A Letters Journal Exploring the Frontiers of Physics

January 2020

EPL, **129** (2020) 20002 www.epljournal.org
doi: 10.1209/0295-5075/129/20002

Mixed distribution model of human communication and its impacts on the spreading process

Shengfeng Wang[1], Muhua Zheng[2,3(a)] and Jinghua Xiao[1(b)]

[1] *School of Science, Beijing University of Posts and Telecommunications - Beijing, China*
[2] *Departament de Física de la Matèria Condensada, Universitat de Barcelona - Martí i Franquès 1, 08028 Barcelona, Spain*
[3] *Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona - Barcelona, Spain*

received 6 November 2019; accepted in final form 5 February 2020
published online 24 February 2020

PACS 02.50.-r – Probability theory, stochastic processes, and statistics
PACS 89.75.Kd – Patterns
PACS 12.40.Ee – Statistical models

Abstract – The heterogeneous inter-event time of human contacts may fundamentally alter spreading dynamics. This generally assumes that the inter-event time distribution can be depicted with power-law–like decays. In empirical human communication data, the shape of the inter-event distribution is more complicated. Particularly, both the head and the tail of the inter-event distribution deviates from the power-law–like decay. In this paper, we examine two communication databases and propose a mixed distribution to depict the inter-event distributions, which agrees better with the empirical data. We then show how the inter-event distributions shape the co-evolved SIR spreading. Especially, the SIR dynamical equations are extended to adapt to the general inter-event distribution via introducing the new infected rate. By a numerical analysis of the newly infected individuals at each time step, we illustrate how the spreading size is determined by the inter-event time distribution and the parameters of the SIR dynamic.

Introduction. – Over the decades, due to the available digital data of the individual active behaviors, it has become possible to explore the statistic characteristics of the individual activity and its effects on the collective phenomena involved. Communication behavior as the most important human activity attracts a large amount of attention in the science community. In general, the activity in digital records of the communication behavior can be formatted as tuples, *i.e.*, the id of the sender, the id of the receiver and the timing of this special communication event. Based on this types of records, in the level of individuals, the distribution of the inter-event time τ of the consecutive communication events implies the underlying communication dynamic. In particular, the inter-event time τ of the consecutive event for individuals follows a power-law distribution [1,2].

These non-trivial individual activity patterns have impacts on the co-evolving dynamics across the population. The co-evolved dynamics ought to have the approaching time scale with the individual activity patterns. For the spreading dynamic, Iribarren and Moro *et al.* based on an experiment addressed the failure of Poissonian approximation and indicated that the large heterogeneity found in the response time slows down the information spreading [3], and also Vazquez *et al.* drew a similar conclusion by considering the power-law distribution of inter-event time and the structure of the spreading process to be tree-like [4]. The model combining individuals following power-law inter-event time is applied to the meme spreading over the online social media [5]. Apart from the spreading process, the effects of the inter-event time were considered on the random walk over the network [6]. Other dynamics like opinion dynamics also considered the influence of the inter-event time distribution recently [7].

The study of how the individual communication dynamics affects the co-evolving collective dynamics elucidates a central tension on how to reasonably model the individuals' inter-event time distribution. Different individual inter-event time distribution models may quantitatively change the results of the dynamics among the individuals. Previous researches consider the individual

(a) E-mail: zhengmuhua163@gmail.com
(b) E-mail: jhxiao@bupt.edu.cn

inter-event time distribution as the power-law distribution and the power-law distribution with a cutoff [8]. Some other forms of distribution such as the log-normal distribution are an alternative for the power-law distribution and even perform better with some empirical data [9]. On the other hand, when applied to the mutual communication process, the inter-event time distribution is more complicated than a single power-law distribution, and needs to combine the power-law pattern with other mechanisms [10]. Therefore, for understanding how human communication behavior affects the spreading process, a more accurate inter-event time distribution model ought to be introduced, in order to then examine its effects on the spreading process.

In this paper, we study the two usually used datasets and model the inter-event time distribution as a mixed distribution. In particular, we discriminate the interrelation of any two successive events in event series as either being related or unrelated, and thus, a mixed distribution seems to be the natural model. Based on the fact that both the head and tail of empirical distributions deviate from the power-law fitting, a log-normal distribution is considered to model the related inter-event time, while an exponential distribution models the unrelated or memoryless portion. Along with the two distributions, we consider one parameter a representing the probability of whether or not an event responds to the previous event. Given these assumptions, the inter-event time distribution can be expressed as a functional form of a mixed distribution which shows a good agreement with the empirical data. Furthermore, we study the effectiveness of the mixed distribution by the variances of the estimated parameters of the mixed distribution model over the two datasets. For investigating the effects of the inter-event time distribution on the co-evolving SIR spreading dynamics, the dynamical equations of the SIR process are modified to allow for the general inter-event time distribution. Such an extended form is built on decomposing the newly infected individuals at each time step as contributions from all previous time. In this framework, we show the effect of the distribution on the spreading process both by the numerical solution and the stochastic simulation.

