Long Short Memory Process: Modeling Growth Dynamics of Microscopic Social Connectivity

Chengxi Zang*
Computer Science Department, Tsinghua University
zangcx13@mails.tsinghua.edu.cn

Peng Cui
Computer Science Department, Tsinghua University
cuip@tsinghua.edu.cn

Christos Faloutsos
Computer Science Department, Carnegie Mellon University
christos@cs.cmu.edu

Wenwu Zhu
Computer Science Department, Tsinghua University
wwzhu@tsinghua.edu.cn

ABSTRACT

How do people make friends dynamically in social networks? What are the temporal patterns for an individual increasing its social connectivity? What are the basic mechanisms governing the formation of these temporal patterns? No matter cyber or physical social systems, their structure and dynamics are mainly driven by the connectivity dynamics of each individual. However, due to the lack of empirical data, little is known about the empirical dynamic patterns of social connectivity at microscopic level, let alone the regularities or models governing these microscopic dynamics.

We examine the detailed growth process of "WeChat", the largest online social network in China, with 300 million users and 4.75 billion links spanning two years. We uncover a wide range of long-term power law growth and short-term bursty growth for the social connectivity of different users. We propose three key ingredients, namely average-effect, multiscale-effect and correlation-effect, which govern the observed growth patterns at microscopic level. As a result, we propose the long short memory process incorporating these ingredients, demonstrating that it successfully reproduces the complex growth patterns observed in the empirical data. By analyzing modeling parameters, we discover statistical regularities underlying the empirical growth dynamics. Our model and discoveries provide a foundation for the microscopic mechanisms of network growth dynamics, potentially leading to implications for prediction, clustering and outlier detection on human dynamics.

KEYWORDS

Human dynamics; Social dynamics; Social network; Power law growth; Burst;

*Tsinghua National Laboratory for Information Science and Technology (TNList)

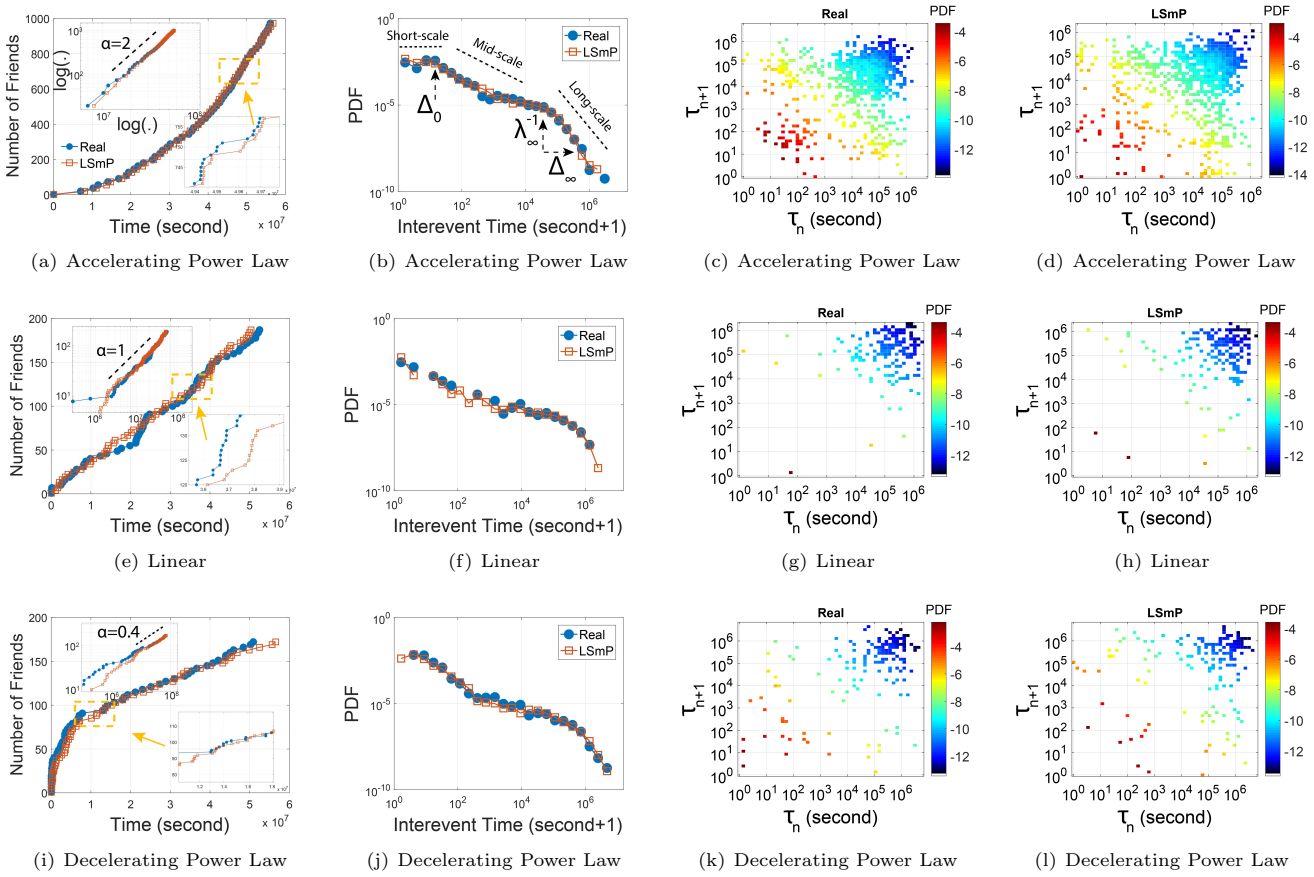
KDD '17, August 13-17, 2017, Halifax, NS, Canada

DOI: 10.1145/3097983.3098055

1 INTRODUCTION

In recent years, much effort has been made to model the growth of social networks, ranging from reproducing the statistical properties of structures [4], to modeling the growth dynamics of nodes and links [25], aiming to enhance our understanding on complex social systems. In microscopic view, the structures and dynamics of complex social systems are mainly driven by the growth dynamics of individual's social connectivity. However, little is known about this microscopic dynamics in reality, mainly due to the lack of large scale datasets recording the detailed social network evolution. Thus, fundamental questions as follows are largely unknown: What are the dynamic growth patterns of an individual's connectivity in social networks? What are the plausible regularities governing these microscopic growth dynamics? Can we unify the connectivity growth dynamics of largely heterogeneous individuals into a simple mathematic model? Answering to these questions not only advances our understanding on the evolution of complex social systems, but also delivers models to capture human dynamics at microscopic level, leading to applications for predicting future growth, clustering typical modes of human dynamics, and detecting outliers.

In literature, the studies on microscopic connectivity dynamics are mainly conducted theoretically. The Barabási-Albert model [4] which generates scale-free network, predicts the scaling relationship between the connectivity and the time with exponent 0.5. The fitness model [5] extends this exponent to a fitness function with range < 1. However, these theoretical hypotheses ignore the ubiquitous densification phenomenon [15], and the empirical growth dynamics may have exponent value ≥ 1. Recently, a paucity of empirical studies are conducted. Ref.[14] finds the rate of connectivity grows with current degree in a mean-field setting, indicating a superlinear growth in average, but they ignore the large heterogeneity among different users. Ref.[2] reports that the rate of connectivity is constant, indicating a linear growth in average, but the result is also from a mean-field setting and the connectivity timestamps are simulated from user behavior logs rather than explicitly recorded. Thus, the growth dynamics of social connectivity of each individual has not been empirically explored or theoretically studied.



