

A Universal Model for Growth of User Population of Products and Services

Choujun Zhan and Chi K. Tse
The Hong Kong Polytechnic University, Hong Kong
(*e-mail*: michael.tse@polyu.edu.hk)

Abstract

We consider a network of interacting individuals, whose actions or transitions are determined by the states (behaviour) of their neighbours as well as their own personal decisions. Specifically, we develop a model according to two simple decision-making rules that can describe the growth of the user population of a newly launched product or service. We analyse 22 sets of real-world historical growth data of a variety of products and services, and show that they all follow the growth equation. The numerical procedure for finding the model parameters allows the market size, and the relative effectiveness of customer service and promotional efforts to be estimated from the available historical growth data. We study the growth profiles of products and find that for a product or service to reach a mature stage within a reasonably short time in its user growth profile, the user growth rate corresponding to influenced transitions must exceed a certain threshold. Furthermore, results show that individuals in the group of celebrities having numerous friends become users of a new product or service at a much faster rate than those connected to ordinary individuals having fewer friends.

Contents

1 Introduction	1
2 User Growth Model	3
3 Analysis of Datasets	6
3.1 Relationship Between the Measured Data and User Growth Process	6
3.2 Application Examples of First/Second-Order Growth Models	7
3.3 Fitting 21 Real Data by First-Order Growth Model	10
3.4 Analysis of Growth Span of Products	14
4 Conclusion	15
References	16

1 Introduction

The study of the spreading of certain behaviour in a connected community has been conveniently modelled in terms of a network of individuals whose behaviour change as a result of mutual influence. The transitional behaviour of individuals, such as spreading of an idea or adoption of a new product or service in a networked community, partly resembles the

process of epidemic spreading where contacts among individuals, i.e., influenced transitions, remain a deciding factor for the growth in the number of infected individuals. Thus, the conventional SI model would capture the behaviour of the spread of certain behaviour or the growth of the user population of a product or service, associated with influence of an individual to its other connected individuals which is similar to “transmission” of virus from an infected individual to a connected susceptible individual. In the world of business and a real social network, however, personal preference and educated decision of individuals do play an equally important role in determining how behaviour spreads or how a product popularises in the user community.

There is a wealth of literature devoted to the study of behaviour spreading on social networks, growth of specific business sectors, and popularisation of products and services, including both theoretical analysis and empirical research on real data. In the theoretical work by Campbell (2013), a model of demand, pricing and advertising with individuals engaging in word-of-mouth communications has been reported. Effects of incorporating heterogeneity into several broad classes of models have also been studied by Young (2009). Moreover, empirical research on real data has been actively pursued. In the work of Leskovec *et al.* (2007), a network of 4 million people making recommendations on half a million products has been used to analyse user behaviour, propagation of opinions and scale of cascade of opinions. Empirical research on the cascade of messages in networks has revealed network structures of very low clustering, and showed several features not observed in other social dynamic processes (Iribarren and Moro, 2011). Furthermore, the predictive performance of new models based on *function regression* has been compared with several other models (Sood *et al.*, 2009).

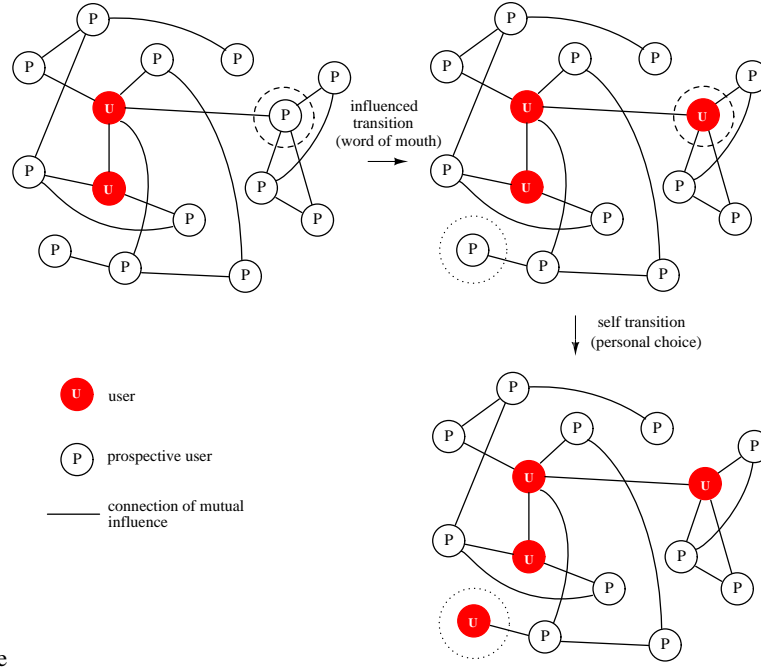
In this paper, we analyse the growth of the user population of a product or services in terms of a connected community or a network of users (active agents) and prospective users (susceptible agents), similar to a network comprising nodes and edges (Strogatz, 2001; Albert and Barabási, 2002; Barabási and Albert, 1999), representing a set of relationships across the edges through which individuals exert their influence on others. A prospective user (P) does not use the product or service at the present time, but may transit into a user (U) of the product or service at a later time. We consider two transition rules here. The first rule is a peer-influenced transition, and the second one is a self (independent) transition. These two rules represent two basic types of human decision behaviour, namely, by word of mouth and personal choice (Geroski, 2000; Goldenberg *et al.*, 2001; Campbell, 2013). Individuals' decisions are often influenced by the decision of their neighbours, which is common in social networks. The peer-influence rule corresponds to a prospective user who, after interacting with other users and being informed about the benefits of using the product or service, becomes a user. The transition driven by personal choice is a spontaneous decision of the particular prospective user and has nothing to do with other users whom the prospective user connects with. A prospective user may be positively informed about the product or service via the media, advertisements, or other means unrelated to the users of the product or service whom the prospective user knows. Then, the growth of users in the network can be studied in terms of stochastic processes. This model has the basic structure of the SIS model of epidemiology (Pastor-Satorras and Vespignani, 2001; Lopez, 2008; Hethcote, 2000; Jackson and Rogers, 2007), however, the transitions of the states of individuals are dependent upon factors not limited to influence represented by the set

of relationships across the edges. Moreover, our goal in this paper is to derive from first principle an ordinary differential equation (ODE) model that describes the user growth profile in continuous time. Two cases are considered. First, if the social network is a general uncorrelated network, the user growth profile can be described by a set of ODEs. It can be shown theoretically that nodes with high degrees, namely, prospective users having many connections, transit to users at a faster rate. Second, if the social network is a homogeneous uncorrelated network, the user growth model can be represented by a nonlinear first-order ODE, which has a simple closed-form solution.