Empirical patterns. – We consider two datasets that come in the form of lists of messages from one person to another at a time stamp. Database I comes from the operator company [10] with a total number of 548182 time stamp short message records and 44430 users over 30 days. Database II of e-mail logs [11] has a total number of 643502 communication events for 72146 users over 112 days. The time stamp of both datasets has a resolution of 1 s. We examine the individual event series in such datasets. For the validity of statistics, we extract the individuals' event series satisfying the total event number ranging from 100 to 3000 in Database I and ranging from 100 to 10000 in Database II. As a result, 2133 individuals of Database I and 1125 individuals of

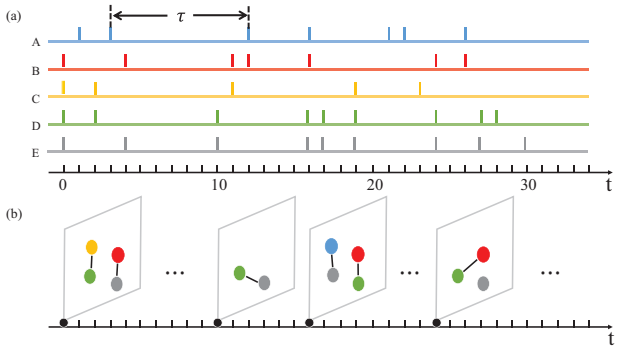

Fig. 1: The sketch map of an example containing 5 individual event series. (a) A, B, C, D, E represent the event series of five different individual, respectively. A bar in the series is an event of the individual being active at that time and contacting another active node. (b) A possible temporal network constructed by the individual even series. Here, the structures of 4 time points are shown as an example. The color of nodes corresponds to the individuals in (a). For example, at time $t = 0$, four individuals are active and contacts occur between them.

Database II are used for the following analysis. Formally, the communication event series of an individual can be denoted by $\{(E_1, t_1), (E_2, t_2), \ldots, (E_n, t_n)\}$ where E_i is an event at time point t_i. In the context of messages communication, E_i represents individuals sending or receiving a text message. We do not distinguish between the sender or the receptor of the communication event, which means the individual event purely represents a contact behavior no matter whether he or she is the sender or not. Therefore, an individual event series is reduced to a set of time stamps, $\{t_1, t_2, \ldots, t_n\}$. Generally, of interest is the inter-event time $\tau_i = t_{i+1} - t_i$ and its possibility density function $f(\tau)$.

Figure 1 gives a sketch map of a database only containing event series of 5 individuals. In fig. 1(a), A, B, C, D, E represent the event series of five different individuals, respectively. A bar in the event series represents a communication event meaning that the individual is active and contacts another active individual at this time point. The distribution of inter-event time $\tau_i = t_{i+1} - t_i$ gives the active pattern of individuals. Figure 1(b) shows a possible temporal structure constructed by the event series of 5 individuals. Notice that at time point $t = 24$, 3 individuals are active including a pair of interconnected nodes and an unpaired node. At time $t = 1$, only one individual is active and is not paired with the other 4 individuals. These unpaired individuals may indicate that the unpaired node has contact with other nodes outside of the database we considered. In this paper, we neglect the temporal interaction structure of the individuals or nodes and assume the active nodes are paired at random. With such an assumption, at each time step, active nodes randomly choose another active node and connect with each other until the next time step. Although there are cases

