# Network Embeddedness and Content Sharing on Social Media Platforms

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

January 5, 2016

**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 Embeddedness and Content Sharing on Social Media Platforms


### *ABSTRACT*

We study the impact of network embeddedness – the overlap in network neighbors between two users – on content sharing in directed social media platforms. We propose a hazard model that flexibly captures the impact of three different measures of embeddedness on content sharing and apply it individual-level data from Digg. Our results indicate that all three measures of embeddedness have an impact on the amount of content sharing. Specifically, a receiver is more likely to share content from a sender if they share more common followees, common followers or common mutual followers after accounting for the other measures. Additionally, the effect of common followers and common mutual followers is positive when the content is novel but decreases and may become negative when many others in the network have already adopted it. Our findings are managerially relevant for targeting customers for content propagation in social networks.



Keywords: social media, content sharing, embeddedness, multiple senders, hazard models

## *INTRODUCTION*

Social media platforms are a popular medium for firms to reach out to customers (Schweidel and Moe 2014; Stephen and Toubia 2010). A recent survey suggests that three quarters of advertisers had used social media for advertising, and 64% of them planned to increase their social advertising budgets (Nielsen 2013). One likely reason for the growing emphasis on social advertising is the promise that users who engage with the ad content might spread information about new products to their social network connections (Aral and Walker 2011; Aral and Walker 2012; Aral and Walker 2014; De Bruyn and Lilien 2008; Leskovec et al. 2007; Trusov et al. 2010).

A primary requirement for the propagation of content in a social network is that receivers in turn share the information that they obtain from their sender/s. However, empirical evidence for such information cascades is limited (Goel et al. 2015). For instance, the average number of retweets per tweet is often less than 20.[1] Thus, it becomes even more important to understand the underlying drivers for the propensity of a receiver to share information. Such analysis will also be useful for marketers to improve their targeting of customers within social networks (Kempe et al. 2003; Richardson and Domingos 2002; Watts and Dodds 2007).

Extant work indicates that one important driver of the propensity of a receiver to share information is the embeddedness or network overlap between a dyad i.e. a sender-receiver pair (Aral and Walker 2014). Network embeddedness or network overlap[2] is a shared characteristic between users in a network and has been associated with effective knowledge transfer between individuals (Reagans and McEvily 2003), extent of information sharing among users (Aral and Van Alstyne 2011) and adoption of applications by users (Aral and Walker 2014). In the context of firms, network embeddedness has been associated with trust between firms (Uzzi 1997) and their economic actions (Granovetter 1985).

Network embeddedness is broadly defined as the number of common neighbors between two users (Easley and Kleinberg 2010). Its operationalization depends on whether the network is directed or not. In undirected networks (e.g., Facebook), embeddedness or network overlap simply means the number of common friends between two users. In directed networks (e.g., Twitter and Digg[3]), by interpreting a

---

[1] Social engagement benchmark report (salesfore 2015): https://www.exacttarget.com/system/files_force/deliverables/etmc-socialengagementbenchmarkreport-tw.pdf?download=1&download=1
[2] We use embeddedness and network overlap interchangeably in this paper
[3] Digg maintained an internal directed network before August 2012, but now it uses the external social networks of users on Twitter and Facebook instead.

neighbor as a followee (outgoing link), follower (incoming link) or mutual follower (bidirectional link), embeddedness can be characterized by three different metrics: the numbers of common followees, common followers, and common mutual followers. 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.

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) |
| **Embeddedness** | |
| 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 | Digg an ad or retweet a tweet |
| 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 ad/tweet |

The nature of overlap in the network connections between two users can reveal the motivation to share content. For example, followees of a user have a high persuasive influence (Haenlein 2013; Hall and Valente 2007) on the user and can represent user's interests and expertise. Furthermore, people tend to share content that signals their expertise (Packard & Wooten, 2013). Thus, more common followees between a sender-receiver pair may suggest similar expertise and higher propensity for a receiver to share the content obtained from the sender. Likewise, more common followers between the sender and the receiver may suggest that their followers share a similar taste. In this case, a receiver may consider content to be more suitable for her audience and may have a higher propensity to share content. Additionally, a higher number of common mutual followers may represent higher trust (Burt 2001; Granovetter 1973) and social bonding (Alexandrov et al. 2013; Ho and Dempsey 2010; Wiatrowski et al. 1981) and may also increase the propensity of a receiver to share information. Finally, as the activities of users on social media platform are visible to others, factors such as uniqueness of content can play a role. It is well documented that users (consumers) have a strong desire for uniqueness and sharing novel content can satisfy such a need (Alexandrov et al. 2013; Cheema and Kaikati 2010; Ho and Dempsey 2010; Lovett et al. 2013). Thus, if the information to be communicated is not novel, a receiver will be less likely to do so.

The purpose of this article is to assess the impact of embeddedness on the level of content sharing in directed networks. We do so using a micro-level model for content sharing within sender-receiver dyads. Our work complements extant work on the role of influential users on product adoption (Trusov et al. 2010) and information diffusion (Susarla et al. 2012; Yoganarasimhan 2012). Other studies have described a user's propensity to adopt a product and share related information based on unitary attributes of adopters such as their demographic and behavioral characteristics (Bapna and Umyarov 2015; Haenlein 2013; Iyengar et al. 2011; Katona et al. 2011; Nitzan and Libai 2011; Rand and Rust 2011). Some other studies have considered shared characteristics of a sender and a receiver but largely in undirected networks (Aral and Walker 2014) where there is a single metric for the overlap among users, i.e., the number of common friends. Finally, in the case of directed networks, to the best of our knowledge, only the effect of reciprocity in the connections between a sender and a receiver has been considered (Shi et al. 2014).

A dyadic level study imposes stringent requirements on the data: the availability of users' profile information, social graph information, and time-stamped, highly granular, individual-level information about sharing activities. In order to meet these requirements we collected a dataset, which represents sharing of sponsored ads on Digg in a month long period in 2012. We also validate our results using a second dataset that captures sharing of tweets posted by Fortune 500 companies on Twitter in a month long period in 2015. At the time the data were collected, both websites maintained a directed social network, allowing users to follow others to keep themselves informed about their activities.

A dyadic level study introduces a methodological challenge as well: multiple senders may share the same content with a focal user and the lack of information regarding the contribution of each sender makes it difficult to identify the impact of dyadic characteristics on receiver's sharing propensity. For example, in the dataset from Digg, 32% of receivers who decided to share had multiple co-senders. In response to this problem, we propose 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 diffusion in networks.

We emerge from the analysis with three key findings. First, we establish that embeddedness plays a role in information sharing in directed networks. That is, the propensity of receiver to share information depends on all three measures of embeddedness (i.e. common followees, common followers and common mutual followers). Second, the effect of embeddedness on content sharing varies across the three metrics suggesting that they may have differing underlying drivers. Third, the effects of common followers and common mutual followers are moderated by the novelty of content. Their effects are positive only when the information is relatively novel (i.e., not shared by many others). When many others have shared the content, the positive effects may decrease and may even become negative, likely due to users' need for uniqueness. This finding suggests a boundary condition for the positive impact of embeddedness found in previous work.

The rest of the paper is organized as follows. We begin with a discussion of related literature and propose specific hypotheses about the impact of the three embeddedness metrics on content sharing. Then, we describe the proposed model and the dataset from Digg that we use in the application. Next, we discuss the results of model estimation and several robustness checks including validation of our results with the dataset from Twitter. Finally, we conclude with theoretical and managerial implications.

## *RELATED LITERATURE*

Our work relates to the broad literature on the role of network characteristics on user actions and outcomes in a social network. These include studies of information sharing (Shi et al. 2014; Susarla et al. 2012; Yoganarasimhan 2012), product adoption (Aral and Walker 2014; Bapna and Umyarov 2015; Iyengar et al. 2011; Katona et al. 2011), and customer churn (Haenlein 2013; Nitzan and Libai 2011).

Some studies have investigated the role of unitary network characteristics of the sender on the overall extent of adoption in the network. For example, Yoganarasimhan (2012) studies how the size and structure of the local network of a user affect product diffusion in undirected networks. The specific network characteristics investigated include the numbers of first- and second-degree friends, the clustering coefficient and the 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 the user influence based on the cascade size associated with a user's extended network. While these studies consider the effect of sender's local and extended network on their effectiveness in spreading product adoption behavior, they do not consider an individual receiver's propensity to adopt these products.

