Towards the Understanding of Human Dynamics

Tao Zhou, Xiao-Pu Han and Bing-Hong Wang

Quantitative understanding of human behaviors provides elementary but important comprehension of the complexity of many human-initiated systems. A basic assumption embedded in previous analyses on human dynamics is that its temporal statistics are uniform and stationary, which can be properly described by a Poisson process. Accordingly, the interevent time distribution should have an exponential tail. However, recently, this assumption is challenged by extensive evidence, ranging from communication to entertainment to work patterns, that human dynamics obeys non-Poisson statistics with heavy-tailed interevent time distribution. This chapter reviews and summarizes recent empirical explorations on human activity pattern, as well as the corresponding theoretical models for both task-driven and interest-driven systems. Finally, we outline some open questions in the studies of statistical mechanics of human dynamics.

12.1 Introduction

Human behavior, as an academic issue in science, has a history of about one century since the time of Watson [1913]. As a joint interest of sociology, psychology and economics, human behavior has been extensively investigated during the last decades. However, due to the complexity and diversity of our behaviors, the in-depth understanding of human activities is still a long-standing challenge thus far. Actually, up to now, most of academic reports on human behaviors are based on clinical records and laboratorial data, and most of the corresponding hypotheses and conclusions are only qualitative in nature. Therefore, we have to ask at least two questions: (1) Could those laboratorial
observations properly reflect the real-life activity pattern of us? (2) Can we establish a quantitative theory for human behaviors?

Barabási [2005] provided a potential start point in answering those questions, that is, extracting the statistical laws of human behaviors from the historical records of human actions. Traditionally, the individual activity pattern is usually simplified as a completely random point-process, which can be well described by a Poisson process,\(^1\) leading to an exponential interevent time\(^2\) distribution [Haight, 1967]. That is, the distribution of time difference between two consecutive events should be almost uniform, and a long gap is hardly to be observed. However, recently, both empirical studies and theoretical analyses display a far different scenario: our activity patterns follow non-Poisson statistics, characterized by bursts of rapidly occurring events separated by long gaps. These new findings have significant scientific and commercial potential. As pointed out by Barabási [2005], models of human activities are crucial for better resource allocation and pricing plans for telephone companies, to improve inventory and service allocation in both online and “high street” retail.

Barabási and his colleagues have opened up a new research area, namely, Human Dynamics. Still in its infancy but motivated by both theoretical and practical significances, the studies of human dynamics attract more and more attention. In this chapter, we summarize recent progresses on this topic, which may be helpful for the comprehensive understanding of the architecture of complexity [Barabási, 2007]. This chapter is organized as follows. In the next section, we show the empirical evidence of non-Poisson statistics of human dynamics. In Sections 12.3 and 12.4, the task-driven and interest-driven models are introduced. Finally, we outline some open problems in studying the statistical mechanics of human dynamics.

### 12.2 Non-Poisson Statistics of Human Dynamics

\(^1\) A Poisson process is defined by Eq. (12.2) below.

\(^2\) Interevent time is the time duration between two consecutive events of the same nature, to be specified in each concrete case under discussion. See examples in Section 12.2.
Previously, it is supposed that temporal statistics of human activities can be described by a Poisson process. In other words, for a sufficiently small time difference \( \Delta t \), the probability that one event (i.e., one action) occurs in the interval \([t, t+\Delta t]\) is independent of time \( t \), and has approximately a linear correlation with \( \Delta t \) such as
\[
P(t, t+\Delta t) \approx \lambda \Delta t, \quad \lambda > 0, \tag{12.1}
\]
where \( \lambda \) is a constant. Under this assumption, the interevent time between two consecutive events, denoted by \( \tau \), obeys an exponential distribution [Haight, 1967]:
\[
P(\tau) = \lambda e^{-\lambda \tau}. \tag{12.2}
\]
Equation (12.2) represents an exponential decay, implying that a long gap without any event (i.e., a very large \( \tau \)) should be rarely observed. However, we will show below extensive empirical evidence in real human-initiated systems, where the distributions of interevent time have much heavier tails than the one predicted by Poisson process. Hereinafter, we will review the empirical results for different systems one at a time.

1. **Email communication**
   The data set contains the email exchange between individuals in a university environment for three months [Eckmann et al., 2004]. There are in total 3,188 users and 129,135 emails with second resolution. Denote by \( \tau \) the interevent time between two consecutive emails sent by the same user, and \( r_w \) the response time taking for a user to reply a received email. As shown in Fig. 12.1, both the distributions of interevent time and response time obey a power law\(^3\) with exponent approximately equal to -1. Although the exponent differs slightly from user to user, it is typically centered on -1.

