Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2003 Proceedings

European Conference on Information Systems (ECIS)

2003

A Session Based Empirical Investigation of Web Usage

Mario Christ
Humboldt-University, Berlin, christ@wiwi.hu-berlin.de

Steffan Baron
Humboldt-University, Berlin, sbaron@wiwi.hu-berlin.de

Ramayya Krishnan
Carnegie Mellon University, rk2x@cmu.edu

Daniel Nagin
Carnegie Mellon University, dn03@andrew.cmu.edu

Oliver Günther
Humboldt-Universität zu Berlin, guenther@wiwi.hu-berlin.de

A Session-based Empirical Investigation of Web Usage

Mario Christ
Graduate School of Distributed Information Systems Berlin-Brandenburg
Humboldt University of Berlin, 10178 Berlin, Germany
+49 (30) 2093-5662
christ@wiwi.hu-berlin.de

Steffan Baron
Institute for Information Systems
Humboldt University of Berlin, 10178 Berlin, Germany
+49 (30) 2093-5740
sbaron@wiwi.hu-berlin.de

Ramayya Krishnan
Heinz School of Public Policy and Management
Carnegie Mellon University, Pittsburgh, USA
rk2x@cmu.edu

Daniel Nagin
Heinz School of Public Policy and Management
Carnegie Mellon University, Pittsburgh, USA
dn03@andrew.cmu.edu

Oliver Günther
Institute for Information Systems
Humboldt University of Berlin, 10178 Berlin, Germany
+49 (30) 2093-5741
guenther@wiwi.hu-berlin.de

Abstract
This paper reports the results of a study of Web usage of 139 users over a 8 month period of time. It uses a longitudinal Web log analysis of the URLs accessed during 33916 user-days of Web usage. It aims to detect changes in Web usage associated with increased experience of using the Web. Specifically, it answers the question whether or not users shift from undirected browsing in the Web to directed access of Web sites as they gain expertise in using the Web. We used a session-based approach to measure individual Web usage. The results of our study have several important implications both for Business to Consumer electronic commerce and for public policy as it pertains to the digital divide.

1 Introduction

Currently, over 147 million hosts are connected to the Internet [IDS02] and 544 million people [NUA02] have access to billions of pages of content. Therefore, the Web is clearly one of the most important communication innovations in history. Although built for academia, the Web turned into one of the most important tools of communication for other sectors. Therefore, the social and economical implications of the Net are substantial.

A lot of research has been conducted on the relationship between a user's use of information technology and his experience. Specifically, the level of expertise of a given user is supposed to affect usage of electronic information systems [39,40]. However, little is known about the extent to which individuals utilize the Web and gain expertise over time. This article reports the results of an analysis of eight months of longitudinal data on residential Web usage. This data was assembled as part of the HomeNet project at Carnegie Mellon University.

Using data from the HomeNet project [Kraut96], [Christ01] found that as individuals gain more Web browsing experience, the number of distinct Web sites the same individuals visit per month decreases to a saturation level that depends on whether the individual belongs to the group of low-rate, moderate, heavy, or very heavy users. In this regard, [Christ01] focused like [Cockburn01] on the changes in the user's vocabulary of Web sites over time. Fig. 1 depicts the results of that analysis. All users started using the Web in month 1. Therefore, we consider the longitudinal development in Fig. 1 'learning curves' after a natural starting point of having first access to the Web.

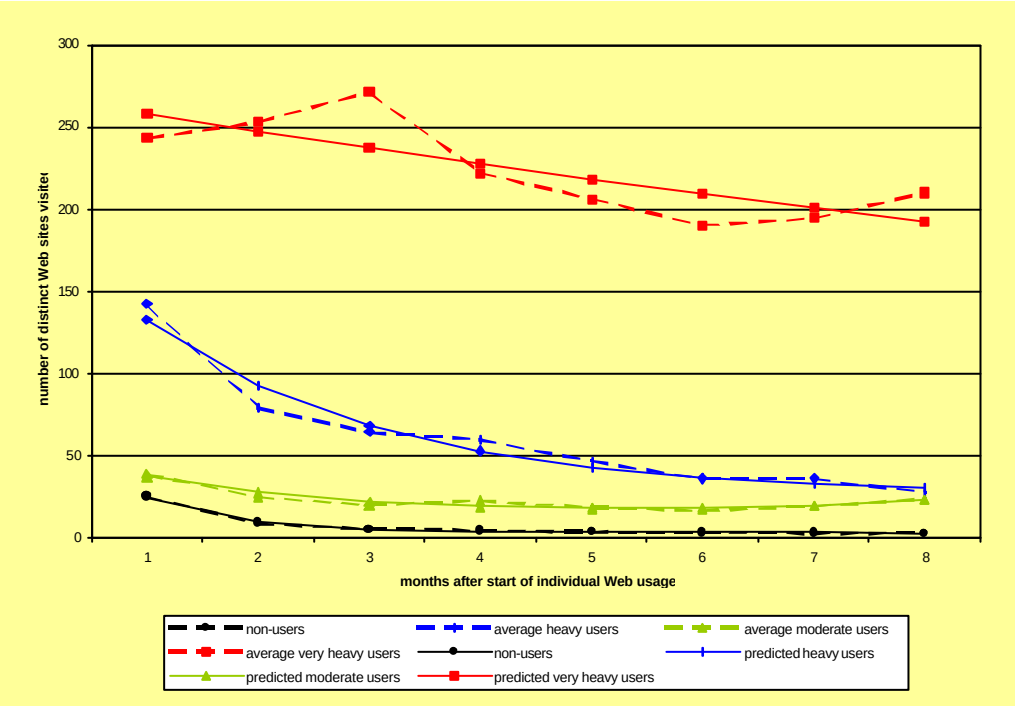

Fig. 1: Trajectories of Web usage [Christ00]

However, the measure of distinct Web sites visited per month is inaccurate in the sense that it does not take into account Web usage behavior within distinct Web sessions. Also, a decrease

in the number of distinct sites visited per month may be due to a variety of reasons. For example, users may develop loyalty to Web sites and converge to a favorite set of sites. On the other hand, users may spend less time in the Internet. Therefore, we wanted to measure time spent in the Web more accurately using a session-based approach. Thus, we followed the lead of [Cooley00] by dividing the individual click stream into Web sessions. We computed five key measures of Web usage within Web sessions for each of the subgroups of users identified in [Christ01].

