Published in "Physica A: Statistical Mechanics and its Applications
390(12): 2395-2400 , 2011" which should be cited to refer to this work.

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

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).

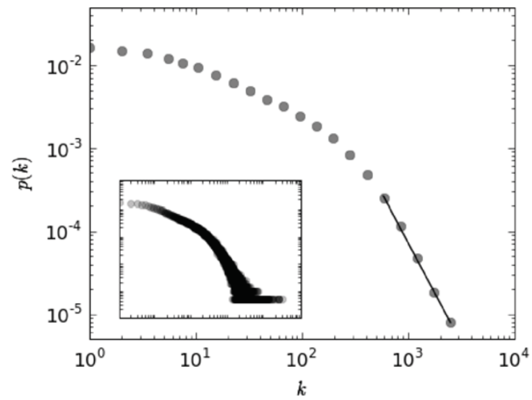

Fig. 1. The distribution $p(k)$ of the number of bookmarks collected by individual users in Delicious. The distribution is shown with a log-bin of k, and the decay exponent is 2.41. Inset: the distribution $p(k)$ as shown with a linear-bin of k.

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

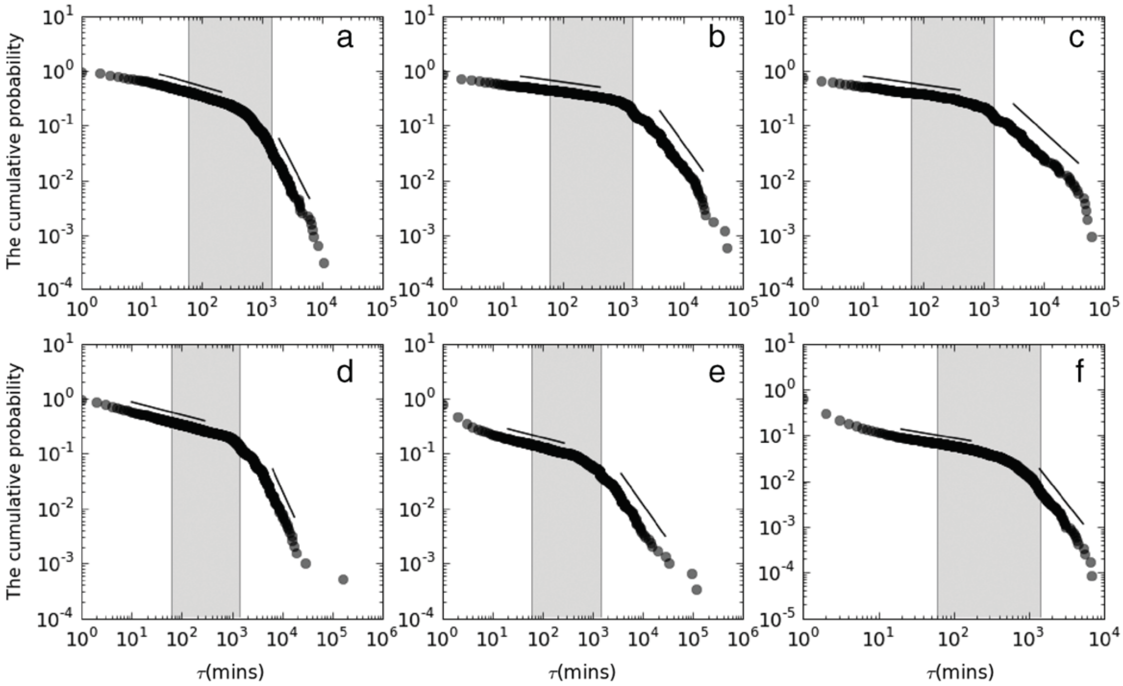

Fig. 2. The cumulative distribution of inter-event time of six random individuals. Their corresponding numbers of bookmarks are 3104, 1689, 1047, 1946, 2983, and 11,892. The shaded areas correspond to the span of τ in the range of 60 min (1 h) $< \tau <$ 1440 min (1 day). The exponents of these cumulative distributions in the intra-day and inter-day ranges (β_{intra}, β_{inter}) are (a) (0.31, 2.15), (b) (0.15, 1.53), (c) (0.15, 1.0), (d) (0.23, 2.02), (e) (0.23, 1.29), and (f) (0.28, 2.09).