2. **Surface mail communication**

---

\(^3\) A power laws means two variables \( y \) and \( x \) are related by the relationship, \( y \propto x^a \), where the exponent \( a \) is a constant.
The data set used for analysis contains the correspondence records of three great scientists: Einstein, Darwin and Freud. The sent/received numbers of letters for those three scientists are 14,512/16,289 (Einstein), 7,591/6,530 (Darwin), and 3,183/2,675 (Freud). The dataset is naturally incomplete, as not all letters written or received by these scientists were preserved. Yet, assuming that letters are lost at a uniform rate, they should not affect the main statistical characteristics. As shown in Fig. 12.2 [Oliveira & Barabási, 2005; Vázquez et al., 2006], the distributions of response time follow a power law with exponent approximated to -1.5. The readers should be warned that the power-law fitting for Freud is not as good as that for Einstein or Darwin. Recently, we analyze the correspondence pattern of a Chinese scientist, Xuesen Qian, and find that both the distributions of interevent time and response time follow a power law with exponent -2.1 [Li et al., 2007].

Fig. 12.1. Heavy-tailed activity patterns in email communications [Barabási, 2005]. (a) Distribution of interevent time; (b) distribution of response time. Data shown in these two plots are extracted from one user. Both solid lines have slope -1 in the log-log plot.

3. Short-message communication

The data set contains the records of short-message communications of a few volunteers using cell phones in China [Hong et al., 2008]. Figure 12.3 shows two typical distributions of interevent time between two consecutive short messages sent by an individual. Both distributions can
be well fitted by a power law with different exponents. Actually, in the data set, almost all distributions of interevent time can be well approximated by a power law, and there is apparently correlation between the average numbers of short messages sent out per day and the power law exponents.

![Graph showing distributions of response times for letters replied by Einstein, Darwin and Freud, respectively.](image1)

**Fig. 12.2.** Distributions of response times for letters replied by Einstein, Darwin and Freud, respectively [Vásquez et al., 2006]. All three distributions are approximated by a power law tail with exponent -1.5.

![Graph showing distributions of time interval of sending short-messages.](image2)

**Fig. 12.3.** Typical examples of distributions of time interval of sending short-messages on log-log plots [Hong et al., 2008]. The x axis denotes time interval (in hour); y axis, the probability. The two distributions are approximated by a power law with exponent -1.52 and -1.70, respectively.

4. **Web browsing**
Web browsing history can be automatically recorded by setting cookies. The data set contains the visiting records of 250,000 visitors to the site www.origi.hu from Nov. 8 to Dec. 8 in the year 2002, with about 6,500,000 HTML hits per day [Dezső et al., 2006]. Figure 12.4 shows the interevent time distribution of a single user, which can be approximately fitted by a power law with exponent -1.0. Although the power law exponents for different users are slightly different, they centered around -1.1. Actually, the exponent distribution obeys a Gaussian function with characteristic value approximate to 1.1 [Vázquez et al., 2006]. Figure 12.5 reports the interevent time distribution of all the users [Dezső et al., 2006], which can be well fitted by a power law with exponent -1.2. Recently, we have analyzed the browsing history through the portal of an Ethernet of a university (University of Shanghai for Science and Technology) in 15 days, with about 4,500,000 URL requirements per day [Zhao et al., 2008]. Similar to the above results, we demonstrate the existence of power-law interevent time distribution in both the aggregated and individual levels. However, the exponents, ranged from -2.1 to -3, are much different from the visiting pattern to the single site www.origi.hu.

Fig. 12.4. The distribution of interevent time between two consecutive web visits by a single user [Vázquez et al., 2006]. \(N(\tau)\) stands for the frequency. The solid line has slope -1.0.

5. Library loan
The data set contains the time books or periodicals were checked out from the library by the faculty at the University of Notre Dame during three years [Vázquez et al., 2006]. The number of borrowers is 2,247, and the total transaction number is 48,409. The interevent time is the time difference between consecutive books or periodicals checked out by the same patron. Figure 12.6 presents the interevent time distribution of a typical user, which can be well fitted by a power law with exponent about -1.0. The exponents for different users are different, ranging from -0.5 to -1.5, with the average around -1.0.

![Graph](image1.png)

Fig. 12.5. Distribution of interevent time between two consecutive web visits by all users [Dezső et al., 2006]. $N(\tau)$ stands for the frequency. The solid line has slope -1.2.

