

## Human dynamics revealed through Web analytics

Bruno Gonçalves\*

*Physics Department, Emory University, Atlanta, Georgia 30033, USA*

José J. Ramasco†

*Complex Systems Lagrange Laboratory, Complex Networks (CNLL), ISI Foundation, Viale S. Severo 65, I-10133 Turin, Italy*

(Received 21 May 2008; published 26 August 2008)

The increasing ubiquity of Internet access and the frequency with which people interact with it raise the possibility of using the Web to better observe, understand, and monitor several aspects of human social behavior. Web sites with large numbers of frequently returning users are ideal for this task. If these sites belong to companies or universities, their usage patterns can furnish information about the working habits of entire populations. In this work, we analyze the properly anonymized logs detailing the access history to Emory University's Web site. Emory is a medium-sized university located in Atlanta, Georgia. We find interesting structure in the activity patterns of the domain and study in a systematic way the main forces behind the dynamics of the traffic. In particular, we find that linear preferential linking, priority-based queuing, and the decay of interest for the contents of the pages are the essential ingredients to understand the way users navigate the Web.

DOI: [10.1103/PhysRevE.78.026123](https://doi.org/10.1103/PhysRevE.78.026123)

PACS number(s): 89.75.Hc, 89.70.-a

### I. INTRODUCTION

Access to the Internet has become increasingly popular during the past decade. However, despite its importance, much is still unknown about the intrinsic properties of the Web, the way people interact with it, and how it impacts our culture [1–4]. Several theoretical approaches have been proposed in the past few years [5–12], but some fundamental issues are yet to be fully understood. In this work, we will focus on answering the following question: Do any laws govern the way and frequency with which a person visits a given Web site, or is each individual intrinsically unique? From a sociological point of view, we would expect that, although the behavior of a single individual is ultimately personal and unpredictable, many inferences can be obtained about the most common behaviors [2,13]. A better understanding of the way an individual uses a given Web site has important economic consequences, as it can help the developers of the site optimize it in a way that facilitates its use, and monetization. Apart from the utilitarian point of view, the activity patterns on the sites provide also important information on the dynamics of a population. The interaction with electronic devices or virtual instruments, such as social sites or mobile phones, opens promising research avenues in this direction [14–20].

The sheer size and diversity of the World-Wide Web renders any attempts to characterize it on a global scale barely feasible. Still several works have recently centered on describing from a statistical perspective the structure of the Web [21,22]. If instead of understanding its structure the goal is to track how users navigate it, the challenge becomes even greater. A solution consists in ignoring the identity of the users, focusing only on the number of visitors per site

and on the number of clicks on its hyperlinks [18,23]. Another possibility is to concentrate the attention onto a group of volunteers [24] or onto the users of a social site that are usually well identified [25–27]. Our aim here is to follow the activity of individually trackable Web surfers in a relatively open environment and characterize the way in which the interaction between users and Web sites occurs. This is the reason why we analyze the logs of the Web server of Emory University. These logs registered the requests by Internet users, internal or external, of Web pages in the second level of the Emory domain ([www.emory.edu](http://www.emory.edu)). The data comprehend a period that goes from Apr. 1, 2005 to Jan. 17, 2006. Each time a computer connects to the Internet, it is assigned a unique IP address that identifies it. When a user requests a page from a Web site, the IP, the page requested (URL), the time at which the request occurred, and several other details are registered by the Web server. In our case, to preserve privacy, the data have been anonymized in a coherent way, allowing us to follow the behavior of each IP by a single ID number but masking the real identity. The log structure is represented schematically in Fig. 1. On the left, we have the anonymized IP addresses that connect to the URLs on the right. To avoid the consideration of different elements of a Web such as photos or logos as independent pages, we have restricted our definition of URL to *(s)htm(l)*, *cfm*, *php*, *asp(x)*, *jsp*, and *txt* documents. Each line of the logs corresponds to a different connection, which is time-stamped with the date and time at which it took place. During our observation period, the domain received over 3 million visitors to about 2.5 million pages, for a grand total of over 53 million clicks.

### II. ACTIVITY PATTERNS OF THE POPULATION

Let us start by taking a view of the collective behavior of the entire population during the time period for which we have data. Intuitively, we expect the activity on a domain to

\*[bgoncalves@physics.emory.edu](mailto:bgoncalves@physics.emory.edu).

†[jramasco@isi.it](mailto:jramasco@isi.it).

