# AN EMPIRICAL ANALYSIS OF WEB SITE STICKINESS

**Mario Christ**[•]

Institute of Information Systems
Humboldt University of Berlin
10178 Berlin, Germany
Phone: +49 (30) 2093-5662
Fax: +49 (30) 2093-5741
christ@wiwi.hu-berlin.de

**Ramayya Krishnan**

Heinz School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213
rk2x@andrew.cmu.edu

**Daniel Nagin**

Heinz School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213
dan03@andrew.cmu.edu

**Oliver Günther**

Institute of Information Systems
Humboldt University of Berlin
10178 Berlin, Germany
guenther@wiwi.hu-berlin.de

## ABSTRACT

*Even though we have seen an exponential growth in the number of Web sites and the number of users, little is known about Web usage at the level of the individual. This paper aims to overcome this lack of knowledge on individual usage patterns. Based on previous findings on saturation of Web usage, we use data from 1995-1998 on residential Web usage conducted as part of the HomeNet project to examine if groups of Web users differ in loyalty to Web sites. We also measure the stickiness of the most popular Web sites in the HomeNet sample. The results help us to understand how one should think of Internet usage and have important implications for Internet marketing and strategy.*

---

[•] Also with Heinz School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, U.S.A.

## 1. INTRODUCTION

While it is well known that there is an exponential growth of the Web if measured in number of domains or number of users, little is known about Web usage at the level of the individual. To overcome this lack of knowledge on individual Web usage, [1] addressed the issue of intensity of individual WWW usage and how it evolves over time. In view of the exponential growth in Web sites available (see figure 1), it was reasonable to expect that this increase in number of Web sites has increased visiting opportunities that in turn might have increased web usage by individuals.

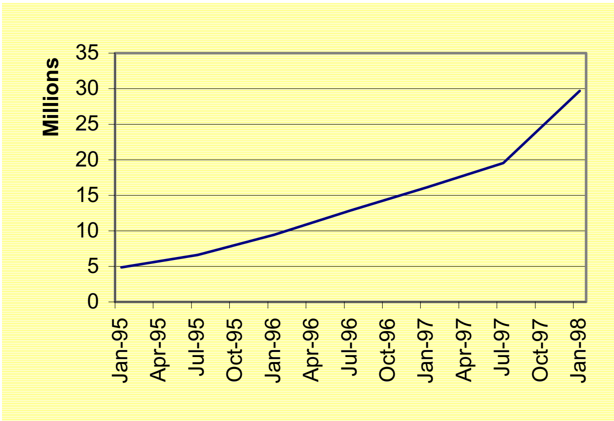

Figure 1: Number of hosts advertised in the DNS (Source: Internet Domain Survey, July 1998)

To test this hypothesis, [1] analyzed the number of distinctive Web sites accessed per week as a measure of the user's interest in the World Wide Web.[1] Groups of individual users with different levels of usage were identified – each with a distinctive trajectory of the development of their Web usage over time. According to this research, Web users in the HomeNet sample [2] can be clustered into four groups with distinct trajectories of use. Figure 2 and figure 3 display the actual and predicted trajectories of the four identified groups of users, which are labeled "very heavy users", "heavy users", "moderate users", and "non-users". The dashed lines represent actual behavior and the solid lines represent predicted behavior.[2]

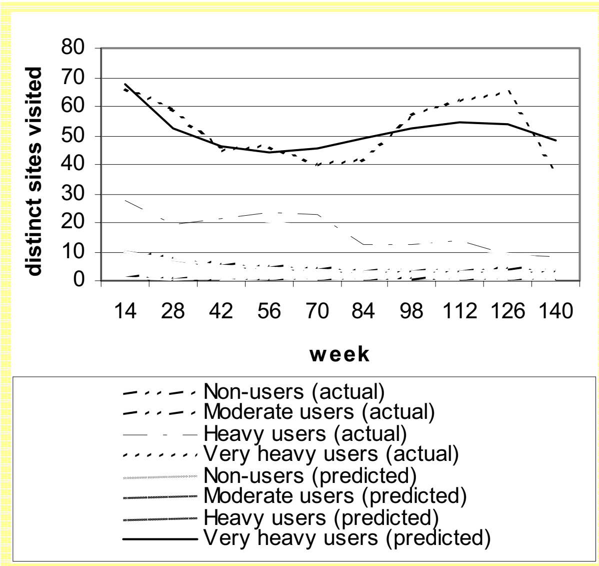

Figure 2: Residential use of the Web measured in number of distinctive Web sites accessed over time

