Applied Network Science

RESEARCH

Open Access

# Information cascade final size distributions derived from urn models

Kazumasa Oida[1*]

*Correspondence:
oida@fit.ac.jp

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

**Abstract**

Bipolarization is a phenomenon in which either a large or very small information cascade appears randomly when the retweet rate is high. This phenomenon, which has been observed only in simulations, has the potential to significantly advance the prediction of final cascade sizes because forecasters need only focus on the two peaks in the final cascade size distribution rather than considering the effects of various details, such as network structure and user behavioral patterns. The phenomenon also suggests the difficulty of identifying factors that lead to the emergence of large-scale cascades. To verify the existence of bipolarization, this paper theoretically derives mathematical expressions of the cascade final size distribution using urn models, which simplify the diffusion behavior of actual online social networks. Under the assumption of infinite network size, the distribution exhibits power-law behavior, consistent with the results of existing diffusion models and previous Twitter analytical outcomes. Under the assumption of finite network size, bipolarization is observed.

**Keywords:** Information diffusion, Online social network, Cascade size distribution, Bipolarization, Twitter

## Introduction

A large-scale information cascade is a phenomenon in which attractive (viral) content spreads to a large number of online social network (OSN) users. The problem of predicting the final size of a cascade when its size is small has been studied for more than twenty years, where the size is the number of users who have shared the information. Some predicting approaches focus on community properties (Weng et al. 2013, 2014; Junus et al. 2015; Bao et al. 2017). Other approaches focus on the details of the cascade and network structures (Zhao et al. 2015; Yu et al. 2015; Li et al. 2015; Krishnan et al. 2016; Cheung et al. 2017). Recent approaches often employ deep neural network technologies (Bourigault et al. 2016; Wang et al. 2017, 2018; Horawalavithana et al. 2020).

If the prediction problem can be solved, then (1) content delivery systems can be made more efficient by moving viral contents to servers close to the viewers, (2) information that people are interested in can be used more quickly for stock investments and product development, and (3) prediction technologies can be applied to viral marketing, which creates large-scale cascades on purpose.

Large-scale cascades rarely occur (Goel et al. 2012; Cheng et al. 2014), so available datasets for their prediction are currently insufficient. As such, it is not always easy to apply machine learning techniques and verify the statistical significance for proposed methods. Meanwhile, simulation-based forecasting research is progressing. A phenomenon in which large and very small cascades appear randomly under a high retweet rate has been reported in Oida (2021). This phenomenon is referred to as bipolarization because the two peaks of the cascade size distribution move apart as the retweet rate increases.

Bipolarization is universal in that it emerges regardless of network topologies (small-world (Watts and Strogatz 1998), scale-free (Barabási et al. 1999), Erdős–Rényi (Batagelj and Brandes 2005), and existing OSNs), existence of various types of communities (Forestier et al. 2015; Baldesi et al. 2018), and user behavioral patterns (social reinforcement (Weng et al. 2013, 2014), user response times (Zhou et al. 2017; Xie et al. 2011), and repeated exposure (Zhou et al. 2015)). This phenomenon was observed in event-driven simulations and was reproduced by the urn model, which is a model for simplifying the Twitter-type information diffusion mechanism (Oida 2021). Bipolarization appears when the network size is finite.

This paper mathematically formulates the urn model to theoretically verify this novel discovery. The contributions of this paper can be summarized as follows:

1. This paper deals analytically with the case where the network size is finite (although the case where the size is infinite is much simpler). This is a realistic approach because existing network sizes are all finite. The effects of network properties and user behavior can be rigorously assessed through partially modifying the derived equations in this paper.
2. This paper is the first to theoretically prove the bipolarization phenomenon.
3. The formula for infinite network size can be applied to the case where most of the cascades are sufficiently smaller than the network size. The cascade size distribution derived from the formula shows a power law over a certain range. This is consistent not only with the results of previous Twitter data analyses (Bakshy et al. 2011) (because most existing cascades are small) but also with those of existing diffusion models (Wegrzycki et al. 2017; Gleeson et al. 2020).

The remainder of this paper is organized as follows. Section  describes related studies. Section  introduces the urn model. Section  formulates the model as a Markov chain. Section  proves many propositions derived from the Markov chain. Section  numerically evaluates the derived equations to investigate the shape of the cascade final size distributions under the finite and infinite network size assumptions. Section  discusses the implication of the findings, and finally, Sect.  concludes the paper.

## Related work

The term “urn models” generally refers to the systems of one or more urns containing objects of various types (colored balls in the usual setting) (Mahmoud 2008). The systems evolve in time, subject to rules of drawing balls and throwing balls into the urns. These models have helped reveal various phenomena, including the evolution of species

in biology, particle systems in chemistry, and formation of social networks in sociology (Mahmoud 2008).

A variety of urn models have been proposed for years, especially in the field of mathematics (Pemantle 2007). In recent reinforcement models, one or more balls are extracted and returned to the urn with additional balls. The number and colors of the additionally returned balls are determined by the colors of extracted ones (Rafik et al. 2019; Crimaldi et al. 2022). These studies focus primarily on asymptotic properties, such as the limit value of the proportion of balls of a certain color in an urn. The model in this paper is simple in that one ball is extracted and one ball is returned; therefore, it is not a reinforcement model. The uniqueness of this study is that a ball to be returned may not be an extracted one and that the number of trials of extracting a ball is increased, not the number of balls to be returned.

The reinforcement model is also used to model various information diffusion phenomena or to predict them through numerical computation. In (Hino et al. 2016), Pólya's urn model was introduced to observe phase transitions in the information cascade. The model was also used for reproducing trajectories of innovation diffusion (Dosi et al. 2019) and for predicting statistical laws for the rate at which novelties (e.g., discovery of new songs and ideas) happen through social interaction (Tria et al. 2014; Di Bona et al. 2022).