![Graph](image2.png)

Fig. 12.6. Distribution of interevent time between two consecutive books or periodicals checked out by the same user [Vázquez et al., 2006]. $N(\tau)$ stands for the frequency. The solid line has slope -1.0.

6. **Financial activities**
The data set contains all buy/sell transactions initiated by a stock broker at a Central European bank between June 1999 and May 2003, with an average of ten transactions per day and a total of 54,374 transactions [Vázquez et al., 2006]. The interevent time represents the time between two consecutive transactions by the broker, with the gap between the last transaction at the end of one day and the first transaction at the beginning of the next trading day ignored. Different from the empirical systems above, the interevent time distribution for stock transactions obviously departs from a power law (Fig. 12.7). A recent empirical analysis on a double-auction market [Scalas et al., 2006] also shows that the interevent time between two consecutive orders do not follow the power-law distribution. Although the interevent time distributions in these two financial systems cannot be considered as power laws, they do display clearly heavy-tailed behaviors, not in the shape of an exponential function.

Fig. 12.7. Distribution of interevent time between two consecutive transactions initiated by a stock broker [Vázquez et al., 2006]. $N(\tau)$ stands for the frequency. The solid line represents a truncated power law, $\tau^{-\alpha}\exp(-\tau/\tau_0)$, with $\alpha = 1.3$ and $\tau_0 \approx 76$ min.

7. **Movie watching**

The data were collected by a large American company for mail orders of DVD-rental (www.netflix.com). The users can rate movies online. In total, the data comprises of 17,770 movies, 447,139 users and about 96.7
Towards the Understanding of Human Dynamics

millions of records. Tracking the records of a given user $i$, one can get $k_i - 1$ interevent times, where $k_i$ is the number of movies user $i$ has already watched. The time resolution of the data is one day. Figure 12.8 reports the interevent time distribution based on the aggregated data of all users. The distribution follows a power law for more than two orders of magnitude, with an exponent approximately equal to -2.08.

![Fig. 12.8. Distribution of interevent time at the population level, indicating a power law.](image)

Although the Poisson process is widely used to quantify the consequences of human actions, yet, an increasing number of empirical results indicate non-Poisson statistics of the timing of many human actions. Specifically, the interevent time $\tau$ or the response time $\tau_w$ obeys a heavy-tailed distribution. Besides what have been described above, more evidence could be found in the on-line games [Henderson & Nhatti, 2001], the Internet chat [Dewes et al., 2003], FTP requests initiated by individual users [Paxson & Floyd, 1996], timing of printing jobs submitted by users [Harder & Paczuski, 2006], and so on. While the majority of these heavy tails can actually be well approximated by a power law, there also exist debates on the choice of the fitting function in
the case of interevent time distribution in email communications [Stouffer et al., 2005; Barabási et al., 2005]. Another candidate, the log-normal distribution, has been suggested [Stouffer et al., 2005] for describing the non-Poisson temporal statistics of human activities. After choosing a power law as the fitting function, the next problem is how to evaluate its exponent [Goldstein et al., 2004]. In most academic reports, the exponent is directly obtained by using a linear fit in the log-log plot. However, it is found—numerically and theoretically—that this simple linear fit will cause remarkable error in evaluating the exponent [Newman, 2005]. A more accurate method, strongly recommended by us, is to use the (logarithmic) maximum likelihood estimation [Goldstein et al., 2004; Zhou et al., 2008a].

Based on the analytical solution of the Barabási queuing model [Barabási, 2005], Vázquez et al. [2006] claimed the existence of two universality classes for human dynamics, with characteristic power-law exponents being -1 and -1.5, respectively. Email communication, web browsing and library loan belong to the former, while surface mail communication belongs to the latter. However, thus far, there are increasing empirical evidence, as shown above, against the hypothesis of universality classes for human dynamics. For example, we [Zhou et al., 2008a] sort the Netflix users by activity (i.e., the frequency of events of an individual, here meaning the frequency of movie ratings) in a descending order, and then divide this list into twenty groups, each of which has almost the same number of users. As shown in Fig. 12.9, we observe a monotonous relation between the power-law exponent and the mean activity in the group, which suggests that the activity of individuals is one of the key ingredients determining the distribution of interevent times. And the tunable exponents controlled by a single parameter indicate a far different scenario against the discrete universality classes suggested by Vázquez et al. [2006]. A similar relationship between activity and power-law exponent is also reported in the analysis of short-message communication [Hong et al., 2008].

