

RESEARCH

Open Access

Impact of network density on cascade size and community growth



Kazumasa Oida[†]

Correspondence: oida@fit.ac.jp

[†]Kazumasa Oida contributed equally to this work.

Department of Computer Science and Engineering, Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Higashi-ku, Fukuoka, Japan

Abstract

This paper addresses two critical questions related to (1) the co-evolutionary dynamics between information diffusion and network topology and (2) the relationship between viral content and social reinforcement. A recurrence relation model is developed to formulate the growth dynamics of a community centered on a dominant user using the tweet-retweet-follow (TRF) and exogenous link-creation events. This model illustrates several fundamental relations among parameters and quantities that are critical to answering the questions. The model reproduces social reinforcement and structural trapping effects, including empirical evidence that viral content does not require strong social reinforcement. In addition, the model demonstrates that the community growth rate is influenced not only by the rate of retweets but also by the network density of the community.

Keywords: Network density, Community, Twitter, TRF events, Recurrence relation, Co-evolutionary dynamics, Information cascade

Introduction

Online social networks (OSNs) have become indispensable vehicles of communication, information access, and advertisement. Content on OSNs may have economically, politically, or socially strong influences when shared among a large number of OSN users. Numerous papers in the literature have documented the manner in which such phenomena emerge; however, the mechanism that causes the phenomena may not be completely understood. This paper presents new insights into this field of study, shedding light on certain mechanisms that work in OSNs.

Every once in a while, large-scale information cascades emerge. Because of their rare occurrence, statistical or machine learning approaches present risks associated with overdependence on small samples. Therefore, analytical approaches must verify pieces of empirical evidence and derive unknown causal relationships based on the evidence. This paper contributes to the literature in the field of information diffusion by answering two important questions.

Question 1 (co-evolution) *Does the information diffusion co-evolve with the network topology?*

It has been established that the underlying network topology affects information diffusion. Conversely, can information diffusion significantly alter the topology? To answer

this question, this paper evaluates the community growth rate while a diffusion process is underway and demonstrates that the growth rate is greatly influenced by the network density of the community.

Question 2 (viral content) *Why does viral content not require strong social reinforcement?*

The seminal paper by Weng et al. (2013) presented a discussion on several community-related features that exhibit significant influences on information diffusion, including *social reinforcement*, which indicates that each additional exposure significantly increases the chance of adoption. One of the key findings of the paper was that non-viral content requires strong social reinforcement, whereas viral content does not. From an analytical perspective, this paper illustrates that large-scale cascades emerge *only if* the network density of the community is low.

These questions are directly related to improvements in future cascade-size prediction. Therefore, answering these questions can help, for example, develop more effective viral marketing strategies and realize more efficient content delivery network (CDN) resource deployment for viral content.

The model in this paper presents a community structure for revealing the mechanisms that yield large-scale information cascades because the community-related features are considered to exhibit a strong influence on future content popularity (Weng et al. 2014). An important factor that must be considered while designing an analytical model is that deviations from the empirical evidence must be small. Based on the evidence that a large number of adoptions occurred within only one degree of a few dominant individuals (Goel et al. 2012), this paper presents a scenario in which one dominant user, who has a significant influence on some online application or some social issues, is followed by a large number of Twitter users. This paper incorporates link-creation events, called tweet-retweet-follow (TRF), into the model because TRF link-creation events are orders of magnitude more likely than link-creation events caused by exogenous reasons (Antoniades and Dovrolis 2015). In addition, the paper also presents a discussion on the effects of exogenous link-creation events.

This paper focuses on the two Twitter features (a dominant user and TRF events) mentioned previously and emphasizes their importance by reproducing the community-related effects (social reinforcement and structural trapping) and the evidence describing that non-viral (viral) content (does not) requires strong social reinforcement (Question 2). After that, the paper presents a discussion on the manner in which the parameters in the model, particularly the network density, relate to the two questions.

The paper is organized as follows. “[Related work](#)” section presents the related work. “[The model](#)” section defines the retweet diffusion model, and “[Results](#)” section derives the fundamental relations among parameters and quantities based on the computational results, and presents some insights into Questions 1 and 2. “[The extended model](#)” section extends the model by detailing user behavior. “[Discussion](#)” section presents a discussion on the implications of the findings and the limitations of this approach. “[Conclusions](#)” section concludes the paper.

Related work

Information diffusion on OSNs has been a subject of intensive study. It has been widely recognized that information diffusion is a more complex contagion than the spread of infectious diseases. Several empirical studies indicate that connectivity distributions among online users are highly inhomogeneous (Goel et al. 2012; Goel et al. 2015; Bose and Shin 2008). Prior studies have examined the influence of underlying network properties, such as scale-free, small-world, and community structures, on the process of information diffusion under the assumption that the networks are static (Weng et al. 2013; Moore and Newman 2000; Watts 2002; Moreno et al. 2004). The influence maximization problem in social networks, under a variety of influence mechanisms and network properties, is an ongoing focus of research (Kempe et al. 2003; 2005; Aral and Dhillon 2018).

Current OSNs include a variety of communities such as language communities (Jin 2017); therefore, many studies have examined or applied the characteristics of information diffusion processes as information is transmitted within/across OSN communities. For example, the authors in (Li et al. 2015) categorized nodes into six community-aware types and evaluated their diffusion capabilities. This paper investigates the dynamics of a single Twitter community, assuming all community members are homogeneous and uniformly connected with each other.

Recent work (Tur et al. 2018) evaluated the effect of social reinforcement using a threshold model that assumes that additional contacts are not redundant but reinforce the probability of contagion. The model in this paper also assumes that additional retweet arrivals are not redundant; the retweet rate is constant independent of the number of arrivals. A fundamental difference is that the model in Tur et al. (2018) considers static networks. Another similar study (Banisch and Olbrich 2019) introduced positive and negative social feedback, and discussed the role of network density in the emergence of opinion polarization. The model in Banisch and Olbrich (2019) also assumes static underlying networks.

There are numerous studies that focus on predicting large-scale information diffusion (Weng et al. 2013; Szabo and Huberman 2010; Cheng et al. 2014; Zhao et al. 2015; Cheung et al. 2017; Krishnan et al. 2016). Incorporating the community structure into diffusion models is expected to improve model accuracy (Bao et al. 2017). Several studies utilize community properties for prediction, such as predicting the future popularity of a meme (Weng et al. 2014) and the virality timing of content (Junus et al. 2015). Contrary to these results, this paper discusses the intrinsic difficulty associated with the prediction.