Next, using appropriate parameter estimation tools, we show that this model permits analysis of the growth profile of user populations of real-world products and services. Here, we analyse a total of 22 historical user-growth (sales) datasets, including online social network services, instant messaging services, online payment services, video game console sales, automobile sales, mobile apps, microblogging subscribers, and so on. In particular, we show that our user growth equation fits these historical growth data, and in each case, the key parameters of the model, namely, market size and growth rates, can be estimated from the given historical growth data. Thus, a very useful application of the model can be conceived. We find that the average magnitude of micro-level word-of-mouth component of the growth rate of 21 different products or services is of order about 10^{-3} (i.e., 1 out of 1000 transits to become user of the product per day), while the personal-choice component is about 10^{-5} . Here, we borrow the concept of settling time from control theory (Phillips and Habor, 1995), and show that if a product or service succeeds to acquire a steady (mature) user population within a few years (5 to 10 years), the word-of-mouth component of growth rate should range from $10^{-3.6}$ to $10^{-2.8}$, which is consistent with the estimation from the real data. This result clearly suggests that a successful product or service with a growth span (the time taken to reach a mature user population) between 5 to 10 years must have a word-of-mouth component of the growth rate larger than a certain threshold (here, numerically 10^{-4}), and its significance begins to saturate when it increases beyond $10^{-2.8}$. Furthermore, for a heterogeneous network, individuals can be roughly divided into two groups: one contains the majority who have relatively few friends, and the other includes celebrities and very popular people who have numerous friends. In this case, the dynamic growth can be model by a second-order growth model. We apply the second-order growth model in two specific cases, namely, the historical data of the number of registered users of Facebook and the donation collected by ALS Association during the Ice Bucket Challenge in July-August 2014. We find that individuals in the group of celebrities have about 7.14 times more friends (connections) than those in the ordinary group, and the group of celebrities becomes Facebook users at a much faster rate than the ordinary group, which is in agreement with the historical growth data of Facebook and the analytical results.

2 User Growth Model

Consider a network $G = (V, E)$ of users and prospective users of a launched product or service, where V and E denote the sets of nodes and edges, respectively. A social network can be modelled as a dynamic network in which each node assumes one or more possible states and may transit from one state to another as time elapses (Pastor-Satorras



centerline
 Fig. 1. Transition network model. A prospective user may become user under peer influence. A prospective user may also make his own independent decision to become user.

and Vespignani, 2001; Buldyrev *et al.*, 2010; González-Bailón *et al.*, 2011). Here, each node represents an individual who may assume one of two possible states: U and P . A node in state U corresponds to a user of the product or service, whereas a node in state P represents a prospective user who may become a user at a future time. Here, a link between two nodes indicates that the two individuals know each other, e.g., being friends, colleagues, relatives, etc. In our model, a prospective user transits into a user according to two rules, as illustrated in Figure 1. A detailed description of the two rules is as follows:

- *Influenced Transition (Word of Mouth)*: A prospective user may be positively informed about a product or service by other users whom the prospective user knows. The result of peer influence may lead to transition of the prospective user into a user (Goldenberg *et al.*, 2001). In the network context, the effect of a connection between a node in state P and a node in state U is that the node in state P may transit to state U with a probability c_1 . The upper part of Figure 1 illustrates this transition, which can be represented by the following *transition channel*:

$$T_1 : (P-U) \xrightarrow{c_1} (U-U), \quad (1)$$

where T_1 denotes a transition channel, and “—” denotes a connection between two nodes. Here, $P-U$ represents a node in state P being connected with a node in state U . The arrow indicates a transition direction, c_1 being the stochastic rate of transition

for channel T_1 . Also, $(P \rightarrow U)$ denotes the set of prospective transition links and $(U \rightarrow U)$ is the set of resulting links after transition.

- *Self Transition (Personal Choice)*: In reality, prospective users are often informed about a product or service through advertisements, sale promotions or personal research (Nelson, 1974), and the decision to use a product or service is a pure personal choice. The transition of a node from state P to U is thus a self transition, which is independent of the nodes in state U that are connected to it. Suppose the transition probability is c_2 . The right side of Figure 1 illustrates this self transition, and the transition channel can be written as

$$T_2 : (P) \xrightarrow{c_2} (U). \quad (2)$$

where T_2 denotes this self transition channel, (P) is the set of prospective transition links (in this case just the set of nodes in state P) and (U) is the set of resulting links after transition.

Here, we assume that transition channels T_μ ($\mu = 1, 2$) are independent and all transition links of transition channels T_μ is homogeneous and exclusive. Then, each prospective transition link of transition channel T_μ has the same transition rate c_μ . Therefore, $c_\mu \Delta t$ is the probability that a prospective transition link of channel T_μ at time t will make a transition in the next infinitesimal time interval $(t, t + \Delta t)$. For simplicity, we assume that c_μ is constant, but it can be made time-varying with no significant effect on the result of the analysis.

Applying the mean-field approach, in a social network $G = (V, E)$ of users and prospective users of a product or service, we first assume the following:

- There are N nodes in this network;
- Each node has a degree $v_i \in \{k_1, k_2, \dots, k_I\}$, where $k_1 < k_2 < \dots < k_I$ and $I \leq N$. Here, k_i represents the number of connections from an individual to others;
- The number of nodes with degree k_i is N_i .

Hence, $N = \sum_{i=1}^I N_i$ and the average node degree $\langle k \rangle = \sum_{i=1}^I k_i N_i / N$. Suppose at time t , the number of users with degree k_i is n_i . An individual $i \in N$ can only exist in two states: U and P .

Assume that this network is uncorrelated, namely, the probability that a link points to a node with s connections is equal to $sp(s)/\langle k \rangle$ (Pastor-Satorras and Vespignani, 2001), where $p(s)$ is the probability of a node having degree s . Then, applying the mean-field approach, we can readily derive the following ***I*th-order growth equation** (details provided in the Supplementary Material):

$$\frac{\partial x_i}{\partial t} = \frac{c_1 k_i}{N \langle k \rangle} \left(N_i \sum_{j=1}^I k_j x_j - x_i \sum_{j=1}^I k_j x_j (1 + \delta_{ij}) \right) + c_2 (N_i - x_i), \quad (i = 1, 2, \dots, I), \quad (3)$$

where $x_i \triangleq E[n_i]$ and $\delta_{ij} = \frac{\text{cov}[n_i, n_j]}{E[n_i, n_j]}$. For a very large network, $\delta_{ij} \ll 1$ generally holds, leading to

$$\frac{\partial x_i}{\partial t} = \frac{c_1 k_i}{N \langle k \rangle} (N_i - x_i) \sum_{j=1}^I k_j x_j + c_2 (N_i - x_i). \quad (4)$$

