Supplementary Information for

Emergence of scaling in human-interest dynamics

Zhi-Dan Zhao,[1,2] Zimo Yang,[1] Zike Zhang,[1,3] Tao Zhou,[1] Zi-Gang Huang,[2,4] and Ying-Cheng Lai[2,*]

[1] *Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China.*
[2] *School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA*
[3] *Institute for Information Economy, Hangzhou Normal University, Hangzhou 310036, China*
[4] *Institute of Computational Physics and Complex Systems, Lanzhou University, Lanzhou, Gansu 730000, China*

(Dated: October 1, 2013)

Contents

*email: Ying-Cheng.Lai@asu.edu

SUPPLEMENTARY NOTE 1: CHARACTERIZING COLLECTIVE HUMAN-INTEREST ACTIVITY

We investigate the collective human interest dynamics by averaging the patterns of many individuals. To provide direct evidence for non-Markovian type of activity patterns, we extract the size l of the consecutive interest length of users in the three systems explored, as shown in Figs. S1(a-c). We see that the distributions of l can all be well described by a power-law, which is consistent with the interest-activity pattern exhibited by any typical individual. This result suggests the algebraic, heavy-tailed type of distribution of the time interval that one maintains the same interest as a universal phenomenon. The distributions of the return time τ at the collective level also exhibit power-law scaling, as shown in Figs. S1(d-f). The collective rank distributions are demonstrated in Figs. S1(g-i), where the r-th interest means the average value of the r-th most browsed interest of all individuals. We also calculate the probability density function $p(\bar{i},\bar{j})$ of the users for the three systems where, for each system, $p(\bar{i},\bar{j}) \equiv \sum_{n=1}^{N_u} p(i,j)/N_u$, and N_u is the number of users in the system. If an individual has not demonstrated more than i (or j) distinct interests, we set $p(i,j) = 0$. From Figs. S1(j-l), we observe that, at the collective level, the interest-transition patterns exhibit the same features as those at the individual level.

SUPPLEMENTARY NOTE 2: CLICK DWELLING TIME VERSUS REAL DWELLING TIME

In our work, the three datasets record the click events of users, and our analysis of human-interest dynamics is thus based on the click behaviors. The dwelling time in one given interest is measured by the length of continuous-click events, which we named click-dwelling time. However, it is also interesting to know the dwelling behavior of user in real time. To compare the click-dwelling time with the real dwelling time, we plot the relation between the real dwelling time t (in seconds) and the corresponding average click-dwelling time $\langle l \rangle$ for Douban and Taobao, as shown in Figs. S2(a-b), respectively. Take the data of Douban in Fig. S2(a) as an example. We observe a positive correlation between $\langle l \rangle$ and t (with saturation) before 3000 seconds, indicating that, in this case, the individual's continuous clicks focused on one interest has the upper limit of approximately one hour. Similar behaviors have been observed for majority of the individuals. Due to the positive correlation between $\langle l \rangle$ and t on average, the conclusions from our work based on the click-dwelling time would be qualitatively the same as those based on the real time.

Nonetheless, for time longer than this upper limit, deviation from the behavior in Figs. S2(a-b) can occur. For example, the subgraphs in Figs. S2(a-b) show the click-dwelling time versus the actual time on a logarithmic plot for a time span longer than one hour, where the abnormal decrease in $\langle l \rangle$ can be attributed to the pseudo-continuous clicks in data introduced by the two identical clicks before and after a long resting time (sack time). In this case, the dwelling time is overestimated while the number of clicks l is small. Majority of the cases in data with dwelling time over the circadian human behavior (24h or $8 * 10^4$s) are found to be related to the pseudo-continuous clicks. As the fraction of pseudo-continuous clicks increases with the real dwelling time, the average value $\langle l \rangle$ decreases. Thus the abnormal decrease of $\langle l \rangle$ toward the tail part (merely caused by data splitting) is not evidence against the positive correlation. between t and l.

To provide further support for the equivalence between the click-dwelling time and the actual time, we compare their probability distributions, as shown in Figs. S3(a,b) for Douban and Taobao, respectively. In the real-time representation, the power-law scaling of human-interest dynamics can also be observed in the marked region (light green), with deviation at both ends. In general, we find that the statistics based on the click time are more robust. (For the MPR data, due to the intrinsic low-time resolution, it is not feasible to study the statistics based on the real time.)

SUPPLEMENTARY NOTE 3: CONSISTENCY OF SCALING EXPONENTS ACROSS INDIVIDUALS

To examine the consistency of human-interest activities across different individuals, we calculate the ranges of the scaling exponents α, β, and γ. As shown in Figs. S4(a-c), we obtain $\alpha = 2.20 \pm 0.46$, 2.57 ± 0.55 and 2.92 ± 0.39 for all users in Douban, Taobao and MPR, respectively. The robustness of the power-law fit can be assessed by the standard Kolmogorov-Smirnov test [1]. Figures S4(d-f) show the distributions of D, the quantifier of the Kolmogorov-Smirnov test, for the three online systems in Figs. S4(a-c), respectively. We observe that in all three cases, the values of D are narrowly distributed, indicating the robustness and consistency of the power-law fitting for $P(l)$ across vast majority of the individuals in these systems. The corresponding behaviors for the scaling exponents β and γ are shown in Figs. S5 and S6, respectively.

SUPPLEMENTARY NOTE 4: NECESSITY OF BOTH PREFERENTIAL RETURN AND INERTIA IN MODELING HUMAN-INTEREST DYNAMICS

In the main text, we state that the basic dynamical ingredients of human-interest dynamics are (1) preferential return associate with the hopping processes among different interests, (2) inertia characterizing the tendency to dwell in the same interest, and (3) exploration of new interest. We can establish the necessity of preferential return and inertia by comparing the outcomes of our model with those from two modified models: one without inertia (denoted as NI) and another without preferential return (denoted as NPR). In the NI model, the distribution of the consecutive interest interval is no longer a power law, as shown in Fig. S7. In the NPR model, individuals select their interests with equal probability. In this case, the distribution of the consecutive interest intervals continues to exhibit the power-law feature, as shown in Fig. S7. The corresponding patterns of the transition probability are shown in Figs. S8(a-d), where Fig. S8(a) is for the NI model, and Figs. S8(b-d) are for the NPR model. We see that for the NI model, the diagonal elements of the transition probability no longer dominate as compared with off-diagonal elements [Fig. S8(a)]. For the NPR model, the off-diagonal elements of the transition probability are essentially uniform [Figs. S8(b-d)]. These behaviors are contradictory to those exhibited by real data, implying that both preferential return and inertia are necessary ingredients in any quantitative description and modeling of the human-interest dynamics.

SUPPLEMENTARY NOTE 5: THEORETICAL ANALYSIS OF SCALING LAWS ASSOCIATED WITH HUMAN-INTEREST DYNAMICS

Our extensive analysis of real data establishes inertia, preferential return, and exploration of new interest as the basic ingredients in modeling human-interest dynamics. Mathematically, inertia characterizes the tendency to dwell in the same interest, which can be assessed from the *click-event series* of browsing data and described by an excited random-walk process [2]. Preferential return among visited interests, and exploration of new interests can be modeled by the *hopping-event series*, which is reduced from the *click-event series* by merging a cluster of clicks on an identical interest. In the hopping-event series, interests are denoted by the rank in the order of the hopping, i.e., the first-passage hopping number.