(a) Accelerating Power Law (b) Accelerating Power Law (c) Accelerating Power Law (d) Accelerating Power Law

(e) Linear (f) Linear (g) Linear (h) Linear

(i) Decelerating Power Law (j) Decelerating Power Law (k) Decelerating Power Law (l) Decelerating Power Law

Figure 1: ***Microscopic social connectivity features various power law growth in the long term and bursts in the short term. LSmP fits reality well.*** **(a)(e)(i) plot three instances of different growth patterns: the accelerating power law growth, linear growth, and decelerating power law growth, respectively. The same plots on log-log scale are shown in the upper insets, while the lower insets illustrate short-term bursts. Each row describes different aspects of the same empirical dynamics. (b) (f) (j) plot inter-event time (IET) distributions. (c)(g)(k) plot the empirical joint distributions of consecutive IETs, while (d)(h)(l) are the counterparts generated by LSmP. Our model fits the reality well on all the aspects.**

In this paper, we study the detailed evolution traces of WeChat from zero to 300 million users and 4.75 billion friendships spanning two years. The dataset explicitly records when and how a link is established between whom and whom, serving as one of the largest datasets which support this study. From this dataset, we find the growth dynamics of social connectivity exhibit rich complexities which are far beyond our current understanding. Figure 1 showcases the growth dynamics of three individuals over two years, we find: *i)* a wide range of power law growth in the long term, and *ii)* stochastic bursts in the short term. Based on these empirical observations, we propose three ingredients, i.e., average-effect, multiscale-effect, and correlation-effect, which govern the growth dynamics. We then propose a stochastic model, **L**ong-**S**hort-**m**emory **P**rocess (**LSmP**), incorporating these ingredients, to characterize the empirical dynamics. The effectiveness of LSmP is validated in multiple aspects. By analyzing the modeling parameters, we find statistical regularities underlying these growth dynamics. By clustering in the parameter space, we find three typical modes of social connectivity growth in the long term, and two typical modes in the short term. We also illustrate that outliers can be easily detected with clear explanation in the parameter space. In summary, the contributions of this work are as follows:

- **Findings:** The growth dynamics of social connectivity features long-term power law growth and short-term bursts. We find three ingredients, i.e., average-effect, multiscale-effect, and correlation-effect accounting for the power law and bursty growth dynamics.

- **LSmP model:** A stochastic model captures above findings. It is parsimonious, and the parameters have clear physical meanings.
- **Accuracy:** LSmP accurately fits the growth dynamics.
- **Usefulness:** LSmP provides insightful understandings on human dynamics, and implies applications in prediction, clustering, pattern discovery, and outlier detection, etc.

Reproducibility: Our code is open-sourced at github.com/calvin-zcx/LSMP

The outline of the paper is the typical one: survey, proposed method, experiments, and conclusions.

2 RELATED WORK

As the investigated problem is closely related to network evolution and human behavior dynamics, we mainly review the related works in these two fields.

Network evolution. The Barabási-Albert model (BA) [4] predicts the dynamics of individual's connectivity scales against time with exponent $\alpha = 0.5$ ($k(t) \sim t^{1/2}$). The fitness model [5] extends the BA model to heterogeneous connectivity dynamics, with scaling exponent $\alpha < 1$. However, we find empirical connectivity dynamics are more complex: the dynamics with scaling exponent $\alpha >= 1$ are prevalent. Recently, a paucity of empirical studies are conducted. [14] finds the rate of connectivity grows with current degree in a mean-field setting, and ignores the large heterogeneity among different users. [2] reports that the rate of connectivity is constant, but the result is also from a mean-field setting and the connectivity timestamps are simulated from user behavior logs rather than explicitly recorded. [25] examines several large social networks and finds the power law growth for both nodes and links over time, implying that the average connectivity scales with time. In all, how the connectivity dynamics grows at *an individual level* is largely unaddressed.

Human dynamics. Human dynamics often exhibit bursts and heavy-tailed inter-event time (IET) distribution [3]. Priority queue model [23] and modulated Poission model [16] may account for IET's fat-tail nature. Recently, evidences of multiscale nature of IET distribution [27] have been found, including bimodal distribution [24] at long-scale, and quick actions at short-scale [27]. However, they treat the IETs independent identically distributed, ignoring their correlations [12]. Two lines of works try to capture the bursts and correlations: the self-exciting process [7, 11, 20], and the self-feeding process [1, 10, 22]. However, none of them can generate both power law growth with arbitrary exponent and mutiscale IET distribution as we observed in the empirical dataset.

3 PROPOSED METHOD

In this section, we present the ingredients accounting for the long-term power law growth and short-term bursts, and then propose our stochastic model based on these ingredients.

Table 1: Symbols and Definitions

Symbols	Definitions
$n(t)$	Cumulative number of events by time t for a specific user
t_i	The time of event i
τ_i	The interevent time (IET) between event i and $i-1$
$\mathcal{H}_t$	The history of previous events by t, a list of event times
$\lambda(t\|\mathcal{H}_t)$	The hazard function of LSmP
$\Phi_\infty(t)$	The hazard function for long-term memory
λ_∞	The long-term event rate
α	The long-term growth exponent
Δ_∞	The long-term time scale
$\Phi_0(t)$	The hazard function for short-term memory
λ_0	The short-term event rate
θ	The short-term fizzling exponent
Δ_0	The short-term time scale
m	Memory length of short-term memory

3.1 Model Intuition

Here we introduce the intuition and backgrounds to motivate the model design.

Ingredients for power law growth. Zang et al. [25] proposed a dynamic model to capture the growth dynamics of the number of nodes $n(t)$ in social networks: $\frac{dn(t)}{dt} = \frac{\beta}{t^\theta} n(t)(N - n(t))$, where N is the ceiling size and β is the growth rate. It generates a variety of growth patterns, including power law growth as a special case when $\theta = 1$ and $n(t) \ll N$. By defining $\alpha = \beta N$, we get the equation which generates power law growth for $n(t)$:

$$\frac{dn(t)}{dt} = \frac{\alpha}{t} n(t). \tag{1}$$

Thus, when the fizzling exponent $\theta = 1$, this equation indicates the physical meaning at microscopic level: *the rate of making friends* $\frac{dn(t)}{dt}$ *at present is proportional to the long-term average rate* $\frac{n(t)}{t}$, and this long-term memory is kept by the behavior of making $n(t)$ friends so far. We name it **average-effect** which generates the power law growth.

Ingredients for bursts. Human behaviors often exhibit multiscale IET distribution [27] as shown in Fig.1b, taking on flat short-scale, fat-tailed middle-scale, and the exponential-like long-scale. When we i.i.d. sample IETs from a multiscale distribution, a sequence of small value IETs occur in short-scale, followed by large value IET in the middle-scale and long-scale, generating intense activities followed by long vacation, namely, bursts. Thus, the multiscale IET distribution, which captures the short-term memory since last event, generates bursts. We denotes this ingredient as **multiscale-effect**.

Given a Poisson process exhibiting non-bursty growth, if we rearrange the IET sequence as follows: make small value IET followed by smaller IET, and make large value IET followed by larger IET, and then we also get bursty behaviors. Thus, the correlations between IETs can also generate bursts. Figures 1cg&k plot the joint distribution of two consecutive IETs from real data. If the IETs are independent, the points on joint distribution should spread out, while in reality the dots are clustered, e.g., red dots, which support our conjecture of correlated IETs. Indeed, the corresponding growth dynamics in Figs. 1ae&i, along with

the lower insets, exhibit bursts in the short-term. We name this ingredient as **correlation-effect**.