If the network is homogenous, namely, $k_1 \approx k_2 \approx \dots \approx k_I \approx \langle k \rangle$, we can treat $I = 1$, and we get the corresponding **first-order growth model** as:

$$\dot{x}(t) = \underbrace{\frac{c_1 \langle k \rangle}{N} (Nx - (1 + \delta(t))x^2)}_{\text{Word of mouth}} + \underbrace{c_2(N - x)}_{\text{Personal choice}}, \quad (5)$$

where $\delta(t) = \text{var}[X(t)]/E[X(t)]^2$ with $X(t) = \sum_{j=1}^I n_j$. With $\delta(t) \ll 1$, we have the following **simplified version of the first-order growth model**:

$$\dot{x}(t) = \left(\underbrace{\frac{c_1 \langle k \rangle}{N} x}_{\text{Word of mouth}} + \underbrace{c_2}_{\text{Personal choice}} \right) \times \underbrace{(N - x)}_{\text{Market size}} \quad (6)$$

The first-order growth model allows convenient interpretations of the physical meanings of the involving parameters, i.e.,

- $\langle k \rangle$: average number of links connecting people in the entire community;
- N : network size or total number of prospective users;
- c_1 : stochastic rate that determines how likely a prospective user transits into a user under the influence of other users connected to him (i.e., word of mouth);
- c_2 : stochastic rate that determines how likely a prospective user transits into a user on his own accord (i.e., personal choice).

Note that the term $c_1 \langle k \rangle / N$ can be treated as the macro-level rate that describes the speed of prospective users transiting into users through peer influence, whereas c_1 is the micro-level stochastic rate. It is readily shown that the influenced transition is dependent on the network parameters, namely, the network size N and the average node degree $\langle k \rangle$. If the user network has a large average node degree $\langle k \rangle$, i.e., individuals have more friends on average, the term $\frac{c_1 \langle k \rangle}{N}$ will be larger and the user growth will be more dependent on peer influence. However, the larger the network size, the smaller the macro-level word-of-mouth component of the transition rate $\frac{c_1 \langle k \rangle}{N}$. Furthermore, the results based on real data listed in Table 2 support this finding. To avoid confusion, we use the term $c_1 \langle k \rangle$ to reflect the effect of peer influence at the micro-level.

For algebraic convenience, we may rewrite the above first-order growth equation as $\dot{x}(t) = k_0 + k_1 x - k_2 x^2$ by putting $k_0 = c_2 N$, $k_1 = c_1 \langle k \rangle - c_2$ and $k_2 = \frac{c_1 \langle k \rangle}{N} (1 + \delta)$. A closed-form solution can also be found for this equation, i.e.,

$$x(t) = a + \frac{b}{c + d e^{-bt}}, \quad (7)$$

where $a \triangleq \frac{k_1 - b}{2k_2}$, $b \triangleq \sqrt{k_1^2 + 4k_0 k_2}$, $c \triangleq k_2$ and $d = \frac{b}{x_0 - a} - c$ and x_0 is the number of initial users.

3 Analysis of Datasets

A total of 22 sets of user growth data for a range of products and services, including online social networking services, instant messaging services, online payment services, microblogging websites, video game consoles, automobile, and mobile apps, are analysed

against the growth equation. These data are publicly available and include Facebook, LinkedIn, Sina Weibo (a popular microblogging website in China), Tencent Weibo (a popular microblogging website in China), XBOX, Playstation 2, Nintendo GameCube, Game Boy Advance SP, Tencent QQ (online messaging service), Line (a popular instant messaging app in Japan), WeChat (a popular instant messaging app in China), Kakao Talk (a popular instant messaging app in Korea), App store, Google Play, electric vehicles, US passenger automobile sales in China, Alipay, Evernote (a software for note taking and archiving), US hospital accounts on YouTube and global internet users.

3.1 Relationship Between the Measured Data and User Growth Process

These datasets provide the number of registered accounts (subscribers) for online services, or the sales volume of particular products. Strictly speaking, these data are not exactly the number of users $x_i(t)$ defined earlier, but are highly indicative of $x_i(t)$ for the products or services being analysed. We can modify the model to enhance its practicality. For example, a user of Playstation 2 may own more than one unit of the game console. Suppose the cumulative number of units already sold to an individual until time t is $M(t)$. Then, the probability that one user owns m units of the game console when the total number of users with degree k_i is $x_i(t) = n_i$ is given by $P(M(t) = m | x_i(t) = n_i)$. Thus, the total sales volume at time t is

$$\sum_{m=1}^{\infty} mP(M(t) = m | x_i(t) = n_i) \times n_i = n_i \left(\sum_{m=1}^{\infty} mP(M(t) = m | x(t) = n_i) \right). \quad (8)$$

Then, on average, one player will buy $\bar{M}_{n_i}(t) = \sum_{m=1}^{\infty} mP(M(t) = m | x(t) = n_i)$ units. Note that in reality, the number of users n_i is very large, such as in millions. So, it is reasonable to assume that $P(M(t) = m | x(t) = n_i) \approx \lim_{n_i \rightarrow \infty} P(M(t) = m | x(t) = n_i) = P_m$ is constant. Hence, $\bar{M}_{n_i}(t) \approx \lim_{n_i \rightarrow \infty} \bar{M}_{n_i}(t) = \beta$ and we have the conclusion

$$\bar{x}_i(t) \approx \bar{M}_{\infty}(t) \cdot n_i \Rightarrow \bar{x}_i(t) \approx \beta x(t), \quad (9)$$

where $\bar{x}(t)$ is the cumulative number of units already sold to a group of individuals with degree k_i and $\beta \geq 1$ is a constant. Combining (9) with the user-growth model (4), we have

$$\frac{\partial \bar{x}_i}{\partial t} = \frac{c_1 k_i}{\beta N \langle k \rangle} \left(\beta N_i \sum_{j=1}^I k_j \bar{x}_j - X_i \sum_{j=1}^I k_j \bar{x}_j (1 + \delta_{ij}) \right) + c_2 (\beta N_i - \bar{x}_i). \quad (10)$$

This growth equation has exactly the same form as Eq. (4) with a market size βN . We therefore conclude that the user growth equation presented in the paper is valid for cases where a user may own multiple units of the product.