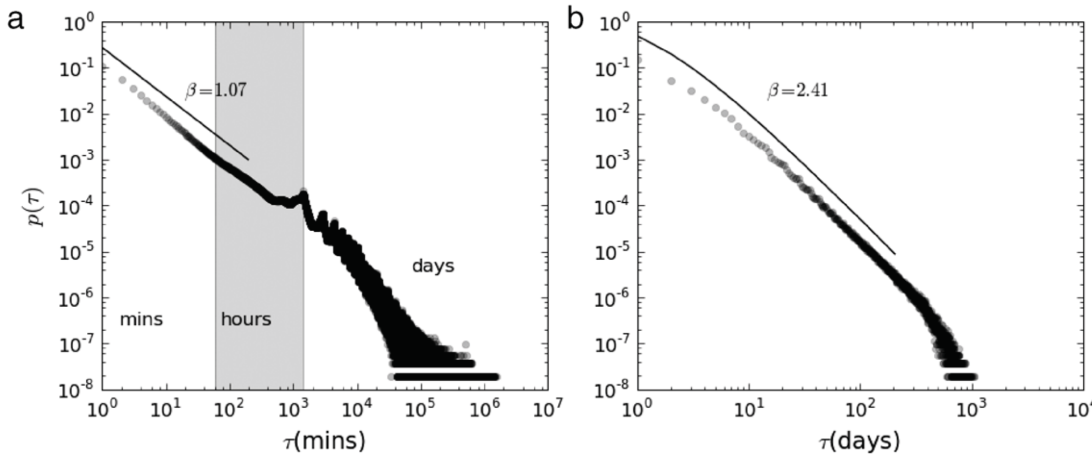

Fig. 3. The global distribution of inter-event time with precision in (a) one minute and (b) one day.

mentioned above. The exponent of the inter-day distribution is very large compared to other systems [7,8,15], which makes the following analysis different from the others. It should be noted that we fit the inter-day distributions with the so-called "shifted power law" (SPL) [22,23]:

$$y \sim (x + h)^{-\beta}. \tag{1}$$

The SPL can be shown by a linear line with a slope β on the $\ln p(x) - \ln(x+h)$ plane. The exponent β of inter-day distributions can thus be fitted by the method of least square error in the plot of $\ln p(x)$ against $\ln(x+h)$ [23], given a value of h. The value of h is determined by a variational approach, which leads to the best fitting results. When the parameter h increases from 0 to ∞, the distribution varies from a power-law distribution to an exponential distribution [23]. Actually, when h is larger than 100, the SPL shows a rather good linear line on a linear–logarithmic plane, indicating approximately an exponential distribution. In our case, the distribution in Fig. 3(b) is fitted by Eq. (1) with $h \approx 3.32$. We will further discuss in the next section the fitting values of h for distributions with different individual *activity*.

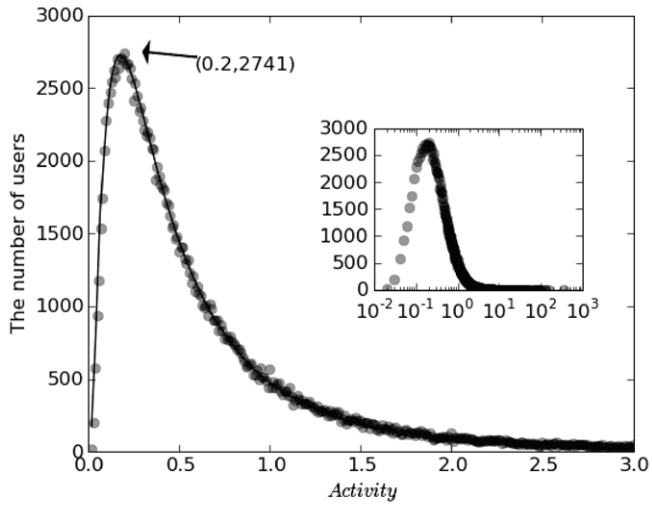

Fig. 4. The distribution of *activity*. The solid line corresponds to the fitting of the log-normal distribution $\ln \mathcal{N}(\mu, \sigma^2)$ with $\mu = \ln(0.42)$ and $\sigma = 0.93$.

