Modeling the Relationship Between Social Network Activity, Inactivity, and Growth

Bruno Ribeiro[1]

[1]*School of Computer Science, University of Massachusetts Amherst, USA*

Online Social Networks (OSNs) are multi-billion dollar enterprises. However, little is known about the mechanisms that drive them to growth, stability, or death. This study sheds light on these mechanisms. We are particularly interested in OSNs where current subscribers can invite new users to join the network (e.g., Facebook, LinkedIn). Measuring the relationship between subscriber activity and network growth of a large OSN over five years, we formulate three hypotheses that together describe the observed OSN subscriber behavior. We then provide a model (and extensions) that simultaneously satisfies all three hypotheses. Our model predicts four distinct subscriber activity behaviors and provides deep insights into the dynamics of subscriber activity, inactivity, and network growth rates. Finally, we present activity data of nearly thirty OSN websites, measured over five years, and show that the observed activity is well described by one of the four time series predicted by our model.

I. INTRODUCTION

There are no publicly available statistics on the survival of online social network (OSN) startups. A quick glance at Techcrunch's deadpool [1], however, reveals an endless list of OSNs that either died quickly or remained active by catering to small population niches. A few, however, make it big like Facebook and LinkedIn [2]. The mechanisms by which OSNs grow, remain stable, or die are unknown to date. This gap in our understanding of OSN dynamics is likely due to the lack of rich enough datasets that can shed light on the matter.

In this work we make a positive step towards understanding the mechanisms behind OSN growth (in subscribers), death, and sustainability (of activity). Observing the relationship between subscriber activity and network growth of a large OSN over five years, our study shows that the growth rate increases linearly with the number of active subscribers in the network. Moreover, observing that the activity lifespan of a subscriber is also linearly related with the subscriber degree, and that the time series of its growth rate is bell-shaped, we develop a model (and extensions) consistent with the multiple hypotheses that explain our data. In our model, active subscribers incite inactive subscribers to become active and also invite a population os susceptible users to join the network, making the network growth rate a function the number of active subscribers. We also consider network growth to be dependent on other external stimuli such as marketing campaigns or media exposure.

Our model predicts that OSNs fall into four distinct categories of network evolution with respect to subscriber activity and network growth. Moreover, our model providing deep insights into the dynamics of OSN evolution, showing that OSNs eventually become stable or die of inactivity. Interestingly, we predict a sharp threshold between critical OSN inactivity and OSN stability, where the OSN maintains a constant level of subscriber activity. We then select thirty OSN websites and, through measurements of their number unique website visitors over a period of five years, we show that they qualitatively fall into one of the four classes predicted by our model. Finally, we use our model to explain that activity reminders sent to subscribers ("Here is some activity you may have missed") can help OSNs near the critical OSN inactivity limit slightly tip the scale towards their survival.

Ouline: The outline of this work is as follows. Sec. II presents the related work. Sec. III presents five years of subscriber activity data of a large OSN. Sec. IV presents our model and Sec. V predicts its behavior classifying OSNs according to its parameters in the model. Sec. VI shows that real OSNs behave as one of the four types of OSNs predicted by our model. Finally, Sec. VII presents our conclusions and future work.

II. RELATED WORK

The growth rate of businesses – both in revenue and number of employees – is known to depend on its current value, a principle know in economics as path dependence growth [3]. This growth rate has been conjectured to be dependent on the business' hierarchical structure [4]. Interestingly, our work shows that OSN growth rates are not directly related to the OSN size but rather dependent on the size of its active subscriber population. Further studies should reveal whether the distribution of OSN growth rates also follow the observed tent-shaped distribution of revenue and employees of "traditional" companies [4] or whether OSN businesses present a new paradigm. In economic theory the diffusion of innovations in a population, such as the adoption of a product like an OSN, is modeled as a compartmental Susceptible-Infected (SI) epidemic, resulting in the number of adoptions in the population being described by a logistic function, which seems to replicate well the available observed data [5, 6].

Human online activity is bursty in nature [7]. Part of this burstiness can be attributed to human mobility [8] and circadian patterns [9, 10]. Recent studies, however, have shown that part of this burstiness can be attributed to the nature that subscribers interact [11]. For instance, the activity between subscribers in Internet chat rooms creates an interaction pattern correlation between these subscribers that is long lasting, along with a persistence of their emotional state throughout their chat sessions [11]. Anecdotally, at least, the above conclusions seem to match our intuition. The activity of our friends in the OSN incites us to login and become active which, in turn, incites our friends to either become active or stay active.

The empirical growth of OSNs has been the subject of a number of studies [12–15]. In these studies various aspects of the evolution of the OSN structure have been considered, from local node characteristics such as degree to network-wide metrics such as diameter. Lacking, however, is an in-depth study of the relationship between network growth and network size or the dynamics of subscriber activity. It is known that network growth rate is its size are correlated [16].

The evolution of the size of the active subscriber base is also of interest in evaluating the market value of an OSN startup. Valuation of OSNs and OSN-dependent businesses (e.g., facebook.com and zynga.com, respectively) is a challenging task. The value of an Internet business such as an OSN is prone to high volatility (sometimes bubbles) [17] and, to date, there is no consensus on how to assess the market value of such companies. We believe that a better understanding of the stochastic process that drives OSN growth and its active subscriber base is key to accurately assessing market risk valuations of these businesses.

The question of what takes to have a large active subscriber base is of great interest to the industry and a variety of metrics exist to measure OSNs [18]. Many factors help determine when and why subscribers join and leave an OSN, and complex factors such as cultural and racial trends are among them [19]. In our work, however, we do not analyze or model complex societal interactions and trends. Rather, we opt for a first order analysis that we believe can be used as a foundation to build upon, providing valuable insights into the relationship between subscriber activity, inactivity, and network growth, as seen next.

III. SUBSCRIBER ACTIVITY AND NETWORK GROWTH

Online social network websites rarely allow access to their subscriber activity data. In this work we use two complementary sources of data. The dataset gives the activity of 1.2 million random myspace.com subscribers that joined MySpace from 2006 to 2008, which was publicly available in January 2009 when our data was collected [20].

A Short History of MySpace. MySpace was founded in 2003 and from 2005 until early 2008, MySpace was the most visited social networking site in the world. In June 2006 MySpace surpassed Google as the most visited website in the United States. By late 2007 MySpace was first reported having a significant loss of its teenager subscriber base to Facebook [21] and in April 2008 Facebook usage overtook MySpace [2].

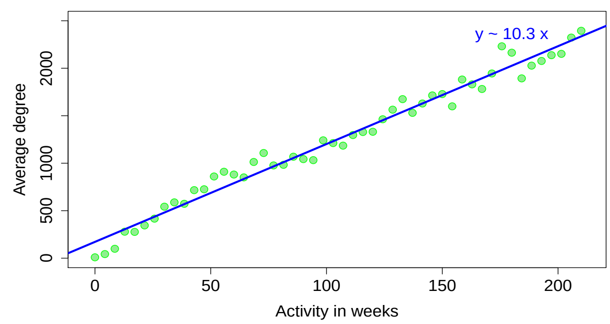


FIG. 1. MySpace.com average number of friends v.s. number of weeks of activity. Green points show the empirical average. The blue line is the regression showing $y \propto 10.3x$.

