

# Identifying Web Browsing Trends and Patterns

Alan L. Montgomery and Christos Faloutsos  
Carnegie Mellon University

Since it entered the popular culture in 1994, the World Wide Web has grown from approximately two million servers to more than 110 million in 2001, according to the Internet Software Consortium. Jupiter Media Metrix, an Internet research company, estimates that during this same period, the number of US home Web users has likewise increased from 3 million to more than 89 million. Has this phenomenal growth resulted in fundamental changes in the way users browse the Web?

To address this question, we analyzed a sample of more than 20,000 Internet users who accessed the Web from July 1997 through December 1999. This nationally representative data set collected by Jupiter Media Metrix is unique because it was gathered from users' computers, not servers, thus it is immune to caching problems present in other Web usage studies.

## EXPONENTIAL GROWTH

One straightforward measure of Internet activity is the number of viewings by a Web user per month. A *viewing* means that the user's browser window displays a URL. Due to caching, the number of viewings always equals or exceeds the number of requests or hits that the server logs. For example, the median number of viewings per month in December 1999 was 310, more than twice the 150 viewings recorded in July 1997. During this same period, the number of Web users increased 120 percent, from 28 million to 62 million.

Our study suggests not only that the number of Web users is increasing, as expected, but also that each user is viewing more pages. As the logarithmic scale in Figure 1 shows, the increase in viewings per month is fairly linear, indicating exponential growth. If the median continues to grow at the current rate of 2.4



**Although Internet usage has increased exponentially since the World Wide Web's inception, many browsing behaviors have remained surprisingly stable.**

percent per month, the median user's viewings will double every 30 months. This trend resembles Moore's law, which states that transistor density on integrated circuits will double every 18 months.

## MORE, NOT LONGER, SESSIONS

Examining Web usage on a session-by-session basis rather than monthly presents a different picture. A *session* is a period of sustained Internet usage; a new session begins when the user has not accessed the Web for more than two hours. The median number of sessions during December 1999 was four. If the present growth rate of 1.4 percent per month continues, the average number of sessions will double in a little more than four years. However, the median number of viewings per session in December

1999 was 48, only a slight increase from 42 in July 1997. By decomposing the number of viewings into sessions and the number of viewings per session, we found that growth in Web activity is primarily due to more frequent, not longer, browsing sessions.

## SMALL BUT PERSISTENT DOMAIN VIEWING

Contrary to expectations, the explosive growth in Web sites and the online information they provide have not encouraged users to examine a wider range of sources. In fact, the data indicates that users look at a fairly small number of sites relative to the number of viewings. The median number of domains viewed per session is six. Although the median number of hosts visited increased steadily from 14 in July 1997 to 25 in December 1999, the ratio of viewings to domains viewed remained fairly steady. Therefore, we can attribute the increase in domains viewed to the increased number of sessions, not to an

increased propensity of users to search out new domains within a session.

## FREQUENT REVISITS

Users revisit the same Web sites frequently. For example, during December 1999, users revisited 54 percent of URLs at least once during a session. Of the URLs that users revisited, 35 percent were viewed consecutively (two viewings of the same page occurred in a row), 22 percent had one viewing in between (the user likely returned to the page through the back button), and the remaining 43 percent had more than one viewing in between.

Users revisit Web domains even more frequently than pages. Our study revealed that users revisit a domain within one viewing 75 percent of the time. During

the study's entire two-and-one-half-year period, the probability of a user revisiting a domain during a session remained between 89 and 91 percent.