In addition, Vázquez et al. [2006] suggest that the waiting time distribution of tasks could in fact drive the interevent time distribution, and that the waiting time and the interevent time distributions should decay with the same scaling exponent. Our empirical study on
correspondence pattern [Li et al., 2007] supports this claim. However, the real situation should be more complicated, and a solid conclusion is not yet achieved.

Fig. 12.9. Dependence of power-law exponent $\gamma$ of interevent time distribution on mean activity of each group [Zhou et al., 2008a]. Each data point corresponds to one group. All the exponents are obtained by using maximum likelihood estimation [Goldstein et al., 2004] and can pass the Kolmogorov-Smirnov test with threshold quantile 0.9.

In short, abundant and in-depth empirical analyses are required before a completely clear picture about the temporal statistics of human-initiated systems can be drawn.

12.3 The Task-Driven Model

What is the underlying mechanism leading to such a human activity pattern? One potential starting point to arrive at an answer is to consider the queuing of tasks. A person needs to face many works in his/her daily life, such as sending email or surface mail, making telephone call, reading papers, writing articles, and so on. Generally speaking, in our daily life, we are doing these works one by one with some kind of order. In the modeling of human behaviors, we can abstract these activities on human life as tasks. Accordingly, Barabási [2005] proposed a model based on queuing theory.
In this model [Barabási, 2005], an individual is assigned a list with \( L \) tasks. The length of list mimics the capacity of human memory for tasks waiting for execution. At each time step, the individual chooses a task from the list to execute. After being executed, it is removed from the list, and a new task is added. Each task is assigned a priority parameter \( x_i \) \((i = 1, 2, \cdots, L)\), which is randomly generated by a given distribution function \( \eta(x) \). Here the individual is facing three possible selection protocols for these tasks:

The first one is the first-in-first-out (FIFO) protocol, wherein the individual executes the tasks in the order that they were added to the list. This protocol is common in many service-oriented processes [Reynolds, 2003]. In this case, the waiting time of a task are determined by the cumulative executing time of tasks added to the list before it. If the executing time of each task obeys a bounded distribution, the waiting time, representing the length of time steps between the arrival and execution, of tasks is homogeneous.

The second one is to execute the tasks in a random order independent of their priority and arriving time. In this case, the waiting time distribution of tasks is exponential [Gross & Harris, 1985].

The last but most important one is the highest-priority-first (HPF) protocol. In this case, the tasks with highest priority are executed first, even though they are added later in the list. Hence, the tasks with lower priority could wait for a long time before being executed. Such protocol exists widely in human behaviors; for instance, we usually do the most important or the most urgent works first, and then the others.

The Barabási model [2005] focuses on the effect of the HPF protocol. At each time step, it assumes that the individual executes the task with the highest priority with probability \( p \), and executes a randomly chosen task with probability \( 1 - p \). Obviously, if \( p \to 0 \), the model obeys the second protocol, and if \( p \to 1 \), it displays a pure HPF protocol.

Simulation results with \( \eta(x) \) being a uniform distribution in \([0, 1]\) are shown in Fig. 10.10. For \( p \to 0 \) (random chosen protocol), the waiting time distribution \( P(\tau) \) decays exponentially; for \( p \to 1 \) (HPF protocol), it follows a power-law distribution with exponent \(-1\), which agrees well with the empirical data of email communication. The results shown in Fig. 12.10a are generated with list length \( L = 100 \); however, the tail of
the waiting time distribution $P(\tau)$ is independent of $L$, and the observed heavy-tailed property holds even for $L = 2$. Its exact analysis for the case $L = 2$ and different $p$ is given in [Vázquez, 2005; Vázquez et al., 2006; Gabrielli & Caldarelli, 2007]. As shown in Fig. 12.11, it is not necessary to have a long priority list for individuals. If an individual can balance at least two tasks, the heavy-tailed property of the waiting time distribution will emerge. These results imply that the HPF protocol could be an important mechanism leading to the non-Poisson statistics of human dynamics.

