Telematics and Informatics 62 (2021) 101623

Contents lists available at ScienceDirect

Telematics and Informatics

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

Decreasing social contagion effects in diffusion cascades: Modeling message spreading on social media

Hai Liang

School of Journalism and Communication, The Chinese University of Hong Kong, HKSAR, China

ARTICLE INFO

Keywords:
Information diffusion
Peer influence
Social contagion
Diffusion cascades
Homophily
Social network

ABSTRACT

Modeling retweeting behaviors is important for understanding and predicting how information spreads on social media platforms. The present study contributes to the literature by examining the decreasing social contagion and increasing homophily effects with the depth of diffusion cascades. To test the hypotheses, the study proposes a matching-on-followers method by combining choice and cascade models. More specifically, the study examines the impacts of interaction frequency, multiple exposures, and interest similarity between parent users and potential retweeters on the likelihood of retweeting. The study also incorporates the depth of diffusion cascades and network structures into the model. By using a random sample of original tweets, their retweets, and potential retweeters ($N = 87{,}139$), the study found that cascade depth is negatively associated with social contagion effects (interaction and multiple exposures) and positively associated with the effect of interest similarity on message sharing. These results indicate that influence-based and homophily-driven diffusion operate differently in cascades with different diffusion structures.

1. Introduction

Information diffusion on social media has been modeled in various ways to satisfy different application needs, such as influence maximization in viral marketing (Agarwal and Mehta, 2020; Zhang et al., 2019), detecting political events (Ansah et al., 2018), and misinformation control (e.g., Liu et al., 2019). Two major analytic frameworks have been developed for detecting and explaining information diffusion patterns along social ties. One is to analyze individual message-sharing decisions based on choice models like logistic/hazard regression models (e.g., An et al., 2014; Aral and Walker, 2014) and machine learning models (e.g., Chen and Deng, 2020). The other is to analyze diffusion structures by tracking diffusion cascades (e.g., Goel et al., 2016; Vosoughi et al., 2018). Both approaches provide valuable insights for understanding message diffusion mechanisms, and mathematical models, such as independent cascade, linear threshold, and susceptible-infected models have been developed based on the empirical findings of these approaches (e.g., Liu et al., 2015; Zhan et al., 2018; Zhang et al., 2016). However, few empirical studies have tried to integrate the two approaches.

Choice models are typically user-centric in formulating message diffusion as a series of independent decisions regarding message sharing: whether and under what conditions a focal user will share a given message from another user. Under this framework, one of the most important explanations of diffusion is the social contagion theory, which argues that communication between peers could serve as a mechanism that exposes individuals to messages (Burt, 1980, 1987). Multiple exposures can increase the likelihood of

E-mail address: hailiang@cuhk.edu.hk.

https://doi.org/10.1016/j.tele.2021.101623
Received 7 December 2020; Received in revised form 24 February 2021; Accepted 31 March 2021
Available online 7 April 2021

message sharing (Centola and Macy, 2007; Marin et al., 2020; Romero et al., 2011), so the likelihood of sharing between two users is a function of social interaction and the number of exposures. However, this kind of analysis generally neglects the structural dynamics of diffusion paths. Messages could spread in multiple steps and follow different structural patterns; because the choices within a diffusion cascade are interdependent, contagion effects could be influenced by diffusion structures.

As a complementary approach, cascade models are typically message-centric in formulating message diffusion as a series of diffusion paths of a message: how and with what structural characteristics a given message spreads in a social network. In cascade models, a diffusion cascade is the collection of diffusion paths over which a given message spreads within a social network. Although this approach is convenient for analyzing diffusion structures like the differences between broadcast and viral diffusion models (Goel et al., 2016; Zhang et al., 2020), its analyses are based solely on sharing actions (e.g., retweets), making it difficult to draw multivariate inferences due to the lack of unshared cases.

To solve this problem, the present study proposes a new analytical framework by combining user-centric choice models with message-centric cascade analytics, which also makes it feasible to examine how social contagion effects could vary in different diffusion structures. In particular, the study tests the moderating role of cascade depth (how many steps a message spreads in social networks) on social contagion effects in the diffusion process.

1.1. *Contagion effects in information diffusion*

In social network analysis, social contagion has been used interchangeably with peer influence, which is characterized by similarity driven by influence and transmitted through peer connections (e.g., Aral and Walker, 2011). Social contagion effects in information diffusion involve two major factors: interaction and number of exposures. The interaction hypothesis states that whether a given individual shares a message from another individual depends on the social interactions between those two people. Social contacts and communications can make individuals socially proximate and thus increase peer influences (Burt, 1987). In empirical studies, the frequency of interaction is also a proxy for tie strength. Bakshy et al. (2012) found that tie strength, as measured by frequency of both online and offline interactions, was positively associated with an individual's probability of sharing a Facebook link if their friends had previously shared that link. Liang and Fu (2019) found that tie strength, measured as mentions and replies, was positively associated with the retweeting probability between users. Therefore, we posit this hypothesis as a starting point for analysis in the next section:

H1: Social media users are more inclined to share a message if they have interacted more frequently with the author or sharers (e.g., retweeters on Twitter) of that message.

The exposure hypothesis states that whether an individual shares a message depends on the number of exposures to multiple sources. Contagions based on exposures can be either simple or complex; simple contagion refers to situations where a single activated source is sufficient for transmission, as with the spread of infectious diseases. However, many social behaviors are costly, unfamiliar, or controversial, requiring social affirmation or reinforcement from multiple sources to spread (Centola and Macy, 2007). Therefore, the complex contagion principle posits that successful transmission of these behaviors depends on interaction with multiple contacts rather than the frequency of interaction with a single contact. Many studies have equated the impact of the number of exposures with the social contagion effect (e.g., Aral et al., 2009; Aral and Walker, 2014; Marin et al., 2020; Ugander et al., 2012).

Although the complex contagion theory was originally proposed to explain the diffusion of costly collective behaviors, it has been widely applied to the study of message diffusion on social media platforms. For example, repeated exposures on Twitter can increase the probability of users mentioning specific hashtags (Romero et al., 2011) or URLs (Hodas and Lerman, 2015). However, studies have also found that this relationship can vary across topics (Romero et al., 2011) and the number of a user's followees (Hodas and Lerman, 2015). One reason cited is that information diffusion on social media differs from other collective behaviors like the adoption of innovations. Normally, online message sharing is neither costly nor risky (Guilbeault et al., 2018), so social reinforcement in this situation is much less important. Furthermore, repeated exposures indicate a high level of information redundancy, which can inhibit message sharing on social media (Liang and Fu, 2019). Given these contradictory predictions, the present study poses the following research question:

RQ: Are social media users more inclined to share a message if they are exposed to more online friends who shared that message?

Social contagions and homophilous diffusion are generically confounded in social networks (Shalizi and Thomas, 2011). The degree to which social media friends are more similar makes them more likely to have a similar strength of preference for sharing the same message even if they do not influence one another (Aral et al., 2009). Homophily – in which individuals are more inclined to interact with others who share similar backgrounds and tastes – should be carefully controlled when estimating contagion effects. Homophily in information diffusion creates a selection bias because neither interactions nor exposures are randomly assigned: users are more likely to be exposed and contacted because of their similarity with their friends. As to information diffusion on social media specifically, homophily can manifest itself in two ways.

