The Economics of Surfing

Eytan Adar and Bernardo A. Huberman

Xerox Palo Alto Research Center
Palo Alto, CA 94304

Abstract

We have established that depending on the domain of inquiry, users display different and regular patterns of surfing. This difference can be exploited in order to benefit information providers. We propose mechanisms for implementing temporal discrimination in surfing by dynamically configuring sites and versioning information services.

Introduction

One of the most important aspects of the explosive growth of the World Wide Web is the potential for electronic commerce that the web offers. The novelty of the medium and the ease of access not only lead to interesting market structures[Ad99], but to different ways of searching for services and interacting with consumers. Over the past four to five years we have seen the appearance of several important WWW services for electronic commerce, of which a relatively new instance are the portal sites. Some important examples are Yahoo[Ya99], Excite[Ex99], and Lycos[Ly99]. Portals attempt to act as a starting point for users on the web and therefore lead consumers to electronic commerce activities, such as travel and consumer electronics. Portals are in fact second generation services, and are essentially a refinement of the web search engine service.

The business model of portal services consists of two parts. The first one is to have the consumer buy goods directly from the portal or through a partner site. The other strategy involves satisfying the user's information needs locally and in the process present advertising banners that result in direct revenues to the portal. The mechanics of this approach are relatively simple. Users enter a portal site, search for information, see ads, and then usually leave.

While this model seems to have worked for various companies, it is highly paradoxical in the sense that it confronts the portal with a dilemma. On the one hand, the service strives to generate better and faster results for the users. However, if the results are presented at the entry point, leading users to another site, there is an opportunity cost for the provider. The user never travels deeply into the site, thus missing the advertising and consumer goods that he potentially could buy. Witness the case of the Lycos search service, which in order to solve this dilemma has now implemented mechanisms that work to prevent users from leaving to other search services, such as Excite[Hu99]. Until recently, when a user searched for the word "excite" on the Lycos site, the results attempted to dissuade the user from leaving by presenting a link back to the Lycos search page.

A recent heuristic solution to this problem resorts to the notion of stickiness, whereby the provider attempts to keep consumers at their site by displaying potentially attractive links that only point into their site. This is much in line with standard marketing practices in the physical world. An alternative solution, which we here explore, attempts to capture the existence of different surfing patterns in order to solve this dilemma. Instead of blocking surfing paths, one increases the depth to which users surf when accessing an e-commerce provider.

In this paper, we show that depending on the domain of inquiry, users display different and regular patterns of surfing users when accessing different kinds of information goods or services. This difference can be exploited in order to benefit information providers. We propose mechanisms for implementing temporal discrimination in surfing by dynamically configuring sites and versioning information services**.** This is done by exploiting the fact that surfing patterns are extremely regular and described by a universal law[Hub97]. We also propose mechanisms for extracting consumer surplus by dynamically configuring sites and versioning information services[Va99].

Temporal Discrimination

Consider the hyperlink structure of a site as shown in Figure 1. It is clear that it allows for several ways through which a user reaches the desired information goal. For each path there is an associated time, which will depend on the number of links and the amount of time that the user spends at each page. In the context of attention economics[Go97], time is a proxy for the price a consumer pays to access the information.

As shown in the figure by link 1, a user can get to the goal in one step or, because of ignorance about the site organization, through several links, thus making it costlier to access the desired information (for example link 2 and then 3). Alternatively, a user may store a link pointer as a bookmark, thereby having direct access to the desired page, avoiding any navigation. If this is the case, it implies that different users willing to pay different prices for the same informational good .

To verify this hypotheses, we show in Figure 2 what amounts to a demand curve in the time domain. The figure depicts the value of 1 minus the cumulative distribution function for the amount of time (in number of clicks) that users are willing to spend for a given service need. This depends on the link structure of the information provided by the site[1]. This data was produced by analyzing the surfing patterns of 525,000 unique visitors to a major web portal, and is a reflection of the fact that there is a law of surfing which determines in a universal way the surfing patterns of visitors to any web site[Hub97].

The law of surfing, which has been verified empirically in a number of cases, determines the probability distribution *P(L)* of the number of pages *L* that a user visits within a Web site. It is given by Equation 1 below.

---

[1] The existence of such a demand curve does not have a unique explanation, for it is not clear whether or not this is due to the particular organization of the website, or the demographic characteristics of the user population.

where L is the number of links, and λ and μ are parameters which determine the mean and variance of the distribution.

