Agents, Bookmarks and Clicks: A topical model of Web navigation

Mark R. Meiss[1,3]* Bruno Gonçalves[1,2,3] José J. Ramasco[4]
Alessandro Flammini[1,2] Filippo Menczer[1,2,3,4]

[1]School of Informatics and Computing, Indiana University, Bloomington, IN, USA
[2]Center for Complex Networks and Systems Research, Indiana University, Bloomington, IN, USA
[3]Pervasive Technology Institute, Indiana University, Bloomington, IN, USA
[4]Complex Networks and Systems Lagrange Laboratory (CNLL), ISI Foundation, Turin, Italy

ABSTRACT

Analysis has shown that the standard Markovian model of Web navigation is a poor predictor of actual Web traffic. Using empirical data, we characterize several properties of Web traffic that cannot be reproduced with Markovian models but can be explained by an agent-based model that adds several realistic browsing behaviors. First, agents maintain bookmark lists used as teleportation targets. Second, agents can retreat along visited links, a branching mechanism that can reproduce behavior such the back button and tabbed browsing. Finally, agents are sustained by visiting pages of topical interest, with adjacent pages being related. This modulates the production of new sessions, recreating heterogeneous session lengths. The resulting model reproduces individual behaviors from empirical data, reconciling the narrowly focused browsing patterns of individual users with the extreme heterogeneity of aggregate traffic measurements, and leading the way to more sophisticated, realistic, and effective ranking and crawling algorithms.

Categories and Subject Descriptors: H.3.4 [Information Storage and Retrieval]: Systems and Software–*Information networks*; H.4.3 [Information Systems Applications]: Communications Applications–*Information browsers*; H.5.4 [Information Interfaces and Presentation]: Hypertext/ Hypermedia–*Navigation*

General Terms: Algorithms, Measurement

Keywords: Web links, navigation, traffic, clicks, browsing, entropy, sessions, agent-based model, bookmarks, back button, interest, topicality, PageRank, BookRank

1. INTRODUCTION

Despite its simplicity, PageRank [5] is a well-established model that characterizes Web browsing as a random surfing activity. As people spend more and more time online, their Web traces provide an increasingly informative window into human behavior, enabling systematic testing of PageRank's underlying navigation model [12]. Traffic patterns aggregated across users reveal that the key assumptions of uniform random walk and teleportation are widely violated, making PageRank a poor predictor of traffic. While the intent of PageRank is to gauge the importance of pages rather than predict traffic, traffic is a direct empirical indication of the esteem in which users hold a page. Here we expand on our previous empirical analysis [12, 11] by considering *individual* traffic patterns [8], which suggest the need for an *agent-based* model with more realistic features, such as memory and topicality, to account for individual traffic patterns observed in real-world data.

We previously introduced *BookRank*, a browsing model that adds a memory mechanism to PageRank [2]. Here we introduce a novel agent-based model that also accounts for the topical interests of users. We compare the traffic patterns of these models with individual Web traffic data from a field study of 1,000 users. Our main contributions include:

- We show that the diversity of pages visited by individual users is not well-predicted by either PageRank or BookRank, suggesting that users have focused interests and recurrent habits. The diversity apparent in aggregate measures of traffic follows from the diversity across individual interests.
- Using *logical sessions* (cf. §3), we find including a simple memory mechanism (as in the BookRank model) is insufficient to capture broad distributions of session size and depth.
- We present *ABC*, an agent-based model with three key ingredients: (1) *bookmarks* are used as teleportation targets, defining session boundaries and capturing the diverse popularity of starting pages; (2) a *back button* accounts for branching observed in empirical traffic; and (3) *topical interests* drive whether agents continue browsing or start new sessions, yielding diverse session sizes. ABC also incorporates *topical locality*, so that an interesting page is likely to link to other such pages.
- Finally, we demonstrate that ABC outperforms both PageRank and BookRank in modeling individual statistics such as entropy and size and depth of sessions.

2. BACKGROUND

There have been many empirical studies of Web traffic patterns, most commonly through analysis of Web server logs. This methodology allows us to distinguish individual users though their IP addresses, thus capturing *individual* traffic patterns [8]. However, the choice of target server biases both the sample of users and the sample of the Web graph being observed. An alternative source of

*Corresponding author. Email: mmeiss@indiana.edu

HT'10, June 13–16, 2010, Toronto, Ontario, Canada..