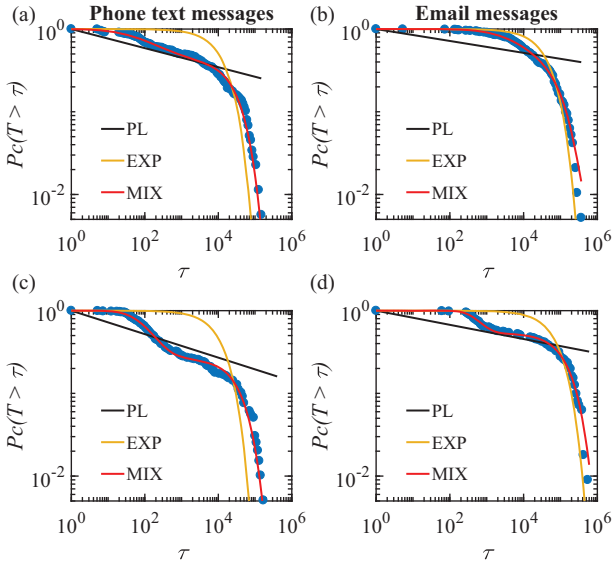

Fig. 2: Complementary cumulative distribution function (CCDF) $Pc(T > \tau)$ of the inter-event time for 4 typical individuals. The first column shows examples of the distribution of phone text messages. The second column is the distribution of e-mail messages. $Pc(\tau)$ is binned in the log-log scale. Blue dots show the empirical data. The red line is the fitting curve of the mixed distribution function (MIX). The black line and yellow line are the power-law fitting (PL) and exponential fitting (EXP), respectively. All the fitting lines are obtained by the least squares estimation (LSE). The mixed distribution parameters are (a) $a = 0.507$, $\lambda = 2.0\times10^{-5}$, $\mu = 5.12$, $\sigma = 2.588$; (b) $a = 0.581$, $\lambda = 1.148 \times 10^{-5}$, $\mu = 8.756$, $\sigma = 1.745$; (c) $a = 0.649$, $\lambda = 1.777 \times 10^{-5}$, $\mu = 5.319$, $\sigma = 1.342$; (d) $a = 0.465$, $\lambda = 5.795 \times 10^{-6}$, $\mu = 6.129$, $\sigma = 0.908$.

of unpaired active individuals, the effects can be ignored when the population is large enough. In this framework, the interaction between active individuals is random, and we study the active pattern of individuals and how the active pattern affects the spreading dynamic. The empirical distributions are shown to follow power-law decays for intermediate τ, and deviate from it at the tail and the head, as shown in fig. 2. Especially, the deviation of the tail can be divided into two cases. In the first case, the probability continuously decreases as shown in figs. 2(a) and (b), while in the second case, the individuals' event series follow the distributions where a hump occurs at the tail, examples shown in fig. 2(c) and (d). The complexity of the inter-event distribution implies a non-trivial mechanism underlying these communications behaviors.

In the context of one-to-one communication, we model individual communication behaviors as follows. One event would trigger the next with probability a, or in other words, the event would be responded to with probability a. The corresponding inter-event time is assumed to follow a log-normal distribution. On the contrary, a new event would occur with probability $1 - a$, and the corresponding inter-event time follows an exponential distribution. Here, we consider the log-normal distribution based on the observation of the derivation of the head from the body straight line in the log-log plot. Indeed, works in the literature presented the relation between the power-law and log-normal distributions [12] and these two distributions are prevalent in empirical data [2,13–15]. Our model corresponds to a hidden Markov model with two hidden states. Particularly, the individual activity is assumed to transform between these two states with a probability of a and $1 - a$. Due to the special state transformation rule, state change depends only on the current state and is not affected by the previous state and thus the inter-event time series can be modeled by a renewal process, so that event series could just be sampled randomly from the predefined distribution, which leads to a mixed distribution,

$$Pc(T > \tau) = (1 - a)e^{-\lambda\tau} + a\left(\frac{1}{2} - \frac{1}{2}erf\left(\frac{\ln(\tau) - \mu}{\sqrt{2}\sigma}\right)\right), \quad (1)$$

where $erf(x)$ is the error function used to simplify the integral expression of the log-normal distribution. Contrasting with the broadly used power-law distribution with cutoff $f(\tau) \propto \tau^{-\alpha}e^{-\lambda\tau}$, the mixed distribution model possesses flexibility in modifying whether the tail is over the extrapolation of the power-law fitting.

In fig. 2, the red line represents the fitting of the mixed distribution. The black line and the yellow line as benchmarks represent the power-law and exponential fitting, respectively. Our results show that the mixed distribution has a good fitting to the empirical data. We need to notice that the previous broadly used power-law with cutoff distribution, $f(\tau) = \tau^{\alpha}e^{-\lambda\tau}$, is not good to fit the case where the tail of the distribution displays an anomalous behavior (figs. 1(c) and (d)). On the other hand, because of our sufficient parameters within the mixed distribution model, in some senses, it appears a natural consequence to well approximate to the empirical data. However, applying this mixed distribution model to the empirical data, we found some of the parameters reusable among individuals. We use RSD (relative standard derivation) to measure the variation of parameters among the corresponding population. A large RSD of the parameter indicates corresponding parameters in the population are wildly distributed, and vice versa. Table 1 shows the RSD and the averages of the three parameters in the two databases. Both in the SMS and e-mail datasets, λ varies a lot among the populations. The other two parameters own a relative small RSD meaning that there is similarity among individuals. By parameter variances, the inter-event distribution of one particular communication behavior has not as many varying parameters as we supposed.