Others have investigated the role of unitary network characteristics of the receiver on her individual adoption behavior. For instance, Iyengar et al. (2011) consider the impact of user characteristics such as opinion leadership (captured by the number of ties and self-reported measures) on the adoption of prescription drugs. Katona et al. (2011) investigate the effect of a receiver's network characteristics on their adoption or registration at a site. Similarly, Bapna and Umyarov (2015) consider the effect of the receiver's network size on her propensity to subscribe to a music site. Rand and Rust (2011) evaluate the role of local network on the adoption behavior using an agent based model. Nitzan and Libai (2011) and Heinlein (2013) investigate the role of the network neighbors' churn behavior on the retention behavior of an individual. However, none of the above studies considers the impact of shared network characteristics between the receiver and the sender on the former's adoption behavior.

Some recent studies do focus on the role of shared characteristics on a focal user's actions albeit in undirected networks. For example, Centola (2010) shows that users are more likely to adopt when they receive social reinforcement from multiple neighbors and, as a result, the behavior spreads more in a clustered network than a random network. While a clustered network can represent higher network overlap with neighbors, this overlap is artificially created in the experiment and does not directly capture the shared characteristics between two users. Aral and Van Alstyne (2011) investigate the role of embeddedness on the sender's incentive to share information with a particular receiver but not the receiver's propensity to, in turn, share the content with all her followers. Aral and Walker (2014) examine the effect of network embeddedness more directly and find that it has a positive effect on the adoption of an application on Facebook (an undirected network). Finally, while Shi et al. (2014) study information sharing in a directed network, they primarily focus on the role of reciprocity between senders and receivers.

In summary, there is clearly much interest in understanding how users' network characteristics affect product diffusion and information sharing in networks. While literature has focused on either aggregate network measures or unitary characteristics of senders and receivers, an emerging stream of work has started to highlight the role of such dyadic attributes as network embeddedness. This literature, to the best of our knowledge, has considered undirected networks. In this paper, we fill the gap and evaluate how network embeddedness affects information sharing in directed networks. Table 2 provides a summary of existing literature.

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

| Study | Network Characteristics | Network Type | Context and User Actions |
|---|---|---|---|
| | **Dyadic network characteristics** | | |
| **Present study** | Three network embeddedness metrics between dyads | Directed | Online content sharing |
| Aral and Walker (2014) | Network embeddedness and interaction intensity between dyads | Undirected | Facebook app adoption |
| Aral and Van Alstyne (2011) | Network embeddedness between dyads | Undirected | Information sharing by sender with individual receiver |
| Shi et al (2014) | Reciprocity between dyads | Directed | Online content sharing |
| | **Unitary network characteristics** | | |
| Yoganarasimhan (2012) | Network characteristics of sender | Undirected | Diffusion of Youtube videos and related information |
| Susarla et al. (2012) | Network characteristics of sender | Directed and Undirected | Diffusion of Youtube videos and related information |
| Katona et al. (2011) | Network characteristics of receiver | Undirected | Registration (Adoption) of social networking site |
| Bapna and Umyarov (2015) | Network size of receivers | Undirected | Subscription (Adoption) of Last.fm |
| Iyengar et al. (2011) | Opinion leadership of receiver | Directed | Adoption of prescription drug |
| Bakshy et al. (2011) | Network characteristics of sender | Directed | Information diffusion |
| Rand and Rust (2011) | Network characteristics of receiver | Undirected | Adoption behavior using an agent based model |
| Irit and Libai (2011) | Churn of behavior of neighbors | Undirected | Churn behavior of receiver |
| Hanlein (2013) | Churn behavior of ingoing and outgoing connections of receiver | Directed | Churn behavior of receiver |
| | **Overall network structure** | | |
| Centola (2010) | Overall structure of network (clustered vs. random) | Undirected | Registration (Adoption) of online health forum |

## *THEORETICAL BACKGROUND AND HYPOTHESES*

Consumers typically share content to satisfy multiple goals. Users may share content with others in a social network for the purpose of impression management (Berger 2014). Further, factors such as trustworthiness of a sender (Burt 2001; Granovetter 1973) may play a role in a user's propensity to share any content received from a sender. Users may have additional motives as well to share content such as social bonding (Alexandrov et al. 2013; Berger 2014; Ho and Dempsey 2010; Travis 2002; Wiatrowski et al. 1981) and the need for uniqueness (Cheema and Kaikati 2010; Grier and Deshpandé 2001; Ho and Dempsey 2010; Lovett et al. 2013; Snyder and Fromkin 1980). Next, we outline these motivations in more detail and how they relate to our main construct of network embeddedness.

*Impression Management.* Users share content to shape others' impression about them. 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 and enhance their social status (Alexandrov et al. 2013; Grier and Deshpandé 2001; Lovett et al. 2013).

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 characteristics, knowledge base or expertise (Packard & Wooten, 2013). Further, content sharing is a social exchange process (Aral and Van Alstyne 2011). To increase social acceptance or social recognition, users may selectively share information of interest to their audience (Aral and Van Alstyne 2011; Wu et al. 2004), as sharing information perceived to be unsound or irrelevant can hurt their reputation (Barasch and Berger 2014; Bock et al. 2005).

*Trust*. Trust is a key determinant of social information exchange process (Burt 2001; Granovetter 1973). The trust of users on the source (i.e., senders) can alleviate the receivers' concern on the quality of the content and hence increase the probability of sharing (Camarero and San José 2011). The trust between two users often increases with common mutual connections (bandwidth) between them (Aral and Van Alstyne 2011; Burt 2001).

*Social bonding*. Social control theory suggests that people have a need to bond with others and maintain relationships (Travis 2002; Wiatrowski et al. 1981). Social bonding is also referred to as "need to belong" (Alexandrov et al. 2013; Ho and Dempsey 2010). The formation of a bond between individual and a group requires frequent interactions with others in the group (Alexandrov et al. 2013). On social media platforms, as the user actions are visible, one way to interact with others is to further share the content shared by others. The closer two users are, the stronger obligation they may have in sharing content shared by each other.

*Need for uniqueness*. The theory on self-presentation suggests that users are intrinsically motivated to achieve uniqueness (Tajfel and Turner 1979; Turner and Oakes 1986) 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 word of mouth for differentiated brands (Lovett et al. 2013). Need for uniqueness has also been observed for other user generated content such as reviews (Ludford et al. 2004) and photographs (Zeng and Wei 2013). Thus, in order to establish a unique identity on social media platforms, a user may resist sharing content that have already shared by many others.

In the case of content received from a sender, we posit that the characteristics shared between the user and the sender are an important contextual feature that can moderate how likely a user will satisfy one or more of the above mentioned goals and, thereby, influence their propensity to share content. We use network overlap or embeddeness between users in a social network to operationalize the shared characteristics. Next, we discuss our hypotheses on how the three metrics of embeddedness can impact content sharing.

**Common Followees**

In a directed social network, people follow others to keep themselves informed about their activities. Followees of a user have a high persuasive influence on the user (Haenlein 2013; Hall and Valente 2007). Thus, the composition of one's followees largely reflects her topical interest or taste. In addition to taste, the composition of one's followees may also reflect her expertise, as people may selectively follow others with similar expertise. In order to signal online identities and create an impression, users tend to share content falling into their area of expertise or interest (Berger 2014; Packard & Wooten, 2013). This is likely irrespective of the type of content, including popular content. Therefore, the more common followees two users have, the more likely they have similar expertise and taste due to homophily, and the more likely they will share the content shared by each other. So 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 audience is an important factor that users are likely to consider while sharing content (Berger 2014). 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.

On the other hand, an alternative driver that may lower the propensity of a receiver to share content obtained from a sender with whom the receiver has a lot of common followers is the need for uniqueness. Sharing redundant (i.e., duplicated) content that has already been seen by their followers from other sources can harm the perception of the receiver as a unique source of information. Thus, novelty of content can play a role in moderating the impact of common followers on the propensity of a receiver to share content. Less popular or novel information is more valuable due to its scarcity (Aral and Van Alstyne 2011). When the content is not as popular yet, the novelty of the content will make it relatively easier for a receiver to distinguish herself from others. In such a case, the sharing decision of the receiver should be primarily driven by impression management rather than by her need for uniqueness (as it is being satisfied by sharing novel content). When the content is popular, the need for uniqueness may be strong enough to outweigh impression management. 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 information.