3.2 Application Examples of First/Second-Order Growth Models

The user growth equation (6) is used to fit real datasets. The model has a set of intrinsic parameters which can be estimated via constrained nonlinear programming (NLP), with the objective of finding an estimated trajectory of growth that fits the measured data (Moles *et al.*, 2003; Mendes and Kell, 1998). The inverse method is used to find a feasible set of

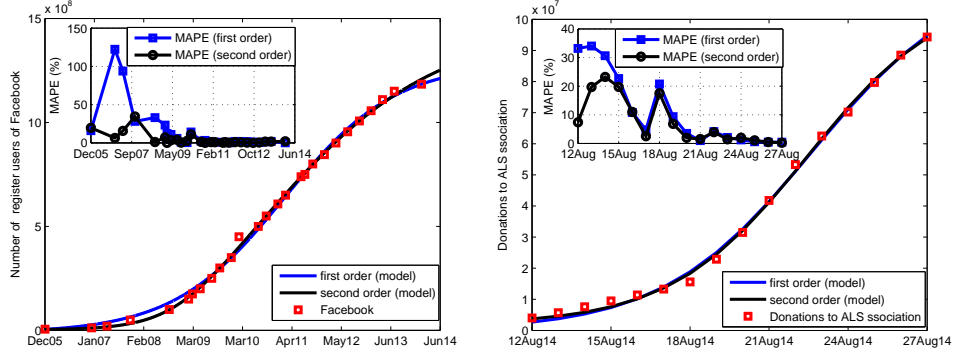


Fig. 2. Model generated growth trajectories by the first-order model (blue solid curves), the second-order model (black solid line) and historical data (squares) for registered users of Facebook (left panel) and donations to ALS Association during the Ice Bucket Challenge in July–August 2014 (right panel); MAPE of the prediction data generated by the first-order model and the second-order model (inset).

parameters (Zhan and Yeung, 2011). Finally, we generate the estimated trajectory with the identified parameter.

First, we compare the fitting ability of Second and First order growth model to test how the fitness of the data depends on the order of the growth model used. We use Mean Absolute Percentage Error (MAPE) as another criterion for evaluating the performance. The MAPE is defined as

$$\text{MAPE}(t_i) = \frac{|\tilde{x}(t_i) - x(t_i)|}{x(t_i)} \times 100\%, \quad (11)$$

where $\tilde{x}(t_i)$ is the estimated number of users at time t_i . Here, we use Facebook as an example. Figure 2 shows that the MAPE of the first-order growth model gives a good fitting of the measured data after 2008. However, in the earlier period (before 2008), the first-order user growth model cannot capture the user growth dynamics accurately. For instance, in January 2007, the MAPE is almost 130%. Now, let us use the second-order user growth model ($I = 2$) to fit the data, i.e.,

$$\begin{cases} \dot{x}_1(t) = \frac{c_1 k_1}{N(k)} (N_1 - x_1) (k_1 x_1 + k_2 x_2) + c_2 (N_1 - x_1), \\ \dot{x}_2(t) = \frac{c_1 k_2}{N(k)} (N_2 - x_2) (k_1 x_1 + k_2 x_2) + c_2 (N_2 - x_2). \end{cases} \quad (12)$$

In the second-order user growth model, all the Facebook users can be roughly divided into two groups: one group is the majority (ordinary) group of people who have relatively few friends, i.e., k_1 is small; the other group includes celebrities and very popular people who have numerous friends, i.e., k_2 is large. Also, N_1 and N_2 are the number of people in the first and second groups, respectively. The estimated growth trajectory of Facebook users using (12) and the corresponding MAPE are given in Figure 2. From these results, we see that the second-order user growth model can capture the real data more accurately for the entire timeframe. Furthermore, the results of parameter identification reveal that $\frac{k_2}{k_1} \approx 7.14$, $N_1 = 1.39 \times 10^9$ and $N_2 = 7.05 \times 10^8$, i.e., the size of the ordinary group is about 10^9 while the size of celebrities and popular people is 7×10^8 . However, one popular person may have

Table 1. Estimated model parameters from fitting of historical datasets

Data	Market size (N_1)	Market size (N_2)	Micro-level growth rate by word of mouth ($c_1 \langle k_1 \rangle$)	Micro-level growth rate by word of mouth ($c_1 \langle k_2 \rangle$)	Growth rate by personal choice (c_2)	k_2/k_1
Facebook	1.39×10^9	7.09×10^8	1.42×10^{-3}	1.02×10^{-2}	1.32×10^{-6}	7.14
Donations (Ice Bucket Challenge)	9.43×10^4	1.09×10^8	1.799×10^{-1}	3.988×10^{-1}	4.52×10^{-21}	2.21

7.14 times more friends than an average ordinary person. The estimated parameter seems reasonable. As another example, the growth of donations to ALS (Amyotrophic Lateral Sclerosis) Association during the Ice Bucket Challenge, which went viral on social media during July to August 2014, also supports our findings. The estimated parameters of the second-order user growth model are given in Table 1.

Consider the I th-order growth equation (10) and let z_j be the number of remaining prospective users of degree k_j . Thus, $z_j = N_j - x_j$ ($j = 1, 2, \dots, I$). At the initial time, the number of prospective users of degree k_i is $z_{i,0}$. Then, we may state that if the growth of the user population of a product or service is described by the I th-order model (4), the dynamic profile of remaining prospective users of degree k_i is given by

$$z_i(t) = z_{i,0} (g(t))^{c_1 k_i} e^{-c_2 t}, \tag{13}$$

where $g(t) = e^{\frac{\Psi(t)}{N \langle k \rangle} - t}$ and $\Psi(t) = \int_0^t \sum_{j=1}^I k_j z_j(s) ds$. The derivation of (13) is given in the Appendix.

Now, applying (13), the number of prospective users of Facebook follows:

$$\begin{cases} \tilde{z}_1(t) \approx (g(t))^{c_1 k_1} e^{-c_2 t}, \\ \tilde{z}_2(t) \approx (g(t))^{c_1 k_2} e^{-c_2 t} \end{cases} \tag{14}$$

where $0 < g(t) < 1$ and $0 \leq \tilde{z}_i(t) = z_i(t)/N_i \leq 1$. Hence, we have $\log(\tilde{z}_2(t)/\tilde{z}_1(t)) \approx 7$, i.e., the group of celebrities becomes Facebook users at a much faster rate than the ordinary group. This result reveals some key features of Facebook growth:

- At the beginning, only ordinary users with small k_1 signed up. The growth was slow before 2008.
- As time elapses, the group of celebrities and popular people signed up. As this group has more friends (a large k_2), the word-of-mouth transition became significant and the number of sign-ups increased drastically from 2008 to 2014.
- From 2014, as most people in the popular group and their friends have signed up as Facebook users, the word-of-mouth influence began to subside, and the growth slowed down.

This phenomenon is in perfect agreement with our analytical results.

Equation (14) shows that the group of prospective users with more links will transit into users at a faster rate. A similar phenomenon has been observed in the long-run behaviour of the model by Paster-Satorras and Vespignani (2001). In the real world, a new product or activity exhibits very rapid growth if it is promoted by celebrities having a lot of connections. Our analytical result coincides with the spreading phenomenon observed in the growth of users of a product or participants of an activity, such as the Ice Bucket Challenge in 2014.