Moreover, there are studies on the dynamic nature of the OSNs (Leskovec et al. 2008; Myers and Leskovec 2014). The work in (Myers and Leskovec 2014) empirically demonstrated that large information diffusion has the potential to generate sudden influxes of edge creation or deletion events. In contrast to the assumption that the network topology is static while a diffusion process is ongoing, some papers have attempted to understand the diffusion phenomena on the basis of the co-evolutionary dynamics, where network topology and information diffusion influence each other reciprocally. In (Weng et al. 2013), several link-creation mechanisms were introduced and their roles in the diffusion process were examined. The authors in Farajtabar et al. (2017) formulated a stochastic process expressing the joint dynamics of information diffusion and network evolution. In Antoniadou and Dovrolis (2015), the authors studied the importance of a link-creation event called tweet-retweet-follow (TRF). They demonstrated that

TRF events can create dense user communities. There have been several studies that focused on link-creation events such as triadic closure (Leskovec et al. 2008), shortcuts (Weng et al. 2013), homophily (Hours et al. 2016; Aragón et al. 2017), and social influence (Aragón et al. 2017). This paper formulates a retweet diffusion model based on TRF events because the events express general content sharing and link creation dynamics on Twitter. Instead of introducing stochastic processes or differential equations, this paper uses recurrence relations to express TRF dynamics because the recurrence relations derived are simple and ensure extensibility.

The model

The model includes an initiator, a large number of Twitter followers of the initiator (referred to as supporters), and Twitter users who are not supporters (referred to as non-supporters). The model assumes that supporters are closely related to each other on the basis of the follower-followee relationship and form a community. Therefore, the model is a community centered on the initiator. This section introduces recurrence relations to express the retweet propagation phenomena that originate from the initiator to supporters and non-supporters according to the TRF events.

Table 1 presents the parameter values in the model. N_i denotes the number of supporters at the time initiator tweets, N_f is the number of followers a user (except the initiator) has, and $\alpha \in (0, 1]$ is the ratio of supporters in N_f followers of a supporter; e.g., $\alpha = 0$ (1) indicates that all N_f followers are non-supporters (supporters). Parameter α is an indicator of the network density of the community because a higher α implies a higher network density. Therefore, α is also referred to as the network density indicator of the community. This paper fixes the value of N_f at 500 because most of the Twitter analysis reports on the Internet present three-digit numbers for the average number of followers.

The model is designed under the following assumptions:

Asm 1 *Supporters and non-supporters are homogeneous in that they do not possess any unique attributes.*

- *Supporters (non-supporters) retweet with rate λ_s (λ_n). $\lambda_s \geq \lambda_n$.*
- *Supporters and non-supporters have the same N_f followers.*
- *Supporters have αN_f followers who are supporters ($0 < \alpha \leq 1$), while the followers of non-supporters are all non-supporters.*

Asm 2 *Users synchronously post retweets after receiving (re)tweets. The synchronization points are called steps.*

Table 1 Parameter values in the model

Parameter	Description	Default value
N_i	Number of supporters at step 0	10,000
N_f	Number of followers a user has	500
α	Ratio of supporters in N_f followers	(0, 1]
λ_s	Retweet rate of supporters	0.01
λ_n	Retweet rate of non-supporters	0.0015

Asm 3 The community does not have any specific network topology. At each step, αN_f followers of each supporter are uniformly selected from all supporters.

Asm 4 All non-supporters who post retweets become supporters (i.e., follow the initiator).

Asm 4 is relaxed in “The extended model” section; however, the assumption provides most of the essential results in this paper. The model is simplified to focus on two Twitter properties: TRF events and locally scale-free structure (one dominant user and other users with a relatively small number of followers). The importance of the properties is confirmed by reproducing the community-related effects (social reinforcement and structural trapping) and (non-)viral content behavior.

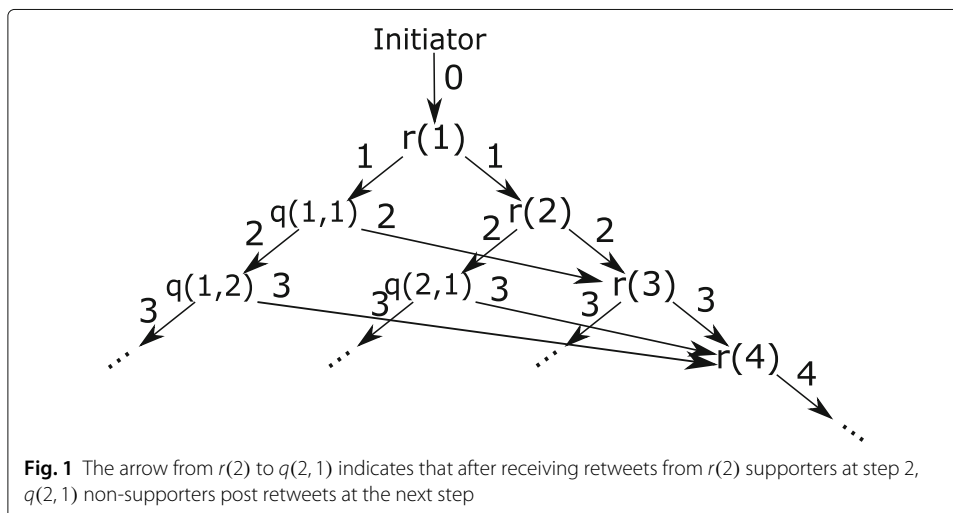
Figure 1 illustrates a retweet cascade originated from the initiator at step 0. The recurrence relation model can be described as follows. Let $r(i)$ ($\bar{q}(i)$) be the number of supporters (non-supporters) who retweet at step i . At step 0, the initiator tweets. At step 1, $r(1)$ supporters retweet the tweet. At step 2, $r(2) + q(1, 1)$ followers of the $r(1)$ supporters post retweets, where $q(i, j)$ denotes the number of non-supporters who post retweets at step $i + j$ after receiving retweets from $q(i, j - 1)$ non-supporters if $j > 1$ or from $r(i)$ supporters if $j = 1$ (see Fig. 1).

From Asm 4, $q(1, 1)$ followers become supporters and the new $q(1, 1)$ supporters influence $r(3)$. Therefore, $q(i, j)$ also denotes the number of new supporters at step $k > i + j$. At step 3, $r(3) + q(1, 2) + q(2, 1)$ users retweet. At step i , $r(i) + \bar{q}(i)$ users retweet, where $\bar{q}(1) = 0$ and for $i > 1$,

$$\bar{q}(i) = \sum_{k=1}^{i-1} q(k, i - k). \tag{1}$$

Using the parameters in Table 1, $r(1) = \lambda_s N_i$, $q(1, 1) = \lambda_n (N_f - \alpha N_f) r(1)$, and $q(1, 2) = \lambda_n (N_f - \alpha N_f) q(1, 1)$. In general, for $i \geq 1$ and $j \geq 0$,

$$\begin{aligned} q(i, j + 1) &= \lambda_n (N_f - \alpha N_f) q(i, j) \\ &= \lambda_n^{j+1} (N_f - \alpha N_f)^{j+1} r(i). \end{aligned} \tag{2}$$