**Common Mutual Followers**

The number of common mutual followers characterizes the mutual accessibility of two users through third-parties, which may be the most appropriate counterpart to the embeddedness defined in undirected networks. 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. Therefore, the level of trust between two users should increase with the number of common mutual followers. In addition, the more common mutual followers two users have, the more likely they belong to the same social group, and the more likely they feel obligated to propagate content shared by each other in order to maintain a strong social bond. Both drivers on trust and social bonding suggest that the number of common mutual followers should have positive effect on content sharing. More common mutual followers may also suggest a common taste for audience. Finally, more common mutual followers suggests higher similarity in taste and expertise due to homophily even after accounting for the effect of other embeddedness metrics. This would further increase the receiver's propensity to share content.

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 a 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 already satisfied and the receiver would have a higher propensity to share content due to high number of common mutual followers. 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 positive effect of common mutual followers on the receiver's propensity to share content from sender decreases with the popularity of the information.

Table 3 summarizes the drivers associated with the three embeddedness 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 drive the effect of common followees, as followees represent sources rather than the audience of a focal user. That different drivers are associated with the three metrics illustrates the nuanced role of embeddedness on information sharing in directed networks.

Table 3. Drivers Associated with the Three Embeddedness Metrics

| **Embeddedness Metric** | **Positive Driver** | **Negative Driver** |
|---|---|---|
| Common followees | Impression management | |
| Common followers | Impression management | Need for uniqueness |
| Common mutual followers | Trust, social bonding, impression management | Need for uniqueness |

## *MODEL*

Our objective is to evaluate the impact of network embeddedness 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 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 embeddedness. 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 effect 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 information 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 ad content over the social media platform, Digg.com (as it is the context of our primary dataset). On Digg, when a user (sender) diggs (shares) an ad (content), 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 diggs the same ad. In addition to social feeds, users can also see the ad on the front page of Digg. Therefore, there are two types of shares on Digg: those driven by social sources (i.e., feeds from followees) and others driven by non-social sources (i.e., the front page). Other platforms such as Twitter have a similar process for information sharing between users connected in a social network.

Let $i$, $j$, and $k$ index senders, receivers, and ads, respectively. Let $t$ be the time elapsed since the creation of an ad. 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 Digg). Let $X_{ij}$ represent the dyadic attributes concerning sender $i$ and receiver $j$ (e.g., embeddedness measures), $X_{ik}$ represent sender $i$'s attributes that are specific to ad $k$ (e.g., the time sender $i$ diggs the ad $k$), and $X_{jk}$ represent receiver $j$'s attributes that are specific to ad $k$ (e.g., number of receiver $j$'s followees that have shared ad $k$). Let $\lambda_{ijk}(t)$ represents the dyadic level hazard of sender $i$ causing receiver $j$ to adopt ad $k$ at time $t$. Let $\lambda_{k0}(t)$ represents the baseline hazard for ad $k$. The dyadic level hazard, stratified on ads, is given by

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

$\lambda_{k0}(t)$ captures the baseline hazard for each ad. Note that the above semi-parametric formulation allows $\lambda_{k0}(t)$ to change arbitrarily over time and across ads and allows us to capture static ad-specific effects such as the ad content and time-varying effects such as overall reduced tendency to share a specific ad with time. For example, $\lambda_{k0}(t) = 0$ represents a case when an ad 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 ad from multiple senders.

Note that $X_{ik}$ and $X_{jk}$ include variables representing when a sender shares (which accounts for decaying effect) 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 information in H3 and H5. To test these effects, we consider interaction of the popularity of ads 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 content received from another user in the social network, or after receiving it directly from the platform or an external source. The latter type of sharing is termed as a spontaneous sharing and occurs via a non-social source (e.g., platform or external site). In order to incorporate the impact of non-social sources (e.g., the front page of Digg) in our study, we treat them as a special sender and use a dummy variable to capture their effect on the hazard rate:

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

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 $\beta_0$). The parameter $\beta_0$ captures the combined effect of all non-social sources, as compared to a sender whose attributes may be 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 ad $k$ 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 ad $k$ 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 ad $k$.

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 standard 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 ad $k$ 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 ad $k$ 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 $\tau_{jk}$, 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 (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 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 A.1 in Appendix A).

**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 ad-related content might be more likely to share ads 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[4] and the receiver-level, which allow each sender and receiver to have a random intercept that captures the main effect of unobserved characteristics. 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 estimate models with fixed effects at the sender level (fixed effects allow for 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 information sharing across a dyad is that a receiver often sees the same information from multiple senders before sharing, and the quantitative contribution of each co-sender may

[4] 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.

be unclear. We address this challenge statistically by proposing a novel proportional hazards model that determines the contribution of each co-sender based on their characteristics.

## *DATA*

We seek to understand how embeddedness between a sender and a receiver connected in a social network impacts the sharing behavior of the receiver. A dyadic level study imposes stringent requirements on the data. First, we need a sample of marketing-related messages or content generated on a social media platform by firms.[5] 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 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., portal pages), we can also consider them as receivers.

We collected a dataset with the desired information from Digg.com, one of the largest online social news aggregation websites. On the website, 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, which 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's users. Diggable ads were seamlessly integrated with organic stories and displayed at three fixed positions of the eighteen slots available on the front page. At

[5] This is important as we can establish the implications of our results for firms utilizing social media to reach out to consumers.

the time we collected the data, Digg maintained a Twitter like social network structure (see footnote 1), allowing users to follow each other.

Initially ads are only shown on the front page. Users can digg up or down 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 including mutual 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.[6] 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 embeddedness 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.[7] We define a user as active if she has dugg at least 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.[8] In robustness analysis, we also consider users who have dugg an ad in the past but have no activity during the focal period and find that our results are similar.

For each potential receiver, we generate one dyadic observation for her if one of her followees shares the ad. Since everyone has access to the front page[9], we generate one additional dyadic observation for each potential receiver, with the front page being the sender. The act of digging allows the user to share the ad with her followers. One converts from a receiver to a sender immediately after the sharing activity, implying

---

[6] Identification of the effect of network drivers is easier for diggable ads as opposed to that for organic content. Diggable ads are guaranteed to be displayed on the front page before running out of budget, whereas whether an organic story is displayed on the front page depends on many factors, including the diggs the story receives and the freshness of the story. Therefore, the spontaneous hazard of organic stories may change radically over time due to their unstable visibility on the front page, which makes it difficult to mod

[7] 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).

[8] 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.

[9] On Digg, the front page is the primary non-social source for the sharing of ads. Another possibility, albeit rare in our context, is that users may discover the content through a search engine. For ease of exposition, we refer all non-social sources as the front page.

that senders are a subset of receivers. A nuanced issue in our context is that ads stop showing up on the front page after a certain period and as a result, the spontaneous hazard becomes zero. To ensure, that is not the case, we choose a censoring time for each ad as the last time when the ad was shared spontaneously by a user. The censoring time of an ad ranges from 1.4 days to 7 days, after its creation. The average censoring time is approximately 5 days. This resulted in a sample of 8,164 users and 95,144 dyads. Table 4 shows the summary information of the dataset. The table shows that 32% of shares have more than one co-sender (excluding the special sender "front page"), and the average number of co-senders is 2.82, including the front page.

Table 4. Summary Statistics

| | |
|---|---|
| Number of ads/tweets | 31 |
| Number of sharing user (senders) | 1,058 |
| Number of potential sharing user (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 shares (diggs) | 2,810 |
| Number of spontaneous shares | 1,438 (51.2%) |
| Number of potential influenced shares | 1,372 (48.8%) |
| Percentage with more than one co-senders (excluding special sender) | 32.1% |

We used the APIs provided by Digg to collect the social graph of all potential users who could share the sample ads. Due to the rate limit on API calls, it took 19 days (June/7/2012- June/26/2012) to collect a single snapshot of the complete set of followers and followees for these users. One concern with this data extraction process is that network of users may have changed even during the sample period. However, the extent of network changes is small in our setup.[10] Thus, changing network is not likely to significantly impact our results. As a further check, we split our sample in two subsamples and repeat the analysis for each (see robustness checks). All our substantive findings are robust.