Theoretically, a higher-order growth model can capture the real data more accurately. However, for a higher order growth model, the parameter estimation problem has to be formulated into a constraint nonlinear programming problem (NLP) with differential-algebraic constraints, which is often multimodal (non-convex) and has many local minima. In a noisy environment, the problem becomes more difficult to tackle. Moreover, there is no closed form solution for a higher order growth model. Hence, an ODE solver has to be used during the process of identification and has to be executed thousands, even millions of times. The computational time spent on finding a reliable optimisation solution is quite excessive (typically days or weeks using ordinary computers). However, for parameter estimation of the first-order growth model, we are privileged by the existence of a closed form solution. The cost function has only algebraic equation constraints and does not require an ODE solver. Our test shows that the computational time for identifying a second-order model is more than 100 times more than the computational time for identifying the first-order model. Hence, in order to strike a balance between computational time and model accuracy, we use the first-order growth model for fitting the data and analysis.

3.3 Fitting 21 Real Data by First-Order Growth Model

Here, we use the Normalised Relative Mean Square Error (NRMSE) (Shcherbakov *et al.*, 2013) as a criterion to evaluate the first-order growth model. The NRMSEs of the 21 products are given in Table 2, which shows that the first-order growth model can capture the measured data accurately. The trajectories generated by the user growth equation are also shown in Figures 3 and 4. Historical user growth data are plotted in squares, and growth trajectories generated by our model are plotted with solid curves. Note that the 21 products or services were launched by different enterprises in different countries, belonging to different markets and attracting users from different communities. Some datasets contain over a decade of user growth information, such as Tencent QQ, Facebook and etc. In the past decade, many factors have contributed to the business growth of online products and services. However, their user growth profiles are all governed by a growth equation.

The estimated parameters of the 21 products or services are given in Table 2. Let us take a look at $c_1 \langle k \rangle$ and c_2 , which are the micro-level growth rate corresponding to influenced transition (word of mouth) and personal choice rate, respectively. For example, the $c_1 \langle k \rangle$ of Tencent QQ is 1.51×10^{-3} . Here, we assume $\langle k \rangle = 10$, which means that, on average, one QQ (prospective) user connects with 10 other (prospective) users. Then, $c_1 = \frac{c_1 \langle k \rangle}{10} = 1.51 \times 10^{-4}/\text{day}$, which shows that, on average, 1.51 out of 10 thousand prospective QQ users will transit into a user each day by word of mouth.

Table 2. Estimated model parameters from fitting of historical datasets. Value is fractional number of transitions to users per day, e.g., $c_1 = 2.25 \times 10^{-3}$ means that 2.25 out of 1000 prospective users become users per day.

Data	Market size (N)	Macro growth rate by word of mouth ($\frac{c_1 \langle k \rangle}{N}$)	Micro growth rate by word of mouth ($c_1 \langle k \rangle$)	Growth rate by personal choice (c_2)	NRMSE (%)
Facebook	1.29×10^9	1.77×10^{-12}	2.27×10^{-3}	2.25×10^{-5}	5.22
LinkedIn	4.42×10^8	3.63×10^{-12}	1.60×10^{-3}	7.83×10^{-6}	4.42
Tencent QQ	2.70×10^8	5.60×10^{-12}	1.51×10^{-3}	9.33×10^{-13}	11.6
Evernote	1.35×10^8	2.40×10^{-11}	3.23×10^{-3}	2.31×10^{-6}	5.09
Line	3.87×10^8	1.29×10^{-11}	4.99×10^{-3}	1.44×10^{-4}	3.63
Kakao Talk	3.86×10^8	4.03×10^{-12}	1.56×10^{-3}	1.26×10^{-4}	3.27
Playstation 2	1.72×10^8	2.67×10^{-12}	4.60×10^{-4}	3.19×10^{-4}	3.80
Gameboy	4.18×10^7	4.66×10^{-11}	2.04×10^{-3}	8.11×10^{-4}	6.93
Advantage SP					
XBOX	2.98×10^7	4.52×10^{-11}	1.34×10^{-3}	4.36×10^{-4}	8.18
Nintendo	2.22×10^7	1.09×10^{-10}	2.44×10^{-3}	1.27×10^{-4}	9.86
GameCube					
Available iOS apps	1.66×10^6	8.67×10^{-10}	1.44×10^{-3}	1.24×10^{-4}	8.72
Downloaded iOS apps	1.02×10^{11}	2.32×10^{-14}	2.36×10^{-3}	3.48×10^{-5}	7.33
Available Android apps	1.36×10^6	2.14×10^{-9}	2.91×10^{-3}	1.09×10^{-4}	5.50
Sales of US brand cars in China	3.32×10^7	2.27×10^{-11}	7.56×10^{-4}	1.58×10^{-5}	2.48
Sales of plug-in vehicles in US	6.73×10^6	1.63×10^{-10}	1.10×10^{-3}	8.13×10^{-5}	3.84
US hospital accounts on Youtube	1.39×10^2	4.30×10^{-5}	5.93×10^{-3}	1.38×10^{-4}	4.62
Wechat	6.16×10^8	8.99×10^{-12}	5.53×10^{-3}	8.76×10^{-5}	2.54
Sina Weibo	5.95×10^8	7.06×10^{-12}	4.19×10^{-3}	1.23×10^{-4}	9.80
Tencent Weibo	3.63×10^8	2.36×10^{-11}	8.57×10^{-3}	2.37×10^{-25}	12.4
Alipay	4.24×10^8	7.38×10^{-12}	3.12×10^{-3}	3.01×10^{-26}	7.20
Internet users	4.93×10^9	8.91×10^{-12}	4.39×10^{-4}	2.42×10^{-5}	6.47

Furthermore, comparison of $c_1 \langle k \rangle$ and c_2 can provide crucial information about the relative importance of peer influence (purchase incentivised by word of mouth) and promotional efforts (purchase based purely on personal decision of buyers). The model generated from Tencent QQ, for example, has negligible c_2 compared to $c_1 \langle k \rangle$, which clearly suggests the importance of word of mouth and peer influence on the prospective users. On the other hand, the decision to buy a Playstation 2 is more a matter of personal choice, which is likely a result of aggressive promotional efforts, as suggested by the relatively larger c_2 compared to $c_1 \langle k \rangle$. Here, we show how these parameters determine the shape of the growth curve. From the closed form solution of the first-order model (7), we know that

