Saturated total-population dependent branching process and viral markets

Khushboo Agarwal* and Veeraruna Kavitha
{agarwal.khushboo, vkavitha}@iith.ac.in, IEOR, IIT Bombay, India

Abstract—Interesting posts are continually forwarded by the users of the online social network (OSN). Such propagation leads to re-forwarding of the post to some of the previous recipients, which increases as the post reaches a large number of users. Consequently, the effective forwards (after deleting the re-forwards) reduce, eventually leading to the saturation of the total number of copies. We model this process as a new variant of the branching process, the ‘saturated total-population-dependent branching process’, and analyse it using the stochastic approximation technique. Notably, we obtain deterministic trajectories which approximate the total and unread copies of the post ‘asymptotically and almost surely’ over any finite time window; this trajectory depends only on four parameters related to the network characteristics. Further, we provide expressions for the peak unread copies, maximum outreach and the life span of the post. We observe known exponential growth but with time-varying rates. We also validate our theory through detailed simulations on the SNAP Twitter dataset.

I. INTRODUCTION

Social media offers a global platform for communication; people share their content, forward interesting posts among the ones shared with them, etc. It plays a significant role in marketing, e.g., a post advertising a specific product/service can get viral (reach a large number of users). It is also used for spreading propaganda, sharing a piece of knowledge, influencing the beliefs of a large population, etc.

The content propagation (CP) over an online social network (OSN) can be outlined as follows: a) a post is shared with an initial set of users, called seed users; b) the seed users forward the post to their friends/followers if they like it; c) the recipients follow suit and this continues. The content either gets extinct in the initial phase or gets viral, and eventually, copies get saturated, and the propagation ceases.

There are several approaches for studying CP, and we use branching processes (BPs) to analyze the same. Further, since a post can witness a huge surge in the shares in a short duration, we consider continuous-time BPs (e.g., [1]–[3]). We discuss and compare other techniques at the end.

Branching processes: A variety of BPs have been studied before (e.g., [4]–[6] are few strands of them). We discuss a few relevant varieties and provide some details to model the CP process using BPs. The course of any BP largely depends upon the expected number of offsprings (Γ). We have super-critical BPs when the expected offsprings £[Γ] > 1 (see e.g., [4]); such BPs have positive probability of exploding (population grows exponentially) and can mimic the viral CP. In the critical/sub-critical regime (£[Γ] ≤ 1), the BP gets extinct with probability one, i.e., the population eventually declines to zero. The current population in a BP represents the live population, while the total population also includes the dead ones. In the context of CP, the number of unread copies (the corresponding recipients are yet to view/read the post) represents the current population. While the total number of recipients, including the ones that already read the post, represents the total population.

Saturation: When a post is viral and is already forwarded to a noticeable fraction of the network, a significant fraction of further forwards (by future users) can overlap a part of the network that already received the post. The expected number of effective forwards (after deleting re-forwards) represents the expected offsprings when one attempts to model CP using a BP. The number of re-forwards depends on total copies (number of users that already received the post) and not just on the current population/unread copies. We are only aware of current-population dependent BPs (e.g., [6]).

Thus the existing BP models are insufficient to mimic a saturated CP due to two imperative factors: a) the overlaps in forwards/offsprings depend upon the total copies/population, and b) the usual content propagation process traverses from super-critical to sub-critical regime before getting extinct. We thus consider a new variant of continuous-time total-population dependent Markovian BP, named as saturated total-population dependent BP (STP-BP). The total copies either increase with time or saturate; if re-forwards are proportional to the total copies, then the resultant value of expected offsprings decreases with time. This also ensures the desired transition from super-critical to sub-critical regime. The saturated BP well mimics the CP.

Recently, we analyzed many new variants of BPs, using a new approach based on stochastic approximation techniques, e.g., attack and acquisition BP (competing viral markets [5]), and proportion-dependent BP (fake news on OSNs [7]). Further modifying the said approach to address the required finite horizon analysis allows us to analyze saturation resulting from total population dependency.

Key contributions: At first, we formally analyze the STP-BP. We derive an appropriate ordinary differential equation (ODE) and its solution, which (time asymptotically) almost surely approximates the embedded chain of the CP process over any finite time window. These deterministic solutions depict exponential growth and linear fall for unread copies. Secondly, we model and fit an appropriate total-copies dependent piece-wise linearly decreasing function for the expected offsprings, having two different slopes. We further derive important metrics like the peak number of unread
copies, time asymptotic value of the total copies, and others. The growth of the total copies is exponential and depends on the reduction rates in the expected forwards.