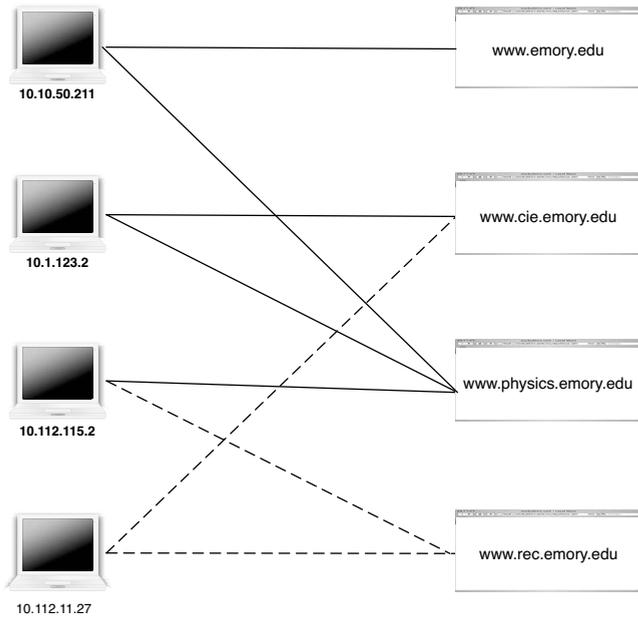


FIG. 1. Schematic representation of the interactions between users and Web pages. The system is dynamic; to provide a more visual impression of its variability, dashed lines represent new added connections.

vary from day to day, week to week, and even month to month. In particular, it should be possible to observe variations in the activity, seen as the number of requests, due to weekends, holidays, and other major events that disrupt the normal life of the university. The traffic at Emory is dominated by students and professors in the course of their professional activities, and hence the major events in the course of the school year, such as the beginning and end of a semester, breaks, or holidays, should be noticeable in the Web traffic. In order to check this idea, the number of page requests detected per day is shown in Fig. 2 as a function of the observation date. One obvious feature of the figure is a clear oscillatory behavior with a period of one week. It also displays different trends for two special times of the year: one at the later part of August, corresponding to the begin-

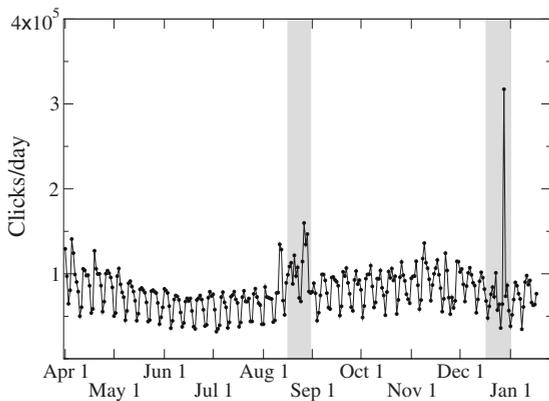


FIG. 2. Total number of clicks registered per day during the whole period of traffic observation. The gray bands correspond to the beginning and the end of the semester: from Aug. 16 to Aug. 31, and from Dec. 16 to Dec. 31.

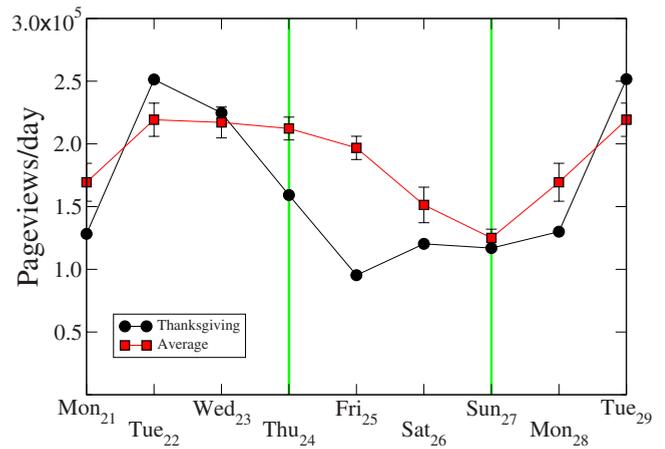


FIG. 3. (Color online) Comparison between the average week activity and activity during Thanksgiving week. The green vertical lines represent the beginning and end of the official Thanksgiving break at Emory University. The error bars for the average are calculated as two times the standard deviation,  $2\sigma$ , or the 95% confidence interval.

ning of the school year, and the other at the end of December, when the semester finishes.