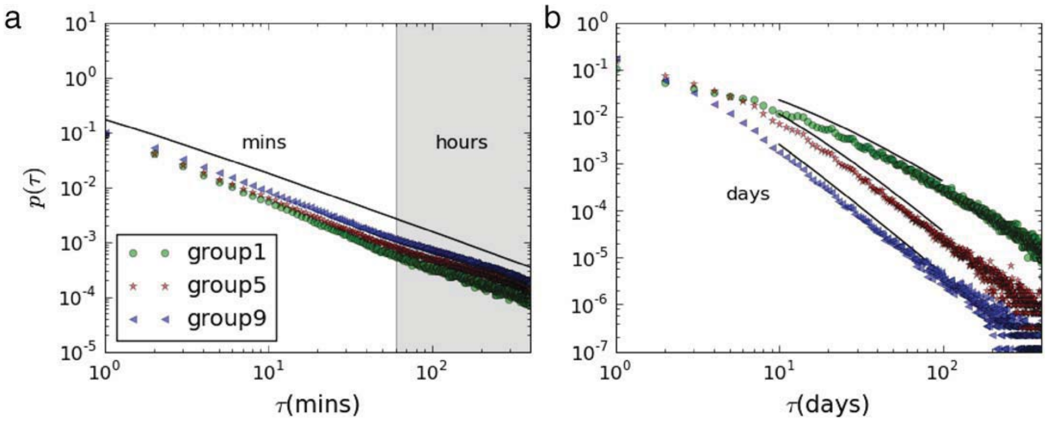

Fig. 5. (Color online) The exponent dependence on *activity*. The inter-event time distributions of groups 1, 5, and 9 are shown in this figure. As a comparison, the slope of the straight line in (a) is 1.07, which we obtained from the global distribution. In (b), the exponents we obtained are 2.15 for group 1, 2.91 for group 5, and 2.90 for group 9.

5. Activity and exponents

Based on individual heterogeneity, we investigate both the intra-day and inter-day distributions. First, we define the average *activity* A_i of user i as

$$A_i = \frac{n_i}{T_i}, \tag{2}$$

where n_i is the total number of bookmarks saved by user i and T_i is the time interval between the first and the last bookmarks of user i. We consider only users with at least 20 bookmarks and T_i larger than 10. There are 173108 users who satisfy these conditions. Fig. 4 shows the distribution of *activity* A_i, which is also heavy tailed, and approximately follows a log-normal distribution, as shown by the fitting line. The value of A_i of most users is between 0.01 and 1, with most probable value around 0.2 per day.

To examine inter-event time distributions in relation with user *activity*, we sort users in an ascending order of A_i and divide the entire population into 10 groups, each of which has M users ($M \approx N/10$, where N is the total number of users). Accordingly, the mean activity of each group obeys the inequality $\langle A \rangle_1 < \langle A \rangle_2 < \cdots < \langle A \rangle_{10}$. We then investigate the dependence of the exponent on *Activity* in both the intra-day and inter-day ranges. In Fig. 5, we plot the inter-event time distribution of groups 1, 5, and 9 (which respectively correspond to $\langle A \rangle = 0.09$, 0.37, and 1.12 per day). In the intra-day range, we find that the exponents are only weakly dependent on $\langle A \rangle$ (as shown by Fig. 5(a), a slight decrease is observed with $\langle A \rangle$). In contrast, in the inter-day range, the exponents increase from 2.15 to 2.91 with $\langle A \rangle$ increasing from 0.37 to 1.12. This result shows that behavioral heterogeneity in intra-day and inter-day ranges is also evident from the exponent dependence on $\langle A \rangle$.

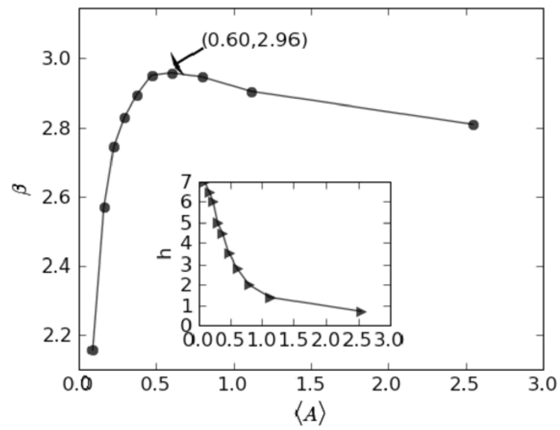

Fig. 6. β of each group is plotted as a function of average *Activity*. The inset shows h as a function of average *Activity*.