We corroborate our theoretical results by performing Monte-Carlo simulation on SNAP Twitter-dataset [8]. The description of the theoretical trajectories depends only on four parameters of the OSN (e.g., two rates of reduction of the expected forwards) and the attractiveness of the post.

Related work: The study of CP on OSNs has been a topic of interest for a long time, and several approaches are used for its analysis. Random graph models are widely used to analyze CP on OSNs (e.g., see [9]–[11]); in particular, [11] considers re-exposure of users with a post on OSN. However, such models can not capture aspects like virality, as discussed in [1]–[3], while BPs facilitate virality analysis like growth patterns, extinction/virality probability etc.

The epidemiology-based models (in particular SIR) are also used to study CP, which succeed in capturing saturation (e.g., [12]–[14]). The current and total copies are respectively modelled as infected and (infected+recovered) populations. We argue that this approach is not suitable for analyzing viral markets on OSNs/email platforms: (a) this set of literature directly starts with an appropriate ODE (e.g., [13] considers exponentially diminishing infection rate/interest in the post, while [12] considers standard SIR ODE), (b) they do not delve into the details of the random dynamics (e.g., the chance encounters between various individuals), and more importantly, (c) on OSNs, majority of users share the post to a subset of their friends ‘only once’ after viewing and lose interest immediately after; in contrast, in SIR based models, infected individuals keep infecting/spreading the information for a random/prefixed time before recovering/losing-interest. In other words, SIR-based models well capture the typical behaviour of word-of-mouth dynamics (individuals remain interested in gossip, keep sharing it, and then lose interest) and are insufficient for viral markets over OSNs.

There is a brief indirect mention of the saturation effect in [3], where using BPs, the authors predict the future progress of CP using the available history of an ongoing campaign; they use the well-known Kolmogorov’s backward equations (for PGFs in BPs) to achieve this. In contrast, we provide a theoretical study of a new relevant variant of BP, facilitating an exhaustive study of the saturated CP on OSNs.

II. PROBLEM DESCRIPTION AND BACKGROUND

Consider an OSN where the content of interest is forwarded by its users. At the start of the propagation, the post is shared by its content provider to an initial set of users, called seed users. These users view the post on the OSN at random time instances. Then, they forward it to some or all of their friends, depending on how much they like the post. This subset of users further forward the post when they visit the OSN, and the post propagation continues likewise (more details in [1], [3], [5]). The random instances at which users visit the OSN are called wake-up times. The time gap between the wake-up and the post-reception times of any recipient is exponentially distributed with parameter \( \lambda \). Let \( \mathcal{F} \) be the (random) number of friends of a typical user, with finite mean. The attractiveness of the post is captured by factor \( \rho \), which specifies a subset of friends (\( \Gamma \)) to whom the post is forwarded. Naturally, \( \rho \) determines the growth of the post.

We will see that the content propagation (CP) majorly depends on the expected number of such forwards, which in turn depends on the then total recipients of the post (say \( a \)). That is, \( E[\Gamma] := m(a) \), where \( m \) is some function which will be discussed in detail in the coming sections. The study of such a dependency (using SNAP dataset) and its influence on the CP is the key differentiating feature of our work. As one may anticipate again, such dependency will eventually lead to the reduction in the effective forwards of the post, in turn leading to the saturation of shares. This effect has been observed majorly through numerical studies in the past, e.g., see [1, Fig. 6] and [15, Fig. 1] for total and current shares, respectively. We aim to provide an analytical explanation of saturation in terms of an approximating deterministic trajectory and some relevant performance measures. We next describe the new variant of ‘saturated’ BP, STP-BP.

A. Saturated Content Propagation and Branching process

Let \( A(t) \) denote the total number of copies of the post on the network, i.e., all the copies which have been received (viewed or not, forwarded or not) by the users on the OSN till time \( t \). We briefly refer this number as total shares. Further, let \( C(t) \) be the unread/live/current number of copies of the post till time \( t \). These represent the copies that have been received but not yet viewed by the users till time \( t \), and are solely responsible for further propagation of the post; we refer to this number as current shares.

Let \( A(0) = C(0) = a_0 \) be the number of seed users. By previous discussion, it is clear that the number of forwards have a direct dependence on the total shares, i.e., \( \Gamma = \Gamma(A(t)) \). Observe that \( \Gamma \) does not depend on the current shares, as is usually considered in the BP literature.