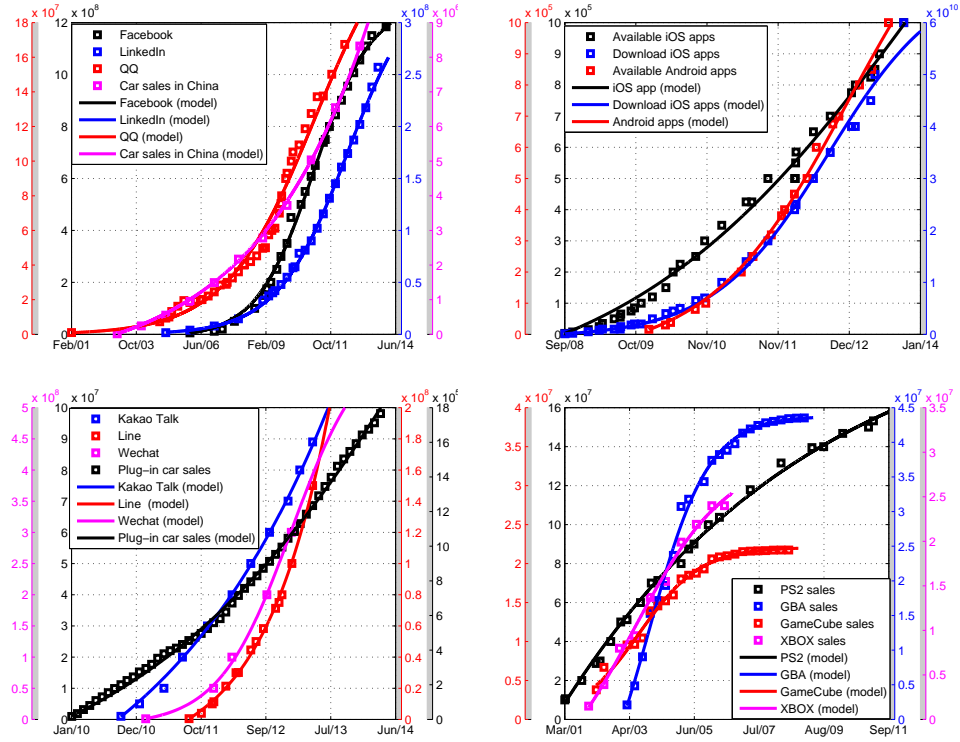


Fig. 3. Model generated growth trajectories (solid curves) and historical datasets (squares) for register users of Facebook, LinkedIn, QQ and sales of US brand cars in China (upper left panel); available iOS apps, downloaded iOS apps and available Android apps (upper right panel); register users of Kakao Talk, Line, WeChat and sales of plug-in vehicles in US (lower left panel); sales of Playstation 2, Gameboy Advantage, GameCube and XBOX (lower right panel).

$x'(t)$ is always positive and $x(t)$ is monotonically increasing. The shape of the curve is related to the second derivative of $x(t)$, which is $x''(t) = \frac{b^3 de^{-bt}(de^{-bt}-c)}{(c+de^{-bt})^3}$. Since $b^3 de^{-bt} > 0$ and $(c+de^{-bt})^3 > 0$, the sign of $x''(t)$ is decided by the term $(de^{-bt}-c)$, which is monotonically decreasing. Here, we may categorise the shape of growth curves into two general classes:

- S-shape curve: In this case, $x''(t_0) \geq 0$, then, $x''(t) < 0$ with $t > \frac{\log d - \log c}{b}$. The final value is $x''(+\infty) = 0$.
- Concave curve: In this case, $x''(t_0) < 0$ and $x''(t) < 0$ at all times. The final value is $x''(+\infty) = 0$.

Hence, the shape of growth curve is determined by $x''(t_0)$. Without loss of generality, we can set the initial time $t_0 = 0$. Then, we have $de^{-bt_0} - c = d - c$. From the definitions of a , b , c , d , k_1 and k_2 , we can write

$$\begin{cases} x''(t_0) \geq 0 \Leftrightarrow N \frac{c_1 \langle k \rangle - c_2}{2c_2(1+\delta)} \geq x_0, & \text{S-shape} \\ x''(t_0) < 0 \Leftrightarrow N \frac{c_1 \langle k \rangle - c_2}{2c_2(1+\delta)} < x_0, & \text{Concave} \end{cases}$$

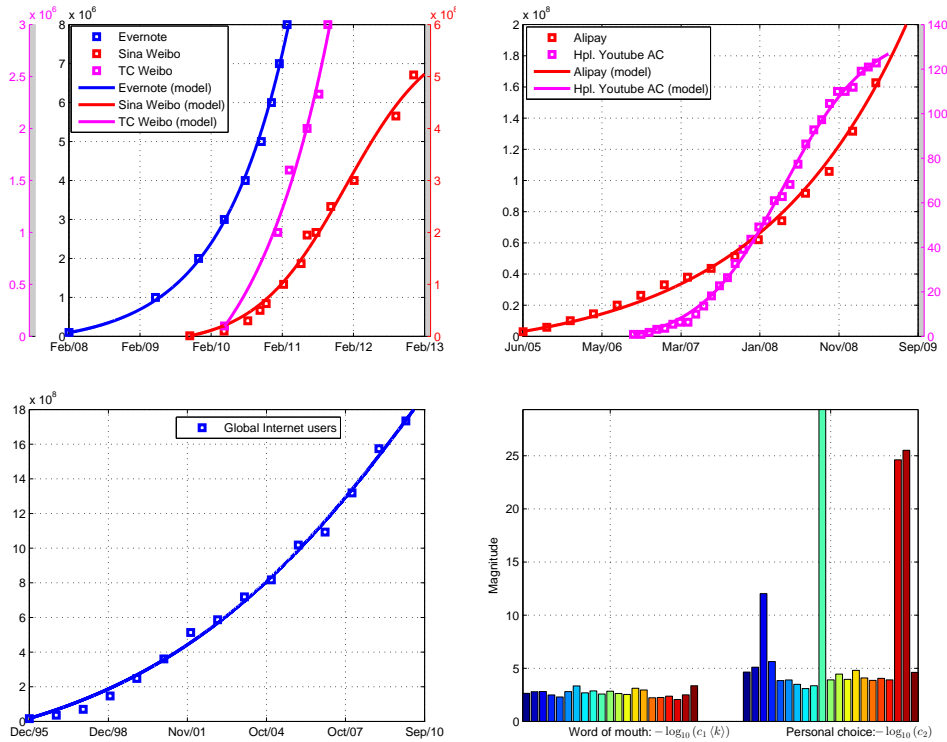


Fig. 4. Model generated growth trajectories (solid curves) and historical datasets (squares) for registered users of Evernote, Sino Weibo and Tencent Weibo (upper left panel); registered users of Alipay and YouTube accounts of US hospitals (upper right panel); global internet users (lower left panel); the micro-level word-of-mouth transition rate and personal-choice rate of the 21 markets (lower right panel).