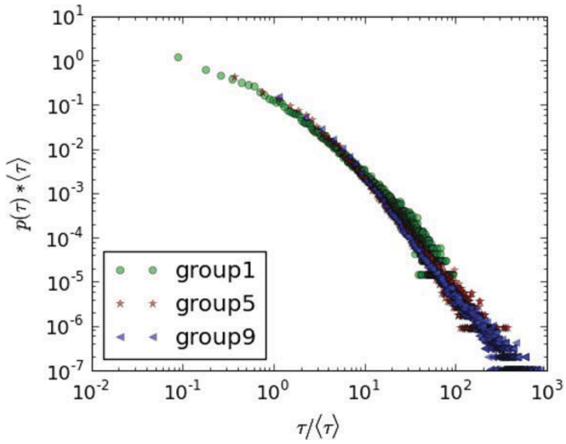

Fig. 7. (Color online) Scaling of the inter-event time distributions.

In Fig. 6, we show the inter-day exponents of the inter-event time distribution as a function of A_i. It is noted that the exponents here increase much more quickly than the ones in previous studies in spite of the same tendency [7,8,15]. The exponent of group 1 is 2.15 and that of group 3 already increases to 2.74. The reason for this steep increase may correspond to the large exponent of 2.41 of the global distribution, which leads us to observe the dependence of exponents on a broader range. As we can see, the exponent reaches the maximum at group 6 and then it decreases slightly. This suggests a limiting value of $\beta \simeq 3$ for decay exponents. Actually, in Radicchi's study [7], the exponents of the last group of America On-Line and Wikipedia also decrease. We further plot in the inset of Fig. 6 the fitted values of h from Eq. (1) for the distributions of the 10 groups, which shows a monotonic decrease of h with $\langle A \rangle$. For instance, $h \approx 7$ for group 1 and $h \approx 0.7$ for group 10, indicating a substantial decrease of h. As mentioned above, this shows that the inter-event time distributions of less active groups are closer to an exponential distribution than those of more active groups.

Instead of considering the bare value for the inter-event time τ, we take into account the activity of each single user and consider the rescaled variable $\tau/\langle\tau\rangle$. $\langle\tau\rangle$ represents the average inter-event time between two actions performed by the same user. Interestingly, the simple scaling of $\langle\tau\rangle P(\tau)$ versus $\tau/\langle\tau\rangle$ allows us to find a data collapse between curves corresponding to populations with different *activity*. In Fig. 7, for example, we plot the quantity $\langle\tau\rangle P(\tau)$ versus $\tau/\langle\tau\rangle$ for the three groups in Fig. 5(b). The same result can be observed in other groups. It should be noted that these observations have already been reported in other human-driven systems [7,24,25].

6. Discussion

In this paper, we have shown that the distributions of inter-event time at both individual and population levels are heavy tailed. Our results further verified heterogenous human dynamics in intra-day and inter-day ranges. On the one hand, there is a significant difference between the exponents in the intra-day and inter-day distributions, which are 1.07 and 2.41, respectively. On the other hand, the inter-day exponents are strongly dependent on the individual's *activity*, while this dependence is absence in the intra-day range. Moreover, our study suggests that there is a maximum value of $\beta \approx 3$ for the increase of exponent with *activity*. It should be noted that similar results were also reported on Wiki-revising and blog posting [15], which shows that heterogenous scaling in the intra-day and inter-day ranges may be a common feature of human activities. This phenomenon is apparently related to human circadian rhythms, and one possible explanation is the time scale in scheduling activities. For instance, we can plan our daily schedule carefully according to our personal preference or need, but we hardly plan for every minute.

For the intra-day behavior, the distributions seem to be in the universality class of human dynamics characterized by $\beta = 1$ [2,6,7,11,13], if we ignore the weak dependence of exponents on *activity*. It is well known that the priority-queue model of Barabási [1] can produce a waiting time distribution with $\beta = 1$, which can also be used to explain the inter-event time distribution according to the arguments of [2]. The other models include the so-called "adaptive interest" model and the zero-crossing model, which also give similar inter-event time distributions [26,27].

The strong dependence of the exponent on *activity* in the inter-day range implies that it is inappropriate to classify the distributions in this range by only their exponents. However, human behavior in this range still shows some common features. One of them corresponds to the exponents of these distributions being variable in a wide range with a strong correlation with an individual's *activity* [7,8,15]. Moreover, the values of these exponents are often larger than or equal to 2 and smaller than 3 [7,8,15]. Though exponent dependences on *activity* are observed in this range, successful data collapse of these distributions suggests that the dynamics at different *activity* is driven by the same mechanism. Similarly, in the nonhomogeneous Poisson model, the same mechanism leads to inter-event distributions with different exponents based on different system parameters [3]. A more interesting result is obtained from the temporal-preference model in which even with the same system parameter different exponents are observed in different parts of the time series with heterogeneous *activity* [11]. This result can explain the data collapse between different distributions in some human-driven systems [7,24,25]. However, the current temporal-preference model is oversimplified, although it can give a preliminary explanation for the inter-day dynamics. In order to improve the model and give better agreements with most systems, it is key to understanding the relation between action repetition and memory [28,29].