Since accesses to the Emory domain are mostly work-related, traffic can be used as an indirect measure of the university “productivity.” Busier days would result in larger amounts of traffic, while during holidays and weekends the number of page requests is overall smaller, thus rendering the relative changes in traffic significant. The averages of page requests by the day of the week during the complete observation period are plotted, together with their corresponding 95% confidence intervals, in Fig. 3. Our results support the old adage that after Wednesday, the hardest part of the week is already behind us, with the activity slowly decreasing from then on to the weekend. Sundays are the least active day of the week. It is also interesting to note the not so active behavior of Mondays, only slightly more active than Saturdays. Armed with an estimate of how activity evolves over the week, we are now in a position to evaluate the effects of a break. In the same figure, we also represent the data for the days surrounding Thanksgiving, one of the major holidays in the United States. Traditionally, Thanksgiving recess goes from Thanksgiving Thursday until Sunday, so one might expect any decreases in activity to be most noticeable during this period. This is what we observe, but we find other effects as well. Both the Monday before and after Thanksgiving seem to be less productive than normal. This is, however, complemented with busier than usual Tuesdays before and after the break.

Intraday variations, with some times of the day being busier than others, are also seen. By averaging the activity observed at a given hour over all the weekdays in our data set, we obtain Fig. 4. The most active period is between 7am and 6pm. The large dip between 11am and 2pm is due to the lunch break. After lunch, the activity peaks reach the higher level of the day. After 6pm, activity levels off until 10pm, marking the end of the workday. Saturdays do not differ significantly from other days of the week, only Sundays dis-

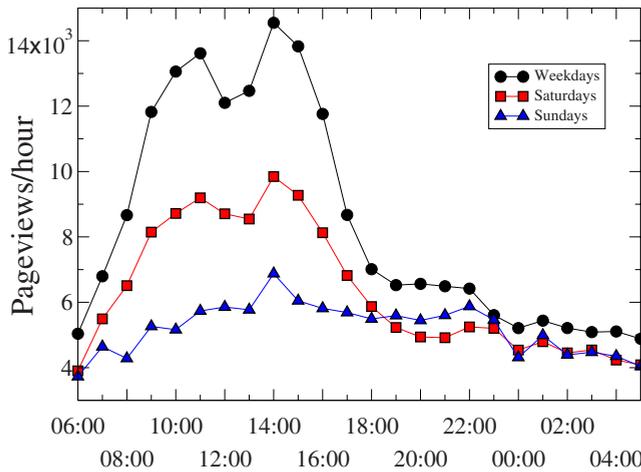


FIG. 4. (Color online) Average hourly activity in the complete Emory domain as a function of the hour of the day. The curves are averaged over the weekdays (circles), Saturdays (squares), and Sundays (triangles).

play a different activity profile. Similar patterns for human circadian rhythms have been recently reported for other systems in Refs. [15,17,18]. Such ubiquity indicates important universal features (profiles) regarding human habits that Web analytics can help to characterize in a quantitative way.

### III. INDIVIDUAL ACTIVITIES

Although interesting, the analysis of averages taken over the entire population has limitations. The histograms of single user activity are typically very wide, being in some cases well-modeled by power-law distributions with exponents smaller than 2 [23]. When this happens, it is difficult to identify a “typical” user based on such metrics: while most users only visit the domain sites a few times, a significant fraction of individuals (as identified by their IP addresses) accumulate large numbers of page requests. This variability deserves greater attention since it can carry important information. Figure 5 shows the activity patterns of three users. We do not know the actual IPs, but it is possible to deduce the intention of the visit based on the particular URL accessed and on the profile of the activity. In Figs. 5(a) and 5(b), the users are computer programs. One, the case shown in (a), corresponds to a malicious attack on a finance service Web page of Emory. It took place on April 4th. The profile of the number of access attempts per unit of time displays a very peculiar shape, quite regular as occurs for most automatic navigators, with a very high number of requests concentrated in a short period of time. Other, more friendly, robots are those corresponding to updating programs. An example can be seen in Fig. 5(b), where a software site in Emory is regularly visited presumably in search for new updates. Finally, human users show a very different activity profile from that of the machines. The activity of a human user selected at random can be seen in Fig. 5(c). In this case, the URL is an administrative site that demands manual introduction of data. The activity congregates in some days followed by relatively long periods of time without any request.