Web traffic data is browser toolbars, which gather data based on the surfing activity of many users. Such a method is still biased by users who have opted to install the software, and the data are not generally available to researchers. Adar *et al.* [1] used this approach to study patterns of page revisitation without regard to sessions. A related approach is to identify a panel of desirable users and have them install tracking software, which eliminates many sources of bias but incurs significant experimental costs. Such an approach has been used to describe the exploratory behavior of Web surfers [3]. These studies did not propose models to explain the observed traffic patterns. The methodology of our study captures traffic data directly from a running network, an approach first adopted by Qiu *et al.* [15], who used captured HTTP packet traces to investigate how browsing behavior is driven by search engines.

We particularly focus on the statistical characterization of browsing sessions. A common assumption is that long pauses correspond to breaks between sessions. Based on this assumption, many researchers have relied on timeouts as a way of defining sessions, a technique we have found to be flawed [11], motivating the definition of time-independent *logical sessions,* based on building session trees rooted at pages requested without a referrer. The model we present is in part aimed at explaining the broad distributions of size and depth empirically observed for these logical sessions.

Other researchers have suggested more plausible models to capture features of real Web browsing such as the back button [10, 4]. The interplay between user interests and page content in shaping browsing patterns has also been studied. Huberman *et al.* proposed a model in which visited pages have interest values described by a random walk; navigation continues while the current page has a value above a threshold [9]. This kind of model is closely related to algorithms designed to improve topical crawlers [14].

In preliminary results, we proposed a model in which the user maintain a list of bookmarks from which they start new sessions [2]. We called this model *BookRank,* since bookmark selection is controlled by a ranking based on the frequency of visits. This model reproduces many characteristics observed in empirical traffic data, including page and link traffic distributions, but fails to account for features of the navigation patterns of individual users, such as entropy and session characteristics. In the remainder of this paper, we extend the BookRank model to address these shortcomings.

3. EMPIRICAL TRAFFIC DATA

The HTTP request data were gathered from one of the undergraduate dormitories at Indiana University under conditions described in previous work [11]. This data set consists of (referrer, target) pairs for HTTP requests associated with actual page fetches; it contains roughly 30 million requests, 1,000 users, 2.5 million distinct URLs. Using this filtered set of HTTP requests ("clicks"), we organize each user's clicks into sessions. These sessions are not based on a simple timeout, which previous analysis has shown to be arbitrary and misleading [11]. Instead, we organize the clicks into tree-based *logical* sessions according to an algorithm described formally in our previous work [11]. The key notions are that new sessions begin with requests with an empty referrer; that each request represents a directed edge from a referring URL to a target URL; and that requests are assigned to the session in which their referring URL was most recently requested.

These session trees mimic the multitasking behavior of users with modern browsers: a user can have several active sessions. The key properties of the trees, such as size and depth, are also relatively insensitive to an additional timeout constraint [11]. In this analysis, we include such a timeout: a click cannot be associated with a session tree that has been dormant for thirty minutes.

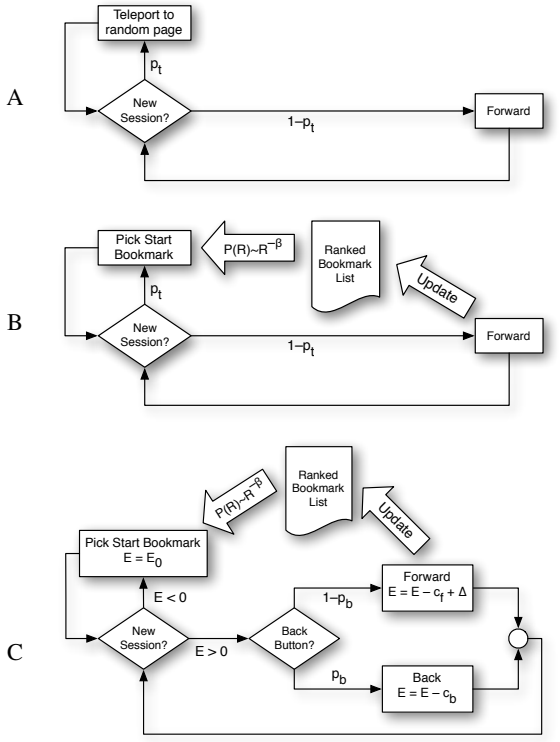

Figure 1: Schematic illustrations of the PageRank (A), BookRank (B), and ABC (C) models.