Apply this result to growth data, we have the following observations:

- When $c_1 \langle k \rangle \gg c_2$ and x_0 is small, we have $x''(t_0) > 0$. In this case, the number of initial users is small and the product is mostly promoted through word-of-mouth of users. For instance, Tencent QQ, which is an instant messaging software service developed by Chinese company Tencent Holdings Limited and released in 1999, had a market value less than 50,000 USD and about 100,000 users in 2000. Back then, Tencent could hardly afford extensive advertisement. At the beginning, QQ was mostly promoted by word-of-mouth in high schools and colleges. Later users of QQ spread all over China. The growth curve is an S-shape curve. Another typical example is Facebook, which was launched by a small company (x_0 being small) having very limited fund for promotion (c_2 being small). However, this product is a revolutionary product and gains reputation rapidly ($c_1 \langle k \rangle$ being large). The growth curve is also S-shape as consistently described by the model.
- When $c_1 \langle k \rangle \approx c_2$ and x_0 is large, we have $x''(t_0) < 0$. In this case, personal choice plays an important role. For instance, the PS2 game console, which is Sony's second installment in the PlayStation Series, made extensive advertisement and overwhelming news before it was launched. Sony, which is a giant companies, could afford

very substantial promotional cost to advertise its PS2 all over the world to make PS2 the best-selling video game console in history. In this case, the growth curve is a concave curve. Other examples are Gameboy, XBOX, etc. Hence, a product from a financially well positioned company could attract a lot of attention before it is launched (x_0 being large). The company invests substantially to promote it (c_2 being large). The growth curve is concave.

Furthermore, let us take a look at the estimated user size N of two video game consoles, XBOX (Microsoft) and PS2 (Sony). Our results show that the estimated PS2 potential user population is almost 6 times larger than that for XBOX. In fact, Microsoft first set into the video console market and developed XBOX in 2001, but at that time, the video game console market was already monopolised by Sony. XBOX did not manage to change the market trend, and PS2 remained one of most popular video game consoles. The estimated potential user populations of XBOX and PS2 are in full agreement with the real market data.

3.4 Analysis of Growth Span of Products

The results of parameter identification support our analytical findings. Specifically, the macro-level word-of-mouth component of the growth rate is dependent on the network structure. For a larger market size N , the macro-level word-of-mouth component of the growth rate is smaller (see Table 2, column 3). For instance, the iOS apps have a large market size of about $N = 1.02 \times 10^{11}$, while their macro-level user growth rate is the smallest (about 2.32×10^{-14}). However, the magnitude of the micro-level word-of-mouth component of the growth rate $c_1 \langle k \rangle$ is of order 10^{-3} (18 products), and 10^{-4} (3 products). This finding is noteworthy, as the 21 products or services belong to different markets, such as video games, automobile, smartphones, etc. The lower right panel of Figure 4 shows the 21 values of $-\log_{10}(c_1 \langle k \rangle)$ and $-\log_{10}(c_2)$. Note that the larger the value of $-\log_{10}(c_1 \langle k \rangle)$ or $-\log_{10}(c_2)$, the smaller the value of $c_1 \langle k \rangle$ or c_2 . This indicates that the micro-level word-of-mouth component of the growth rate varies very little, while the personal-choice component varies widely from product to product.

A question arises at this point. Why do the 21 datasets have such consistent (similar) word-of-mouth component of the growth rate $c_1 \langle k \rangle$ (all being ranged from $10^{-2.8}$ to 10^{-4})? Here, we borrow the concept of settling time from control theory, which is used for analysis of the dynamic property of a system (Phillips and Habor, 1995). Specifically, we define the *growth span* of a product or service as the time required for its user population to grow from a small initial value $x(t_0) = \alpha N$ at $t = t_0$ to reach its final mature value, namely, $x(t_e) = \beta N$ at $t = t_e$, within a small tolerance range (e.g., 5%) around $x(t_e)$ for the first time. The time duration $T_s = t_e - t_0$ is the growth span of a product.

Thus, the growth span is basically the settling time commonly used in control theory, and in our case, if the user network is homogeneous and uncorrelated network, and assuming $\delta \approx 0$, the growth span of a product is given by

$$T_s(c_1 \langle k \rangle, c_2) \approx \frac{\log(1 - \alpha) + \log(\beta c_1 \langle k \rangle + c_2) - \log(1 - \beta) - \log(\alpha c_1 \langle k \rangle + c_2)}{c_1 \langle k \rangle + c_2} \quad (15)$$

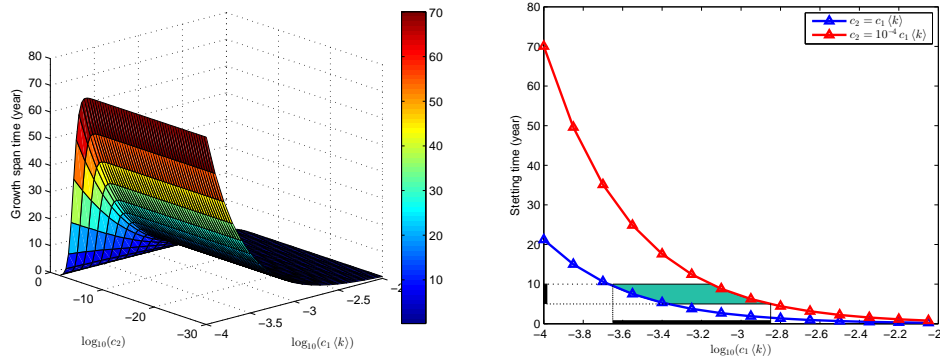


Fig. 5. Growth span of products versus $c_1 \langle k \rangle \in [10^{-4} \ 10^{-2}]$ and $c_2 \in [10^{-30} \ 10^{-2}]$ (left panel). Lifespan of products for $10^{-4} \leq c_1 \langle k \rangle \leq 10^{-2}$ and $10^{-4}c_1 \langle k \rangle \leq c_2 \leq c_1 \langle k \rangle$. Green area is the desirable region for a product with growth span of 5 to 10 years (right panel).

The derivation of the above result is given in the Appendix. In this paper, we set $\alpha = 0.05$ and $\beta = 0.95$.

Figure 5 shows the growth span versus $c_1 \langle k \rangle$ and c_2 . Here, we analyze a class of products or services, of which the growth span is about 5 to 10 years (i.e., $5 \leq T_s \leq 10$). The personal-choice component of the growth rate is given as $c_2 = \lambda c_1 \langle k \rangle$, where λ is the ratio of the personal-choice component compared to the word-of-mouth component of the growth rate. Here, we set $10^{-4} \leq \lambda \leq 1$. In particular, $\lambda = 1$ means that personal choice is as important as word of mouth, while $\lambda = 10^{-4}$ means that the impact of personal choice is negligible. Figure 5 shows that with a constant λ , the smaller the value of c_2 , the longer the lifespan of a product. The green area in the left panel of Figure 5 is the desirable region with lifespan T_s ranging from 5 to 10 years (black segment in the y-axis). Note that under this condition, the suitable range of $c_1 \langle k \rangle$ is about $10^{-3.8}$ to $10^{-2.8}$ (black segment in the x-axis), which agrees with the result shown in column 4 of Table 2 and the lower right panel of Figure 4. Thus, we clearly see that $c_1 \langle k \rangle$ for a successful product or service must exceed 10^{-4} , but its significance saturates which it reaches $10^{-2.8}$.