The first occurs when sharers of a message are similar to one another in terms of social attributes, interests, tastes, and so on: sharing the same messages may reflect users' similar preferences. Aral et al. (2009) found that the adopters of a mobile service application were more like their adopter friends than their non-adopter friends. By conditioning matches on a vector of observable characteristics, Aral et al. (2009) found that the impact of the number of exposures on adoption was largely overestimated compared to the random matching method. Matching on observed variables has two disadvantages for studying information diffusion on social media. First, the method assumes that the observed covariates carry all latent homophily effects; otherwise, unobserved variables could bias the estimates (Shalizi and Thomas, 2011). Therefore, the estimated social contagion effect is the upper bound. Second, user attributes like demographics are usually unavailable on social media. Nevertheless, one of the observable variables could be easily controlled for on most social media platforms. Similar users are more likely to post on similar topics (reflecting their interest similarity) and thus more likely to share the same message (Chen and Deng, 2020; Hu et al., 2018). For example, the semantic similarity between users' tweets is positively associated with the likelihood of retweeting between those users (Liang and Fu, 2019). The present study uses the following hypothesis to re-examine this relationship:

H2: Message sharing is more likely between users who post similar topics.

Second, sharers of the same message are usually well connected in social networks. Individuals tend to form closer relationships with similar peers: similarity breeds connections (McPherson et al., 2001). Therefore, controlling for network structures is another way to control for homophily (Weng et al., 2013), which is particularly useful when observed attributes are lacking. By controlling for the network structure of adopter friends, Ugander et al. (2012) found that the number of exposures was negatively associated with the likelihood of adoption. In an adoption study, Aral and Walker (2014) found that network embeddedness, as measured by the number of shared friends, was positively associated with social contagion. Researchers also incorporated homophily and network structures to measure peer influence on social media (Li et al., 2020).

Empirical studies have also confirmed the role of network structures in information diffusion. For example, users are more likely to share messages from others with reciprocal ties and those who share more followees and followers (An et al., 2014). By using representative ego networks, Liang and Fu (2019) found that structural redundancy, a measure of the proportion of shared and mediated followees between two users, was positively related to retweeting between those two users. Network redundancy is used as a measure of the level of redundant contacts. In a structurally redundant ego network, neighbors are themselves tightly connected (Burt, 1992). In this sense, the measure of structural redundancy is empirically consistent with structural diversity in Ugander et al. (2012) and embeddedness in Aral and Walker (2014).

Although network structures are related to homophily, structural factors have also been interpreted as social contagions. As Granovetter (1973, 1983) argues, strong ties usually exist in triads, and shared friends in tightly connected networks can either induce or reflect strong ties. Aral and Walker (2014) argue that embeddedness is likely to conduct greater peer influence because it increases the level of trust between embedded peers. Liang and Fu (2019) also report that structural redundancy is correlated with both tie strength and information similarity. To avoid this controversy, the present study treats structural factors as control variables to estimate social contagion effects.

1.2. Contagion effects in diffusion cascades

Previous analyses of diffusion cascades (or diffusion networks) are generally descriptive. A diffusion path is a chain of sharing actions; a message can spread through different intermediaries to reach other individuals through multiple paths. A collection of diffusion paths is called a diffusion cascade. In a diffusion network, nodes are the individuals who share specific information, while the edges are information paths over which individuals transmit that information to others. Diffusion cascade analysis is often used to provide a quantitative differentiation between viral and broadcast diffusion structures (Goel et al., 2016). A broadcast structure indicates that all people share the message directly from the seed user. On the other hand, diffusions following the viral model have many intermediaries, and their diffusion trees are composed of numerous person-to-person diffusion paths. Cascade depth is the number of steps from the seed user (source) that the information has spread. A large depth value suggests a long chain of information diffusion and thus implies viral spreading. Under this framework, empirical studies have repeatedly found that the broadcast diffusion model, which features short-depth diffusion trees, is dominant in many online systems like Twitter, Facebook, Digg, and Weibo (see a review by Zhang et al., 2016). Nevertheless, it remains meaningful to study the dynamics of viral diffusion. For example, recent studies have used diffusion structures to characterize the spreading of rumors and fake news on Twitter and Facebook (e.g., Del Vicario et al., 2016; Liu et al., 2019; Vosoughi et al., 2018). In addition, viral diffusion might be associated with larger cascade size (see Figure s1) and person-to-person diffusion could be more pervasive in some situations (Anderson et al., 2015).

Beyond descriptive analysis, Liang (2018) found that the probability of retweeting between different political ideologies was positively associated with cascade depth. The tentative explanation provided in Liang (2018) implies that the increase in cross-ideological retweeting over cascade depth might be caused by a decrease in peer influence and contagion effects with cascade depth. Sharers at deeper steps in the diffusion cascades are less susceptible to normative pressures to share their friends' posts. They might be socially distant from the seed users (e.g., unfamiliar or not directly connected with them) and feel less obliged to share friends' reposts than original posts. Given the social contagion effects stated in *H1* and *RQ*, the present study explicitly examines decreasing contagion effects, using the following hypothesis:

H3: The contagion effects of (a) social interaction and (b) the number of exposures on message sharing (*H1 & RQ*) are negatively associated with cascade depth.

Relational factors are not the only reason for message sharing; informational factors are also relevant (Liang and Fu, 2017; Xu et al., 2013). As cascade depth increases, individuals will rely less on relational factors like peer influence and social contagion. Instead, the sharing decision will be more likely to rely on message content. Users may share a given message simply because they are genuinely interested in that message, even if they are subject to peer influence. An et al. (2014) found that people who are interested in a given topic are more inclined to retweet items with which they disagree than those with which they agree, implying that retweeters may rely more on the content of a message when peer influence plays a less important role.

This phenomenon is also related to the homophilous diffusion explanation. Anderson et al. (2015) found that edgewise homophily (interest similarity between dyads) was insufficient to explain cascade-level homophily. Instead, new adopters' attributes were governed not only by their parents but also by their parents' parents in the diffusion cascade, suggesting that homophily could have a cumulative effect along the diffusion tree and that homophily effects in diffusion cascades might thus increase with cascade depth. Taken together, these observations indicate that sharing based on interest similarity will increase with cascade depth. Therefore, we propose the following hypothesis:

H4: The interest similarity effect on message sharing stated in *H2* is positively associated with cascade depth.

2. Method

2.1. Analytical framework: matching on followers

When a message is initially posted on a social media platform (Twitter in this study), the main audience of the message is the followers of the seed user (Chen and Deng, 2020; Myers et al., 2012). Although we assume that most followers will read the message, it is unlikely that all viewers will eventually retweet it. As diffusion continues, it is important to know what kinds of followers will be "selected" to retweet in the next step. In other words, it is a matter of selecting actual retweeters from all potential retweeters. Comparing the differences between retweeters and non-retweeters is a plausible approach to identifying the important factors associated with message diffusion. This modeling strategy is widely known as the conditional logistic regression (Hosmer et al., 2013), but it ignores the cascade dynamics that appear as a message spreads to the seed user's followers, changing the potential pool of retweeters to followers' followers. If a message spreads in multiple steps, the potential retweeters will change accordingly. If the diffusion follows a broadcast model, it reduces to a conditional logistic problem.