Most importantly, the tree structure allows us to infer how users backtrack as they browse. Because modern browsers use caching mechanisms to improve performance, unless overridden by HTTP options, browsers generally do not repeat a recent request. We thus do not observe multiple links pointing to the same page (within a single logical session), giving us *direct* way of determining when a user backtracks. However, session trees allow us to *infer* backward traffic: if the next request in a tree comes from a URL other than the most recently visited one, the user must have navigated to that page, or else opened it in a separate tab.

To characterize the properties of our traffic data and evaluate the models proposed later in this paper, we examine several distributions that focus on the properties of individual users and sessions. For an individual user j, the *Shannon information entropy* is defined as $S_j = -\sum_i \rho_{ij} \log_2 \rho_{ij}$, where ρ_{ij} is the fraction of visits of user j to site i aggregated across sessions. The *session size* is defined as the number of unique pages visited in a logical session. The *session depth* is the maximum tree distance between the starting page of a session and any page visited within the same session.

Note that we have already characterized aggregate distributions such as page and link traffic in preliminary work [2, 11]. Another feature sometimes used to characterize random browsing behavior is the distribution of return time, which in this case would be the number of clicks between two consecutive visits to the same page by a given user [8, 2]. However, cache behavior and overlapping sessions mean that this information cannot be retrieved in a reliable way from the empirical data.

To properly analyze these distributions, we compare them with those generated by two reference models based on PageRank-like modified random walkers with teleportation probability $p_t = 0.15$. To obtain a useful reference model for traffic data that is based on

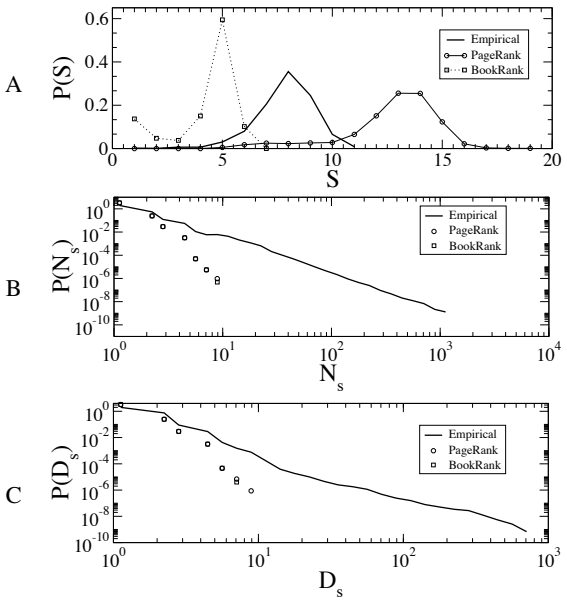

Figure 2: Empirical and baseline distributions of user entropy (A), session size (B), and session depth (C).

individuals, we imagine a population of PageRank random walkers, as many as the users in our study. The first reference model (PageRank) is illustrated in Fig. 1A. Each walker browses for as many sessions as there were empirical sessions for the corresponding real-world user. The PageRank sessions are terminated by the constant-probability jumps, so the total number of pages visited by a walker may differ from the corresponding user. Teleportation jumps lead to session-starting pages selected uniformly at random.

The BookRank model is shown in Fig. 1B. The key realistic ingredient differentiating this model from PageRank is related to memory: agents maintain individual lists of bookmarks chosen as teleportation targets based on the number of previous visits. Initially, each agent randomly selects a starting page (node). With probability $1 - p_t$, the agent navigates locally, following a link from the present node selected with uniform probability. Unless previously visited, the new node is added to the bookmark list. The frequency of visits is recorded, and the list of bookmarks is kept ranked from most to least visited. With probability p_t, the agent teleports (jumps) to a previously visited page (bookmark), initiating a new session. The bookmark with rank R is chosen with probability $P(R) \propto R^{-\beta}$.