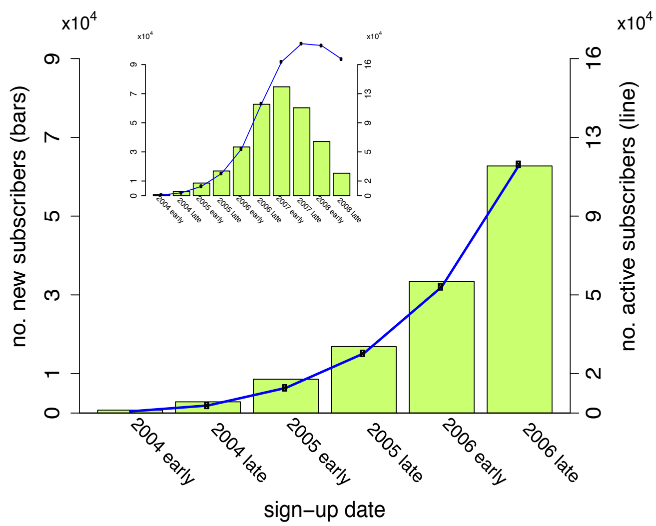


FIG. 2. MySpace.com growth × activity. Linear relationship between number of observed new subscribers (bars) and active subscribers (line) per semester from 2004 to 2006, before MySpace started competing for usage with Facebook. The inset shows the same plot including the years 2007 and 2008 where we observe that while Facebook's competition significantly reduces user activity on MySpace, it has a much milder impact on network growth.

We start our analysis with the relationship between network growth and subscriber activity. The subscriber activity lifespan is defined as the period between the subscriber join date and the subscriber's last login date. A strong correlation between subscriber activity and his or hers number of friends is also observed. In average, a subscriber with activity lifespan of x weeks has $10.3x$ friends (see Fig. 1). Moreover, Fig. 2 shows the linear relationship between subscriber activity (lifespan) and the network monthly growth between the first semester of 2004 (early 2004) until the end of the second semester of 2006 (late 2006). The green bars show the number of new subscribers of each year observed in our random sample while the black line shows the number of active subscribers (left vertical axis). If we assume that the activity rate of a subscriber grows linearly with his or her activity lifespan, then at each semester between 2006 and 2007 for every active "old" subscriber MySpace got a new subscriber.

The inset of Fig. 2 shows the main plot including years 2007 and 2008. Note that the network growth rate (the bars) decreases significantly in 2007 and 2008 when compared to previous years (2004-2006). Estimating the lifespan distribution between subscribers that joined in 2006 and 2007 using the Kaplan-Meier [22] estimator reveals little change in their lifespan distribution. Another interesting aspect of the 2007/2008 decline is that – as seen later in this work in Figure 5(h) and other indicators – Facebook's competition seems to have only significantly affected MySpace by late 2008. The unchanging lifespan distribution and the constant activity of subscribers in 2007 both indicate that there should be no decline in network growth rates. Interestingly, as we see next, an explanation for this phenomenon is found in the shape of the growth rates (bars): it resembles an asymmetric bell-shaped time series.

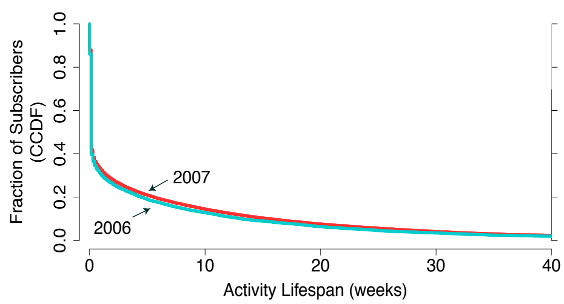


FIG. 3. Kaplan-Meier estimated CCDF of MySpace subscriber lifespans.

The above data suggests three hypotheses about the subscriber dynamics:

($\mathcal{H}_1$) Active subscribers acquire friends at constant rate of 10.3 new friends per week (with a fraction of those being new subscribers); and

($\mathcal{H}_2$) the lifespan of a subscriber increases with the number of friends, that is, an active MySpace subscriber with y friends has lifespan $y/10.3$ weeks.

($\mathcal{H}_3$) the arrival process of new subscribers in MySpace can be approximated by the economic model of adoption of innovations [5, 6, 23]. This model predicts the bell shaped rate of new adoptions observed on MySpace (Fig. 1 inset).

Interestingly, hypothesis $\mathcal{H}_1$ also offers a simple interpretation to the linear relationship between the number of active subscribers and the rate of new subscribers shown in Fig. 2: a fraction of the new edges are the result of invitations to "susceptible users" which accept the invitation, creating a new edge to the newly joined subscriber. The term "susceptible user" denotes potential users that are part of the "true" social network of an active subscriber – friends, acquaintances, or family members – that have not yet joined the OSN.

In what follows we present a model that accurately describes the relationship between subscriber activity, inactivity, and network growth and is simultaneously consistent with hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$. Moreover, our model – through parametrization – remains valid even either hypothesis $\mathcal{H}_1$ or hypothesis $\mathcal{H}_2$ are false.

IV. MODEL

In this section we introduce our model, showing that, by construction, it is consistent with the observations presented in Sec. III.

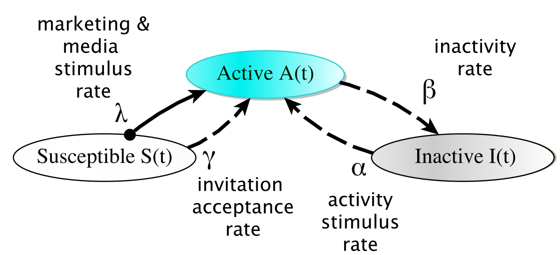
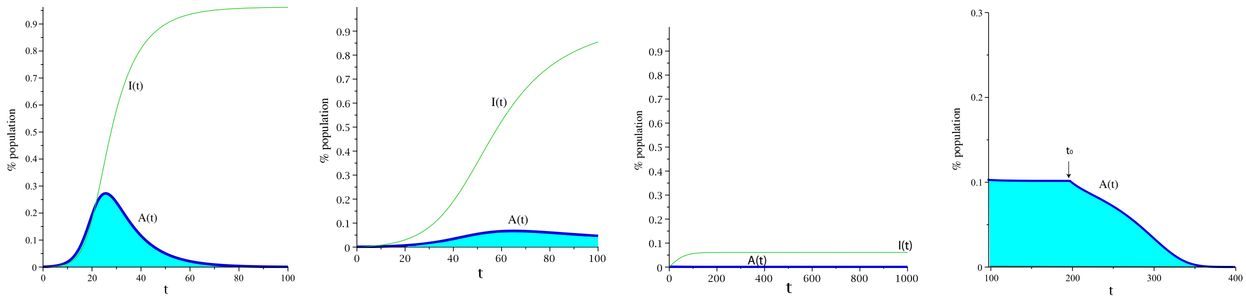