To address this challenge, the present study proposes a matching-on-followers method (Fig. 1). The black nodes in Fig. 1 are the retweeters who retweeted the original message (1–8) while the gray nodes (*a*–*f*) are non-retweeters. The edges indicate the following relationships (e.g., nodes 8 and *f* are both followers of node 7). Edges between black nodes also indicate retweeting actions (e.g., node 8 retweeted from node 7). Taken together, the black nodes and the edges between them constitute a diffusion cascade. The number of generations indicates cascade depth, and the nodes between retweeters and seed users are the intermediaries (e.g., node 2). The seed user and intermediaries are the parent users of their retweeters, who retweet the message directly from them. For example, node 3 is the parent of nodes 6, 7, and *d*, while the parent of node 3 is the seed user *S*.

The proposed approach seeks to match retweeters with non-retweeters randomly selected from potential retweeters at different diffusion steps. Since sharing on social media platforms is usually based on follower-followee networks (Chen and Deng, 2020; Myers et al., 2012), it is reasonable to assume that all followers of the retweeters in a given cascade are themselves potential retweeters. For example, the followers of the seed user in Fig. 1 are the potential retweeters in step 1, and the followers of nodes 1–4 are the potential retweeters in step 2 (nodes 5–*e*). All potential retweeters are exposed to the original message at least once and thus have the chance to retweet it. Typically, most potential retweeters do not retweet the message. A central goal of the present study is to examine the differences between retweeters and non-retweeters. Given this, a sample of non-retweeters could be randomly selected from potential retweeters to match with the retweeters. Given that there are different parents (intermediaries) in a diffusion cascade, retweeters should be matched separately by parent. Therefore, a given parent node's retweeters were matched with its followers who did not retweet. As illustrated in Fig. 1, nodes 6, 7, and *d* together form a matched set at step 2, while retweeter 5 has no matched non-retweeters.

Comparing the differences in characteristics between retweeters and matched non-retweeters within each matched set enables us to test the relative effects of different variables on sharing probability between parents and potential retweeters. Therefore, *H1* and *H2* can be formally tested. Beyond a simple logistic framework, this comparison can be conducted at different steps in cascades, allowing

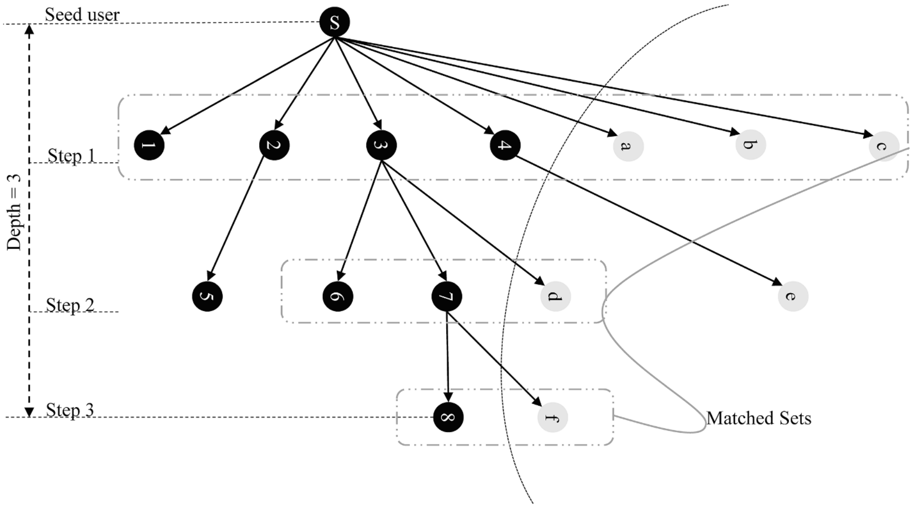


Fig. 1. Matching on followers in diffusion cascades. The edges indicate following relationships. The black nodes are retweeters, and the collection of all black nodes and the edges between them represent a diffusion cascade. The gray nodes are non-retweeters. The retweeters and non-retweeters of the same parent user comprise a matched set. In this figure, retweeter 8 and non-retweeter f are both parent 7′s followers and thus form a matched set. In each step and for different parent users, different matched sets were constructed.

for cascade depth to be included as a moderator to test the conditional effects of social contagion and homophily on sharing probability (*H3* and *H4*).

The proposed approach has two advantages. First, it combines choice and cascading models. Instead of using a conditional logistic regression that does not allow within-strata constant variables (here, cascade depth), the present study uses a multilevel logistic model to incorporate cascade depth. More detail is introduced in the section below on statistical models. Second, matching-on-followers may control for latent structural and homophily effects. Since comparisons are conducted within the matched set and the treated (retweeters) and control (non-retweeters) users are exposed to the same parent user, the impact of following relationships and network structures is generally minimalized. Given that homophily effects are intertwined with the formation of following relationships (McPherson et al., 2001), we should be able to control for some latent homophily effects related to the formation of following relationships.

2.2. Data collection

Data were collected from Twitter using its public application programming interfaces (APIs). First, in order to collect a random sample of tweets, Twitter's streaming API was used to download random tweets between February 21 and 28, 2019; a total of 18,382,174 tweets (including retweets and replies) were collected, of which 909,375 were original tweets. In order to collect all the

Table 1

Multilevel Logistic Regression Coefficients Predicting Retweeting in Diffusion Cascades.

	Model I	Model II	Model III
	Fixed Effects (Log-Odds with SE)		
Intercept	−2.78(0.03)**	−3.00(0.04)**	−3.48(0.05)**
Interaction frequency	1.28(0.03)**	1.19(0.02)**	0.94(0.02)**
Number of exposures	−1.41(0.05)**	−1.53(0.05)**	−1.40(0.05)**
Interest similarity	1.07(0.04)**	1.08(0.04)**	0.72(0.03)**
Cascade depth	−0.51(0.03)**	−0.29(0.03)**	−0.27(0.03)**
Time elapsed	0.15(0.02)**	0.15(0.02)**	0.16(0.02)**
Interaction frequency × depth		−0.35(0.02)**	−0.26(0.02)**
Number of exposures × depth		1.03(0.04)**	0.88(0.04)**
Interest similarity × depth		0.22(0.03)**	0.16(0.03)**
Structural Factors			
Reciprocity			−0.12(0.04)*
Structural redundancy			0.05(0.02)*
Retweeter Attributes			
Retweeting inertia			0.82(0.02)**
Number of followees			−0.74(0.03)**
Number of followers			−0.02 (0.04)
Number of statuses			1.53(0.03)**
Account age			−0.28(0.02)**
Parent Attributes			
Number of followees			−0.08(0.03)*
Number of followers			0.86(0.03)**
Number of statuses			−0.53(0.03)**
Account age			0.11(0.02)**
Group Means			
Interaction (parent level)	0.17(0.05)*	0.10(0.05)	−0.38(0.06)**
Exposure (parent level)	0.25(0.06)**	0.16(0.06)	0.33(0.06)**
Similarity (parent level)	−0.98(0.07)**	−0.91(0.07)**	−0.64(0.07)**
Interaction (cascade level)	−0.14(0.10)	−0.26(0.10)*	−0.26(0.10)*
Exposure (cascade level)	0.79(0.08)**	0.76(0.08)**	0.48(0.09)**
Similarity (cascade level)	−0.26(0.08)*	−0.22(0.09)*	−0.21(0.09)
	Random Effects (Variance)		
σ^2 – Residual variance	3.29	3.29	3.29
τ – Cascade level/parent level			
Intercept	0.10/0.13	0.14/0.23	0.23/0.06
Interaction frequency	0.07/0.36	0.07/0.28	0.07/0.12
Number of exposures	0.30/1.39	0.32/0.64	0.30/0.50
Interest similarity	0.49/0.21	0.38/0.20	0.23/0.05
	Model Summary		
Marginal R^2	55.3%	59.7%	73.9%
Conditional R^2	63.2%	65.7%	75.9%
AIC	50,062.23	48,744.03	39,499.41
Sample size	967 cascades/5,028 parents/87,139 observations		