This mechanism mimics uses of frequency ranking in modern browsers, such as URL completion in the address bar and suggested starting pages in new windows. The functional form $P(R)$ is motivated by data on selection among ranked lists of results [7].

In our simulations, browsing occurs on scale-free networks with N nodes and degree distribution $P(k) \sim k^{-\gamma}$, generated using the growth model of Fortunato *et al.* [16]. We used $N = 10^7$ nodes, ensuring a network larger than the number of pages visited in the empirical data. We also set $\gamma = 2.1$ to match the Web and our data set. To prevent dangling links, we construct this graph with symmetric links. Within a session, we simulate the browser's cache by recording traffic only when the target page is new to a particular session. This allows us to measure the number of unique pages visited in a session, which corresponds to the empirical session size. We assume that that cached pages are reset between sessions.

Preliminary work examined aggregate system properties; our focus here is on characterizing individual users and sessions. The simplest hypothesis is that broad distributions of aggregate behavior reflect extreme variability within the traffic generated by single users, none of whom are typical in the sense of overall traffic. To capture the diversity of behavior across users, we examine the entropy of each user's traffic, which directly measures the focus of a user's interests. Given an arbitrary number of visits N_v, the entropy is maximum ($S = N_v \log(N_v)$) when N_v pages are visited once, and minimum ($S = 0$) when all visits have been paid to a single page. The distribution of entropy across users is shown in Fig. 2A. The reference PageRank model produces higher entropy than observed in the empirical data; a PageRank walker picks starting pages with uniform probability, whereas a real user is more likely to start from a previously visited page and thus revisit neighboring pages. BookRank encourages such behavior, and we indeed observe lower entropy values in Fig. 2A. However, BookRank underestimates the entropy and its variability across users.

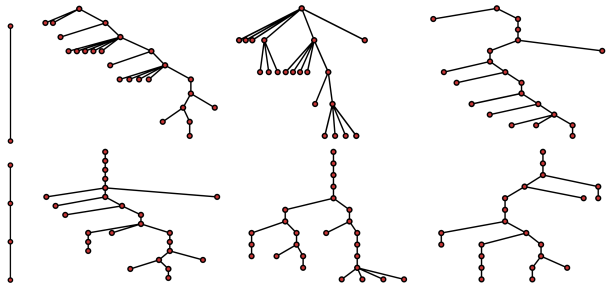

Figure 3: Representation of a few typical and representative session trees from the empirical data (top) and from the ABC model (bottom). Animations are available at `cnets.indiana.edu/groups/nan/webtraffic`.

We can also consider the distributions that characterize logical sessions, namely the size (number of unique pages) and depth (distance from a session's starting page) distributions. Figs. 2B and C show that both empirical distributions are rather broad, spanning three orders of magnitude, revealing a surprisingly large proportion of very long sessions. In contrast, both the PageRank and BookRank reference models generate very short sessions. The probabilistic teleportation mechanism that determines when a PageRank walker starts a new session is incapable of capturing broadly distributed session sizes. In fact, session size is upper-bounded by the length ℓ (number of clicks) of a session, which exhibits a narrow, exponential distribution $P(\ell) \sim (1 - p_t)^{\ell}$. Note that the exponentially short sessions are not inconsistent with the high entropy of PageRank walkers (Fig. 2A), which is a result of the frequent jumps to random targets rather than the browsing behavior.

4. ABC MODEL

The empirical analysis in the previous section demonstrates that a more sophisticated model of user behavior is needed to capture individual navigation patterns. We build upon the BookRank model by adding two additional ingredients.

First, we provide agents with a backtracking mechanism, needed to capture the tree-like structure of sessions (cf. top row of Fig. 3). Our data indicate that the incoming and outgoing traffic of a site are seldom equal, but have a ratio distributed over many orders of magnitude [12]. Teleportation alone cannot explain this violation of flow conservation, demonstrating that users' browsing sessions have many branches. Finally, our prior results show that the average

branch factor of session trees is almost two. All of these observations are consistent with the use of tabs and the back button.