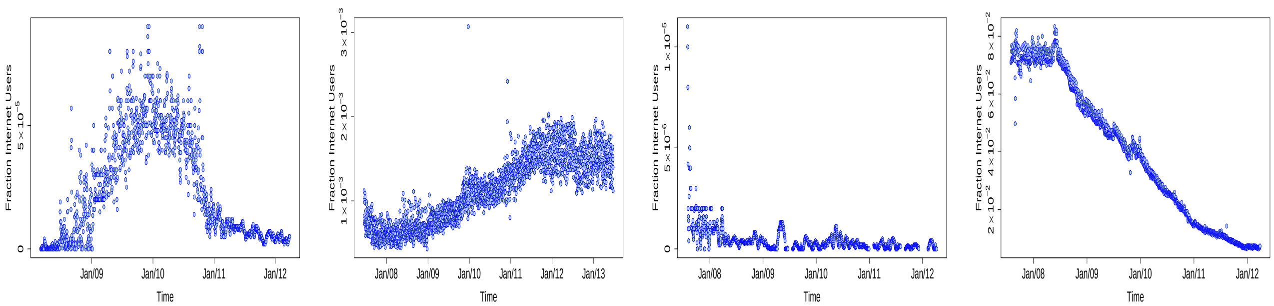
FIG. 4. Model dynamics. In an infinitesimal interval dt, a susceptible user joins the OSN either through an active subscriber invitation with probability γdt or convinced by marketing campaigns (or media exposure) with probability λdt. Inactive subscribers become active from stimuli received from active subscribers with probability αdt. Active subscribers become inactive with probability βdt. Dashed arrows indicate the dependence of the flow on $A(t)$.

A. Basic Model

We start our analysis building a compartmental model of subscriber activity, inactivity, and network growth. Our model accounts for stimuli, such as marketing campaigns, that persuade susceptible users to join the network independently of the network size. These "marketing campaigns" are modeled as have a per unit time probability λ of convincing a susceptible user to join the OSN. Initially, consider a population of universe of susceptible users and OSN subscribers. Let $S(t)$ be the fraction of susceptible users at time t in this universe. Let $A(t)$ be the fraction of active subscribers, and $I(t)$ be the



(a) Marketing intensive OSN **(b)** Self-sustaining OSN **(c)** Unsustainable OSN **(d)** Preyed OSN



(e) 12seconds.tv **(f)** community.babycenter.com **(g)** fiveacross.com **(h)** myspace.com

FIG. 5. (Model behaviors $\times$ Real OSN behaviors) (a)-(d) Four different types OSNs behaviors predicted by our model with respect to subscriber activity. Green line shows the fraction of inactive subscribers $I(t) = 1 - A(t) - S(t)$. (e)-(h) Real subscriber daily usage patters of distinct OSNs measured from October 2007 until January 2012.

fraction of inactive subscribers in this universe. Individuals in this population universe can only be in one of the above three states, making $S(t) + A(t) + I(t) = 1$. In what follows let *network growth rate* denote the number of newly joined subscribers per unit time.

Consider the flow of users during an infinitesimally small time interval dt, illustrated in Fig. 4. The number of susceptible users monotonically reduces over time due to two effects. First, it reduces due to "marketing campaigns": If λdt is the probability that a susceptible user becomes a subscriber because of marketing efforts, then $S(t)$ decreases by $S(t)\lambda dt$. The second effect is due to invitations from active subscribers. An active subscriber sends a *friend invitation* to a randomly chosen user in the universe population (recall the universe population consists of susceptible users together with active and inactive subscribers), which is accepted with probability γdt (we neglect double invitation). If the friend invitation is accepted by a susceptible user, this user joins the OSN as a new active subscriber. The invitation sent to users that are already subscribers has no effect. For now we will disregard the impact of the "true" social network topology in our model.

The latter together with the marketing campaign term yields

$$\frac{dS(t)}{dt} = -S(t)A(t)\gamma - S(t)\lambda \,, \tag{1}$$

which describes the rate of decrease of susceptible users over time.

Let's interpret (1) in the light of the observations of Sec. III. In the beginning, at time t_0, $S(t_0) \approx 1$ and the relationship between the number of active subscribers and the network growth rate is linear, which is consistent with our MySpace observations in Sec. III. More specifically, the above model satisfies the following: (1) the observations in Fig. 2 showing that the number of new subscribers grows linearly with the number of active subscribers; (2) hypothesis $\mathcal{H}_1$, which states that active subscribers make new friends (edges) at a constant rate (by sending friend invitations to a randomly chosen users in the universe population).

The dynamics of subscriber activity is more complex, however. In light of hypothesis $\mathcal{H}_2$, inactive subscribers become active as a consequence of the influence of other active subscribers (this corresponds to the flow Inactive$\rightarrow$Active in Fig. 4). Let αdt be the probability that an active subscriber incites activity in a random user in the universe population. The interaction between active subscribers and susceptible users is already accounted for in parameter γ, so we can safely ignore this event. The interaction among two active subscribers can also be ignored. If an active subscriber incites activity in an inactive subscriber, the inactive subscriber becomes active. We denote α the *attention stimulus* parameter. The increase in $A(t)$ resulting from the Inactive $\rightarrow$ Active subscriber flow is $I(t)A(t)\alpha dt$.

Active subscribers also eventually become inactive after some time. Let βdt be the probability that an active subscriber become inactive during interval dt. The decrease in $A(t)$ caused by the Active $\rightarrow$ Inactive flow is $A(t)\beta dt$. Another increase in the Active $\rightarrow$ Inactive flow is due to susceptible users that newly join the network, thus making the contribution towards $A(t)$ be $S(t)A(t)\gamma dt + S(t)\lambda dt$.

Gathering the four terms described above yields the subscriber activity equation

$$\frac{dA(t)}{dt} = S(t)A(t)\gamma + S(t)\lambda - A(t)\beta + I(t)A(t)\alpha\,. \quad (2)$$

Last, as $S(t) + A(t) + I(t) = 1$, the chance in subscriber inactivity is

$$\frac{dI(t)}{dt} = A(t)\beta - I(t)A(t)\alpha\,. \quad (3)$$

Note that the flows Inactive $\rightarrow$ Active and Susceptible $\rightarrow$ Active in (2) are proportional to the fraction of active subscribers, thus, an increase in the fraction of active subscribers also increases these flows. Not surprisingly, online social network websites have recently taken active instance against subscriber inactivity. Facebook, for instance, in recent years introduced notification messages of the form: "Here's some activity you may have missed" (Facebook subscribers may opt out of such messages), even if almost no activities are reported or the subscriber consistently shows no interest in them. Network growth is not the only reason why online social networks want to keep it subscribers active. OSN revenue also often depends on the number of active subscribers [17]. But, most importantly, next we see that the survival of the network may depend on whether the ratio α/β is greater or smaller than one, creating a great incentive to increase α by sending activity reminders to inactive subscribers.

B. Model Discussion

In what follows we discuss some properties and shortcomings of our model.