Individual active rate. – In this section, we study the individual active rate $P_e(t)$ which represents the expected possibility of an individual being active at time t provided that the individual is active at time $t = 0$. It has been shown that the special hidden Markov model means the inter-event time series comes from a renewal process, so that the active rate of individuals $P_e(t)$ is capable of

Table 1: The mean of parameters for the mixed distribution in the empirical data. The value within the round brackets indicates the relative standard derivation (RSD). All parameters are estimated by the non-linear least square method. The statistic results are from the fitting results whose R-squared is greater than 0.99. In the phone messages data, 1902/2133 individuals are included. In the email messages data, 965/1125 individuals are included.

Parameters	Database I	Database II
N	1902	965
a	0.507(0.428)	0.301(0.640)
λ	7.761×10^{-5}(3.334)	2.772×10^{-5}(1.455)
μ	6.210(0.248)	7.721(0.283)
σ	1.909(0.630)	2.081(0.833)

being directly derived from the inter-event distribution. For convenience, let us make clear the symbols used hereafter. The random variable T_i denotes the time interval between the $(i-1)$-th and the i-th event of this process. $N(t)$ represents the number of events that have happened by the time t. Then the time of the n-th event S_n can be expressed as: $S_n = \sum_{i=1}^{n} T_i$.

In the discrete-time framework, the probability of an event occurring or the active rate at timing t reads

$$P_e(t) = \sum_{i=1}^{t} P(S_i = t), \tag{2}$$

where $P(S_i = t) = f_i(t)$ with $f_i(t)$ the i-th–order convolution of the inter-event time distribution $f(\tau)$.

In spite of the complex form of $P_e(t)$, the limiting property of $P_e(t)$ shows that

$$P_e(t) \to 1/\langle\tau\rangle \text{ as } t \to \infty,$$

where $\langle\tau\rangle$ is the expected value of the inter-event distribution $f(\tau)$. This can be shown as $m(t) = E[N(t)] = \sum_{i=1}^{\infty} F_i(t)$, $F_i(t)$ is the cumulative distribution of $f_i(t)$. $m(t) = \int_0^t P_e(y)\mathrm{d}y$. By the elementary theorem of the book [16], $\frac{m(t)}{t} \to 1/\langle\tau\rangle$ as $t \to \infty$. Therefore, $\frac{\int_0^t P_e(y)\mathrm{d}y}{t} \to 1/\langle X\rangle$ as $t \to \infty$. Using the *L'Hospital's rule*, $P_e(t) \to 1/\langle X\rangle$ as $t \to \infty$.

The limiting property of $P_e(t)$ also implies that for comparing different individual active patterns in the same pages, the expectations of the inter-event time distribution should equal each other. According to the formulation of the active rate, $P_e(t) = 1/\langle\tau\rangle$ if the inter-event distribution $f(\tau)$ is exponential. Actually, only the exponential distribution has a constant active rate. For the power-law active pattern $f(\tau) \sim \tau^{-\alpha}$, the expectation $\langle\tau\rangle$ is not convergent if $\alpha < 2$. That means that the individual being active at $t = 0$ would not be active at $t \to \infty$. For comparison of different active patterns, we reproduce the mixed distribution pattern using the parameters estimated from Database I. While considering the singularity

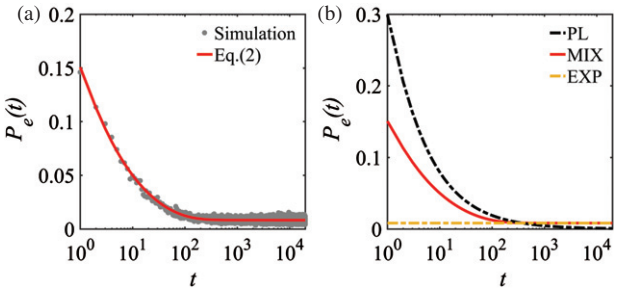

Fig. 3: The active rate $P_e(t)$ for the three inter-event distribution patterns. (a) shows the simulation result of active rates can be well depicted by the $P_e(t)$. (b) $P_e(t)$ gradually converges to $1/\langle\tau\rangle$. The mixed distribution pattern and the power-law distribution pattern have large $P_e(t)$ at initial time. The parameters of the mixed distribution pattern are estimated from Database I, while that of the power-law and the exponential patterns are estimated by fitting the generated mixed distribution. Simulation results are obtained by computing the probability of the synthetic event series whose inter-event time are drawn from the predefined distribution.

of the power-law distribution pattern, we obtain the exponent of the power-law distribution from the generated mixed distribution, and the parameter of the exponential distribution can be directly obtained from the expectation of the mixed distribution. Figure 3(a) shows the numerical solution of $P_e(t)$ approaches the simulation results well. Figure 3(b) compares the active rate $P_e(t)$ for the three patterns. At the initial time, the mixed and the power-law distribution pattern have large values then gradually converge to the limiting property $1/\langle\tau\rangle$.