For the exploration state, our empirical analysis indicates that the exploration of new interest is algebraically related to the number of hopping n, as shown in Fig. S9. In particular, the number of interests already visited by a given user, denoted by S, increases by one at the n-th hopping with exploring probability $p_{new} = \rho n^{-\lambda}$, where $\lambda \approx 0.5$ and the parameter ρ can be obtained from Fig. S9 through fitting. The form of p_{new} implies that the growth rate of S is reduced as the number of hopping n increases. Under the continuous approximation, we have

$$\frac{dS}{dn} = p_{new} = \rho n^{-\lambda}, \tag{S1}$$

from which we get the dependence of S on the number of hopping n as

$$S \sim n^{-\lambda+1}. \tag{S2}$$

As shown in Fig. S9, the dependence of S on n calculated from real data exhibits precisely this dependence with the scaling exponent $-\lambda + 1 \simeq 0.5$.

For preferential return, which is the complementary event to "exploration" at the hopping-event level occurring with probability $1 - p_{new}$, we notice that users revisit one particular interest according to the prior visiting probability:

$$p_i = \frac{\omega_i}{\sum_{j=1}^{S} \omega_j} \equiv \Pi(\omega_i), \tag{S3}$$

where ω_i is the frequency of visiting interest i, i.e., the accumulated dwelling time (total number of click-events) at interest i.

In the following, we provide a detailed theoretical analysis of our model of human-interest dynamics in terms of A) the distribution $P(l)$ of consecutive interest, B) the distribution of the interest-return interval $P(\tau)$, and C) the rank distribution of frequencies of visiting different interest categories.

Probability distribution of interest interval l

We can see from the click-event series of browsing data that the phenomenon of dwelling in the same interest is quite common. As described in the main text [Figs. 1(a-c)], the length of the dwelling time l, the *consecutive interest interval*, obeys a power-law scaling distribution. For simplicity, we assume that the dwelling processes in different interests are similar, i.e., interests share the same distribution $P(l)$. This approximation holds for majority of the interests, although data reveals that the hub interests

tend to have larger values of l. Here we provide a theoretical derivation of this power-law scaling relation. The basic idea is that, dwelling in any particular interest can be treated as an excited random-walk (ERW) process [2], which ends when the walker returns to the origin. The distribution $P(l)$ can then be obtained as $P(l) \sim l^{-2+\zeta}$, where ζ is the biased probability for the walker to move.

Here we adopt the detailed analysis of ERW in Ref. [2], in which the walker at a site that contains cookies eats one cookie and then hops to the right with probability ζ and to the left with probability $1-\zeta$. If the walker hops onto an empty site, there is no bias. To gain insights, we examine an unbiased random walk in a finite interval $[0, x]$ with absorbing boundaries. The Laplace transforms of the first-passage probabilities for a random walk that starts at x_0 to exit at the left and the right edges of interval $[0, x]$, in the continuum limit, are [2], respectively,

$$\ell_x(x_0, s) = \frac{\sinh\sqrt{2s}(x - x_0)}{\sinh\sqrt{2s}x}, \;\; r_x(x_0, s) = \frac{\sinh\sqrt{2s}x_0}{\sinh\sqrt{2s}x}, \tag{S4}$$

with s the complex variable in the Laplace domain.

Consider the case where the walk arrives at site x for the first time at time t_0. Let $L_x(t)$ be the probability of arriving at site 0 for the first time at time $t_0 + t$ without visiting $x + 1$, and $R_x(t)$ be the probability of arriving at $x + 1$ for the first time at time $t_0 + t$, without visiting the site 0. In the Laplace domain, the two first-passage probabilities are

$$\begin{aligned} L_x &= (1-\zeta)(1-s)\ell_{x+1}(x-1, s), \;\; \text{and} \\ R_x &= \zeta(1-s) + (1-\zeta)(1-s)r_{x+1}(x-1, s). \end{aligned} \tag{S5}$$

For L_x, the site $x+1$ is not visited, so the first step must be to the left, hence the factor $(1-\zeta)(1-s)$. The factor $\ell_{x+1}(x-1, s)$ is then the first-passage probability to the origin from $x-1$, without visiting $x+1$. Similarly, for R_x, the term $(1-\zeta)(1-s)$ accounts for the transition $x \to x+1$ in a single step. The factor $(1-\zeta)(1-s)$ in the second term accounts for a single step to the left after which the walk is at $x-1$ while the last cookie is at $x+1$. Then the factor $r_{x+1}(x-1, s)$ gives the first-passage probability to $x+1$ without visiting the origin, starting from the site $x-1$.

The Laplace transform of the first-passage probability to the origin can be obtained, which begins at $x_0 = 1$ with the system initially full of cookies, by summing over all paths that contain $0, 1, 2, ...$ first passages to the last cookie before the origin is reached. This gives

$$F(s) = L_1 + R_1(L_2 + R_2(L_3 + R_3(L_4 + \cdots = L_1 + R_1L_2 + R_1R_2L_3 + R_1R_2R_3L_4 + \cdots . \tag{S6}$$

In the limit $s = 0$, this Laplace transform is the integral of the first-passage probability over all time or, equivalently, the probability of eventual return to the origin. It can be verified that this return probability equals one. In particular, the auxiliary probabilities R_x and L_x are

$$R_x = \zeta + (1-\zeta)\frac{x-1}{x+1} = 1 - \frac{2(1-\zeta)}{x+1} \;\; \text{and} \;\; L_x = \frac{2(1-\zeta)}{x+1} = 1 - R_x. \tag{S7}$$

We then have

$$F(w=0) = 1 - R_1 + R_1(1-R_2) + R_1R_2(1-R_3) + R_1R_2R_3(1-R_4) + \cdots . \tag{S8}$$

This expression is equal to one for all $\zeta < 1$, while for the extreme case of $\zeta = 1$, one gets $F(s=0) = 0$.

The sum of the first m terms in Eq. (S6) can be expressed as

$$F_m(s) = \sum_{x=1}^{m} L_x \prod_{j=1}^{x-1} R_j. \tag{S9}$$

In the limit $s \to 0$, according to Eq. (S7), we have,

$$F_m(s=0) = \frac{2(1-\zeta)}{\Gamma[2-2(1-\zeta)]} \sum_{x=1}^{m} \frac{\Gamma[x+1-2(1-\zeta)]}{\Gamma(x+2)}, \tag{S10}$$

which approaches unity as $m \to \infty$ [3]. The $s = 0$ forms R_x and L_x from Eq. (S7) remain applicable for m on the order of $1/\sqrt{s}$ or less. The main contribution in Eq. (S9) to $F(s)$ thus is assumed to come from terms with $m < 1/\sqrt{s}$ [2]. That is,

$$F(s) \approx \frac{2(1-\zeta)}{\Gamma[2-2(1-\zeta)]} \sum_{x=1}^{1/\sqrt{s}} \frac{\Gamma[x+1-2(1-\zeta)]}{\Gamma(x+2)}. \tag{S11}$$

This finite sum can be written as the difference between the infinite sum (which equals 1) and the sum from $1/\sqrt{s}$ to ∞. In the latter sum, we have $x \gg 1$ for all terms as $s \to \infty$. We can thus replace the ratio of the gamma functions by $x^{-2(1-\zeta)-1}$ and the sum by an integral. These approximations lead to

$$1 - F(s) \sim \int_{1/\sqrt{s}}^{\infty} dx\, x^{-2(1-\zeta)-1} \sim s^{(1-\zeta)}. \tag{S12}$$

This result for $F(s)$ implies that for large times the first-passage probability to the origin indeed exhibits a power-law behavior as $F(t) \sim t^{-2+\zeta}$. Thus the dwelling time l of user in one given interest, which is determined by the ERW process as analyzed above, obeys the distribution $P(l) \sim l^{-2+\zeta}$.

