Location and time do matter: A long tail study of website requests

Chetan Kumar, John B. Norris, Yi Sun

1. Introduction

There has been a tremendous growth in the amount and range of information available on the Internet. Cisco Systems report forecasts that global Internet Protocol (IP) traffic will increase six folds between 2007 and 2012, from fewer than 7 exabytes per month in 2007 to 44 exabytes per month in 2012. ComScore Inc. estimates that total global Internet users have surpassed 1 billion visitors in December 2008. The increase in Internet traffic is aided because making information available online is becoming relatively inexpensive, and as more people have Internet access demand for information increases. In addition, new format and content of Web 2.0 technologies such as video, social networking and collaboration applications prompted more interest in online deliveries. The trend of increasing Internet traffic is likely to continue.

The visitation of users to websites or online product purchase can be captured by a long tail model coined by Chris Anderson, shown in Fig. 1. Anderson used this model to explain the success of Amazon book and Netflix DVD rental recommendations system to promote obscure products. Brynjolfsson et al. had earlier noted this effect due to lower search costs in the digital economy. A few popular websites enjoy a high number of visitations. Interestingly there are also a large number of infrequently requested websites. The former is shown by the steep end of the curve, while the latter forms the tapering long tail.

Before the Internet age economic scale favored services that catered to a large amount of customers. For example, books that potentially attract more readers will be more likely published than those targeting niche markets. However the inexpensive online medium and reduced intermediaries lowered the hurdle of entrance. Websites with potentially small audiences can also exist because of inexpensive hosting costs for the information service provider. All sorts of information has a more equal chance to be present on the Internet and, with the assistance of efficient search engines such as Google, to be found by users. In this study we investigate this phenomenon by using real world data and show how users' location and time of access (weekdays versus weekends) affects this long tail model.

Our results can be used to design better online marketing strategies, affiliate advertising models, and Internet caching algorithms. According to Interactive Advertising Bureau Internet ad revenues grew to $21 billion in 2007, up 25% over 2006. However as online advertising is still only 10% of all US ad spending it has considerable room to grow. Internet based marketing, advertising, and content delivery is becoming increasingly important for businesses and institutions. Understanding patterns in user request behavior can provide guidelines for customizing advertising pricing for Internet portals such as Yahoo and Google. In addition they can be exploited for Internet caching algorithms to reduce user delays on the Internet. Therefore we believe this study contributes to an important area for Information Systems research.

The rest of this paper is organized as follows. We first discuss literature related to online visitation and marketing, followed by a description of methodology of analysis. Next we present the model...
and data analysis. Finally we conclude with discussion and areas for future research.

2. Related literature and methodology

2.1. Literature review

Researchers have long been interested in client-side study of human behavior in the context of the Internet. Some focused on the demographics attributes of Internet users and how these attributes affect users' browsing behavior [19]. Hu et al. [17] suggested a model to predict users' gender and age from their web browsing behavior. Others studied the way that users navigate the Internet [9]. Tauscher and Greenberg [29] and Cockburn et al. [11] demonstrated that the probability of re-visititation of a website is very high. Deborah [14] studied Internet searching habits and suggested that many users only use one or two search engines most of the time. Sen et al. [28] investigated the determinants of online search strategies and found that buyers' attitudes toward the price, their perception of online price dispersion, and their awareness of shopping agents affected their choice of online search strategy significantly. Breslau et al. [6], Almeida et al. [1] and Glassman [16], among others, modeled online visitations using the classical Zipf distribution [33] that relates object request frequency to rank. In his influential study Zipf [33] demonstrated an empirical law, that given some corpus of natural language such as English, the frequency of word occurrence is approximately inversely proportional to its rank in the frequency table. For example, in English the most frequently occurring word with about 7% of all occurrences. Following the Zipf distribution of “of,” which is the second ranked word, occurs about 3.5% of the time. More details of the Zipf distribution are provided in model and data analysis Section 3. Zipf distribution [33] is among the set of power law distributions such as the well known Pareto distribution [25] which characterizes the “80–20” rule, where 20% of the population controls 80% of the wealth. Brynjolfsson et al. [8] have shown that Pareto distribution can be modified with regard to concentration of product sales on the Internet due to effect of search costs. Along these lines in our study we first empirically demonstrate the long tail of website requests using the Zipf distribution [33]. Our study then goes beyond previous online visitation studies by focusing on the differences in user browsing behavior due to differing geographical locations and time of access. There are several reasons to warrant the study. Users from different areas have different demographic characteristics. For example, in certain regions in the United States (US) there are higher percentages of immigrant residents or high-skilled workers. Residents in these locations tend to have more diversified interests and the differences may exhibit in diversified visitations for online information. We also expect that time of access has an effect on diversity of website requests. For example, users may have different browsing behavior on weekdays versus weekends.