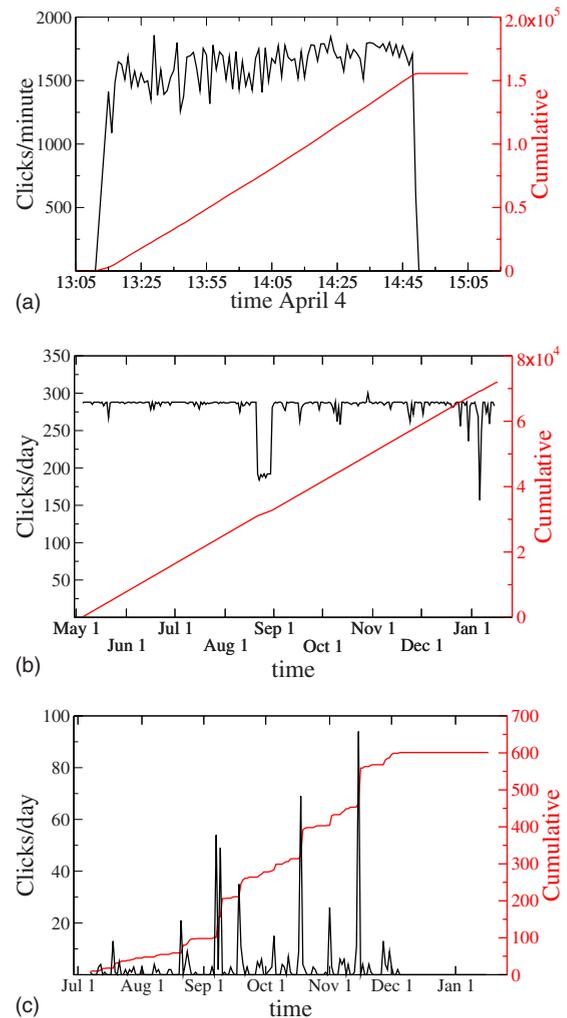


FIG. 5. (Color online) Activity history of several individuals: (a) what seems to be a malicious attack on a finance Web page of the University, (b) an automatic software update program, and (c) a human user filling data in an administration site. The red curves represent the cumulative number of clicks. To facilitate the visualization, the scale of the cumulative and temporal number of clicks is different. The axis on the right side of each plot displays the scale for the cumulative number of clicks.

Given the strong variability in the activity of human users, it is interesting to measure some statistics about it. In Fig. 6, we have represented the histograms of the duration of the periods between requests for two different scenarios: in Fig. 6(a) for the time between consecutive visits of the same user to the same URL,  $P(\tau_v)$ , and, in Fig. 6(b), for the time between clicks by the same user to any of the sites in Emory’s domain (not necessarily to the same URL),  $P(\tau_c)$ . Both distributions are rather wide. The distribution  $P(\tau_c)$  can be well fitted by a power-law decaying function of the type  $P(\tau_c) \sim \tau_c^{-1.25}$ . The distribution of time between consecutive visits,  $P(\tau_v)$ , decays even more slowly with an exponent of value  $-1$ . This latter value can be understood thanks to a model on human dynamics recently proposed by Barabási [28] (see also [17,27,29–31]). In this model, an agent has to perform a set of tasks each with a random priority assigned. A step consists in the selection of the task with the highest priority

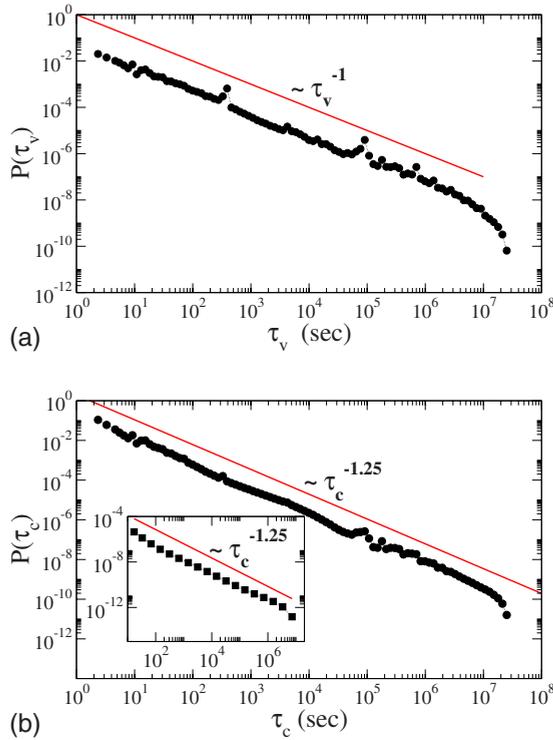


FIG. 6. (Color online) Distribution of times between consecutive clicks: (a) visits of the same user to the same URL, and (b) the same user to any page of the Emory domain. The straight lines correspond to the power law  $f(\tau) \sim \tau^{-1}$  in (a) and to  $f(\tau) \sim \tau^{-1.25}$  in (b). In the inset of (b), the distribution of time in the queue is plotted for a variation of Barabási's model [28] (see text) with a number of executed tasks per unit of time of  $\nu=3$ , with probability of choosing a task according to priority  $p=0.99999$ , a total of 100 tasks, and  $10^7$  time steps.