Extended SIR dynamic for a general inter-event time distribution. – Each communication event represents individuals being active, that is potential for the co-evolving diffusion processes. The classical SIR diffusion model has been widely used to understand the effects of the active pattern of individuals. In the context of the SIR diffusion process, the individual states would be transformed from the susceptible (unaware of the information) to the infected state (knowing and spreading the information) and finally, the infected state would transform into the recovery state (knowing yet not spreading the information). In the following, we consider this co-evolving process via simulation and numerical solution of the derived equations.

In this paper, we focus our analysis on the active pattern through neglecting the interconnection structure though it impacts on the co-evolving dynamics as well. Therefore, at any snapshot of the temporal network, the active nodes are paired randomly. We assume that there are $n(t)$ active individuals at time t. For even $n(t)$, an active individual connects with another one with equal probability until all individuals are paired. For odd $n(t)$, a random active individual would not connect to other active individuals because only one link per node is allowed in this model. In the thermodynamic limit, the one rest node

can be ignored. After time t, the connections at t among active individuals are destroyed and new connections are created among the new active individuals. This generated temporal structure is illustrated by fig. 1(b). The computational simulation is built on this framework. To analyze the SIR process over this kind of random paired temporal population, we extend the SIR process to allow for individual following general inter-event time distributions. For the sake of simplicity and convenience, we consider the discrete SIR process which in general is given by

$$\begin{aligned} S(t) - S(t-1) &= -Ia(t), \\ I(t) - I(t-1) &= Ia(t) - \gamma I(t-1), \\ R(t) - R(t-1) &= \gamma I(t-1), \end{aligned} \tag{3}$$

where $Ia(t)$ represents the newly infected individuals at time t. γ is the recovery rate of the SIR process. This set of equations can be solved numerically so long as the only unknown term $Ia(t)$ can be expressed as the function of $I(t')$ and $S(t')$ with $t' < t$. Consider a textbook situation, if the individuals are continuously active during the whole time range, in the fully mixed population, $Ia(t) = \beta S(t-1)I(t-1)$ with β representing susceptible individuals being affected by the infectious individuals. When it comes to the situation where individuals are only active at some special time points, and only active individuals are potential for propagating information, we analyze this case by unfolding the $Ia(t)$ into the contributions from all previous discrete time. Suppose $x < t$, then $Ia(x)(1-\gamma)^{t-x}P_e(t-x)$ will be the active infected individuals at time t contributed by active individuals at time x. $(1-\gamma)^{t-x}$ and $P_e(t-x)$ represent the probability of individuals still being infected and active at time t, respectively. Then, in general, $Ia(t)$ can express the accumulation rate as $Ia(t) = \beta(\sum_{i=1}^{t-1} Ia(i)(1-\gamma)^{t-i}P_e(t-i)) \times \frac{(P_e(t+W)-\sum_{i=1}^{t-1} Ia(i)P_e(t-i))}{P_e(t+W)}$, where $P_e(t+W)$ represents the active fraction of population at time t and W is the elapsed time before the spreading process happens. And since $P_e(t)$ converges to a constant, W occurs here for removing the influence of the early time. Formally, the new individuals $Ia(t) = \beta(\sum_{i=1}^{t-1} Ia(i)(1-\gamma)^{t-i}P_e(t-i)) \times \frac{(P_e(t+W)-\sum_{i=1}^{t-1} Ia(i)P_e(t-i))}{P_e(t+W)}$ for all $t \geq 2$. Given $Ia(1)$ a fraction of the seeds, the set of equations (3) can be solved iteratively. Also, since $S(t) = 1 - \sum_{i=1}^{t} Ia(i)$ for all t, $R(t) = 1 - S(t) = \sum_{i=1}^{t} Ia(i)$ in the limit $t \longrightarrow \infty$. According to these reasonings, the new adding infected nodes $Ia(t)$ at every time step can also represent the SIR process.