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

[10] By comparing the profile information of users on June 7 and July 9, we found that the both the follower and followee numbers changed less than 5% on the log scale for 85% of users and the mean absolute relative change on the log scale is less than 2.5%

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 control variables.

Table 5. Descriptions of Independent Variables

| **Independent Variable** | | **Description** |
|---|---|---|
| $X_i/X_j$ | | **Attributes of sender $i$ / receiver $j$** |
| Network attributes | followees | Number of followees (out-degree) |
| | followers | Number of followers (in-degree) |
| | mutual | Number of mutual followers |
| Engagement levels | diggs | Total number of diggs |
| | comments | Total number of comments |
| | submissions | Total number of submissions |
| | avgDiggs | Average number of diggs per month |
| | avgComments | Average number of comments per month |
| | avgSubmissions | Average number of submissions per month |
| Others | gender | Male, female, or missing |
| | isSocial ($s_i$) | 1 if sender $i$ is a social source (i.e., followee), otherwise 0 |
| | isSubmitter | 1 if the sender is the submitter of the ad, otherwise 0 |
| $X_{ij}$ | | **Attributes of a sender-receiver dyad** |
| Dyadic network attributes | isMutual | Does 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 |
| $X_{ik}$ | | **Sender-specific attributes of an ad** |
| Sharing timing | wday | Day of a week when sender $i$ dugg ad $k$ |
| | hour | Hour of a day when sender $i$ dugg ad $k$ |
| | shareTime | Hours taken for sender $i$ to adopt since the creation of ad $k$, 0 for the front page |
| $X_{jk}$ | | **Receiver-specific attributes of an ad** |
| | co-senders | Number of followees (co-senders) of the receiver who have already shared |
| $X_k$ | | **Attributes of ads $k$ (only interaction with other variables can be identified)** |
| | popularity | Number of diggs on an ad 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 | 141 | 268.0 | 423.7 | 0 | 118 | 10122 |
| Number of followers | 146 | 386.3 | 1091.0 | 0 | 136 | 29331 |
| Number of mutual | 424 | 114.4 | 203.8 | 0 | 36 | 4598 |
| **Dyadic Network Attributes of Sender-Receiver Dyads** | | | | | | |
| isMutual (1– reciprocal, 0 – non-reciprocal) | 63733 | 0.27 | 0.44 | 0 | 0 | 1 |
| Number of common followees | 4736 | 41.1 | 52.5 | 0 | 23 | 814 |
| Number of common followers | 2182 | 100.7 | 334.2 | 0 | 26 | 9812 |
| Number of common mutual followers | 19805 | 17.1 | 35.3 | 0 | 4 | 594 |
| **Popularity of Ads** | | | | | | |
| Number of diggs | 0 | 93.4 | 86.2 | 4 | 95 | 295 |

Table 7 outlines the correlation among dyadic network characteristics. As discussed earlier, 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 embeddedness metrics are not very 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.

Table 7. 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 |

In order to understand how ads were shared over time, we plot the Kaplan-Meier survival curve for some sample content (see Figure W1 in the Web Appendix). Note that the sharing activities on most ads basically ceased at the censoring time. The sharing graphs for two sample ads with average popularity are shown in figure W2 in the Web appendix. These graphs demonstrate that path length is short (around 2 on average) for content as they propagate through the user network. This is in agreement with the observation made by Goel et al. (2015) about short path lengths for diffusion in online social networks. Note that our model assumes that the effect of co-senders can either increase or decrease. This may not accurately capture the aggregate diffusion pattern especially when the network is saturated and the effect of co-senders is very likely to decrease. However, path lengths for our data suggest that the network is not saturated and alleviates such concern. Next, we discuss our results on the role of embeddedness on the sharing propensity of the receiver.

## *RESULTS*

### Main Results

Table 8 summarizes the results of four model specifications.[11] Our main model of interest is model 4 that includes interaction terms representing the moderating effect of ad popularity on common followers and common mutual followers. We have also estimated models with no interaction terms or including only

[11] We omit the coefficients on control variables for clarity. Please see Appendix B for the complete set of parameter estimates.

one of the two interaction terms (Models 1-3, respectively). Likelihood ratio tests suggest that model 4 is preferred over models 2 and 3 ($p<0.05$). The following discussion is based on the estimates from model 4 unless otherwise specified.

Table 8. Parameters Estimates of Different Model Specifications

| | **Model1** | **Model2** | **Model3** | **Model4** |
|---|---|---|---|---|
| **Embeddedness** | | | | |
| logCommonFollowees | 0.23*** | 0.174** | 0.175** | 0.175** |
| logCommonFollowers | 0.845*** | 1.364*** | 0.829*** | 1.074*** |
| logCommonMutuals | -0.245*** | -0.19*** | 0.799*** | 0.418** |
| **Interactions with Popularity** | | | | |
| logCommonFollowers:logPopularity | | -0.153*** | | -0.071** |
| logCommonMutuals:logPopularity | | | -0.258*** | -0.16*** |
| **Fitness** | | | | |
| logLikelihood | -22661 | -22623 | -22620 | -22618 |
| AIC | 45401 | 45326 | 45320 | 45317 |

* Significance levels: p<0.001 (***), p<0.01 (**),p<0.05 (*), and p<0.1 (.). The main effect of logDiggNum cannot be identified as everyone sees the same digg number at a given time point, the effect of which is cancelled out in the likelihood. Model 2 is chosen as our main model based on fitness.

*Common followees*: The number of common followees has a positive effect on the sharing propensity of the receiver. This validates H1. The number of common followees reflects the similarity between the sender and the receiver's tastes and expertise. For the purpose of impression management, users tend to share content representing their taste or expertise (Berger 2014; Packard & Wooten, 2013). Thus, the more common followees the receiver has with the sender, the more likely the receiver will also share the content from the sender. 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 ads is low. This finding validates H2. As discussed earlier, the number of common followers reflects the similarity between the sender and receiver's audiences. Users tend to share content of interest to their audience to impress them (Aral and Van Alstyne 2011; Wu et al. 2004). Therefore, if the receiver has a similar audience with the sender, 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 findings that indicate 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. The existence of neighbors mutually connected to two users expands the bandwidth of communication between them and increases their trust in each other (Aral and Van Alstyne 2011; Burt 2001). The negative interaction of common mutual followers with popularity confirms H5. This finding shows a boundary condition for the positive effect of embeddedness previously reported in undirected networks (Aral and Walker 2014; Bapna et al. 2015). Specifically, the effect of common friends might be positive only when the information to be communicated is relatively novel (or not as popular).

In sum, all our proposed hypotheses find support from data. Our results show that the effect of embeddedness in directed networks varies across different types of "neighbors". Moreover, the impact of common followers and common mutual followers are negatively moderated by content novelty. The 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. 2015) and also observed in our dataset (Figures W1 and W2).

In addition to the findings on the three embeddedness metrics, 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 decreases with the number of co-senders (though the overall effect of all senders may increase). Second, the effect of shareTime is positive[12], suggesting that the later a co-sender shared, the stronger effect the co-sender has on the receiver. This documents a recency effect for co-senders consistent with previous findings that social effects decay over time (Bakshy et al. 2012; Haenlein 2013; Nitzan and Libai 2011; Trusov et al. 2009).

[12] It can be easily proved 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.

### Robustness Checks

*Unobserved Characteristics*. A potential concern with our analysis is that sharing of content could be driven by unobserved characteristics at the sender, the receiver, and even the dyad level. The dyadic observations with the same sender, receiver or dyad may not be independent because of common unobserved characteristics. As a robustness check, we consider sender-specific, receiver-specific and dyad-specific random effects. 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 Digg.

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

| | **none** | **rs** | **fs** | **rs-rr** | **fs-rr** | **rs-rr-rd** | **fs-rr-rd** |
|---|---|---|---|---|---|---|---|
| **Embeddedness** | | | | | | | |
| logCommonFollowees | 0.175** | 0.146. | 0.233*** | 0.2* | 0.165*** | 0.118 | 0.152* |
| logCommonFollowers | 1.074*** | 1.077*** | 1.078*** | 0.674*** | 0.658*** | 1.784*** | 0.861*** |
| logCommonMutuals | 0.418** | 0.364. | 0.193 | 0.679** | 0.361* | 1.023*** | 1.119*** |
| **Interactions with Popularity** | | | | | | | |
| logCommonFollowers:logPopularity | -0.071** | -0.068* | -0.07** | -0.054. | -0.082** | -0.166*** | -0.34*** |
| logCommonMutuals:logPopularity | -0.16*** | -0.153*** | -0.132*** | -0.178*** | -0.098* | -0.226** | -0.186** |
| **Fitness** | | | | | | | |
| logLikelihood | -22618 | -22538 | -22278 | -19974 | -19628 | -20226 | -19960 |
| AIC | 45317 | 45163 | 46727 | 40038 | 40932 | 40544 | 41597 |

* Note: 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. Therefore, "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" is the main model used in this paper. The model "none" doesn't include random or fixed effects on any subject.