with probability  $p$  or of a random one with probability  $1-p$ . After the execution, a new task occupies the free spot in the queue. This group of rules is extremely simple but is able to reproduce a distribution of waiting times for the tasks in the queue that, in the limit of small  $p$ , decays as  $\sim 1/\tau$ . It can be argued that consecutive visits to the same site in Emory are equivalent to one of these tasks since many of the visits are related to work or studies, and probably bear an inherent sense of priority for each user. Also returning immediately to the same URL and reloading it is not a common practice, at least not among humans. It is important to note that if the user pushes the back button in the browser, typically we are not able to detect such a move because it does not leave a trace in the logs of the server due to browser caching. If each entrance is seen as a fresh start of a different task, the parallelism between the rules of the model and the way users return to the same pages can be justified.

The question is then whether there is a way to understand also the exponent  $-1.25$  of  $P(\tau_c)$ . The answer is yes, if one considers that a single click on the domain does not necessarily have to be related to the realization of a task. Many tasks will require a (fast) sequence of clicks on different sites of the domain for their completion. This is why we propose the following modification of the model: in each time step, instead of a single task, a group of  $\nu$  tasks is selected for

execution. The selection of each of them is done as before: by priority with probability  $p$ , and at random otherwise. We have performed a systematic numerical study of this model and found that provided that  $\nu > 2$ , the exponent of the distribution of the time of permanence in the queue decays always as  $\sim \tau^{-1.25}$ . An example with  $\nu=3$  is shown in the inset of Fig. 6(b). These two models are oversimplifications but seem able to capture some of the essential features present in the dynamics of a large community of users leading to the existence of universal exponents.

#### IV. ATTRACTIVENESS AND PREFERENTIAL LINKING

Another aspect that is worth exploring in the dynamics of our database is whether the new connections or new clicks follow a preferential rule. Preferential linking or the “rich get richer” effect is a relatively old concept considered originally in a socioeconomic context by Simon [32]. In the area of graphs theory, it was introduced in 1999 [7] with a model inspired in the hyperlinks of the Web (see also [33,34]). A few years have passed, and although several attempts have been made to check the existence of preferential linking in a variety of systems [26,35–37], as far as we know, a systematic study of preferentiality on the user-Web relationship is still missing. To be precise, if the variable under consideration,  $x$ , can change in time for each element of the system, it is said that it shows linear preferentiality if the variation follows on average an expression of the type

$$\langle \Delta x \rangle \approx Ax + B, \quad (1)$$

where the average  $\langle \cdot \rangle$  is taken over all elements  $i$  of the system with  $x_i=x$ , and  $A$  and  $B$  are constants. This mechanism supposes that if the update refers to quantities such as the number of connections or the number of clicks of a site, the probability that a particular site is chosen to update is proportional to the number of connections or clicks that it has previously accumulated. More popular sites concentrate thus higher attention, leading to an agglomeration process that, after a while, produces a very wide distribution of values of  $x$ . If the relation of Eq. (1) is linear, the distribution  $P(x)$  can be approached by a decaying power-law function with an exponent depending on the values of  $A$  and  $B$  [4]. If it is not linear, two simple scenarios can occur. Either  $\Delta x$  grows with  $x$  faster than linear and the most popular element will eventually congregate a finite fraction of all the available value of  $x$ , or it is sublinear and the distribution of values of  $x$  will not be wide (stretched exponential instead of a power law) [4,38,39].

In our case, the “elements” of the system are Web pages and IPs, and the quantity  $x$  can be, among other things, the number of clicks of a certain user on a given URL, which we call  $w$ , the number of different users that an URL receives,  $k_{\text{URL}}$ , or the number of different sites that an IP visits,  $k_{\text{IP}}$ . We have also performed a similar study for the activity of the URLs and IPs (defined as the number of requests received or sent), but the results are similar. We will focus our attention, therefore, only on  $k_{\text{URL}}$ ,  $k_{\text{IP}}$ , and  $w$ . The variation of each of these variables  $\Delta x$  in a single day is measured after having accumulated the values of  $x$  for a full week. Then an average