Model framework: temporal point process. The process that a user will add friends at a specific time epoch is a (stochastic) temporal point process. Mathematically, we want to know the time t of the next event, and let $f(t|\mathcal{H}_t)$ be the conditional probability density function of the time of the $(n+1)^{th}$ event conditioned on all the events in history $\mathcal{H}_t$, and its corresponding cumulative distribution function is $F(t|\mathcal{H}_t)$. This temporal point process is intuitively and uniquely specified by the hazard function $\lambda(t|\mathcal{H}_t)$ [8]. The hazard function $\lambda(t|\mathcal{H}_t) = \mathbb{E}[n(dt)|\mathcal{H}_t]/dt$ describes the rate of the occurrence of the event given the history $\mathcal{H}_t = (t_1, ..., t_{n-1}, t_n)$ of previous events, where $n(dt)$ denotes the number of events falling in $[t, t+dt)$ and $n(t)$ stands for the cumulative number of events falling in $(-\infty, t)$. The relationship between hazard function and the conditional probability density function is $\lambda(t|\mathcal{H}_t) = \frac{f(t|\mathcal{H}_t)}{1-F(t|\mathcal{H}_t)}$ and $f(t|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)e^{-\int_{t_n}^{t} \lambda(s|\mathcal{H})_t ds}$.

3.2 The LSmP model

Here we propose the **L**ong-**S**hort-**m**emory **P**rocess (LSmP), which is based fundamentally on the average-effect, multiscale-effect, and correlation-effect to capture the long-term power-law growth and short-term bursts of social connectivity. The hazard function specifying the LSmP is:

$$\begin{aligned}\lambda(t|\mathcal{H}_t) = & \underbrace{\lambda_\infty \alpha (\frac{t}{\Delta_\infty}+1)^{\alpha-1}}_{\text{Long-term memory } \Phi_\infty(t|\mathcal{H}_t)} + \underbrace{\sum_{i=n(t)-m+1}^{n(t)} \lambda_0 (\frac{t-t_i}{\Delta_0}+1)^{-\theta}}_{\text{Short-term memory } \Phi_0(t|\mathcal{H}_t)} \\ & = \Phi_\infty(t|\mathcal{H}_t) \ + \ \Phi_0(t|\mathcal{H}_t)\end{aligned} \tag{2}$$

3.2.1 Long-term memory. We first show the average-effect generates power-law stochastic growth, taking on long-term memory.

LEMMA 3.1. *The average-effect, namely the present hazard rate $\Phi_\infty(t|\mathcal{H}_t)$ is the the mean value of its previous hazard rate from $-\Delta_\infty$, generates power law dynamics with arbitrary exponent α.*

PROOF. We formulate the average-effect as:

$$\Phi_\infty(t|\mathcal{H}_t) = \frac{\alpha}{t+\Delta_\infty} \int_{-\Delta_\infty}^{t} \Phi_\infty(s|\mathcal{H}_t) ds, \tag{3}$$

and replace the integral term by the equation $\int_{-\Delta_\infty}^{t} \Phi_\infty(s) ds = \lambda_\infty \Delta_\infty (\frac{t}{\Delta_\infty}+1)^\alpha$, we get:

$$\Phi_\infty(t|\mathcal{H}_t) = \lambda_\infty \alpha (\frac{t}{\Delta_\infty}+1)^{\alpha-1}. \tag{4}$$

□

The $\Phi_\infty(t|\mathcal{H}_t) = \lambda_\infty \alpha (\frac{t}{\Delta_\infty}+1)^{\alpha-1}$ captures the long-term power law growth dynamics of adding friends in the temporal point process framework, serving as the stochastic counterpart of the dynamic differential equation 1 in Sec. 3.1.

Justification of the model:

Various power law growth. As shown in Sec. 3.1, the α is the average number of new friends introduced by one existing friend in $n(t)$ at time interval $(t, t+dt)$. When $\alpha > 1$, the accelerating power law growth appears, implying a rich-get-richer phenomenon: more friends you have, much more friends you will have in the future. When $\alpha = 1$, $\Phi_\infty(t|\mathcal{H})$ equals to λ_∞, indicating that $n(t)$ is supposed to have a constant growth rate without long-term memory. In contrast, when $\alpha < 1$, $n(t)$ exhibits decelerating power law growth. Three instances are shown in Figs. 1ae&i.

Long-scale of multiscale-effect. The range of long-scale is controlled by λ_∞ and α. When $\alpha = 1$, $\Phi_\infty(t|\mathcal{H}_t)$ degenerates to the Poisson process, which captures the routine activities of adding friends. Specifically, an event is expected to occur with an average inter-event time λ_∞^{-1} with deviation controlled by Δ_∞, as shown in Fig. 1b. When α increases, the exponential mode in the tail part moves towards left, implying a smaller τ and thus an accelerating $n(t)$. In contrast, when a decreases, the second mode of $f(\tau)$ moves to right, implying a larger τ and thus a decelerating $n(t)$.

Long-term correlation of IETs. When $\alpha \neq 1$, the long-term memory appears, taking on long-term IETs' correlations. Indeed, when $\alpha \geq 1$, IETs tend to become shorter and shorter, while when $\alpha \leq 1$, IETs become larger and larger, implying an increasing and a saturated circle of friends, respectively.

3.2.2 Short-term memory. The short-term memory $\Phi_0(t|\mathcal{H}_t)$ captures the multiscale-effect and correlation-effect, which generate bursty dynamics.

LEMMA 3.2. *The $\lambda(t|\mathcal{H}_t)$ generates constant short-scale, power law middle-scale and exponential long-scale when $\alpha = m = \theta = 1$.*

PROOF. The conditional probability density function of the time of the next event, say $(n+1)^{th}$ event, is (refer to Sec. 3.1):

$$\begin{aligned} f(t|\mathcal{H}_t) &= \lambda(t|\mathcal{H}_t) e^{-\int_{t_n}^{t} \lambda(s|\mathcal{H}_t) ds} \\ &= (\Phi_\infty(t|\mathcal{H}_t) + \Phi_0(t|\mathcal{H}_t)) e^{-\int_{t_n}^{t} \Phi_\infty(s|\mathcal{H}_t) ds} e^{-\int_{t_n}^{t} \Phi_0(s|\mathcal{H}_t) ds} \\ &= [\lambda_\infty \alpha (\frac{t}{\Delta_\infty}+1)^{\alpha-1} + \sum_{i=n-m+1}^{n} \lambda_0 (\frac{t-t_i}{\Delta_0}+1)^{-\theta}] \\ &\times e^{-\lambda_\infty \Delta_\infty [(\frac{t}{\Delta_\infty}+1)^\alpha - (\frac{t_n}{\Delta_\infty}+1)^\alpha]} \\ &\times \prod_{i=n-m+1}^{n} e^{-\int_{t_n}^{t} \lambda_0 (\frac{s-t_i}{\Delta_0}+1)^{-\theta} ds} \end{aligned}$$

The condition $\alpha = 1$ excludes the effect of long-term memory. When $m = 1$, the short-term memory is solely determined by the fizzling effect, indicating the independency between any two consecutive events. Thus $\lambda(t|\mathcal{H}_t)$ describes a **renewal process** where $f(t|\mathcal{H}_t) = f(\tau_{n+1}) = f(\tau)$. When $\theta = 1$,

$$\begin{aligned} f(t|\mathcal{H}_t) &= [\lambda_\infty + \lambda_0 (\frac{t-t_n}{\Delta_0}+1)^{-1}] e^{-\lambda_\infty (t-t_n)} (\frac{t-t_n}{\Delta_0}+1)^{-\lambda_0 \Delta_0} \\ &\approx \lambda_0 (\frac{t-t_n}{\Delta_0}+1)^{-(1+\lambda_0\Delta_0)} e^{-\lambda_\infty (t-t_n)}, \quad \text{when } \lambda_\infty \ll \lambda_0 \end{aligned}$$

Thus, when $\alpha = m = \theta = 1$, the IET distribution is:

$$f(\tau) \approx \lambda_0 (\frac{\tau}{\Delta_0}+1)^{-(1+\lambda_0\Delta_0)} e^{-\lambda_\infty \tau}, \tag{5}$$

which exhibits constant rate λ_0 at short scale ($\tau < \Delta_0$), power law distribution with exponent $1+\lambda_0\Delta_0$ at middle scale, and the exponential mode at long scale (τ in the vicinity of λ_∞^{-1}) . □

Justification of the model:

Multiscale IET distribution. The multiscale pattern of IET distribution also holds in other parameter settings:

i) The range of short scale is controlled by Δ_0, as shown in Fig. 1b. When Δ_0 increases, the range of short-scale expands to right, and vice verse. When $\Delta_0 \to 0$, the short-scale of IET distribution disappears.

ii) The slope of the middle range is primarily controlled by θ. When $\alpha = m = 1$, the the situation when $\theta = 1$ is shown in lemma 3.2; when $\theta \neq 1$, the $e^{-\int_{t_n}^{t} \lambda_0 (\frac{s-t_n}{\Delta_0}+1)^{-\theta} ds}$ is a stretched exponential distribution which exhibits fat-tail IET distribution at middle scale. Thus, the larger θ is, the sharper slope is.