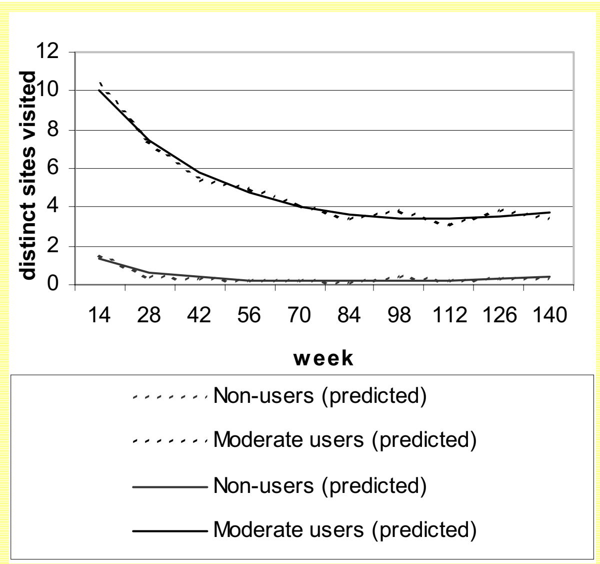

Figure 3: Number of distinctive Web sites visited over time; non-users and moderate users only

[1] Intensity of web usage is measured by number of distinctive web site accessed per months. Consider a Web usage pattern where a users visits three sites {A, A, B} in a given week. This user accesses three sites but only two are distinctive.

[2] Predicted behavior is calculated as the expected value of the random variable depicting each group's behavior. Expected values are computed based on model coefficient estimates. Actual behavior is computed as the mean behavior of all persons assigned to the various groups identified in estimation. See [1,3] for a detailed description of the method.

Note that in contrast to the exponential growth in Web sites available as shown in figure 1, there is actually a decline in residential Web usage intensity if measured in terms of the number of distinctive Web sites accessed per week. Moreover, all the groups follow a downward path and achieve - after a period of 'surfing around' and 'exploring' the Web - saturation in their extent of Web usage as measured by the average number of distinctive Web sites visited per week. The increase of available Web sites did not lead to an increase of WWW usage on the individual level. The saturation levels are 0, 4, 10, and 50 sites per week for the group of 'non-users', 'moderate users', 'heavy users', and 'very heavy users' respectively

If one considers the Web as a marketplace, the number of Web users multiplied by each individual's saturation level of Web usage determines the size of the market. Clearly, the size of the market affects the nature of competition. Therefore, [1] came to the conclusion that the Web is a highly competitive entity whose degree of competition is likely to become even higher when eventually the growth of the Web in numbers of new users accessing the Web slows down. Following this logic, there is the need to further analyze the extent of loyalty of individual users to web sites. Such analysis should identify the demographic characteristics of loyal and disloyal user groups and the identity and characteristics of the web sites that engender the most loyalty. Terms such as "churn" and "stickiness" have been used to describe loyalty on the Web. We introduce precise ways of measuring loyalty on the web and characterize loyalty empirically using HomeNet data.

Table 1: Overview of characteristics of users in the various groups

| | all users | non-users | Moderate users | Heavy users | Very heavy users |
|---|---|---|---|---|---|
| Percentage | 100% | 49.9% | 35.5% | 10.2% | 4.3% |
| | | | | | |
| Adult | 59.3% | 63.2% | 57.5% | 51.5% | 41.7% |
| Female | 55.1% | 61.6% | 54.2% | 39.4% | 33.3% |
| Minority | 29.8% | 39.7% | 19.8% | 15.2% | 16.7% |
| | | | | | |
| Role in family: | | | | | |
| Mom | 26.8% | 31.6% | 25.0% | 18.2% | 0.0% |
| Dad | 19.5% | 16.1% | 25.0% | 21.2% | 8.3% |
| Daughter | 23.6% | 23.0% | 25.8% | 18.2% | 25.0% |
| Son | 18.9% | 16.7% | 17.5% | 33.3% | 33.4% |
| Other | 11.2% | 12.6% | 6.7% | 9.1% | 33.3% |
| Avg. age | 30.7 | 31.6 | 30.6 | 28.3 | 25.2 |

[1] identified the demographic factors that distinguish different user groups and the estimated proportion of the population belonging to each of these groups, as shown in Table 1. Clearly, the results in Table 1 speak to the digital divide debate [6]. We wanted to answer the question if these groups also differ in loyalty to Web sites. For example, it is reasonable to expect that users with lower Web utilization rates are more loyal to Web sites than heavy users, because heavy users may be visiting a large number of distinct sites infrequently and moderate users may be visiting relatively few sites with high frequency In this regard, we continue the work of [1] by measuring churn of Web users and stickiness of Web sites in the same data sample. The results have important implications for Web site operators from a business perspective. The paper is organized as follows: Section 2 introduces precise quantitative ways of measuring the loyalty of Web users to Web sites over time. It analyzes whether a given level of WWW usage intensity is directed to one site or many sites. It thereby answers the question if users converge over time to a set of 'favorite' Web sites. Section 3 paves the way for measuring popularity of Web sites, which influences the probability that a given Web site will be in a users set of favorite sites. Section 4 actually measures 'stickiness' of the most popular Web sites, which determines the ability of these sites to actually remain in this set of favorite domains over time. Section 5 brings together the results in a 'popularity-stickiness map'. Finally, sections 6 deals with open research issues, summarizes, and discusses the implications for electronic commerce.