Stability. The diffusion of innovations model [5, 6, 23] predicts that after a certain (possibly large) time $t^\star$, the OSN nearly exhausts its pool of susceptible users: $S(t') \approx 0, \forall t' > t^\star$. In what follows we analyze our model after the above critical time, $t > t^\star$. Approximating $I(t) = 1 - S(t) - A(t) \approx 1 - A(t)$ and substituting in (2)

$$\begin{aligned} \frac{dA(t)}{dt} &= \cancelto{0}{S(t)}A(t)\gamma + \cancelto{0}{S(t)}\lambda - A(t)\beta + \cancelto{(1-A(t))}{I(t)}A(t)\alpha, \\ \frac{dA(t)}{dt} &\approx -A(t)\beta + (1 - A(t))A(t)\alpha. \end{aligned} \quad (4)$$

In equilibrium, $dA(t)/dt \approx 0$, (4) yields for $t > t^\star$,

$$A(t) \approx \begin{cases} 0 & \text{if } \alpha/\beta \leq 1, \\ 1 - \beta/\alpha & \text{if } \alpha/\beta > 1. \end{cases}$$

The above results are also the stationary solutions of the compartmental SIS epidemic model [24] with infection rate α and recovery rate β (see Keeling [24] for details). In epidemiology the value $r_c = \alpha/\beta$ is known as the epidemic threshold. Note that when $r_c < 1$ the OSN is expected to eventually wither and die, regardless of the amount of marketing effort λ.

Parametrization. Regrettably, OSNs do not provide data that can be used to parameterize our models; our datasets are not rich enough to be directly used to parameterize the model either. We believe, however, that under certain conditions one should be able to infer the model parameters – with reasonable accuracy – by just observing the evolution of $A(t)$. Parameter estimation is outside the scope of this work, however. Instead, we focus on qualitative results, leaving the important task of assessing how accurate can our model forecast the evolution of OSNs for future work. Forecasting should benefit from richer models that take into account topological information, such as the model presented in Sec. IV D.

Susceptible User Population. The diffusion of innovations model assumes the final number of adopters of the innovation is known, which is one of the factors goes into determining the normalized rate of adoption λ. For OSNs this means knowing how many subscribers will eventually join the network. In practice, one can estimate this value by estimating the "bell" shape of the new subscriber rate curve, i.e., determine the shape of the green bars in Fig. 2's inset, and then integrate the curve to find the size of the susceptible user population.

Arrivals and Departures. Our model does not take into account (a) arrivals of new susceptible users or (b) departures of inactive subscribers canceling their accounts. These arrivals and departures can, however, be easily accounted for by (a) adding a drift term ω to $dS(t)/dt$ in (1) and (b) subtracting a drift term η from $dI(t)/dt$ in (3). As arrivals and departures introduce two new parameters in the model, a variety of new possible combinations between all parameters unleash a host of distinct $A(t)$ time series behaviors. We believe, however, that OSNs subscribers rarely close their accounts. Having $\eta \ll 1$ introduces a complication in the model: arrivals without departures creates an unbounded growth in population, increasing the universe of users as $(\omega-\eta)t$. Thus, we choose to leave the analysis of our model with arrivals and departures as future work.

C. Modeling Sudden Changes

Sudden changes in the OSN landscape may significantly affect dynamics of a given OSN. The most important sudden change is the appearance of competing OSNs. Marketing campaigns and the impact of media exposure are other important types of sudden changes. A prime example of the first is the case of MySpace and Facebook. Facebook is, arguably, the most successful OSN startup in the Internet's short history [2]. Since its inception in 2004, Facebook's subscriber activity grew fast and steadily until the daily subscriber activity leveled up at about 45% of all daily Internet users in early 2012 (see Fig. 8(b)).

Opposite to Facebook's fast success we find Myspace's quick downfall. Figure 5(h) shows Alexa.com's traffic report estimate of MySpace's daily subscriber activity from late 2007 until early 2012 (MySpace was launched in 2004 but our data only covers the 2007–2012 period, see Sec. VI for more details). Note that from late 2007 until approximately late 2008, MySpace subscriber activity remains constant. However, after late 2008 – and through the end of our measurement – MySpace suffers a steady decline in its active subscriber base. By April 2008 Facebook usage overtook MySpace in the number of unique website visitors [2].

We model such sudden changes by means of two functions. First, we use a Heaviside step function to mark the beginning of the diffusion of the new subscriber behavior, like the birth of Facebook or when Facebook reaches a critical mass that enables it to influence MySpace subscribers. Then, after the new behavior starts, we use a logistic function to describe the diffusion of the adoption of the new behavior in the subscriber population, such as the adoption of Facebook among MySpace subscribers. As revisited in Sec. II, the logistic function is widely used in economic theory to describe the diffusion of innovations in a population [5, 6, 23]. In our Facebook v.s. MySpace example, the logistic function describes the process by which MySpace subscribers discover Facebook.

In what follows we model how the change in subscriber behavior (opening also a Facebook account) affects the parameters of our model. In the specific case of MySpace and Facebook, after late 2007 MySpace subscribers are reported to have started fleeing to Facebook [19], although it seems that only by late 2008 the Facebook competition really impacted MySpace. We believe, however, that a subscriber is not going to close his or her MySpace account to become a Facebook subscriber, which would be a naive oversimplification of the true process. Instead, MySpace subscribers open also a Facebook account and start sharing their attention between the two OSNs. Moreover, susceptible users may now choose to join Facebook instead of MySpace. This change of behavior affects γ, α, and λ but not β, changing equations (1), (2), and (3), yielding

$$\begin{aligned}
\frac{dS(t)}{dt} &= -(S(t)A(t)\gamma - S(t)\lambda)\psi(t)\,, \\
\frac{dA(t)}{dt} &= (S(t)A(t)\gamma + S(t)\lambda)\psi(t) \\
&\quad - A(t)\beta + I(t)A(t)\alpha\psi(t)\,, \\
\frac{dI(t)}{dt} &= A(t)\beta - I(t)A(t)\alpha\psi(t)\,,
\end{aligned} \tag{5}$$

where

$$\psi(t) = 1 - H(t-t_0)\phi(t-t_1),$$

and

$$\phi(t') = \frac{1}{1+e^{-\theta t' + C}}$$

is the logistic function,

$$H(t') = \begin{cases} 0 & \text{if } t' < 0, \\ 1 & \text{if } t' \geq 0, \end{cases}$$

is the Heaviside step function, t_0 is the time that the new behavior starts, and t_1 is a time parameter related to the diffusion of the new subscriber behavior. Henceforth, we refer to the model in (5) as the *preyed OSN* model. Preyed emphasizes that this models an OSN with a strong competitor.

Fig. 5(d) shows a numeric solution of $A(t)$ in (5) using α, β, γ, and λ parameters of a "self-sustaining OSN" described in details in Sec. V). We set $t_0 = 200$ and $t_1 = 250$. Note that for $t < t_0$ the OSN is active and stable and $S(t_0) \approx 0$ (which greatly simplifies (5)). After $t > t_0$ the OSN starts to feel the effects the new behavior, losing active subscribers almost linearly until there are virtually no active subscribers left at time $t = 350$.

