Physica A 390 (2011) 2395–2400

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Heterogenous scaling in the inter-event time of on-line bookmarking

Peng Wang [a,b,*], Xiao-Yi Xie [a], Chi Ho Yeung [b], Bing-Hong Wang [a,c]

[a] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China

[b] Department of Physics, University of Fribourg, Fribourg, 1700, Switzerland

[c] The Research Center for Complex System Science, University of Shanghai for Science and Technology and Shanghai Academy of System Science, Shanghai, 200093, China

ARTICLE INFO

Article history:
Received 19 October 2010
Received in revised form 4 January 2011
Available online 3 March 2011

Keywords:
Inter-event time distribution
Power law
Human dynamics

ABSTRACT

In this paper, we study the statistical properties of bookmarking behaviors in *Delicious.com*. We find that the inter-event time (τ) distributions of bookmarking decay in a power-like manner as τ increases at both individual and population levels. Remarkably, we observe a significant change in the exponent when the inter-event time increases from the intra-day range to the inter-day range. In addition, the dependence of the exponent on individual *activity* is found to be different in the two ranges. Instead of monotonically increasing with *activity*, the inter-day exponent peaks around 3. These results suggest that the mechanisms driving human actions are different in the intra-day and inter-day ranges. We further show that the global distributions of less active users are closer to an exponential distribution than those of more active users. Moreover, a universal behavior in the inter-day range is observed by considering the rescaled variable $\tau/\langle\tau\rangle$. Finally, the possible causes of these phenomena are discussed.

1. Introduction

With increasing availability of data from Internet applications, recent years have witnessed expanding interest in characterizing and modeling human behavior. Many on-line human activities such as email communications [1–4], web surfing [5–7], movie rating [8], playing on-line games [9,10], and blog posting [11] and off-line activities such as letter communications [4,12,13] and text messages [14] are under active investigation to provide understanding of our society. One of the main results of these empirical studies is the heavy-tailed nature of the inter-event time distribution: the time interval between two consecutive human actions, which we denote as τ, follows a power-law distribution, i.e., $p(\tau) \sim \tau^{-\beta}$. Moreover, some studies have claimed that there exist a few universality classes in human dynamics characterized by universal exponents [2], which has led to scientific debates [3,4,7,8]. Other studies show that the exponents of inter-event time distributions depend on *activity* (the frequency with which an individual takes actions), which implies that the exponent of an individual is not a good representation of human behavior [7,8], but a universal behavior can nevertheless be found by considering the rescaled variable $\tau/\langle\tau\rangle$ [4,7]. It is noted that this strong dependence can only be observed in the inter-day range; it becomes much weaker in the intra-day range [15]. These results also suggest that we may classify human activities by different time ranges.

In this paper, we study in detail inter-event time statistics in, which is a typical web 2.0 application. Through Delicious, users save and manage bookmarks, while sharing interesting bookmarks with friends. It should be noted that there is a close relation between web surfing and bookmarking: in most cases, we surf on the web, bookmark interesting web pages, and

* Corresponding author at: Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China.
E-mail addresses: wangpenge@gmail.com (P. Wang), bhwang@ustc.edu.cn (B.-H. Wang).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.02.026

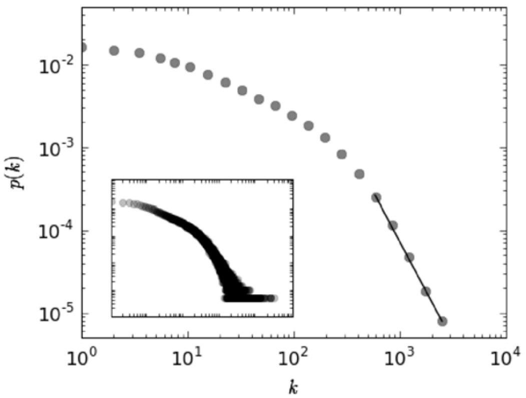

Fig. 1. The distribution $p(k)$ of the number of bookmarks collected by individual users in Delicious. The distribution is shown with a log-bin of k, and the decay exponent is 2.41. Inset: the distribution $p(k)$ as shown with a linear-bin of k.

continue surfing. Heavy tails have already been observed in the distribution of the time interval between consecutive visits to URLs [6,7]. On the other hand, the data set of Delicious is widely adopted as a training set for recommender systems [16–18]. Understanding the temporal pattern in Delicious may give us insight to devise time-aware recommender algorithms, which utilize the time stamps of data to increase the recommendation accuracy [19,20].

The paper is organized as follows. In Section 2, we provide detailed descriptions of the data set studied. In Section 3, we give examples of individual inter-event time distributions which show heavy-tailed nature and heterogenous scaling in the intra-day and inter-day ranges. In Section 4, we give the global inter-event time distribution in these two ranges and distinguish them by estimating the respective decay exponents. In Section 5, through comprehensive analysis of the dependence of the exponent on *activity*, we show that different trends are observed in the intra-day and inter-day ranges. Then, a data collapse among the inter-day distributions is observed by considering the rescaled variable $\tau/\langle\tau\rangle$. Finally, we summarize the results and discuss the possible causes in Section 6.

2. Data description

Our data set consists of 54,204,641 bookmarking activities by 220 867 users over a period of 31 months (between 01/04/2004 and 01/11/2007). Here we use only the identifier (ID) of the users and the time when the bookmarks were saved. The resolution of time stamps is in seconds. We denote k to be the number of bookmarks saved by a user, and $p(k)$ to be the distribution of k among users, which is shown in Fig. 1. As we can see, $p(k)$ is broad, and the tail of the distribution decays as a power law as k increases, giving $p(k) \sim k^{-2.41}$. This result resembles the distribution of the number of messages in Ebay [7], and is significantly different from the distributions of the number of log-in actions in Wikipedia (which follows a power law over the whole range [7]) and the number of posts in blog (which is the so-called "double power law" [11,21]). Interestingly, in spite of the difference in these distributions, the statistics on their inter-event times are very similar, as we will see below.

3. Inter-event time distribution for individuals

In our context, the inter-event time τ is defined as the time interval between consecutive bookmarks by the same user. Fig. 2 shows the cumulative distribution of inter-event time obtained from six users. As we can see, all curves show a crossover around $\tau \approx 1$ day, which corresponds to a change in exponent between the intra-day and inter-day ranges. Although power-law decays are observed in both ranges, the change in exponent (which is also noticed in other systems [15]) suggests that the mechanism driving intra-day and inter-day activities are different. Moreover, changes in exponent are observed even within the intra-day range for some users. As shown in Fig. 2(e) and (f), a slight increase in the decay exponent is observed at $\tau \approx 1$ h.

4. The global distribution of inter-event time

The global distribution of inter-event times is plotted in Fig. 3. In order to have a clear picture in the intra-day range, we express τ in Fig. 3(a) with a resolution of minutes. In Fig. 3(b), we express it with a resolution of days, where the circadian oscillations are masked, which makes the decay in the inter-day range clearer. Both distributions in the intra-day and inter-day ranges present a power-like decay, with exponents $\beta_{\text{intra}} \simeq 1.07$ for the intra-day range and $\beta_{\text{inter}} \simeq 2.41$ for the inter-day range. This significant difference between the exponents of the two ranges is consistent with the results obtained from the distribution of individuals in Section 3. The exponent in the intra-day distribution in this case is similar to the one obtained from consecutive visits to URLs [6]. This is reasonable, since bookmarking often follows web surfing, as we