[Christ01] also identified the demographic factors that distinguish different user groups and the estimated proportion of the population belonging to each of these groups. The statistically significant determinants are age, gender, and race. In this paper, the higher accuracy of measuring Web usage using a session-based approach helps to identify significantly more determinants of Web usage. We identify the determinants by applying formal regression models on the HomeNet data.

Our study advances research on Web usage in three important aspects. Many prior studies (e.g., [Tauscher97], [McKenzie01], and [Catledge95]) have relied on highly non-representative samples (e.g, individuals who worked or studied in computer science departments). Our study relies on a sample of households that is more closely representative of the general population. Second, much of the current work on Web usage is 'cross-sectional'. It compares groups of users at the same point in time. We follow the lead of [Christ01] by adopting a longitudinal-developmental approach to examine the changes in behavior of individual Web users. In this regard, we monitor the same individuals over a (comparatively long) period of time. In our judgment an extended period of observation is required to calibrate credibly patterns of change in Web use. Third, we address specifically the issue of Web sessions of individual users.

Conducting such a study on Web usage and the determining demographic characteristics of usage is necessary for marketing departments, especially in the information and communication industry. The goal is to describe how average citizens (as represented in this study) use the Web and to predict the determinants of usage. Many of the findings in this work will be useful to stakeholders in the new Information Age, in particular policy makers. In this regard, the results have important implications for Internet marketing strategy and public policy pertaining to the digital divide.

Section 2 describes our source of data – the HomeNet project at Carnegie Mellon University – and our five key measures of Web usage in Web sessions. In section 3 we present the results of our analysis: the development of individual Web usage in Web sessions. Also, we report the results of formal regression analyses, which reveal the individual characteristics (e.g., demographics of users) that determine Web usage. Finally, in section 4 we discuss the results and their implications for electronic commerce and public policy.

2 Data Source and Method

2.1 Data Source – The HomeNet project at Carnegie Mellon

Our study is based on individual use records from the HomeNet project [Kraut96]. HomeNet is a field trial at Carnegie Mellon University, which aims at understanding people's home usage of the Internet. Starting in 1995, it provided about 150 people in the Pittsburgh area with hardware and Internet connections and carefully documented their residential usage of on-line services such as electronic mail, computerized bulletin boards, chat groups, and the

World Wide Web. We used computer-generated use records of Web sites visited for our analyses.

The HomeNet data set we used consists of 139 family members in 56 households. These users were performing 1187325 http requests between 11-6-1995 and 4-28-1997. The available fields in the data set are, among others, a unique user ID and the URL accessed. Additionally, demographic data like age, race, and gender were assembled.

2.2 Measurements of Web Usage

We adopted a longitudinal-development approach by monitoring the same individuals at different points in time. Over time, each individual gains experience and expertise (see [Fidel98, Khan98] for the relationship between experience and expertise). In this longitudinal-developmental study, we took repeated measures of Web usage in respect of each individual.

There are several conceptually reasonable alternatives to take repeated measures of Web usage. Broadly, they may be classified into frequency-based measures and time-based measures.

Time-based measures use the time spent by an individual in the Web as an indicator of the utility received from utilizing the Web. The HomeNet project provides data on individual log-on and log-off activities. However, this time-based measure is not confident in the sense that it does not necessarily reflect time spent on the computer. Users may download a Web site but only actively attend to the Web site for a small fraction of time in which the site was displayed (i.e., leave the computer unattended for hours or even days). Therefore, it was necessary to construct a more sophisticated measure that accurately reflects actual time spent actively interacting with a Web site. Thus, we followed the lead of [Cooley00] by compartmentalizing the individual click stream into Web sessions. We applied the following sessionizing criterion: A Web session is considered finished after a period of 30 minutes of user inactivity. Whenever a page request occurs more than 30 minutes after the previous request, it is considered the starting point of a new session. [Cooley99] show that this approach leads to sufficient accuracy in identifying Web sessions. Note that identifying user sessions is a non-trivial problem. See [Cooley00] for a discussion of this issue.

There are also frequency-based measures such as the number of distinct Web sites visited by a given user per time period or Web session and total Web site visits per time period or Web session. The first measure does not count repeated visits to the same Web site whereas the latter measure does count such visits. The count of distinct Web sites visited is an interesting measure because at the level of the individual the diversity of Web sites visited provides an indication of individual-level willingness to search the exponentially expanding set of visiting opportunities in the Web. However, because this measure does not count repeated visits to a given Web site, it also important to examine total Web site visits as an alternative utilization intensity measure. This permits an analysis of whether users who visit a few distinct Web sites in a given period of time are more intensive users of these sites than users who visit a larger number of sites but use each of them less intensively. Further, the number of repeated visits to a site per user is equivalent to the pages downloaded per site. This is an important measure with relevance for advertising online using banner advertisements. These advertisements are served as part of a downloaded Web page and priced per thousand impressions of the advertisement (also known as cost per thousand impressions or CPM) (see http://www.iab.net/measuringsuccess/index.html for more on online advertising). Frequency-

based measures have been discussed widely. For example, [Christ01] have measured the number of distinct Web sites visited by individuals in the HomeNet sample.

We computed a variety of time-based and frequency-based metrics. In order to gain insights into the development of Web usage in Web sessions, we continued the work of [Christ01] and applied five key measures to the data from the HomeNet project. Specifically, we combined frequency-based measures with time-based sessionizing by tracking:

a) The development of the number of Web sessions over time

b) The development of the number of distinct sites *per Web session* over time

c) The development of the number of page views *per Web session* over time

d) The development of the number of page views per Web site *in Web sessions*

The measure d) 'pages viewed per Web site within sessions' deserves special attention because it speaks to the issue of Web loyalty, which is of utmost importance to electronic marketing. While the count of distinct Web site visited provides an indication of individual-level willingness to search the exponentially expanding set of visiting opportunities, the number of pages viewed per distinct Web site permits an analysis of how intense users of these sites make use of these sites. In a case of complete loyalty where an individual's page view capacity is directed to one Web site, this measure would be higher than in a case where an individual page view capacity is directed to many sites.