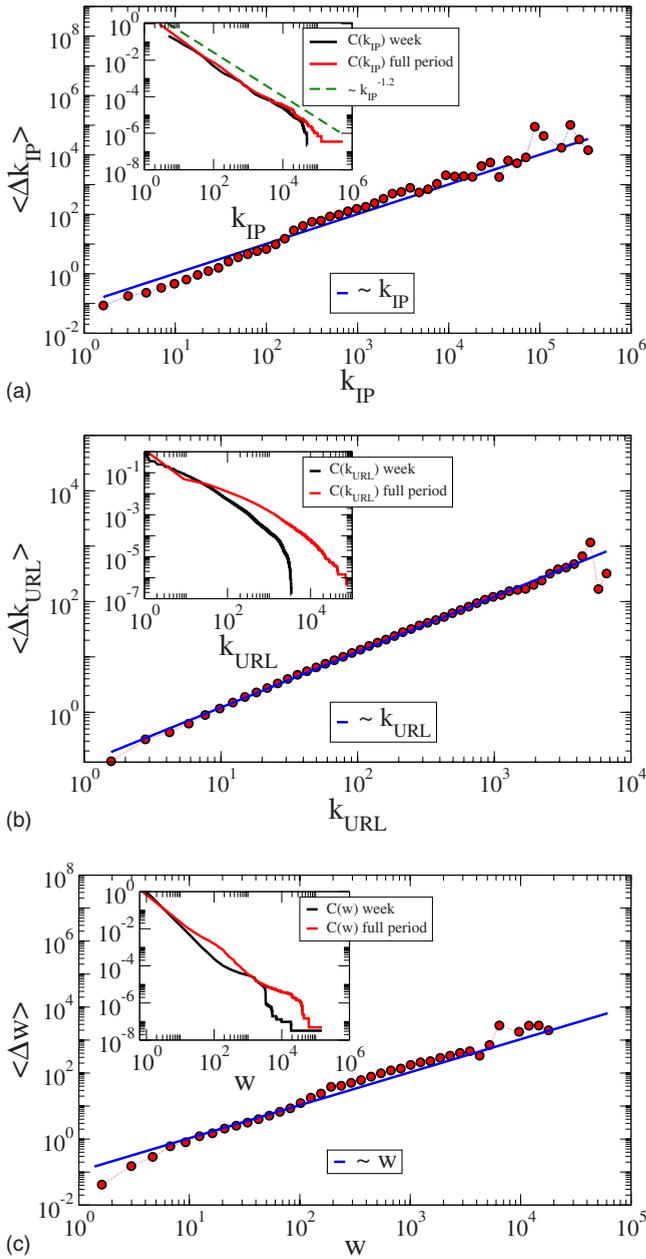


FIG. 7. (Color online) (a) Average variation in a single day of the number of different visited sites,  $\langle \Delta k_{IP} \rangle$ , as a function of the number of sites already seen during the previous week,  $k_{IP}$ . (b) The same type of function but for the number of visitors to a URL,  $\langle \Delta k_{URL} \rangle$ . (c) The average day variation of the number of clicks on each connection IP-URL as a function of the clicks accumulated during the previous week,  $\langle \Delta w \rangle(w)$ . The insets display the cumulative distributions for each quantity, the black curves are obtained by splitting the database in one-week periods and averaging over all of them, while the red ones are the distributions for the full 292-day period.

is taken over all the weeks of the database. The results displaying  $\Delta x$  as a function of  $x$  are depicted in Fig. 7. The variation of  $k_{IP}$ ,  $k_{URL}$ , and  $w$  can be well approached by linear preferential functions similar to Eq. (1) (straight lines in the main plots). This means that the rate at which users explore the Web ( $\Delta k_{IP}$ ), the rate at which popular pages at-

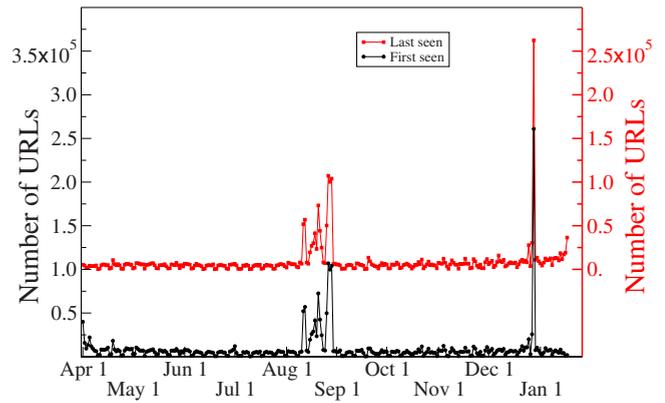


FIG. 8. (Color online) The number of URLs that are first (last) seen as a function of time. The two major “extinction” and “creation” events correspond to the beginning and end of the semester and closely match the peaks detected in Fig. 2.