4 Conclusion

A growth equation for any product, service or participation in an activity is derived rigorously from a network model and consideration of the physical processes of the transitions involved. The final form of this model is simple and appears as a product of several factors which have very clear physical meanings. Specifically, given a set of historical data of the user growth profile, this model allows the prospective market size to be predicted, and at the same time identifies the relative importance of customer service and promotional efforts. These estimated parameters, namely N , $c_1 \langle k \rangle$, and c_2 , generated from historical data, can thus provide very useful information for making decisions in business. From the business perspective, parameter N is effectively the market size, whereas parameters $c_1 \langle k \rangle$ and c_2 are measures of the importance of word of mouth and personal choice, respectively, which translate to effectiveness of customer service and promotional efforts. Furthermore, $(N - x)$ gives the remaining market size which is an important indicator of the growth potential of

the user population of a product or service. We should stress that we present here a basic model, which fits historical growth data and generates parameters of practical importance. There is still plenty of room for further study. For instance, we may improve the model by extending the model to accommodate competing products, over-supply quantities, etc. In terms of parameter estimations, more efficient and reliable optimisation methods will need to be developed for finding the parameters which have important business implications. Using this model, together with sufficient historical data and efficient parameter estimation methods, effective marketing and timing strategies can thus be developed to ensure the success of the products or services in question or the timely initiation and launching of new products or services.

Appendix

Proof of Equation (13)

Let z_j be the number of prospective users of degree k_j , and $z_j \triangleq N_j - x_j$ ($j = 1, 2, \dots, D$). Hence, (4) can be simplified as

$$\frac{\partial z_i}{\partial t} = -z_i \left(\frac{c_1 k_i}{N \langle k \rangle} \sum_{j=1}^I k_j N_j + c_2 \right) + \frac{c_1 k_i}{N \langle k \rangle} z_i \sum_{j=1}^I k_j z_j. \quad (16)$$

Define

$$\begin{aligned} \beta_i &= \frac{c_1 k_i}{N \langle k \rangle}, \\ \gamma_i &= \frac{c_1 k_i}{N \langle k \rangle} \sum_{j=1}^I k_j N_j + c_2 = c_1 k_i + c_2, \end{aligned}$$

Then, (16) can be simplified as

$$\frac{\partial z_i}{\partial t} = -\gamma_i z_i + \beta_i z_i \sum_{j=1}^I k_j z_j. \quad (17)$$

Let $\phi = \sum_{j=1}^I k_j z_j$, $\Psi(t) = \int_0^t \phi(s) ds$ and $X_i(t) = z_i e^{-\beta_i \Psi(t)}$. Then, from (17), we have

$$\frac{\partial X_i(t)}{\partial t} = -\gamma_i z_i e^{-\beta_i \Psi} = -\gamma_i X_i(t). \quad (18)$$

Note that since $\Psi(0) = 0$, $X_i(0) = z_{i,0}$, where $z_{i,0}$ is the initial number of prospective users of degree k_i . Hence, (18) has an analytical solution given by

$$\frac{\partial X_i(t)}{\partial t} = -\gamma_i X_i(t) \Rightarrow X_i(t) = z_{i,0} e^{-(c_1 k_i + c_2)t}. \quad (19)$$

Finally, from (19), we obtain the prospective user profile as

$$z_i(t) = z_{i,0} (g(t))^{c_1 k_i} e^{-c_2 t} \quad (20)$$

where $g(t) = e^{\frac{\Psi(t)}{N \langle k \rangle} - t}$. Q.E.D. \square

Derivation of Growth Span of Products

Without loss of generality, we assume that $t_0 = 0$, then, there exist $x(0) = \alpha N$ and $x(T_s) = \beta N$. From (7),

$$\alpha N = x(0) = a + \frac{b}{c+d}, \quad \text{and} \quad \beta N = x(T_s) = a + \frac{b}{c + d e^{-b T_s}}. \quad (21)$$

As $\delta \approx 0$, from the previous equation, we can derived that

$$Ts(c_1 \langle k \rangle, c_2) = \frac{\log(1 - \alpha) + \log(\beta c_1 \langle k \rangle + c_2) - \log(1 - \beta) - \log(\alpha c_1 \langle k \rangle + c_2)}{c_1 \langle k \rangle + c_2} \quad (22)$$

Q.E.D. \square

Acknowledgement

This work was supported by Hong Kong Polytechnic University Grants G-YBKB and G-YBAT.

References

- Albert, R. and Barabási, A-L. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 (1): 47.
- Barabási, A-L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, 289 (5439): 509–512.
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. and Havlin, S. (2010) Catastrophic cascade of failures in interdependent networks. *Nature*, 464 (7291): 1025–1028.
- Campbell, A. (2013) Word-of-mouth communication and percolation in social networks. *American Economic Review*, 103 (6): 2466–2498.
- Geroski, P. A. (2000) Models of technology diffusion. *Research Policy*, 29 (4): 603–625.
- Goldenberg, J., Libai, B. and Muller, E. (2001) Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12 (3): 211–223.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A. and Moreno, Y. (2011) The dynamics of protest recruitment through an online network. *Scientific Reports*, 1: 197.
- Hethcote, H. W. (2000) The mathematics of infectious diseases. *SIAM Review*, 42 (4): 599–653.
- Jackson, M. O. and Rogers, B. W. (2007) Relating network structure to diffusion properties through stochastic dominance. *B.E. Journal of Theoretical Economics*, 7 (1): 1–16.
- Iribarren, J. L. and Moro, E. (2011) Affinity paths and information diffusion in social networks. *Social Networks*, 33 (2): 134–142.
- Leskovec, J. and Adamic, L. A. and Huberman, B. A. (2007) The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1 (1): 1-46.
- López-Pintado, D. (2008) Diffusion in complex social networks. *Games and Economic Behavior*, 62 (2): 573–590.
- Mendes, P. and Kell, D. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14 (10): 869–883.
- Moles, C. G., Mendes, P. and Banga, J. R. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13 (11): 2467–2474.
- Nelson, P. (1974) Advertising as information *Journal of Political Economy*, 82 (4): 729–754.
- Phillips, C. L. and Habor, R. D. (1995) *Feedback Control Systems*. Simon & Schuster.
- Pastor-Satorras, R. and Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Physical Review Letters*, 86 (14): 3200.
- Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A. and Kamaev, V. A. (2013) A survey of forecast error measures. *World Applied Sciences Journal (Information Technologies in Modern Industry, Education & Society)*, 24: 171–176.
- Sood, A., James, G. M. and Tellis, G. J. (2009) Functional regression: A new model for predicting market penetration of new products. *Marketing Science*, 28 (1):36–51.

Strogatz, S. H. (2001) Exploring complex networks. *Nature*, 410 (8625): 268–276.

Young, H. P. (2009) Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99 (5): 1899–1924.

Zhan, C. and Yeung, L. F. (2011) Parameter estimation in systems biology models using spline approximation. *Bioinformatics*, 5 (1): 14.