![Fig. 12.10. Waiting time distribution predicted by the queuing model [Barabási, 2005]. Priorities were chosen from a uniform distribution $x_i \in [0, 1]$, and numerical simulation monitors a priority list of length $L = 100$ over $T = 10^6$ time steps. a, Log-log plot of the tail of probability $P(\tau)$ that a task spends $t$ time on the list obtained for $p = 0.99999$ (where $p$ is the probability that the highest-priority task will be first executed) corresponding to the deterministic limit of the model. The straight line in the log–log plot has slope -1, in agreement with numerical results and analytical predictions. The data were log-binned, to reduce the uneven statistical fluctuations common in heavy-tailed distributions, a procedure that does not alter the slope of the tail. b, Linear-log plot of the $P(\tau)$ distribution for $p = 0.00001$, corresponding to the random choice limit of the model. The fact that the data follows a straight line on a linear-log plot indicates that $P(\tau)$ decays exponentially.](image)
In further discussions about the Barabási model [Barabási, 2005; Vázquez et al., 2006], a natural extension is introduced: assuming tasks arrives at rate $\lambda$ and executed at rate $\mu$, and allowing the length of the list to change in time. Let $\rho = \lambda/\mu$; obviously, here are three different cases need to be discussed [Vázquez et al., 2006]:

The first case is the subcritical regime, where $\rho < 1$, namely the arrival rate is smaller than the execution rate. In this case, the list will be often empty, and most tasks are executed soon after their arrival, thus the long term waiting time is limited. Simulations indicate that the waiting time distribution exhibits an exponential decay when $\rho \to 0$, and when $\rho \to 1$ it is close to a power-law distribution with exponent $-3/2$ and an exponential cutoff.

The second case is the critical regime, where $\rho = 1$. Here, the length of the list is a random walk in time. Different from the case with a fixed
Towards the Understanding of Human Dynamics

L, the fluctuation in the list length will affect the waiting time distribution. Simulation results indicate that the waiting time distribution obeys a power law with exponent \(-3/2\).

The third case is the supercritical regime, where \(\rho > 1\), namely the arrival rate is larger than the execution rate. Thus the length of the list will grow linearly, and a \(1 - 1/\rho\) fraction of tasks are never executed. Simulations indicate that the waiting time distribution in this case obeys a power law with exponent \(-3/2\), too. A problem is how to understand such a growing list. Let us think over the case of replying regular mails. When we receive a mail, we put it on desk and piled it up with early mails, and we usually choose a mail from the pile to reply. If the received mails are too many and we do not have enough time to reply all of them, the mails in the pile will become more and more. We do not need to remember the list, because all the mails are put on there. Therefore, the list of mails waiting for reply is unlimited. Simulation results for this case are in agreement with the empirical data of surface mail replies of Darwin, Einstein and Freud. All these three great scientists have many mails never being replied. Numerical results are shown in Fig. 12.12 [Vázquez et al., 2006], which is in accordance with the above analysis.

The above model only considers the behavior of an isolated individual. Actually, every person is living in a surrounding society with countless interactions to others; such interactions may affect our activities, such as email communication, phone calling and all collaborated works. In a recent model on human dynamics, based on the queuing theory, the simplest case taking into account the interactions between only two individuals is considered [Oliveira & Vázquez, 2007]. This model only considers two individuals: A and B. Each individual has two kinds of tasks, an interacting task (I) that must be executed in common, and a aggregated non-interacting task (O) that can be executed by the individual himself/herself. Each task is assigned a random priority \(x_{ij}\) (\(i = I, O; j = A, B\)) extracted from a probability density function \(\eta(x)\). At each time step, both agents select a task with highest priority in their list (with length \(L_A\) for A and \(L_B\) for B). If both agents select task I, then it is executed; otherwise each agent executes a task of type O.

As shown in Fig. 12.13, numerical simulations of this model indicate that the interevent time distribution of the interacting task I is close to a
power law, having a wide range of exponents. This result extends the
range of the Barabási model, and highlights a potential way to
understand the pattern of the non-Poisson statistics in the interacting
activities of human.