Comparing Figs. 5(d) and 5(h) we observe that our model qualitatively describes the evolution observed in the data. It is crucial to point out, however, that the virtual disappearance of active subscribers near $t = 350$ in our model happens when $\alpha(t) = 0.8 \times \alpha$ (see Fig. 6). Thus, the competing OSN objective is to reduce the *attention stimulus* of its competitors' subscribers just enough to get the competitor's feedback loop between active and inactive subscribers (illustrated in Fig. 4) to tip towards inactivity.

The above creates a new hypothesis that explains how OSNs such as MySpace and others die after Facebook appearance (see observations of other OSNs in Fig. 10). MySpace OSN subscribers canceled their accounts and switch to Facebook. Rather, Facebook attracted MySpace subscriber attention just enough to tip MySpace's active-inactive dynamics towards inactivity. The above analysis also provides a plausible explanation to why in 2011, during the time of Google+ explosive growth, Facebook began sending activity reminders – "Here is some

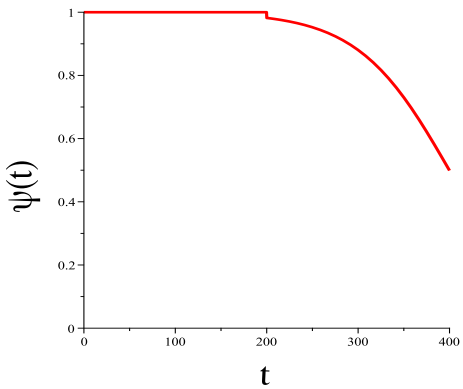


FIG. 6. Evolution of $\psi(t)$.

activity you have missed" – to its subscribers[32]: According to our hypothesis, Facebook was actively pushing against its critical inactivity tipping point.

Marketing campaigns and media exposure can also be modeled by replacing λ in equations (1) and (2) with $\lambda(1+\mu\delta(t-t_0))$, where $\delta(\cdot)$ is the Dirac delta function, t_0 is the time that the marketing campaign starts, and μ is the intensity of the campaign.

In the presence of richer network data our basic model can be further refined. In what follows we present a refined model that takes into account network topology information.

D. Topology-aware Model

For the sake of completeness, we extend our model to take into account network topologies. We consider the topologies of the OSN and the "true" social network by which OSN subscribers invite new prospective users. However, as we do not have access to datasets with network topology information, we leave the qualitative validation of our topology-aware model against real data as research to be done when such datasets are available. The following equations show how our basic model can be refined in the presence of extra side information about the networks.

Let $\mathbf{A}^{(real)}$ be the adjacency matrix of the true social network that subscribers use to invite susceptible users. Let $\mathbf{A}^{(OSN)}$ be the OSN adjacency matrix. Let $\mathbf{s}_u(t)$, $\mathbf{a}_u(t)$, and $\mathbf{i}_u(t)$ be the probability that user u is in the susceptible non-subscriber, subscriber active, and subscriber inactive states, respectively. As user u can only be in one of the above three states, $\mathbf{s}_u(t)+\mathbf{a}_u(t)+\mathbf{i}_u(t)=1$. Let $\mathbf{s}(t)$, $\mathbf{a}(t)$, and $\mathbf{i}(t)$ be the vectors that represent the states susceptible non-subscriber, subscriber active, and subscriber inactive of all potential users of the OSN, respectively.

The following set of equations describe the evolution of $\mathbf{s}(t)$, $\mathbf{a}(t)$, and $\mathbf{i}(t)$ of user u:

$$\begin{aligned}
\frac{d\mathbf{s}_u(t)}{dt} &= -\gamma\,\mathbf{s}_u(t)\,(\mathbf{A}^{(real)}\,\mathbf{a}(t))_u - \mathbf{s}_u(t)\lambda\,, \\
\frac{d\mathbf{a}_u(t)}{dt} &= \gamma\,\mathbf{s}_u(t)\,(\mathbf{A}^{(real)}\,\mathbf{a}(t))_u + \mathbf{s}_u(t)\lambda \\
&\quad + \alpha\,\mathbf{i}_u(t)\,(\mathbf{A}^{(OSN)}\,\mathbf{a}(t))_u - \beta\mathbf{a}_u(t)\,, \\
\frac{d\mathbf{i}_u(t)}{dt} &= \beta\mathbf{a}_u(t) - \alpha\,\mathbf{i}_u(t)\,(\mathbf{A}^{(OSN)}\,\mathbf{a}(t))_u\,.
\end{aligned} \tag{6}$$

The equations in (6) can be greatly simplified if assumptions can be made about the structure of $\mathbf{A}^{(real)}$ and $\mathbf{A}^{(OSN)}$. For instance, if $\mathbf{A}^{(real)}$ and $\mathbf{A}^{(OSN)}$ are fully connected networks, then (6) can be well approximated by our basic model. Other examples are Avrachenkov et al. [25] and Pastor-Satorras and Vespignani [26], but these simplifications are outside the scope of this work. In the presence of topological information, the calculation of r_c can be refined using recent results on SIS models on random networks [27].

V. PREDICTED BEHAVIOR

In this section we show that our models predict four distinct behaviors of the $A(t)$ time series. In this section we consider our basic model presented in Sec. IV A and its extension in Sec. IV C. We choose to use the evolution of $A(t)$ as a metric of interest because (a) it is through active subscribers that OSNs monetize their service, (b) the number of active subscribers is proportional to the rate of network growth; and finally, (c) we have access to subscriber activity data of a variety of OSNs, which we later compare against the behaviors predicted by our model.

Varying the OSN parameters of our model we predict four distinct behaviors in the evolution of $A(t)$:

- ♦ **Marketing intensive OSN**. A marketing intensive OSN is characterized by a large λ – that is responsible for a sizable fraction of the initial network growth – and a small ratio $\alpha/\beta < 1$. The fast initial arrival of new subscribers, propelled by the marketing campaign, creates a significant active subscriber base. However, the active subscriber base is not sustainable over time. The marketing campaign saturates and exhausts the potential subscriber base. With $\alpha/\beta < 1$ the active subscribers are not able to elicit enough inactive subscribers into activity to sustain any positive level of activity in the network. In these scenarios, the fraction of active subscribers, $A(t)$, has the characteristic asymmetric bell shaped curve shown in Fig. 5(a).

- ♦ **Self-sustaining OSN**. A self-sustaining OSN may or may not be marketing intensive. The main characteristic of a self-sustaining OSN is having the ratio $\alpha/\beta > 1$, which guarantees that the fraction of active subscribers is stable and positive. The asymptotic fraction of active subscribers is $\lim_{t\to\infty} A(t) =$

$1-\beta/\alpha$. A self-sustaining OSN may also have a large λ but, unlike *marketing intensive OSNs*, it can sustain a non-negligible level of interest below the marketing peak. In the absence of strong marketing, the fraction of active subscribers, $A(t)$, has a characteristic sigmoid shape, as shown in Fig. 5(b).