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.

*The Growth of Network Structure*. Another concern with our analysis is that the network structure among users may change over time but we used a static snapshot. Note that Digg users often establish new ties but rarely break old ties. The direct consequence of the inaccurate network structure information is that the number of observed co-senders for a receiver could be larger or smaller than the actual number of co-senders for the receiver, depending on whether the receiver dugg the ad before or after the time her network information was collected by us. In our dataset, almost all the ads were posted on three days: May 24, June 1, and June 25. In order to test the sensitivity of our results to this issue, we split the dataset into two subsets: one focusing on ads created between May 24 and June 1, and another focusing on ads created on June 25. Recalling that the network structure is collected during June 7- June 26, the number of co-senders is likely to be overestimated on the first dataset as the network structure is collected afterwards. In the second dataset, the number of co-senders is likely to be underestimated as most of the digging activities take place after the network structure is collected. If overestimation or underestimation of the number of co-senders causes a substantial bias on our estimates, the results on these two subsets should be very different from that on the full dataset. Table 10 summarizes the results on the two subsets, respectively. The results show that the estimates on the two subsets are highly consistent with that on the full dataset.

Table 10. Parameters Estimates on the Two Subsets

| | **May24-June1 (1879 Events)** | **June 25 (931 Events)** |
|---|---|---|
| **Embeddedness** | | |
| logCommonFollowees | 0.233*** | 0.204* |
| logCommonFollowers | 0.858*** | 1.147*** |
| logCommonMutuals | 0.654*** | 0.643** |
| **Interactions with Popularity** | | |
| logCommonFollowers:logPopularity | -0.055. | -0.038 |
| logCommonMutuals:logPopularity | -0.186*** | -0.274*** |
| **Fitness** | | |
| logLikelihood | -15149 | -7427 |
| AIC | 30379 | 14935 |

*Inactive Users*. In our main analysis, we only consider active users as candidates for sharing. We also re-estimate our model by including data for users who have dugg an ad in the past but are not active during the panel period. Results are included in Table W1 in the Web Appendix and are qualitatively similar to our main analysis.

**Generalizability to Other Social Networks**

To test whether our findings generalize to other directed networks, we collected an additional dataset from Twitter. In the context of Twitter, the act of sharing is retweeting. Similar to Digg, the sharing is spontaneous if a user shares a tweet before any of her followees do. Otherwise, the sharing is considered as sharing influenced by others. To make sure that the content of the Twitter dataset is similar to that of the Digg dataset and also to improve the managerial relevance of our study, we focus on the sharing of brand-authored tweets.

We focus on nine brands listed by Fortune magazine as the top fortune 500 companies using social media.[13] We first collect the tweets authored (or retweeted in rare cases) by these brands in the past 10 days.[14] 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 all retweeters (including the author) of the tweet. These users represent the set of senders for the focal tweet. Next, we collected the social graph information for the followers (receivers) of the senders. Since the density and network size of Twitter users is much higher than that of Digg users[15], collecting data for all followers of every sender is not feasible due to API restrictions.[16] In order to control the data size, for every sender, we consider all followers who retweet. However, we randomly sample other followers from the sender's ego network using the risk set sampling approach (Langholz and Borgan 1995; Langholz and Goldstein 1996). Specifically, depending on popularity of each brand, we sample 5~20 followers from the ego network of each sender (sample size is smaller for popular brands with more data to collect).[17] We then collect the profile information for all the identified users. Similar to the Digg dataset, we focus on the receivers who are still active in the past three months. We focus on 4740 sharing activities on 74 tweets with more than 20 retweets in our analysis.[18] Further description and statistics for the Twitter dataset are shown in Tables W2-W5 in the Web Appendix. Table W6 shows the complete set of results for the Twitter dataset.

Table 11 summarizes the parameter estimates for the three embeddedness metrics for Twitter dataset. Our main model of interest is model 4 and has the best fit. The results show that the findings on the Twitter dataset are consistent with that on the Digg dataset. The coefficient of common followees is positive and

[13] http://fortune.com/2014/06/02/500-social-media/
[14] We collected two sets of tweets for each brand in about six weeks.
[15] In our sample, a user on Digg, on an average, has around 400 followers whereas a user on Twitter has around 19000 followers.
[16] https://dev.twitter.com/rest/public/rate-limits
[17] In order to ensure that the number of followers sampled does not affect our results, we tried to increase the sample size to as many as 50 followers for each retweeter on some brands and find the estimates are rather robust to the sample size.
[18] We also tried using tweet samples with other popularity thresholds, such as 5, 10, 30 and 40 and our results are qualitatively similar.

significant. The coefficients of common followers and common mutual followers are also positive and significant. And, the coefficients of the terms capturing interaction of these variables with popularity are negative and significant. This pattern of results demonstrates the generalizability of our findings from Digg to other directed social media platforms like Twitter. Unlike in the Digg dataset, however, we cannot effectively estimate random/fixed effects on the Twitter dataset as the reoccurrences of each sender, receiver, and dyad are substantially lower.

Table 11. Parameter Estimates on Twitter Dataset

| | **Model1** | **Model2** | **Model3** | **Model4** |
|---|---|---|---|---|
| **Embeddedness** | | | | |
| logCommonFollowees | 0.294*** | 0.311*** | 0.299*** | 0.309*** |
| logCommonFollowers | -0.127*** | 0.227*** | -0.132*** | 0.144*** |
| logCommonMutuals | -0.035 | 0.013 | 0.643*** | 0.284*** |
| **Interactions with Popularity** | | | | |
| logCommonFollowers:logPopularity | | -0.105*** | | -0.081*** |
| logCommonMutuals:logPopularity | | | -0.174*** | -0.075*** |
| **Fitness** | | | | |
| logLikelihood | -29378 | -29324 | -29340 | -29319 |
| AIC | 58822 | 58717 | 58747 | 58708 |

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. However, our model can effectively handle the case when only one of the co-senders has a significant impact. Therefore, this should not bias our estimates, especially given that only 7% of retweeters in our sample have more than one co-sender. 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.

## *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 information 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 information sharing are nascent and predominantly focus on undirected networks.

In this paper we study the effect of a dyad's network embeddedness on information 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 embeddedness in directed networks varies across different types of "neighbors". The number of common followees is positively associated with receiver's propensity of sharing. Other embeddedness 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 shared content. Thus, our study provides insight into consumer behavior in online information sharing and adds to the existing literature highlighting the role of uniqueness in social consumption (Cheema and Kaikati 2010; Zeng and Wei 2013). It is possible that uniqueness concerns may be preventing users from sharing the information received from others once the information becomes less novel. This in turn might be causing small cascades. Thus, our results provide a potential explanation for the relatively small size of information cascades that have been observed in online social networks (Goel et al. 2015)

We make 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.[19] 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; Trusov et al. 2010) or does not focus on the identification of the effect of shared characteristics (Sharara et al. 2011; Trusov et al. 2010).

[19] 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.

Our approach makes no such assumptions and, as a consequence, can help to better tease apart the effect of the 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[20], Twitter may want to consider how many common neighbors 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 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 sponsored ads and 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 type of content being shared. Further, we considered sponsored ads and brand-authored tweets. It is possible that the user behavior may be different for organic content. Future studies should investigate the role of content characteristics in moderating the effect of network attributes on information sharing. 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 latent model to capture the effect of impressions (Kang et al. 2013).  Finally, the assumption that the

[20] 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.

existence of one co-sender does not cannibalize or reinforce the effects of other co-senders is restrictive. In our analysis, we address this problem by allowing the hazard of a co-sender to change with the number of co-senders (i.e., shared followees in Table B.1 of Appendix B). 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." Department of Sociology, University of Pennsylvania.

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. 2011. "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science* (57:9), pp. 1623-1639.