## 2. MEASURING CHURN IN WEB SITES VISITED

The results from [1] tell us that different groups of people reach different levels of saturation in terms of how many distinctive Web sites they visit over time. These saturation levels differ across groups. However, it is important to keep in mind that individuals do not necessarily visit the same distinct

Web sites from week to week. Indeed, there might be considerable churn in the specific Web sites visited over time. Therefore, we are interested in the loyalty of users in the different groups to the Web sites they visit. By measuring the degree of loyalty of Web users to Web sites over time and analyzing whether a given level of Web usage intensity is directed to one site or many sites, one could answer the related question about the demographics of the loyal users on the Web. In case of low loyalty or high churn, there would be limited overlap over time in the identities in the specific Web sites visited. When it comes to measuring churn over time, two extreme cases are possible:

- No churn

  When people reach saturation, they visit the same set of 4, 8, or 50 specific Web sites (for moderate, heavy, and very heavy users respectively) over each time period (e.g., every week or month).

- 100% Churn

  When people reach saturation, they visit 4, 8, or 50 Web sites (depending on group membership) per week but do not visit the same sites from one week to another.

In a 'no churn' scenario, people would find their right set of Web sites they stick to after a period of 'exploring' the Web. It would be very easy to detect the successful Web sites that 'survived' the exploration period of a given user by simply identifying the Web sites that remain in the user's set in the last period of observation.

However, it is reasonable to presume that the truth lies somewhere between the two extremes. Therefore, it is important to find the right measurement of churn over time, which involves a variety of issues. The fact that there may be sites to which users are loyal to should increase the measurement of overall loyalty of the given user. On the other hand, the fact that there may be sites to which users not loyal should decrease a measurement of overall loyalty. There are already some existing approaches of measuring churn in the WWW related literature. For example, [4] use simply the percentage of revisits to Web pages over time. We propose another approach, which we exemplify in the tables on this page.

In the example in table 2, a given user visits 4 distinct Web sites {A,B,C,D} in the first time period $t$=1, 3 distinct Web sites {A,B,D} in the second time period $t$=2, and three distinct Web sites {A,F,G} in the third time period $t$=3. Apparently, this user is loyal to Web site A, which he visited in all of the three periods of time. On the other hand, Web site C was only visited once in $t$=1 but not in $t$=2 and $t$=3, indicating disloyalty to this Web site. Table 3 depicts the case of complete loyalty. Every Web site is revisited in the period of time followed by period in which the site first appeared. On the other hand, table 4 shows the case of total disloyalty, where there are no revisits at all.

Table 2: Fictitious Web usage

| t=1 | t=2 | t=3 |
|---|---|---|
| A | A | A |
| B | B | F |
| C | D | G |
| D | | |

Table 3: Total loyalty

| t=1 | t=2 | t=3 |
|---|---|---|
| A | A | A |
| B | B | B |
| C | C | C |

Table 4: Total disloyalty

| t=1 | t=2 | t=3 |
|---|---|---|
| A | D | G |
| B | E | H |
| C | F | I |

We apply the following method of measuring churn for each given user:

$$c_{i,t} = \frac{s_{i,t \ldots t+T}}{\sum_{time=t}^{t+T} s_{i,time}},$$

where $c_{i,t}$ is the churn of a given user $i$ in a given period of time $t$, the numerator is the number of Web sites visited by the same user in a time window that starts at $t$ and ends at $t+T$, and the denominator is the sum of numbers of visits to distinct Web sites in the periods of observation $t$, $t$+1, …, $t+T$, of which the time window $t...t+T$ is comprised of. $T$ is the fixed length of this time window. For example, $T$ in table 2 equals 3, the numerator is 6 (distinct Web sites), and the denominator is 4+3+3=10.

Applying this measure to the examples in tables 2-4 leads to the following results: In table 4 – the case of total disloyalty – the churn for the given user is:

$$c_{1,1} = \frac{9}{3+3+3} = 1 .$$

In table 3 – the case of total loyalty – the churn for the given user 1 is:

$$c_{1,1} = \frac{3}{3+3+3} = \frac{1}{3} .$$

The churn for the given user in table 2 is:

$$c_{1,1} = \frac{6}{4+3+3} = 0.6 .$$

Note that the upper bound of $c$ is 1 and the lower bound of $c$ is 1 divided by the length of the time window, which is 1/3 in our example. In other words: $(1/T) \leq c_{i,t} \leq 1$, where $c_t$=1 for the least loyal user and $c_t$=$1/T$ for the most loyal user.

We define $cn_{i,t}$ as the normalized measurement of churn with $0 \leq cn_{i,t} \leq 1$:

$$cn_{i,t} = 1 - \left( \frac{1}{1 - \mathrm{T}^{-1}} \times \left( 1 - \mathrm{c}_{\mathrm{i,t}} \right) \right)$$

In example A, B, and C, given a time window of $T$=3, $cn_{i,t}$ equals 0.4, 0, and 1 respectively.

The time windows with $T$=3 or any other length can be used as a sliding time window to capture the development of churn over time. We compartmentalized data in the HomeNet sample by using 1-month periods. We also used a sliding time window with an arbitrary chosen length of $T$=3 (months) to analyze churn in the HomeNet data over time. Figure 4 depicts the average normalized churn of given groups of users in the HomeNet sample over a period of 14 months.