Probability distribution of return interval τ

Our analysis of the real data indicates that a typical user tends to revisit frequently the interests they have explored before, implying some kind of memory effect in user's online activities. The scaling behavior of the return time is thus an important quantity to characterize the human-interest dynamics. Here, we provide a framework to analyze the distribution of the return time interval τ, in terms of the distribution $P(l)$ of the interest interval we have discussed.

We consider the case where a user returns to one given interest I_i after τ click events since the last click, where the click events is composed of m inertial processes in interests (other than I_i) introduced by m hopping events. The probability for interest I_i to be revisited after τ clicks can be expressed as

$$\mathbb{P}_i(\tau) = \sum_{m=1}^{\tau} F_{ii}(m) P_m(\tau), \tag{S13}$$

where $F_{ii}(m)$ stands for the probability at the hopping-event level that the user returns to interest I_i for the first time after m hops among other interests. The function $F_{ii}(m)$ is nothing but the i-th diagonal element of the matrix $\mathbf{F} = (\mathbf{T}\mathbf{I}')^{m-1}\mathbf{T}$, where $\mathbf{I}'$ is the identity matrix but with the i-th diagonal element being zero, as the i-th interest is set to be the sink in the first-return process. The matrix $\mathbf{T}$ is the transition probability matrix among interests at the hopping-event level, with every row summing to 1. The quantity $P_m(\tau)$ in Eq. (S13) is the probability for user to perform τ clicks in these m hopping events, which is defined as the following m-fold convolution of the function $P(l)$:

$$P_m(\tau) = \sum_{l_1, l_2 \cdots l_{m-1}} P(l_1) \cdot P(l_2) \cdots P(\tau - \sum_{j=1}^{m-1} l_j). \tag{S14}$$

Equation (S13) shows the probability for the first return click-step to be τ for a given interest. Then, regarding all the interests as a whole, the probability for the first return click-step to be τ is given by

$$\mathcal{P}(\tau) = \sum_i p_i \mathbb{P}_i(\tau), \tag{S15}$$

where p_i is the probability for the i-th interest to be selected in the preferential return processes defined in Eq. (S3). It appears not feasible to obtain analytic formulas for the probabilities $\mathbb{P}_i(\tau)$ and $\mathcal{P}(\tau)$. It may be possible to calculate these probabilities numerically, provided that the distribution $P(l)$ at the click-event level and the transition-probability matrix $\mathbf{T}$ at the hopping-event level are available.

Frequency-based rank of interests

As discussed, the rank distribution of interests based on the visiting frequency ω_i obeys power-law scaling [Figs. S1(g-i)]. Here we analyze this scaling from the asymptotic behavior in the frequency evolution.

For convenience, we denote the hopping number n to be n_i if at that hopping the interest I_i is explored. The frequency ω_i of interest I_i at its first exploration, which occurs at the n-th hopping of user, is the corresponding dwelling time denoted by l_{n_i}, where l_{n_i} obeys the distribution $P(l)$ and is the initial value of ω_i. Subsequently, the frequency ω_i increases as soon as interest I_i is selected again through preferential return that occurs with the probability $1 - p_{new}$ at the hopping-event level, in which I_i is selected with probability p_i. Then, in general, the evolution of the frequency ω_i is governed by

$$\frac{d\omega_i}{dn} = (1 - p_{new}) p_i l_n, \tag{S16}$$

with l_n being the dwelling time at the n-th hopping. We see from Eq. (S1) that, for $\lambda > 0$, $p_{new} \to 0$ as n increases. It is thus useful to consider the following asymptotic equation:

$$\frac{d\omega_i}{dn} = p_i l_n, \tag{S17}$$

with initial condition $\omega_i(n_i) = l_{n_i}$.

From the definition of p_i in Eq. (S3) and the approximation that $P(l)$ assumes the same form for different interests, we get the expected value of $p_i l_n$ as

$$\overline{p_i l_n} = \frac{m_i \bar{l}}{\sum_{j=1}^{S} m_j \bar{l}} \bar{l} \doteq \frac{\omega_i}{n}, \tag{S18}$$

where m_i denotes the total number of hopping towards the i-th interest, with $\sum_{i=1}^{S} m_i = n$. Then the growth rate of frequency ω_i in the long run is

$$\frac{d\omega_i}{dn} \doteq \frac{\omega_i}{n}, \tag{S19}$$

and we have the approximate relation $\omega_i \doteq C_i n$. Using the initial condition $\omega_i(n_i) = l_{n_i}$, we get the parameter $C_i = l_{n_i}/n_i$, and finally obtain

$$\omega_i \doteq \frac{l_{n_i}}{n_i} n, \tag{S20}$$

indicating that the earlier an interest is explored (smaller n_i), the higher visiting frequency ω_i it will have (namely, $\omega_i \sim {n_i}^{-1}$). Consequently, the frequency-based ranking of interests I, denoted by $r(I)$, is nothing but the rank in the order of first exploring time, i.e., $r(I_i) = i$. Also, the number S of visited interests at the n_i-th hopping (when the i-th interest has just been explored) is $S(n_i) = i$. Then we get $r(I_i) = S(n_i)$. From the function $S(n)$ in Eq. (S2), we find

$$n_i \sim r(I_i)^{1/(1-\lambda)} \tag{S21}$$

Finally, from Eq. (S20), we get the approximate relation between the frequency ω_i and the frequency-based rank r as

$$\omega_i \sim n_i^{-1} \sim r^{-1/(1-\lambda)}. \tag{S22}$$

This theoretical result shows that the scaling exponent of the frequency rank denoted by $-\gamma$ in Fig. S1(g-i) is related to the scaling exponent $-\lambda$ associated with the exploration phase [Eq. (S1)] as $\gamma = 1/(1-\lambda)$. The values of exponents γ (in Fig. S1) and λ (in Fig. S9) calculated from real data validate our theory.

In each preferential return hopping, the frequency of the selected interest ranked at i will increase by approximately $\Delta\omega_i = l/(n\bar{l})$. For the case that $\Delta\omega_i$ is larger than the frequency difference of the $(i-1)$-th and the i-th interests (denoted by $\delta\omega_i \equiv \omega_{i-1} - \omega_i$), the rank is altered. Thus, the stability of frequency rank requires $l < \delta\omega_i n\bar{l}$. This means that, when a user has a large dwelling time l by chance, or hops to an interest at the tail of the rank (where the frequency difference $\delta\omega_i$ is small), or the hopping number n is small (implying somewhat immature browsing experience of the user), the rank is less stable.

SUPPLEMENTARY NOTE 6: FINITE SIZE EFFECT OF S

To demonstrate the finite-size effect, we calculate f_r with different S values for both real data [Fig. S10(a)] and model [Fig. S10(b)]. We observe that both show the curve at the bottom of the distribution. The range in the real data is from 10 to 40, and in model it is between 30 and 1000.

SUPPLEMENTARY NOTE 7: ENTROPY AND PREDICTABILITY

Entropy is the most fundamental quantity characterizing the degree of predictability from time series [4, 5]. A lower value of the entropy implies higher predictability, and vice versa. Here we discuss how to measure the actual interest sequence entropy of individuals over their past history. For this purpose we use the estimator based on the Lempel-Ziv data compression [6, 7], which is known to rapidly converge to the real entropy. Given an interest series with N_a actions, the entropy can be estimated by

$$E \approx (\frac{1}{N_a} \sum_j \Lambda_j)^{-1} \ln N_a. \tag{S23}$$

where Λ_j is the length of the shortest substring starting at position j, which does not previously appear from position 1 to $j-1$. Kontoyianis [7] *et al.* proved that E converges to the actual entropy as N_a approaches infinity. For Douban, Taobao and MPR, the distributions of entropy values are shown in Figs. S11(a-c), respectively.