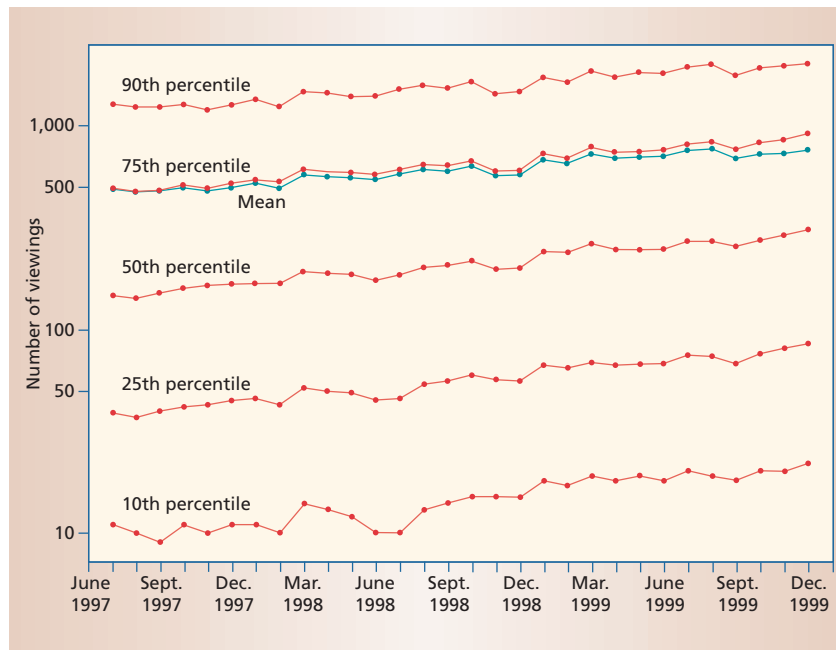
### WEB BROWSING PATTERNS

Our study showed that URL browsing patterns follow power laws, which are characterized by the relationship  $y = x^a$ , in which  $x$  and  $y$  are the variables of interest—for example, force and distance in Newton's law of gravity. A popular distribution with this property is the Zipf distribution. The Zipf distribution describes several Web browsing behaviors that are highly skewed, including how often a user revisits a URL within a session and the number of viewings a user makes before revisiting a URL or a domain. These distributions also remained stable during the study period in spite of the strong growth in usage.

The Zipf distribution of URL revisitations is somewhat analogous to the Zipf word usage pattern. We repeat certain words such as “the” extensively, while we use other unique words infrequently. These infrequently used words manifest a long tail in the Zipf distribution. Similarly, URLs form the Web user's vocabulary. The huge diversity of Web pages offers a large vocabulary for use in browsing the Internet. But browsers constantly reuse certain pages, perhaps as a navigational tool for moving to other pages, and they use other unique pages less frequently.

The empirical trends we observed yielded several surprises. Many Web browsing patterns, such as the number of pages and domains a user views during a computer session, have remained stable. Although overall Internet usage has grown exponentially, the time that an individual browses online has increased at a slower rate. Also, despite dramatic changes in its size and content, the way people interface with the Web has not significantly changed. Users remain loyal to hosts and persistent in their viewing habits.

These findings demonstrate that it is important to examine how users interact with the Internet as well as changes in the



**Figure 1. Distribution of monthly Web viewings on a logarithmic scale. Because the mean exceeds the median, a reverse-J shape with mass in the tail characterizes the percentile distribution. The mean is very close to the 75th percentile, indicating that heavy users are weighted more highly. The smaller percentiles are growing at a slightly faster rate than larger ones, implying that growth in Web viewings is stronger for light users than for heavy users.**

medium itself. In addition, future statistical measurements of browsing patterns must move beyond simple averages, which heavy users tend to skew, to capture the Internet population's growing diversity. \*

*Alan L. Montgomery is an associate professor at the Graduate School of Industrial Administration at Carnegie Mellon University. Contact him at alan.montgomery@cmu.edu.*

*Christos Faloutsos is an associate professor at the School of Computer Science at Carnegie Mellon University. Contact him at christos@cs.cmu.edu.*

**Editor: Ron Vetter, University of North Carolina at Wilmington, Department of Computer Science, 601 South College Rd., Wilmington, NC 28403-3297; voice +1 910 962 7192, fax +1 910 962 7457; vetterr@uncwil.edu**

# The IEEE Computer Society

publishes over 150 conference proceedings a year.

For a preview of the latest papers in your field, visit

[computer.org/proceedings/](http://computer.org/proceedings/)