Let us next discuss previous studies focusing on information cascade sizes. The authors of Wegrzycki et al. (2017) presented a theoretical proof that the sizes of cascades generated by the cascade generation model (CGM) (Leskovec et al. 2007) follow a power-law distribution. In Gleeson et al. (2020), information spread was modeled with a branching process, which also presented power-law behavior over a limited range. Both approaches are practical in that their models were effective for fitting empirical data. However, they did not clearly quantify the effects of finiteness of the network size on cascade sizes. This paper extends their work by articulating the impact of the finiteness from a different perspective.

## Proposed model

Figure 1(left) shows the urn model that represents the mechanism of spreading retweets to followers. Table 1 describes the symbols in the model and corresponding OSN quantities. There are $N$ balls in the urn, and each is either black or white. The OSN quantity

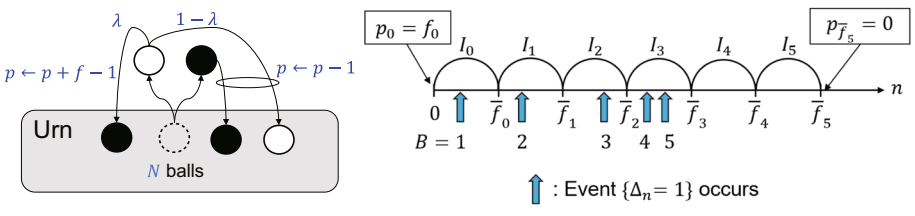


**Fig. 1** Left: In the urn model, a trial of extracting and returning a ball is repeatedly conducted until $p$ reaches zero. $p$ is increased by $f-1$ if a white ball is extracted and a black ball is returned with probability $\lambda$. In all other cases, $p$ is decremented by one after each trial. In the figure, "$p \leftarrow p+f-1$" denotes that the new value of $p$ is $p+f-1$. Right: One of the cases of $B_\tau = 5$. Event $p=0$ occurs at the end of one of the intervals $I_0, I_1, I_2, \ldots$

**Table 1** Symbols in the urn model and corresponding OSN quantities

| Symbol | Urn model | OSN |
|---|---|---|
| $p$ | Number of remaining trials | Number of unread retweets |
| $f$ | Number of trials to be increased | Number of followers |
| $N$ | Total number of balls in the urn | Total number of users (network size) |
| $B$ | Number of black balls in the urn | Number of users who have retweeted |
| $\lambda$ | Probability to return a black ball | Retweet probability |

corresponding to a black (white) ball is a user who has (not) posted a retweet. Thus, the initial condition is that all balls are white. The model repeats a trial in which a ball is randomly extracted from the urn and then a ball is returned to the urn until $p = 0$, where $p$ is the number of remaining trials, and its initial value is $f(> 0)$. As shown in Table 1, $p$ corresponds to the number of retweet messages that have not been read yet and $f$ the number of followers.

In the OSN diffusion mechanism, a user posts a retweet with probability $\lambda$ after reading an arrival retweet if the user has not posted a retweet yet. In this case, the number of unread retweets $p$ increases by $f - 1$ (because the user has read the retweet and the number of followers is $f$), and the number of users who have retweeted $B$ increases by one. If a retweet message arrives at a user who has already posted a retweet or at a user who decides not to post a retweet, $p$ decreases by one and $B$ is unchanged. The number $B$ at $p = 0$ corresponds to the final size of the cascade.

The urn model in Fig. 1(left) follows the above-mentioned OSN diffusion mechanism. If a white ball is taken out of the urn, a black ball is returned to the urn with probability $\lambda$ (this ball-swapping rule represents the process in which a user who has not retweeted turns into a user who has retweeted). If this happens, $p$ is incremented by $f - 1$. In all other cases, the extracted ball is returned and $p$ is decreased by one.

## Formulation

This section defines a stochastic process from the urn model. Let $p_n$ and $B_n$ be $p$ and $B$ immediately after the $n$-th trial, respectively. The stopping time $\tau$ of the trial is defined as

$$\tau := \inf\{n \geq 1 | p_n = 0\}. \tag{1}$$

$B_\tau$ is determined by the Markov chain $X_n = (B_n, p_n)$ given by

$$X_0 = (0, f_0), \tag{2}$$

$$X_{n+1} = (B_n + \Delta_{n+1}, p_n + \Delta'_{n+1}), \tag{3}$$

where $\Delta_n \in \{0, 1\}$ and $\Delta'_n \in \{-1, f - 1\}$ represent increases in $B$ and $p$ just after the $n$-th trial, respectively. Note from the previous section that $\{\Delta_n = 0\} = \{\Delta'_n = -1\}$, $\{\Delta_n = 1\} = \{\Delta'_n = f - 1\}$, and

$$P(\Delta_{n+1} = 1 | B_n = i) = 1 - P(\Delta_{n+1} = 0 | B_n = i) = \frac{N - i}{N}\lambda. \tag{4}$$

This paper assumes that $\lambda$ is constant and $f$ may vary with time. Let $f_k$ be the $k$-th value of $f$ ($f_0$ is the initial value of $p$). In the OSN setting, $f_k$ corresponds to the number of followers of the user who posted the $k$-th retweet. Let

$$I_k := (\bar{f}_{k-1}, \bar{f}_k],\ k = 0, 1, \ldots, \tag{5}$$

where $\bar{f}_k := \sum_{i=0}^{k} f_i$ and $I_0 := (0, \bar{f}_0]$. Figure 1(right) shows one of the cases where $B_\tau = 5$ occurs. This event occurs if and only if events $\Delta_n = 1$ do not occur at $n \in I_5$, they occur five times at $n \in \cup_{k=0}^{4} I_k$, at least four times at $n \in \cup_{k=0}^{3} I_k$, at least three times at $n \in \cup_{k=0}^{2} I_k$, at least twice at $n \in \cup_{k=0}^{1} I_k$, and at least once at $n \in I_0$.