Algorithm 1 Calculate $r(i)$

```

1:  $r' \leftarrow 0$ 
2: for  $j = 1$  to  $r(i-1) + \bar{q}(i-1)$  do
3:   for  $k = 1$  to  $\alpha N_f$  do
4:      $P_{unret} \leftarrow \frac{N_i - r' - \sum_{1 \leq \ell \leq i-1} r(\ell)}{N_i - 1 - (k-1) + \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)}$ 
5:      $r' \leftarrow r' + \lambda_s P_{unret}$ 
6:   end for
7: end for
8:  $r(i) \leftarrow r'$ 

```

On the other hand, $r(i)$ is computed according to Algorithm 1. The algorithm can be explained as follows. At step $i-1$, $r(i-1) + \bar{q}(i-1)$ supporters retweet and each supporter sends αN_f retweets to supporters. Because each retweet message arrives at uniformly selected followers (Asm 3), the probability (P_{unret}) that the k -th retweet arrives at a supporter who has not posted any retweets is given by

$$P_{unret} = \frac{N_i - r' - \sum_{1 \leq \ell \leq i-1} r(\ell)}{N_i - 1 - (k-1) + \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)}, \quad (3)$$

where the numerator denotes the number of supporters who have not posted any retweets, and the denominator denotes the number of supporters who have the possibility to receive the k -th retweet. In (3), r' is the number of supporters who have retweeted at step i . In the denominator, “ -1 ” indicates that retweeters do not receive their own retweets, and “ $-(k-1)$ ” indicates that a user does not receive multiple retweets from the same user. Because P_{unret} is also the mean of the binomial distribution $B(1, P_{unret})$, the mean value of the number of supporters who retweet after receiving the k -th retweet is given by $\lambda_s P_{unret}$. For all $(r(i-1) + \bar{q}(i-1)) \times \alpha N_f$ retweets, $\lambda_s P_{unret}$ values are computed and added to r' in line 5. $r(i)$ is then derived in line 8.

The computation of $r(i)$ and $\bar{q}(i)$ stops at $i = T + 1$ if

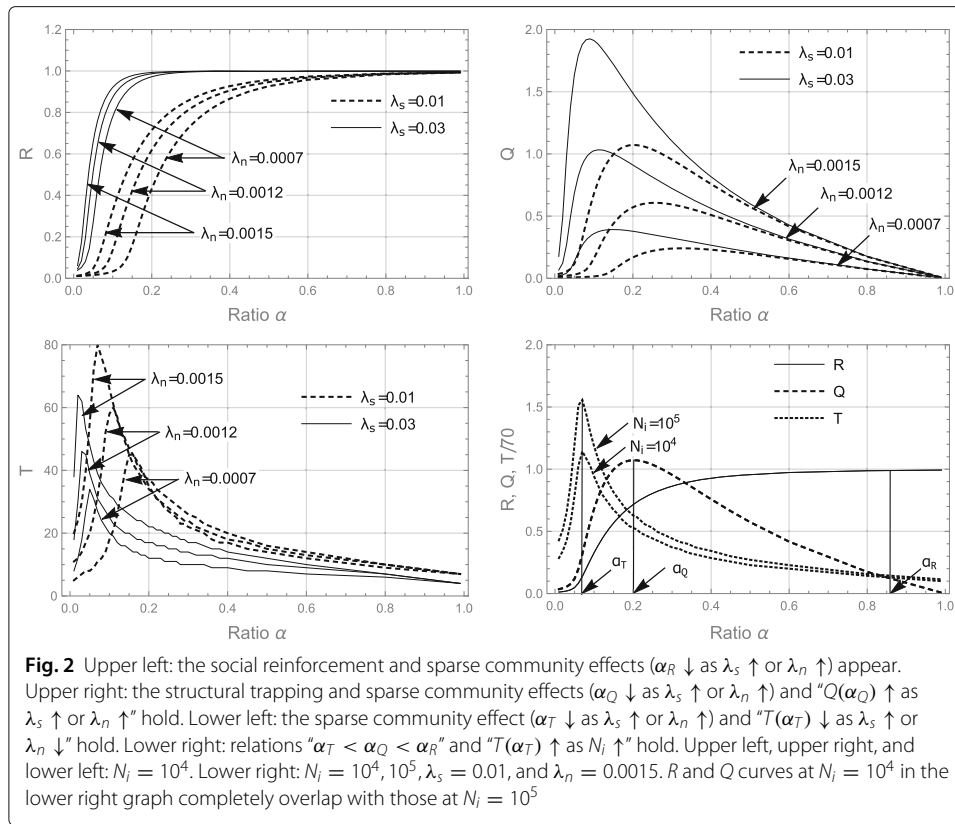
$$r(T+1) + \bar{q}(T+1) < 1, \quad (4)$$

which indicates that the number of users who post retweets at step $T+1$ is less than one. This paper defines T (steps) as the duration of retweet propagation.

Results

Sparse community effect

This section discusses the impact of the ratio α on $R = \sum_{k=1}^T r(k)/N_i$, $Q = \sum_{k=1}^T \bar{q}(k)/N_i$, and T , where R denotes the relative number of original supporters (supporters at step 0) who have retweeted until step T and Q is the relative number of non-supporters who become supporters until step T . R satisfies $0 \leq R \leq 1$ and $R = 1$ indicates that all original supporters have retweeted. Q can grow larger than one and $Q > 1$ means that the number of supporters has more than doubled. Figure 2 illustrates R , Q , and T as functions of α . From Fig. 2(upper left), R exhibits a curve that is similar to a sigmoid curve, while Fig. 2(upper right) and (lower left) illustrate that Q and T have uni-modal curves. Therefore, there are three characteristic ratios: α_R , α_Q , and α_T . α_Q and α_T denote the ratios at which T and Q exhibit peaks, respectively. α_R is the smallest ratio satisfying $|R - 1| < \delta$, where δ is a positive and sufficiently small real number (e.g., 10^{-4}).



Through intensive calculations, the following four fundamental relations are derived. The first two findings are related to the behavior of α_R, α_Q , and α_T . From Fig. 2(lower right),

$$\alpha_T < \alpha_Q < \alpha_R. \tag{5}$$

The second finding is that α_T, α_Q , and α_R decrease as λ_s or λ_n increases (see Fig. 2(upper left), (upper right), and (lower left)); i.e.,

$$\alpha_T \downarrow, \alpha_Q \downarrow, \alpha_R \downarrow \text{ as } \lambda_s \uparrow \text{ or } \lambda_n \uparrow. \tag{6}$$

This paper terms this phenomenon the *sparse community effect*. The effect implies that sparse communities, in which supporters have many (few) links to non-supporters (supporters), yield the largest cascades ($R + Q$) and the most long-lived cascades (T) when the rates of retweet are high. The sparse community effect presents a case in which the network structure affects content virality.