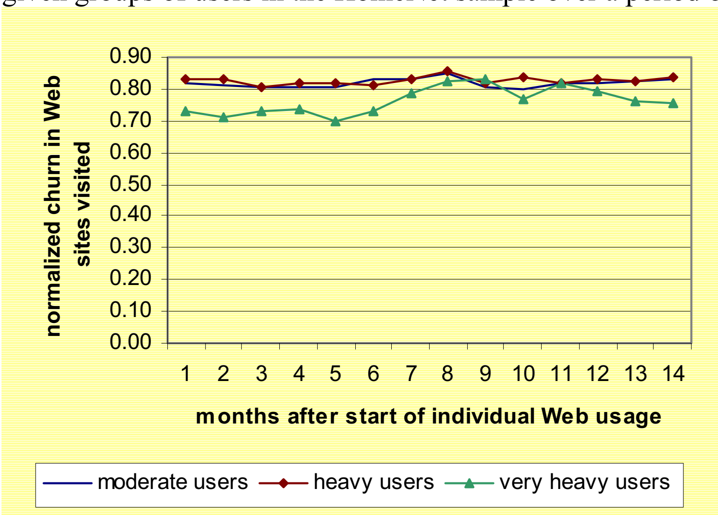


Figure 4: Normalized churn in Web sites visited

While we find considerable churn across all groups in the sample, the surprising result is that over time, churn stays almost constant around 0.8 and is independent of group membership.

The results imply that for moderate users that visit about 4 sites a week that users are loyal to about 1 in 4 sites. Heavy users are loyal to about 2 sites out of 7-8 sites they visit each week and very heavy users are loyal to about 8 sites out of the 50 they visit each week. In particular, group membership seems to have only a negligible impact on churn. We expected that moderate users would show a lower degree of churn. However, our findings contradict this hypothesis. The group of very heavy users seems to be even a little more loyal than the other two groups. In general, high churn or little loyalty, as shown in Figure 4, indicates that the process of exploring the Web continues even in late periods of observation. Users do not converge to a favorite set of Web sites. Even though there may be a set of favorite Web sites for given users, their percentage in the set of visited Web sites must be rather small. The use of these Web sites does not lead to a decrease in the number of new Web sites visited over time, which results in almost constant churn over time. Given the fact that the Web itself is not static, as shown in Figure 1, this is a reasonable result.

Fortunately for Web site operators, a churn of about 0.8 means a loyalty of 0.2. As long as loyalty is not zero, users do not visit Web sites at random. Moreover, even if there are apparently only minor differences between the different groups of people percentagewise, there are substantial differences in the actual number of Web sites people are loyal to. For example, according to [1], the group of heavy users visits considerably more distinct site per week than the group of moderate users. Both groups do have – percentagewise – about the same degree of disloyalty of 0.8. Given a loyalty of approximately 0.2 for both groups, the number of Web sites users are loyal to is much larger for the group of very heavy users than for the group of moderate users. In general, there are - to a different extent depending on group membership - Web sites that are less affected by churn and stay in the set of Web sites more permanently. Ways to identify these successful sites will be the discussed in the following sections.

## 3. POPULARITY OF WEB SITES

One implication of the results of the previous section, which indicated considerable churn across subgroups of residential users, is the opportunity to attract/acquire new customers (of course, retaining these customers is the difficult problem). We analyze the ability of web sites in the HomeNet sample to acquire customers and the ability to retain them. Since users have a fixed number of sites that they are willing to visit in any give time period, we model both the popularity of a web site (its acquisition ability) and its stickiness (its ability to retain customers).

As a measure of popularity of Web sites we simply order the data by the number of users who accessed a given Web site within the whole period of observation. In this regard, when we speak of popularity we mean the short-term popularity that derives from attracting users at least once without saying anything about the ability of the site to make users visit the same site again. Table 5 shows the most popular Web sites in the HomeNet sample in terms of the proportion of users who actually accessed the site at least once.

Table 2: The most popular Web sites in the HomeNet sample

| domain | Overall popularity |
|---|---|
| HOMENET.ANDREW.CMU.EDU | 0.96 |
| HOME.NETSCAPE.COM | 0.93 |
| YAHOO.COM | 0.79 |
| CS.CMU.EDU | 0.59 |
| EXCITE.COM | 0.54 |
| INFO.CERN.CH | 0.49 |
| PATHFINDER.COM | 0.47 |
| INFOSEEK.COM | 0.45 |
| PITT.EDU | 0.37 |
| LYCOS.COM | 0.37 |
| W3.ORG | 0.34 |
| MIT.EDU | 0.29 |
| PITTSBURGH.NET | 0.28 |

Note that domains of banner ad sites and web hosts have been removed from table 5 because they skewed the results. Furthermore, note that the data show particular characteristics of the HomeNet sample. Users in this sample are people from the Pittsburgh area, which explains the high popularity of some local Web sites (e.g., www.pittsburgh.net). This also supports the hypothesis that a large share of Web activity is ‘local’, even if the Internet itself is ‘global’.