Fig. 12.12. Waiting time distribution for tasks in the extended queuing model with
continuous priorities [Vázquez et al., 2006]. Numerical simulations were performed as
follows. At each step, the model generates an arrival $\tau_a$ and service time $\tau_s$ from an
exponential distribution with rates $\lambda$ and $\mu$, respectively. If $\tau_a < \tau_s$ or there are no tasks in
the queue, a new task is then added to the queue, with a priority $x \in [0, 1]$ from uniform
distribution, and update the time $t \rightarrow t + \tau_a$. Otherwise, the model removes from the
queue the task with the largest priority and update the time $t \rightarrow t + \tau_s$. The waiting time
distribution is plotted for three $\rho (= \lambda / \mu)$ values: $\rho = 0.9$ (circles), 0.99 (squares) and 0.999
(diamonds). The data has been rescaled to emphasize the scaling behavior $P(\tau_w) = \tau_w^{-3/2} f(\tau_w / \tau_0)$, where $\tau_0 \sim (1 - \rho^{1/2})^{-2}$. The inset shows the distribution of waiting times for
$\rho = 1.1$, after collecting up to $10^5$ (plus) and $10^7$ (diamonds) executed tasks, showing that
the distribution of waiting times has a power law tail even for $\rho > 1$ (supercritical regime).
Note that in this regime a high fraction of tasks are never executed, staying forever on the
priority list whose length increases linearly with time, a fact that is manifested by a shift
to the right of the cutoff of the waiting time distribution.
The motivations of our behaviors are extremely complex [Kentsis, 2006]. Therefore, setting up the rules in modeling by simplifying or coarse-graining the real world is the main way (even the only possible way) in the study of human dynamics [Oliveira & Barabási, 2006]. Although the queuing models do obtain a great success in explaining the heavy tails in human dynamics, they have their own limitations. Actually, the core and fundamental assumption of the queuing models is that the behaviors of humans are treated as executing tasks; however, this assumption could not fit all human behaviors. Some real-world human activities could not be explained by a task-based mechanism, but they could still exhibit similar statistical law (heavy-tailed interevent time distribution), such as browsing webs [Dezsö et al., 2006], watching on-line movies [Zhou et
al., 2008a], playing on-line games [Henderson & Nhatti, 2001], and so on. Clearly, these activities are mainly driven by personal interests, which could not be treated as tasks needing to be executed. In-depth understanding of the non-Poisson statistics in these interest-driven systems requires new insight and ideas beyond the queuing theory [Han et al., 2008].

Before introducing the rules of an interest-driven model (HZW model for short), let us think over the changing process of our interests or appetites on many daily activities. For example, you had eaten a certain kind of food (hamburger, say) and found it tasting good a long time ago. Because for such a long time you have not tasted it, you almost forgot about it. However, after you accidentally eat it again and find it tasty, you will recall its good taste last time, and then your appetite about this food will get stronger; the frequency of you eating hamburger in the future will get higher and higher, until you feel you have eaten too much hamburger. Then your good feeling of hamburger disappears, and your appetite for it also weakens after a long period of time. A similar daily experience can be found in Web browsing. If a person has not browsed the Web in a long time, an accidental browsing event may give him a good feeling and wake up his old interest on Web browsing. Next, during the activity the good feeling is durative, and the frequency of Web browsing in future may increase. Then, if the frequency is too high, he may worry about the habit and reduce those browsing activities. We can also find similar changing processes of interest or appetite in many other daily activities, such as game playing, movie watching, and so on. In short, we usually adjust the frequency of our daily activities according to our interest: greater interest will lead to higher frequency, vice versa. In other words, our interests are history dependent or adaptively changing.

To mimic these daily experiences, the following simple assumptions in the modeling of interest-driven systems are extracted: (1) each activity will change the current interest for a give interest-driven behavior, while the frequency of activities depends on the interest. (2) We assume the interevent time $\tau$ has two thresholds: when $\tau$ is too small (i.e., events happen too frequently), interest will be depressed and thus the interevent time will increase; while if the time gap is too long, we will impose an occurrence to mimic a casual action.
According to the assumptions above, the rules of the HZW model are listed as follows:

1. Time is discrete and labeled by \( t = 0, 1, 2, \ldots \); the occurrence probability of an event at time step \( t \) is denoted by \( r(t) \). The time interval between two consecutive events is called the interevent time and denoted by \( \tau \).

2. If the \((i+1)\)th event occurred at time step \( t \), the value of \( r \) is updated as \( r(t + 1) = a(t)r(t) \), where \( a(t) = a_0 \) if \( \tau_i \leq T_1 \), \( a(t) = a_0^{-1} \) if \( \tau_i \geq T_2 \), and \( a(t) = a(t - 1) \) if \( T_1 < \tau_i < T_2 \).

If no event occurs at time step \( t \), we set \( a(t) = a(t - 1) \); namely \( a(t) \) remains unchanged. Here \( T_1 \) and \( T_2 \) are two thresholds satisfied by \( T_1 \ll T_2 \); \( \tau_i \) denotes the time interval between the \((i+1)\)th and the \(i\)th events; \( a_0 \) is a parameter controlling the changing rate of occurrence probability (\( 0 < a_0 < 1 \)). If no event happens, the value of \( r \) will not change. Clearly, simultaneously enlarging (with the same multiplication factor) \( T_1 \), \( T_2 \) and the minimal perceptible time will not change the statistics of this system. Therefore, without lost of generality, the lower boundary \( T_1 \) is set to 1.