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*: ACM, 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*: ACM, pp. 519-528.

Bapna, R., Gupta, A., Rice, S., and Sundararajan, A. 2015. "Trust, Reciprocity and the Strength of Social Ties: An Online Social Network Based Field Experiment," *Working Paper*.

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* (forthcoming).

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.

BIA/Kelsey. 2013. "U.S. Social Ad Revenues to Reach $11b in 2017."

Bock, G.-W., Zmud, R.W., Kim, Y.-G., and Lee, J.-N. 2005. "Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate," *MIS quarterly* (29:1), pp. 87-111.

Braun, M., and Moe, W.W. 2013. "Online Advertising Campaigns: Modeling the Effects of Multiple Ad Creatives," *Marketing Science* (32:5), pp. 753-767.

Brown, J.D. 1986. "Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments," *Social Cognition* (4:4), pp. 353-376.
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.
Camarero, C., and San José, R. 2011. "Social and Attitudinal Determinants of Viral Marketing Dynamics," *Computers in Human Behavior* (27:6), pp. 2292-2300.
Centola, D. 2010. "The Spread of Behavior in an Online Social Network Experiment," *Science* (329:5996), pp. 1194-1197.
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.
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.
Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets*. Cambridge University Press.
Goel, S., Anderson, A., Hofman, J., and Watts, D.J. 2015. "The Structural Virality of Online Diffusion," Management Science (forthcoming).
Granovetter, M. 1985. "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology* (91:3), pp. 481-510.
Granovetter, M., and Soong, R. 1986. "Threshold Models of Interpersonal Effects in Consumer Demand," *Journal of Economic Behavior & Organization* (7:1), pp. 83-99.
Granovetter, M.S. 1973. "The Strength of Weak Ties," *American Journal of Sociology* (78:6), pp. 1360-1380.
Grier, S.A., and Deshpandé, R. 2001. "Social Dimensions of Consumer Distinctiveness: The Influence of Social Status on Group Identity and Advertising Persuasion," *Journal of Marketing Research* (38:2), pp. 216-224.
Grubb, E.L., and Grathwohl, H.L. 1967. "Consumer Self-Concept, Symbolism and Market Behavior: A Theoretical Approach," *Journal of Marketing* (31:4), pp. 22-27.
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.
Hall, J.A., and Valente, T.W. 2007. "Adolescent Smoking Networks: The Effects of Influence and Selection on Future Smoking," *Addictive behaviors* (32:12), pp. 3054-3059.
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.
Jones, S.C. 1973. "Self-and Interpersonal Evaluations: Esteem Theories Versus Consistency Theories," *Psychological Bulletin* (79:3), p. 185.
Kang, J.-H., Lerman, K., and Getoor, L. 2013. "La-Lda: A Limited Attention Topic Model for Social Recommendation," in *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, pp. 211-220.
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.
Kempe, D., Kleinberg, J., and Tardos, E. 2003. "Maximizing the Spread of Influence through a Social Network," in: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. Washington, D.C.: ACM, pp. 137-146.
Korgaonkar, P.K., and Wolin, L.D. 1999. "A Multivariate Analysis of Web Usage," *Journal of advertising research* (39:2), pp. 53-68.

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.
Leadtail. 2013. "B2b Social Marketing Report: Marketing Executives on Twitter."
Leskovec, J., Adamic, L.A., and Huberman, B.A. 2007. "The Dynamics of Viral Marketing," *ACM Transactions on the Web* (1:1), p. 5.
Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), pp. 74-89.
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*: ACM, pp. 631-638.
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.
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*.
Nielsen. 2013. "Paid Social Media Advertising - Industry Update and Best Practices."
Nitzan, I., and Libai, B. 2011. "Social Effects on Customer Retention," *Journal of Marketing* (75:6), pp. 24-38.
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.
Richardson, M., and Domingos, P. 2002. "Mining Knowledge-Sharing Sites for Viral Marketing," in: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. Edmonton, Alberta, Canada: ACM, pp. 61-70.
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), 2014/08/01, 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*.
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.
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.
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.
Tajfel, H., and Turner, J.C. 1979. "An Integrative Theory of Intergroup Conflict," *The Social Psychology of Intergroup Relations* (33), p. 47.

Therneau, T.M. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer.
Toubia, O., Goldenberg, J., and Garcia, R. 2014. "Improving Penetration Forecasts Using Social Interactions Data," *Management Science* (60:12), forthcoming, pp. 3049-3066.
Travis, H. 2002. *Causes of Delinquency*. Transaction Publishers.
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.
Turner, J.C., and Oakes, P.J. 1986. "The Significance of the Social Identity Concept for Social Psychology with Reference to Individualism, Interactionism and Social Influence," *British Journal of Social Psychology* (25:3), pp. 237-252.
Uzzi, B. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness," *Administrative Science Quarterly* (42:1), pp. 35-67.
Volinsky, C.T., and Raftery, A.E. 2000. "Bayesian Information Criterion for Censored Survival Models," *Biometrics* (56:1), pp. 256-262.
Watts, D.J., and Dodds, P.S. 2007. "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research* (34:4), pp. 441-458.
Whiting, A., and Williams, D. 2013. "Why People Use Social Media: A Uses and Gratifications Approach," *Qualitative Market Research: An International Journal* (16:4), pp. 362-369.
Wiatrowski, M.D., Griswold, D.B., and Roberts, M.K. 1981. "Social Control Theory and Delinquency," *American Sociological Review* (46:5), pp. 525-541.
Wu, F., Huberman, B.A., Adamic, L.A., and Tyler, J.R. 2004. "Information Flow in Social Groups," *Physica A: Statistical Mechanics and its Applications* (337:1), pp. 327-335.
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.

## *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. (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.[21] 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. The prefix "r" indicates covariates on the receiver side. Enclosed in parentheses are the standard deviations of the relative errors.

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) |

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

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

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 average number of co-senders on a receiver.

## *APPENDIX B: COMPLETE RESULTS*

Table B1. Complete Parameter Estimates for Models in Table 7

| | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| **Characteristics of Sender** | | | | |
| isSocialTRUE | 1.409** | 1.577*** | 1.602*** | 1.482** |
| isDiggAdsTRUE | 0.36 | -0.434 | 0.277 | 0.005 |
| logFollowees | -0.003 | 0.016 | -0.015 | -0.004 |
| logFollowers | -0.872*** | -0.759*** | -0.827*** | -0.798*** |
| logMutuals | 0.002 | 0.011 | 0.026 | 0.028 |
| logDiggs | -0.32* | -0.285* | -0.269. | -0.239 |
| logComments | -0.191. | -0.037 | -0.183. | -0.155 |
| logSubmissions | -0.009 | -0.148 | -0.005 | -0.069 |
| logAvgDiggs | 0.585*** | 0.445*** | 0.444*** | 0.417*** |
| logAvgComments | 0.008 | -0.105 | 0.025 | -0.001 |
| logAvgSubmissions | 0.04 | 0.121 | 0.038 | 0.081 |
| genderf | 0.222 | -0.047 | -0.096 | -0.089 |
| genderm | 0.274* | 0.274* | 0.209. | 0.231. |
| **Characteristics of Receiver** | | | | |
| logFollowees | -0.238*** | -0.248*** | -0.246*** | -0.248*** |
| logFollowers | -0.208*** | -0.206*** | -0.211*** | -0.209*** |
| logMutuals | -0.101*** | -0.109*** | -0.102*** | -0.104*** |
| logDiggs | 0.166*** | 0.169*** | 0.171*** | 0.17*** |
| logComments | -0.156*** | -0.16*** | -0.163*** | -0.162*** |
| logSubmissions | -0.152*** | -0.153*** | -0.155*** | -0.154*** |
| logAvgDiggs | 0.438*** | 0.435*** | 0.435*** | 0.435*** |
| logAvgComments | 0.218*** | 0.217*** | 0.224*** | 0.221*** |
| logAvgSubmissions | 0.029 | 0.031 | 0.031 | 0.03 |
| genderf | 0.12** | 0.129** | 0.129** | 0.13** |
| genderm | 0.137*** | 0.143*** | 0.143*** | 0.144*** |
| **Sharing Timing of Sender** | | | | |
| wday1 | -0.462** | -0.324* | -0.448** | -0.382* |
| wday2 | 0.119 | -0.117 | -0.11 | -0.109 |
| wday3 | 0.263* | 0.073 | 0.122 | 0.088 |
| wday4 | 0.129 | 0.048 | 0.045 | 0.039 |
| wday5 | -0.103 | 0.03 | -0.002 | 0.021 |
| wday6 | 0.117 | 0.265 | 0.004 | 0.148 |
| hour(5,11] | -0.348** | -0.281* | -0.259* | -0.268* |
| hour(11,17] | -0.256* | -0.165 | -0.196 | -0.179 |
| hour(17,23] | 0.044 | 0.006 | 0.003 | -0.002 |
| shareTime | -0.023 | 0.099** | 0.097** | 0.116** |
| **Number of Co-senders** | | | | |
| co-senders | -0.082*** | -0.057*** | -0.058*** | -0.056*** |
| **Dyadic Characteristics** | | | | |
| isMutualTrue | -0.645*** | -0.499*** | -0.526*** | -0.5*** |
| logCommonFollowees | 0.23*** | 0.174** | 0.175** | 0.175** |
| logCommonFollowers | 0.845*** | 1.364*** | 0.829*** | 1.074*** |
| logCommonMutuals | -0.245*** | -0.19*** | 0.799*** | 0.418** |
| logCommonFollowers:logPopularity | | -0.153*** | | -0.071** |
| logCommonMutuals:logPopularity | | | -0.258*** | -0.16*** |
| **Fitness** | | | | |
| logLikelihood | -22661 | -22623 | -22620 | -22618 |
| AIC | 45401 | 45326 | 45320 | 45317 |