This data points to the fact that if information providers satisfy consumer needs at a fixed number of clicks, they would be missing the additional time that some users are willing to spend in order to get that same information.

$$P(L) = \sqrt{\frac{\lambda}{2\pi L^3}}\left[\frac{-\lambda(L-\mu)^2}{2L\mu^2}\right]$$

Equation 1: Law of Surfing

Measuring Surfing Patterns in the Web

While an aggregate demand curve seems indicative of different user preferences in accessing information, it is of interest to know whether or not specific surfing patterns depend on the information being accessed. In order to answer this, we ran a set of experiments on user logs provided by a large web portal and Excite[2].

The web portal data consisted of anonymized usage logs for an eleven day period between September 14, 1998 and September 25, 1998. The logs contain unique identifier for users, timestamps, unique (but anonymous) identifier for the requested page within the directory hierarchy and demographic information (when provided by the user). The data was filtered for requests only within portal's directory service (versus financial news, weather, etc.). The resulting data set contained over five million impressions, or page views, generated by over five hundred thousand unique visitors.

The measured function (one minus the cumulative distribution function (CDF) of the depth L) is shown in Figure 3a, where we plot the quantity 1-CDF as a function of depth of surfing, L, for three top level categories in the site, and genders. Figure 3a was obtained by fixing the level of surfing L and measuring how many users reached that level. The difference between any two levels, L and L' reflects the stop rate, i.e. the number of people who stop surfing at the first depth L.

---

[2] To our knowledge neither Excite nor the web portal (which has requested to remain anonymous) use the techniques described in this paper.

An alternative representation of this data is provided by Figure 3b, which depicts the same three category classification in terms of the demand curve. Once again, there is a noticeable shift among different categories being explored.

As the measurements show, there is a marked difference in the typical depth to which users surf depending on the category, while only a slight difference was detected among surfers of different gender. Moreover, the tails of the distributions for each category are markedly different, which implies that certain categories are surfed deeper than others.

As to the slight gender differences that we observed we suggest that they are due to the fact that the data has been aggregated at a top level category, thus being unable to finely discriminate between gender. Had a finer resolution been available, as for example between fast cars and female cosmetics, we would have noticed a gender difference in surfing patterns.

Another experimental test of a shift in the maximum of the surfing curves was provided by an analysis of surfing data from Excite that discriminated between adult and non-adult type sites. In 1997 Excite provided the research community with a small, anonymized usage log, containing over fifty thousand page views from over ten thousand unique visitors. The log entry contained the unique user id, a timestamp, and the query terms applied to search engine. Each entry provides an indication that the user viewed a single result page containing, by default, ten matches. Two entries in the log for the same search by the same user indicate that the user viewed twenty results.

Discrimination was obtained by devising a filter for the search queries that separated any explicitly sexual query from the rest. The results are shown in Figure 4a, where we again plot the quantity 1-CDF as a function of surfing depth, L, for adult and non-adult sites. As can be seen there is a slight shift towards the right (indicating increased demand) for the adult queries as compared to non-adult. We believe that this shift will be more noticeable when adult sites are compared to more specific sites, such as weather or local news. In analogy to our analysis of the portal data, in Figure 4b we present the demand curves for the adult and non-adult information.

Mechanisms for Temporal Discrimination

In what follows we propose three specific strategies that can be used to extract maximum surf depths from consumers exploring a portal. The first one is a general procedure that works with any commercial web site, whereas the second is ideally suited for information providers, such as search engines and shopping directories. Lastly, we discuss the use of coupons and their strategic placement within a site.

Dynamic reconfiguration of web links.

One can construct a web site that changes its link structure to lengthen the path traversed by a given user, thereby making him visit many more pages. For example, if there is a short route (in the number of clicks) to a given page, one may wish to turn that off if the user is likely to visit more pages in between. An example of such dynamic organization is described in Figure 4.

Notice that a user entering node 1 only has the choice of entering node 2 or leaving the site. At Node 2, a user now has three choices. If one can predict with some degree of certainty (based on the user's state) that the user's trajectory will pass through node 4 or 5, and there is a user willingness to continue clicking further, one can remove the direct links from nodes 2 to 4 and 5. This can be done based on statistically derived confidence levels that determine how many intermediate nodes a user is willing to travel through to arrive at her eventual goal.