## 4. STICKINESS OF WEB SITES

The popularity of a Web site does not say anything about the actual ‘stickiness’ of the same site, its ability to attract users again. Popularity of a Web site could be the result of many users visiting this Web site only once without ever coming back.

Therefore, more subtle measures of a Web site's success are needed. In this regards, we calculate the 'stickiness' of a given Web site as:

$$s_{i,url} = \frac{\# a_{i,domain}}{\# p_{i,domain}},$$

where $s_{i,domain}$ is the stickiness of a Web site *domain* for a given user *i*, $\#p_{i,domain}$ is the number of months left in the sample period after user *i* accessed site *domain* first, and $\#a_{i,domain}$ is the number of months after the user accessed the Web site first in which the users actually accessed the given Web site.

For example, table 6 depicts the stickiness data of www.yahoo.com for a subset of users. Zeros denote months in which a given user did not access this Web site. 'Ones' denote months in which the user actually accessed the site. Missing data is denoted by dots. In this example, user *i*=52 accessed the Web site domain='www.yahoo.com' first in period 2. After that there remain $\#p_{i,domain}$=12 periods of observation (t3-t14). User 52 accessed the given site in $\#a_{i,domain}$=4 of the remaining 12 periods of time (namely in t4, t7, t8, and t13). Stickiness is calculated as $s_{i,domain}$=4/12=0.33.

Table 6: Stickiness table for Yahoo!

| user | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 | T11 | T12 | T13 | T14 | Stickiness |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ... | | | | | | | | | | | | | | | |
| 50 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | **0.54** |
| 51 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | **0.92** |
| 52 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | **0.33** |
| 53 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | **0.46** |
| 54 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | . | . | **0** |
| 55 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | . | . | **0** |
| 56 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | . | **0.33** |
| ... | | | | | | | | | | | | | | | |
| **avg** | **0.58** | **0.39** | **0.23** | **0.27** | **0.31** | **0.30** | **0.31** | **0.25** | **0.23** | **0.26** | **0.30** | **0.25** | **0.25** | **0.23** | **0.46** |

The share of the users who actually accessed this Web site in a given month is shown in the line at the bottom of table 6 (e.g., there is an average popularity of 0.23 in period t3). Furthermore, the number in the lower right corner of table 6 shows the average stickiness of 'www.yahoo.com' across all users in the sample (0.46). Users who did not visit the Web site at all were dropped from the calculation of this average stickiness (missing data).

Table 7: Stickiness and popularity of Web sites

| domain | t1 | t2 | t3 | t4 | t5 | t6 | t7 | t8 | t9 | t10 | t11 | t12 | t13 | t14 | stickiness | Overall popularity | Average popularity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HOMENET.ANDREW.CMU.EDU | 0.94 | 0.69 | 0.66 | 0.68 | 0.59 | 0.65 | 0.64 | 0.57 | 0.54 | 0.50 | 0.61 | 0.63 | 0.61 | 0.69 | 0.71 | 0.96 | 0.64 |
| HOME.NETSCAPE.COM | 0.90 | 0.74 | 0.63 | 0.71 | 0.65 | 0.67 | 0.60 | 0.56 | 0.57 | 0.46 | 0.61 | 0.58 | 0.53 | 0.56 | 0.73 | 0.93 | 0.63 |
| YAHOO.COM | 0.58 | 0.39 | 0.23 | 0.27 | 0.31 | 0.30 | 0.32 | 0.25 | 0.23 | 0.26 | 0.30 | 0.25 | 0.25 | 0.23 | 0.46 | 0.79 | 0.30 |
| CS.CMU.EDU | 0.38 | 0.30 | 0.22 | 0.24 | 0.10 | 0.15 | 0.13 | 0.11 | 0.08 | 0.05 | 0.14 | 0.04 | 0.08 | 0.12 | 0.39 | 0.59 | 0.15 |
| EXCITE.COM | 0.27 | 0.25 | 0.16 | 0.17 | 0.11 | 0.08 | 0.12 | 0.07 | 0.10 | 0.05 | 0.08 | 0.08 | 0.05 | 0.04 | 0.33 | 0.54 | 0.12 |
| INFO.CERN.CH | 0.28 | 0.13 | 0.11 | 0.11 | 0.04 | 0.05 | 0.07 | 0.08 | 0.06 | 0.07 | 0.06 | 0.03 | 0.03 | 0.08 | 0.32 | 0.49 | 0.09 |
| PATHFINDER.COM | 0.25 | 0.20 | 0.14 | 0.13 | 0.04 | 0.07 | 0.08 | 0.07 | 0.03 | 0.05 | 0.06 | 0.04 | 0.03 | 0.04 | 0.30 | 0.47 | 0.09 |
| INFOSEEK.COM | 0.35 | 0.08 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.03 | 0.00 | 0.05 | 0.10 | 0.19 | 0.45 | 0.05 |
| PITT.EDU | 0.15 | 0.09 | 0.07 | 0.11 | 0.12 | 0.08 | 0.14 | 0.14 | 0.09 | 0.10 | 0.11 | 0.04 | 0.14 | 0.12 | 0.42 | 0.37 | 0.11 |
| LYCOS.COM | 0.16 | 0.04 | 0.03 | 0.04 | 0.02 | 0.02 | 0.02 | 0.06 | 0.05 | 0.00 | 0.10 | 0.03 | 0.12 | 0.12 | 0.25 | 0.37 | 0.06 |
| W3.ORG | 0.20 | 0.09 | 0.08 | 0.07 | 0.04 | 0.03 | 0.03 | 0.03 | 0.02 | 0.04 | 0.03 | 0.01 | 0.03 | 0.00 | 0.28 | 0.34 | 0.05 |
| MIT.EDU | 0.15 | 0.09 | 0.07 | 0.05 | 0.05 | 0.04 | 0.01 | 0.07 | 0.00 | 0.01 | 0.03 | 0.01 | 0.02 | 0.02 | 0.29 | 0.29 | 0.04 |
| PITTSBURGH.NET | 0.14 | 0.08 | 0.10 | 0.08 | 0.03 | 0.02 | 0.04 | 0.02 | 0.01 | 0.01 | 0.10 | 0.01 | 0.05 | 0.04 | 0.31 | 0.28 | 0.05 |