Note. All predictors except reciprocity were standardized to have a mean of 0 and a standard deviation of 1. Count variables (interaction frequency, number of exposures, cascade depth, time elapsed, number of tweets, number of followers, number of followees, and account age) were log-transformed before standardization. *$p < .01$, **$p < .001$.

retweets of the original tweets, the study revisited the original tweets twice, one week and two weeks later. 28,957 such tweets were retweeted at least 5 times, of which only 58 (0.2%) original tweets received more than 100 retweets; they were excluded from the formal analysis due to the API's limit of downloading only the100 most recent retweets retrospectively. The present study treats these extremely popular tweets as outliers and focuses instead on the patterns of average tweets. A sensitivity analysis suggests that selection based on the number of retweets did not influence the main findings as presented in the result section (see Appendix Fig. s2). Then, the study randomly sampled 1,000 original tweets from the 28,957 tweets. Of these, 12 tweets no longer existed 2 weeks later and were excluded from the study, because it could not be determined whether all retweets were captured in these cases. For the sampled tweets, 21,329 unique retweets posted by 20,930 unique users were obtained using the public API (GET statuses/retweets/:id).

Second, the study obtained all followees (N = 39,678,526) and followers (N = 236,222,070) of the retweeters (using the GET friends/ids and GET followers/ids APIs). Combining them with the retweet dataset enabled us to successfully reconstruct 988 diffusion cascades using a method from Liang (2018) and Vosoughi et al. (2018). In the diffusion cascades, 5,173 parent users (i.e., users who are retweeted by child users; in this case, they are seed users and intermediaries) were retweeted by other users. For each parent user, the study randomly selected 20 followers who did not retweet the message in the diffusion cascade. For parent users with fewer than 20 followers (84%, M = 15, Mdn = 16, Min = 3), all followers that met the condition were selected. The followers selected were matched users who received the tweets but had not retweeted them by the time the data were collected. Ultimately, 98,678 unique matched users were obtained. All followees of the matched users who did not protect their accounts were obtained via the public API (GET friends/ids, N = 579,699,564).

Third, the study collected the profiles (i.e., the numbers of tweets, followers, followees, and time of registration) and most recent tweets (up to 3,200) from all retweeters and the matched non-retweeters, using GET users/lookup and GET statuses/user_timeline, respectively. Ultimately, the data contain complete profiles of 109,056 unique users and 190,550,745 tweets and retweets posted by 97,484 unique users. Replication datasets can be obtained on GitHub (https://github.com/rainfireliang/TwitterDiffusionData).

2.3. Measures

The dependent variable is *retweeting*, a binary variable that indicates whether a user retweeted a message from its parent in the matched dataset described above. In the matched datasets, there were 101,943 non-retweeting cases. For the 5,173 parent users, 5,109 users had matched followers who did not retweet the original messages. As to followers who opted to protect their user accounts, it was nearly impossible to obtain either following relationships or tweets using the public APIs. Those users were thus removed so that the final matched dataset contains 87,139 cases, including 16,099 retweets and 71,040 matched non-retweets (overall retweeting proportion = 18.5%). In total, 967 diffusion cascades and 5,086 parent users had matched cases and served as the final dataset used in the analyses presented below.

Interaction frequency between a potential retweeter and its parent was measured by the number of unique posts, including original tweets, replies, and retweets, that mentioned the parent user. This measures how frequently a potential retweeter replied to, retweeted from, or mentioned the parent user. For the 87,139 retweeter-parent cases, 62.5% had not previously interacted with each other; the mean frequency is 9 (Mdn = 0, SD = 65.89). To improve estimation efficiency in regression models, the variable was log-transformed and then standardized to have a mean of 0 and a standard deviation of 1 (a similar transformation was performed for other variables in Table 1).

Interest similarity between users was also measured at the dyadic level by calculating the cosine similarity of hashtags used by retweeters and their parent users. There are two advantages of choosing hashtags instead of raw text to calculate similarity. First, hashtags are user-defined topics and could thus be a better proxy than the bag-of-words technique to measure topic interest similarity. Second, the hashtag approach avoids the challenges of multilingual text analysis that may be entailed with users who use a first language other than English. Similarity scores were calculated separately for different matched sets. First, all tweets posted by a parent user's followers were extracted. All hashtags were selected and aggregated by user. Then, a document-term matrix was constructed, with rows representing users, columns unique hashtags, and values the frequencies of the hashtags mentioned by the users. The raw frequencies were further adjusted by the inverse document frequencies to account for the relative importance of the hashtags. Finally, a user's topic interest was represented by a vector of weighted frequencies of hashtags. Interest similarity between two users could then be measured by the cosine similarity between the users' topic vectors (see Liang and Fu, 2017). Theoretically, the cosine similarity score ranges from 0 (completely dissimilar) to 1 (exactly the same). The mean of the similarity is 0.068 (Mdn = 0.008, SD = 0.155). Alternatively, word embedding could also be used to measure semantic similarity between users (see Appendix Table s3).

Number of exposures of a potential retweeter was measured by the number of followees who retweeted the original tweet. Given that the dataset was constructed based on following relationships, the minimal number of exposures is 1. On average, the potential retweeters were exposed to 3.269 followees (Mdn = 1, SD = 6.310). Most retweeters retweeted the messages after the initial exposure (82.4%), while 42.4% of the matched non-retweeters were exposed to the messages at least twice.

Cascade depth was measured based on diffusion cascades. It was determined by the number of intermediaries between seed users and potential retweeters. For a message retweeted directly from the seed user, the cascade depth is 1. If there is one intermediary, the depth is 2. For matched non-retweeters, the depth of the followers of the seed users is 1. The depth of the followers of the intermediaries is the depth of the intermediaries plus 1. The maximum depth in the matched data is 37 (M = 3.601, Mdn = 2, SD = 4.104). The distribution could be found in Figure s1.

Network redundancy of a potential retweeter with its parent was measured by two indicators: the proportion of shared followees and the proportion of mediated followees between the retweeter and the parent user of the total number of the retweeter's followees. The mediated followees are the followees of the potential retweeter's followees and the followers of the parent user. As these two indicators are highly correlated (Spearman's Rho = 0.82, $p < .001$), the average was used as a measure of network redundancy. Theoretically, the minimum value is 0 (non-redundant) and the maximum value 1 (purely redundant). As a result, the mean of network redundancy is 0.096 (*Mdn* = 0.048, *SD* = 0.123) in the matched data set. Another structure variable is *reciprocity*, which was determined by whether the following relationship between the potential retweeter and the parent user was mutual. In the matched dataset, 46.5% of relationships are reciprocal; for the actual retweeters, the percentage is somewhat smaller (38.8%).

Several covariates derived from the user profiles were included as control variables that have been suggested as influences on the retweeting probability. For each potential retweeting action, the study included the attributes of both the potential retweeters and their parents: number of tweets, number of followers, number of followees, and account age (years since registration: 0 indicates accounts registered in 2019). In addition, the study measured the *retweeting inertia* of potential retweeters by the percentage of retweets in their timelines. Retweeting inertia indicates the baseline likelihood of retweeting others' messages. The average retweeting inertia for all unique users in the matched data is 47.8% (*Mdn* = 46.5%, *SD* = 32.7%); the variable was then standardized. Finally, cascade depth might be positively associated with the *elapsed time* since the message was posted. The study calculated the median elapsed time (in minutes) as a control variable for different cascade depths in different cascades.

