Web Portals: Evidence and Analysis of Media Concentration

RAJIV M. DEWAN, MARSHALL L. FREIMER,
ABRAHAM SEIDMANN, AND JIE ZHANG

RAJIV M. DEWAN is an Associate Professor of Computers and Information Systems at the William E. Simon Graduate School of Business Administration at the University of Rochester, where he teaches courses on financial information systems and the economics of information management and e-commerce. His research interests include management of information systems in organizations, markets for information goods and services, and the economics of electronic commerce. His research has appeared in *Management Science, Journal of Management Information Systems, Communications of the ACM, IEEE Transactions on Computers, International Journal of Electronic Commerce, INFORMS Journal of Computing,* and other journals. He has won best paper awards at the Hawaii International Conference on System Sciences (in 1998 and 2002), and at the International Conference on Information Systems (in 1995). He recently cochaired the Workshop on Information Systems and Economics (WISE) in Seattle, Washington, in December 2003.

MARSHALL L. FREIMER is a Professor of Management Science and Computers and Information Systems at the William E. Simon Graduate School of Business Administration at the University of Rochester. He utilizes applied probability and decision sciences to analyze problems in information management, electronic commerce, marketing, and health care. His papers have appeared in engineering, management, economics, mathematics, and statistics journals.

ABRAHAM SEIDMANN is the Xerox Professor and Area Coordinator of Computers and Information Systems, Electronic Commerce, and Management Science at the William E. Simon Graduate School of Business Administration at the University of Rochester. Current research interests include information economics, business process reengineering, work flow systems, operations management, and strategic interorganizational information systems. Professor Seidmann is the author of numerous research articles. He is also an Associate or Area Editor for *IIE Transactions, International Journal of Flexible Manufacturing Systems, Production Planning and Controls, Journal of Intelligent Manufacturing, Journal of Management Information Systems, Management Science,* and *Production and Operations Management.* His current research and consulting activities include Business Process Reengineering, Strategic Manufacturing Systems, Information Economics, Healthcare Management, Stochastic and Performance Modeling. Professor Seidmann has consulted with many of the leading industrial and service corporations and presented research or executive seminars in four continents.

JIE (JENNIFER) ZHANG is an Assistant Professor of Information, Operations, and Technology Management in the College of Business Administration at the University of

Journal of Management Information Systems / Fall 2004, Vol. 21, No. 2, pp. 183–201.

0742–1222 / 2004 $9.50 + 0.00.

Toledo. She received her Ph.D. in Computer Information Systems in 2003. She employs analytical and empirical techniques to examine a number of issues in electronic commerce and management of information systems. Her research appears in *Journal of Management Information Systems* and numerous conference proceedings.

ABSTRACT: Although the Web has grown to several billion pages over the past few years, just a few of the Web sites get most of the visits. Such sites, called portals, attract visitors and advertisers and provide a lot of valuable content at no charge to the visitors. The portals attract a disproportionate amount of the Internet advertising dollars and have the ability to influence the success of new electronic commerce ventures. Using monthly audience data, we examine relative market shares of Web sites in search engines, travel, financial, news, and other categories. We find clear evidence of increasing disparity in page views with the top Web sites getting an increasing share of the market. Using economic modeling, we show that this disparity is a result of a development externality that exists in this industry: the sites that have more viewers get more revenues; this in turn allows them to develop more content and attract an even greater number of viewers.

KEY WORDS AND PHRASES: advertisement policy, electronic commerce, media concentration, page views, portals, value index, World Wide Web.

THE EXPONENTIAL GROWTH OF THE NUMBER of Web pages on the Internet presents its users with an unprecedented depth and breadth of information from thousands of online sources. Although there is some disagreement on the exact numbers, Google, in early 2004, claims to do its search over 4 billion pages [8]. With this number of pages online and the ease with which an individual user can visit any one, it is surprising that just a handful of sites garner a majority of visits from users. Some of these sites are visited by over 50 percent of all active users. Nielsen/NetRatings reports that 58 percent and 56 percent of all active browsers visited Yahoo and MSN, respectively, in January 2004 [13]. This trend is present not only among Web sites in general but also in segmented categories such as search, finance, travel, and others. Further, this lopsided viewership has only increased over time. In this paper, we present some evidence of growing disparity in audience measures and provide a theory to explain this phenomenon.

Over the past few years, several other sites have emerged as easy and organized gateways to the Web, or to a set of Web pages addressing specific topics. Yahoo, for example, appeals to many individual users because it provides structured information on many popular categories. Similarly, CNET appeals to users who are interested in information technology developments, investments, and products. Users find it easier to log on first to one of these sites (such as Yahoo or CNET) and from there to be hyperlinked to other pages of interest on the Internet. These sites, called *portals,* have become major gateways into the Internet. To attract traffic, portals provide valueadded services to the users. They continuously scan the World Wide Web for relevant and timely information, screen and prioritize the links, and provide a consistent interface to the ever-changing Web.