**Proposition 1** *For any integer $k \geq 0$,*

$$\{B_\tau = k\} = \{\tau = \bar{f}_k\}. \tag{6}$$

***Proof***

*As shown in Fig. 1(right), event $p = 0$ occurs only at the end of one of intervals $I_0, I_1, I_2, \ldots$. If $B_\tau = k$, the trials are not conducted at $\bar{f}_j$, $j > k$ and $\tau$ is not smaller than $\bar{f}_k$ because if $\tau = \bar{f}_j$, $j < k$, $B_\tau$ must be smaller than $k$. Thus, $\{B_\tau = k\} \subset \{\tau = \bar{f}_k\}$. If $\tau = \bar{f}_k$, $B_{\bar{f}_k} = k$. Thus, $\{\tau = \bar{f}_k\} \subset \{B_\tau = k\}$.* □

## Calculation of $P(B_\tau)$

### Infinite network size

This section calculates the probability of $B_\tau$ when the total number of balls $N$ (referred to as the network size) is infinite. Proposition 1 indicates that $\{B_\tau = k\}$ depends only on the first $\bar{f}_k$ trials $\Delta_1, \ldots, \Delta_{\bar{f}_k}$. Because $\Delta_n \in \{0, 1\}$, $\omega = (\Delta_1, \ldots, \Delta_{\bar{f}_k})$ takes one of $2^{\bar{f}_k}$ different binary sequences. Let $\binom{a}{b\ c}$ be the number of events $\omega$ that satisfy the condition that $a$ trials generate $b$ events $\Delta_n = 1$ and $\tau = c$. Then,

$$|\{\omega | B_{\tau(\omega)} = k\}| = \binom{\bar{f}_k}{k\ \bar{f}_k}. \tag{7}$$

If $N$ is infinite, the right-hand side of (4) is $\lim_{N\to\infty} \frac{N-i}{N}\lambda = \lambda$, so $P(\Delta_{n+1} = 1 | B_n)$ is constant $\lambda$ regardless of $B_n$. Thus, by using $X_0 = (0, f_0)$,

$$P(B_\tau = 0) = P(B_0 = 0)P(\Delta_1 = 0 | B_0 = 0) \cdots P(\Delta_{f_0} = 0 | B_{f_0 - 1} = 0) = (1 - \lambda)^{f_0}, \tag{8}$$

and for $k \geq 0$,

$$P(B_\tau = k) = \binom{\bar{f}_k}{k\ \bar{f}_k} \lambda^k (1 - \lambda)^{\bar{f}_k - k}, \tag{9}$$

where $\binom{\bar{f}_0}{0\ \bar{f}_0} = |\{\omega | B_\tau = 0\}| = 1$.

**Proposition 2** *For any integer* $k > 0$,

$$\begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix} = \begin{pmatrix} \bar{f}_{k-1} \\ k \end{pmatrix} - \sum_{m=0}^{k-2} \begin{pmatrix} \bar{f}_{k-1} \\ k \ \bar{f}_m \end{pmatrix}. \tag{10}$$

***Proof***

*From Proposition 1,*

$$\{B_\tau = k\} = \{B_{\bar{f}_k} = k\} \cap \{\tau = \bar{f}_k\}. \tag{11}$$

Because $p_n$ becomes zero only at $n = \bar{f}_0, \bar{f}_1, \bar{f}_2, \ldots$,

$$\{B_{\bar{f}_k} = k\} = \{B_{\bar{f}_k} = k\} \cap \bigcup_{m=0}^{k} \{\tau = \bar{f}_m\}. \tag{12}$$

Note that $\{\tau = \bar{f}_m\}$, $m = 1, 2, \ldots, k$, are disjoint sets. From (11) and (12),

$$\{B_\tau = k\} = \{B_{\bar{f}_k} = k\} \setminus \{B_{\bar{f}_k} = k\} \cap \bigcup_{m=0}^{k-1} \{\tau = \bar{f}_m\}. \tag{13}$$

Because

$$|\{B_{\bar{f}_k} = k\}| = \begin{pmatrix} \bar{f}_{k-1} \\ k \end{pmatrix}, \tag{14}$$

$$|\{B_{\bar{f}_k} = k\} \cap \bigcup_{m=0}^{k-1} \{\tau = \bar{f}_m\}| = \sum_{m=0}^{k-1} \begin{pmatrix} \bar{f}_{k-1} \\ k \ \bar{f}_m \end{pmatrix}, \tag{15}$$

$$\begin{pmatrix} \bar{f}_{k-1} \\ k \ \bar{f}_{k-1} \end{pmatrix} = 0, \tag{16}$$

(10) holds. □

**Proposition 3** *For any integers* $k$ *and* $m$ *satisfying* $0 < m < k$,

$$\begin{pmatrix} \bar{f}_{k-1} \\ k \ \bar{f}_m \end{pmatrix} = \begin{pmatrix} \bar{f}_m \\ m \ \bar{f}_m \end{pmatrix} \begin{pmatrix} \bar{f}_{k-1} - \bar{f}_m \\ k - m \end{pmatrix}. \tag{17}$$

***Proof***

$\begin{pmatrix} \bar{f}_{k-1} \\ k \ \bar{f}_m \end{pmatrix}$ *is the number of binary sequences* $(\Delta_1, \ldots, \Delta_{\bar{f}_{k-1}})$ *in which events* $\Delta_n = 1$ *occur* $k$ *times and* $\tau = \bar{f}_m$. *The number of binary sequences* $(\Delta_1, \ldots, \Delta_{\bar{f}_m})$ *satisfying* $\tau = \bar{f}_m$ *is* $\begin{pmatrix} \bar{f}_m \\ m \ \bar{f}_m \end{pmatrix}$, *and the number of sequences* $(\Delta_{\bar{f}_m+1}, \ldots, \Delta_{\bar{f}_{k-1}})$ *in which events* $\Delta_n = 1$ *occur* $k - m$ *times is* $\begin{pmatrix} \bar{f}_{k-1} - \bar{f}_m \\ k - m \end{pmatrix}$. □

Propositions 2 and 3 indicate that $\begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix}$ is obtained by using $\begin{pmatrix} \bar{f}_m \\ m \ \bar{f}_m \end{pmatrix}$, $m = 0, 1, \ldots, k-2$, where $\begin{pmatrix} f_0 \\ 0 \ f_0 \end{pmatrix} = 1$. Therefore, from (9), $P(B_\tau = k)$ can be derived in ascending order of $k$.