Versioning information goods

We now describe an experimental technique to extract additional clicks for a given search engine type web site. This technique is based on the existence of the law of surfing, which determines the number of users that will surf for a given depth within a site[Hub97].

The basic scheme consists in collecting surfing data for each site for a fixed level of versioning or quality of service. For a search engine this means a set precision/recall [Sa68], and it will lead to surfing curves shown schematically in Figure 6. When these surfing distributions are expressed in terms of one minus their cumulative distributions, they are consistent with the data shown in Figures 3 and 4. Thus the typical number of links a user will surf determines when to provide the good quality information or the incentives to go to other pages.

In order to extract the consumer surplus indicated by the shifted maxima in the curves, one can offer a second version of the service with different quality characteristics. This new version can be integrated into the service along with decision rules -trained by the usage logs- that decide which version should be offered to the user. Thus, users can now continue to surf along this new version and therefore spend more time in a given provider's site.

Coupons

Another mechanism for optimal placement of advertisement banners borrows from the marketing technique of "seducible moment". Basically it amounts to presenting the banner at the point where the information needs of the user are satisfied, and where a user is more likely to notice the advertisement and therefore continue to surf the site. This can be done by using the demand curves described here, as they allow sites to predict with increased accuracy where the seduction points will be. Furthermore, advertisements presented at this moment can be sold to advertisers at a premium, as click through rates (the percentage of people clicking on an ad) will likely increase.

Conclusion

In this paper we presented evidence for differences in surfing behavior conditioned on the domain of inquiry by examining the paths of over five hundred thousand unique visitors to two popular web services. We established that depending on the domain being explored, consumers display different surfing characteristics which, while described by the same law, have different modes and tails.

One surprising result of our study was the lack of any noticeable gender dependence on the surfing patterns of so many users. While we expected to see large differences in behavior when coupling gender differences to domain specifics, so far this large body of the data does not support this conjecture. Further research may uncover some differences when analyzing domains at a finer structural level.

Nevertheless, the noticeable shift in the demand curves for different domains, suggested three possible mechanisms by which providers can exploit it. In all these cases, what is attempted is to capture the willingness of consumers to surf at a given site for a longer period of time than the average one. This approach is to be contrasted with those based on standard marketing techniques, which rely on notions of stickiness to keep consumers at a given site. While our measurements provided a first cut at this interesting problem, they are not by any means exhaustive. It is to be hoped that a more refined analysis of surfing data over specific sites will uncover other indicators.

Acknowledgements

We thank Lada Adamic, Rajan Lukose, Jim Pitkow and Tad Hogg for discussions.

References

[Ad99] Lada A. Adamic and Bernardo A. Huberman, "The Nature of Markets in the World Wide Web", proceedings of the Conference for Computational Economics, Boston, June 24-26, 1999.

[Ex99] Excite Inc., http://www.excite.com

[Fr90] David D. Friedman, Price Theory, South Western Publishing, 1990

[Go97] Michael H. Goldhaber, "The Attention Economics and the Net", First Monday, 1997.

[Hu99] Jim Hu, “Lycos’s loop has users reconsider rivals,” CNET News.com, April 2, 1999. http://www.news.com/News/Item/0%2C4%2C34588%2C00.html?st.ne.fd.gif.i

[Hub97] Bernardo A. Huberman., Peter L. T. Pirolli, James E. Pitkow, and Rajan M. Lukose, “Strong Regularities in World Wide Web Surfing,” *Science*, April 3, 1997, 280(5360), pp. 95-97.

[Ly99] Lycos Inc., http://www.lycos.com

[Sa68] G. Salton, and M. E. Lesk, “Computer Evaluation of Indexing and Text Processing,” *Journal of the Association for Computing Machinery*, 1968, 15, pp. 8-36.

[Va99] Hal R. Varian, "Market Structure in the Network Age", in *Understanding the Digital Economy*, Department of Commerce, Washington, DC 1999.