Figures 4(a), (c), (e) show the SIR process evolving with the time t. In the short term of the spreading process, the mixed distribution pattern and the power-law distribution pattern accelerate the spreading shown by the location of the hump of $i(t)$, the fraction of the infected nodes. Figures 4(b), (d), (f) are $Ia(t)$, the corresponding new adding infected nodes of the SIR process. Here we show the figures from time step $t = 2$ given that the seed of the infected individuals of the three different patterns

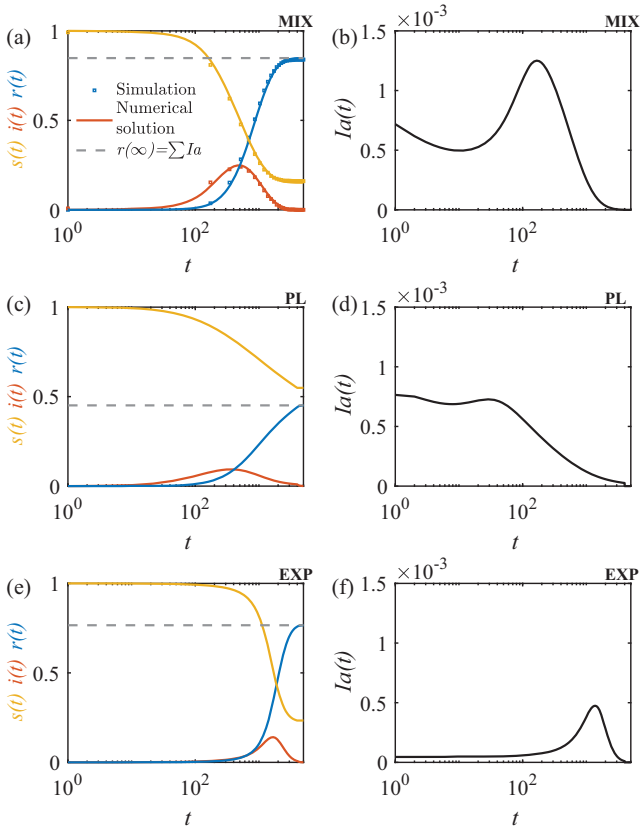

Fig. 4: The evolution of the SIR process over the temporal population. (a), (c), (e) correspond to the results of the mixed distribution pattern, the power-law distribution pattern and the exponential distribution pattern, respectively. Solid lines are the numerical solutions of the SIR spreading process. (b), (d), (f) are the corresponding new adding infected nodes $Ia(t)$ at time t. The location of the hump of $Ia(t)$ shows the early feature of the spreading process. The integral of $Ia(t)$ indicates the evolution results of the SIR process. The simulation runs over the network with size $N = 4000$. The parameters of the SIR process are $\beta = 0.7$ and $\gamma = 0.03$.

are equal. The $Ia(t)$ for the exponential distribution pattern is relatively small in the short term and then has a peak representing bursty increasing of $Ia(t)$. Compared with the exponential distribution pattern, the power-law distribution pattern has a large $Ia(t)$ at the initial time step. The mixed distribution has a bursty increasing of $Ia(t)$ and the bursty duration occurs earlier than that of the exponential distribution pattern. The $Ia(t)$ of the mixed distribution pattern is a combination of the power-law distribution pattern and the exponential distribution pattern. From the observation of $Ia(t)$, the differences between different inter-event distributions are clearly shown. Next, we study the threshold and the spreading size of the SIR process. In fig. 5, the phase diagram is applied to show the correlations between the spreading size and the parameters of the SIR process. First, let us put some thought into the threshold of the SIR process which separates the phase diagram into two regimes in which the

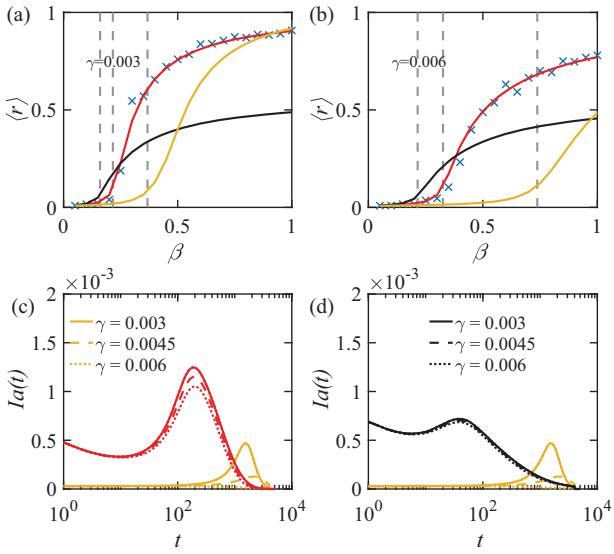

Fig. 5: The phase diagram of the SIR process and the evolving of $Ia(t)$ for different distribution patterns under different γ. The red line, the black line, and the yellow line represent the results of the mixed distribution pattern, the power-law distribution pattern and the exponential distribution pattern, respectively. (a) and (b) correspond to the different recovery rates. The blue cross symbol represents the simulation results. The solid lines are obtained by the integral of $Ia(t)$. The parameters of the inter-event distribution used in this figure are from Database I in table 1. The gray dashed lines correspond to the threshold computed by eq. (5). Different γ's greatly affect the spreading size of the exponential distribution pattern. (c) and (d) show how $Ia(t)$ is affected by γ. As the increasing of $Ia(t)$ for the power-law distribution pattern and the mixed distribution pattern is at the early time of the spreading process, the large γ has less impacts on these two patterns.