#### Finite network size

Let us next consider the case where $N$ is finite, and let $\lambda_i := \frac{N-i}{N}\lambda$. In this case, probability $P(B_\tau = k)$ becomes a complex formula, and its calculation time rises sharply as $k$ increases. Therefore, the upper and lower bounds of $P(B_\tau = k)$ are derived as a first step.

Assume that events $\Delta_n = 1$ occur $k$ times at $n = n_1, n_2, \ldots, n_k$, where $1 \le n_1 < n_2 < \cdots < n_k \le \bar{f}_{k-1}$. Let $c = \begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix} \prod_{i=0}^{k-1} \lambda_i$ and

$$L(n_1, n_2, \ldots, n_k) := \prod_{i \in \{1,2,\ldots,\bar{f}_k\} \setminus \{n_1,\ldots,n_k\}} P(\Delta_i = 0 | B_{i-1}). \tag{18}$$

**Proposition 4** *For $k > 0$, $P(B_\tau = k)$ satisfies*

$$cL(\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_{k-1}) \le P(B_\tau = k) \le cL(1, 2, \ldots, k). \tag{19}$$

***Proof***

*Let $\omega^*$ ($\omega_*$) be one of the $\omega$ values in $\{\omega | B_{\tau(\omega)} = k\}$ that maximizes (minimizes) probability $P(\omega)$. From (7),*

$$\begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix} P(\omega_*) \le P(B_\tau = k) \le \begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix} P(\omega^*). \tag{20}$$

In the following, $P(\omega^*)$ and $P(\omega_*)$ are derived. From (4),

$$P(\Delta_{n_1} = 1 | B_{n_1 - 1} = 0) \cdots P(\Delta_{n_k} = 1 | B_{n_k - 1} = k - 1) = \prod_{i=0}^{k-1} \lambda_i, \tag{21}$$

which is independent of $(n_1, \ldots, n_k)$. For any $\omega$, $P(\omega)$ is given as

$$P(\omega) = L(n_1, n_2, \ldots, n_k) \prod_{i=0}^{k-1} \lambda_i. \tag{22}$$

From (4), for any $k$-tuple $(n_1, \ldots, n_k)$, $L(n_1, \ldots, n_k)$ strictly decreases as $n_i \in \{n_1, \ldots, n_k\}$ increases. Note that $B_\tau = k$ if $(n_1, \ldots, n_k)$ is equal to $(1, 2, \ldots, k)$ or $(\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_{k-1})$, and $B_\tau \ne k$ if $n_j > \bar{f}_{j-1}$ for any $j, 1 \le j \le k$. Accordingly, $L$ is minimized (maximized) if events occur at $(n_1, \ldots, n_k) = (\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_{k-1})$ ($(n_1, \ldots, n_k) = (1, 2, \ldots, k)$). □

Note that

$$L(1, 2, \ldots, k) = (1 - \lambda_k)^{\bar{f}_k - k}, \tag{23}$$

$$L(\bar{f}_0,\bar{f}_1,\ldots,\bar{f}_{k-1}) = \left(\prod_{i=0}^{k-1}(1-\lambda_i)^{f_i-1}\right)(1-\lambda_k)^{f_k}. \tag{24}$$

This paper further improves the upper and lower bounds in (19) by using the following propositions. Let

$$\delta(N,\lambda,j) := \frac{1}{N(\lambda^{-1}-1)+j}. \tag{25}$$

**Proposition 5** *For* $0 < j \le k$, $L$ *defined in* (18) *satisfies*

$$\frac{L(n_1,\ldots,n_j+1,\ldots,n_k)}{L(n_1,\ldots,n_j,\ldots,n_k)} = 1-\delta(N,\lambda,j), \tag{26}$$

$$\frac{L(n_1,\ldots,n_j-1,\ldots,n_k)}{L(n_1,\ldots,n_j,\ldots,n_k)} = 1+\delta(N,\lambda,j-1). \tag{27}$$

***Proof***

*From (4) and (18),*

$$\begin{aligned}
&\frac{L(n_1,\ldots,n_j+1,\ldots,n_k)}{L(n_1,\ldots,n_j,\ldots,n_k)}\\
&= \frac{(1-\lambda_0)^{n_1-1}\cdots(1-\lambda_{j-1})^{n_j}(1-\lambda_j)^{n_j-2}\cdots(1-\lambda_{k-1})^{n_{k-1}-1}(1-\lambda_k)^{n_k}}{(1-\lambda_0)^{n_1-1}\cdots(1-\lambda_{j-1})^{n_j-1}(1-\lambda_j)^{n_j-1}\cdots(1-\lambda_{k-1})^{n_{k-1}-1}(1-\lambda_k)^{n_k}}\\
&= \frac{1-\lambda_{j-1}}{1-\lambda_j}\\
&= \left(1-\frac{N-(j-1)}{N}\lambda\right)\Big/\left(1-\frac{N-j}{N}\lambda\right)\\
&= 1-\frac{1}{N(\lambda^{-1}-1)+j}.
\end{aligned}$$

Similarly,

$$\frac{L(n_1,\ldots,n_j-1,\ldots,n_k)}{L(n_1,\ldots,n_j,\ldots,n_k)} = \frac{1-\lambda_j}{1-\lambda_{j-1}} = 1+\frac{1}{N(\lambda^{-1}-1)+j-1}.$$

□

Proposition 5 implies that an increase in $n_j$ by one results in a decrease in $L$ by $\delta(N,\lambda,j)\times 100\%$. The function $\delta$ rises (falls) with an increase in $\lambda$ ($N$ or $j$). Because $N$ and $\lambda$ are constant in this urn model, $j$ determines $\delta$. From (25), however, the impact of $j$ on $\delta$ may be small because $j < N$ and $\lambda \ll 1$.