Apart from using frequence-based measures, we also applied a time-based measure:

e) The development of the duration of individual Web sessions.

We used these measures to gain insights in how information searching behavior changes over time. Information searching behavior is supposed to change with increasing expertise in information technology. In this regard, we were interested in analyzing whether there actually is a trend to directed access of Web sites in contrast to undirected Web browsing. With respect to Web sessions, users might visit many distinct Web sites in the beginning of their Web learning experience when they start exploring the Web. The same users might visit less distinct Web sites per session later on as expertise in Web usage increases. It was reasonable to expect that less experienced users consume the Web in a few large chunks, whereas more experienced users consume Web sites in many small chunks. As expertise increases and users gain knowledge in using the Web, they might start a Web session for the very purpose of visiting one specific Web site (for example, checking a bank account using the Internet). Fig. 2 illustrates the shift from undirected browsing to directed access of Web sites. The left hand of Fig. 2 reports individual Web usage of a fictitious user, which consists of eight page views at three distinct Web sites: amazon.com, yahoo.com, and ft.com. Without applying a session-based approach we do not know whether the same user develops loyalty to Web sites. For example, on the left hand, the user shows limited loyalty to Web sites in the sense that he directs his Web usage to three Web sites, whereas the right hand of Fig. 2 reveals that the same user shows complete loyalty to Web sites within each session. Therefore, in order to identify loyalty in sessions, we use the new measure of average page views per site in sessions. In this respect, users develop loyalty to Web sites if the ratio of pages viewed per Web site within sessions increases over time.

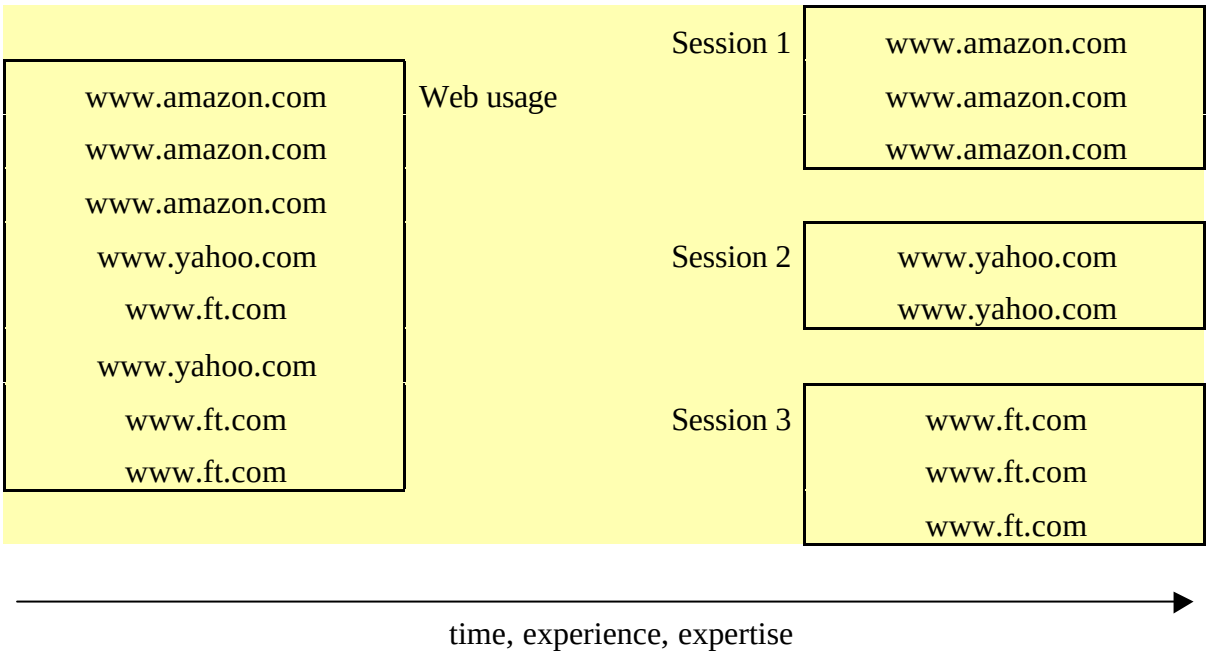

Fig. 2: Hypothetical learning experience over time

It was reasonable to expect that a hypothetical learning experience such as in Fig. 2 has the following impact on how our measures a) to e) develop over time. Specifically, we expected:

a) an increase in the number of Web session

b) a decrease in the number of distinct Web sites visited per session

c) a decrease in the number of Web pages viewed per session

d) an increase in the number of page views per Web site in Web *sessions* (an increase in loyalty to Web sites within sessions)

e) a decrease in the duration of Web sessions

In this paper, we wanted to test if we actually see this in the HomeNet data. Section 3 reports the results.

Fig. 2 also illustrates how aggregating Web usage in a given period of time conceals patterns of Web usage in Web sessions. For example, suppose that the clickstream on the left hand of Fig. 2 is the aggregated clickstream of a given user in a given month across all sessions, such as described in [Christ01]. Further analysis reveals that overall monthly Web usage can be divided into three distinct sessions. In each of these sessions the given user visits one distinct site only. Clearly, a measure such as ‘page views per site’ in a given month conceals loyalty in sessions.

3 Results

3.1 Five key measures of Web usage in Web sessions

Using the sessionizing approach described in the previous section revealed that the 139 users in the HomeNet project use the Web in 6612 distinct Web sessions during the first 8 months of the project after their initial Web usage. The average duration of a session was 26.38

minutes. The average number of distinct sites visited per session was 6.47. Users visited on average 19.93 pages per session. *Tab. 1* reports these descriptive statistics.

Tab. 1: descriptive statistics on Web usage in Web sessions

total number of sessions	6612			
	Average	10th percentile	50th percentile	90th percentile
number of Web sessions	7.87	0	4	21
duration of Web session (minutes)	26.38	6	22	50
number of distinct sites visited per session	6.39	2.15	5.5	11.43
number of pages viewed per session	18.27	5	15	34.23
Ratio page views / site per session	3.03	2.5	2.57	4.33

Before we move on and show how these measures evolve over time, we wanted to answer the question if subgroups of users differ in how they divide overall Web usage into Web sessions. For example, do subgroups of users show different loyalty to Web sites in Web sessions? Do some users stay in the Web longer than other users? In this regard, we were interested in how our five key measures of sessionized Web usage differ across subgroups of people.