2.4. Statistical models

Multilevel logistic regression models were employed to test the relationships between the predictors and retweeting probability in diffusion cascades. There are three levels in the matched datasets: the cascade level (level 3, $N = 967$), the parent level (level 2, $N = 5{,}086$), and the potential retweeter level (level 1, $N = 87{,}139$). Parent-level variables include cascade depth and all of each parent's attributes. Potential retweeter level (level 1) variables include interaction frequency, number of exposures, interest similarity, and all of each retweeter's attributes.

In different cascades, the cascade-level factors (e.g., the content characteristics and popularity of the tweet) may influence diffusion dynamics. However, these were not measured and are not the focus of the present study; a random-effects model could help control for these unobserved variables (Snijders and Bosker, 2012). Similarly, the characteristics of the parents are expected to influence retweeting probability. Some characteristics, such as the number of followers of the parents, were measured in this study, but many others were not. In order to control for these higher-level confounding variables, the study fitted random-intercept and random-slope multilevel models by including the group means (IGM) at the cascade and parent levels. Specifically, the present study calculated the means of interaction frequency, the number of exposures, and interest similarity by 5,086 parent users and 967 cascades, respectively. Then, the six aggregated variables were included as level 2 and 3 predictors in the multilevel models. IGM has been demonstrated to produce unbiased level 1 estimates because the correlations between level 1 variables and higher-level confounding variables are fully controlled for (Hanchane and Mostafa, 2012; Huang, 2016). Given that interaction frequency, the number of exposures, and interest similarity are all level 1 variables, their estimated effects in the multilevel models including group means will be unbiased even when omitting higher-level confounding variables.

As discussed in the literature review, social contagion or peer effects involve two factors: interaction and exposure. In the multilevel models predicting retweeting, the estimated coefficients of interaction frequency and the number of exposures indicate the effect size of social contagion. Meanwhile, the coefficient of interest similarity measures the homophily tendency. Although homophily effects could be measured based on different attributes, interest similarity represents one of these effects. Unless all attributes are controlled for in the regression models, latent homophily at the retweeter level could confound the main effects of interaction frequency, the number of exposures, and interest similarity on retweeting probability. However, as the present study controlled for the following relationships by matching-on-followers, the latent homophily effects via followings are fully accounted for.

Furthermore, non-retweeters were matched with the retweeters because they were followers of the same parents. Non-retweeting cases serve as the reference group for retweeting cases. As suggested in Aral et al. (2009), it is possible to formally match retweeters and non-retweeters on a vector of user attributes (number of followers, number of followees, number of statuses, and retweeting inertia) to estimate treatment effects more accurately. Instead of using a matched study design, the present study employed regression models by controlling for these characteristics. Regression models (as the parametric version of matching design) have been demonstrated to be more appropriate than matched studies when the baseline characteristics are imbalanced between groups (Brazauskas and Logan, 2016). The distributions of the numbers of followers, followees, and statuses in this study are highly skewed and thus imbalanced. Furthermore, as discussed above, a multilevel regression model could help control for higher-level confounders. Therefore, a multilevel logistic regression with random slopes and random intercepts including group means was selected to compare the within-stratum differences between retweeting and non-retweeting cases in the matched sets.

3. Results

3.1. Bivariate analysis

Compared with the matched non-retweeters, retweeters interacted more frequently with their parents (1.07 vs. −0.24, $p < .001$). They were more similar to their parents in terms of interest similarity (0.45 vs. −0.10, $p < .001$). Regarding retweeting inertia, they were more inclined to retweet than non-retweeters (0.68 vs. −0.15, $p < .001$), and their parents were more structurally redundant to them (0.16 vs. −0.04, $p < .001$). However, retweeters were exposed to the messages fewer times than non-retweeters (–0.37 vs. 0.08, $p < .001$) and were less likely to have reciprocal following relationships with their parent users (38.8% vs. 48.2%, $p < .001$).

In diffusion cascades without matched non-retweeters, cascade depth was positively associated with interest similarity

(Spearman's Rho = 0.26), the number of exposures (0.64), structural redundancy (0.30), and percentage of reciprocal following relationships (0.69), but it was only weakly correlated with retweeting inertia (0.04). Given the large sample size, all coefficients are significant. To correct for the large-*N* bias, Fig. 2 presents the correlation coefficients for the matched non-retweeters as a reference group; it shows a very similar pattern between retweeters and non-retweeters, which implies that the correlations were partially due to chance. Nevertheless, considering the differences between the two series, Fig. 2 suggests that retweeters at the deeper level were more similar to but interacted less with their parents. They were exposed to messages more times, and their parents were more structurally redundant.

The bivariate analysis illustrates that both interaction and interest similarity are positive predictors of retweeting probability. However, the number of exposures and reciprocity are negatively associated. In addition, the correlation analysis with cascade depth suggests that, as messages spread more deeply, retweeters become more similar and share more friends but interact less often. This implies that diffusion patterns could influence contagion effects.

3.2. *Multilevel logistic regression*

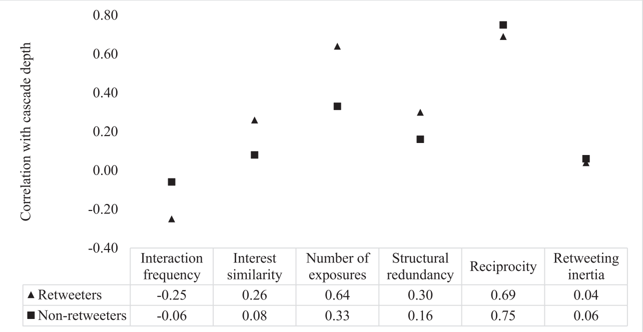
A series of three-level random-slope models was conducted to test the hypotheses (see Table 1). According to the R^2 values, the models can explain retweeting behaviors in diffusion cascades very well. The marginal R^2 is an indicator of the variance explained by fixed factors, while conditional R^2 is an indicator of variance explained by both fixed and random factors (Nakagawa and Schielzeth, 2013). The main effects of interaction frequency, the number of exposures, interest similarity, and cascade depth can explain 55.3% of the variance in Model I. The interaction effects with cascade depth in Model II are all significant, though with limited R^2 improvement (Model II – Model I = 4.4%). The control variables improved the goodness of fit a great deal (Model III – Model II = 14.2%).

Since all continuous variables were standardized to have a mean of 0 and a standard deviation of 1, it is possible to interpret the coefficients as effect sizes and thus compare them directly. According to Model III, the most influential factor (except for control variables) is the number of exposures. However, unlike the hypothesis of complex contagions, the direction is negative ($B = -1.40$, $SE = 0.05$). Users exposed to more followed retweeters are less likely to retweet. If the number of exposures increases by a single standard deviation, the odds of retweeting the message will decrease by 75.3%. Regarding *RQ*, the finding is more than consistent with the information redundancy explanation. The second influential factor is interaction frequency, which is positively associated with the likelihood of retweeting. Given a user who posts a message, followers who interact more frequently with that user are more likely to retweet the message ($B = 0.94$, $SE = 0.02$). If the interaction frequency increases by a single standard deviation, the odds of retweeting the message will increase by 154.7%. Therefore, *H1* is confirmed.