Short-term correlation of IETs. Both correlation-effect and multiscale-effect have impact on the short-term correlation of IETs:

i) When $\alpha = m = 1$, the LSmP denotes a renew process, indicating no correlation between IETs. When $m > 1$, the correlation appears. When $m = \infty$, $\Phi_0(t|\mathcal{H}_t)$ behaves like Hawkes process with power-law-like kernel, and the supercritical state ($\int \Phi_0(s|\mathcal{H}_t)ds > 1$) generates exponential growth, implying the strongest correlation.

ii) The larger value of θ, the quicker temporal fizzling of previous influence, and thus the weaker IETs' correlation.

Bursts. The bursts are governed by the multiscale effect and the correlation-effect:

i) As shown in Fig. 1b, Δ_0 is critical point between small IET (quick activities) in short-scale and fat-tailed IET in middle-scale. Thus, the larger Δ_0 is, the more bursty $n(t)$ is. The smaller θ, the more bursty of $n(t)$. Because the smaller θ, the fatter IET distributions at middle-scale, and thus more IETs with relatively small values appear.

ii) A longer memory length induces the stronger short-term IETs' correlation due to the superposition influence of previous events, leading to a sudden large rate/burst.

3.3 Parameter estimation

We learn the parameters of LSmP by maximizing log-likelihood function. The log-likelihood function of observing a point process $\{t_1, ..., t_{n-1}, t_n\}$ on time period $[0, T)$ is given by:

$$\begin{aligned}\log L(t_1, ..., t_n) &= -\int_0^T \lambda(t|\mathcal{H}_t)dt + \int_0^T \log \lambda(t|\mathcal{H}_t)dN(t) \\ &= -\lambda_\infty \Delta_\infty [(\frac{t_n}{\Delta_\infty}+1)^\alpha - 1] \\ &- \sum_{i=1}^{n} \frac{\lambda_0 \Delta_0}{1-\theta}[(\frac{t_{i+m}-t_i}{\Delta_0}+1)^{1-\theta} - 1] \\ &+ \sum_{i=1}^{n} \log[\lambda_\infty \alpha (\frac{t_i}{\Delta_\infty}+1)^{\alpha-1} + \lambda_0 A(i)]\end{aligned} \tag{6}$$

where $A(i) = \sum_{t_{i-m} \le tj < t_i} (\frac{t_i - t_j}{\Delta_0}+1)^{-\theta}$ for $i \ge 2$, and t_i denotes the time of occurrence of the ith event, and $A(1) = 0$. For convenience, we assume $T = t_n$. Maximizing Eq. 6 regarding $\{\lambda_\infty, \alpha, \Delta_\infty, \lambda_0, \theta, \Delta_0, m\}$, subject to $\{\alpha, \Delta_\infty, \theta, \Delta_0, \lambda_\infty, \lambda_0 \ge 0; m \in \mathcal{N}\}$ leads to estimated modeling parameters.

A good advantage of LSmP is that all the parameters have closed-form gradients. The gradients for the long-memory part are:

$$\frac{\partial \log L}{\partial \lambda_\infty} = -\Delta_\infty [(\frac{t_n}{\Delta_\infty}+1)^\alpha - 1] + \sum_{i=1}^{n} \frac{\alpha(\frac{t_i}{\Delta_\infty}+1)^{\alpha-1}}{D(i)} \tag{7}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \alpha} =& -\lambda_\infty \Delta_\infty (\frac{t_n}{\Delta_\infty}+1)^\alpha \ln(\frac{t_n}{\Delta_\infty}+1) \\ &+ \sum_{i=1}^{n} \frac{\lambda_\infty (\frac{t_i}{\Delta_\infty}+1)^{\alpha-1}[\alpha \ln(\frac{t_i}{\Delta_\infty}+1)+1]}{D(i)}\end{aligned} \tag{8}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \Delta_\infty} =& -\lambda_\infty [(\frac{t_i}{\Delta_\infty}+1)^\alpha - 1] + \lambda_\infty \alpha \frac{t_n}{\Delta_\infty}(\frac{t_n}{\Delta_\infty}+1)^{\alpha-1} \\ &- \sum_{i=1}^{n} \frac{\lambda_\infty \alpha(\alpha-1)\frac{t_i}{\Delta_\infty^2}(\frac{t_i}{\Delta_\infty}+1)^{\alpha-2}}{D(i)}\end{aligned} \tag{9}$$

where $D(i) = \lambda_\infty \alpha (\frac{t_i}{\Delta_\infty}+1)^{\alpha-1} + \lambda_0 A(i)$. When $\theta \neq 1$, the gradients for the short-memory part are:

$$\frac{\partial \log L}{\partial \lambda_0} = -\frac{\Delta}{1-\theta}\sum_{i=1}^{n}[(\frac{t_{i+m}-t_i}{\Delta}+1)^{1-\theta} - 1] + \sum_{i=1}^{n}\frac{A(i)}{D(i)} \tag{10}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \Delta_0} =& -\frac{\lambda_0}{1-\theta}\sum_{i=1}^{n}[(\frac{t_{i+m}-t_i}{\Delta_0}+1)^{-\theta}(\theta\frac{t_{t+m}-t_i}{\Delta_0}+1) - 1] \\ &+ \sum_{i=1}^{n}\frac{\lambda_0 B(i)}{D(i)}\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \theta} =& -\frac{\lambda_0 \Delta_0}{(1-\theta)^2}\sum_{i=1}^{n}[(\frac{t_{i+m}-t_i}{\Delta_0}+1)^{1-\theta} - 1 \\ &- (1-\theta)(\frac{t_{i+m}-t_i}{\Delta_0}+1)^{1-\theta}\ln(\frac{t_{i+m}-t_i}{\Delta_0}+1)] \\ &- \sum_{i=1}^{n}\frac{\lambda_0 C(i)}{D(i)}\end{aligned} \tag{12}$$

where $B(i) = \sum_{t_{i-m} \le tj < t_i} \frac{\theta(t_i - t_j)}{\Delta_0^2}(\frac{t_i-t_j}{\Delta_0}+1)^{-\theta-1}$ for $i \ge 2$ and $B(1) = 0$, and $C(i) = \sum_{t_{i-m} \le tj < t_i}(\frac{t_i-t_j}{\Delta_0}+1)^{-\theta}\ln(\frac{t_i-t_j}{\Delta_0}+1)$ for $i \ge 2$ and $C(1) = 0$. The counterparts when $\theta = 1$ can easily get, and we leave it for brevity. In numerical calculation when $\theta = 1$, we could add a very small error, say 10^{-8}, to θ to reuse above formulas.

For a specific m, we solve the constrained optimization problem by the trust-region-reflective algorithm [6], and then choose the best results by enumerating m according to the maximum likelihood principle [14]. For reproducibility, we open our code, see Section 5.