tract new users ( $\Delta k_{URL}$ ), and the rate at which users revisit Web pages ( $\Delta w$ ) depend linearly on the previous week’s performance. It should also imply that the distributions  $P(k_{IP})$ ,  $P(k_{URL})$ , and  $P(w)$  are wide and well fitted by a power law. In order to check this last point, we have measured the cumulative distributions  $C(x) = \int_x^\infty dy P(y)$  for the three quantities. The cumulative distribution  $C(x)$  is the probability of having a value of the variable greater than  $x$  and usually exhibits better statistics than  $P(x)$ . Note that if  $P(x)$  goes as  $P(x) \sim x^{-\gamma}$ , then  $C(x) \sim x^{1-\gamma}$ . The results are shown in the insets of Fig. 7. In these plots, we have also included the cumulative distributions estimated aggregating the values of  $k_{URL}$ ,  $k_{IP}$ , and  $w$  for the whole period of the database (292 days). The comparison of the cumulative distributions obtained for the two time windows reveals an important surprise. For  $C(k_{IP})$ , the two curves overlap and can be fitted with a power law of exponent  $\gamma \approx 2.2$ . However, this is not true for the popularity of the URLs,  $k_{URL}$ , or for  $w$ . This difference in the output depending on the extension of the time window has important consequences for modeling the dynamics of the system. Its origin is related to the fact that in a university, the time during which a site, or more specifically its content, is relevant closely tracks the evolution of the academic year. In general, a similar rule should apply to all the Web sites. The lifetime can be more flexible, depending also on the number of visitors, but a certain loss of interest as time passes after the first online publication can be expected [12]. After this time, the page does not attract new users or visits from the old ones at the same rhythm (if it attracts any at all). This breaks one of the implicit assumptions of preferential linking: new elements are added at a constant rate, while the old ones keep attracting attention indefinitely. It also implies that linear preferential linking is not valid for longer time windows for  $k_{URL}$  and  $w$ , and that their distributions cannot be modeled as simple (stable over time windows) power laws.

To visualize the life story of URLs, we represent in Fig. 8 the number of pages first seen or last seen in the system as a function of time. We will say that a certain URL  $\mathcal{U}$  is first seen at time  $t$  if it receives its first request at  $t$ . Complementarity, the time in which  $\mathcal{U}$  is last seen, disappearing from the

database, is when it receives the last registered visit. Note that, although similar in look, this plot is different from Fig. 2, where we plot the activity measured as the total number of clicks on the Emory domain as a function of time. Two large peaks can be seen in Fig. 8. The time of these peaks coincides with the end and beginning of the semester. Many Web pages seem thus to have a relatively short life, probably being set up by professors or students who abandon them at the end of the semester. In many cases, even the *http* addresses are no longer maintained.

## V. DISCUSSION AND CONCLUSIONS

Web server logs have proven to be an important source of information regarding human dynamics. Here we have offered an extensive study on the medium-sized Web domain of Emory University, tracking the users in a consistent way for 292 days. A clear signal of human circadian rhythms has been obtained as well as activity patterns that seem to be universal since they are in agreement with previous results on mobile phone records or email posting in social sites. In addition, by comparing the traffic of an ideal average week with the week containing Thanksgiving, we have shown how major holidays induce changes in the structure of the activity of the whole population, not only during the vacation time but also days before and after. Some of these days concentrate an important level of activity, much higher than the usual one, and others fall quite behind.

After the characterization of activity at the whole university scale, we moved our focus to the study of statistics of single users. The difference in the navigation patterns between humans and automatic processes, either malicious or friendly, has been highlighted. Humans are in general more unpredictable, although a similar behavior might be reproduced by sophisticated automatic means. In particular, for

human users, it is important to analyze the statistics of the times between events (clicks) and compare them with recently introduced models based on priority queues. We have shown that indeed such models are able to explain the inter-clicks period distribution if the dyad user-site is considered. Furthermore, a simple modification, in which the number of tasks to execute in a short interval of time is higher than one, can also account for the statistics of times between requests of the same user on the whole Emory domain.