In the simulations, we set the initial interest, \( r_0 = r(0) \), equal to 1, which is also the maximum value of \( r(t) \) in the whole process. As shown in Fig. 12.14, the succession of events predicted by the HZW model exhibits very long inactive periods separated by bursts of rapidly occurring events, and the corresponding \( r(t) \) shows clearly seasonal changing property. Fig. 12.15 presents simulation results with tunable \( T_2 \) and \( a_0 \). Given \( a_0 = 0.5 \), if \( T_2 >> T_1 \), the interevent time distribution generated by the present model displays a power law with exponent -1; while if \( T_2 \) is not sufficiently large, the distribution \( P(\tau) \) will depart from the power law, exhibiting a cutoff in the tail. Correspondingly, given sufficiently large \( T_2 \), the effect of \( a_0 \) is very slight, which can thus be ignored. The power-law exponent -1 can also be analytically obtained under the circumstance of \( T_2 >> T_1 \); detailed mathematical derivation can be found in [Han et al., 2008].

Differing from the queuing models discussed in Section 12.3, this model is driven by the personal interest. In this model, the frequency of events is determined by the interest, while the interest is in turn affected
by the occurrence of events. This intertwining mechanism, similar to that in an active walk\(^4\) [Lam, 2005, 2006], is a generic origin of complexity in many real-life systems. The rules of the model are extracted from the daily experiences of people, and the simulation results agree with many empirical observations, such as the activities of Web browsing [Dezsö et al., 2006]. This model indicates a much simple activity pattern of human behaviors; that is, people could adaptively adjust their interest on a specific behavior (for example, TV watching, Web browsing, on-line game playing, etc.), which leads to a quasi-periodic change of interest, and this quasi-periodic property eventually gives rise to the departure from Poisson statistics. This simple activity pattern could be universal in many human behaviors.

---

\(^4\) In an active walk, a particle (the walker) is coupled to a deformable landscape, influencing each other as the particle moves or the landscape changes.
There are also many other models which are neither based on queuing theory nor the interest-driven mechanism. One important model, proposed by Vázquez [2007], takes into account the memory of past activities by assuming that a person reacts by accelerating or reducing their activity rate based on the perception of their past activity rate. Let $\lambda(t)\,dt$ be the probability the individual performs the activity between time $t$ and $t + dt$. Based on this assumption, the equation for $\lambda(t)\,dt$ can be written as follows:

$$\lambda(t) = a \int_0^t dt \lambda(t) ,$$

(12.3)

where $a > 0$ is the only parameter in this model. When $a = 1$, $\lambda(t) = \lambda(0)$ and the process is stationary. On the other hand, when $a \neq 1$ the process is non-stationary with acceleration ($a > 1$) or reduction ($a < 1$).
Note that Eq. (12.3) includes a latent assumption on the starting time \( t = 0 \). As indicated in [Vázquez, 2007], it is a reflection of our bounded memory, meaning that we do not remember or do not consider what took place before that time. For instance, we usually check for new emails every day after arriving at work no matter what we did the day before.

From Eq. (12.3), one can obtain the function of interevent time distribution via mathematical derivations; details are given in [Vázquez, 2007]. The general conclusion of the model is as follows:

When \( a = 1 \), the resulting interevent time distribution is an exponential function. Let \( \tau_0 = (a \lambda_0)^{-1} \), where \( \lambda_0 \) is the mean number of events in the considered time period \( T \). When \( a > 1 \) (acceleration) and \( \tau_0 \ll \tau < T \), the interevent time distribution generated by the model is close to a power law with exponent \( -2 - (a - 1)^{-1} \). On the other hand, when \( 0 < a < 1/2 \) (reduction) and \( \tau << \tau_0 \), its interevent time distribution is close to a power law with exponent \( -1 + a/(1 - a) \).

Comparing with the cumulative number of regular mails sent out by Darwin and Einstein as a function of time and the interevent time distribution of these regular mails, results generated by this simple model are in agreement with the empirical data, as shown in [Vázquez, 2007].