3.4 Simulation of LSmP dynamics

We design the simulation method of LSmP by the inverse method [8] (P260, Algorithm 7.4.III.) due to unbounded nature of $\lambda(t|\mathcal{H}_t)$ when $\alpha > 1$. The basic operation of the inverse method is solving the equation $\log u + \int_{t_n}^{t} \lambda(s|\mathcal{H}_t)ds = 0$ for t where u is generated from uniform distribution $U(0, 1)$. When $\alpha \le 1$, the adapting Ogata's thining algorithm [19] can be applied for speeding up. For brevity, we only present the inverse method here. We open our simulation code, see Section 5.

4 EXPERIMENTS

In this section, we evaluate the effectiveness of LSmP on the real data at large scale. We introduce the dataset in Sec.4.1. In Sec.4.2, we validate the accuracy of matching real data in three aspects. What's more, our LSmP can be applied in

Algorithm 1: Simulation algorithm for the Long-Short-Memory Process (LSmP)

Input : Hazard function of $\text{LSmP}\lambda(t|\mathcal{H}_t) = \Phi_\infty(t|\lambda_\infty, \Delta_\infty, \alpha) + \Phi_0(t|\lambda_0, \Delta_0, \theta)$, total event number N
Output : $\{t_1, ..., t_N\}$

Set current number of events $n = 1$, and current time $t = 0$;
while $n \leq N$ **do**
  Sample $u \sim Uniform([0, 1])$;
  Solve $\log u + \int_t^x \lambda(s|\mathcal{H}_t)ds = 0$ for x by Algorithm 2.;
  $t = t + x$;
  $t_n = t$;
  $n = n + 1$;
end

Algorithm 2: Newton's iterative method

Input : Equation $F(x) = \log u + \int_{t_n}^x \lambda(s|\mathcal{H}_t)ds$.
Output : x

Set $\epsilon = 10^{-8}$, $x = t_n - t_{n-1}$;
while $|F(x)| \leq \epsilon$ **do**
  $x = x - \frac{F(x)}{F'(x)}$;
end

many data mining tasks. In Sec.4.2.3, by analyzing the modeling parameters, we find statistical regularities underlying the empirical growth dynamics. In Sec.4.4, we cluster typical growth patterns of human connectivity and detect outliers.

4.1 Datasets

Our experiments are conducted on WeChat, which is the largest online social network in China with more than 806 million monthly active users by June 30, 2016. We collected the history data of WeChat consisting of complete records of all the node and link growth from January 21, 2011 (the release day of WeChat), to January 16, 2013 when the registered users reached 300 million. In total, there are 300 million registered users and more than 4.75 billion friendships. The records are (u, t_u) and $(u, v, t_{u,v})$ where u and v represent the users' ID, and t_u, $t_{u,v}$ are the timestamps of the birth of node and link respectively. In order to get enough friendships to study connectivity dynamics for each individual, we select the first 1 million users sorted by degree (with degree ≥ 143) and random select 10% of this population for the following experiments. All the WeChat data that we could access were anonymized for strict privacy policy.

4.2 Accuracy

We validate the accuracy of our LSmP by answering if our model can capture the empirical growth dynamics for each individual. The experiments are conducted by validating the accuracy of following three aspects: *(i)* the growth shape, *(ii)* the multiscale IET distribution, and *(iii)* the IETs' correlation.

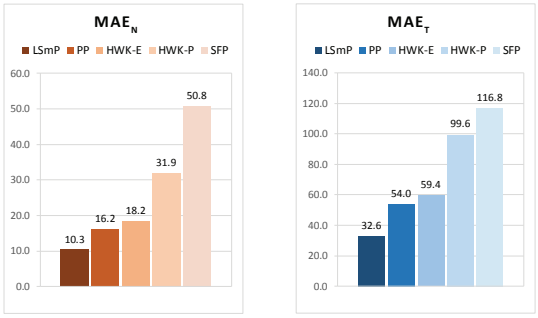
4.2.1 Baselines for human dynamics. We consider following four representative human dynamic models for comparison:

1) Poisson process (PP) has a constant hazard function $\lambda(t|\mathcal{H}_t) = \lambda$, which generates linear growth.

2) Hawkes process with exponential kernel (HWK-E) has hazard function $\lambda(t|\mathcal{H}_t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$ [9].

3) Hawkes process with power law kernel (HWK-P) has a fat-tailed kernel is used in HWK-P. The hazard function of this model is $\lambda(t|\mathcal{H}_t) = \mu + \sum_{t_i<t} \alpha(t - t_i)^{-\beta}$ [13]. HWK-P owns critical phenomena [7].

4) Self-feeding process (SFP) [1, 22] captures the Markovian dependency of IETs. The hazard function of this model is $\lambda(t|\mathcal{H}_t) = \frac{1}{\mu/e+\tau_i}$.



(a) (b)

Figure 2: *LSmP fits reality much better than competitors in both settings.* Median value of the Mean Average Error (MAE) for cumulative number of events over time $n_i(t)$, i.e., $MAE_N(i)$, and the event time epochs t, i.e., $MAE_T(i)$ are shown in (a) and (b), respectively.

4.2.2 Accuracy of growth curve shape. We evaluate the shape accuracy between the empirical dynamics $n_i(t) = |\{t_j < t | t_j \in \mathcal{H}_t = (t_1, ..., t_{n_i})\}|$ and the estimated dynamics $\hat{n_i}(t) = |\{\hat{t_j} < \hat{t} | \hat{t_j} \in \hat{\mathcal{H}}_t = (\hat{t_1}, ..., \hat{t_{n_i}})\}|$ of each individual i by considering the average mean error of the cumulative count: $MAE_N(i) = \frac{\sum_{j=1}^{n_i} |n(t_j) - \hat{n(t_j)}|}{n_i}$, and the average mean error of the event time epochs: $MAE_T(i) = \frac{\sum_{j=1}^{n_i} |t_j - \hat{t_j}|}{n_i}$. For each individual, we compute $MAE_N(i)$ and $MAE_T(i)$, and Figure 2 plots the median of MAEs for the $n_i(t)$.

The proposed LSmP model fits the reality much better than the baselines for both the metrics. In average, LSmP captures the number of event $n(t)$ at each event time epoch t_i with error 10.3. In contrast, the MAE_N given by the PP, HWK-E, HWK-P, SFP are 1.6, 1.8, 3.1, 4.9 times of the LSmP's, respectively. Similar results go with the error of the event time t_i: the MAE_T given by the PP, HWK-E, HWK-P, SFP are 1.7, 1.8, 3.1, 18.7 times of the LSmP's, respectively. The worst results are given by the SFP due to the fact that it may generate extremely overestimated IET. The HWK-P gives the second largest error due to the fact that the Hawkes process with power law kernel is easily in the supercritical regime, which generates exponential growth. The PP and the HWK-E give similar results because they generate linear growth.

Table 2: KS-Test of models. Our LSmP model captures the multiscale IET distribution. Winner in bold.

	KS-Test (Sig. $\alpha = 5\%$)				
τ	LSmP	PP	HWK-E	HWK-P	SFP
Pass Rate	**87.1%**	7.8%	62.9%	37.0%	2.4%
Error Statistic	**0.098**	0.256	0.134	0.162	0.291
$\log(\tau)$	LSmP	PP	HWK-E	HWK-P	SFP
Pass Rate	**87.1%**	7.8%	62.9%	37.0%	2.4%
Error Statistic	**0.098**	0.256	0.134	0.162	0.291

4.2.3 Accuracy of multiscale IET distribution. We then demonstrate that the LSmP accurately captures the multiscale IET distribution of the growth dynamics. We compare the IETs of each individual's connectivity dynamics $\tau_i = \{\tau_1, ..., \tau_{n_i}\}$ and the generated counterpart $\hat{\tau}_i = \{\hat{\tau}_1, ..., \hat{\tau}_{n_i}\}$ by testing whether τ_i and $\hat{\tau}_i$ are from the same continuous distribution. The two-sample Kolmogorov-Smirnov test (KS -Test) is applied for this standard statistical test mission. The fitting results of the model are shown by the indicator of passing the test, and a statistic representing the extent of difference between empirical data and the fitting results.