We computed our key measures for each of the subgroups of users identified in [Christ01], who clustered users in the HomeNet sample into four distinct groups: low-rate users, moderate users, heavy users, and very heavy users, depending on how many distinct sites users visit per month. *Tab. 2* reveals that the number of sessions increased as overall Web usage increased. This relationship between overall Web usage and the number of Web sessions is almost tautological: users who use the Web rarely have fewer opportunities to visit Web sites. With respect to the duration of Web sessions, we see a substantial increase as overall Web usage increases. Note that the average duration of Web sessions of people who belong to the group of very heavy users was twice as long as the duration of sessions of low-rate users. Therefore, heavy users not only log on to the Web more often, they also spent more time online and consume Web sites in larger chunks.

Tab. 2: Key measures of Web usage in Web sessions across subgroups of users

	low-rate-users	moderate users	heavy users	very heavy users
Average number of sessions per month	2.6	7.6	13.3	36.1
Average duration of Web sessions (minutes)	21.0	26.8	32.5	40.7

Next we compared the key measures distinct sites visited, pages viewed, and our loyalty measure 'pages viewed per Web site' with the related monthly measures. *Tab. 3* reports results of this comparison.

Tab. 3: Key measures of Web usage in Web sessions across subgroups of users

		low-rate-users	moderate users	heavy users	very heavy users
monthly	Average number of distinct sites visited per month	3.2	18.5	34.1	183.8
	Average number of page views per month	10.8	92.3	153.8	817.5
	Ratio of average monthly page views per distinct Web site	3.0	5.1	4.6	4.4
Per Web session	Average number of distinct sites visited per session	5.3	5.9	8.1	11.5
	Average number of page views per session	14.3	17.6	24.4	31.9
	Ratio of average of monthly page views per distinct Web site per session	2.9	3.3	3.2	2.7

Notice that the number of distinct sites visited per month, which is the key measure [Christ01] use as a basis for clustering four user groups, differ substantially. However, there is much less of a difference if we consider the number of distinct sites visited per session. For example, low-rate users and moderate users visit almost the same number of distinct sites per session (5.3 and 5.9). The difference in monthly Web usage is apparently rather due to the lower number of Web sessions per month for low-rate users than due to lower Web utilization in sessions. The same argument holds for the measures 'average number of pages viewed per session' and 'pages viewed per Web site'.

Note that our measure of Web loyalty, the ratio of page views per site in sessions reported in *Tab. 3* is lower than the ratio of page views per site and month, as reported in *Tab. 1*. We consider this an indicator of (though limited) loyalty of Web users to Web sites. Keep in

mind that Web usage within a given month usually consists of many distinct Web sessions. Under the assumption that users do not visit Web sites randomly and there is an overlap in the identity of Web sites visited across sessions, the vocabulary set of Web sites visited by a given user does not increase in a linear way as the number of sessions increases. Therefore, the ratio of page views per Web site and month is supposed to be larger than the same ratio per session if a user shows loyalty to Web sites at least to some extent. In a case of complete loyalty, in which user's keep returning to the same Web sites, the domain vocabulary would be a constant whereas the number of page views increases in a linear way. The ratio of page views per site and month would be the ratio pages viewed/distinct Web sites per session multiplied with the number of sessions for this users in the given month. For example, the monthly ratio of pages viewed per Web site of a moderate user who is completely loyal to Web sites would be 3.3x7.6=25.1. In *Tab. 1*, this ratio is 5.1. The comparison of these two measures revealed that there is only very limited loyalty in Web sites visited.

3.2 Results of the longitudinal analysis

Next, we were interested in how our five key measures evolve over time. Fig. 3-6 report the longitudinal development of these measures.

Fig. 3 shows the duration of Web sessions across subgroups of users over time. Keep in mind that all users started using the Web in month 1. In this regard, we consider the longitudinal development of our key measures 'learning curves' after a natural starting point of having first access to the Web. Fig. 3 reveals that the duration of Web sessions stays almost constant over time. We see only a slight decrease across subgroups of users. Most importantly, we do not see a group that follows an upward path. This seems to confirm our hypothesis that Web sessions become shorter as expertise of users increases.