The third finding is that $Q(\alpha_Q)$ rises as λ_s or λ_n grows; whereas $T(\alpha_T)$ falls (rises) as λ_s (λ_n) grows; i.e.,

$$T(\alpha_T) \downarrow, Q(\alpha_Q) \uparrow \text{ as } \lambda_s \uparrow, \tag{7}$$

$$T(\alpha_T) \uparrow, Q(\alpha_Q) \uparrow \text{ as } \lambda_n \uparrow. \tag{8}$$

The fourth finding is illustrated in Fig. 2(lower right). $T(\alpha_T)$ increases with N_i ; i.e.,

$$T(\alpha_T) \uparrow \text{ as } N_i \uparrow. \tag{9}$$

Note that in Fig. 2(lower right), R and Q do not change with N_i (these are quantities relative to N_i).

Structural trapping and social reinforcement

Figure 2 presents other findings that agree with the empirical phenomena. It can be observed from Fig. 2 (upper right) that Q approaches zero as α approaches one; i.e.,

$$Q \downarrow 0 \text{ as } \alpha \uparrow 1, \tag{10}$$

which represents the *structural trapping* effect (Weng et al. 2013) because a larger α , indicating denser communities with fewer outgoing links, lowers Q , indicating a smaller amount of propagation outside of the community. While, Fig. 2 (upper left) illustrates the *social reinforcement* effect, which indicates that in the presence of high clustering, any additional adoption is likely to produce more multiple exposures than in the case of low clustering (Weng et al. 2013), because R is a strictly increasing function of α ; i.e.,

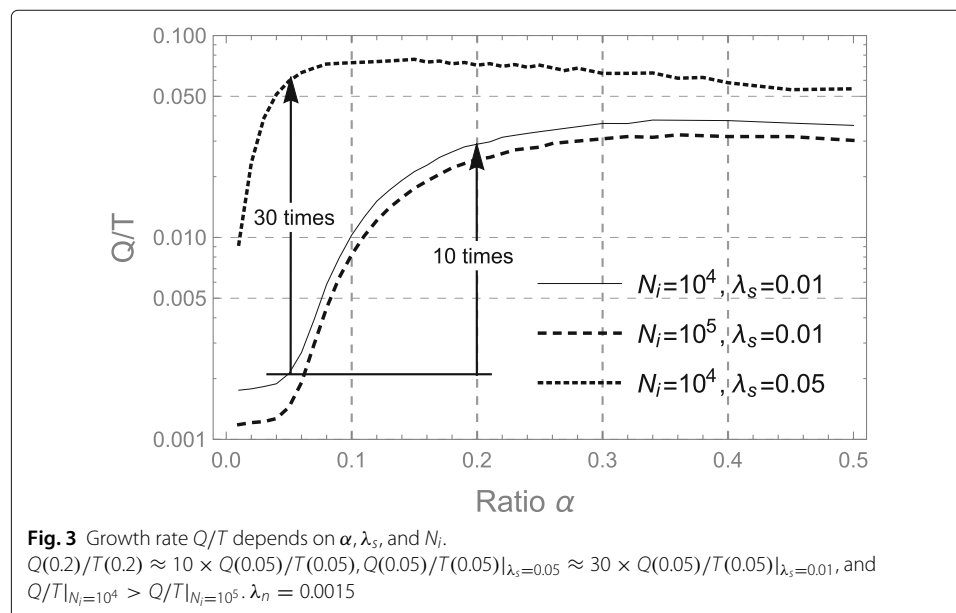
$$R \uparrow \text{ as } \alpha \uparrow. \tag{11}$$

Answering two questions

Let us first discuss Question 1 (co-evolution) in “Introduction” section using the community growth rate measured by Q/T . A large Q/T implies dynamic community growth. The following demonstrate under what conditions it is dynamic (or static). Because $\alpha_T \neq \alpha_Q$, the growth rate Q/T must greatly vary in the range of $[\alpha_T, \alpha_Q]$. It can be anticipated that

$$\frac{Q(\alpha_T)}{T(\alpha_T)} < \frac{Q(\alpha_Q)}{T(\alpha_Q)}. \tag{12}$$

Figure 3 compares the growth rates at $\alpha = 0.05$, which is close to α_T , and at $\alpha = 0.2(\approx \alpha_Q)$. From the figure, $Q(0.2)/T(0.2)$ is approximately 10 times greater than $Q(0.05)/T(0.05)$.



Meanwhile, (7) suggests that Q/T monotonically increases with λ_s ; i.e.,

$$Q/T \uparrow \text{ as } \lambda_s \uparrow. \quad (13)$$

Figure 3 illustrates the case where λ_s increases from 0.01 to 0.05. From the figure, $Q(0.05)/T(0.05)$ at $\lambda_s = 0.05$ is approximately 30 times greater than that at $\lambda_s = 0.01$. This implies that an attractive tweet from the initiator has the potential to greatly enlarge the growth rate. Note from Fig. 3 that $Q(0.05)/T(0.05)$ at $N_i = 10^4$ and $\lambda_s = 0.05$ is approximately 0.06, and this implies that after 10 steps, the average community size increases by 60%.

Because Q is independent of N_i (see Fig. 2(lower right)), it follows from (9) that

$$Q/T \downarrow \text{ as } N_i \uparrow. \quad (14)$$

In other words, communities with a large number of supporters grow slowly. Figure 3 illustrates that Q/T at $N_i = 10^5$ is slightly smaller than that at $N_i = 10^4$. Accordingly, the growth rate is particularly sensitive to α and λ_s .

Let us next discuss Question 2 (viral content) using the cascade size W , which is measured by $W = R + Q$. From (6)–(8),

$$Q(\alpha_Q) \uparrow, \alpha_Q \downarrow, \alpha_R \downarrow \text{ as } \lambda_s \uparrow \text{ or } \lambda_n \uparrow, \quad (15)$$

which implies that cascade size W exhibits a higher peak at a smaller α as λ_s or λ_n increases; i.e.,

$$W(\alpha_W) \uparrow, \alpha_W \downarrow \text{ as } \lambda_s \uparrow \text{ or } \lambda_n \uparrow, \quad (16)$$

where α_W is the ratio α that maximizes W . Namely, a large number of users post retweets (i.e., viral content emerges) at small α values where the social reinforcement effect is considered to be small.

Let us intuitively explain why (16) occurs. Community members (i.e., supporters) exhibit a greater tendency to post retweets than non-members (i.e., non-supporters) because of the social reinforcement effect and $\lambda_s > \lambda_n$. When the value of λ_s is high, for example, members easily retweet; therefore, by reducing α , more retweets must be sent outside of the community, providing non-members more chances to retweet; otherwise, the structural trapping effect becomes more pronounced. In other words, to maximize W , the ratio α_W controls the numbers of retweet messages flowing inside and outside the community such that fewer (more) messages flow inside (outside) the community (i.e., α_W is reduced) when the value of λ_s or λ_n increases (the sparse community effect).