To predict an individual's future interests, we measure the probability Π subject to Fano's [4, 5, 8] inequality. That is, if an individual with entropy E moves between S interests, his/her predictability $\Pi^{\max}(E, S)$ is determined by

$$E = -[\Pi^{\max}\log_2\Pi^{\max} + (1-\Pi^{\max})\log_2(1-\Pi^{\max})] + (1-\Pi^{\max})\log_2(S-1). \tag{S24}$$

We calculate $\Pi^{\max}$ separately for each user in all three online systems. We find the behavior that $P(\Pi)$ has large values, peaking about $\Pi = 0.93$, 0.84 and 0.83 for Douban, Taobao and MPR, as shown in Figs. S11(d-f), respectively. These highly localized distributions indicate that the individual's interest trajectory is not totally random, and in fact has an unexpectedly high degree of potential predictability.

---

[1] Corder, G. W. & Foreman, D. I. *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* (Wiley, 2009).

[2] Antal, T. & Redner, S. The excited random walk in one dimension. *J Phys. A* **38**, 2555 (2005).

[3] Graham, R. L., Knuth, D. E. & Patashnik, O. *Concrete mathematics: a foundation for computer science*, vol. 2 (Addison-Wesley Reading, 1989).

[4] Navet, N. & Chen, S. H. On predictability and profitability: Would gp induced trading rules be sensitive to the observed entropy of time series? *Natural Computing in Computational Finance* 197–210 (2008).

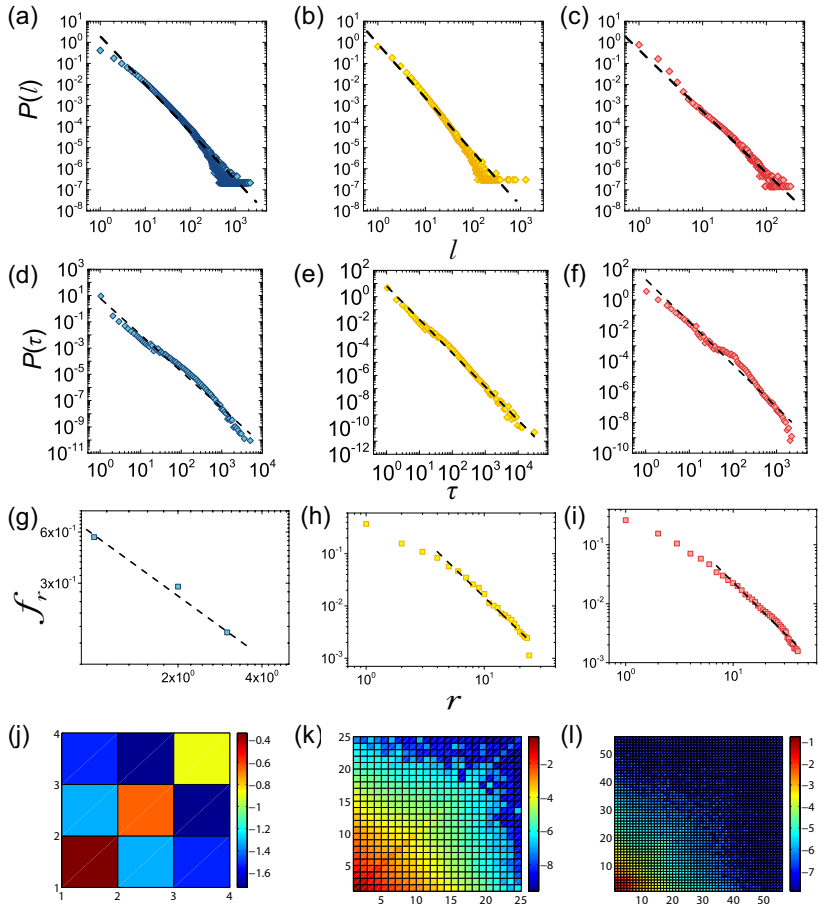
[5] Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).

[6] Ziv, J. & Lempel, A. A universal algorithm for sequential data compression. *IEEE T. Inform. Theory* **23**, 337–343 (1977).

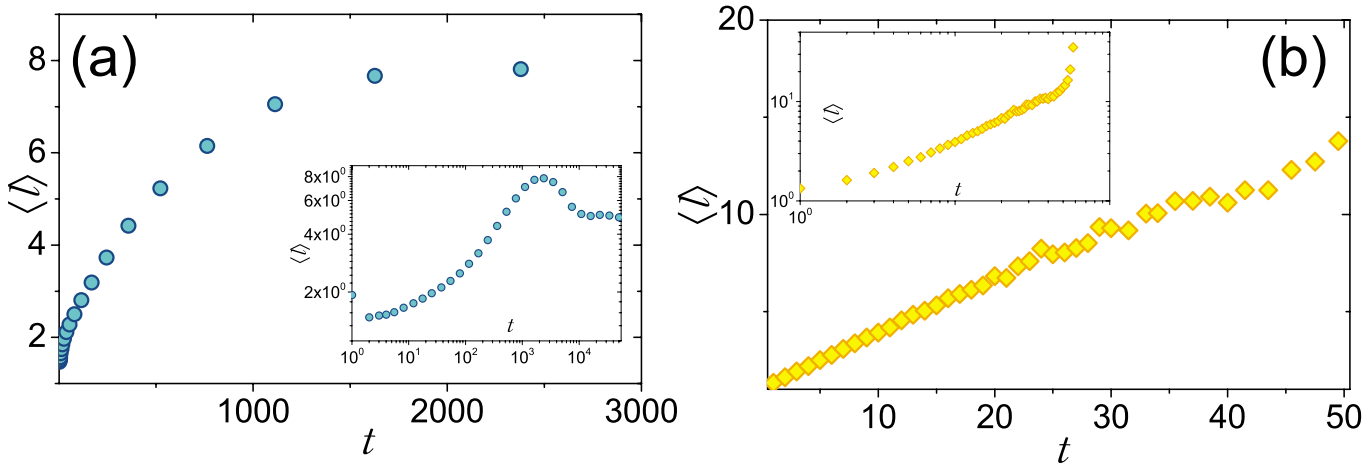
[7] Kontoyiannis, I., Algoet, P. H., Suhov, Y. M. & Wyner, A. J. Nonparametric entropy estimation for stationary processes and random fields, with applications to english text. *IEEE T. Inform. Theory* **44**, 1319–1327 (1998).

[8] Fano, R. & Hawkins, D. Transmission of information: A statistical theory of communications. *Am. J. Phys.* **29**, 793 (1961).