infected nodes of the SIR process either spread out or die out. This spreading size is written as $R(\infty) = \sum_{t=1}^{\infty} Ia(t)$. Generally, around the threshold, the quadratic terms and higher-order terms of $Ia(i)$ can be ignored. Therefore, the recursive equations can be simplified as

$$\sum_{t=1}^{\infty} Ia(t) = \beta \sum_{t=1}^{\infty} \sum_{i=1}^{t-1} Ia(i)(1-\gamma)^{t-i} P_e(t-i) \qquad (4)$$

Through exchanging the order of the summation over t and i, the spreading threshold β is expressed by

$$\beta_c = \frac{1}{\sum_{i=1}^{\infty}(1-\gamma)^i P_e(i)}. \qquad (5)$$

The exponential decay form of $(1-\gamma)$ in eq. (5) means that the spreading threshold β_c is basically determined by the early feature of $P_e(t)$. The mixed distribution and the power-law distribution lead to a large $P_e(t)$ at the initial time and then gradually converges to $1/\langle\tau\rangle$. Indeed, the phase diagram in figs. 5(a) and (b) shows that the mixed distribution pattern and the power-law distribution pattern have a smaller threshold than the exponential distribution pattern. By the combination of eq. (5) and the feature of $P_e(t)$ for the exponential pattern and supposing $\beta = 1$, we can derive the upper bound of the recovery rate as $\gamma = 1 - \frac{1}{\langle\tau\rangle}$. Moreover, figs. 5(a) and (b) show the phase diagram under different recovery rate γ. In this figure, the red line, the black line and the yellow line represent the results of the mixed distribution pattern, the power-law distribution pattern and the exponential distribution pattern, respectively. The numerical solution is coherent with the simulation results. For the small recovery rate $\gamma = 0.003$ and large spreading rate β, the spreading size of the power-law distribution pattern is less than that of the exponential distribution pattern. On the contrary, if the recovery rate $\gamma = 0.006$ or for small spreading rate β, the spreading size of the power-law distribution pattern is greater than that of the exponential distribution pattern. This feature would come from the quadratic terms of the recursive equation of $Ia(t)$ which can be neglected in computing the threshold as in eq. (5). According to our previous equations, the spreading size $R(\infty)$ is the integral of the new adding infected nodes $Ia(t)$. Figures 5(c) and (d) show how the recovery rate of the SIR process γ affects $Ia(t)$. For the exponential distribution pattern, the bursty increasing duration has been put off and the hump of the burst drops down with the increasing of the recovery rate γ, while for the mixed distribution pattern and the power-law distribution pattern the descending of the $Ia(t)$ is quite slight. The increasing of γ affects greatly the spreading size of the exponential distribution pattern as its bursty duration of $Ia(t)$ needs a long wait.

Discussion and conclusion. – In the modeling of the inter-event time distribution of human communication behavior, the model based on the hidden Markov chain was shown to well depict the real data. We showed in the study of two empirical databases, by ignoring the type of the communication event, that the communication process can be modeled by the hidden Markov model with two states. This simple case allows an analytical inter-event distribution which is useful in the study of the effects on the co-evolving dynamic. A mixed distribution derived from the hidden Markov model as $f(\tau) = (1-a)\lambda e^{-\lambda\tau} + a\frac{1}{\tau\sigma\sqrt{2\pi}} e^{-\frac{(\ln\tau-\mu)^2}{2\sigma^2}}$ was shown to have a good agreement with the empirical inter-event distributions. Here, the Markov assumption appears to be reasonable in a one-to-one communication behavior, because people hardly respond to the event that occurs far from the latest event. This is in contrast with the collective behavior involving many people, like editing articles of Wikipedia in which any event can be triggered by all previous events [17]. Moreover, although our model has four parameters to determine the empirical distribution, among the population, three of them are relatively stable and only one parameter λ varies a lot. The parameter λ explains the shape of the tail, and a small λ corresponds to the case where a hump occurs around the tail of the distributions. This justification is in the spirit coherent with the bimodal

model [10] in which the parameter of the Poisson process affects the shape of the tail of the distribution.