Finally, we have explored another mechanism that has been proposed as an important ingredient in the development of the WWW, namely “preferential attachment.” Linear preferential attractiveness is detected in all aspects of the traffic contemplated: the rate of exploration of new sites by users, the capture of new visitors by the sites, or the new clicks received on each connection user-Web page. In all these cases, the linear relation holds for a short period of time. For longer periods, the lifetime of the Web pages must be taken into account, complicating substantially the scenario. Preferential linking, priority queuing, and Web page aging seem thus to be essential factors for any model aimed at characterizing Web surfing.

## ACKNOWLEDGMENTS

The authors would like to thank Alain Barrat, Stefan Boettcher, Ciro Cattuto, Helmut Katzgraber, Filippo Menczer, Muhittin Mungan, Filippo Radicchi, and in general the members of the Cx-Nets collaboration for useful discussions and comments. We would also like to thank the IT service of Emory University for access to the database. Funding from the Lagrange Project of the CRT Foundation (Torino, Italy) and from the National Science Foundation under Grant No. 0312150 was received. The use of computer resources provided by the Open Science Grid supported by the NSF and by the Office of Science of the U.S. Department of Energy is acknowledged.

- 
- [1] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner, *Science* **313**, 769 (2006).
  - [2] D. J. Watts, *Nature* **445**, 489 (2007).
  - [3] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004).
  - [4] S. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
  - [5] D. J. Watts and S. H. Strogatz, *Nature* **393**, 409 (1998).
  - [6] B. A. Huberman, P. L. Pirolli, J. E. Pitkow, and R. M. Lukose, *Science* **280**, 95 (1998).
  - [7] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
  - [8] F. Menczer, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14014 (2004).
  - [9] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **63**, 056125 (2001).
  - [10] C. Cattuto, V. Loreto, and V. D. P. Servedio, *Europhys. Lett.* **76**, 208 (2006).
  - [11] M. V. Simkin and V. P. Roychowdhury, *Europhys. Lett.* **82**, 28006 (2007).
  - [12] F. Wu and B. A. Huberman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17599 (2007).
  - [13] *Encyclopedia Of Sociology—Volume I*, edited by E. F. Borgatta and R. J. V. Montgomery, 2nd ed. (Macmillan Reference, New York, 2000).
  - [14] J.-P. Onnela *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).
  - [15] S. Golder, D. Wilkinson, and B. A. Huberman, e-print arXiv:cond-mat/0611137.
  - [16] J. Candia *et al.*, *J. Phys. A* **41**, 224015 (2008).
  - [17] A. Vázquez, *Physica A* **373**, 747 (2007).
  - [18] M. R. Meiss, F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani, *Ranking Web Sites with Real User Traffic*, *Proc. WSDM* (Sheridan Printing, Alpha, NJ, 2008).
  - [19] T. Zhou, X.-P. Han, and B.-H. Wang, arXiv:0801.1389.
  - [20] T. Zhou, H.-A. T. Kiet, B. J. Kim, B.-H. Wang, and P. Holme, *Europhys. Lett.* **82**, 28002 (2008).
  - [21] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
  - [22] S. Dill *et al.*, *Biosens. Bioelectron.* **2**, 205 (2002).

- [23] M. R. Meiss, F. Menczer, and A. Vespignani, *On the Lack of Typical Behavior in the Global Web Traffic Network*, Proc. WWW (Association for Computing Machinery, Denver, 2005).
- [24] L. D. Catledge and J. E. Pitkow, *Comput. Netw.* **27**, 1065 (1995).
- [25] C. Cattuto, V. Loreto, and L. Pietronero, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1461 (2007).
- [26] A. Capocci *et al.*, arXiv:physics/0602026.
- [27] A. Vázquez, J. G. Oliveira, Z. Dezso, K. I. Goh, I. Kondor, and A. L. Barabasi, *Phys. Rev. E* **73**, 036127 (2006).
- [28] A.-L. Barabási, *Nature* **435**, 207 (2005).
- [29] J. G. Oliveira and A.-L. Barabási, *Nature* **437**, 1251 (2005).
- [30] A. Vázquez, *Phys. Rev. Lett.* **95**, 248701 (2005).
- [31] J. G. Oliveira and A. Vázquez, arXiv:0710.4916.
- [32] E. H. Simon, *Biometrika* **42**, 425 (1955).
- [33] S. Bornholdt and H. Ebel, *Phys. Rev. E* **64**, 035104(R) (2001).
- [34] S. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, e-print arXiv:cond-mat/0009090.
- [35] A.-L. Barabási *et al.*, *Physica A* **311**, 590 (2002).
- [36] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [37] S. Redner, *Phys. Today* **58**, 49 (2005). See arXiv:physics/0506056 for a more extensive version.
- [38] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [39] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001).