Table 2 presents the average pass rate and the error statistic for LSmP and the baselines with the significance level $\alpha = 5\%$. Our LSmP model outperforms all the baselines significantly in both the metrics: 87.1% of the generated dynamics by LSmP passes the KS-Test with average error 0.098. The pass rate of LSmP increases to 94.1% if we loose the significance level to 1%. We also conduct the hypothesis test on the logarithmically transformed IET $\log(\tau)$. We get same results for the $\log(\tau)$ as the τ. All the baselines fail to capture the multiscale IET distribution: *i)* the PP, which only generates exponential distribution in the long-scale, fails the short-scale and middle-scale; *ii)* the HWK-E exhibits a mixture of two exponential distributions at short-scale and long-scale, neglecting the fat-tail in middle-scale; *iii)* the HWK-H misses the flat short-scale; *iv)* the SFP cannot generate flat short-scale and exhibits large bias in the middle scale.

4.2.4 Accuracy of IETs' correlation. We further show that our LSmP captures the correlation between the IETs by comparing the joint distribution of consecutive IETs $p(\tau_{i,j}, \tau_{i,j+1})$ from empirical data, and the generated results $p(\hat{\tau}_{i,j}, \hat{\tau}_{i,j+1})$. We apply the two-dimensional version of the Kolmogorov-Smirnov test (2D-KS-Test) [21] to test whether $p(\tau_{i,j}, \tau_{i,j+1})$ and $p(\hat{\tau}_{i,j}, \hat{\tau}_{i,j+1})$ are same. We set the significance level to 5%. Table 3 presents the average pass rate and the error statistic for LSmP and other baselines. Again, LSmP beats all the baselines significantly in both metrics. We further examine to what extent the goodness-of-fit by the 2D-KS-Test are dominated by the marginal distributions. By shuffling the IET of the generated growth dynamics, we test the shuffled $p_s(\hat{\tau}_{i,j'}, \hat{\tau}_{i,j'+1})$ against the empirical $p(\hat{\tau}_{i,j}, \hat{\tau}_{i,j+1})$. We find LSmP gets the largest pass rate 77.5% (as shown in the second panel of Table 3) due to the nice characterization of the marginal distribution. What's more, LSmP also has the maximum additional increase 8% (=85.5% − 77.5%) due to the characterization of the correlation between IETs. Same conclusions can be drawn from the logarithmically transformed IET $\log(\tau)$.

Table 3: 2D-KS-Test of models. Our LSmP model captures the IETs' correlation. Winner in bold.

	2D-KS-Test (Sig. $\alpha = 5\%$)				
τ	LSmP	PP	HWK-E	HWK-P	SFP
Pass Rate	**85.5%**	7.9%	61.1%	34.1%	2.0%
Error Statistic	**0.145**	0.329	0.189	0.220	0.371
Shuffled τ	LSmP	PP	HWK-E	HWK-P	SFP
Pass Rate	**77.5%**	7.9%	57.2%	27.3%	1.9%
Error Statistic	**0.159**	0.329	0.194	0.231	0.384
$\log(\tau)$	LSmP	PP	HWK-E	HWK-P	SFP
Pass Rate	**85.4%**	6.7%	62.0%	34.0%	2.6%
Error Statistic	**0.145**	0.329	0.189	0.220	0.371

In all, only our LSmP correctly approximates the empirical growth dynamics of individuals' connectivity with respect to the shape of curves, the multiscale IET distribution and the IETs' correlation.

4.3 Discovering statistical regularity by parameter analysis

One good advantage of LSmP is that all the modeling parameters have clear physical meanings. In this section, we find statistical regularities underlying the empirical growth patterns through parameter analysis. To our best knowledge, this is the first empirical study of the profound exponents, α and θ, which govern human dynamics.

Figure 3 plots the distributions of the six parameters of LSmP. We fit the distributions of each parameter by Levenberg-Marquardt algorithm [17] and report their coefficients' values with confidence bounds 95%. We get following findings for the human dynamics of adding friends in the long-term:

Rich complexities and heterogeneities in power law growth. Figure 3a plots the distribution of the exponent α of power law growth, we find that it's a mixture of three distributions as follows, taking on complexity:

i) A skewed normal distribution ($\mu = 0.70$, $\sigma = 0.57$, $s = 2.6$) near 1 where μ is the mean location, the σ is the standard deviation and s is the skewness. The location value indicates in average users show decelerating power law growth with exponent $\alpha \sim t^{0.7}$. The positive skewness value shows that there exist a large number of linear growth($\alpha = 1$) and accelerating power law growth ($\alpha > 1$). A relatively large variance implies the heterogeneity of human growth patterns in the long-term. Besides, these findings at microscopic level are consistent with our previous finding [25] that at macroscopic level nodes grow over time $n(t) \sim t^{2.15}$, links grow over time $e(t) \sim t^3$ and thus average connectivity grows $\frac{e(t)}{n(t)} \sim t^{0.85}$.

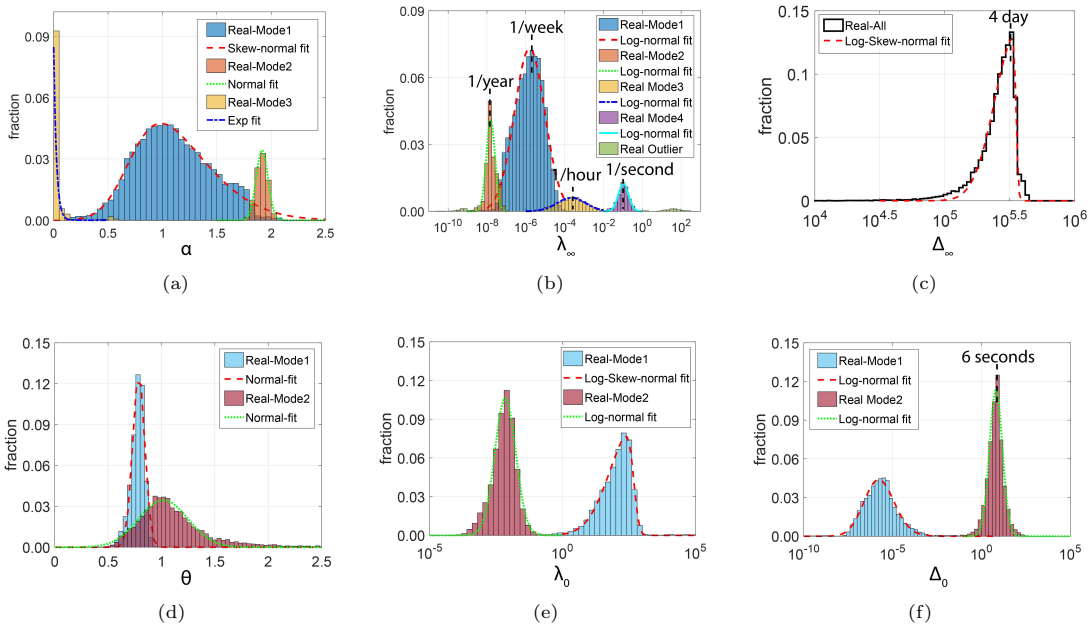


Figure 3: ***Parameter distributions.*** **Short-term parameters: (a) α, (b) λ_∞, (c) Δ_∞, and long-term parameters: (d) θ, (e) λ_0, (f) Δ_0. The colored bars represent the sub-clusters, and the same colors denotes same population. The overall distribution is marked by black stairs. The fitting curves for each cluster are denoted by colored lines.**

ii) A normal distribution ($\mu = 1.9$, $\sigma = 0.077$) near 2 implies that there exist some users with large power law exponents with upper-bound near ≈ 2. The reason of this bound might origin from the fact that the maximum number of links to be linked in a social network is proportional to the square of the number of nodes so far.

iii) An exponential distribution ($\alpha \propto e^{-39.4x}$) near 0, indicates that the growth dynamics in this regime decelerate significantly in the long-term. The finite size effect of friendships, possibly derived from the limited human cognition may account for this deceleration.