The impacts of interest similarity between potential retweeters and their parents on the retweeting probability are positive ($B = 0.72$, $SE = 0.03$). Potential retweeters sharing more interest similarity with their parents are more likely to retweet the message. A single standard deviation increase in interest similarity is associated with a 105.9% increase in the odds of retweeting. Therefore, *H2* is supported, but the effect size is smaller than the contagion effects.

Structural redundancy and reciprocity are both network variables that indicate close relationships between potential retweeters and their parents. However, Table 1 suggests that the two variables have different effects on the likelihood of retweeting. Users are more inclined to retweet messages if their parents in the diffusion cascades are more structurally redundant to them ($B = 0.05$, $SE = 0.02$). However, reciprocity has a negative effect ($B = \text{-}0.12$, $SE = 0.04$) when controlling for all other variables. The effect sizes of both structural factors are small. It might be caused by the matching procedure. In a matched set, all retweeters and non-retweeters are followers of the same parent users and thus might be in similar positions in following networks.

More importantly, cascade depth can moderate the contagion and homophily effects in diffusion cascades. According to Table 1, cascade depth attenuated the contagion effects; the impacts of social interactions and the number of exposures decreased with cascade depth. Fig. 3A&B indicate that the (absolute) values of the coefficients decrease with cascade depth. As Fig. 3D presents, when depth = 1, the difference of retweeting probabilities between two and no interactions is 3.8% ($p < .001$); when depth = 5, the difference decreases to 1.3% ($p < .001$). Similarly, the probability difference between two exposures and one exposure at step 1 is −13.7% ($p <$



	Interaction frequency	Interest similarity	Number of exposures	Structural redundancy	Reciprocity	Retweeting inertia
▲ Retweeters	-0.25	0.26	0.64	0.30	0.69	0.04
■ Non-retweeters	-0.06	0.08	0.33	0.16	0.75	0.06

Fig. 2. Correlation coefficients (Spearman's Rho) with cascade depth; variables were in the original scales. All correlation coefficients are statistically significant ($p < .001$). For reciprocity, the coefficient is the correlation between cascade depth and the proportion of reciprocal ties in the corresponding step.

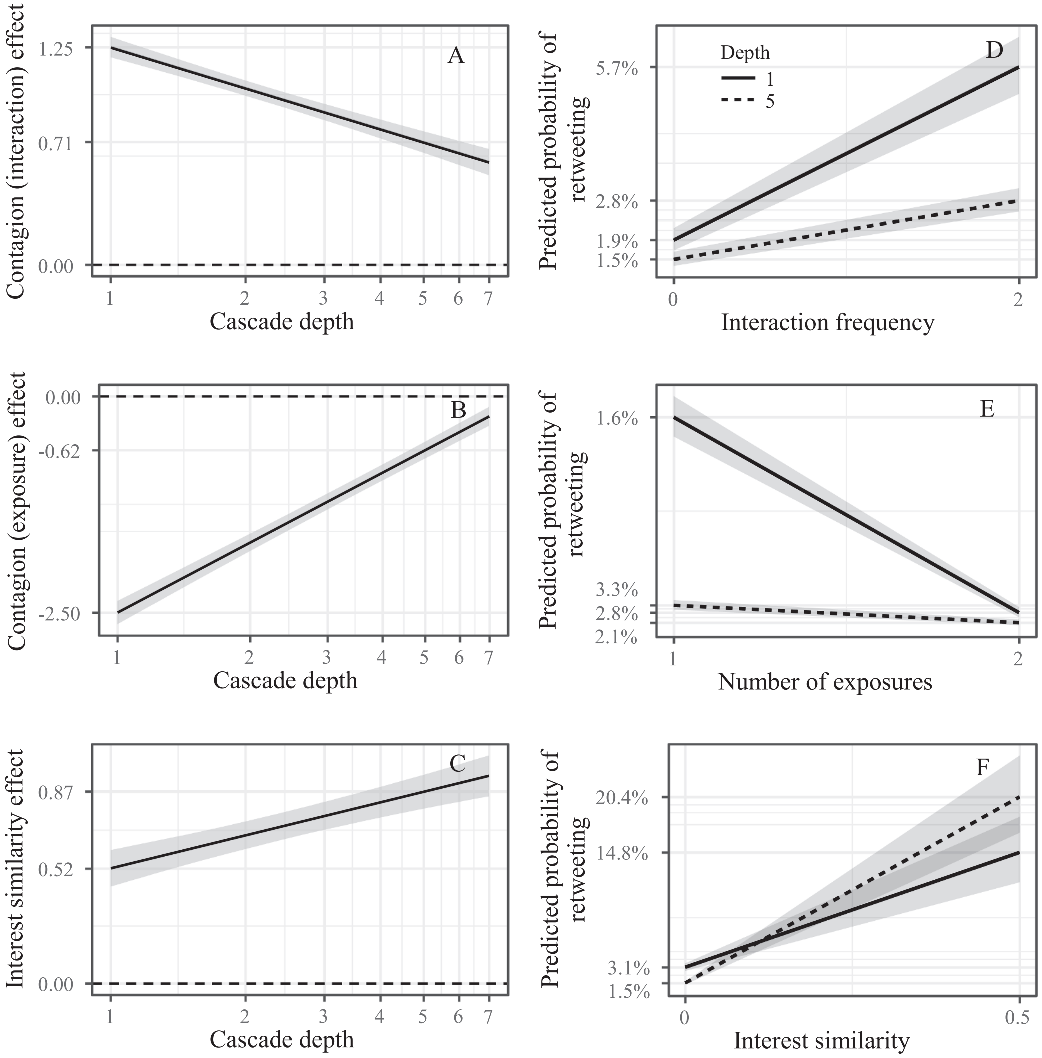


Fig. 3. Marginal effects of the interaction terms in Model III in Table 1. Figures A-C visualize the changes in the coefficient of one term in a two-way interaction conditioned by the other terms. Figures D-F visualize the conditional predicted probabilities for cascade depth = 1 and 5, respectively. Shaded areas are 95% CIs.

.001), while it decreases to −1.2% ($p < .001$) at step 5 (see Fig. 3E). Therefore, both *H3a* and *H3b* are confirmed.

By contrast, cascade depth amplified the effect of interest similarity on retweeting probability. As Fig. 3C shows, at a deeper level, interest similarity increases the probability to a larger extent. The difference of retweeting probabilities between 0.5 interest similarity and 0 interest similarity is 11.6% ($p < .001$) at step 1 and increases to 18.8% ($p < .001$) at step 5 (see Fig. 3F). Therefore, *H4* is confirmed. In summary, as a message spreads more deeply, the similarity of topic interest becomes more important than social interactions and the number of exposures, which are initially the most important factors.

4. Discussion and conclusions

The present study contributes to information diffusion research in two ways. Theoretically, the study demonstrates how diffusion structures are associated with the degree of peer influence and social contagion effects in information diffusion cascades. Users' retweeting decisions relied less on peer influence and social contagions (*H3*) and more on interest similarity (*H4*) as diffusion cascades became deeper. Methodologically, it proposes an analytical framework by leveraging matching-on-followers and multilevel modeling to model user message-sharing decisions and diffusion structures (here, cascade depth) together. In addition to combining choice and cascade models, the proposed method has the merits of controlling for latent homophily and structural factors.