We created tables similar to table 6 for the more popular Web sites given in table 5. We focused on the popular sites because the smaller the number of users a Web site attracts, the sparser the data in such a table becomes. Table 7 depicts the summary lines of all tables similar to table 6 for the other popular Web sites in the HomeNet sample. Note that the smaller the stickiness, the higher the difference between overall popularity as reported in Table 5 and average popularity over time.

One might think that popularity and stickiness measure the same latent construct. This is not necessarily the case, especially because we measure the overall stickiness of a Web site only as an average of individual stickiness of the users who actually access the site. Even though the two measures are related, a popular site is not necessarily sticky and a sticky site is not necessarily popular. However, notice that – on average – there is a relation between popularity and stickiness if you observe a fixed number of users over time. Given a fixed set of users, even most popular Web sites will loose their popularity over time if they do not attract users again. In this regard, note also that local Web sites such as pitt.edu sustain a constant popularity over time.

Figure 5 displays the development of popularity of some Web sites over time. All Web sites in Figure 5 suffer a loss of popularity after the start of the project. However, some Web sites such as Yahoo! manage to sustain a quite constant popularity over time, which indicates that they have properties that make users come back to the site. In this regard, note that www.yahoo.com has a very high stickiness in table 7 and a high overall popularity in table 5. In a world with a fixed set of users (139 users in the HomeNet project), high stickiness leads to constantly high levels of popularity. Another explanation for this is the word of mouth: if people have reasons to stick to a Web site (high

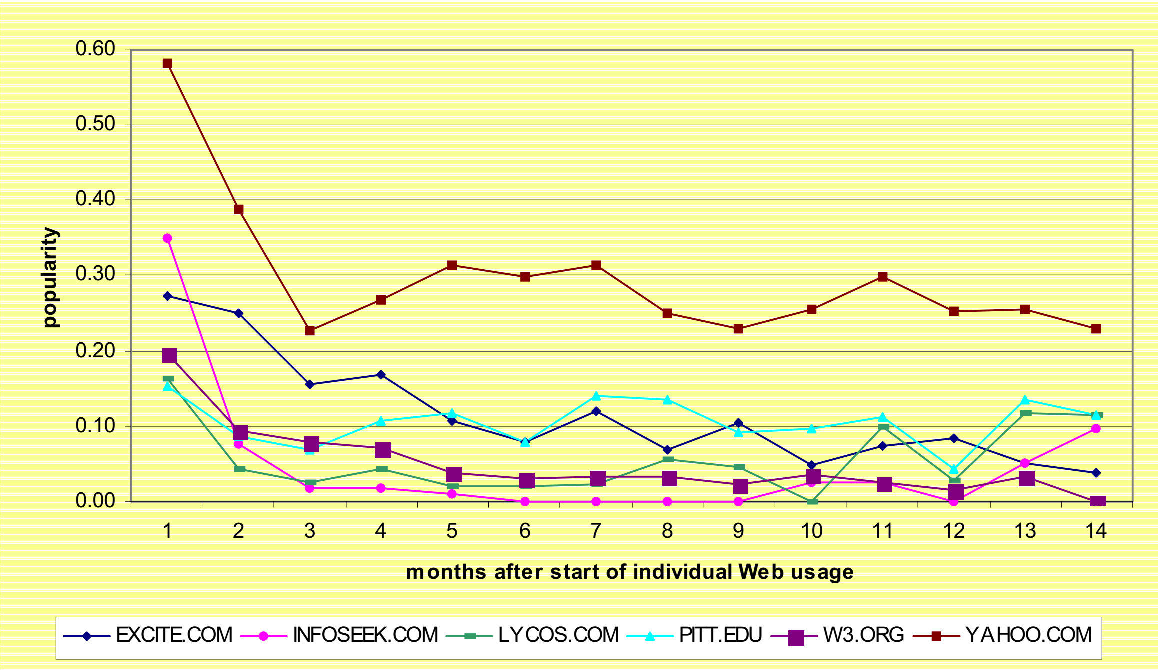


Figure 5: Popularity of Web sites in the HomeNet sample over time

stickiness), they might tell their friends to visit this site, thereby increasing the number of users at this site (high popularity).