* The three levels for gender are: m – male, f – female, and u – unknown. For wday, Monday is coded as 0. Hour of a day is grouped into four bins. For dummy variables, the missing levels are the reference levels.

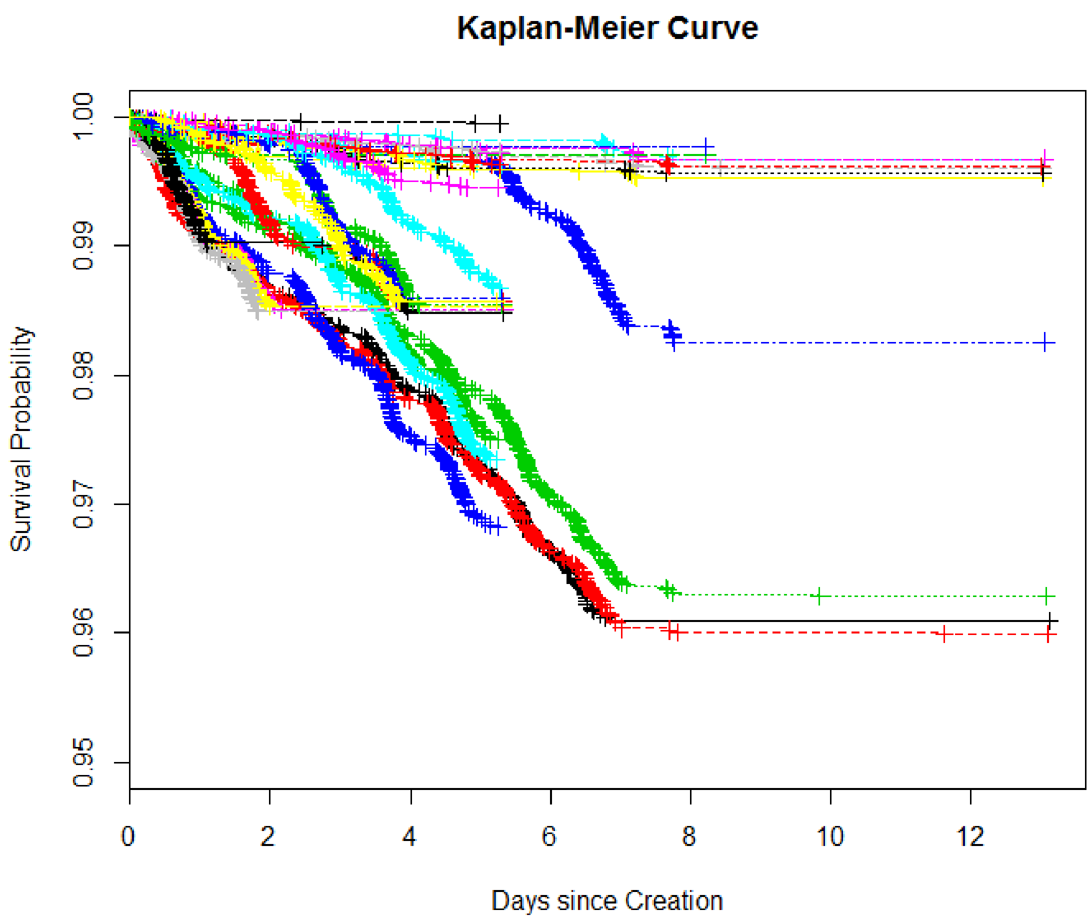


Figure W1. Kaplan-Meier Survival Curve for Digg Ads[22]

[22] The KM curve is computed based on the average survival probability of all receivers who are at risk over time

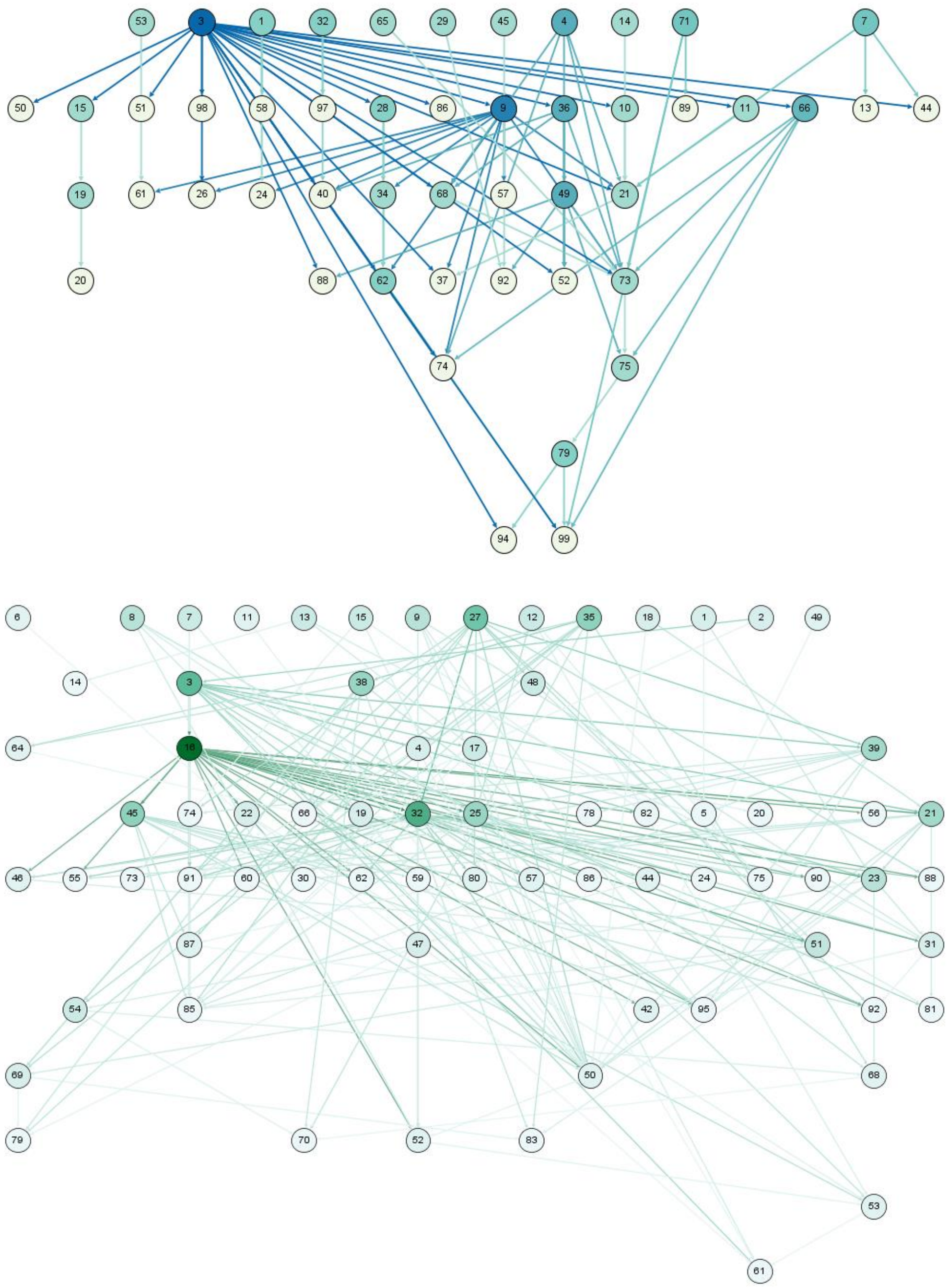


Figure W2. Sharing Graphs for Ads 1 & 2.[23]

[23] Arrow represents information flow. Nodes without incoming links share spontaneously. Nodes are labeled based on the order they share the ad. The darkness of the color of a node is proportional to her outgoing links in the graph. The color of an arrow is consistent with the source node.