In light of these findings, several points are worth further discussion. First, although it appears that the proposed framework is a natural extension of choice models by adding cascade depth as a moderator, it is important to note their conceptual difference. In choice models, researchers use contagion, homophily, and structural factors to predict users' decisions about message sharing. When situating choices in diffusion cascades, the question can be rephrased as follows: given that a message will be shared at step *N*, which viewers will share the message next? Information diffusion via social contacts on social media platforms (in contrast to mass media, for example) is not a process of how users select one of many competing messages or senders to share. Instead, their selections are constrained by and within concrete diffusion cascades, and the sharing probability is conditioned on facets of the local diffusion process like cascade depth.

Second, retweeters interacted more frequently with their parents than with the matched non-retweeters. This also implies that message sharing occurs through strong ties in diffusion cascades. Although the number of exposures did have a significant effect on retweeting, the direction was contrary to the complex contagion hypothesis and some previous empirical studies on Twitter (e.g., An et al., 2014). There are several possible reasons for this outcome. The retweeting function makes sharing behavior less effortful (Guilbeault et al., 2018), which makes the complex contagion theory inapplicable. The topic is another factor; as Romero et al. (2011) reported, political messages (e.g., An et al., 2014) are more likely to spread as complex contagions. However, the present study used a random sample of tweets across diverse topics. Finally, it is also related to the matching-on-followers method. Previous studies selected the reference group in different ways. Some studies selected non-retweeters randomly from the entire available population rather than the followers (e.g., An et al., 2014), while others selected non-retweeters from a predefined closed system or simply used all retweeters in the diffusion cascades (e.g., Bakshy et al., 2012; Hodas and Lerman, 2015). When matching on following relationships, structural factors may somehow be controlled for, and the number of exposures thus is negatively associated with retweet probability, which is consistent with Ugander et al. (2012).

Third, the small effect sizes of network structures should be interpreted with caution. The results are strikingly different from previous studies, which generally found strong effects of network structures in information diffusion (e.g., Aral and Walker, 2014; Liang and Fu, 2019; Ugander et al., 2012; Weng et al., 2013). The major reason is that the current matching strategy is not appropriate for testing network structure effects. The comparisons are between retweeters and non-retweeters that are followers of the same parent user. Matching on the following relationships naturally controlled for many structural characteristics, and the structural effects in Table 1 were underestimated. However, the results empirically confirmed that the matching-on-followers method could control for some latent homophily and structural factors. In addition, the structural effects could also apply at the parent level. Parents who are opinion leaders may be associated with certain structural characteristics such as large coreness that make them more influential than others in triggering large cascades (Kitsak et al., 2010).

4.1. Implications

The findings imply that peer influences may be more important than homophily-driven diffusion in triggering high information popularity. Information diffusion in social systems is usually generated by a combination of broadcast (one-to-many) and viral (person-to-person) spreading. Still, empirical studies have found that broadcast is more pervasive than viral spreading in many systems, including social media platforms (Goel et al., 2016). The present study found that social contagion effects become weaker when sharing takes place at deeper steps. This implies that peer influence might be stronger in broadcast diffusion, while homophily might be stronger in viral diffusion. Taken together, these indications suggest that peer influences are more common than homophily effects in popular diffusion cascades.

Although broadcast diffusion (e.g., mass media) associated with strong peer influences could generate large-scale information cascades, viral diffusion has its own advantages. Cascade depth might be associated with less bias caused by peer influences and a greater emphasis on message content in the diffusion process. For example, viral diffusion decreases partisan selective bias in political communication (Liang, 2018), and sharing decisions will rely more on users' interest in the content of the messages (An et al., 2014). In this sense, cascade depth may be associated with political deliberation. It is also consistent with the argument that the depth of online conversation networks is a prerequisite for political deliberation (Gonzalez-Bailon et al., 2010). However, reality might be more complicated. If homophily effects in general (not just interest similarity) increase with cascade depth, other attributes like gender and race could play equally important roles. If this is indeed the case, viral diffusion could also amplify gender or racial biases.

4.2. Limitations and future research

Despite its substantial contributions, the present study has certain limitations that need to be addressed in future research. First, like most studies using social media data, its generalizability may be limited due to platform variations. Two variables are particularly relevant here. Twitter is a news-sharing website and thus encourages the spreading of novel information. The negative effect of exposures might differ in other platforms that place a greater emphasis on social networking. The retweet button, which lessens the effort required for message sharing, can alter social contagion effects (Guilbeault et al., 2018). In addition, messages can spread in other ways on Twitter, including modified retweets, replies, and likes. The current study only considered official retweets, which might underestimate both cascade depth and size. The observed shortest path may only be weakly correlated with the actual shortest path. A sensitivity analysis performed to test whether the existence of likes could bias the findings indicated that the number of likes did not influence the main findings in Table 1, though some effects were underestimated when an original tweet had many likes (see Appendix Fig. s3).

Second, the matching-on-followers method was developed intuitively according to the natural spreading process in diffusion cascades (i.e., spreading via contact), which has several limitations. First, it does not consider external influences, such as mass media, other than peer influences. Second, the matching method has been demonstrated to reduce structural variations. However, it is unclear which part and how much was reduced. Future studies should formally investigate these problems by using simulated or ground truth data. Another limitation is that the retweeting probability is conditioned on the parent users, which assumes that the diffusion cascades will continue. For users who did not have followers or were not retweeted by any followers, no matched set was constructed. However, solving this problem requires answering a separate question – the sustainability of diffusion cascades – that merits a series of future studies. Finally, the proposed method could also be combined with experimental design (e.g., Monsted et al., 2017) to control for other retweeter-level confounding variables and make causal inferences more accurate in the future.

Funding

This study was supported in part by National Social Science Foundation of China (Major Program: 19ZDA324).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tele.2021.101623.

References

Agarwal, S., Mehta, S., 2020. Effective influence estimation in twitter using temporal, profile, structural and interaction characteristics. Inf. Process. Manage. 57 (6), 102321. https://doi.org/10.1016/j.ipm.2020.102321.

An, J., Quercia, D., Cha, M., Gummadi, K., Crowcroft, J., 2014. Sharing political news: the balancing act of intimacy and socialization in selective exposure. EPJ Data Sci. 3 (1) https://doi.org/10.1140/epjds/s13688-014-0012-2.

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., Tiwari, M., 2015. Global diffusion via cascading invitations: Structure, growth, and homophily, 24th International Conference on the World Wide Web, Florence, Italy. International World Wide Web Conferences Steering Committee, pp. 66-76.

Ansah, J., Kang, W., Liu, L., Liu, J., Li, J., 2018. Information propagation trees for protest event prediction, 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Melbourne, Australia, pp. 777–789.

Aral, S., Muchnik, L., Sundararajan, A., 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc. Natl. Acad. Sci. 106 (51), 21544–21549. https://doi.org/10.1073/pnas.0908800106.

Aral, S., Walker, D., 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. Manage. Sci. 57 (9), 1623–1639. https://doi.org/10.1287/mnsc.1110.1421.

Aral, S., Walker, D., 2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. Manage. Sci. 60 (6), 1352–1370. https://doi.org/10.1287/mnsc.2014.1936.

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L., 2012. The role of social networks in information diffusion, the 21st International Conference on the World Wide Web. ACM, Lyon, France, pp. 519–528.

Brazauskas, R., Logan, B.R., 2016. Observational studies: Matching or regression? Biol. Blood Marrow Transplantation 22 (3), 557–563. https://doi.org/10.1016/j.bbmt.2015.12.005.