Let $L^* := L(1,2,\ldots,k)$ and $L_* := L(\bar{f}_0,\bar{f}_1,\ldots,\bar{f}_{k-1})$, and let us abbreviate $\delta(N,\lambda,j)$ as $\delta_j$. Proposition 5 yields the following.

**Proposition 6**

$$L(n_1, n_2, \ldots, n_k) = L^*(1-\delta_1)^{n_1-1}(1-\delta_2)^{n_2-2}\cdots(1-\delta_k)^{n_k-k} \tag{28}$$

$$= L_*(1+\delta_0)^{\bar{f}_0-n_1}(1+\delta_1)^{\bar{f}_1-n_2}\cdots(1+\delta_{k-1})^{\bar{f}_{k-1}-n_k}. \tag{29}$$

***Proof***

*By iteratively applying (26),*

$$\begin{aligned} L(n_1, n_2, \ldots, n_k) &= L(n_1-1, n_2, \ldots, n_k)(1-\delta_1) \\ &= L(n_1-2, n_2, \ldots, n_k)(1-\delta_1)^2 \\ &= L(1, n_2, \ldots, n_k)(1-\delta_1)^{n_1-1} \\ &= L(1, 2, \ldots, k)(1-\delta_1)^{n_1-1}(1-\delta_2)^{n_2-2}\cdots(1-\delta_k)^{n_k-k}. \end{aligned}$$

(29) is obtained in the same way. □

Let $\Lambda_j := \{j, j+1, \ldots, \bar{f}_{j-1}\}$ and $\Lambda'_j := \{\bar{f}_{j-1}+1, \bar{f}_{j-1}+2, \ldots\}$.

**Proposition 7** *Assume that events* $\Delta_n = 1$ *occur at* $n = n_1, n_2, \ldots$, *where* $0 < n_1 < n_2 < \cdots$.

$B_\tau = k$ *is equivalent to*

$$(n_1, n_2, \ldots, n_k, n_{k+1}) \in \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_k \times \Lambda'_{k+1}. \tag{30}$$

***Proof***

*Assume that (30) does not hold. This happens when at least one of* $n_j$, $1 \le j \le k$, *satisfies* $n_j \notin \Lambda_j$ *or when* $n_{k+1} \notin \Lambda'_{k+1}$. $n_j \notin \Lambda_j$ *implies* $n_j > \bar{f}_{j-1}$ *because* $n_j \ge j$ *due to* $0 < n_1 < n_2 < \cdots$. *If* $n_j > \bar{f}_{j-1}$, $\tau \le \bar{f}_{j-1}$; *therefore,* $B_\tau \le j-1 < k$. *This result does not depend on* $n_{k+1} \in \Lambda'_{k+1}$. *If* $n_i \in \Lambda_i$ *for* $i = 1, 2, \ldots, k$ *and* $n_{k+1} \notin \Lambda'_{k+1}$, $p_{\bar{f}_i} > 0$ *for* $i = 0, 1, \ldots, k$; *therefore,* $B_\tau > k$.

Assume conversely that $B_\tau \neq k$. If $B_\tau = j < k$, $\tau = \bar{f}_j$; therefore, $n_{j+1} > \bar{f}_j$, which is inconsistent with (30). If $B_\tau > k$, $p$ must satisfy $p_{\bar{f}_k} > 0$; accordingly, $n_{k+1} < \bar{f}_k$, which is inconsistent with (30). □

Proposition 7 and (7) imply that $|\bar{\Lambda}_k| = \begin{pmatrix} \bar{f}_k \\ k \end{pmatrix}$, where

$$\bar{\Lambda}_k := \{(\ell_1, \ell_2, \ldots, \ell_k) \in \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_k | \ell_1 < \ell_2 < \cdots < \ell_k\}. \tag{31}$$

Accordingly, $P(B_\tau = k)$ is exactly given as

$$P(B_\tau = k) = \prod_{i=0}^{k-1} \lambda_i \sum_{(n_1, n_2, \ldots, n_k) \in \bar{\Lambda}_k} L(n_1, n_2, \ldots, n_k). \tag{32}$$

The improved upper and lower bounds, denoted as $B^*$ and $B_*$, respectively, are given by

$$B^* = \prod_{i=0}^{k-1} \lambda_i \left( \left( |\bar{\Lambda}_k| - |S^*(k,\epsilon)| \right) L^* + \sum_{(n_1,n_2,\ldots,n_k)\in S^*(k,\epsilon)} L(n_1,n_2,\ldots,n_k) \right), \tag{33}$$

$$B_* = \prod_{i=0}^{k-1} \lambda_i \left( \left( |\bar{\Lambda}_k| - |S_*(k,\epsilon)| \right) L_* + \sum_{(n_1,n_2,\ldots,n_k)\in S_*(k,\epsilon)} L(n_1,n_2,\ldots,n_k) \right), \tag{34}$$

where

$$S^*(k,\epsilon) := \{(\ell_1,\ldots,\ell_k) \in \bar{\Lambda}_k | \frac{\ell_i - i}{\bar{f}_{i-1} - i} \geq \epsilon, i = 1,2,\ldots,k\}, \tag{35}$$

$$S_*(k,\epsilon) := \{(\ell_1,\ldots,\ell_k) \in \bar{\Lambda}_k | \frac{\bar{f}_{i-1} - \ell_i}{\bar{f}_{i-1} - i} \geq \epsilon, i = 1,2,\ldots,k\}. \tag{36}$$

The upper and lower bounds in Proposition 4 correspond to the case of $\epsilon = 1$, at which $B^*$ ($B_*$) takes the maximum (minimum) value. $P(B_\tau = k) = B^* = B_*$ when $\epsilon = 0$. The computation time rises sharply as $\epsilon$ decreases, so $\epsilon$ should be decreased gradually from one.

