attributes of All Users						
Number of followees	141	268.0	423.7	0	118	10,122
Number of followers	146	386.3	1,091.0	0	136	29,331
Number of mutuals	424	114.4	203.8	0	36	4,598
Dyadic Network Attributes of Sender-Receiver Dyads						
isMutual (1– reciprocal, 0 – non-reciprocal)	63,733	0.27	0.44	0	0	1
Number of common followees	4,736	41.1	52.5	0	23	814
Number of common followers	2,182	100.7	334.2	0	26	9,812
Number of common mutual followers	19,805	17.1	35.3	0	4	594
Popularity of Ads						
Number of diggs	0	93.4	86.2	4	95	295

Table G4. Correlation among Dyadic Network Characteristics

	isMutual	logCommonFollowees	logCommonFollowers	logCommonMutuals
isMutual	1.00	0.16	0.07	0.53
logCommonFollowees		1.00	0.53	0.46
logCommonFollowers			1.00	0.46
logCommonMutuals				1.00

Table G5. Complete Model Estimates on the Digg Dataset

	Model1	Model2	Model3	Model4
Characteristics of Sender				
isSocialTRUE	1.982***	0.936*	1.424***	0.967**
isDiggAdsTRUE	-0.953	-2.257*	-2.711	-0.194
logFollowees	0.030	-0.050	0.027	-0.051
logFollowers	-1.214***	-0.267***	-0.336***	-0.235**
logMutuals	-0.288**	-0.017	-0.121	0.023
logDiggs	0.309**	-0.087	-0.055	-0.075
logComments	-0.200**	-0.034	0.053	-0.009
logSubmissions	-0.012	-0.096.	-0.083	-0.057
logRegDays	-0.684***	-0.212***	-0.141*	-0.152**
genderf	0.210	-0.184	-0.202	-0.169
genderm	0.000	0.242	0.080	0.163
Characteristics of Receiver				
logFollowees	-0.178***	-0.250***	-0.258***	-0.277***
logFollowers	-0.223***	-0.209***	-0.221***	-0.215***
logMutuals	-0.092***	-0.100***	-0.105***	-0.098***
logDiggs	0.520***	0.517***	0.513***	0.518***
logComments	-0.016	-0.036**	-0.033**	-0.036**
logSubmissions	-0.144***	-0.145***	-0.147***	-0.149***
logRegDays	-0.495***	-0.458***	-0.456***	-0.450***
genderf	0.031	0.040	0.032	0.029
genderm	0.078.	0.100*	0.079.	0.081.
Sharing Time of Sender				
wday1	-0.685*	0.266	-0.107	0.000
wday2	0.411	0.238	-0.196	-0.134
wday3	0.871**	0.323*	0.229	0.171
wday4	0.785*	0.684***	0.340*	0.394*
wday5	0.331	0.805**	0.316	0.421
wday6	0.499	1.190***	0.034	0.714**
hour(5,11]	-0.132	0.258	0.202	0.146
hour(11,17]	-0.232	0.103	-0.029	-0.015
hour(17,23]	0.245	0.107	-0.069	-0.122
shareTime	-0.310***	0.170***	0.189***	0.222***
Number of Co-Senders				
co-senders	-0.008	0.013	-0.015	-0.010
Dyadic Characteristics				
isMutualTrue	-1.189***	-0.574***	-0.911***	-0.674***
logCommonFollowees	0.192.	0.212**	0.073	0.191**
logCommonFollowers	1.125***	2.287***	0.154**	1.520***
logCommonMutuals	-0.079	0.034	2.739***	0.768***
logCommonFollowers:logPopularity		-0.517***		-0.339***
logCommonMutuals:logPopularity			-0.589***	-0.171***
Fitness				
Log Likelihood	-20115.5	-19753.2	-19693.3	-19674.7
BIC	40548.7	39831.9	39712.2	39683.0

The three levels for gender are: m – male, f – female, and u – unknown. For wday, Monday-Sunday are coded as 1-7. Hour of a day is grouped into four bins. For dummy variables, the missing levels are the reference levels.