- **Unsustainable OSN**. An unsustainable OSN is not marketing intensive and has ratio $\alpha/\beta < 1$. The ratio $\alpha/\beta < 1$ guarantees that the fraction of active subscribers is unstable and that it asymptotically goes to zero. In the absence of strong marketing, the fraction of active subscribers, $A(t)$, never takes off and even the fraction of inactive users remains small, growing slowly only due to $\lambda \ll 1$, as shown in Fig. 5(c).

- **Preyed OSN**. The sudden appearance of a competitor OSN may chip away subscriber attention. We use (5) to model this competition. Figure 5(d) shows the numerical solution to (5) with the following parameters: $t_0 = 200$, $t_1 = 250$, $\theta = 0.02$, and $C = \log(20)$. The competition will eventually annihilate the OSN activity after the OSN reaches the critical point $\alpha\psi(t)/\beta < 1$ (Fig. 6 shows the evolution of $\psi(t)$ over time, showing that $\psi(t) > 0.5, \forall t < 400$). Note that when the competition starts, the fraction of active subscribers, $A(t)$, decreases initially almost linearly (or super-linearly depending on the parameters of the model) but later takes the shape of an exponential decay, as shown in Fig. 5(d).

Next we compare the above four predicted subscriber activity time series, $A(t)$, against real world data obtained from nearly thirty OSN websites.

VI. REAL OBSERVED BEHAVIOR

In this section we compare our model against real daily subscriber activity obtained from a variety of Internet sites. We select Alexa.com's daily reach per million as our metric of interest. The daily reach per million is measured as the number of unique website visitors per million web visitors in the same day. We note, however, that Alexa.com is unlikely to include the traffic generated by mobile phone users in its data. In the period that we are analyzing, from mid 2007 until early 2012, we believe that mobile users do not significantly affect the measurements. In what follows we give a short description of what is publicly known about Alexa's data collection methodology. We then show that the activity behaviors observed in these datasets match that of our model. Finally, we present a short description of the datasets.

a. A Short Description of Alexa.com Alexa.com is subsidiary of Amazon.com that provides commercial web traffic data. Today, Alexa provides traffic data, global rankings and other information on 30 million websites [28]. Alexa ranks sites based primarily on tracking information of users of its toolbar available for all the Internet Explorer, Firefox and Google Chrome web browsers. Since 2008 Alexa claims to remove self-selection bias – bias related to gathering data of a specific audience subgroup that is more likely to install Alexa's toolbar – by taking into account other data sources "beyond Alexa Toolbar users" [28], but the nature of such data sources and the methodology employed are not disclosed. Nonetheless, because Alexa's report is detailed and widely used in the industry, we believe that Alexa's *unique subscriber* daily traffic reports are a good source of data for our study.

In what follows we use the evolution of the (daily) fraction of unique Internet visitors of an OSN website as a proxy for the evolution of its active subscriber base, denoted as $A(t)$ in our model. Interestingly, in our datasets we observe the same four distinct types of OSN activity evolution that are predicted by our model (a short description of these websites is presented in the next section). For each of the four behaviors, we start by illustrating it using one OSN, followed by all other OSNs that present the same behavior.

- **Asymmetric Bell Shaped Activity.** Figure 5(e) shows the fraction of daily unique Internet visitors of 12seconds.tv (a social video-sharing website) from early 2008 until early 2012. The graph shows an bell shaped growth and descent followed by a steady decrease in unique visitors. The asymmetric bell shape matches the predicted behavior of the *market intensive OSN* parametrization of our model shown in Fig. 5(a). Fig. 7 shows similar activity behavior in a variety of other websites.

- **Sustained Activity.** Figure 5(f) shows the fraction of daily unique Internet visitors of community.babycenter.com (an OSN where new parents share experiences) from early 2008 until early 2013. We see a quite different picture from above. Here the graph shows a steady growth until the curve levels up. This behavior matches the predicted behavior of the *self-sustained OSN* parametrization of our model shown in Fig. 5(b). Fig. 8 shows similar activity behavior in a variety of other websites.

- **Low/Decreasing Activity.** Figure 5(g) shows the fraction of daily unique Internet visitors of fiveacross.com (a social network platform targeting companies trying to connect with their clients and users). Here we observe low activity with spikes that quickly return to the baseline activity level. This behavior matches the predicted behavior of the *unsustainable OSN* parametrization of our model shown in Fig. 5(c). Fig. 9) shows similar activity behavior in a variety of other websites.

- **Sudden Activity Decrease.** Figure 5(h) shows the fraction of daily unique Internet visitors of myspace.com. Here we observe that MySpace's activity was stable until mid 2008 when a steady decay in activity began. The initial decrease in activity is almost linear in the beginning, later transitioning to an exponential decay. This behavior matches the predicted behavior of the *preyed OSN* parametrization of our model shown in Fig. 5(d) Fig. 10 shows similar activity behaviors in a variety of other Facebook competitors such as Friendster, Hi5, Multiply, and Orkut.br.

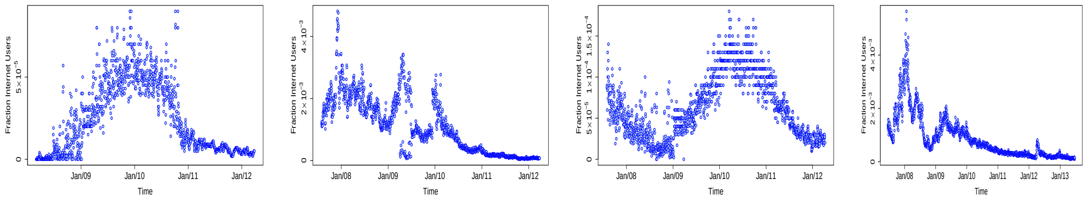
The subscriber activity datasets show great agreement with our model. Real subscriber activity reports are classified into one of the four classes predicted by our model. In what follows we present a detailed account of these datasets.

Description of Other Datasets

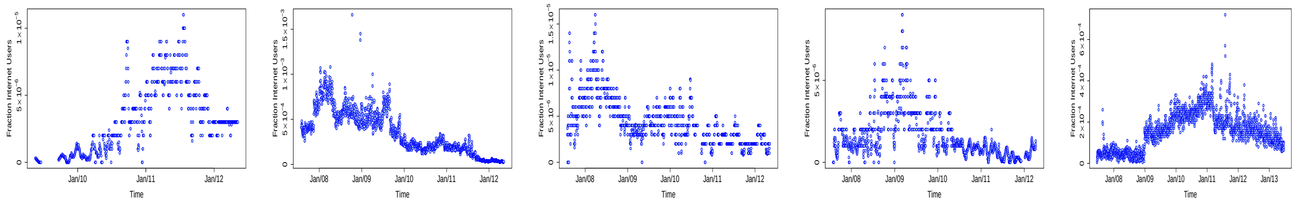
The following is a description of the websites used in Figs. 8, 7, 9, and 10. Together, along OSN websites, we also present subscriber activity of online dating websites. Online dating websites share some common activity and growth traits with OSNs. E.g., (a) the activity of an active subscriber incites activity of other subscribers and (b) the activity of a subscriber influences other people in their "real" social network to also join the network.