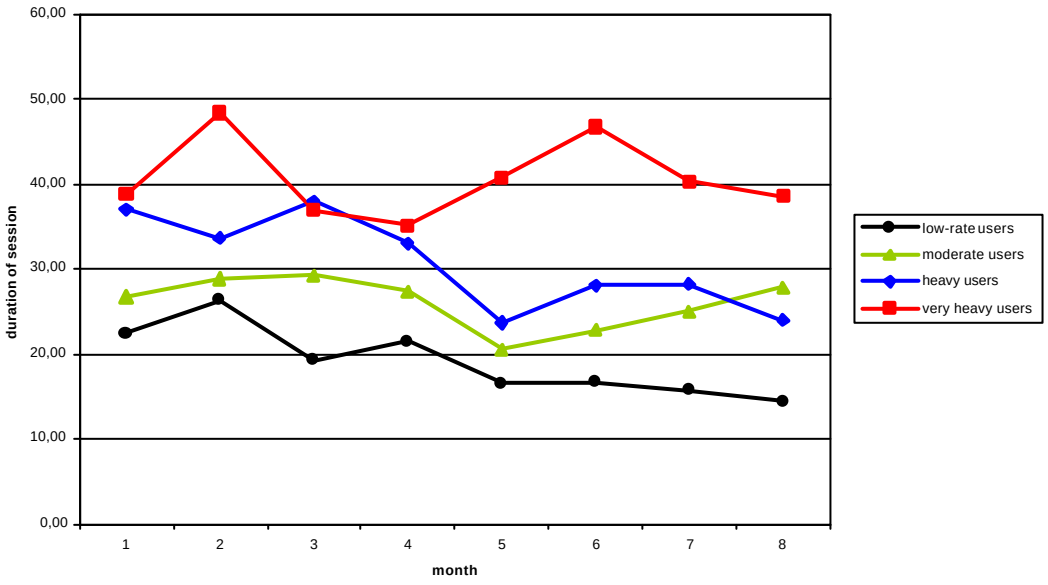

Fig. 3: duration of Web sessions across subgroups of users

Fig. 4 reports the number of distinct sites visited per session by individuals. In contrast to our expectations, the number of Web sites visited within sessions did not decrease over time. In particular, low rate users visit nearly the same number of distinct sites per session as moderate users. In general, there is much less of a difference between the various groups than if measured in overall numbers. This seems to confirm that heavy overall Web usage is rather due to more frequent use of the Web more than due to more intense in sessions. The number of sites 'consumed' per session stays almost constant over time. For each subgroup of users, there is constant Web site consumption in sessions over time. There also seems to be a lower and upper bound for the number of distinct sites visited per session. Even heavy users do not visit more than 12 sites per session. May be, there are cognitive processing limits that are responsible for this [Miller56]. On the other hand, even low-rate users consumed quite a few Web sites in their Web sessions. Overall low utilization rates for this group as described in [Christ01] are apparently due to a very limited number of Web sessions instead of limited Web usage within sessions.

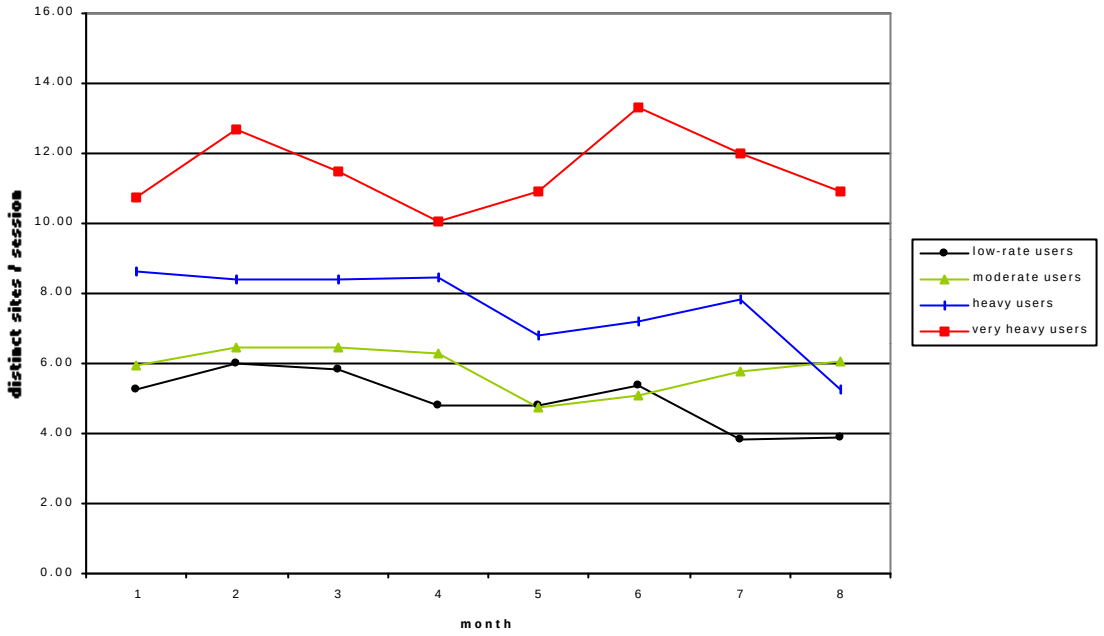

Fig. 4: Distinct Web sites visited within sessions

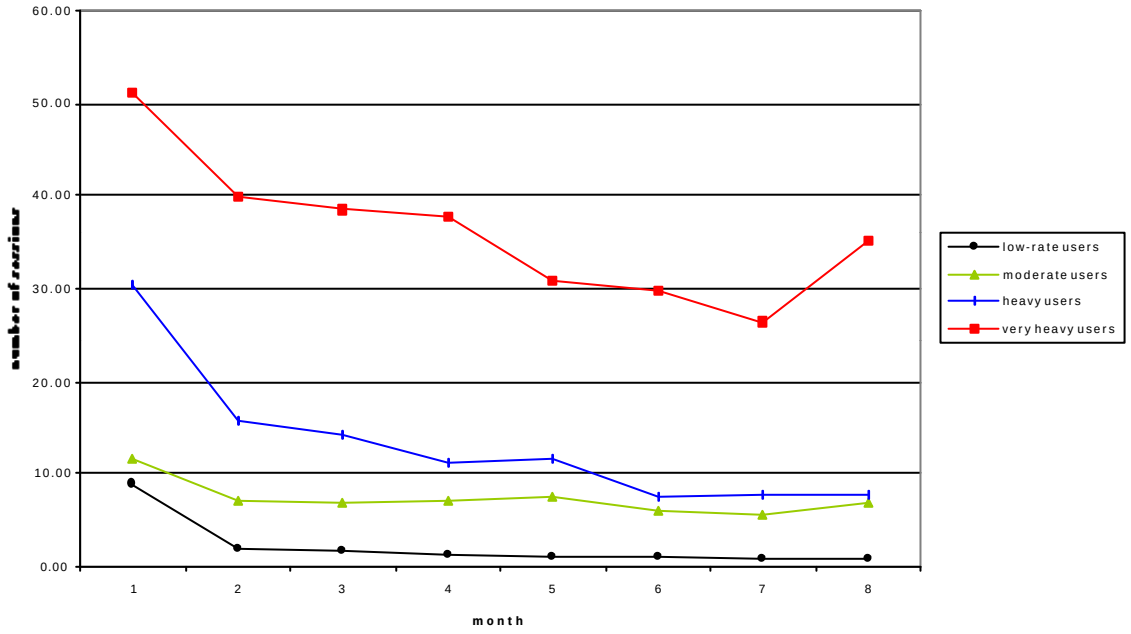

Fig. 5: number of Web sessions over time

Fig. 5 shows the number of sessions and how it evolves over time. Surprisingly, no group of users follows an upward trend. All groups follow a downward path until they reach saturation. Note that there is large similarity between the curves in Fig. 1 and Fig. 5. This seems to confirm that overall Web usage is highly dependent on the frequency of Web usage.