The extended model

Detailed user behavior

Asm 4 in “Results” section assumes that all non-supporters who post retweets become supporters. This section relaxes the assumption and introduces exogenous link-creation, which states that non-supporters become supporters without receiving retweets, by replacing Asm 4 with the following assumptions:

Asm 5 *The ratio of non-supporters who become supporters after retweeting β_r (referred to as the retweeter’s follow rate), which is given by*

$$\beta_r = \frac{\#\{x \in S_n | x \text{ posts a retweet and become a supporter}\}}{\#\{x \in S_n | x \text{ posts a retweet}\}}, \quad (17)$$

satisfies $0 < \beta_r \leq 1$, where S_n is the set of non-supporters.

Asm 6 The ratio of non-supporters who become supporters without retweeting β_n (referred to as the exogenous link-creation indicator), which is given by

$$\beta_n = \frac{\#\{x \in S_n | x \text{ becomes a supporter and does not retweet}\}}{\#\{x \in S_n | x \text{ becomes a supporter}\}}, \tag{18}$$

satisfies $0 \leq \beta_n < 1$.

Asm 7 The ratio of non-supporters who become supporters without retweeting after receiving retweets c , which is given by

$$c = \frac{\#\{x \in S_n | x \text{ receives a retweet, becomes a supporter, and does not retweet}\}}{\#\{x \in S_n | x \text{ becomes a supporter}\}}, \tag{19}$$

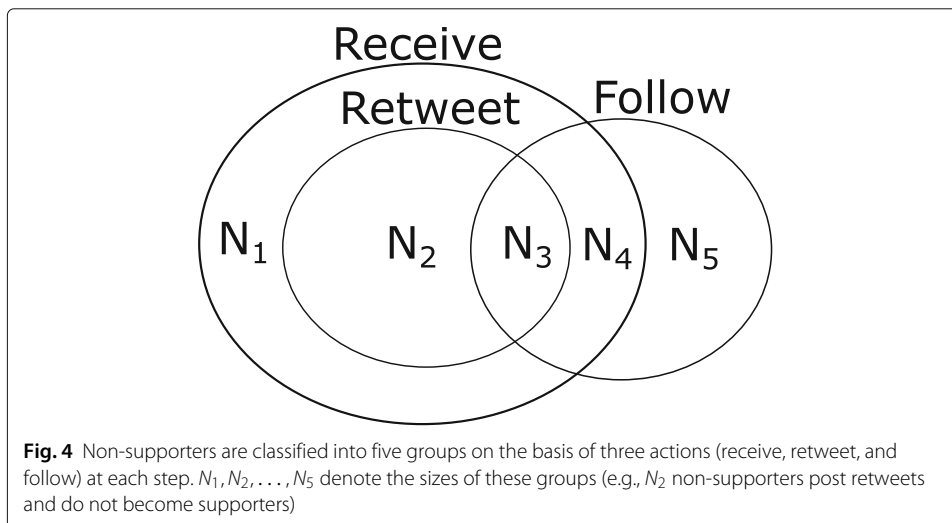
is fixed and $0 \leq c \leq \beta_n$.

Asm 4 corresponds to $\beta_r = 1$ and $\beta_n = 0$. Constant c in Asm 7 does not affect the output of the extended model; while, Asms 5 and 6 require making the following modifications to Algorithm 1:

- In line 2 in Algorithm 1, $r(i - 1) + \bar{q}(i - 1)$ (the number of supporters who retweet at step $i - 1$) is replaced with $r(i - 1) + \beta_r \bar{q}(i - 1)$.
- In line 4, P_{unret} is revised as

$$P_{unret} = \frac{N_i - r' - \sum_{1 \leq \ell \leq i-1} r(\ell) + \beta_r \frac{\beta_n}{1-\beta_n} \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)}{N_i - 1 - (k - 1) + \beta_r \left(1 + \frac{\beta_n}{1-\beta_n}\right) \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)}. \tag{20}$$

The following explains (20). Figure 4 illustrates five non-supporter groups classified on the basis of three actions: receive, retweet, and follow. In the figure, N_k denotes the size of the k -th group at each step. Note that the previous model corresponds to the case of $N_2 = N_4 = N_5 = 0$. From (17)–(19),



$$\beta_r = \frac{N_3}{N_2 + N_3}, \tag{21}$$

$$\beta_n = \frac{N_4 + N_5}{N_3 + N_4 + N_5}, \tag{22}$$

$$c = \frac{N_4}{N_3 + N_4 + N_5}. \tag{23}$$

Because the number of non-supporters who retweet at step ℓ is $\bar{q}(\ell)$,

$$N_2 + N_3 = \bar{q}(\ell). \tag{24}$$

As a modification, the numerator in (3) must include $N_4 + N_5$ (the number of new supporters who have not retweeted) of all previous steps and the denominator requires replacement of N_3 , which corresponds to $\bar{q}(\ell)$, with $N_3 + N_4 + N_5$ (the number of new supporters) for all previous steps. From (21), (22), and (24),

$$N_4 + N_5 = \frac{\beta_n}{1 - \beta_n} \beta_r \bar{q}(\ell), \tag{25}$$

$$N_3 + N_4 + N_5 = \beta_r \bar{q}(\ell) + \frac{\beta_n}{1 - \beta_n} \beta_r \bar{q}(\ell). \tag{26}$$

Therefore, compared with (3), the numerator in (20) includes a new term $\beta_r \frac{\beta_n}{1 - \beta_n} \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)$, and in the denominator, $\sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)$ is replaced by $\beta_r \left(1 + \frac{\beta_n}{1 - \beta_n}\right) \sum_{1 \leq \ell \leq i-1} \bar{q}(\ell)$.

Let us next consider *exogenous link-creation events* (Antoniades and Dovrolis 2015), which indicate that non-supporters become supporters without receiving retweets. Therefore, the ratio of exogenous link-creation events in the link-creation events β_e is given by

$$\beta_e = \frac{N_5}{N_3 + N_4 + N_5}. \tag{27}$$

Using (22), (23), and (27),

$$\beta_n = \beta_e + c, \tag{28}$$

which indicates that β_e linearly increases with β_n . Therefore, β_n is referred to as the exogenous link-creation indicator.

The new assumptions redefine R (the relative number of supporters who retweet until step T), Q (the relative number of non-supporters who become supporters until T), and cascade size W (the relative number of users who retweet until T). Let $\bar{Q} = \sum_{k=1}^T \bar{q}(k)$ and $\bar{R} = \sum_{k=1}^T r(k)$. They are written as

$$R = \bar{R} \left(N_i + \beta_r \frac{\beta_n}{1 - \beta_n} \bar{Q} \right)^{-1}, \tag{29}$$

$$Q = \beta_r \left(1 + \frac{\beta_n}{1 - \beta_n} \right) \bar{Q} / N_i, \tag{30}$$

$$W = (\bar{Q} + \bar{R}) / N_i, \tag{31}$$

where R is normalized such that it satisfies $0 \leq R \leq 1$, and Q and W are quantities relative to N_i . Note that Q and \bar{Q}/N_i are different; \bar{Q}/N_i denotes the relative number of non-supporters who post retweets until T . Note also that the definition of W is the same as before.