Let \( \tau^+ \) and \( \tau^- \) denote the time instances immediately after and before wake-up time \( \tau \) (of any user), e.g., \( A(\tau^+) := \lim_{t \uparrow \tau} A(t) \). Then, the evolution of the system at transition epoch \( \tau \) is:

\[
C(\tau^+) = C(\tau^-) + \Gamma(A(\tau^-)) - 1,
A(\tau^+) = A(\tau^-) + \Gamma(A(\tau^-)).
\]

Above dynamics can easily be placed in a BP framework, when the current shares are modelled as a population. Accordingly, the number of forwards \( \Gamma(A(\tau^-)) \) can be viewed as the offsprings of the population. Further, when the post is shared by a user, the current shares reduce by 1 (see (1)); this is exactly like a death in a BP (see [1], [5], [7] for similar details).

To analyse total population dependent BPs, one needs to study two-dimensional tuple \( \Phi(t) := (C(t), A(t)) \) simultaneously, a realisation of which is denoted by \( \phi = (c,a) \). In contrast, the existing BP models can analyse \( C(t) \) alone if the need is only to analyse the current population. We now
III. DYNAMICS AND ODE APPROXIMATION

To facilitate the study of STP-BP, we analyse the embedded chain corresponding to the underlying continuous-time jump process (CTJP), as in [5], [7]. It is a standard technique to use embedded chains, when transience, recurrence, extinction and similar properties of CTJP are studied. We use it for similar purpose in this paper. In particular, we observe the dynamics in (1) at the time instances when a user with unread copy of the post wakes-up. Let \( \tau_n \) be \( n^{th} \) such transition epoch.\(^3\) Let \( C_n := C(\tau_n^+) \) be the current shares of the post immediately after \( \tau_n \). Similarly define \( A_n \). Note that the time taken by the first user to wake-up after \( n^{th} \) transition epoch, \((\tau_{n+1} - \tau_n)\), is exponentially distributed with parameter \( \lambda C_n \).

Thus, if a user wakes up at \( \tau_n \), then:

\[
C_n = C_{n-1} + \Gamma_n(A_n-1) - 1, \quad A_n = A_{n-1} + \Gamma_n(A_{n-1}). \tag{2}
\]

As in [5], [7], we use stochastic approximation (SA) approach to study the embedded chain. Towards this, define the following fractions of current and total shares respectively, \( \Psi^n_c := C^n / \Psi^n \) and \( \Psi^n_a := A^n / \Psi^n \) for \( n \geq 1 \), with \( \Psi^0 = \Psi^0_a = a_0 \). Also let \( \Upsilon_n = (\Psi^n_c, \Psi^n_a) \). Further, define:

\[
\epsilon_n = \frac{1}{n+1}, \quad t_n := \sum_{k=1}^{n} \epsilon_k, \quad \text{and} \quad \eta(t) := \max \{ n : t_n \leq t \}. \tag{3}
\]

Then, the evolution of \( \Upsilon_n \) can be captured by 2-dimensional SA based updates given below (see (2)).

\[
\begin{align*}
\Psi^n_c &= \Psi^n_c - \epsilon_{n-1} [\Gamma_n(A_n-1) - 1 - \Psi^n_c - 1] \Psi^n_{c-1} > 0, \\
\Psi^n_a &= \Psi^n_a - \epsilon_{n-1} \Gamma_n [\Psi^n_a - \Psi^n_a - 1] \Psi^n_{a-1} > 0. \tag{4}
\end{align*}
\]

We analyse these fractions using SA techniques (e.g., [16]), which helps in approximating the same using the solutions of the ODE (see (3)).

\[
\psi^c = (m(a) - 1 - \psi^c) I, \quad \psi^a = (m(a) - \psi^a) I, \quad I := 1_{\psi^a > 0}, \quad a(t) := \psi^a(t) \eta(t), \quad \text{and} \quad m(a) := E[\Gamma(a)].
\tag{5}
\]

By [17, Lemma 1], the solution for this non-autonomous and non-smooth ODE exists over any finite time interval in the extended sense (satisfies the ODE for almost all \( t \)). In Theorem 1 given below, we will prove that the above ODE indeed approximates (2).

**Approximation result:** The study of continuous time population size dependent BPs has been limited in the literature. In this paper, we use the ODE approximation result to study the saturated BP. Now, for mathematical tractability, we require the following assumption on offspring distributions:

(A) There exists an integrable random variable, \( \Gamma \), such that \( \Gamma(a) \leq \Gamma \) almost surely for every \( a \) and \( E[\Gamma]^2 < \infty \).

(B) The mean function, \( m(\cdot) \) is Lipschitz continuous.