### Approximation

The calculations of $B^*$ and $B_*$ in (33) and (34) still require large computational capacity. This subsection approximates $P(B_\tau = k)$ in (32). Because $|\bar{\Lambda}_k| = \begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix}$, a part of the right-hand side of (32) can be approximately given by

$$\sum_{(n_1,n_2,\ldots,n_k)\in\bar{\Lambda}_k} L(n_1,n_2,\ldots,n_k) \approx \begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix} \frac{L_1 + L_2 + \cdots + L_m}{m}, \tag{37}$$

where $L_1, L_2, \ldots$ are values of the function $L$ calculated using randomly selected $k$-dimensional coordinates $(n_1,n_2,\ldots,n_k) \in \bar{\Lambda}_k$. Both sides in (37) coincide if $m = \begin{pmatrix} \bar{f}_k \\ k \ \bar{f}_k \end{pmatrix}$. The values of function $L$ are obtained from (28) or (29).

## Numerical results

This section numerically evaluates analytical results derived so far under the condition that follower sizes are constant, i.e., $f_1 = f_2 = \cdots = f_k = f$.

### Infinite network size

This subsection discusses the case of $N = \infty$. Figure 2 shows $P(B_\tau = k)$ in (9) when $f\lambda = 1$, where $f\lambda = 1$ represents an intermediate state between expansion and slow-down of cascade growth because $f\lambda$ can be considered as the expected number of future retweets yielded by one retweet. The figure indicates that the tail of $P(B_\tau = k)$ follows a power law $P(B_\tau = k) \propto k^{-1.5}$ as long as $f\lambda = 1$, regardless of the values of $f$ and $\lambda$.

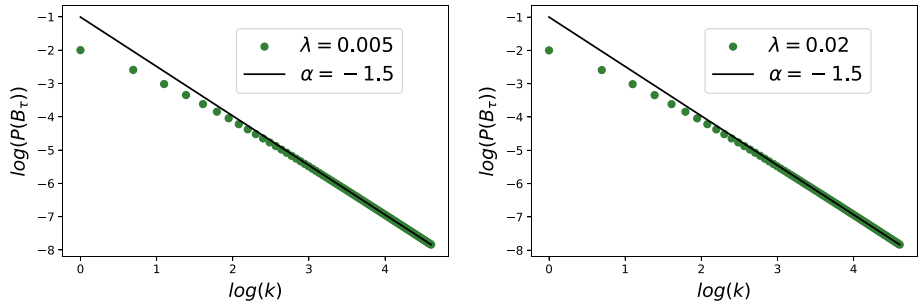


**Fig. 2** The infinite model shows power-law behavior if $f\lambda = 1$. As long as $f\lambda = 1$, the exponent $\alpha$ is $-1.5$ regardless of $f$ and $\lambda$. Left: $f = 200$ and $\lambda = 0.005$. Right: $f = 50$ and $\lambda = 0.002$. The correlation coefficients of $(\log k, \log(P(B_\tau = k)))$, $k \geq 95$, of the two graphs are less than $-0.999999999$

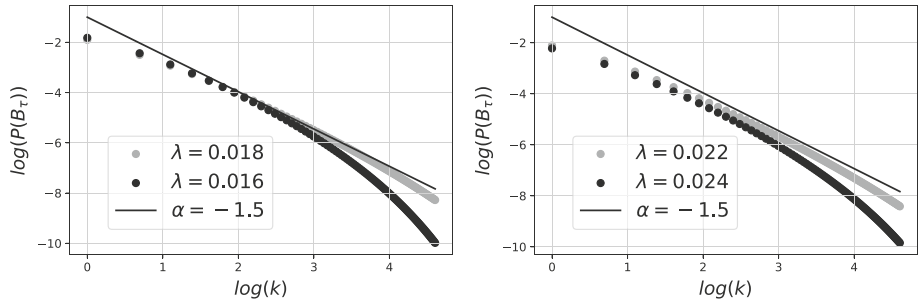


**Fig. 3** The infinite model shows power-law behavior with $\alpha = -1.5$ over a limited range if $f\lambda \neq 1$. $f = 50$. Left: $f\lambda < 1$. Right: $f\lambda > 1$

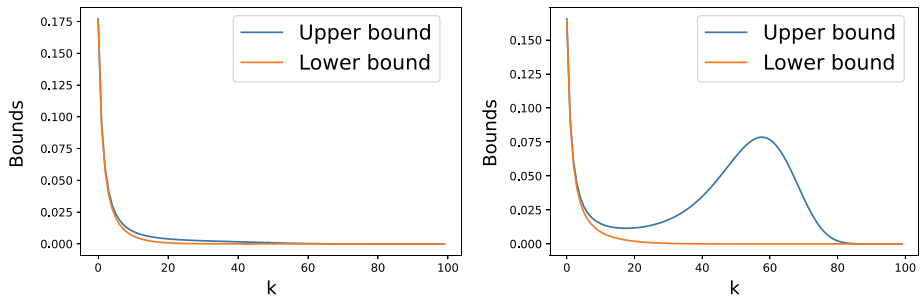


**Fig. 4** Left: The upper and lower bounds monotonically decrease with $k$ when $f\lambda = 0.7$. Right: The upper bound shows another peak, while the lower bound does not when $f\lambda = 0.8$. $N = 100$ and $f = 20$

Figure 3 shows the case when $f\lambda \neq 1$. The tail of $P(B_\tau = k)$ in this case decays faster than that in the case $f\lambda = 1$. It is easy to understand that the tail is short if $f\lambda < 1$ because the cascade size is small in this case. The tail is also short when $f\lambda > 1$ because probability $P(B_\tau = \infty)$ increases. From Figs. 2 and 3, $P(B_\tau = k)$ strictly decreases as $k$ increases, regardless of $f\lambda$, when $N = \infty$.