Based on the individual inter-event distribution, we investigate how the inter-event time pattern quantitatively affects the SIR process. Assuming the population being fully mixed, we derived the dynamic equations of the SIR process by unfolding the newly added infected individuals $Ia(t)$ as being contributed to by all the previous time. Within these equations, the individual active probability $P_e(t)$ is the core to understand the process. As seen in this paper, the mixed distribution has a large $P_e(t)$ at the initial time and then gradually converges to that of the exponential distribution pattern. Actually, these results are relatively robust, and by the definition of $P_e(t)$, only the Poissonian pattern has a constant $P_e(t)$. Therefore any non-Poissonian pattern would have a larger probability at the initial time and then gradually converge to the Poisson pattern when these inter-event time distributions keep the same expectation. According to our analysis, at the initial time, the larger $P_e(t)$ is, the larger $Ia(t)$ is, consequently the SIR spreading process across the population runs much faster at initial times. On the basis of our expansion for the SIR dynamic, we also show that the smaller threshold of the SIR process is guaranteed by the $P_e(t)$ with large value at the initial time. However, if the infectious rate β of the SIR dynamic is far larger than the threshold, this condition would induce a different long-term result of the spreading size for the exponential and the power-law distribution pattern at different recovery rates γ. This difference is illustrated by the impact of the recovery rate γ on $Ia(t)$. Because the bursty increasing of $Ia(t)$ for the exponential distribution occurs later than the power-law and the mixed distribution, large γ significantly affects the hump of $Ia(t)$ for the exponential distribution while it does not affects much the mixed and the power-law distribution.

In summary, under the assumption of the state transition, we analytically derived the form of the inter-event distribution that shows a good agreement with the empirical data. In terms of the active rate $P_e(t)$ that was calculated from the inter-event time distribution, we explored the possible influence of inter-event time distributions on the co-evolving SIR spreading process. In comparison with the exponential distribution, the mixed and the power-law distributions accelerate the spreading at initial times. The threshold of the spreading model is mainly determined by the initial features of the active rate $P_e(t)$. Since the time duration of the increasing of $Ia(t)$ for the exponential distribution pattern was later than that of the mixed distribution pattern and the power-law distribution, the spreading size of the exponential distribution was greatly affected by the recovery rate γ of the SIR process. In these analyses, considering the mixed distribution model quantitatively changes the spreading pattern.

* * *

This research was supported by the National Natural Science Foundation of China (grant No. 11775034) and the Fundamental Research Funds for the Central Universities with contract No. 2019XD-A10.

REFERENCES

[1] Barabasi A.-L., *Nature*, **435** (2005) 207.
[2] Clauset A., Shalizi C. R. and Newman M. E. J., *SIAM Rev.*, **51** (2009) 661.
[3] Iribarren J. L. and Moro E., *Phys. Rev. Lett.*, **103** (2009) 038702.
[4] Vazquez A., Racz B., Lukacs A. and Barabasi A. L., *Phys. Rev. Lett.*, **98** (2007) 158702.
[5] Wang L.-Z., Zhao Z.-D., Jiang J., Guo B.-H., Wang X., Huang Z.-G. and Lai Y.-C., *Chaos*, **29** (2019) 023136.
[6] Hoffmann T., Porter M. A. and Lambiotte R., *Phys. Rev. E*, **86** (2012) 046102.
[7] Li M. and Dankowicz H., *Phys. A: Stat. Mech. Appl.*, **523** (2018) 1355.
[8] Rocha L. E. C. and Blondel V. D., *PLOS Comput. Biol.*, **9** (2013) e1002974.
[9] Stouffer D. B., Malmgren R. D. and Amaral L. A., arXiv preprint, physics/0605027 (2006).
[10] Wu Y., Zhou C., Xiao J., Kurths J. and Schellnhuber H. J., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 18803.
[11] Ebel H., Mielsch L. I. and Bornholdt S., *Phys. Rev. E*, **66** (2002) 035103.
[12] Mitzenmacher M., *Internet Math.*, **1** (2004) 226.
[13] Blenn N. and Van Mieghem P., arXiv preprint, arXiv:1607.02952 (2016).
[14] Sun K., Technical Report, Department of Physics (2004).
[15] Benguigui L. and Marinov M., arXiv preprint, arXiv:1507.03408 (2015).
[16] Ross S. M., *Introduction to Probability Models*, 10th edition (Academic Press, Amsterdam, Boston) 2010.
[17] Zha Y. L., Zhou T. and Zhou C. S., *Proc. Natl. Acad. Sci. U.S.A.*, **113** (2016) 14627.