This knowledge can help business managers design better marketing strategies to allocate website advertising budget given observed patterns in user request behavior. It can also provide guidelines for customizing advertising pricing or affiliate marketing models for Internet portals such as Yahoo and Google. There is growing interest in research on online marketing and content customization. Brynjolfsson et al. [8] compared Internet buying habits to offline catalog purchases. Baye et al. [4] propose a model for pricing products and advertisements online. Sen [27] investigated the differences of search engine optimization and paid placements and proposed models that describe sellers' choice based on cost and consistency of rankings. Business community and academic researchers have also recognized the importance of customizing online content and marketing strategy based on users' preferences and their demographic attributes. Since total Internet advertising revenues is still only 10% of all US ad spending, it has considerable growth potential [3]. A significant area for improvement for Internet advertising is in customizing content to users. Luo and Seyedian [24] discuss the importance of contextual marketing in ecommerce based on different interests of users. Venkatesh and Agarwal [31] show how usability of website improves online purchase rate, and Bhatnagar and Papata [5] demonstrate the benefits of customizing online advertising according to user profile. We expect that the result of our research can enhance current online marketing research.

Furthermore we can use the knowledge to develop more effective caching algorithms. Caching involves storing copies of objects in locations that are relatively close to the user thereby reducing access delays [13,21]. Popular caching algorithms include least recently used (LRU), least frequently used (LFU), and their numerous variations [13,26,32]. However, most existing caching mechanisms do not consider location or time of access differences of the end user. Our results may help develop caching methods with priority given to users' locations and their time of access to further reduce user delays.

2.2. Methodology and hypotheses

In this study we conduct an empirical investigation on the long tail characteristics of website requests at different proxy server locations and times of access. We use web trace data from the IRCache network that maintains proxy servers at nine cities in the US (www.ircache.net). Using this data our first objective is to confirm the overall validity of using Zipf distribution to study long tail effects of online visitations. As discussed by Anderson [2], the adoption of the Internet as a distribution channel has lead to the existence of a long tail in infrequently requested goods and services in the digital medium. Past studies have used Zipf distribution to show relationship of web object rank to request frequency [1,6,16]. Our goal is to empirically demonstrate presence of the long tail effect of a large number of infrequently requested websites in addition to few very popular ones, using the Zipf distribution. Factors that contribute to presence of the long tail include reduced costs of hosting web content and the improved ability to search for niche information online [2,8].

Hypothesis 1. Long tail of website requests. There exists a long tail of large number of infrequently requested websites in addition to a few very popular websites.

We now consider differences in diversity of websites requests due to user's geographical location and time of access. This allows us to go beyond earlier studies by analyzing differences in long tail characteristics due to these two factors. The differences are measured by statistically testing variations in website request diversity. To examine location differences we focus on representative IRCache servers present in the four time zones of continental US; Silicon Valley at NASA Ames Research Center, California (Pacific time); Boulder, Colorado (Mountain time); Urbana-Champaign, Illinois (Central
time); and New York, New York (Eastern time). We then perform statistical tests to confirm differences in website request heterogeneity across different locations. User location is an important factor to consider because different geographical areas in the US have varying demographic profiles [30]. This may result in differences in diversity of website requests at different locations.

**Hypothesis 2.** User location and diversity of website requests. The user’s geographical location has an effect on the diversity of website requests.

Next we consider effect of time of access on diversity of website requests across locations. For this we partition the requests between weekdays and weekends. We then statistically test for differences in request behavior during weekdays and weekends at each location. We are interested to observe if users request diversified information sources at different times of access. At weekends users have more free time to explore personal interests and this may result in variations in website request behavior.

**Hypothesis 3.** User’s time of access and diversity of website requests. The time of access of the user has an effect on the diversity of website requests at every location.

Note that we partition data between weekdays and weekends at each location in order to examine differences due to user time of access. Considering smaller time intervals such as hours or minutes at each location may result in a multicollinearity problem. This is because these two factors may be correlated and can contribute redundant information to the model [18]. For example, users’ requests at different time intervals during a day may be affected by the demographic preferences at their locations’ time zones. This effect is minimized if we partition data for each location only between weekday and weekend. In addition we avoid over counting differences in time of access by not pooling data from all locations. By determining weekday and weekend. In addition we avoid over counting differences minimized if we partition data for each location only between access. Considering smaller time intervals such as hours or minutes at each location in order to examine differences due to user time of access.