The assumption (A) is readily satisfied (details in the next section), while (B) is an extra assumption required for additional affirmation (see Theorem 1(ii)). We will now see that the piece-wise constant interpolation, \( \Omega^n(\cdot) := (\Psi^n_c(\cdot), \Psi^n_a(\cdot)) \) of \( \Upsilon_n \) trajectory defined as:

\[
\Omega^n(t) = \Upsilon_n \text{ if } t \in [t_n, t_{n+1}), \tag{6}
\]

satisfies an almost integral representation as below, with \( I_n := 1_{\psi^n_c(\cdot) > 0} \) (see [17, eqn. (25)] for derivation):

\[
\begin{align*}
\psi^n_c(t) &= \psi^n_c + \int_0^t \left( m(\psi^n_a(s)) - \psi^n_a(s) \right) I_n ds + \varepsilon^n_a(t), \\
\psi^n_a(t) &= \psi^n_a + \int_0^t \left( m(\psi^n_a(s)) - 1 - \psi^n_a(s) \right) I_n ds + \varepsilon^n_a(t).
\end{align*}
\tag{7}
\]

Let \( \Omega^n(\cdot) \) be the solution of ODE (5), with \( \Omega^n(0) = \Upsilon_n \). Observe that \( \Omega^n(\cdot) \) in (7) is similar to \( \Omega^n(\cdot) \), except for the difference term \( \varepsilon^n_a(t) := \varepsilon^n_a(t, \varepsilon^n_a(\cdot)) \). Further, if at all \( \varepsilon^n_a(t) = 0 \) for all \( t \leq T \), then, by uniqueness of the solution (see [17, Lemma 1]), the BP trajectory (4) would have coincided with it, i.e., \( \Upsilon_k = \Omega(t_k) \) for all \( k \) such that \( t_k \leq T \). However, it is not true in general; nevertheless, we will show that \( \|\varepsilon^n\| \rightarrow 0 \) as \( n \rightarrow \infty \) (see norm \( \| \cdot \| \) in (8)).

Thus, we have two operators which are converging towards each other; the first operator including \( \varepsilon^n \) in (7) provides the BP trajectory, while the second operator without \( \varepsilon^n \) in (7) provides the ODE solution. Further, using Maximum theorem, we show that the difference between the two solutions of the operators (7) with and without \( \varepsilon^n \) is small, when \( \|\varepsilon^n\| \) is small. Formally, we state the result as follows:

**Theorem 1.** For any \( \Upsilon(\cdot) = (\psi^c(\cdot), \psi^a(\cdot)) \), define the norm with any finite \( T > 0 \):

\[
\|\Upsilon\| := \max\{\|\psi^c\|, \|\psi^a\|\}, \quad \text{where}
\]

\[
\|\psi^i\| := \sup\{t \in [0, T] : |\psi^i(t)|\} \text{ for any } i \in \{a,c\}. \tag{8}
\]

Under assumption (A), we have the following almost surely:

(i) if \( \|\varepsilon^n\| \rightarrow 0 \) as \( n \rightarrow \infty \), and

(ii) if (B) also holds, then, the difference

\[
\sup_{k : n \geq 0, t_k \leq T} \|\Upsilon^n(t_k) - \Omega^n(t_k)\| \text{ depends upon the magnitude of } \|\varepsilon^n\|.
\]

**Proof** is provided in [17]. \( \square \)

**Remarks:**

(i) For each \( n \), consider the ODE initialised with value of the embedded chain, \( \Upsilon_n \). Then, the embedded chain values at transition epochs, \( k \in \{n, n \eta(t_n + T)\} \), are close to the ODE solution, \( \Omega^n(t_k - t_n) \), at time points \( t_k \in [t_n, t_{n+T}] \). This approximation improves as \( n \) increases, and the result is true almost surely (a.s.) and for all \( T < \infty \).

(ii) **Dichotomy:** We have a ‘modified dichotomy’: either the population gets extinct (\( \Gamma = 0 \) in initial epochs), or the population explodes exponentially as confirmed by ODE-solution (12) in section V. In contrast to classical dichotomy (e.g., [4]), in both the sets of sample paths, the saturated BP eventually gets extinct.

(iii) As the ODE solution approximates the embedded chain (BP) trajectory, one can analyse the latter using the former. In contrast to many existing studies, we would consider the analysis of the ODE trajectories and not the attractors, which is more relevant here. Before that we derive an appropriate function that can represent \( m(\cdot) \), the total-share-dependent expected forwards (TeF) in a typical OSN.

---

\(^3\)If the post gets extinct at \( n^{th} \) epoch, we set \( \tau_k := \tau_n \) for all \( k \geq n \) and the same is true for rest of the quantities.
IV. POPULATION DEPENDENT EXPECTED FORWARDS

It is clear from the ODE (5) and Theorem 1 that the expected number of forwards, \( m(\cdot) \), influence the course of any post; we construct a piece-wise linear function to capture the same. We estimate the parameters and validate the model using the SNAP dataset [8] in section VI. Following aspects are considered for the model:

- Any user on the OSN forwards the post to a subset of friends based on its attractiveness. Among this subset, a fraction of users will have previously received the post. Most likely, such users will not be interested in the same post again. Say \( \kappa \) is the fraction of common friends between any two typical users, then \( \kappa a \) denotes the number of such re-forwards if ‘a’ number of users already had the post. Thus a linearly decreasing function is a suitable choice for \( m(\cdot) \).