In online dating, however, subscriber lifespan is likely short as subscribers find stable relationships and cancel their accounts. While these departures can be easily accounted for in our model, see Sec. IV B, this behavior may significantly differ from the behavior of the remaining datasets. Thus, instead, we report the subscriber activity of dating websites that we believe provide services to subscribers interested in non-committing relationships. In what follows we provide a description of these OSN and online dating websites[33]:

- **12seconds.tv:** "12seconds.tv is a Twitter-like video status service. It gives you 12 seconds to share video moments from your life" [29].
- **alstrasoft.com:** "AlstraSoft provides a range of components for organizations who want to add functionality to their websites. Among these components is E-friends, a downloadable white label social networking platform" [29].
- **fiveacross.com:** "FiveAcross is a Social Network platform targeting companies trying to connect with their clients and users, was acquired by Cisco in February of 2007" [29].
- **cafemom.com:** "CafeMom is a social network site for moms, reaching an audience of more than 20 million users." [29].
- **patientslikeme.com:** Patientslikeme is a social networking site that allows people with similar diseases to share their experiences about treatments, doctors, and seek emotional support.
- **ashleymadison.com:** "Ashley Madison is a Married Dating service and social network for those engaged in relationships but looking to have an affair" [29].
- **linkedin.com**. LinkedIn is a social network website focused on work relationships. As expected, linkedin.com traffic presents different intensities between weekdays and weekends.
- **community.babycenter.com:** The Community Baby Center, launched in 2008, is a social network for parents with young children of all ages to share their experiences.
- **kiva.org:** Kiva uses the power of community and social networking to help someone out of poverty by enabling communities of lenders to finance poor people all over the world with enough money to start a business or improve their standard of living.
- **worlddatingpartners.com:** Dating service plataform.
- **vox.com:** Vox abruptly ceased operations in January 2011. Vox was a web based blogging & social networking platforms." [29].
- **rockyou.com:** "RockYou builds... social experiences on the web through properties like Zoo World, Pieces of Flair, and Birthday Cards" [29]. The recent company push towards the Facebook platform may have reduced the traffic of rockyou.com. Game companies, however, are known to have large peaks of subscriber interest [30] that are similar to large marketing campaigns.
- **sixapart.com:** Six Apart was formed in 2001 as a blog solution provider.
- **true.com:** Online dating website.
- **marriedsecrets.com:** Married Dating service and social network.
- **groupswim.com:** GroupSwim is a provider of on-demand social software for businesses [29].
- **raisingthem.com:** Similar to parentslikeme.com.
- **crowdtilt.com:** "At Crowdtilt, were trying to make it easier for groups to do things together. Whether its a party-bus to the next Phish show or pooling money for a cause you know your close friends are passionate about" [29].
- **goingon.com:** GoingOn provides a service to build online communities.
- **theport.com:** ThePort provides enterprise social network solutions [29].
- **hi5.com:** Founded in 2003, hi5 is a social networking website [29].
- **friendster.com:** Friendster, launched in 2002, is one of the first social networking sites [29].
- **multiply.com:** Multiply is a mix between an e-commerce platform and a social networking website, offering sellers a combination of e-commerce and social communications tools. Multiply is the largest marketplace in Southeast Asia [29].
- **orkut.com.br:** Until recently Orkut.com.br was one of the most visited websites in Brazil. "Originally hosted in California, in August 2008 Orkut control went to Google Brazil in the city of Belo Horizonte" [29].

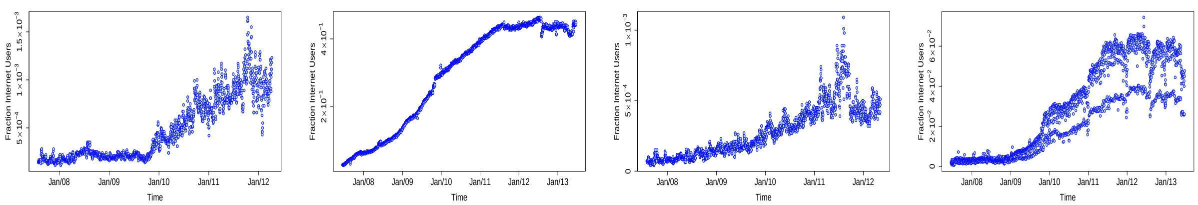


(a) 12seconds.tv **(b)** rockyou.com **(c)** sixapart.com **(d)** flixster.com

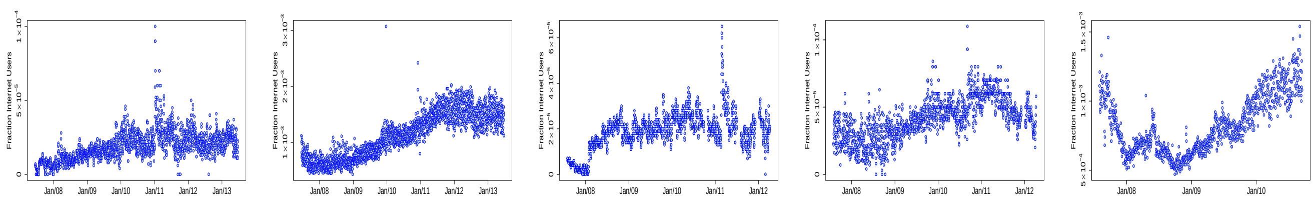


(e) sixsingles.com **(f)** true.com **(g)** marriedsecrets.com **(h)** groupswim.com **(i)** kiva.org

FIG. 7. (Marketing intensive OSN) Websites whose active subscriber time series ($A(t)$) resemble marketing intensive OSNs in our model. Kiva.org, however, may also be a self-sustaining OSN that in 2011 was preyed by another service or a mobile service.



(a) cafemom.com **(b)** facebook.com **(c)** ashleymadison.com **(d)** linkedin.com

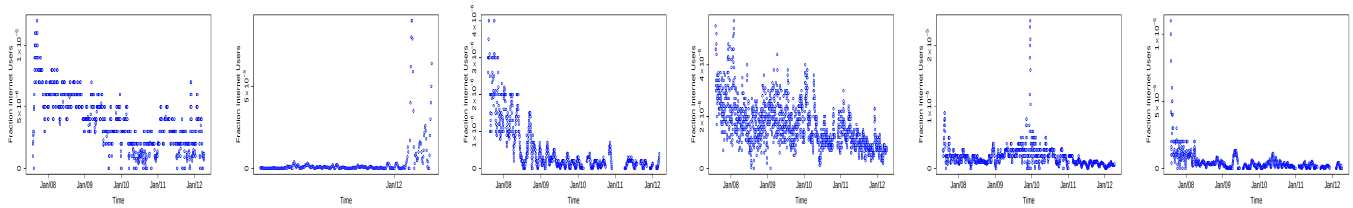


(e) patientslikeme.com **(f)** community.babycenter.com **(g)** worlddatingpartners.com **(h)** alstrasoft.com **(i)** vox.com

FIG. 8. (Self-sustaining OSN) Websites whose active subscriber time series ($A(t)$) resemble self-sustaining OSNs in our model. (Note: vox.com interrupted service in January 2011.) LinkedIn traffic presents different intensities between weekdays and weekends.