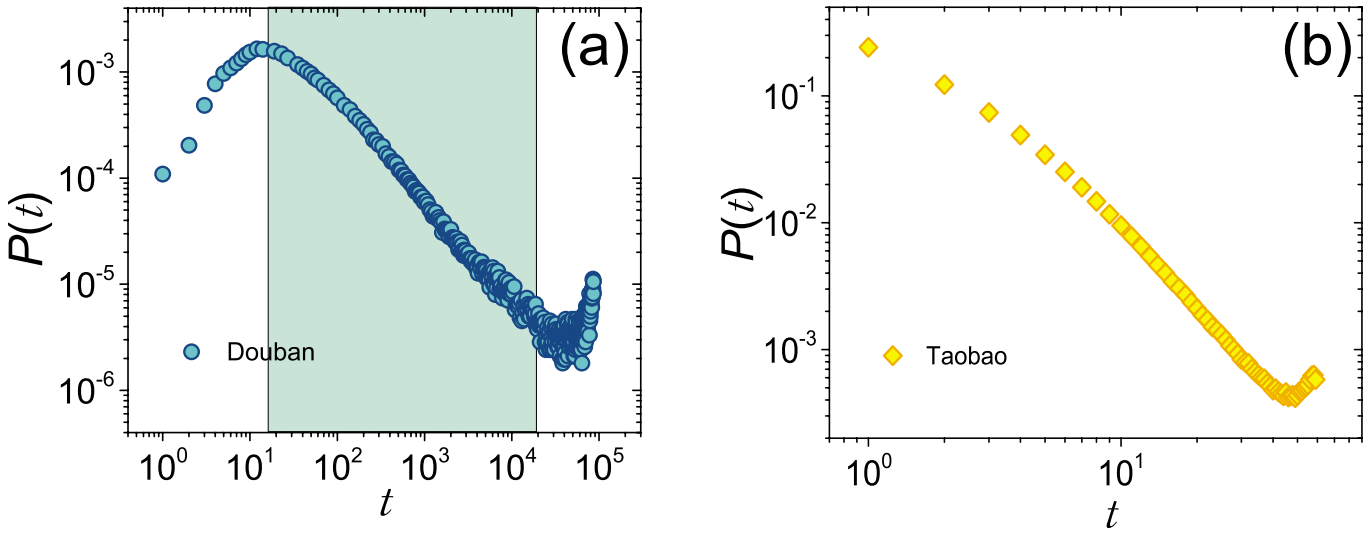
SUPPLEMENTARY FIGURES



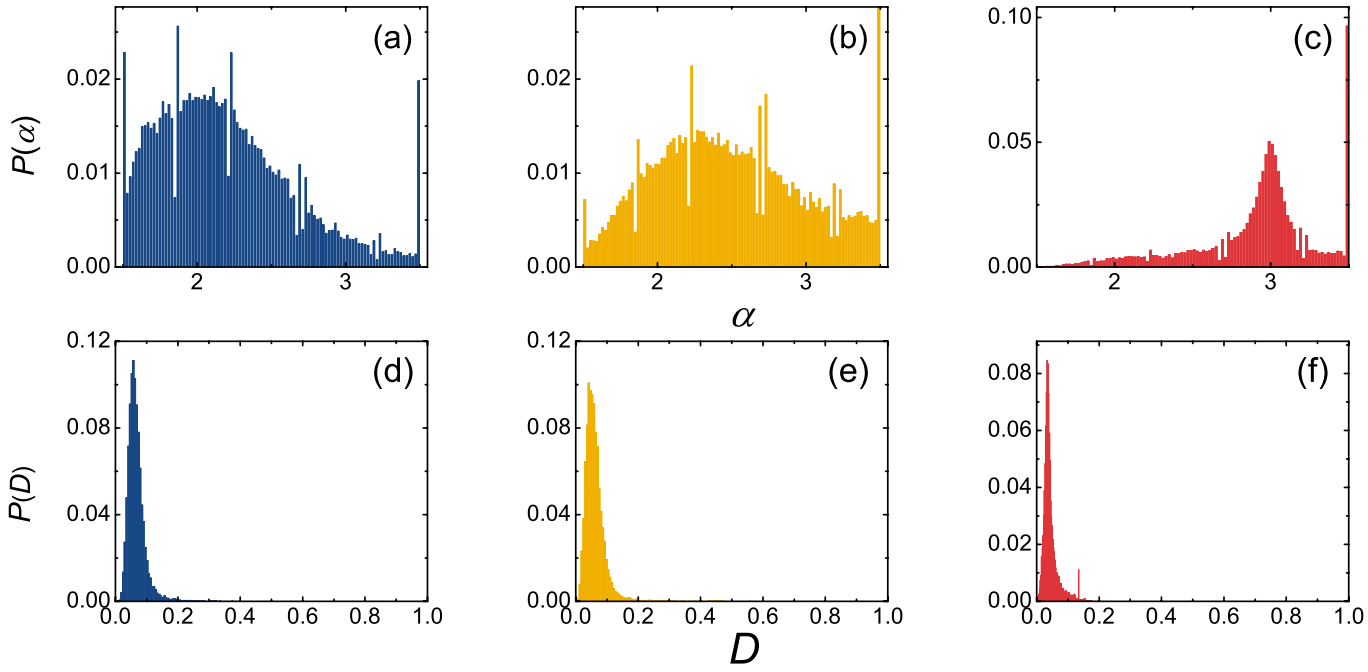
Supplementary Figure S1: Patterns of collective human-interest dynamics. (a-c) Distributions of consecutive interest duration l for users in Douban, Taobao and MPR, respectively, where the parameters are $N_u = 21,148$ and $N_a \geq 300$ for (a), $N_u = 34,330$ and $N_a \geq 100$ for (b), and $N_u = 19,067$ and $N_a \geq 100$ for (c) (N_a is the number of individual click actions). The distributions of l in all three systems can be well fitted by $P(l) = l^{-\alpha}$, with exponents $\alpha = 2.55 \pm 0.02$, $\alpha = 3.00 \pm 0.02$, and $\alpha = 3.20 \pm 0.05$, respectively. (d-f) For the data sets in (a-c), respectively, power-law distributions ($\tau^{-\beta}$) of the time τ taken to revisit the same interest. The values of the fitted exponent β are approximately 2.78 ± 0.04, 2.55 ± 0.02, and 2.78 ± 0.05 for Douban, Taobao and MPR, respectively. (g-i) For the systems (a-c), respectively, the probability f_r for individuals to visit the r-th most browsed interests. This interest rank distribution follows a power-law scaling: $f_r \propto r^{-\gamma}$, where the values of the exponent are $\gamma = 2.18 \pm 0.08$ for Taobao and $\gamma = 1.80 \pm 0.04$ for MPR. (The dash line in Fig. S1(g) is for eye guidance.) (j-l) For the systems in (a-c), respectively, collective mean probability density functions $p(\bar{i}, \bar{j})$ characterizing the interest transitions, where a color represents a certain magnitude of $p(\bar{i}, \bar{j})$ on a logarithmic scale.



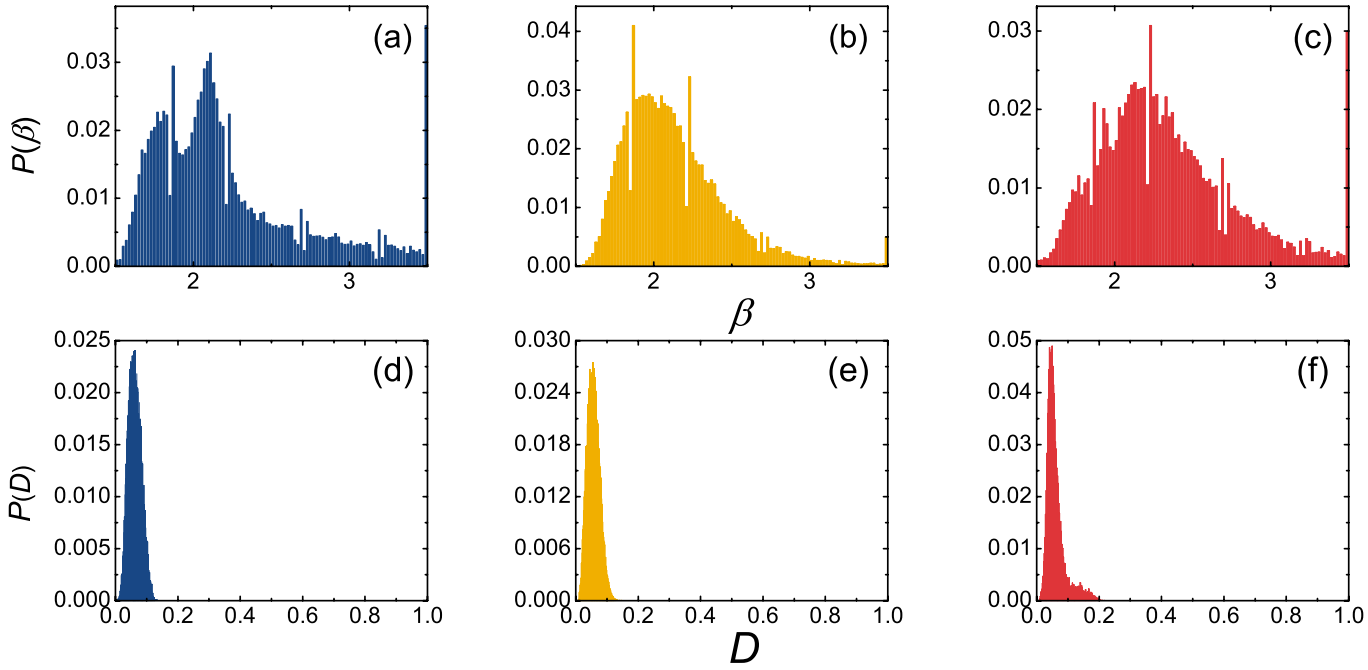
Supplementary Figure S2: Mean click-dwelling time $\langle l \rangle$ (vertical axis) averaged over given real dwelling time (horizontal axis) t from Douban (a) and Taobao (b). Insets are plots on a logarithmic scale in an extended time span.