3. Model and data analysis

We use a variation of the classical Zipf distribution [33] to characterize the long tail behavior of website requests. Zipf’s law in our setting may be stated as follows. The fraction of the time, \( f(R) \), that the \( R \)th most popular website is requested is given by:

\[
f(R) = \frac{\Omega}{R^s}
\]  

(1)

where \( R \) is the rank of websites in terms of number of requests, \( s \) is the exponent characterizing the distribution, \( \Omega = 1 / \left( \sum_{n=1}^{N} 1 / n^s \right) \) is a normalizing constant and \( N \) is the total number of requested websites. In his seminal study Zipf demonstrated that in the English language the frequency of word occurrence is approximately inversely proportional to its rank in the frequency table, with \( s \) value slightly greater than 1 [33]. Parameter \( s \) determines the relative impact that higher ranked objects have in the overall distribution as shown in Fig. 2. Note that the \( f(R) \) vs. rank \( R \) mapping captures the long tail behavior of relatively few sites being requested very frequently and a large number of infrequently requested websites. Some transformations allow us to connect and characterize both Zipf distribution and the long tail behavior using a log-linear model. Performing log transformation on Eq. (1) we obtain:

\[
\log f(R) = \log \Omega - s \log R
\]  

(2)

Eq. (2) is analogous to fitting a log-linear regression relationship between rank \( R \) of site and frequency of requests \( f(R) \) as follows:

\[
\ln f(R) = \beta_0 + \beta_1 \ln |R| + \epsilon
\]  

(3)

Parameter \( s \) of (2) is directly captured by coefficient \( \beta_1 \) of (3), and \( s \) also indirectly contributes to \( \log \Omega \) and intercept \( \beta_0 \) terms of the two expressions, respectively. The log-linear relationship (3) has been used successfully by economists to study distribution of income and wealth [23,25]. More recently Brynjolfsson et al. [8] have used it to show the effect of search costs on concentration of product sales. In our context \( \beta_0 \) can be interpreted a measure of overall demand for unique websites at a server location, while \( \beta_1 \) measures how quickly the share of total number of requests attributed to a particular website falls as site rank increases. A benefit of using expression (3) is that we can test if the differences in \( \beta_1 \) for various locations or time periods are statistically significant.

3.1. Long tail of website requests (H1)

Our first objective is to empirically demonstrate that there exists a long tail behavior of large number of infrequently requested websites in addition to a few very popular ones. For this we fit a log-linear regression model for the comprehensive trace data of the IRCache network. The data was collected from 29 April to 30 June, 2004. It includes the URLs of requests, the request time, the format of object requested, unique identifiers for user IP address, and the elapsed times for serving requests. The comprehensive proxy trace of all nine IRCache server locations, referred as All Servers, includes 926,552 total website requests. Table 1 summarizes the regression results for All Servers, as well as for one IRCache location in each of the four US time zones: Silicon Valley (SV), Boulder (BO), Urbana-Champaign (UC), and New York (NY) (Eastern time). We then statistically test for differences in request behavior during weekdays and weekends at each location. We are interested to observe if users request diversified information sources at different times of access. At weekends users have more free time to explore personal interests and this may result in variations in website request behavior.

![Zipf Distribution](image)

Fig. 2. Zipf distribution and characterizing exponent \( s \).
and New York (NY). The high $R^2$ value of 0.97 for All Servers indicates that there is a good fit between frequency of website requests and rank of the websites. This is also demonstrated in Fig. 3 which shows a good linear fit in a scatter plot between the two variables on a log–log scale. The approximate straight line in the figure, with slope $\beta_1 = -1.365$, displays a Zipf distribution for frequency vs. rank for websites. This in turn confirms Hypothesis 1 that the long tail behavior for website requests does exist at proxy servers.

The absolute value of $\beta_1$ coefficients corresponds to parameter $s$ of the Zipf distribution. A smaller $|\beta_1|$ indicates greater diversity in requests as websites with lower ranks retain a higher share of total requests. Conversely, larger $|\beta_1|$ indicates more homogenous website requests with the most popular websites making up a higher proportion of total requests.