- It is likely that a user with a high number of friends is also a friend to a large number of users; the dataset supports this observation. Thus, a user with more connections is more likely to receive the post earlier. To capture such aspects, we model the friends of a user, \( F(a) \), to be independent across users, but with decreasing expected values, i.e., \( a \mapsto E[F(a)] \) is itself a decreasing function. With this, the assumption (A) is readily satisfied by considering for example, \( \Gamma = F(a_0) \), with \( E[F(a_0)]^2 < \infty \).

- Further, as time progresses, the post would naturally proceed towards saturation. Therefore, it is likely that the slope of \( m(\cdot) \) is drastically different towards the end. To summarize, we model the TeF function \( m(\cdot) \) as a piece-wise linearly decreasing function, with two different slopes, where the initial slope is bigger than that towards the end.

- It is reasonable to assume that the expected number of forwards is proportional to the attractiveness factor \( \rho > 0 \). Hence, we model the TeF function \( m(\cdot) \) as \( m(a) = \rho m_N(a) \forall a \), where \( m_N(\cdot) \) is the TeF with \( \rho = 1 \). Such a factorization of \( m(\cdot) \) is supported by the experiments on the SNAP dataset for a wide range\(^4\) of \( \rho \). The latter function \( m_N(\cdot) \) is determined solely by the characteristics of the network. Basically, \( m_N(\cdot) \) corresponds to a hypothetical situation where every recipient shares with all of its friends, but new shares are only the effective forwards. Conclusively:

\[
m(a) = \rho m_N(a), \quad m_N(a) = (\tilde{m} - \kappa_1a)_{a\leq \tilde{a}} + (\tilde{m} - \kappa_2a)_{a>\tilde{a}}. \tag{9}
\]

Thus, the TeF is determined by four parameters, \( \tilde{m}, \kappa_1, \kappa_2 \) and \( \tilde{a} \). Here, \( \rho \tilde{m} = \rho E[F(a_0)] \) is the expected number of forwards in the beginning, and \( \rho \kappa_1, \rho \kappa_2 \) reflect the slopes of the TeF. Observe that the slope reduces to \( \rho \kappa_2 < \rho \kappa_1 \) when total shares reach a particular value, \( \tilde{a} \). We assume\(^5\) \( \rho \tilde{m} > 1 \).

- We corroborate that the above model well captures the content propagation using an instance of the SNAP dataset: (a) we first obtain the TeF curve by estimating \( m_N(a) \) (i.e., with \( \rho = 1 \)) for some values of \( a \); (b) then fit two slope curve using a naive approach; and (c) use the estimated (four) parameters to derive the solution of the ODE (5) and other theoretical conclusions, for posts with different \( \rho \) values.

Estimating the parameters mentioned above in step (b) is crucial in any such study. For now, we estimate the same using a simple trial-and-error method, as explained in footnote 4 and section VI. However, more sophisticated study is essential to estimate these parameters more accurately, and we leave it for the future. As said earlier, this study aims to provide a theoretical understanding of the CP process.

V. ODE ANALYSIS

Using the TeF (9), we solve the ODE (5). By Theorem 1, this solution (a.s.) approximates the CP process. We also derive the approximate trajectories for \( a(t) \) and \( c(t) \), which show the desirable saturation effect. Let \( \tau_e := \{ t: a(t) = \bar{a} \} \) and \( \tau_c := \{ t: c(t) = 0 \} \), the extinction time.

In extinction sample paths (we have, \( \tilde{Y}_n \equiv 0 \), from ODE (5), \( \tilde{\tau}_n = 0 \forall t \)). We now consider viral sample paths.