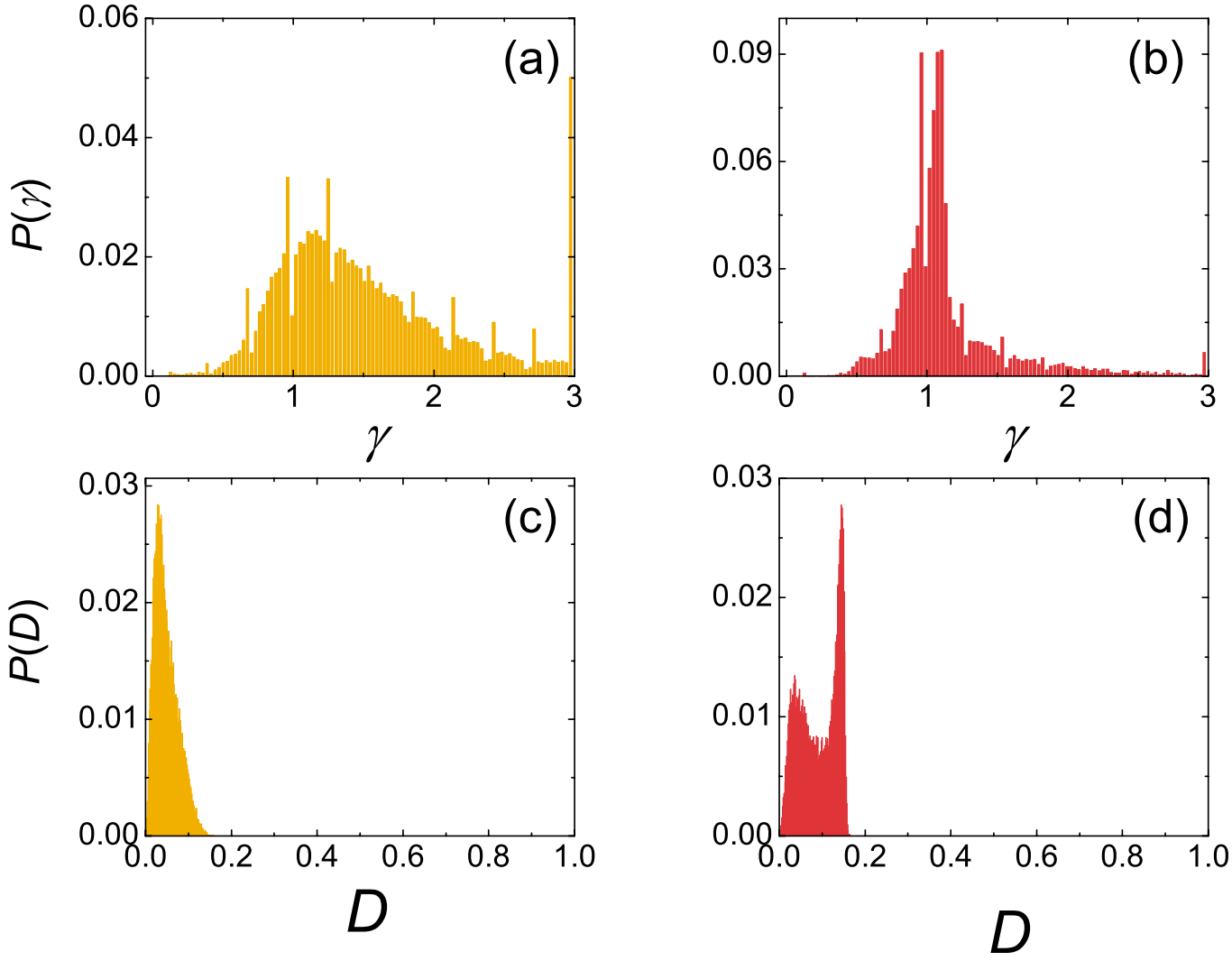
Supplementary Figure S3: Probability distributions of real dwelling time of Douban (a) and Taobao (b).



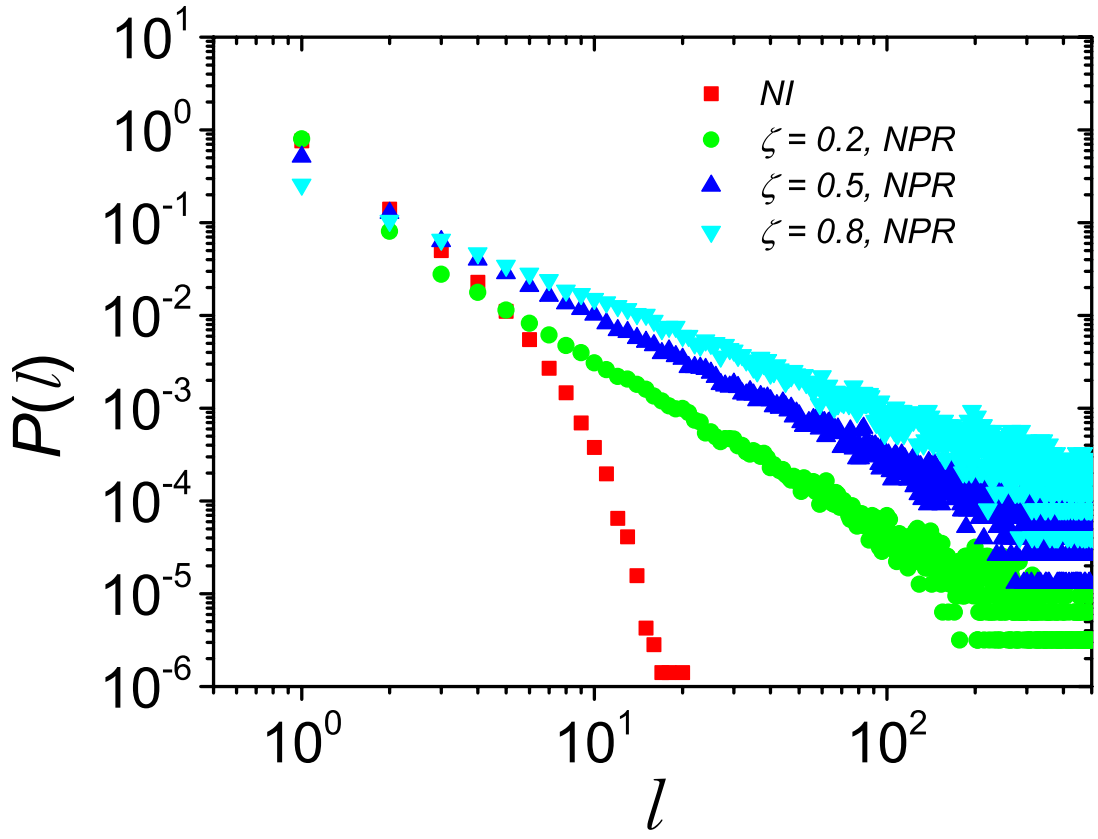
Supplementary Figure S4: (a-c) Distributions of the scaling exponent α for users in Douban, Taobao and MPR, respectively; (d-f) respective distributions of the quantifier D of the Kolmogorov-Smirnov test.