Burt, R.S., 1980. Models of network structure. Annu. Rev. Sociol. 6 (1), 79–141. https://doi.org/10.1146/annurev.so.06.080180.000455.

Burt, R.S., 1987. Social contagion and innovation: Cohesion versus structural equivalence. Am. J. Sociol. 92 (6), 1287–1335. https://doi.org/10.1086/228667.

Burt, R.S., 1992. Structural holes: The social structure of competition. Harvard University Press, Cambridge, MA.

Centola, D., Macy, M., 2007. Complex contagions and the weakness of long ties. Am. J. Sociol. 113 (3), 702–734. https://doi.org/10.1086/521848.

Chen, L., Deng, H., 2020. Predicting user retweeting behavior in social networks with a novel ensemble learning approach. IEEE Access 8, 148250–148263. https://doi.org/10.1109/ACCESS.2020.3015397.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W., 2016. The spreading of misinformation online. Proc. Natl. Acad. Sci. U.S.A. 113 (3), 554–559. https://doi.org/10.1073/pnas.1517441113.

Goel, S., Anderson, A., Hofman, J., Watts, D.J., 2016. The structural virality of online diffusion. Manage Sci. 62, 180–196. https://doi.org/10.1287/mnsc.2015.2158.

Gonzalez-Bailon, S., Kaltenbrunner, A., Banchs, R.E., 2010. The structure of political discussion networks: a model for the analysis of online deliberation. J. Inf. Technol. 25 (2) https://doi.org/10.1057/jit.2010.2.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78 (6), 1360–1380. https://doi.org/10.1086/225469.

Granovetter, M., 1983. The strength of weak ties: A network theory revisited. Sociol. Theory 1, 201. https://doi.org/10.2307/202051.

Guilbeault, D., Becker, J., Centola, D., 2018. Complex contagions: A decade in review. In: Lehmann, S., Ahn, Y.Y. (Eds.), Comput Soc Sci. Springer, Cham, Switzerland, pp. 3–25.

Hanchane, S., Mostafa, T., 2012. Solving endogeneity problems in multilevel estimation: An example using education production functions. J. Appl. Statistics 39 (5), 1101–1114. https://doi.org/10.1080/02664763.2011.638705.

Hodas, N.O., Lerman, K., 2015. The simple rules of social contagion. Sci. Rep. 4 (1) https://doi.org/10.1038/srep04343.

Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression, 3rd ed. John Wiley & Sons, Hoboken, NJ.

Hu, J.Y., Luo, Y.W., Yu, J., 2018. An empirical study on selectivity of retweeting behaviors under multiple exposures in social networks. J. Comput. Sci.-Neth. 28, 228–235. https://doi.org/10.1016/j.jocs.2017.11.004.

Huang, F.L., 2016. Alternatives to multilevel modeling for the analysis of clustered data. J. Experim. Educ. 84 (1), 175–196. https://doi.org/10.1080/00220973.2014.952397.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A., 2010. Identification of influential spreaders in complex networks. Nature Phys. 6 (11), 888–893. https://doi.org/10.1038/nphys1746.

Li, X., Sun, C., Zia, M.A., 2020. Social influence based community detection in event-based social networks. Inf. Process. Manage. 57 (6), 102353. https://doi.org/10.1016/j.ipm.2020.102353.

Liang, H., 2018. Broadcast versus viral spreading: the structure of diffusion cascades and selective sharing on social media. J. Commun. 68, 525–546. https://doi.org/10.1093/joc/jqy006.

Liang, H., Fu, K.W., 2017. Information overload, similarity, and redundancy: Unsubscribing information sources on Twitter. J Comput-Mediat Comm 22, 1–17. https://doi.org/10.1111/jcc4.12178.

Liang, H., Fu, K.W., 2019. Network redundancy and information diffusion: the impacts of information redundancy, similarity, and tie strength. Commun. Res. 46 (2), 250–272. https://doi.org/10.1177/0093650216682900.

Liu, C., Zhan, X.X., Zhang, Z.K., Sun, G.Q., Hui, P.M., 2015. How events determine spreading patterns: information transmission via internal and external influences on social networks. New J Phys 17. https://doi.org/10.1088/1367-2630/17/11/113045.

Liu, Y., Jin, X., Shen, H., 2019. Towards early identification of online rumors based on long short-term memory networks. Inf. Process. Manage. 56 (4), 1457–1467. https://doi.org/10.1016/j.ipm.2018.11.003.

Marin, E., Guo, R., Shakarian, P., 2020. Measuring time-constrained influence to predict adoption in online social networks. Trans. Soc. Comput. 3 (3), 1–26. https://doi.org/10.1145/3372785.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annu. Rev. Sociol. 27 (1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415.

Monsted, B., Sapiezynski, P., Ferrara, E., Lehmann, S., 2017. Evidence of complex contagion of information in social media: An experiment using Twitter bots. Plos One 12, e0184148. https://doi.org/10.1371/journal.pone.0184148.

Myers, S.A., Zhu, C., Leskovec, J., 2012. Information diffusion and external influence in networks, 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, Beijing, China, pp. 33–41.

Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 4, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Romero, D.M., Meeder, B., Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter, 20th International Conference on the World Wide Web. ACM, Hyderabad, India, pp. 695–704.

Shalizi, C.R., Thomas, A.C., 2011. Homophily and contagion are generically confounded in observational social network studies. Sociol. Meth. Res. 40 (2), 211–239. https://doi.org/10.1177/0049124111404820.

Snijders, T.A.B., Bosker, R.J., 2012. Multilevel analysis: An introduction to basic and advanced multilevel modeling, 2nd ed. Sage, London.

Ugander, J., Backstrom, L., Marlow, C., Kleinberg, J., 2012. Structural diversity in social contagion. Proc. Natl. Acad. Sci. 109 (16), 5962–5966. https://doi.org/10.1073/pnas.1116502109.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151. https://doi.org/10.1126/science.aap9559.

Weng, L.L., Menczer, F., Ahn, Y.Y., 2013. Virality prediction and community structure in social networks. Sci. Rep. 3 https://doi.org/10.1038/srep02522.

Xu, B., Huang, Y., Kwak, H., Contractor, N., 2013. Structures of broken ties: Exploring unfollow behavior on Twitter, 2013 Conference on Computer Supported Cooperative Work. ACM, San Antonio, TX, pp. 871–876.

Zhan, X.X., Liu, C., Zhou, G., Zhang, Z.K., Sun, G.Q., Zhu, J.J.H., Jin, Z., 2018. Coupling dynamics of epidemic spreading and information diffusion on complex networks. Appl. Math. Comput. 332, 437–448. https://doi.org/10.1016/j.amc.2018.03.050.

Zhang, Y., Wang, L., Zhu, J.J.H., Wang, X., 2020. Viral vs. broadcast: Characterizing the virality and growth of cascades. EPL (Europhysics Letters). 131, 28002. https://doi.org/10.1209/0295-5075/131/28002.

Zhang, Z., Zhao, W., Yang, J., Paris, C., Nepal, S., 2019. Learning influence probabilities and modelling influence diffusion in Twitter, Companion Proceedings of The 2019 World Wide Web Conference, pp. 1087-1094.

Zhang, Z.K., Liu, C., Zhan, X.X., Lu, X., Zhang, C.X., Zhang, Y.C., 2016. Dynamics of information diffusion and its applications on complex networks. Phys. Rep. 651, 1–34. https://doi.org/10.1016/j.physrep.2016.07.002.