Relaxing sparse community effect

Figure 5 exhibits the effects of the retweeter’s follow rate β_r and the exogenous link-creation indicator β_n . In Fig. 5(left) and (right), the solid lines correspond to the previous model (i.e., $\beta_r = 1$ and $\beta_n = 0$). Because these parameters directly affect Q , it is natural that Q must increase with β_r or β_n . The effects of β_r and β_n can be summarized as follows:

$$\alpha_T \downarrow, \alpha_Q \downarrow, \alpha_R \downarrow \text{ as } \beta_r \uparrow \text{ or } \beta_n \downarrow, \tag{32}$$

$$T(\alpha_T) \uparrow, Q(\alpha_Q) \uparrow \text{ as } \beta_r \uparrow \text{ or } \beta_n \uparrow. \tag{33}$$

From (32), an increase in β_r strengthens the sparse community effect, whereas β_n weakens the effect. The following partly explains (32). An increase in β_r or a decrease in β_n indicates that a percentage of retweeters in the community increases; therefore, to enlarge R and Q , more retweets must flow outside the community, which implies that α_Q and α_R decrease.

The extended model also satisfies the fundamental relations (5)-(9) in “Results” section and exhibits the sparse community, social reinforcement, and structural trapping effects. Therefore, the extended model provides the same discussion on Questions 1 and 2 as in “Answering two questions” subsection.

Figure 6 illustrates the growth rate (Q/T) and the cascade size (W) for different β_r and β_n values. It can be observed from the figure that

$$Q/T \uparrow \text{ as } \beta_r \uparrow \text{ or } \beta_n \uparrow, \tag{34}$$

$$W \uparrow \text{ as } \beta_r \uparrow \text{ or } \beta_n \uparrow. \tag{35}$$

It is worth noting that in Fig. 6(upper left), $Q(0.1)/T(0.1)$ at $\beta_r = 1$ is approximately 50 times greater than that at $\beta_r = 0.1$; i.e., a low retweeter’s follow rate makes networks static. In addition, from Fig. 6(lower right), the high peak of W at $\beta_n = 0.4$ suggests another mechanism that creates large-scale cascades.

Large cascade creation mechanisms

Figure 7 illustrates contour lines of cascade size W and growth rate Q/T on the $\alpha - \lambda_s$ plane. Figure 7(upper left) exhibits the case wherein large-scale cascades emerge because $W = 16$ implies $W \times N_i = 160,000$ retweets. This large cascade creation mechanism at $\beta_n = 0$ is referred to as the *endogenous mechanism*. The figure exhibits that

$$\alpha_W < 0.01. \tag{36}$$

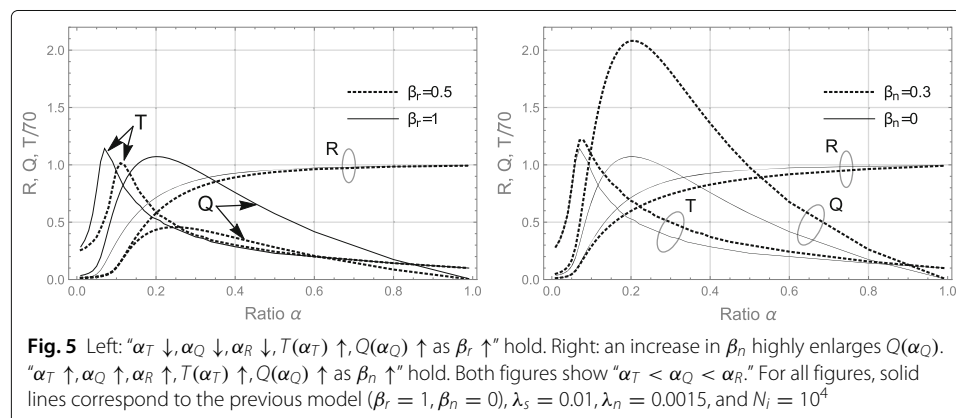
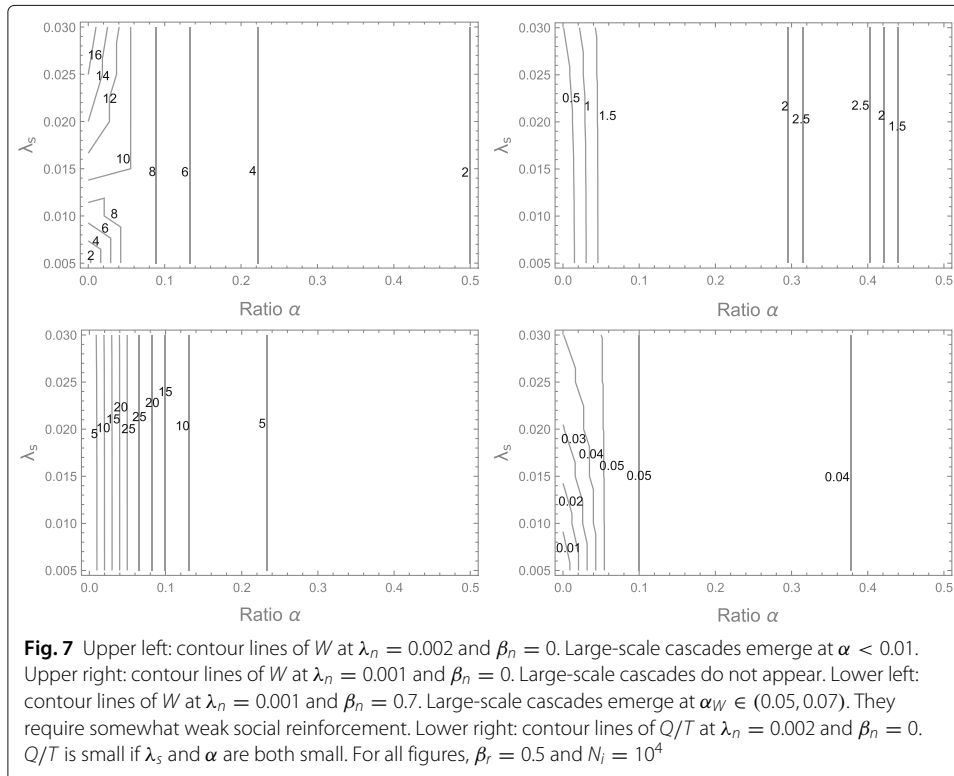
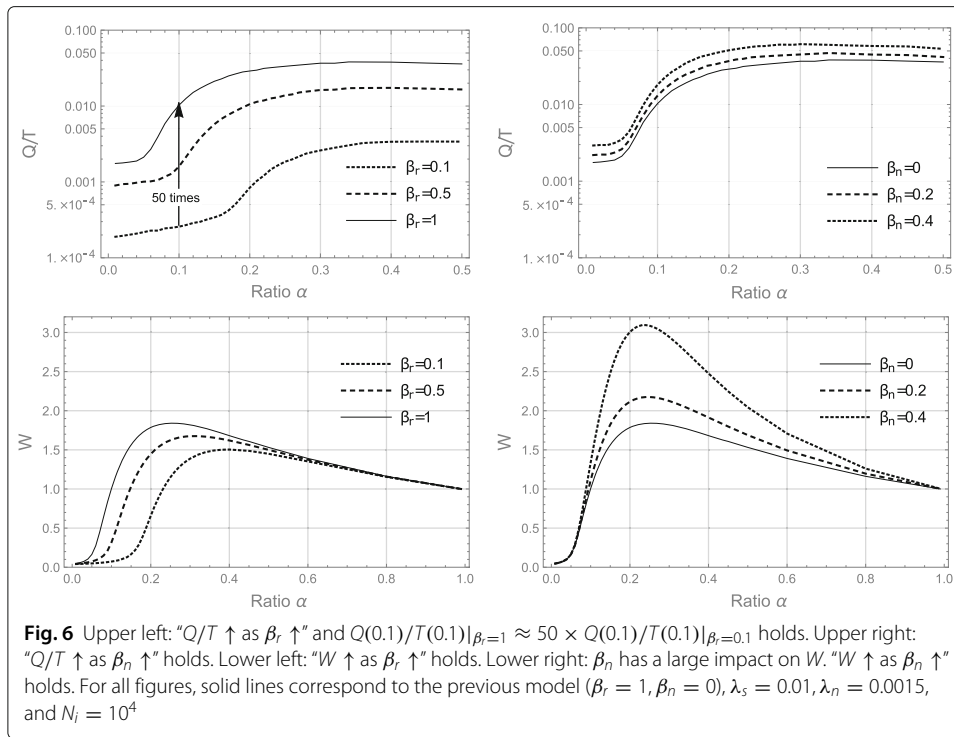