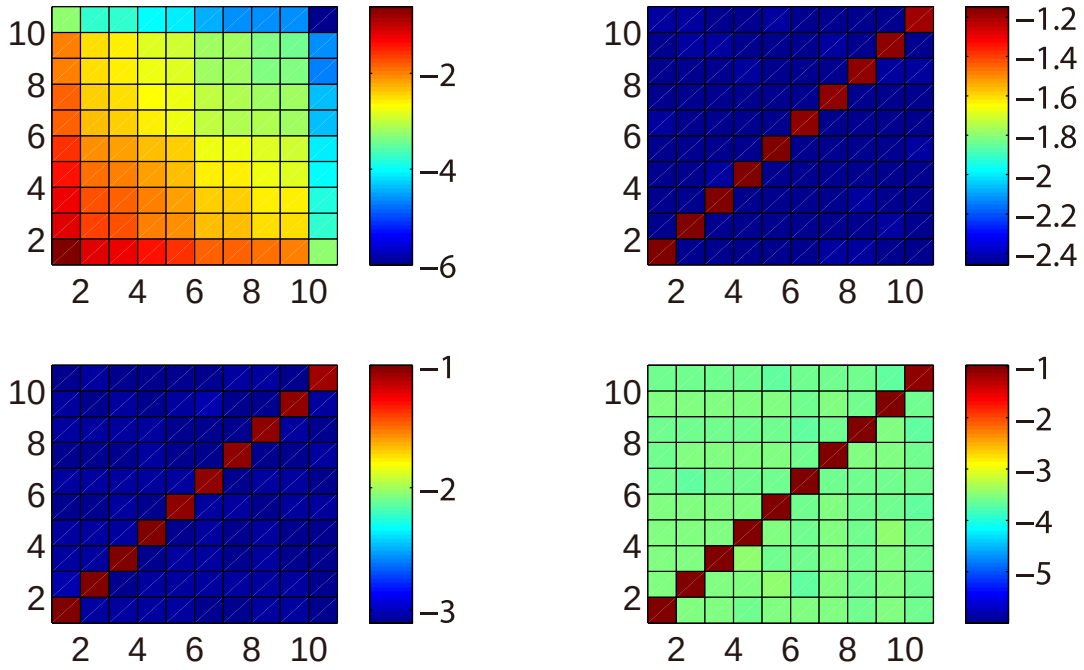
Supplementary Figure S5: (a-c) Distributions of the scaling exponent β for users in Douban, Taobao and MPR, respectively; (d-f) respective distributions of the quantifier D of the Kolmogorov-Smirnov test.



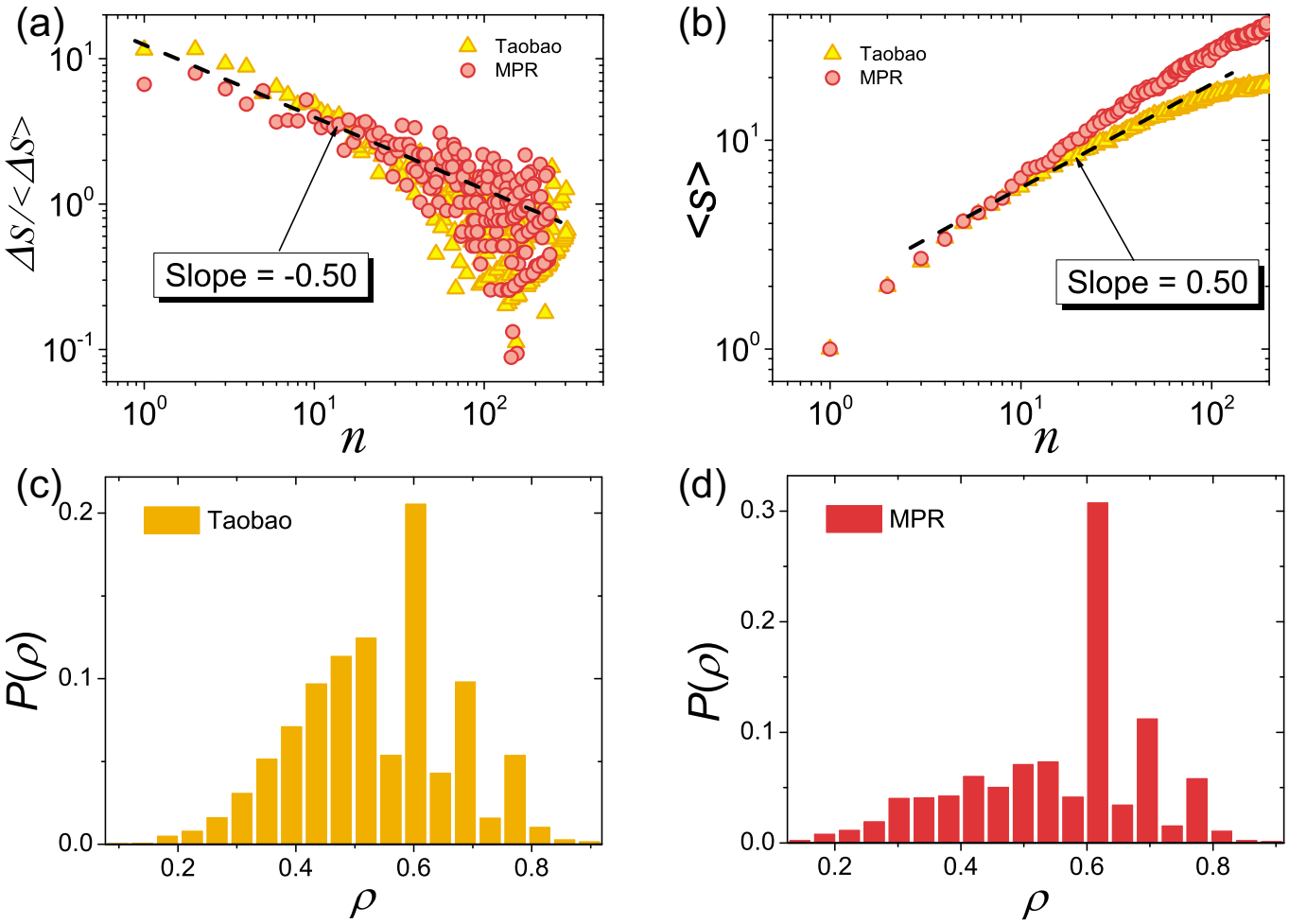
Supplementary Figure S6: (a-b) Distributions of the scaling exponent γ for users in Taobao and MPR, respectively; (c-d) respective distributions of the quantifier D of the Kolmogorov-Smirnov test.