Total and current fractions: The fractions \( (\psi^a, \psi^c) \) are given by the solution of the ODE (5), which can be directly obtained using the integrating factor (IF) approach. For \( t \leq \tau_e \), the closed-form (extended) solution for \( \psi^a \) is given by:

\[
\psi^a(t) = e^{-\int_{t_0}^t u_1(\gamma(s) + 1)ds}(u_2 + u_3 \int_{u_1}^t e^{\int_{s}^{t} u_1(\gamma(s) + 1)ds}ds),
\]

where constants \( u := (u_1, u_2, u_3, u_4) \) in the two phases equal:

\[
u = \begin{cases} (0, c_0, \bar{m}_\rho, \kappa_1 \rho), & 0 \leq t \leq \tau_s, \\ (\tau_s, \psi^c(\tau_s), \bar{m}_\rho, \kappa_2 \rho), & \tau_s < t \leq \tau_e. \end{cases}
\]

After extinction, i.e, for \( t > \tau_e \), from the ODE (5), \( \psi^a(t) = \psi^c(\tau_e) \). Similarly, the current fraction of shares till time \( t \) evaluates to the following, again using IF approach:

\[
\psi^c(t) = e^{(-t - u_2)}(v_2 + \int_{u_1}^t e^{(s - u_4 - u_2)}(v_3 - v_3a(s) - 1)ds),
\]

with \( v = (v_1, v_2, v_3, v_4) = \begin{cases} (0, c_0, \bar{m}_\rho, \kappa_1 \rho), & 0 \leq t \leq \tau_s, \\ (\tau_s, \psi^c(\tau_s), \bar{m}_\rho, \kappa_2 \rho), & \tau_s < t \leq \tau_e. \end{cases} \)

Further, \( \psi^c(t) \equiv 0 \) for all \( t \geq \tau_e \).

Trajectories of shares: It is more relevant to analyse the trajectories corresponding to the current and total shares; and we provide approximate expressions for the same. Recall from (5), \( a(t) = \psi^a(t)\eta(t) \) and \( c(t) = \psi^c(t)\eta(t) \). Thus, we begin with an approximation for \( \eta(t) \), defined in (3):

\[
\eta(t) \approx \max\{n: \gamma + \ln(n) \leq t\} \approx e^{t - \gamma},
\]

where \( \gamma \) is Euler-Mascheroni constant. Henceforth, we use this approximation of \( \eta(t) \) in all the computations. With this approximation and using (5), the ODE for \( a(\cdot) \) is given by:

\[
\dot{a} = \psi^a e^{t - \gamma} + \psi^c e^{t - \gamma}1_{c>0} = m(a)e^{t - \gamma}1_{c>0}. \tag{11}
\]

The solution of the above ODE with \( m(\cdot) \) as in (9), can be obtained using the standard techniques in ODE theory:

\[
a(t) = \begin{cases} w_1 - w_2 e^{-w_3 t}, & 0 \leq t \leq \tau_e, \\ a(\tau_e), & t > \tau_e. \end{cases}
\]

\(^4\)Such a common fit is good mostly for \( \rho \geq 0.4 \), for others we directly derive a good linear fit of \( m(\cdot) \) by trial-and-error, see section VI.

\(^5\)Then, there is a possibility of exponential growth leading to virality.
where the constants are given by the following (with $w = (w_1, w_2, w_3)$, observe by continuity $a(\tau_s) = \bar{a}$):

\[
w = \begin{cases}
  \frac{m}{\tau_1}, & (w_1 - \bar{a})e^{w_1} - \gamma \\
  \frac{m}{\tau_2}, & (w_2 - \bar{a})e^{w_2} - \gamma
\end{cases}, \quad 0 \leq t \leq \tau_s,
\]

\[
w = \begin{cases}
  \frac{m}{\bar{\tau}_1}, & (w_1 - \bar{a})e^{w_1} - \gamma \\
  \frac{m}{\bar{\tau}_2}, & (w_2 - \bar{a})e^{w_2} - \gamma
\end{cases}, \quad \tau_s < t \leq \tau_e.
\]  

(13)

Observe that $a(\cdot)$ saturates as the current shares get extinct. We believe, only STP-BP can capture these effects. Proceeding as in (11), the ODE for $c(t)$ is given by:

\[
\dot{c} = (m(a) - 1)e^{-\gamma}1_{c>0}.
\]  

(14)

By solving, the trajectory of the current shares is given by:

\[
a(t) = \frac{w_1 - w_2e^{-nw_3c}}{w_3c^\gamma}, \quad 0 \leq n \leq n_e,
\]

\[
a(t) = \frac{w_1 - w_2e^{-nw_3c}}{w_3c^\gamma} - n, \quad 0 \leq n \leq n_e.
\]  

(16)

(17)

In the above, (17) follows because $a(t_n) - c(t_n) = n$ from (2). After extinction, $(a(t_n), c(t_n)) = (a(\tau_e), 0)$ $\forall n \geq n_e$.