Another search engine, www.infoseek.com develops differently over time. This Web site has a high popularity in the beginning, which is also reflected by a high overall popularity in table 5, but fails to attract users again, which is also reflected by a low stickiness in table 7. The popularity of this site decreases dramatically as shown in Figure 5.

Keep in mind that this relation (derived from a closed sample of HomeNet users) does not necessarily hold in the real world where there is no fixed set of users but a radical increase of the number of WWW users. Even if individual users do not come back to a site, there are often enough new users to keep the site's popularity on a high level. However, the relation between stickiness and popularity gives us insights into the success or failure of some Web sites with low stickiness when the growth of the Internet in terms of number of new users slows down or even stops. This was also anticipated in [1].

## 5. THE POPULARITY-STICKINESS MAP

We display the different values of popularity and stickiness on a popularity-stickiness map in Figure 6. Observe that this popularity-stickiness map displays only the most popular Web sites. Stickiness is depicted on the horizontal axis; popularity is depicted on the vertical axis in a log scale. Because all the Web sites belong to the group of Web sites with the highest popularity, they are located in the upper half of Figure 6. If we chose to display not only the popular but also all the other Web sites in the sample, the picture would show a non-random distribution of Web sites across all possible values for popularity and stickiness with a cluster of Web sites in the lower left corner.

We divide the popularity-stickiness map into four areas with high or low values for stickiness and popularity respectively. The site that dominates all the other Web sites in Figure 6 is 'www.yahoo.com'. Because both, popularity and stickiness are high, this Web site is located in an area of the upper right corner of the map and belongs therefore to the group of ''Type-1 Web sites'.

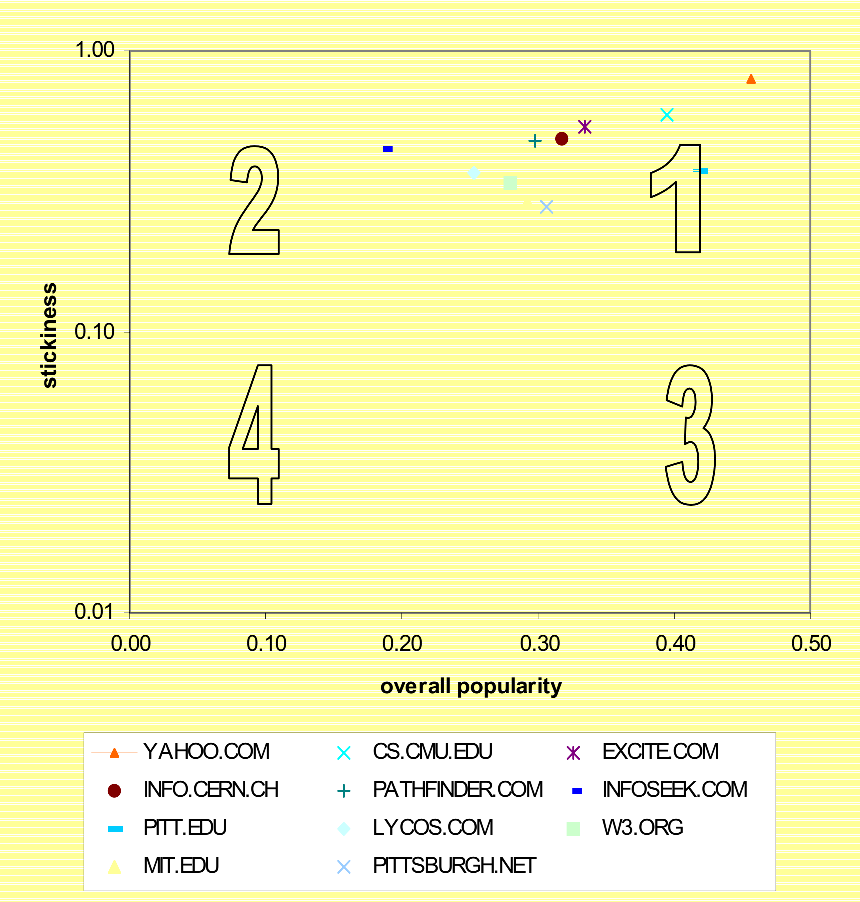


Figure 6: popularity-stickiness map of the more popular Web pages in the HomeNet sample

Direct comparison of 'www.yahoo.com' with other popular web sites reveals that all other sites have both, lower (but still high) popularity and lower stickiness. The Web site that comes closest to a position in the upper left corner is 'infoseek.com'. We call Web sites in this area of the map 'Type-2 Web sites'. They achieve to attract a lot of users but fail to make the same users come back to the site. Note, that the less sticky a site is, the more likely it is to become a 'Type-3 Web site' or even a 'Type 4 Web site' in the future when the growth of the Internet slows down and the Web site operators do not act to increase the Web site's stickiness.