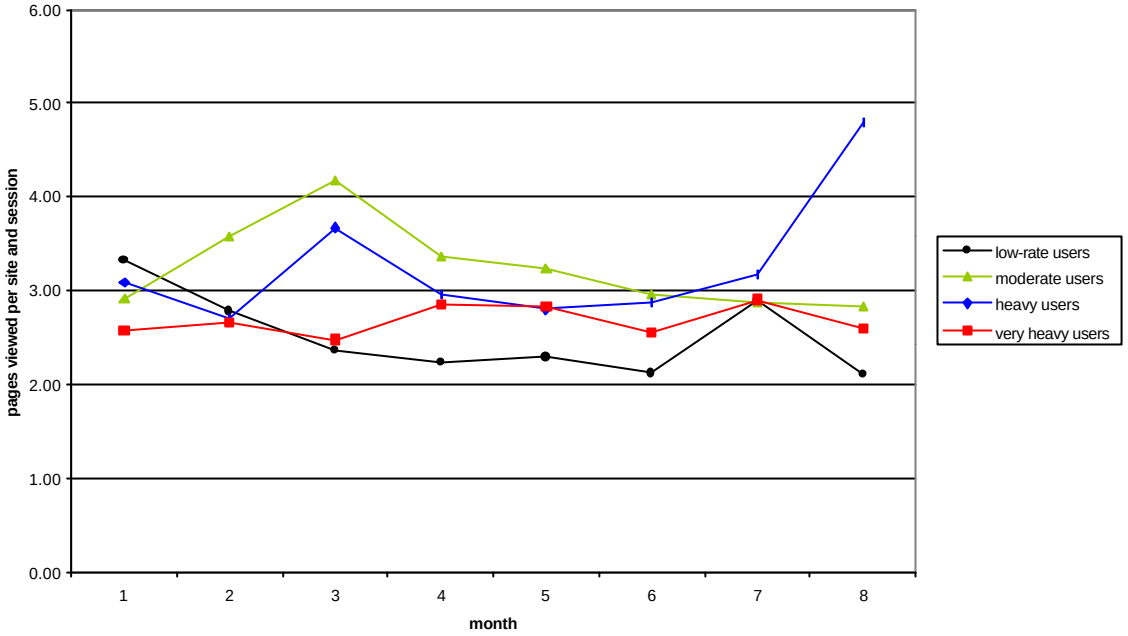

Fig. 6: Pages viewed per Web site within Web sessions

Fig. 6 depicts the development of the number of pages viewed per distinct site within Web sessions. Across subgroups, this measure stays more or less constant at a level of about three. Surprisingly, no group is following an upward path. In this regard, only very limited loyalty to Web sites in sessions develops over time. This conforms with [Christ02], who find a lack of commonality in the sites visited by users.

3.3 Results of the regression analysis

Judging from the results of the previous section, applying our key measures to the four subgroups of users in the HomeNet identified by [Christ01], does not help us in determining the individual characteristics that determine Web usage within Web sessions. Therefore, we also performed a more formal regression analysis. Specifically, we applied several POISSON models for which we chose the dependent variables 'number of sessions per month', number of distinct sites per session', 'page views per session', and 'duration of session'. Using a model which uses a poisson distribution is very common when the dependent variable is a count. The regression aimed to test, which individual characteristics determine Web usage in Web sessions. The following variables are the parameters of our models:

- 'white', which tells something about the ethnic background of individuals,
- 'adult' or 'age', which tells something about the age group of individuals,
- 'female', the gender of individuals,
- 'income', household income,
- 'c-skills', psychometric computer skill level,
- 'mail', mails sent weekly,

- 'phone', time spent using the phone.

We considered mail usage another indicator of Internet usage. Also, we wanted to test if Internet usage can be considered a substitute to telephone usage. For example, users might send email to friends instead of calling them. Also, users might retrieve information from the Web instead of calling somebody to get the needed information. Therefore, it was reasonable to assume that the Internet might be used as a substitute to the telephone.

Tab. 4: Poisson estimates: Determinants of number of Web sessions

```
                                                Prob > chi2     =     0.0000
Log likelihood =  -307.8775                     Pseudo R2       =     0.3426

------------------------------------------------------------------------------
# Websessions|      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |    .283475   .1333055     2.13   0.033      .022201    .5447491
         age |   .0103011   .0034153     3.02   0.003     .0036073     .016995
      female |  -.7061442   .1065865    -6.63   0.000    -.9150499   -.4972385
      income |  -.0080047   .0019257    -4.16   0.000     -.011779   -.0042303
     c-skill |   .1535055   .0467751     3.28   0.001      .061828    .2451829
        mail |    .194439   .0136804    14.21   0.000     .1676259    .2212521
       phone |   .1080724   .0484168     2.23   0.026     .0131773    .2029676
       _cons |   .7840724   .3239362     2.42   0.016     .1491691    1.418976
------------------------------------------------------------------------------
```

Tab. 4 reveals that a white Caucasian ethnical background, age, computer skills, mail and phone usage correlate with the number of Web sessions. Being female and household income have a negative impact on this measure.

Tab. 5: Poisson estimates: Determinants of duration of sessions

```
                                                Prob > chi2     =     0.0000
Log likelihood = -418.39166                     Pseudo R2       =     0.1206
------------------------------------------------------------------------------
    duration |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .1046968   .0602288     1.74   0.082    -.0133496    .2227431
         age |   .0042478   .0017762     2.39   0.017     .0007666     .007729
      female |  -.2456042    .053038    -4.63   0.000    -.3495567   -.1416517
      income |  -.0035302   .0010038    -3.52   0.000    -.0054976   -.0015629
     c-skill |   .1501078   .0254734     5.89   0.000      .100181    .2000347
        mail |  -.0397439   .0111477    -3.57   0.000     -.061593   -.0178948
       phone |   .0364762   .0264663     1.38   0.168    -.0153967    .0883491
       _cons |   2.745716   .1693715    16.21   0.000     2.413754    3.077678
------------------------------------------------------------------------------
```

The results of a regressions analysis of the dependent variable ‘duration of Web sessions’ comes to the following results: white Caucasian ethnical background, age, computer skills, phone usage correlate with the duration of Web sessions. Being female, mail usage and household income have a negative impact on this measure.