We now investigate other important metrics related to CP. **Growth rates:** In standard BPs, the population exhibits dichotomy, i.e., either the population grows exponentially large (at a constant rate), or declines to zero. In the former case, both $C(t_n)$ and $A(t_n)$ grow like $e^{(\rho(n-1))t_n}$, where $\bar{\tau}_n := \lambda t_n$ is independent$^6$ of $\lambda$, as time progresses. While in STP-BP, from (16), which again approximates $A(t_n)$ a.s., and time asymptotically, it is clear that the total shares have exponential growth, however, the rates are different in the two phases. The growth rate in the initial phase is $w_3c^\gamma = \kappa_1 \rho s$, while it decreases to $\kappa_2 \rho s$ in the later phase (since $\kappa_1 > \kappa_2$). Further, the current shares also experience an initial exponential growth, which is further modulated by the growing factor of $n$, in (17), leading to an eventual linear fall. This illustrates the modified dichotomy discussed after Theorem 1. More attractive posts have higher growth rate.

At this point, we would like to admit that discussing the growth rates at transition epochs is a non-standard practice in the BP literature. However, for the saturated BP, that does not grow forever, we believe such a discussion is relevant and important. In future, we plan to include the influence of $\{\tau_n\}$ on these growth rates, by extending the ODE analysis to fractions $\{\tau_n/\bar{\tau}_n\}$; this would help us derive the standard growth patterns discussed in the BP literature for STP-BP.

---

**Peak of current shares:** Define $c^* = \sup_t c(t)$, i.e., the peak (maximum) current shares. It can be obtained from (15) $(c''(t) < 0)$ and equals (recall $n_s = a(\varphi) - c(\varphi)$):

\[
c^* = w_1 - \frac{1 + \ln(w_2w_3c^\gamma)}{w_3c^\gamma}, \quad e^{-\gamma + \varphi} = n_s.
\]  

(18)

where $\varphi$ takes different values in two phases as in (15).

**Life span and Max reach:** Recall $n_c$ is the epoch at which CP of the post terminates, or in other words, it represents the life span of the post. Now, substituting $c(t_n)$ from (17), $n_c$ can be written as the solution of the equation:

\[
w_1 - w_2e^{-n_cw_3c^\gamma} = n_c.
\]  

(19)

where $w$ is given by the second line of (13). Further, by (17), the max reach or saturated total shares $a(t_{n_c})$ equal $n_c$.

**Probability of virality:** When one transitions from super-critical to sub-critical regime, it is clear that the process gets extinct a.s. (see modified dichotomy). One of the important questions related to CP is the chances of virality, where total/current shares grow significantly large. This necessitates the definition of a different probability related to STP-BP, the probability of virality (PoV). In view of the initial exponential growth of current shares (17) and the dichotomy remark after Theorem 1, we define PoV, $p_{\Delta} := P(C(t) > \Delta$ for some $t$) for some threshold $\Delta > 0$. We conjecture that for small $\Delta$, $p_{\Delta} \approx 1 - p_c$, where $p_c$ the probability of extinction of standard population independent BP solves the equation: $f(s) = s$, where $f(\cdot)$ is the PGF of $\Gamma(a_0)$ (see [4]). This is an important aspect for future study.

**VI. NUMERICAL EXPERIMENTS**

We now perform exhaustive Monte-Carlo (MC) simulations on the SNAP Twitter dataset [8] to validate our theory. The dataset consists of inter-connections among 81,306 users, with 29.77 average number of friends. In all the case studies, we consider only viral sample paths.

**MC simulations over the dataset:** Say the content provider initially shares the post with 2 random seed users chosen from the dataset, identified by their user IDs. These users are added to the total and current-shares lists. At any time, one random user from the current-shares list wakes-up and forwards the post to a random subset of its friends, each chosen independently with probability $\rho$. Out of this subset, we ignore the friends who had the post. Then, we delete this user from the current-shares list, and update the two lists with the new effective forwards. The propagation continues in this manner and terminates when the current-shares list is empty, thus completing one sample run of the CP process.

**Estimation of TeF:** To estimate the TeF discussed in section IV, we create bins of equal length (1000 in our case). Then, we simulate the CP process 861 times over the dataset for $\rho = 1$. In each run and bin $n$, we count: (i) the number of transitions that have occurred and (ii) the sum total of effective forwards, while the number of total shares belongs to that bin. These two entries are accumulated bin-wise over 861 viral sample paths. For each bin, we divide the effective forwards by the number of transitions, which represents the estimated TeF curve (see solid black curve in Fig. 2(i)).
We repeat the above routine for $\rho = 0.4, 0.6$ as well, and the corresponding estimates of $m(\cdot)$ are plotted in Fig. 2(i) after dividing by the respective $\rho$ values. The resultant picture gives the confidence that one can derive the individual $m(\cdot)$ curves by using $m_N(\cdot)$ as suggested by (9). We also plot an approximate piece-wise linear curve (see the dashed curve in Fig. 2(i)), with the parameters $\tilde{m} = 21.321042, \kappa_1 = 532 \times 10^{-6}, \kappa_2 = 83 \times 10^{-6}$ and $\tilde{a} = 35000$ obtained using trial and error method. Henceforth, we refer this $m_N(\cdot)$ curve as the common-fit (C-fit) curve. For some sets of simulations, again using the trial and error method, we individually choose best fit $m(\cdot)$ curve for the given case study.