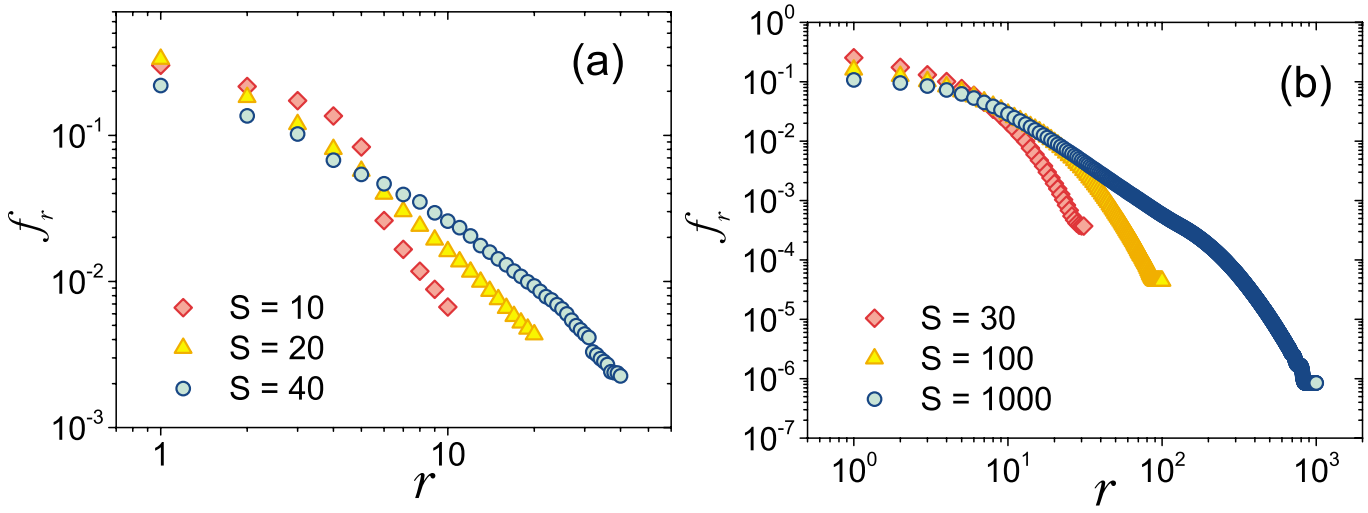
Supplementary Figure S7: Distributions of the consecutive interest interval in the NI and NPR models for different parameters.



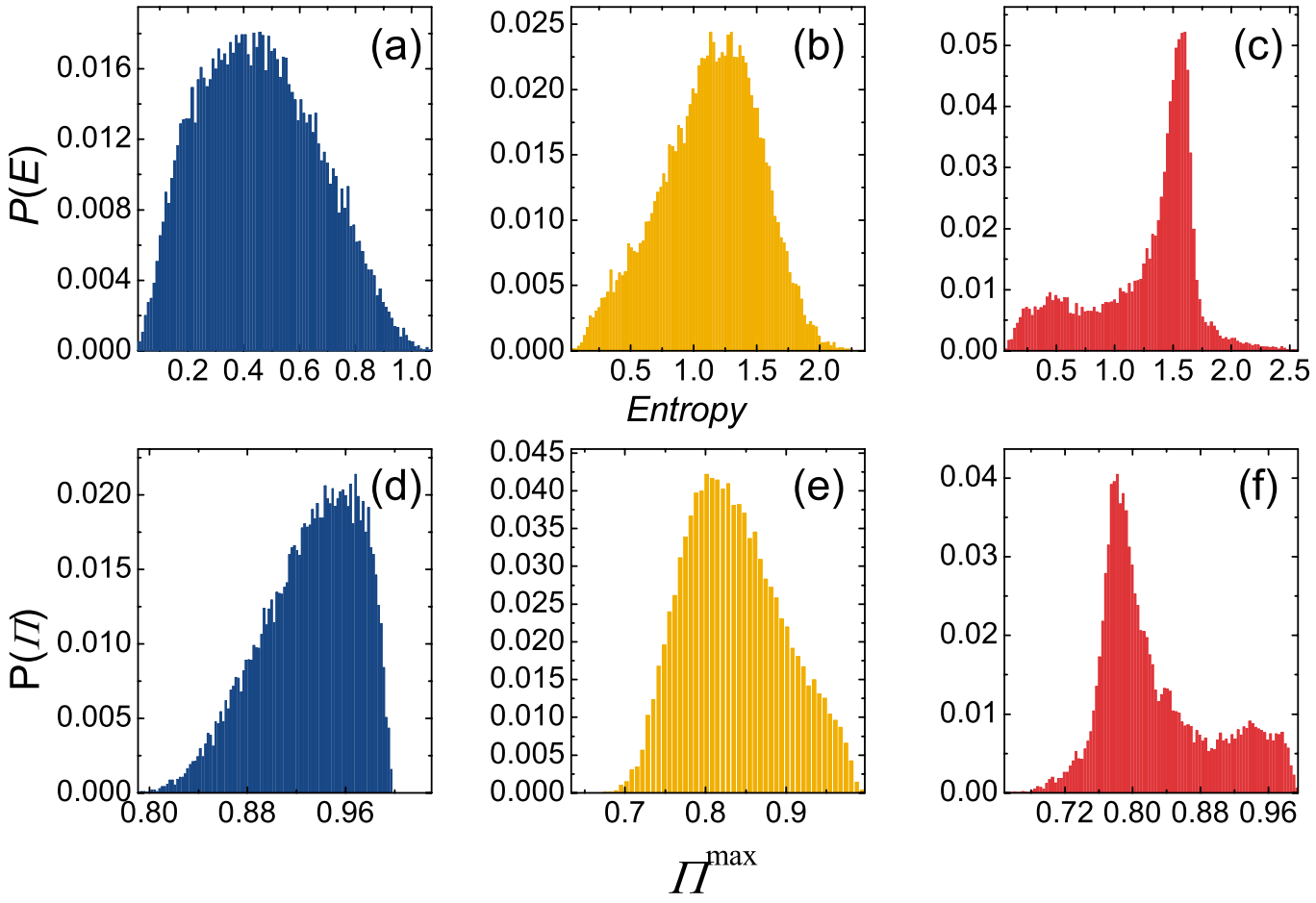
Supplementary Figure S8: Transition probability density for (a) the NI model and (b-d) the NPR model for $\zeta = 0.2$, 0.5 and 0.8, respectively. The parameter N_a is 10^6.



Supplementary Figure S9: For Taobao and MPR, (a) the average increment of the number of visited interests, ΔS, at the n-th hopping [Eq. (S1)]. We have $\Delta S \sim n^{-\lambda}$, where $\lambda \approx 0.5$, (b) relation between the average number of distinct interests $\langle S \rangle$ and number of hops n; (c-d) Probability density functions $P(\rho)$ for the Taobao and MPR user groups, where ρ is a model parameter [Fig. 4(a) in the main text]. The mean values of ρ are approximately 0.58 and 0.55 for Taobao and MPR, respectively. (For Douban, the number of interests is too small for ΔS to be meaningful.)



Supplementary Figure S10: (a) Frequency of the r-th interest from the data set MPR (which has relatively larger maximum value of S = 82 than these in the other data sets). We chose individuals with S values of at least 10, 20 and 40, respectively. (b) Frequency of the r-th interest from our model with values of S at 30, 100 and 1000. The number of users in the model is 1000.



Supplementary Figure S11: For the three data sets Douban, Taobao and MPR, distributions of (a-c) the entropy values and (d-f) the values of the predictability measure Π^{max} across all users, respectively.