3.2. Effect of user location on diversity of website requests (H2)

Next we test if there are significant differences in the characteristics of websites requests depending on the locations of the proxy servers. Fig. 4 demonstrates that there is a good linear fit for log (frequency of requests) vs. log (rank of sites) for the individual servers located in the four cities mentioned earlier. The high $R^2$ values of 0.976, 0.981, 0.983, and 0.978, respectively indicate long tail request behavior at each of the server locations (refer Table 1). We can use the $\beta_1$ coefficients to statistically test for differences in heterogeneity of website requests. Utilizing a paired sample test, we measure if the difference in $\beta_1$ coefficients for any two locations is significant.

Table 2 shows difference in $\beta_1$ corresponding to the pair of locations listed in each row and column. The $t$ statistic for this difference is $\frac{(\beta_1 j - \beta_1 j') \sqrt{(n_j + n_j')}}{\sqrt{\text{Var}(\beta_1 j - \beta_1 j')}}$, where $j$ denotes server location. Using this, we confirm the difference in $\beta_1$ coefficients is significant ($p < 0.001$) for all $4C2 = 6$ pairs of locations. This supports Hypothesis 2 that the geographical location has a significant effect on diversity of website requests.

The distribution of $\beta_1$ coefficients for the four server locations is shown in Fig. 5. It shows that users from different locations do have different interests in online access. We can use this for judging if some locations can be grouped together based on website request patterns. For example, Silicon Valley and New York locations display greater heterogeneity in user’s choice of information than Boulder and Urbana-Champaign. This can be explained partially by the fact that the former two areas have higher percentages of international immigrants and high-skilled workers. We use data from the US Census Bureau [30] to confirm this. The US Census Bureau is a

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
& Boulder & Urbana-Champaign & New York \\
\hline
Silicon Valley & 0.144 & 0.164 & 0.017 \\
Boulder & 0.164 & 0.019 & -0.127 \\
Urbana-Champaign & 0.019 & 0.017 & -0.147 \\
\hline
\end{tabular}
\caption{Difference in $\beta_1$ coefficients for four locations.}
\end{table}

* Significantly different from zero, $p < 0.001$. 

![Fig. 3. Frequency of requests vs. rank of sites for all servers.](image)

![Fig. 4. Frequency of requests vs. rank of sites for four server locations.](image)
For that we partition the trace data to website requests on weekdays and browsing behavior.

3.3. Effect of user’s time of access on diversity of website requests (H3)

Next we consider the effect of time of access for website requests. This can also be compared in Fig. 6 that shows the long tail behavior does exist on both weekdays and weekends. As before, we use the $\beta_1$ coefficients in each case to statistically test for differences in heterogeneity of website requests. The $t$ statistic for this difference is $t = (\beta_{wd} - \beta_{we}) / \sqrt{Var(\beta_{wd} - \beta_{we})}$, where $wd$ and $we$ denote weekday and weekend, respectively. Using this, we confirm that $\beta_1$ coefficients between weekday and weekend is significant ($p<0.001$) for all four locations. This supports Hypothesis 3 that time of access does have a significant effect on diversity of website requests. This can also be compared in Fig. 6 that shows distribution of weekday and weekend coefficients, $\beta_{wd}$ and $\beta_{we}$, for each location. We observe that distributions are indeed different. In Table 4 we observe a trend that $|\beta_{wd}|$ and $|\beta_{we}|$ are lower at each location compared to weekdays. This indicates that there is greater diversity of requests on weekends. We reason this to be because users have more free time to explore personal interests during weekends rather than focus on work requirements during weekdays. This allows them to access diversified information sources during weekends.

These results are important on multiple levels. First, online marketing strategies may be segmented to better target users from different locations. For example, a marketing company for a niche product, such as an exotic oriental vase, may allocate more resources to locations with more diversified website visits. Another example is that, rather than promoting a set of products uniformly across all locations, Internet portal websites may choose to advertise obscure products more in heterogeneous regions while focusing on mainstream product in more homogeneous regions. Second, Internet portals can adjust pricing schedules based on different regions at different times. For example, we propose that web search advertisers may modulate that some locations have greater affinity for niche websites than others. This may indicate to an extent that persons with higher paid jobs tend to have more diversified interests and browsing behavior.