The second ingredient concerns the fact that the BookRank model fails to predict individual statistics: all agents are identical, session size has a narrow, exponential distribution; and the entropy distribution is unsatisfactory. In the real world, the duration of a session depends on the individual intentions (goals) and interests of a user. Visiting relevant pages, those whose topics match the user's interests, will lead to more clicks and thus longer sessions. We therefore introduce agents with distinct *interests* and page *topicality* into the model. An agent spends attention when navigating to a new page and gains attention when visiting pages that match the user's interests. To model this, each agent stores some "energy" (units of attention) while browsing. Visiting a new page incurs higher energy cost than going back to a previously visited page. Known pages yield no energy, while unseen pages increase the energy store by some random amount that depends on the page's relevance to the agent. Agents continue to browse until they run out of energy, whereupon they start a new session.

We call the resulting model *ABC* for its main ingredients: agents, bookmarks and clicks. Clicks are driven by the topicality of pages and agent interests, in a way inspired by *InfoSpiders* [14], which were adaptive Web crawlers driven by similarity between search topics and page content. Better matches led to more energy and more exploration of local link neighborhoods. Irrelevant pages led to agents running out of energy and dying, so that resources would be allocated to more promising neighborhoods. In ABC, this idea is used to model browsing behavior, as shown in Fig. 1C. Each agent starts at a random page with an initial amount of energy E_0.

At each time step, if $E \leq 0$, the agent starts a new session by teleporting to a bookmark chosen as in BookRank. Otherwise, the agent continues the current session, following a link from the present node. With probability p_b, the back button is used, leading back to the previous page, and the agent's energy is decreased by a fixed cost c_b. Otherwise, a forward link is clicked with uniform probability. The agent's energy is updated to $E - c_f + \Delta$, where c_f is a fixed cost and Δ is a stochastic value representing the new page's relevance to the user. As in BookRank, the bookmark list is updated with new pages and ranked by visit frequency.

The dynamic variable Δ in the ABC model is a measure of relevance of a page to a user's interests. The simplest way to model relevance is by a random variable, in which case stored energy behaves as a random walk, and the session duration ℓ (number of clicks until $E = 0$) has a power-law tail $P(\ell) \sim \ell^{-3/2}$ [9]. However, our empirical results suggest a larger exponent [11]: we know that content similarity between two pages is correlated with their graph distance, as is the change that a page is relevant to some given topic [6, 13]. Neighbor pages are likely to be related topically, and the relevance of page t to a user is related to the relevance of a page r that links to t. To capture such *topical locality*, we introduce correlations between the Δ values of consecutively visited pages. For the starting page we use an initial value $\Delta_0 = 1$. Then, when a page t is first visited *in a given session,* we set $\Delta_t = \Delta_r(1 + \epsilon)$, where r is the referrer page, ϵ is a uniform random variable in $[-\eta, \eta]$, and η controls the degree of topical locality. In a new session we assume a page can again provide energy, even if it was visited in a previous session. However, it will yield different energy in different sessions, reflecting changing interests.

5. MODEL EVALUATION

We ran two sets of simulations of ABC using distinct scale-free graphs. One (G1) is the artificial network discussed in § 3; the second (G2) is derived from an independent subset of the Web graph based on the largest component from a traffic network generated by the activity of about 100,000 people[12]. G2 is based on three weeks of traffic in November 2009; it has $N = 8.14 \times 10^6$ nodes and the same degree distribution, with exponent $\gamma \approx 2.1$.

Within each session we simulate the browser's cache so that we can compare the number of unique pages visited by the model agents directly to the empirical session size.

The proposed models have various parameters. In prior work [16], we have shown that the distribution of traffic with empty referrer generated by our models is related to the parameter β. Namely, the distribution is well approximated by a power law $P(T_0) \sim T_0^{-\alpha}$, where $\alpha = 1 + 1/\beta$. Empirical study shows that $\alpha \approx 1.75$ for the Web, so we set $\beta = 1.33$. The back button probability $p_b = 0.5$ is also taken empirically. The initial energy E_0 and the forward and backward costs c_f and c_b are closely related and control session durations. We therefore set $E_0 = 0.5$ arbitrarily and use an energy balance argument to find suitable values of the costs. Empirically, the average session size is close to two pages. The net loss per click of an agent is $-\delta E = p_b c_b + (1 - p_b)(c_f - \langle\Delta\rangle)$ where $\langle\Delta\rangle = 1$ is the expected value from a new page. By setting $c_f = 1$ and $c_b = 0.5$, we obtain an expected session size $1-(1-p_b)E_0/\delta E = 2$. In general, higher costs lead to shorter sessions and lower entropy. We explored the sensitivity of the model to the topical locality parameter η through simulation, settling on $\eta = 0.15$. Smaller values give all pages similar relevance, making session distributions too narrow. Larger values imply more noise (absence of topical locality), making them too broad. The results shown below refer to this combination of parameters.