[Wi93] Robert B. Wilson, *Nonlinear Pricing*, Oxford University Press, 1993

[Ya99] Yahoo Inc., http://www.yahoo.com

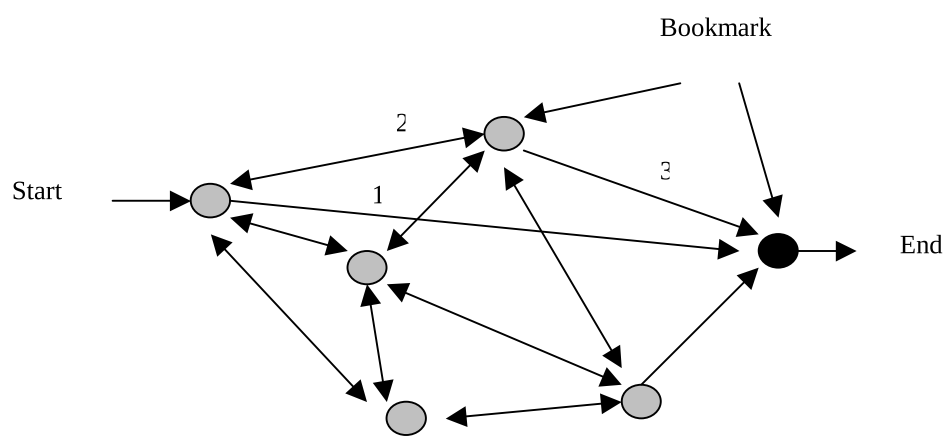


Figure 1: This graph illustrates a basic web site structure. Each circular node represents a web page, and each edge a directed connection between nodes. A user usually enters the site from the start (home) page. In this example, the target is the black node. The time a user takes to reach the target is a function of various things, including structure, and expertise. More interestingly for our purposes is that the number of clicks will also be a function of the amount of time a user is able to spend in reaching the target.

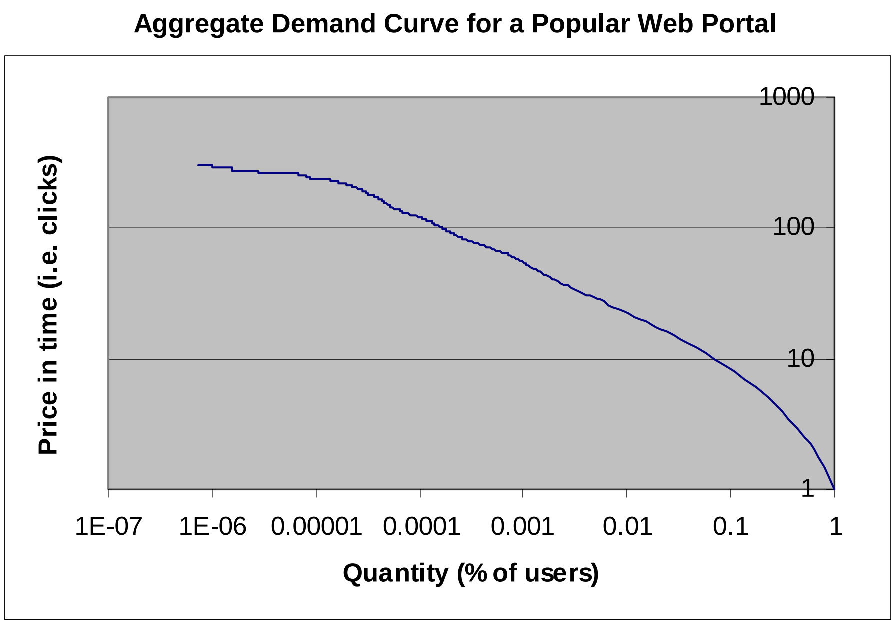
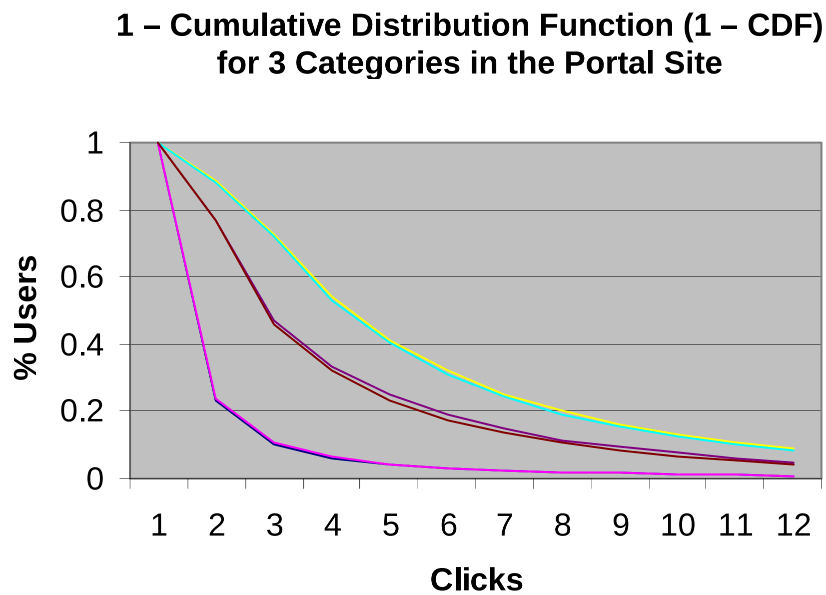


Figure 2: aggregate demand curve for a WWW portal.