### Table 3
Demographic profile for four locations.

<table>
<thead>
<tr>
<th>Server locations (counties where servers are located)</th>
<th>Foreign born persons, percent, 2000</th>
<th>Language other than English spoken at home, percent age 5+, 2000</th>
<th>Per capita money income, 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silicon Valley (Santa Clara)</td>
<td>34.1%</td>
<td>45.4%</td>
<td>$$32,795$</td>
</tr>
<tr>
<td>Boulder (Boulder)</td>
<td>9.4%</td>
<td>13.6%</td>
<td>$$28,976$</td>
</tr>
<tr>
<td>Urbana-Champaign (Champaign)</td>
<td>8.0%</td>
<td>11.8%</td>
<td>$$19,708$</td>
</tr>
<tr>
<td>New York (New York)</td>
<td>29.4%</td>
<td>41.9%</td>
<td>$$42,922$</td>
</tr>
</tbody>
</table>

The Standard Errors are in parentheses, $M =$ total website requests.

* Significantly different from zero, $p<0.001$.

Table 4 Log-linear regression for frequency of website requests onto rank of site for four locations on weekday and weekend.

<table>
<thead>
<tr>
<th>Server locations</th>
<th>Time of access</th>
<th>$\beta_1$</th>
<th>$\beta_0$</th>
<th>$R^2$</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silicon Valley</td>
<td>Weekday</td>
<td>12.855 (0.0054)</td>
<td>-1.124 (0.0005)</td>
<td>0.976</td>
<td>122,490</td>
</tr>
<tr>
<td></td>
<td>Weekend</td>
<td>10.750 (0.0092)</td>
<td>-1.053 (0.0096)</td>
<td>0.969</td>
<td>37,421</td>
</tr>
<tr>
<td></td>
<td>Difference in $\beta_1$</td>
<td>0.0713</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Boulder</td>
<td>Weekday</td>
<td>14.053 (0.0005)</td>
<td>-1.265 (0.00047)</td>
<td>0.980</td>
<td>136,599</td>
</tr>
<tr>
<td></td>
<td>Weekend</td>
<td>12.679 (0.0079)</td>
<td>-1.201 (0.00081)</td>
<td>0.979</td>
<td>46,548</td>
</tr>
<tr>
<td></td>
<td>Difference in $\beta_1$</td>
<td>-0.0618</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Urbana-Champaign</td>
<td>Weekday</td>
<td>14.318 (0.0048)</td>
<td>-1.253 (0.00044)</td>
<td>0.983</td>
<td>140,406</td>
</tr>
<tr>
<td></td>
<td>Weekend</td>
<td>13.350 (0.0072)</td>
<td>-1.264 (0.00073)</td>
<td>0.984</td>
<td>46,342</td>
</tr>
<tr>
<td></td>
<td>Difference in $\beta_1$</td>
<td>-0.0156</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>New York</td>
<td>Weekday</td>
<td>12.984 (0.0051)</td>
<td>-1.131 (0.00048)</td>
<td>0.977</td>
<td>125,551</td>
</tr>
<tr>
<td></td>
<td>Weekend</td>
<td>11.325 (0.0079)</td>
<td>-1.088 (0.00081)</td>
<td>0.976</td>
<td>43,328</td>
</tr>
<tr>
<td></td>
<td>Difference in $\beta_1$</td>
<td>-0.0426</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

The Standard Errors are in parentheses, $M =$ total website requests.

* Significantly different from zero, $p<0.001$.
markets. For example, in Fig. 7, \( p_1 > p_1' \) and \( p_2 < p_2' \). Case 2: for the same product being advertised in two markets, the more demanded product \( m_1 \) may be priced higher \( p_1 \) at less diversified market than \( p_1' \) at more diversified market. On the other hand, lower demand product \( m_2 \) may be priced lower \( p_2 \) at less diversified market and priced higher \( p_2' \) at more diversified market. For example, in Fig. 7, \( p_1 > p_1' \) and \( p_2 < p_2' \). The same reasoning may also be used for customizing Internet advertisements and marketing strategies based on the characteristics of website visitations. This approach is an improvement over current cost-per-click advertising and other online marketing models on the Internet that do not effectively consider location and time of access differences of users.

Finally, our results may help to design better Internet caching algorithms that are aware of the difference in the users’ request patterns based on their locations. An example is as follows. For website locations and time periods with greater diversity of website requests we assign closer to equal priorities for caching different objects. With lesser diversity of requests we assign greater priorities to objects that are more popular. This would be an improvement over current LRU based caching mechanisms, such as those used at IRCache network (www.ircache.net), that do not consider server locations and time of access differences [20].