Fig. 1. Validation of theoretical trajectories against two instances of CP on SNAP dataset, for $\rho = 0.2, 0.4, 0.6$ respectively

In Fig. 1, we obtain the theoretical curves for total and current shares using C-fit TeF (discussed above), which are then compared with two instances of the CP process on the SNAP dataset. The theoretical curves well approximate the dataset curves for $\rho = 0.4, 0.6$. In fact, we observed this is true for $\rho \geq 0.4$, in general. However, for lower values of $\rho$ (e.g., $\rho = 0.2$ in Fig. 1), the best fit curve better approximates. This suggests that one can obtain C-fit curves for higher and lower ranges of $\rho$ separately. At last in Fig. 2(ii), for different $\rho$ values, we show that the peak shares and max. reach from the C-fit curve approximate the respective values for an instance of CP on the dataset with a maximum error of 11.4645% and 2.4837% respectively (for $\rho \geq 0.4$).

Fig. 2. (i) Left: piece-wise linear TeF; (ii) Right: Peak shares (18) and Max reach (19) versus SNAP estimates

VII. CONCLUSIONS

In this paper, we studied the saturation effect experienced by the total shares/copies of the post due to continual re-forwarding on OSNs. We captured the dynamics via newly introduced, saturated total-population dependent branching process. The analysis uses the stochastic approximation technique, which provides an ODE dependent on the post's total expected (effective) forwards. TeF. We modelled TeF as a piece-wise linearly decreasing function with two slopes. The derived trajectories (dependent on four network-specific parameters) asymptotically and almost surely approximate the total and read copies over any finite time interval.

Unlike classical dichotomy (either explosion or extinction), the unread copies observe explosion followed by extinction or direct extinction under saturated BPs. Interestingly, maximum reach (number of users that received the post) and the life span (number of users that read the post) are equal. We showed that derived expressions provide a good fit to the simulated instances of propagation over the SNAP dataset. Here, we fitted the parameters of TeF using a naive method; however, advanced methods of estimation/learning can improve the approximation.

REFERENCES

[1] R. Dhounchak and et al., “A viral timeline branching process to study a social network,” in ITC 29. IEEE, 2017.
[2] J. L. Iribarren and et al., “Branching dynamics of viral information spreading,” Physical Review E, 2011.
[3] R. Van der Lans and et al., “A viral branching model for predicting the spread of electronic word of mouth,” Marketing science, 2010.
[4] K. B. Athreya and et al., “Branching processes,” Handbook of statistics, 2001.
[5] K. Agarwal and et al., “Co-virality of competing content over osns?” in 2021 IFIP Networking. IEEE, 2021.
[6] F. C. Klebaner, “On population-size-dependent branching processes,” Advances in Applied Probability, 1984.
[7] S. Kapaskar and et al., “Controlling fake news by collective tagging: A branching process analysis,” IEEE Control Systems Letters, 2020.
[8] J. J. McAuley and et al., “Learning to discover social circles in ego networks,” in NIPS. Citeseer, 2012.
[9] M. Deijen and et al., “The winner takes it all,” The Annals of Applied Probability, 2016.
[10] J. Zhou and et al., “Cost-efficient viral marketing in online social networks.” World Wide Web, 2019.
[11] X. Lu and et al., “Beyond the coverage of information spreading: Analytical and empirical evidence of re-exposure in large-scale online social networks,” arXiv preprint arXiv:1907.12389, 2019.
[12] H. S. Rodrigues and et al., “Can information be spread as a virus? viral marketing as epidemiological model,” Mathematical Methods in the Applied Sciences, vol. 39, no. 16, 2016.
[13] M. Freeman and et al., “Viral information propagation in the digg online social network,” Physica A: Statistical Mechanics and its Applications, vol. 415, pp. 87–94, 2014.
[14] P. Jiang and et al., “A quantitative model for the spread of online information,” Quality & Quantity, vol. 53, no. 4, pp. 1981–2001, 2019.
[15] J. Hemsley, “Studying the viral growth of a connective action network using information event signatures,” First monday, 2016.
[16] H. Kushner and et al., Stochastic approximation and recursive algorithms and applications. Springer Science, 2003.
[17] K. Agarwal and et al., “Saturated total-population dependent branching process and viral markets,” http://arxiv.org/abs/2203.16239, 2022.