12.5 Discussion and Conclusion

The study of human dynamics, still in its infancy, is currently attracting a lot of attention due to their theoretical and practical significances. Extensive empirical evidence demonstrates non-Poisson statistics in temporal human activities, in opposite to what is expected from the traditional hypothesis [Gross & Harris, 1985] which assumes uniform and stationary timing of human actions. Actually, the majority of real applications of queuing theory in the past are based on the assumption of a Poisson process of event occurrence [Newell, 1982]. From now on, we may have to complement queuing theory by taking into account the possibility of a power-law interevent time distribution. Since the second moment of a power-law distribution with absolute value of exponent smaller than 3 is not convergent, many previous conclusions in queuing theory are invalid when a heterogeneous interevent time distribution is considered. Mathematically speaking, the new findings on non-Poisson...
Towards the Understanding of Human Dynamics

temporal statistics give rise to an important open question in queuing theory. Theoretical interests aside, these new findings have significant potential in practical applications [Barabási, 2005]. For example, in-depth understanding of human activity pattern is indispensable in modeling social structures [Zhu et al., 2008] and financial behaviors [Caldarelli et al., 1997], and is also crucial for better resource allocation and pricing plans for telephone companies [Barabási, 2005].

Thus far, many models aiming at the explanation of the origin of heavy-tailed human activity pattern have been proposed. The majority of previous models before our HZW model are based on queuing theory. Yet, not all human-initiated systems are driven by some tasks. Besides the task-driven mechanism underlying the queuing theory, some other possible origins of human behavior—such as interest [Han et al., 2008] and memory [Vázquez, 2007]—are also highlighted recently. As stated by Kentsis [2006], there are countless number of ingredients affecting human behaviors, and for most of them, we do not know their impacts. We believe, in the near future, more theoretical models will be proposed to reveal the effects of task deadline, task optimization protocol, human seasonality, social interactions, and so on.

Another important issue is how the non-Poisson temporal statistics affect the relative dynamical processes taking place in human-initiated systems. For example, epidemic spreading of diseases, such as AIDS, influenza and SARS, is driven by social contacts between infected and susceptible persons. At the macroscopic level, the effects of social structures (i.e., epidemic contact networks) have been extensively investigated (see review article [Zhou et al., 2006a] and references therein). However, these works lack a serious consideration of the microscopic factor, namely the temporal statistics of epidemic contacts. Actually, prior works either assume the contact frequency of a given person being proportional to his/her social connectivity [Pastor-Satorras & Vespignani, 2003] or assume the same contact frequency for all infected persons [Zhou et al., 2006b]. Recently, based on the statistical reports of email worms, Vázquez et al. [2007] studied the impact of non-Poisson activity patterns on epidemic spreading processes; this work provides a starting point in understanding the role of individual activity pattern in aggregated dynamics.
In addition, heterogeneous traffic-load distribution as well as long-
range correlation embedded in the traffic-load time series have been
observed in many human-initiated systems, such as the Internet traffic
[Park, 2000; Zhou et al., 2006] and air transportation [Guimerà et al.,
2005; Liu & Zhou, 2007]. The observed heavy-tailed timing of email
communication [Barabási et al., 2005] and Web browsing [Dezsö et al.,
2006] as well as long range human travel [Brockmann et al., 2006] may
contribute to non-trivial phenomena. Furthermore, some social dynamics
may also be highly affected by human activity patterns [Castellano et al.,
2007].

All previous studies on human dynamics focus on the distributions of
interevent time and response time. However, its methodology—
extracting statistical laws from historical records of human activities—is
not limited to this issue. For instance, we can also use some of those data
sets to quantify the herd behavior of an individual; that is, a person
follows the opinions of the majority of people in his/her social
surrounding in an irrational way. Although the herd behavior has found
its significant impact on financial markets [Bikhchandani & Sharma,
2000], it is very hard to be quantified outside the laboratorial surrounding
[Asch, 1955]. In many Web-based recommender systems, such as the on-
line movie-sharing system [Zhou et al., 2008b], the user’s records
contain not only the time he/she watch the movies, but also his/her
opinion (i.e., ratings) on those movies. Similar records can also be found
in the systems of Web-based trading, book-sharing, music-sharing, and
so on. Since before the vote is cast by the user, he/she can see previous
ratings assigned by others to this item, the herd behavior may occur.
Quantitatively uncovering the latent bias of human opinion is crucial for
better design of recommender systems [Zhang et al., 2007; Zhou et al.,
2007, 2008a, 2008b, 2008c; Ren, 2008].

References
Nature 435, 207-211.


Reynolds, P. [2003] Call Center Staffing (Call Center School, Lebanon, TN).
Vázquez, A. [2007] “Impact of memory on human dynamics,” Physica A 373,
Towards the Understanding of Human Dynamics

747-752.