Log-normal distribution governing the multiscale long-term rate and human adds a friend per week. Figure 3b plots the distribution of growth rate λ_∞, and we find a mixture of four log-norm distributions spanning more than 8 order of magnitude, ranging from 1/second, 1/hour, 1/week to 1/year. The majority of human growth lie in the mode 1, with best log-norm fitting (for $\log(\lambda_\infty)$, $\mu = -5.80$, $\sigma = 1.11$), indicating that humans add friends in a cycle of 7.3 days in average. The small λ_∞ in mode 2 ($\mu = -3.65$, $\sigma = 1.04$) implies a long period of time without events at initial lifetime, while large λ_∞ in mode 3 ($\mu = -0.97$, $\sigma = 0.37$) and mode 4 ($\mu = -0.97$, $\sigma = 0.37$) imply a burst of events at the initial lifetime. We discuss the correlation between λ_∞ and α and their implications in the next subsection.

Typical time scale in the long-term is 4 days. Figure 3c plots the distribution of Δ_∞, best fitted by a skewed log-norm distribution ($\mu = 5.55$, $\sigma = 0.17$, $s = -12.00$), implying that the typical time scale of human dynamics of adding friends is near 4 days ($= 10^\mu$ s). For example, when you start a new job on Monday and establish a set of connections to your workmates until Friday (four days later!).

The findings for the human dynamics in the short-term are as follows:

Gaussian mixture of fizzling exponent. Figure 3d plots the the fizzling exponent θ, which follows a Gaussian mixture distribution with two modes: mode 1 ($\mu = 0.79$, $\sigma = 0.080$) and mode 2 ($\mu = 1.02$, $\sigma = 0.32$). The case when $\theta = 1$ in mode 2 is in line with the prediction of the priority queue model [23], indicating a power law decay (See Lemma 3.2). But we find rather than the power law decay in the critical state when $\theta = 1$, $\theta \neq 1$ is more general, indicating a stretched exponential decay. Especially we find the mode 1 with mean value 0.79, indicating a fatter tail of the IET distribution.

Bimodal short-term rate. Figure 3e plots the distribution of the short-term growth rate λ_0, which follows a bimodal distribution, where the skewed log-normal distribution ($\mu = 2.68$, $\sigma = 0.88$, $s = -5.63$) best fits the mode 1, and the log-normal distribution ($\mu = -2.16$, $\sigma = 0.56$) best explains the mode 2. The separated two modes show clear two clusters for human dynamics in the short IET scale. We discuss it in next section.

Bimodal critical time in the short-term, and adding friends as soon as 6 seconds. Figure 3f plots the distribution of short-term scale Δ_0, which also follows a bimodal distribution. The mode 2 is best fitted by log-normal distribution ($\mu = 0.81$,$\sigma = 0.51$), with mean value 6 seconds, indicating the critical time between the quick actions and the fat-tailed patterns of the IET distribution. Indeed, the WeChat offers multiple methods for adding friends as soon as possible, e.g.,"shake it". By shaking the mobile phone, user

can add strangers who is shaking at the same, and thus a random link is built. The mode 1 follows log-normal distribution ($\mu = -6.03, \sigma = 1.3$) with mean value $10^{-5.77}$ (~ 0) seconds, indicating that the IET distribution of this users consisting of only middle-scale and long-scale (Fig. 1d). In addition, we find strong correlations between these three short-term parameters: the population in Δ_0-mode1 correspond to the large λ_0 value in λ_0-mode11 and the 0.79 fizzling exponent in θ-mode1.

4.4 Clustering growth dynamics and detecting outliers

The LSmP's ability of pattern finding is further examined in the joint distribution of modeling parameters.

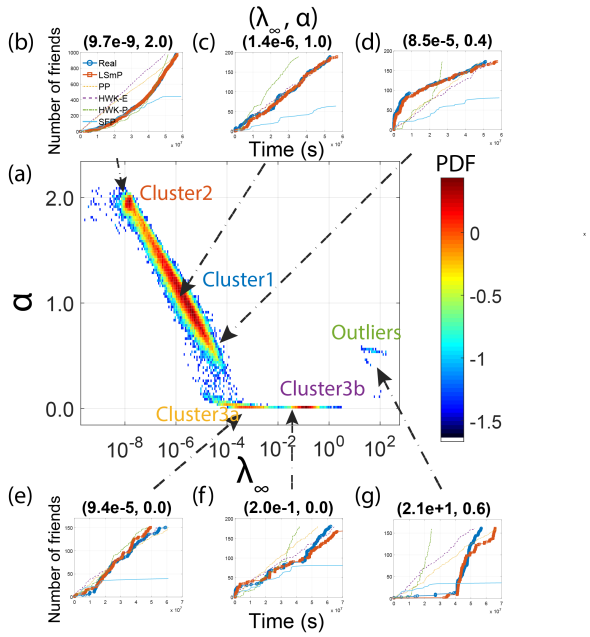


Figure 4: ***Growth modes in the long-term.*** **(a) The joint distribution of the long-term rate λ_∞ and the power law growth exponent α, on a log-linear scale. The corresponding marginal distributions are shown in Figs. 3. Instances of growth are shown in (b-g).**

Due to the prevailing of the (skewed/log) normal distribution in our data, we apply the Gaussian mixture model to the long-term parameter space $(\alpha, \lambda_\infty, \Delta_\infty)$ and short-term parameter space $(\theta, \lambda_0, \Delta_0)$ respectively, and find clusters, as shown in Fig. 3. Bars with different color represent different clusters, while the same color in different marginal distributions represent same cluster. We further examine the cluster patterns in the high-order parameter space [26], as shown in the joint distributions in Fig. 4 and Fig. 5. Different modes in the long-term and short-term are found by these joint distributions.

Long-term growth modes. Figure 4 plots the growth dynamics projected into the long-term (λ_∞, α)-space, we find

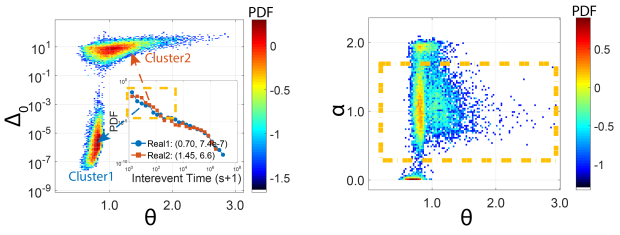


Figure 5: ***Growth modes in the short-term.*** **(a) The joint distribution of the short-term fizzling the exponent θ and the short-term time scale δ_0, on a linear-log scale. The inset plots two representative IET distribuions in cluster 1 and cluster 2, respectively. (b) The joint distribution of θ and α on a linear-linear plot.**

three major growth modes of human connectivity dynamics as our LSmP predicts:

Linear growth is in the vicinity of the $\alpha = 1$ region, consisting of the population near the mode of Cluster1. The $\alpha = 1$ features linear growth as shown in Fig. 4c.

Accelerating power law growth is in the $\alpha > 1$ regime, containing the top half part of the Cluster1, and the quadratic growth dynamics in Cluster2. The quadratic growth ($\alpha = 2$) is a special case of accelerating power law growth, as shown in Fig. 4b.