Fig. 5 Left: “ $\alpha_T \downarrow, \alpha_Q \downarrow, \alpha_R \downarrow, T(\alpha_T) \uparrow, Q(\alpha_Q) \uparrow$ as $\beta_r \uparrow$ ” hold. Right: an increase in β_n highly enlarges $Q(\alpha_Q)$. “ $\alpha_T \uparrow, \alpha_Q \uparrow, \alpha_R \uparrow, T(\alpha_T) \uparrow, Q(\alpha_Q) \uparrow$ as $\beta_n \uparrow$ ” hold. Both figures show “ $\alpha_T < \alpha_Q < \alpha_R$.” For all figures, solid lines correspond to the previous model ($\beta_r = 1, \beta_n = 0$), $\lambda_s = 0.01, \lambda_n = 0.0015$, and $N_i = 10^4$



Inequality (36) coincides with (16). The figure also indicates that W significantly decreases at low ratios α as λ_s decreases to 0.005. This result also agrees with the sparse community effect because (6) is equivalent to

$$\alpha_Q \uparrow, \alpha_R \uparrow \text{ as } \lambda_s \downarrow. \tag{37}$$

It must be noted that the cascade size is determined not by λ_s (content infectivity) but by α (network structure) when $\alpha > 0.1$.

Figure 7(upper right) is obtained using the same parameter values as Fig. 7(upper left) except that λ_n is reduced from 0.002 to 0.001. Large-scale cascades do not appear in this case. The maximum W falls to 2.5 and α_W increases and satisfies

$$0.3 < \alpha_W < 0.4, \tag{38}$$

which also coincides with (16). This result implies that large α values in (38) enlarge small- and middle-scale cascades; i.e., non-viral content requires social reinforcement. The result also demonstrates that large-scale cascades require not only a large λ_s but also a large λ_n .

Figure 7(lower left) is calculated using the same parameter values as Fig. 7(upper right) except that β_n is increased from 0 to 0.7. In this case, large-scale cascades again emerge because the maximum of W is 25. This large cascade creation mechanism is referred to as the *exogenous mechanism*. From the figure,

$$0.05 < \alpha_W < 0.07, \tag{39}$$

which indicates that compared with (36), this mechanism requires somewhat weak social reinforcement. This result agrees with (32).

Figure 7(lower right) illustrates contour lines of growth rate Q/T obtained using the same parameter values as Fig. 7(upper left). It can be observed that the $\alpha - \lambda_s$ plane is separated into dynamic (high Q/T) and static (low Q/T) areas. The static area is located in a small area where λ_s and α are both small. This result agrees with (13), Figs. 3, 6(upper left), and 6(upper right). From Fig. 7(upper left) and (lower right), W and Q/T are not correlated and they have the maximum values at different (α, λ_s) points.

Discussion

Implications of the findings

There are many studies that focus on the prediction of large-scale information cascades. Figure 7(upper left) demonstrates the potential difficulty in predicting such information cascades. From the figure, it can be observed that not only large-scale cascades but also small- and middle-scale cascades appear at low ratios and their future sizes are highly sensitive to α , λ_s , and λ_n . The same result can be seen in Figs. 2(upper left) and 2(upper right), where R and Q vary significantly near ratio zero when λ_s and λ_n are high. It is not easy to measure these rates accurately at the initial stage of a cascade because the sample size at that stage is small. Even if the rates are accurately measured, the future size will, in general, have a certain probability distribution. For example, W at $(\alpha, \lambda_s) = (0.01, 0.01)$ in Fig. 7(upper left) is not deterministically given but will take values such as 4, 6, and 8. Furthermore, ratio α , a network density indicator, itself may vary with time (Antoniades and Dovrolis 2015; Oida and Okubo 2019). This paper considers the ratio as time-invariant; however, if TRF events affect the network densities of Twitter

communities (Antoniades and Dovrolis 2015), the prediction of viral content becomes even more difficult.

This paper incorporated exogenous link-creation events into the model and demonstrated that these events have the potential to create large-scale cascades. According to (Antoniades and Dovrolis 2015), their occurrence probability is significantly low (two or three orders of magnitude lower than the occurrence probability of TRF events). However, big data analytical services, such as personalized recommendation in social networking sites, may boost exogenous link creation in the future.

This paper presents a simplified model, focusing on two Twitter properties: the TRF events and the existence of a dominant user with a large number of followers; nevertheless, the model reproduces social reinforcement and structural trapping effects and presents many interesting causal relations related to Questions 1 and 2. This outcome suggests that these properties are important factors for understanding real network dynamics.

Limitations of this approach

The results in the paper are only hypotheses and must be validated by detailed event-driven simulations or empirical evidence. Because of the uniformly-connected homogeneous agent and synchronous retweet assumptions (Asms 1–3), the model presents the average behavior of the retweet diffusion phenomena. To measure the effects of variations in agent behavior or network structure, one approach is to introduce a heterogeneous agent model with some network topological features. Another approach is to extend the model such that communities or sub-communities inside a community have different attribute values. The latter approach minimizes the complexity of the model and is therefore suited to simulate large-scale diffusion processes. Using the recurrence relations and Algorithm 1 in the paper, the heterogeneous (sub)community model can be easily derived.