The number of users in the simulation, and the number of sessions for each user, are taken from the empirical data. Because the model is computationally intensive, we partitioned the simulated users into work queues of roughly equal session counts, which we executed in parallel on a high-performance computing cluster.

The simulations of the ABC model users generate session trees that can be compared visually to those in the empirical data, as shown in Fig. 3. For a more quantitative evaluation of our model, we compare its results with empirical findings described in § 3. For each of the distributions discussed earlier, we also compare ABC with the reference BookRank model. The latter is simulated on the artificial G1 network.

Let us consider how our model captures the behavior of single users. The entropy distribution across users is shown in Fig. 4A, where the model predictions are compared with the distribution found in the empirical data. The ABC model yields entropy distributions that are somewhat sensitive to the underlying network, but that in any case fit the empirical entropy data much better than BookRank, in terms of both the location of the peak and the variability across users. This result suggests that bookmark memory, back button, and topicality are crucial ingredients in explaining the focused habits of real users.

Having characterized traffic patterns from aggregating across user sessions, we can study the sessions one by one and analyze their statistical properties. In Fig. 4B, we show the distribution of session size as generated by the ABC model. The user interests and topical locality ingredients account for the broad distribution of session size, capturing that of the empirical data much better that the short sessions generated by the BookRank reference model. Agents visiting relevant pages tend to keep browsing, and relevant pages tend to lead to other interesting pages, explaining the longer sessions. We argue that the diversity apparent in the aggregate measures of traffic is a consequence of this diversity of individual interests rather than the behavior of extremely eclectic users who visit a wide variety of Web sites—as shown by the narrow distribution of entropy.

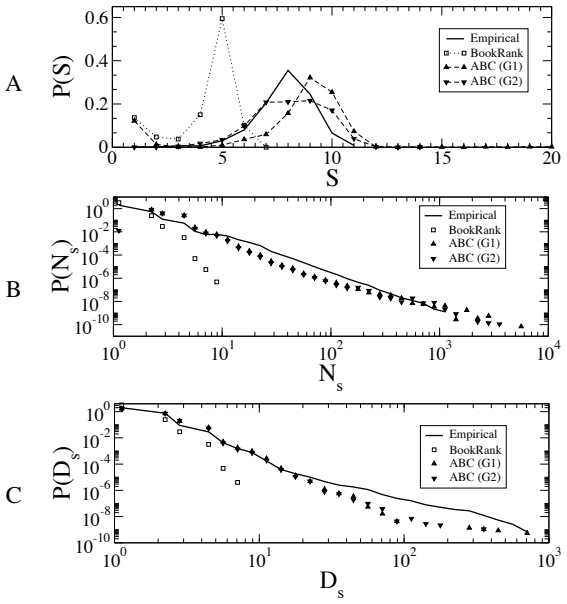

Figure 4: Distributions of user entropy (A), session size (B), and session depth (C) generated by ABC, with baseline comparison.

The entropy distribution discussed above depends not only on session length, but also on how far each user navigates away from the initial bookmark where a session is initiated. One way of analyzing this is by the distribution of session depth, as shown in Fig. 4C. The agreement between the empirical data and the ABC model is excellent and significantly better than the one observed with the BookRank baseline. Once again topicality is shown to be a key ingredient to understand real user behavior on the Web.

6. CONCLUSIONS

Previous studies have shown that Markovian processes such as PageRank cannot explain many aggregate patterns observed in real Web traffic, especially the diversity of session starting points, link traffic, and the session dimensions. Furthermore, despite such diverse aggregate measurements, individual behaviors are quite focused, calling for a non-Markovian agent-based model. Our ABC model is able to reproduce aggregate traffic patterns while offering a mechanism that can generate key properties of logical sessions. We can thus argue that the diversity apparent in page, link, and bookmark traffic is a consequence of the diversity of individual interests rather than the behavior of very eclectic users.