Decelerating power law growth is in the regime with $\alpha < 1$, consisting of the bottom half of the Cluster1 (one instance in Fig. 4d) and Cluster3a and 3b. Figures 4e&f plot two instances in Cluster3a and Cluster3b respectively. Along the λ_∞ axis, the Cluster3a exhibits a large delay at initial time and then decelerating power law growth followed at the small λ_∞ value region, while the Cluster3b shows initial burst then followed by decelerating power law growth growth at the large λ_∞ value region.

Short-term growth modes. Figure 5a plots the joint distribution of the short-term (θ, Δ_0)-space. We find the two modes of the short-term dynamics are separated by whether they show short-scale in IET distributions, as shown in the inset of fig. 5a.

Correlation between long-term and short-term. We further examine the correlation between the long-term parameters and the short-term parameters. The Figure 5b plots the joint distribution of the short-term fizzling exponent θ and long-term power law growth exponent α. We find the majority of dynamics (in Cluster1) show large variance with respect to θ, implying the relatively small correlations between the long-term memory and short-term memory.

Outlier detection. In addition, we showcase the outlier detection in the parameter space. By examining the joint profile for the λ_∞ and α, we also find outliers. Figure 4g plots one growth instance in the outlier point cloud, featuring long time vacation and large burst. A couple of actions for this user within first 1.3 year, he/she then added 50 friends in few minutes. We find the baselines can not capture these phenomena at all, but our LSmP even captures this

outlier to some extent. Furthermore, the parsimonious feature dimension and the clear physical meanings facilitate the design of more elaborated outlier detection method for future work.

5 CONCLUSIONS

In this paper, we studied the growth dynamics of individual's social connectivity in WeChat. We find various power law growth in the long-term and bursts in the short-term. We find the average-effect to account for power law growth and multiscale-effect and correlation-effect to account for bursts. Based fundamentally on these ingredients, we propose LSmP to capture the individual's social connectivity dynamics. Our LSmP gives a unified but parsimonious model which can accurately match and comprehensively explain growth dynamics of microscopic social connectivity. The main contributions are:

- **Novel findings**: We analyze the social connectivity dynamics in real social network at microscopic level and discover power law growth in the long-term and bursts in the short-term. We find three ingredients, i.e., average-effect, multiscale-effect and correlation-effect accounting for the power law and bursty growth.
- **Novel LSmP model**: We propose a stochastic model to capture above findings. LSmP is parsimonious and all the parameters have clear physical meanings.
- **Accuracy**: We present experiments on real social networks. Our LSmP model matches the real growth patterns accurately.
- **Usefulness**: Our LSmP provides insightful understandings on human dynamics, and implies applications in prediction, clustering, pattern discovery, and outlier detection, etc.

Limitation and future work. The sociological theories underlying the observed behaviors need further examination. Due to the ubiquity of growth phenomena, it's also interesting to see if our model can be applied to the growth dynamics of other systems, ranging from the connectome of C.elegans [18] to the growth of a company.

Reproducibility. We open-source our code of the LSmP at https://github.com/calvin-zcx/LSMP.

Acknowledgement

The authors thank anonymous reviewers for many useful discussions and insightful suggestions. This work was supported by National Program on Key Basic Research Project, No. 2015CB352300; National Natural Science Foundation of China, No. 61370022, No. 61521002, No. 61531006 and No. 61210008. Thanks for the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

REFERENCES

[1] Rodrigo Augusto da Silva Alves, Renato Martins Assuncao, and Pedro Olmo Stancioli Vaz de Melo. 2016. Burstiness Scale: A Parsimonious Model for Characterizing Random Series of Events. In *Proceedings of the 22nd ACM SIGKDD*. ACM, 1405–1414.

[2] Valerio Arnaboldi, Marco Conti, Andrea Passarella, and Robin Dunbar. 2013. Dynamics of personal social relationships in online social networks: a study on twitter. In *Proceedings of the first ACM conference on Online social networks*. ACM, 15–26.

[3] Albert-Laszlo Barabasi. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 7039 (2005), 207–211.

[4] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *science* 286, 5439 (1999), 509–512.

[5] Ginestra Bianconi and A-L Barabási. 2001. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)* 54, 4 (2001), 436.

[6] Thomas F Coleman and Yuying Li. 1996. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization* (1996).

[7] Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS* 105, 41 (2008), 15649–15653.

[8] Daryl J Daley and David Vere-Jones. 2003. *An introduction to the theory of point processes: volume I: Elementary Theory and Methods*. Springer Science & Business Media.

[9] Angelos Dassios, Hongbiao Zhao, and others. 2013. Exact simulation of Hawkes process with exponentially decaying intensity. *Electronic Communications in Probability* 18, 62 (2013), 1–13.

[10] Alceu Ferraz Costa, Yuto Yamaguchi, Agma Juci Machado Traina, Caetano Traina Jr, and Christos Faloutsos. 2015. Rsc: Mining and modeling temporal activity in social media. In *SIGKDD'15*. ACM.

[11] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[12] Hang-Hyun Jo, Juan I Perotti, Kimmo Kaski, and János Kertész. 2015. Correlated bursts and the role of memory range. *Physical Review E* (2015).

[13] Patrick J Laub, Thomas Taimre, and Philip K Pollett. 2015. Hawkes processes. *arXiv preprint arXiv:1507.02822* (2015).

[14] Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. 2008. Microscopic evolution of social networks *(KDD '08)*. ACM, 462–470.

[15] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2007. Graph evolution: Densification and shrinking diameters. *TKDD* (2007).

[16] R Dean Malmgren, Daniel B Stouffer, Adilson E Motter, and Luís AN Amaral. 2008. A Poissonian explanation for heavy tails in e-mail communication. *PNAS* 105, 47 (2008), 18153–18158.

[17] Jorge J Moré. 1978. The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*. Springer, 105–116.

[18] Vincenzo Nicosia, Petra E Vértes, William R Schafer, Vito Latora, and Edward T Bullmore. 2013. Phase transition in the economically modeled growth of a cellular nervous system. *PNAS* 110, 19 (2013), 7880–7885.

[19] Yosihiko Ogata. 1981. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* 27, 1 (1981), 23–31.

[20] Tohru Ozaki. 1979. Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* 31, 1 (1979), 145–155.

[21] JA Peacock. 1983. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society* 202, 3 (1983), 615–627.

[22] Pedro Olmo S Vaz de Melo, Christos Faloutsos, Renato Assunção, and Antonio Loureiro. 2013. The self-feeding process: a unifying model for communication dynamics in the web. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1319–1330.

[23] Alexei Vázquez, Joao Gama Oliveira, Zoltán Dezsö, Kwang-Il Goh, Imre Kondor, and Albert-László Barabási. 2006. Modeling bursts and heavy tails in human dynamics. *Physical Review E* 73, 3 (2006), 036127.

[24] Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber. 2010. Evidence for a bimodal distribution in human communication. *PNAS* 107, 44 (2010), 18803–18808.

[25] Chengxi Zang, Peng Cui, and Christos Faloutsos. 2016. Beyond Sigmoids: The NetTide Model for Social Network Growth, and Its Applications. In *Proceedings of the 22Nd ACM SIGKDD (KDD '16)*. ACM, 2015–2024.

[26] Chengxi Zang, Peng Cui, Chaoming Song, Christos Faloutsos, and Wenwu Zhu. 2017. Quantifying Structural Patterns of Information Cascades. In *Proceedings of the 26th International Conference on WWW Companion*. 867–868.

[27] Tianyang Zhang, Peng Cui, Chaoming Song, Wenwu Zhu, and Shiqiang Yang. 2016. A multiscale survival process for modeling human activity patterns. *PloS one* 11, 3 (2016), e0151473.