Some sites that have a much lower popularity can still be as sticky as 'yahoo.com', although for a limited number of users only. We call this group of Web sites ''Type-3 Web sites' because they address only a small subset of the population but achieve high stickiness among its users. In general, candidates for this category are Web sites that focus on a subgroup or niche of WWW users and address the specific needs of these users, thereby providing high utility and achieving high stickiness.

'Infoseek.com' dominates sites with equal small values for stickiness and even lesser values for popularity, which come therefore closer than even less desirable position in the lower left corner, an area in which ' Type-4 Web sites' would be positioned. However, since Figure 6 displays only popular Web sites, none of the sites in this figure actually belongs to this group of Web sites. Candidates for this category are less popular sites with a low stickiness, such as personal homepages.

The most desirable position for Web sites in Figure 6 is a position in the area of 'Type-1 Web sites'. However, given the results [1] and section 3 about saturation in Web usage and a high degree of disloyalty to Web sites, it is impossible that every Web site reaches this position. There will be rather a fierce competition of Web sites for the desirable positions in Figure 6. Given a degree of churn greater than zero, limited capacity of users turns competition into a zero sum game for competitors in the World Wide Web. In other words, whenever some Web sites improve their position in the popularity-stickiness map, others deteriorate their position.

## 6. CONCLUSION

In this paper we developed quantitative measures of loyalty of users and stickiness of Web sites. With respect to loyalty, we find considerable churn in Web sites visited across subgroups. Moreover, we find that the degree of churn is a constant over time across all groups of users. This is a surprising and interesting result and needs to be replicated in larger samples such as the Media Metrix panel usage data. It is important to identify Web sites that are both able to acquire and retain customers (popularity and stickiness) and to identify characteristics that contribute to the features. Given the fact that users reach saturation and show a constant high degree of churn over time, it seems relatively easy to get into a user's set of Web sites. However, high churn also means that it is rather difficult to stay there. Therefore, we analyzed which Web sites have the ability to get into this set of domains (popularity) and the ability to stay there (stickiness). We displayed Web sites in a 'popularity-stickiness map', which we divide the map into four areas.

The rejection of the hypothesis of increasing WWW usage intensity in combination with the insights from our churn analysis is a first indicator that competition among WWW companies for WWW market share is likely to become more intense when the growth in terms of numbers of people accessing the WWW slows down.

We encourage future research that takes into account human context when looking at churn in terms of both why users choose particular sites and what constitutes disloyalty. For example, infrequent use of the same site does not necessarily constitute disloyalty. In this regard, our measure of churn is a simple one in the sense that it does not incorporate the type of Web site. For example, some types of Web sites, such as vacation sites, are by nature visited infrequently. However, users might still be loyal to these sites.

Relatedly, we believe that users may be visiting Websites that are functionally related, e.g. vacation sites. If in fact this is the case, we need to develop methods for modeling churn, which take into account the possibility that Web sites may be complements and substitutes to one other.

Finally, the patterns of WWW usage we found for usage data from 1995-1998 may be different for more recent data. Note that the HomeNet project focuses on individuals at home. A significant part of the population accesses the Internet at work. Therefore, further research is necessary in order to confirm those patterns for all groups of users in the WWW and for data from 1998 on.

## ACKNOWLEDGEMENTS

HomeNet is funded by grants from Apple Computer, AT&T, Bell Atlantic, Bellcore, Intel, Carnegie Mellon University's Information Networking Institute, Interval, the Markle Foundation, the NPD Group, the U.S. Postal Service, and US West. Farallon Computing and Netscape Communications contributed software

Development of the trajectory estimation method and software was supported by the National Science Foundation under Grant No. SBR-9513040 to the National Consortium on Violence and also by separate National Science Foundation grants SBR-9511412 and SES-9911370.

The author Mario Christ was supported by the German Research Society, Berlin-Brandenburg, Graduate School in Distributed Information Systems (DFG grant no. GRK~316). This research was also supported by the TransCoop program of the Alexander von Humboldt Foundation, Bonn, Germany.

The work of Ramayya Krishnan was funded in part by NSF grant CISE/IIS/KDI 9873005.

## REFERENCES

[1] Christ, M., Krishnan, R., Nagin, D., Kraut, R., Günther, O. "Trajectories of individual WWW usage: implications for electronic commerce". Proc. 34th Hawaii International Conference on System Science (HICSS-34). 2001.

[2] Kraut, R. E., Scherlis, W, Mukhopadhyay, T., Manning, J., Kiesler, S. "HomeNet: A field trial of residential Internet services". Communications of the ACM. 39 (12),1996, 55-63.

[3] Nagin, D. "Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach". Psychological Methods. Vol. 4, No. 2, 1999, 139-157.

[4] Tauscher, L., Greenberg, S. „How people revisit web pages: empirical findings and implications for the design of history systems". *International Journal of Human Computer Studies, Special Issue on World Wide Web Usability, 47*, 1997, 97-138

[5] The HomeNet project. [http://homenet.andrew.cmu.edu/progress].

[6] Hoffman, D. L., Novak, T. P. "Bridging the racial divide on the Internet". Science. April 17. 1998.