### Finite network size

This subsection discusses the case of $N < \infty$. Figure 4 shows the upper and lower bounds of $P(B_\tau = k)$ in (19). As shown in Fig. 4(left), both bounds are close and

**Table 2** Percentages of decrease (increase) in the upper (lower) bound. The improvement is small even when $|S^*| = |S_*| = 10^7$. $N = 100$, $f = 50$, and $\lambda = 0.02$

| $\lvert S^*\rvert = \lvert S_*\rvert$ | **0** | $10^4$ | $10^5$ | $10^6$ | $10^7$ |
|---|---|---|---|---|---|
| Upper bound $B^*$ | 0% | 0.016% | 0.076% | 0.096% | 0.126% |
| Lower bound $B_*$ | 0% | 0.037% | 0.135% | 0.173% | 0.211% |

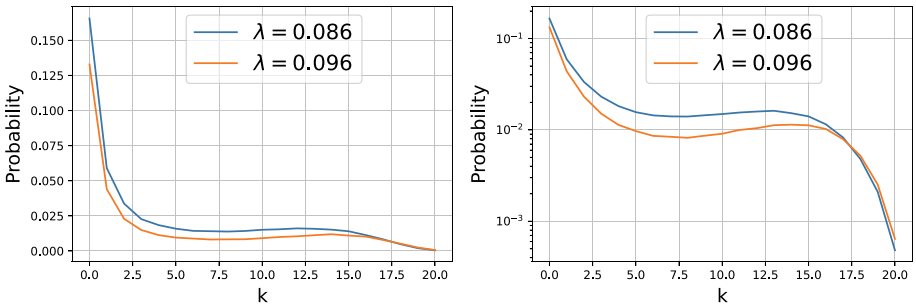


**Fig. 5** Bipolarization phenomenon generated with approximate formula (37). The two graphs are the same except for the scale of the vertical axis. The number of samples, $m$ in (37), is $15 \times k^2$, which provides smooth and stable distribution shape. $N = 20$ and $f = 20$

monotonically decreases with $k$. Accordingly, the distribution of $P(B_\tau = k)$ has a single peak at $k = 0$ when $f\lambda$ is small. If $f\lambda$ rises, as shown in Fig. 4(right), the upper bound shows another peak, while the lower bound does not. The upper bound suggests bipolarization because there are peaks at $k = 0$ and $k = 59$; however, Fig. 4(right) does not prove the existence of the phenomenon because the gap between the two bounds is too wide.

Bounds $B^*$ and $B_*$ in (33) and (34) were obtained to reduce the gap between the upper and lower bounds in (19). Table 2 shows how much these bounds are improved compared with those in (19), which represent the case $|S^*| = |S_*| = 0$. Because the percentages in the table are all small, calculation with larger $|S^*|$ and $|S_*|$ values is needed. However, the calculation requires a powerful computation environment. It took several days to obtain $B^*$ and $B_*$ for $|S^*| = |S_*| = 10^7$ when a new desktop PC was used.

### Approximation

This subsection proves the existence of bipolarization using the approximate formula in (37). Note from Sect.  that bipolarization is identified if there is only one peak at $k \neq 0$. Figure 5 shows that this identification condition holds. The two graphs in the figure are the same except for the scale of the vertical axis. As shown in the figure, the left peak becomes lower and the right peak shifts to the right as $\lambda$ increases. This behavior agrees with the result in Oida (2021).

## Discussion

This paper dealt with two cases of network size $N$: finite and infinite. If network size $N$ is sufficiently greater than cascade size $B$, the infinite model can be used. Contrarily, if $B$ is close to $N$, the effect of finiteness of $N$ becomes dominant. Because the Twitter network

is considerably large and the majority of Twitter cascades are small, the infinite model can be used for comparison with previous Twitter data analytical results.

The infinite model showed that $B_\tau$ follows a power law if $f\lambda = 1$ (Fig. 2) and that even if $f\lambda \neq 1$, the decay exponent is approximately $-1.5$ over a limited range (Fig. 3). According to Bakshy et al. (2011), in which $1.03 \times 10^9$ tweets and $74 \times 10^6$ diffusion events were collected over the two-month period of Sep. 13 to Nov. 15, 2009, the cascade size has a power-law distribution, and interestingly, the exponent is $-1.5$ over the range of $[10^1, 10^3]$.

Some researchers have collected large-scale cascade samples to identify the sources of large-scale cascade emergence (Zhao et al. 2015; Yu et al. 2015; Li et al. 2015; Krishnan et al. 2016; Cheung et al. 2017; Bakshy et al. 2011; Cheng et al. 2014), while some attempted to extract features of large cascade samples using machine learning algorithms (Bourigault et al. 2016; Wang et al. 2017, 2018; Horawalavithana et al. 2020; Zhou et al. 2021). The outcomes of this paper imply that these approaches might bring contradictory or ambiguous consequences regarding the effects of network properties and user behavior on cascade growth due to the stochastic duality, which implies the possibility of two extreme outcomes occurring under the same conditions (i.e., cascades may accidentally live long or may die immediately).

As a next step of this work, the following prediction method is promising. According to Fig. 4, one of the distribution peaks is at $k = 0$. Thus, an appropriate small integer $K > 0$ can be selected such that the distribution of $P(B_\tau = k | \tau > K)$ becomes almost unimodal. The proposed method is to obtain the mean and confidence interval of the final cascade size from this unimodal distribution. This method should yield better results for larger cascades because the finite-size effect (i.e., bipolarization) becomes more pronounced as the cascade size grows.

## Conclusions

To theoretically verify the existence of bipolarization, this paper derived various mathematical equations from an urn model, a model mimicking the fundamental mechanism of Twitter-type information diffusion, and has revealed the followings through numerical computation:

- The infinite network size assumption simplified the final cascade size distribution. The distribution was a strictly decreasing function of the cascade size. The product of the retweet rate ($\lambda$) and number of followers ($f$) determined the shape of the distribution. A power law (over a limited range) appeared if $f\lambda = 1$ ($f\lambda \neq 1$).
- Calculation of the distributions assuming the network size is finite required a very large amount of computation power. The upper and lower bounds of the distribution showed that the distribution was a decreasing function of the cascade size if $f\lambda$ is small. The bounds also suggested that another peak could emerge in the distribution as $f\lambda$ grows.
- The approach of using random numbers to approximate the shape of the distribution revealed the existence of another peak. The two peaks of the distribution moved

apart as $f\lambda$ increased. This result was consistent with that reported in a previous simulation work.