Tab. 6: Poisson estimates: Determinants of number of sites per sessions

```
                                                Prob > chi2     =     0.0005
Log likelihood = -206.61731                     Pseudo R2       =     0.0512
------------------------------------------------------------------------------
   Web sites |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .0720389   .1122298     0.64   0.521    -.1479274    .2920052
       adult |   .2545569   .1365446     1.86   0.062    -.0130657    .5221794
      female |   -.203511   .0970502    -2.10   0.036    -.3937258   -.0132961
     c-skill |   .1105222   .0506675     2.18   0.029     .0112157    .2098287
        mail |  -.0313086   .0229353    -1.37   0.172    -.0762609    .0136437
       _cons |   1.341345   .2492828     5.38   0.000       .85276     1.82993
------------------------------------------------------------------------------
```

Tab. 6 reports that a white Caucasian ethnical background, being adult, and computer skills have a positive impact on the number of Web sites per Web session. Being female and mail usage have a negative impact on this measure.

Tab. 7: Poisson estimates – Determinants of number of page views per session

```
                                                Prob > chi2     =     0.0000
Log likelihood = -339.46156                     Pseudo R2       =     0.1669
------------------------------------------------------------------------------
pages viewed |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .1564213   .0765005     2.04   0.041      .006483    .3063595
       adult |   .2176885   .0816161     2.67   0.008     .0577238    .3776532
      female |  -.3645287      .0638    -5.71   0.000    -.4895744   -.2394831
      income |  -.0018392   .0012002    -1.53   0.125    -.0041916    .0005132
     c-skill |   .1941714   .0305519     6.36   0.000     .1342907    .2540522
        mail |  -.0290614   .0137443    -2.11   0.034    -.0559996   -.0021231
       phone |   .0426757   .0318518     1.34   0.180    -.0197527    .1051042
       _cons |   2.069375   .1890624    10.95   0.000      1.69882    2.439931
------------------------------------------------------------------------------
```

The results of a regressions analysis with the dependent variable ‘number of page views per session’ revealed that white Caucasian ethnical background, being adult, computer skills, and phone usage correlate with the number of pages viewed per Web session. Being female, mail and household income have a negative impact on this measure.

The implications of our analysis for electronic commerce and public policy are discussed in the next section.

4 Conclusion, implications, and future work

The results of our study have several important implications both for business to consumer electronic commerce and for public policy as it pertains to the digital divide. In this paper, we introduced five key measures that help us to gain insights in individual Web usage in Web sessions. Judging from the results in the previous section, Web users spent only limited time in the Web. Regardless of the way of measuring Web usage in Web sessions, only a small group of users uses the Web heavily. We identified characteristics of individuals that influence Web usage in Web sessions, which include ethnic background, gender, household income, phone usage, email usage, and computer skill level. Particularly belonging to a minority group and being female determines low Web usage in Web sessions.

Moreover, we took repeated measures of Web usage in Web sessions over time in order to identify trends in Web usage. Specifically, we tried to identify how Web users change the way they use the Web as their individual level of expertise increases. Surprisingly, we do not see a significant shift from undirected browsing to directed access of Web sites over time. This seems to confirm that users keep exploring that Web even after eight months of Internet experience, which is reasonable if we consider that the Web itself is growing tremendously over our period of observation. However, we expected at least the fraction of directed browsing to increase over time.

4.1 Implications for electronic commerce and electronic marketing

The rejection of the hypothesis of increasing Web usage as described in [Christ01] is a first indicator that competition among Web companies for Web market share is likely to become more intense when the growth in terms of numbers of people accessing the Web slows down. We see this to a similar extent if we consider Web sessions. Further, judging from the results that users do not develop significant loyalty in the Web, it will be difficult for Web site operators to retain customers. Using an approach that is based on Web sessions advances the research on Web usage and helps to identify determinants of Web use that otherwise would not have been identified. Our analysis reveals that individual characteristics such as gender (male) ethnic background (white) determine heavy Web usage in Web sessions significantly. Surprisingly, household income does not play a significant role. Note that age has a positive impact on Web usage in sessions. Prior Work on the HomeNet data such as [Christ01] did find a negative impact of age on Web use. Clearly, using a more subtle session-based approach of measuring usage lead to new conclusions. Our analysis also revealed that Web users seem to have a limited capacity of Web usage in Web sessions. For example, even the group of very heavy users used the Web on average for 40 minutes only. This is twice the duration the group of low-rate users used the Web. In this regard, very heavy users not only used the Web more often but also stayed there longer. However, overall Web usage was rather determined by the number of sessions than by the duration of Web sessions. We see capacity constraints not only with respect to the duration of Web sessions but also with respect to the number of pages viewed and the number of distinct sites visited per Web session. This was already anticipated in [Christ01]. These discussions highlight the need to understand the reasons underlying the capacity limits we observed. It is possible that the limited capacity for Web site visits is due to the current technical shortcomings on the Internet (e.g., ease of use of sites, difficulty in using search engines, ineffectiveness of banner advertisements). Breakthroughs in technology can potentially increase the capacity for Web utilization and in turn the size of the market. For example, recent surveys (see the article "Ads Click" at http://interactive.wsj.com/articles/SB1004115312686358960.htm) demonstrate that ads returned in response to searches are effective in increasing clickthrough rates. Similarly, recent studies undertaken by MSN, Cnet and Doubleclick also demonstrate improved effectiveness of online marketing campaigns using reengineered advertising technology (http://www.iab.net/main/measuringsuccessfinal.pdf). However, it might well be the case that capacity limits on Web utilization are based on cognitive limits and cannot be mediated by technological breakthroughs. Determining the reasons underlying the capacity constraints is an important topic for future research.

4.2 Implications for public policy

As the Internet has grown and become more widely used by government and organizations, concerns have been raised about the digital divide [NTIA00]. The digital divide refers to those members of society who are unable to benefit from the Internet due to their lack of access to it or their inability to make full use of it. Studies such as [Hoffman96] have carefully examined the policy implications of demographic patterns of Web usage. Issues such as the gender gap and the race gap have been discussed and numerous studies [CA98] predict that while the gender gap will likely close over time, the race gap will prevail [Abrams97]. In these discussions of the digital divide [Hoffman98], the usual assumption has been that access to the Web will almost automatically trigger usage and thereby helps close the digital divide. Indeed, a recent report in the Wall Street Journal titled "Closing the Gap"

[Grimes01] discusses government subsidies that have been proposed as part of legislation such as Colorado's Information Technology Education Act for broadband access in rural and urban areas.