4. Discussion and conclusions

The users’ requests for online information can be captured by a long tail model. Our study confirmed this phenomenon using the Zipf
distribution. However previous models do not consider the impact of location and time of access on the diversity of website requests. Therefore we examined the differences in user browsing habits due to location and time of access using an actual proxy trace data. Our tests confirm our hypotheses that server location and time of access indeed have an effect on the heterogeneity of website requests. This can partially be explained by differences in demographic characteristics at locations, due to differing proportions of international immigrants and high-skilled workers, and diverse browsing behavior between weekdays and weekends. Our results can be used for designing better online marketing strategies, affiliate advertising models, and Internet caching algorithms with sensitivities to user location and time of access differences.

This study examines users’ online request patterns using data from multiple proxy servers. We analyzed the data by server locations and time of access. There are several interesting research avenues to pursue in the future. In the subsequent study we plan to expand the analysis to include patterns in user IP addresses across multiple days and locations. In addition our paired comparison method for server locations opens up avenues for using other techniques such as a Tukey–Kramer ANOVA for multiple comparisons. However an ANOVA procedure requires additional assumptions regarding equality of variance across locations. An area of future research is to develop tests for differences both within and across locations. Another line of reasoning is to compare if the extent of difference in weekday and weekend $|β|$ coefficients is influenced by location or other demographic factors. For example, in Table 4 weekday vs. weekend difference in $β$ for SV, BO, UC and NY are $−0.0713$, $−0.0618$, $−0.0156$ and $−0.0425$, respectively. While we have shown that the differences are statistically significant for a particular location, there may be some differences across locations. An area for research can be to develop a statistical measure for testing extent of time of access difference across locations and to provide a rationale for the same. Finally one can also examine alternative approaches that consider smaller time intervals, such as hours or minutes, focused on specific time zones to extract more managerial implications from user browsing patterns.

To the best of our knowledge our research is the first to analyze diversity of website requests based on user location and time of access. There is a significant potential to improve on online marketing, advertising, and content customization by better understanding patterns in user request behavior. In addition they can be exploited for Internet caching algorithms to reduce user delays due to increasing congestion. Therefore we believe that this study contributes to commerce research that is beneficial for users, technology providers, and businesses on the Internet.

List of notations

- $f(R)$: fraction of time that Rth most popular website is requested
- $R$: rank of websites in terms of number of requests
- $s$: exponent characterizing Zipf distribution
- $Q$: normalizing constant for Zipf distribution
- $N$: total number of requested websites ($n = 1$ to $N$)
- $β_0$: intercept of log-linear regression that measures overall demand for unique websites
- $β_1$: coefficient of log-linear regression that measures how quickly the share of total number of requests attributed to a particular website falls as site rank increases
- $β_{1j}$: coefficient for any location $j$
- $β_{1wd}$: coefficient for weekdays $wd$
- $β_{1we}$: coefficient for weekends $we$
- $R^2$: coefficient of determination measure of goodness of fit of log-linear regression

Acknowledgements

We thank the seminar participants of the 2008 INFORMS Annual Meeting and the 2007 Workshop on e-Business (WEB) for valuable comments on this study.

References

[14] F. Deborah, Search Engine Users: Internet Searchers are Confident, Satisfied and Trusting — But They are Also Unaware and Naive, 2005 Available at: http://www.pewinternet.org/pdfs/IPISearchengine_users.pdf.
Chetan Kumar is an Assistant Professor in the Department of Information Systems and Operations Management at the College of Business Administration, California State University San Marcos. He received his PhD from the Krannert School of Management, Purdue University. His research interests include managing computer networks, electronic commerce, web analytics, peer-to-peer networks, caching, and IS strategy for firms. His research has been published in Decision Support Systems journal. He has presented his research at INFORMS, Workshop on e-Business (WEB), Workshop on Information Systems and Economics (WISE), ICIS Doctoral Consortium, and AMCS Doctoral Consortium conferences.

John B. Norris received his PhD from the Quantitative Methods area at the Krannert School of Management, Purdue University. His research interests include web analytics, healthcare management, and decision support tools for student team assignment. His research has been published in Decision Support Systems journal. He has presented his research at AOM, DSI, INFORMS, and POMS conferences.

Yi Sun is an Assistant Professor of Information Systems at the College of Business Administration, California State University San Marcos. His degrees include BA degree from Foreign Affairs College in Beijing, China, and PhD (Management Information Systems) from University of Florida. He has research interests in telecommunications, data mining and artificial intelligence, electronic commerce, and applied operations research. He has published several articles in refereed journals such as Decision Support Systems and European Journal of Operational Research.