Acknowledgements

This work is funded by the National Basic Research Program of China (973 Program No. 2006CB705500); the National Important Research Project (Study on emergency management for non-conventional happened thunderbolts; Grant No. 91024026); the National Natural Science Foundation of China (Grant No. 1097512610635040); the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20093402110032); and Liquid Publications and Qlectives projects (EU FET-open Grants 213360 and 231200).

References

[1] A.-L. Barabási, Nature 435 (2005) 207.
[2] A. Vázquez, J.G. Oliveira, Z. Dezso, K.-I. Goh, I. Kondor, A.-L. Barabási, Phys. Rev. E 73 (2006) 036127.
[3] R.D. Malmgren, D.B. Stouffer, A.E. Motter, L.N. Amaral, Proc. Natl. Acad. Sci. USA 105 (2008) 18153.
[4] R.D. Malmgren, D.B. Stouffer, A.L.O. Campanharo, L.N. Amaral, Science 325 (2009) 1696.
[5] Z. Dezso, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, A.-L. Barabási, Phys. Rev. E 73 (2006) 066132.
[6] B. Goncalves, J.J. Ramasco, Phys. Rev. E 78 (2008) 026123.
[7] F. Radicchi, Phys. Rev. E 80 (2009) 026118.
[8] T. Zhou, H.A.T. Kiet, B.J. Kim, B.-H. Wang, P. Holme, Europhys. Lett. 82 (2008) 28002.
[9] A. Grabowski, N. Kruszewska, R.A. Kosiński, Phys. Rev. E 78 (2008) 066110.
[10] Z.-Q. Jiang, W.-X. Zhou, Q.-Z. Tan, Europhys. Lett. 88 (2009) 48007.
[11] P. Wang, T. Zhou, X.-P. Han, B.-H. Wang, arXiv:1007.4440v2.
[12] J.G. Oliveira, A.-L. Barabási, Nature 437 (2005) 1251.
[13] A. Vázquez, Physica A 373 (2007) 747.
[14] W. Hong, X.-P. Han, T. Zhou, B.-H. Wang, Chin. Phys. Lett. 26 (2009) 028902.
[15] P. Wang, L. Ting, C.H. Yeung, B.-H. Wang, arXiv:1008.3982v1.
[16] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J.R. Wakeling, Y.-C. Zhang, Proc. Natl. Acad. Sci. USA 107 (2010) 4511.
[17] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Phys. Rev. E 76 (2007) 046115.
[18] Z.-K. Zhang, C. Liu, J. Stat. Mech. (2010) P10005.
[19] Y. Ding, X. Li, in: Proc. 14th ACM International Conference on Information and Knowledge Management, 2004, p. 485.
[20] Y. Koren, in: Proc. 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, p. 447.
[21] A. Grabowski, Eur. Phys. J. B 69 (2009) 605.
[22] D.-R. He, Z.-H. Liu, B.-H. Wang, Complex Systems and Complex Networks, Higher Education Press of China, Bejing, 2009.
[23] H. Chang, B.-B. Su, Y.-P. Zhou, D.-R. He, Physica A 383 (2007) 687.
[24] K.-I. Goh, A.-L. Barabási, Europhys. Lett. 81 (2008) 48002.
[25] J. Candia, M.C. González, P. Wang, T. Schoenharl, G. Madey, A.-L. Barabási, J. Phys. A: Math. Theor. 41 (2008) 224015.
[26] X.-P. Han, T. Zhou, B.-H. Wang, New J. Phys. 10 (2008) 073010.
[27] M. Gotz, J. Leskovec, M. McGlohon, C. Faloutsos, in: Proceedings of the Third International ICWSM Conference, 2009.
[28] G. Xue, Q. Dong, C.-S. Chen, Z.-L. Lu, J.A. Mumford, R.A. Poldrack, Science 330 (2010) 97.
[29] J.D. Karpicke, H.L. Roediger, Science 319 (2008) 966.