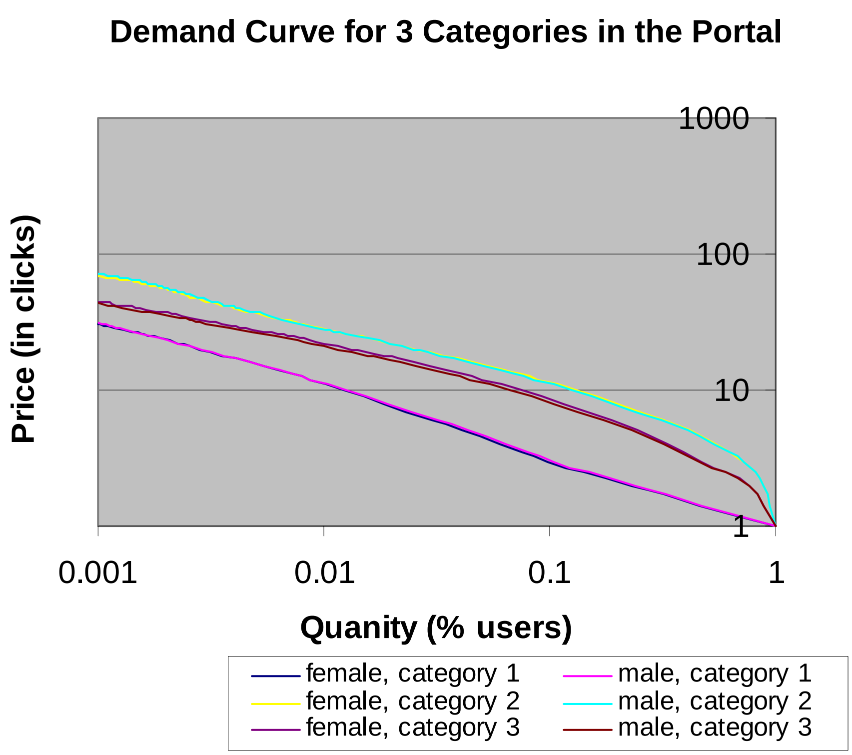
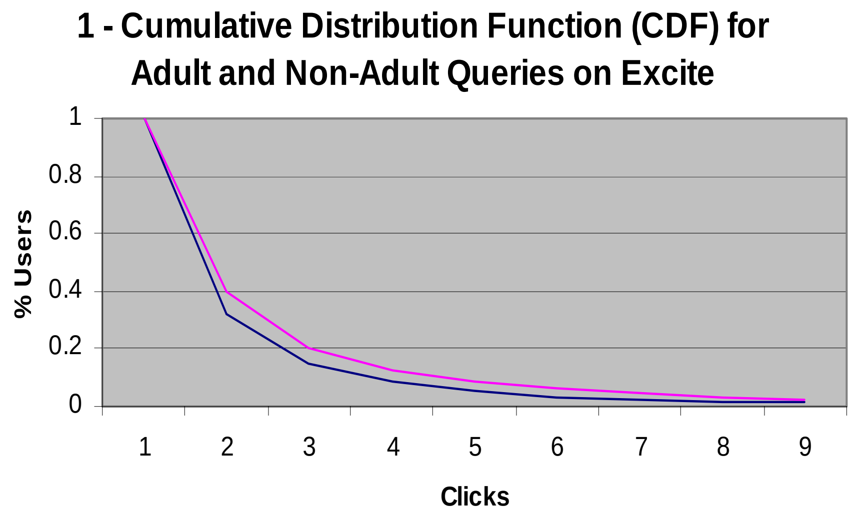


Figure 3a: One minus the cumulative distribution function (1-CDF) for three top level categories in the site's directory, partitioned by gender. Figure 3b represents the demand curves for the same paths.