Conclusions

This paper addressed the following questions: (1) Can the underlying network be considered static while a diffusion process is underway? (2) Why does viral content not require strong social reinforcement? The paper introduced a recurrence relation model that describes the retweet diffusion and community growth dynamics, in which one dominant user (initiator) with a large number of followers posts a tweet, then retweets by followers propagate the tweet, and TRF and exogenous link-creation events generate new followers. Consequently, the community comprising the followers of the initiator grows. The paper focused on the ratio of community members in the followers of each community member (α), which is also a network density indicator of the community. The model not only illustrates the structural trapping and social reinforcement effects but also presents several important relations among three ratios α_T, α_Q , and α_R , three quantities R, Q , and T , and four parameters $\lambda_s, \lambda_n, \beta_r$, and β_n . Above all, the magnitude relation ($\alpha_T < \alpha_Q < \alpha_R$), the sparse community effect, and the result that $Q(\alpha_Q)$ grows and $T(\alpha_T)$ falls as λ_s increases were directly associated with the two questions.

The computational results derived from the model answered the two questions as follows:

(1) The criterion for network variability was the community growth rate. Except for the cases in which the retweeter's follow rate β_r is significantly small or the community size is extremely large, the growth rate is small when retweet rate λ_s and ratio α are both small. If the condition does not hold, the growth rate may rise by more than 10 times; i.e., the networks are potentially dynamic. Therefore, depending on the evaluation conditions of diffusion models, the effect of topological changes must be considered.

(2) There are two mechanisms that create large-scale cascades: the endogenous and exogenous mechanisms. Currently, only the endogenous mechanism is functional in the Twitter network. Both mechanisms demonstrated that the cascade size is maximized at a low ratio α (i.e., at a low network density), at which the social reinforcement effect is considered to be low. Thus, the model reproduced the evidence that viral content does not require strong social reinforcement. This result is closely related to the sparse community effect, which indicates that as the retweet rates increase, the α value that maximizes the cascade size decreases. Therefore, large-scale cascades emerge *only if* the ratio is low.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Author's contributions

The author conducted this study alone. The author read and approved the final manuscript.

Competing interests

The author declare that no competing interests exist.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 13 November 2018 Accepted: 22 April 2019

Published online: 31 May 2019

References

- Antoniades D, Dovrolis C (2015) Co-evolutionary dynamics in social networks: A case study of twitter. *Comput Soc Netw* 2(1):14
- Aragón P, Gómez V, García D, Kaltenbrunner A (2017) Generative models of online discussion threads: state of the art and research challenges. *J Internet Serv Appl* 8(1):15
- Aral S, Dhillon PS (2018) Social influence maximization under empirical influence models. *Nat Hum Behav* 2(6):375
- Banisch S, Olbrich E (2019) Opinion polarization by learning from social feedback. *J Math Sociol* 43(2):76–103
- Bao Q, Cheung WK, Zhang Y, Liu J (2017) A component-based diffusion model with structural diversity for social networks. *IEEE Trans Cybernet* 47(4):1078–1089
- Bose A, Shin KG (2008) On capturing malware dynamics in mobile power-law networks. In: *Proceedings of the 4th International Conference on Security and Privacy in Communication Networks*. ACM. p 12
- Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J (2014) Can cascades be predicted?. In: *Proceedings of the 23rd International Conference on World Wide Web*. ACM. pp 925–936
- Cheung M, She J, Junus A, Cao L (2017) Prediction of virality timing using cascades in social media. *ACM Trans Multimed Comput Commun Appl (TOMM)* 13(1):2
- Farajtabar M, Wang Y, Gomez-Rodriguez M, Li S, Zha H (2017) Coevolve: A joint point process model for information diffusion and network evolution. *J Mach Learn Res* 18:1–49
- Goel S, Anderson A, Hofman J, Watts DJ (2015) The structural virality of online diffusion. *Manag Sci* 62(1):180–196
- Goel S, Watts DJ, Goldstein DG (2012) The structure of online diffusion networks. In: *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM. pp 623–638
- Hours H, Fleury E, Karsai M (2016) Link prediction in the twitter mention network: impacts of local structure and similarity of interest. In: *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference On*. IEEE. pp 454–461
- Jin H (2017) Detection and characterization of influential cross-lingual information diffusion on social networks. In: *Proceedings of the 26th International Conference on World Wide Web Companion*. pp 741–745. International World Wide Web Conferences Steering Committee

- Junus A, Ming C, She J, Jie Z (2015) Community-aware prediction of virality timing using big data of social cascades. In: Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference On. IEEE. pp 487–492
- Kempe D, Kleinberg J, Tardos É (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. ACM. pp 137–146
- Kempe D, Kleinberg J, Tardos É (2005) Influential nodes in a diffusion model for social networks. In: International Colloquium on Automata, Languages, and Programming. Springer. pp 1127–1138
- Krishnan S, Butler P, Tandon R, Leskovec J, Ramakrishnan N (2016) Seeing the forest for the trees: new approaches to forecasting cascades. In: Proceedings of the 8th ACM Conference on Web Science. ACM. pp 249–258
- Leskovec J, Backstrom L, Kumar R, Tomkins A (2008) Microscopic evolution of social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. pp 462–470
- Li C-T, Lin Y-J, Yeh M-Y (2015) The roles of network communities in social information diffusion. In: Big Data (Big Data), 2015 IEEE International Conference On. IEEE. pp 391–400
- Myers SA, Leskovec J (2014) The bursty dynamics of the twitter information network. In: Proceedings of the 23rd International Conference on World Wide Web. ACM. pp 913–924
- Moore C, Newman ME (2000) Epidemics and percolation in small-world networks. *Phys Rev E* 61(5):5678
- Moreno Y, Nekovee M, Pacheco AF (2004) Dynamics of rumor spreading in complex networks. *Phys Rev E* 69(6):066130
- Oida K, Okubo K (2019) Adopter community formation accelerated by repeaters of product advertisement campaigns. *IEEE Trans Comput Soc Syst* 6(1):56–72
- Szabo G, Huberman BA (2010) Predicting the popularity of online content. *Commun ACM* 53(8):80–88
- Tur EM, Zeppini P, Frenken K (2018) Diffusion with social reinforcement: The role of individual preferences. *Phys Rev E* 97(2):022302
- Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
- Weng L, Menczer F, Ahn Y-Y (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
- Weng L, Menczer F, Ahn Y-Y (2014) Predicting successful memes using network and community structure. In: ICWSM
- Weng L, Ratkiewicz J, Perra N, Gonçalves B, Castillo C, Bonchi F, Schifanella R, Menczer F, Flammini A (2013) The role of information diffusion in the evolution of social networks. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. pp 356–364
- Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J (2015) Seismic: A self-exciting point process model for predicting tweet popularity. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. pp 1513–1522

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