Since our data set has demographic data, we analyzed the data to determine if there are race and gender differences in utilization of the Web. As shown in section 3.3, gender and race determines Web usage in sessions significantly. The heavy users in our study are white males, whereas the low-rate-users are minorities and women. This implies that increased utilization of the Web will require more than access. As already stated, all users in the HomeNet panel received free computers with Internet connections and basic training in use of the technology. Informal reports indicate that customized training by gender or race may be needed in addition to access to enable different segments of society to fully benefit from the Internet. Additional work is required to develop policies that will be more successful in promoting utilization of the Web. We observed similar differences in the utilization of the Web by gender. [Shade] make several suggestions for ensuring gender equity to the information infrastructure. For example, to allow for the diverse needs of the women's community to be met with gender-sensitive and specific training. Note that even though the people in the HomeNet sample did receive training, this training was not of gender-sensitive nature, which could explain the still prevailing gender gap with respect to Web usage. Surprisingly, in contrast to the findings of [Christ01] who measured monthly Web usage, using a session-based approach revealed also that age has actually a positive impact on Web usage in sessions.

4.3 Future work

This paper extends the work of prior studies by presenting the results of a long-term study with residential subjects. We encourage future work with respect to three major issues: age of data, length of period of observation, and representativeness of the sample.

The patterns of Web usage we found were based on usage data from 1995-1997. Technical advances, e.g., in the field human computer interfaces in general or personalized recommender systems in particular can affect the intensity of Web usage. One of the main reasons for not using more recent data was to make sure that each user has a natural starting point of individual Web exposure. Further studies on people who did not use the Internet before are necessary to confirm the findings from 1998 onwards.

Our study is distinctive in its use of 8 months of continuous individual Web usage data. In order to gain insights in truly long-term changes in individual access behavior, the analysis of even longer samples of longitudinal data is desirable.

Our study relies on the subject group of people from the HomeNet project, which is close to a truly random residential sample. However, observed development in browsing behavior might arise due to cultural and social peculiarities of the subject group. Also, a significant share of the population accesses the Internet at work. Therefore, future research is necessary in order to confirm the findings for all groups of users. A truly random nationally representative sample is necessary for this work. Further, we encourage conducting this study in various international settings.

References

[Abrams97] Abrams, A. “Diversity and the Internet”. Journal of Commerce. June 26. 1997.

[CA98] CyberAtlas. “Demographics: Who’s on the Net in the US?” [http://www.cyberatlas.com/demographics.html]. 1998.

[Catledge95] Catledge, L., Pitkow, J. “Characterizing browsing strategies in the world wide Web”. In Computer Systems and ISDN Systems. Proceedings of the Third International World Wide Web Conference. 27. 1995.

[Christ01] Christ, M., Krishnan, R., Nagin, D., Kraut, R., Günther, O. “Trajectories of individual Web usage: implications for electronic commerce”. Proc. 34th Hawaii International Conference on System Science (HICSS-34). 2001.

[Christ02] M. Christ, R. Krishnan, D. Nagin, O. Günther, An Empirical Investigation of Web Site Stickiness, Proc. 2002 European Conference on Information Systems (ECIS-02), 2002.

[Cockburn01] Cockburn, A., McKenzie, B., What do Web user do: an empirical analysis of Web use. International Journal of human-computer studies, 54(6), pp. 903-922. 2001.

[Cooley00] Cooley, R., Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data, Ph.D. Thesis, University of Minnesota, 2000.

[Cooley99] Cooley, R., Mobasher, B., Srivastava, J., Data Preparation for Mining World Wide Web Browsing Patterns, Journal of Knowledge and Information Systems, 1(1), 1999.

[Fenichel81] Fenichel, C. H. Online searching: measures that discriminate among users with different types of experiences. Journal of the American Society for Information Science, 32(1), pp. 23-32. 1981.

[Fidel98] Fidel, R., Davies, R.K., Douglas, M.H., Holder, J.K., Hopkins, C.J., Kushner, E.J., Miyagishima, B.K., Toney, C.D., A visit to the information mall: Web searching behavior of high school students. Journal of the American Society for Information Science, 50(1), pp. 24-37. 1998.

[Grimes01] Grimes, A. “Closing the Gap”, The Wall Street Journal, Oct 29, 2001.

[Hoffman96] Hoffman, D. L., William, D., Kalsbeek, D., Novak, T. P. “Internet and Web use in the United States: Baselines for Commercial Development”. Communications of the ACM. 39 (12),1996, 36-46.

[Hoffman98] Hoffman, D. L., Novak, T. P. “Bridging the racial divide on the Internet”. Science. April 17. 1998.

[Hsieh93] Hsieh-Yee, I.. Effects of search experience and subject knowledge of the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 44(3), pp. 161-174. 1993.

[IDS02] “Internet Domain Survey, January 2002, “http://www.isc.org/WWW-200201/index.html”

[Khan98] Khan, K., Locatis, C. Searching thourgh cyberspace: the effects of link display

and link density on information retrieval from hypertext on the World Wide Web. Journal of the American Society for Information Science, 49(2), pp. 176-182. 1998.

[Kraut96] 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.

[McKenzie01] McKenzie, B., Cockburn, A. "An empirical analysis of Web Page revisitation". Proc. 34th Hawaii International Conference on System Science (HICSS-34). 2001.

[Miller56] Miller, G. A., "The magical number seven, plus or minus two: Some limits on our capacity to process information". Psychological Review 63, 1956, 81-97.

[NTIA00] "Falling Through the Net: Toward Digital Inclusion". [http://www.ntia.doc.gov/ntiahome/fttn00/contents00.html].

[Nua02] "Nua Internet How Many Online", February 2002, "http://www.nua.ie/surveys/how_many_online/"

[Shade] Shade, L.R. "Using A Gender-based Analysis in Developing a Canadian Access Strategy: Backgrounder Report". Universal Access Project, http://www.fis.utoronto.ca/research/iprp/ua/gender/GenderBased.html.

[Tauscher97] 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