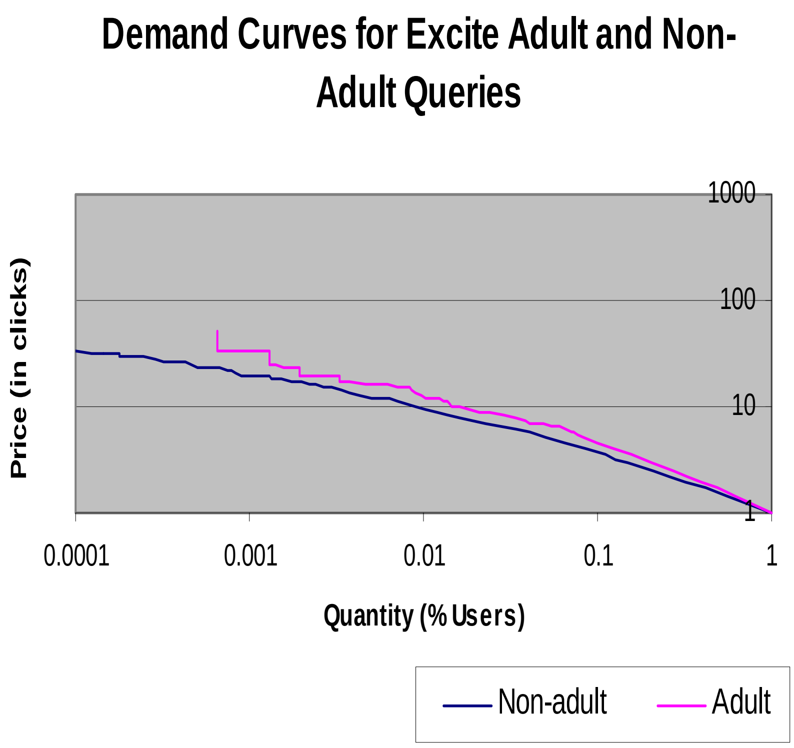


Figure 4a: One minus the cumulative distribution function (1-CDF) for two partitions of Excite queries (adult and non-adult). Figure 4b represents the demand curves for the same paths.

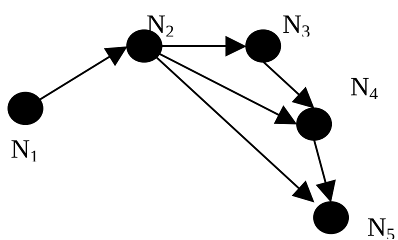


Figure 5a

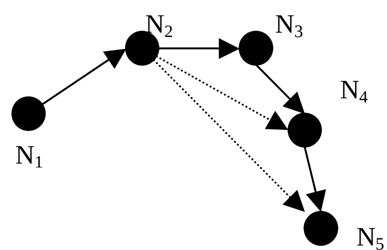


Figure 5b

Figure 5: (a) A directed sub-graph for a web site with various nodes (large circles) connected by various directed links. (b) Same as in (a) with the links from node 2 to 4 and 5 switched on or off.

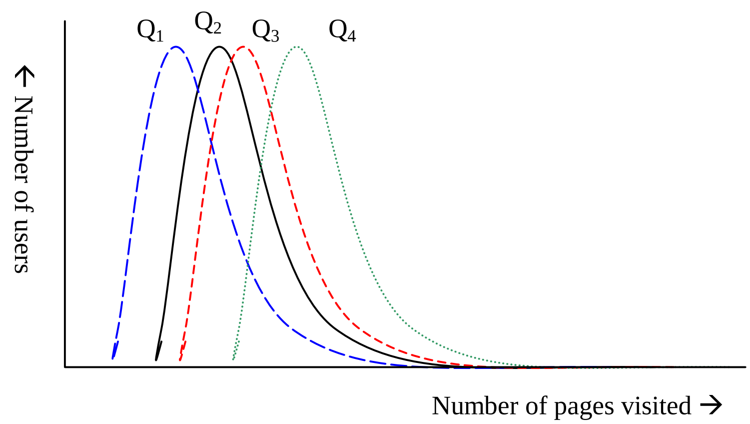


Figure 6: An approximation of user behavior, given by the law of surfing (the solid line). Each curve corresponds to a different domain of inquiry. A site that always performs with a constant quality of service does not take advantage of the willingness of various users to continue surfing.