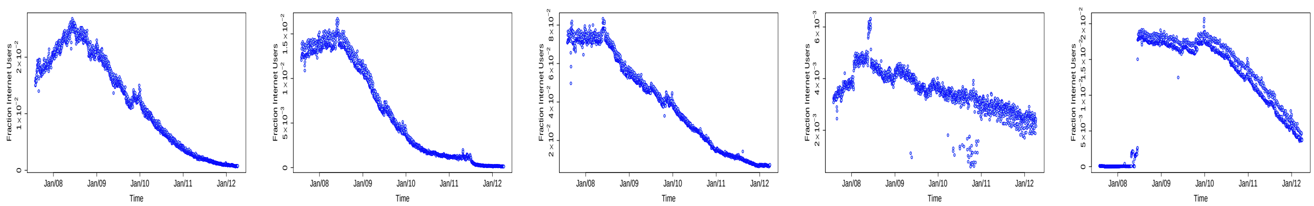
VII. CONCLUSIONS & FUTURE WORK

This study sheds light on the mechanisms related to growth, stability, activity, inactivity, and death of OSNs. Monitoring, over the course of five years, the relationship between subscriber activity and network growth of a large OSN (in number of subscribers), our study observes that the network growth rate increases linearly with the number of active subscribers in the network. We also observe that over five years the network growth rate follows an asymmetric bell-shaped time series. We then formulate three hypotheses that can explain the observed phenomenon. Using these hypotheses we introduced a model that describes the dynamics of OSN subscriber activity, inactivity, and growth, and present two extensions to this model (one including sudden changes in



(a) raisingthem.com **(b)** crowdtilt.com **(c)** elggspaces.com **(d)** goingon.com **(e)** theport.com **(f)** fiveacross.com

FIG. 9. (Unsustainable) Websites whose active subscriber time series ($A(t)$) resemble unsustainable OSNs in our model.



(a) hi5.com **(b)** friendster.com **(c)** myspace.com **(d)** multiply.com **(e)** orkut.com.br

FIG. 10. (Preyed OSN) Websites whose active subscriber time series ($A(t)$) resemble preyed OSNs in our model.

subscriber behavior and the other refining the model using network topology information). We show that these model predict four distinct overall evolution behaviors of the number of active subscribers. We classify these evolution behaviors into four overall OSN types: *marketing intensive OSN*, *Self-sustaining OSN*, *Unsustainable OSN*, and *Preyed OSN*. Finally, using unique visitors activity reports of a variety of OSNs we show that our four OSN behaviors are able to describe the OSN activity in these datasets.

Our work makes a positive step towards modeling the dynamics of OSN activity, inactivity, and growth. Nevertheless, there is much left to do. For instance, the mechanism by which active subscribers recruit new subscribers from the Internet user population is not well understood. Our work bypasses this problem by fixing the size of the susceptible user set. However, it is unclear what determines if an Internet user susceptible to a given OSN. Another interesting direction of study is whether the observed subscriber activity burstiness observed in various studies [7–11] can be explained by the dynamics of our models. Although our model is Markovian and the observed activity burstiness is known to have long-range correlations, it has been shown that even simple Markov chains (of few states) can mimic the behavior of processes with long-range correlations over multiple time scales [31].

Acknowledgments

This work was supported by the NSF grant CNS-1065133 and ARL Cooperative Agreement W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the of official policies, either expressed or implied of the NSF, ARL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

[1] TechCrunch Deadpool, `http://techcrunch.com/tag/deadpool/`.

[2] Wikimedia Foundation, `http://en.wikipedia.org/wiki/List_of_social_networking_websites`.

[3] W. B. Arthur, *Increasing Returns and Path Dependence in the Economy* (U. Michigan Press, 1994).

[4] M. H. Stanley, L. A. Amaral, and S. V. Buldyrev, Nature **379**, 804 (1996).

[5] E. Mansfield, Econometrica: Journal of the Econometric Society pp. 741–766 (1961).

[6] E. Mansfield, The Review of Economics and Statistics **45**, 348 (1963).

[7] J. Stehlé, A. Barrat, and G. Bianconi, Physical review E **81**, 035101 (2010).

[8] A.-L. Barabási, Nature **435**, 207 (2005).

[9] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski, New Jour-

nal of Physics **14**, 013055 (2012).
[10] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, Proceedings of the National Academy of Sciences **106**, 12640 (2009).
[11] A. Garas, D. Garcia, M. Skowron, and F. Schweitzer, Scientific Reports **2** (2012).
[12] G. Kossinets and D. J. Watts, Science **311**, 88 (2006).
[13] J. Leskovec, J. Kleinberg, and C. Faloutsos, TKDD **1**, 2 (2007).
[14] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, in *WOSN* (2008), pp. 25–30.
[15] D. Schiöberg, S. Schmid, F. Schneider, S. Uhlig, H. Schiöberg, and A. Feldmann, in *ACM WebSci* (2012), pp. 265–274.
[16] P. Kavassalis, S. Lelis, M. Rafea, and S. Haridi, Commun. ACM **47**, 50 (2004).
[17] P. Cauwels and D. Sornette, The Journal of Portfolio Management **38**, 56 (2012).
[18] C. C. Aggarwal, *An introduction to social network data analytics* (Springer, 2011).
[19] D. Boyd, in *Race After the Internet*, edited by L. Nakamura and P. A. Chow-White (Routledge, 2011), pp. 203–222.
[20] B. Ribeiro, W. Gauvin, B. Liu, and D. Towsley, in *INFOCOM NetSciCom* (2010), pp. 1–6.
[21] A. Jesdanun, *Myspace popularity with teens fizzles, msnbc* (Nov. 2007).
[22] E. L. Kaplan and P. Meier, Journal of the American Statistical Association **53**, 457 (1958).
[23] M. Rogers Everett, New York (1995).
[24] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011).
[25] K. Avrachenkov, P. Basu, G. Neglia, B. Ribeiro, and D. Towsley, arXiv preprint arXiv:1212.5035 (2012).
[26] R. Pastor-Satorras and A. Vespignani, Physical review letters **86**, 3200 (2001).
[27] R. Parshani, S. Carmi, and S. Havlin, Physical Review Letters **104**, 258701 (2010).
[28] Alexa.com, *http://www.alexa.com/help/traffic-learn-more*.
[29] Crunchbase, *http://www.crunchbase.com*.
[30] Z. Forró, P. Cauwels, and D. Sornette, arXiv preprint arXiv:1112.6024 (2011).
[31] S. Robert and J.-Y. Le Boudec, Performance Evaluation **30**, 57 (1997).
[32] The 2011 Google+ explosive growth is documented in the literature [15]
[33] Our study chose these websites arbitrarily and makes no claim regarding future trends of subscriber activity, viability, or marketability of these websites. Moreover, our model is a first order approximation and, as such, does not model all contributions to the observed activity.