The ABC model is more complex than prior models such as PageRank or BookRank. However, its predictive power suggests that bookmarks, tabbed browsing, and topicality are relevant features of how we browse the Web. In addition to the descriptive power of ABC, our results may lead to more sophisticated and effective ranking and crawling algorithms. While we have attempted to make reasonable and realistic choices for the parameters of ABC, further work is needed to achieve a more complete understanding of parameter space. We know that network size, costs, and topical locality do play a key role in modulating the balance between individual diversity (entropy) and session size. While ABC is a clear step forward, it still shares some limitations present in previous efforts, especially the uniform choice among outgoing links from a page. This may account for the imperfect match between the entropies of our model agents and those of actual users.

Acknowledgments

The authors thank the Advanced Network Management Laboratory and the Center for Complex Networks and Systems Research, both part of the Pervasive Technology Institute at Indiana University; L. J. Camp of the IU School of Informatics and Computing, for support and infrastructure; and the network engineers of Indiana University for support in data collection. This work was produced in part with support from the Institute for Information Infrastructure Protection research program, managed by Dartmouth College and supported under Award 2003-TK-TX-0003 from the U.S. DHS, Science and Technology Directorate. BG was supported in part by grant NIH-1R21DA024259 from the NIH. JJR is funded by the project 233847-Dynanets of the European Union Commission. This material is based upon work supported by the NSF award 0705676. This work was supported in part by a gift from Google. Opinions, findings, conclusions, recommendations or points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security, Science and Technology Directorate, I3P, National Science Foundation, Indiana University, Google, or Dartmouth College.

7. REFERENCES

[1] Eytan Adar, Jaime Teevan, and Susan Dumais. Large scale analysis of web revisitation patterns. In *Proc. CHI*, 2008.

[2] B. Gonçalves, M.R. Meiss, J.J. Ramasco, A. Flammini an d F. Menczer. Remembering what we like: Toward an agent-based model of Web t raffic. *Late Breaking Results WSDM*, 2009.

[3] Thomas Beauvisage. The dynamics of personal territories on the web. In *HT '09: Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 25–34, New York, NY, USA, 2009. ACM.

[4] M. Bouklit and F. Mathieu. BackRank: an alternative for PageRank? In *Proc. WWW Special interest tracks and posters*, pages 1122–1123, 2005.

[5] S Brin and L Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks*, 30(1–7):107–117, 1998.

[6] BD Davison. Topical locality in the Web. In *Proc. 23rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 272–279, 2000.

[7] S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani. Topical interests and the mitigation of search engine bias. *Proc. Natl. Acad. Sci. USA*, 103(34):12684–12689, 2006.

[8] B. Gonçalves and J. J. Ramasco. Human dynamics revealed through web analytics. *Phys. Rev. E*, 78:026123, 2008.

[9] BA Huberman, PLT Pirolli, JE Pitkow, and RM Lukose. Strong regularities in World Wide Web surfing. *Science*, 280(5360):95–97, 1998.

[10] F. Mathieu and M. Bouklit. The effect of the back button in a random walk: application for PageRank. In *Proc. WWW Alternate track papers & posters*, pages 370–371, 2004.

[11] M. Meiss, J. Duncan, B. Gonçalves, J. J. Ramasco, and F. Menczer. What's in a session: tracking individual behavior on the Web. In *Proc. 20th ACM Conf. on Hypertext and Hypermedia (HT)*, 2009.

[12] M. Meiss, F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani. Ranking web sites with real user traffic. In *Proc. WSDM*, pages 65–75, 2008.

[13] F Menczer. Mapping the semantics of web text and links. *IEEE Internet Computing*, 9(3):27–36, May/June 2005.

[14] F Menczer, G Pant, and P Srinivasan. Topical web crawlers: Evaluating adaptive algorithms. *ACM Transactions on Internet Technology*, 4(4):378–419, 2004.

[15] Feng Qiu, Zhenyu Liu, and Junghoo Cho. Analysis of user web traffic with a focus on search activities. In *Proc. 8th International Workshop on the Web and Databases (WebDB)*, pages 103–108, 2005.

[16] A. Flammini S. Fortunato and F. Menczer. Scale-free network growth by ranking. *Phys. Rev. Lett.*, 96:218701, 2006.