**Acknowledgements**
The author acknowledges the support of the Fukuoka Institute of Technology for proofreading and publication costs.

**Author contributions**
The author conducted this study alone. The author have read and approved the final version of the manuscript.

## Declarations

**Competing interests**
The author declares that the author has no competing interests.

Received: 28 February 2023 Accepted: 28 May 2023
Published online: 07 June 2023

**References**

Bakshy E, Hofman JM, Mason WA, Watts DJ (2011)Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on web search and data mining, pp 65–74
Baldesi L, Butts CT, Markopoulou A (2018) Spectral graph forge: graph generation targeting modularity. In: IEEE INFOCOM 2018-IEEE conference on computer communications, pp 1727–1735. IEEE
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
Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512
Batagelj V, Brandes U (2005) Efficient generation of large random networks. Phys Rev E 71(3):036113
Bourigault S, Lamprier S, Gallinari P (2016) Representation learning for information diffusion through social networks: an embedded cascade model. In: Proceedings of the ninth ACM international conference on web search and data mining, pp 573–582
Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J (2014) an cascades be predicted? In: Proceedings of the 23rd international conference on world wide web, pp 925–936. ACM
Cheung M, She J, Junus A, Cao L (2017) Prediction of virality timing using cascades in social media. ACM Trans Multimedia Comput Commun Appl (TOMM) 13(1):2
Crimaldi I, Louis P-Y, Minelli IG (2022) An urn model with random multiple drawing and random addition. Stochast Process Appl 147:270–299
Di Bona G, Ubaldi E, Iacopini I, Monechi B, Latora V, Loreto V (2022) Social interactions affect discovery processes. arXiv preprint arXiv:2202.05099
Dosi G, Moneta A, Stepanova E (2019) Dynamic increasing returns and innovation diffusion: bringing polya urn processes to the empirical data. Ind Innov 26(4):461–478
Forestier M, Bergier J-Y, Bouanan Y, Ribault J, Zacharewicz G, Vallespir B, Faucher C (2015) Generating multidimensional social network to simulate the propagation of information. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 1324–1331. IEEE
Gleeson JP, Onaga T, Fennell P, Cotter J, Burke R, O'Sullivan DJ (2020) Branching process descriptions of information cascades on twitter. J Complex Netw 8(6):002
Goel S, Watts DJ, Goldstein DG (2012) The structure of online diffusion networks. In: Proceedings of the 13th ACM conference on electronic commerce, pp 623–638 . ACM
Hino M, Irie Y, Hisakado M, Takahashi T, Mori S (2016) Detection of phase transition in generalized polya urn in information cascade experiment. J Phys Soc Jpn 85(3):034002
Horawalavithana S, Skvoretz J, Iamnitchi A (2020) Cascade-lstm: Predicting information cascades using deep neural networks. arXiv preprint arXiv:2004.12373
Junus A, Ming C, She J, Jie Z (2015) Community-aware prediction of virality timing using big data of social cascades. In: 2015 IEEE first international conference on big data computing service and applications (BigDataService), pp 487–492 . IEEE
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, pp 249–258 . ACM
Leskovec J, McGlohon M, Faloutsos C, Glance N, Hurst M (2007) Patterns of cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM international conference on data mining, pp 551–556. SIAM
Li C-T, Lin Y-J, Yeh M-Y (2015) The roles of network communities in social information diffusion. In: 2015 IEEE international conference on big data (big data), pp 391–400. IEEE
Mahmoud H (2008) Pólya Urn models. Chapman and Hall/CRC, New York
Oida K (2021) Bi-polarization in cascade size distributions. IEEE Access 9:72867–72880
Pemantle R (2007) A survey of random processes with reinforcement
Rafik A, Nabil L, Olfa S (2019) A generalized urn with multiple drawing and random addition. Ann Inst Stat Math 71(2):389–408
Tria F, Loreto V, Servedio VDP, Strogatz SH (2014) The dynamics of correlated novelties. Sci Rep 4(1):1–8

Wang Z, Chen C, Li W (2018) A sequential neural information diffusion model with structure attention. In: Proceedings of the 27th ACM international conference on information and knowledge management, pp 1795–1798
Wang Y, Shen H, Liu S, Gao J, Cheng X (2017) Cascade dynamics modeling with attention-based recurrent neural network. In: IJCAI, pp 2985–2991
Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440
Wegrzycki K, Sankowski P, Pacuk A, Wygocki P (2017) hy do cascade sizes follow a power-law? In: Proceedings of the 26th international conference on world wide web, pp 569–576
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
Xie J, Zhang C, Wu M (2011) Modeling microblogging communication based on human dynamics. In: 2011 eighth international conference on fuzzy systems and knowledge discovery (FSKD), vol. 4, pp 2290–2294. https://doi.org/10.1109/FSKD.2011.6020045
Yu L, Cui P, Wang F, Song C, Yang S (2015) From micro to macro: uncovering and predicting information cascading process with behavioral dynamics. In: 2015 IEEE international conference on data mining, pp 559–568. IEEE
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, pp 1513–1522. ACM
Zhou C, Zhao Q, Lu W (2015) Impact of repeated exposures on information spreading in social networks. PLoS ONE 10(10):0140556
Zhou C, Zhao Q, Lu W (2017) Cumulative dynamics of independent information spreading behaviour: a physical perspective. Sci Rep 7(1):1–14
Zhou F, Xu X, Trajcevski G, Zhang K (2021) A survey of information cascade analysis: models, predictions, and recent advances. ACM Comput Surv (CSUR) 54(2):1–36

## Publisher's Note