Table W1. Parameter Estimates on the Digg Dataset (including inactive users)

| | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| **Embeddedness** | | | | |
| logCommonFollowees | 0.257*** | 0.200*** | 0.208*** | 0.205*** |
| logCommonFollowers | 0.723*** | 1.275*** | 0.710*** | 1.043*** |
| logCommonMutuals | -0.227*** | -0.179*** | 0.754*** | 0.290** |
| **Interactions with Popularity** | | | | |
| logCommonFollowers:logPopularity | | -0.152*** | | -0.091*** |
| logCommonMutuals:logPopularity | | | -0.237*** | -0.121*** |
| **Fitness** | | | | |
| logLikelihood | -23887 | -23843 | -23845 | -23839 |
| AIC | 47844 | 47757 | 47762 | 47752 |

Table W2. Summary Statistics for Twitter Dataset

| | |
|---|---|
| Number of ads/tweets | 74 |
| Number of sharing users (senders) | 4,209 |
| Number of potential sharing users (receivers) | 36,187 |
| Number of <sender, receiver> dyads | 90,288 |
| Number of <sender, receiver, ad> tuples | 171,685 |
| Number of spontaneous tuples | 80,721 (47%) |
| Number of social tuples | 90,964 (53%) |
| Number of shares (retweets) | 4,740 |
| Number of spontaneous shares | 1,020 (21.5%) |
| Number of potential influenced shares | 3,720 (78.5%) |
| Percentage with more than one co-senders (excluding special sender) | 6.8% |

Table W3. Descriptions of Independent Variables for Twitter Dataset

| **Independent Variable** | | **Description** |
|---|---|---|
| $\boldsymbol{X_i/X_j}$ | | **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 |
| | lists | Number of lists subscribed |
| Engagement levels | statuses | Total number of tweets, including retweets |
| | favourites | Total number of favourites |
| 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 | Does 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 $\boldsymbol{k}$, 0 for the front |
| $\boldsymbol{X_{jk}}$ | | **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 $\boldsymbol{k}$ (only interaction with other variables can be identified)** |
| | popularity | Number of retweets at a given time point |

Table W4. Key Statistics of Main Variables for Twitter Dataset

| | **Zeros** | **Mean** | **SD** | **Min** | **Median** | **Max** |
|---|---|---|---|---|---|---|
| **Unitary Network Attributes of All Users** | | | | | | |
| Number of followees | 14 | 9298.8 | 43743.3 | 0 | 769 | 2422154 |
| Number of followers | 202 | 18843.9 | 348931.6 | 0 | 380 | 59159316 |
| Number of mutuals | 945 | 6278.0 | 31334.0 | 0 | 212 | 1755611 |
| **Dyadic Network Attributes of Sender-Receiver Dyads** | | | | | | |
| isMutual (1– reciprocal, 0 – non-reciprocal) | 40940 | 0.45 | 0.50 | 0 | 0 | 1 |
| Number of common followees | 9596 | 61.1 | 715.5 | 0 | 10 | 79376 |
| Number of common followers | 16015 | 235.3 | 6016.5 | 0 | 5 | 500406 |
| Number of common mutual followers | 26864 | 74.5 | 490.5 | 0 | 2 | 35478 |
| **Popularity of Tweets** | | | | | | |
| Number of retweets | 15 | 20.5 | 41.9 | 0 | 10 | 379 |

Table W5. Correlation among Dyadic Network Characteristics for Twitter Dataset

| | isMutual | logCommonFollowees | logCommonFollowers | logCommonMutuals |
|---|---|---|---|---|
| isMutual | 1.00 | 0.10 | 0.12 | 0.50 |
| logCommonFollowees | | 1.00 | 0.56 | 0.53 |
| logCommonFollowers | | | 1.00 | 0.63 |
| logCommonMutuals | | | | 1.00 |

Table W6. Complete Results on the Twitter Dataset

| | **Model1** | **Model2** | **Model3** | **Model4** |
|---|---|---|---|---|
| **Characteristics of Sender** | | | | |
| isSocialTRUE | -1.575 | -0.762 | -0.742 | -0.633 |
| isAuthorTRUE | 1.459*** | 1.069*** | 1.23*** | 1.076*** |
| logFollowees | -0.764*** | -0.938*** | -0.97*** | -0.997*** |
| logFollowers | 0.399*** | 0.416*** | 0.412*** | 0.421*** |
| logMutuals | 0.811*** | 0.99*** | 0.995*** | 1.042*** |
| logLists | 0.216* | 0.161* | 0.176** | 0.154* |
| logStatuses | -0.142** | -0.147*** | -0.119** | -0.136*** |
| logFavourites | -0.478*** | -0.463*** | -0.467*** | -0.463*** |
| verified | -1.392* | -1.449*** | -1.771*** | -1.604*** |
| logRegMon | 0.482*** | 0.463*** | 0.416*** | 0.432*** |
| **Characteristics of Receiver** | | | | |
| logFollowees | -0.468*** | -0.482*** | -0.473*** | -0.48*** |
| logFollowers | -0.139*** | -0.145*** | -0.143*** | -0.146*** |
| logMutuals | -0.049** | -0.038* | -0.047** | -0.039* |
| logLists | 0.099*** | 0.086*** | 0.091*** | 0.086*** |
| logStatuses | 0.278*** | 0.284*** | 0.281*** | 0.283*** |
| logFavourites | 0.081*** | 0.084*** | 0.084*** | 0.085*** |
| verified | -0.525* | -0.702** | -0.685** | -0.772*** |
| logRegMon | -0.165*** | -0.163*** | -0.164*** | -0.163*** |
| **Sharing Timing of Sender** | | | | |
| wday0 | -1.233 | -0.976 | -0.92 | -0.936 |
| wday1 | -0.049 | -0.147 | -0.177 | -0.186 |
| wday2 | -0.616 | -0.5 | -0.538 | -0.491 |
| wday3 | 2.106** | 1.782** | 1.753** | 1.721** |
| wday5 | 1.387* | 1.266** | 1.468*** | 1.342** |
| wday6 | 0.23 | -0.575 | -0.568 | -0.579 |
| hour(5.75,11.5] | -2.347*** | -2.449*** | -2.479*** | -2.513*** |
| hour(11.5,17.2] | -1.829*** | -1.476*** | -1.559*** | -1.458*** |
| hour(17.2,23] | -0.253 | -0.274 | -0.223 | -0.249 |
| shareTime | 0.054*** | 0.05*** | 0.051*** | 0.05*** |
| **Number of Co-senders** | | | | |
| co-senders | -1.865*** | -2.072*** | -1.993*** | -2.077*** |
| **Dyadic Characteristics** | | | | |
| isMutualTRUE | 0.162* | 0.146. | 0.164* | 0.149* |
| logCommonFollowees | 0.294*** | 0.311*** | 0.299*** | 0.309*** |
| logCommonFollowers | -0.127*** | 0.227*** | -0.132*** | 0.144*** |
| logCommonMutuals | -0.035 | 0.013 | 0.643*** | 0.284*** |
| logCommonFollowers:logPopularity | | -0.105*** | | -0.081*** |
| logCommonMutuals:logPopularity | | | -0.174*** | -0.075*** |
| **Fitness** | | | | |
| likelihood | -29378 | -29324 | -29340 | -29319 |
| AIC | 58822 | 58717 | 58747 | 58708 |

* For wday, Monday is coded as 0. Hour of a day is grouped into four bins. For dummy variables, the missing levels are the reference levels.