Portals attract a large portion of visitors.[1] For instance, 58 percent and 39 percent of all the users of the Web visit Yahoo.com (which is accessible by all Internet users) and the Google search engine portal, respectively [13]. Having so many visitors going through the same few sites attracts advertisers who try to bring users to their own Web sites. This phenomenon is persistent. Not only has this been found in a wide range of categories of Web sites, the disparity is *increasing* over time. This disparity is not just an artifact of immature Web technologies but seems to be a permanent feature of this online economy. This raises some interesting research questions: What is the role of portals in the Internet economy? Why do they arise? Which factors affect the growth of portals? And, are all portals created equal?

We address the issues raised above by analyzing the evolution of Web pages into portals using economics-based models. The models consider the information needs of a heterogeneous user population, the cost of providing quality content on a continuous basis, advertising, and cross-selling revenues. Economies of scale in production and operation have been posited as a cause of media concentration for traditional media such as newspapers [3]. We, instead, examine the impact of competition for traffic and advertisement as the main factors that drive formation of portals on the Web. Over time, as the competition heats up, the Web sites compete for Internet visitors by offering more and more services [5, 6]. Providing greater services is costly for the Web sites. This would eventually lead them to lose money unless they differentiate in quality. However, they can all survive in the marketplace by differentiating their offerings by quality and quantity. A few Web sites offer larger amounts of valuable content and advertising while others specialize in content. This lopsided equilibrium is not unlike that between convenience stores and pharmacies and large grocery superstores. The portals are the superstores of the Web.

Evidence of Growing Asymmetry in Page Views

Data Source and Description

WE OBTAINED PANEL DATA FROM the Nielsen/NetRatings "Audience Measurement" database, which is one of the leading providers of commercial Web traffic databases. The database includes audience measures for all Web properties that meet the "statistical cutoff" for that particular month.[2] We use the total number of monthly page views data as a proxy for Web traffic. The time frame is from February 1999 up to November 2003, totaling 52 time periods (data for November 2001 to April 2002 are missing). We used the classification of the Nielsen/NetRatings database to separate the Web sites into six categories: search and portal, travel, news and information, finance, entertainment, and family and lifestyle. Table 1 provides detailed descriptive statistics for the six categories in the data set.

Table 1. Descriptive Statistics

Categories	Sample size (number of sites)	Page views (millions/month) Mean	Std. dev.	Min	Max
Search and portal	177	292,239	1,132,660.3	202	12,093,977
Family and lifestyle	191	34,000	167,897.3	164	2,951,849
Entertainment	324	19,742	43,009.1	156	856,604
News and information	245	20,536	42,885.2	164	546,400
Finance	136	17,892	29,775.2	171	308,211
Travel	115	13,400	24,604.3	309	219,491

Empirical Evidence of Media Concentration—Existence of Portals

We first turn to the search engine category of portals. The sites in this category are among the most visited sites on the Web. We use a number of different measures to exhibit the disparity in viewership. We picked three measures that are commonly used in economics for measuring disparity in income, health-care access, and so on. These are:

1. *Quintiles:* We partition the set of Web sites into quintiles based on monthly page views. In other words, we rank the Web sites by the monthly page views, and then segment them into five groups such that each group has an equal number of Web sites. We calculate the percentage of page views for each quintile. This is shown in Figure 1 for months ranging from February 1999 to November 2003. Note that the inequality increases over time. The 35 most viewed sites, which make up the top quintile, take 76 percent to 98 percent of traffic in this category from February 1999 to November 2003. By November 2003, the remaining 142 sites only accounted for a total of 2 percent of the page views.
2. *The Lorenz curve:* This method was devised by Lorenz to display income disparity [12]. To show the disparity in a given month, we first rank the sites according to the page views in that month. Next, the cumulative percentage of page views is plotted against the percentage of Web sites in the order of the rank.[3] This curve would be a 45-degree straight line if there were no disparity. Any deviation of the curve from the 45-degree line reflects a disparity in page views for that month. This method of displaying disparity is reliable, and is independent of the mean or the total number of page views.[4] As we can see from Figure 1, the search category exhibits a large disparity in page views. Comparing the curves for different months, we note that the disparity increases with time.
3. *The Gini coefficient:* The Gini coefficient is a statistic of the Lorenz curve that describes the degree of asymmetry. The Gini coefficient is defined as the ratio of the area between the curve and the 45-degree straight line to the total area beneath the 45-degree line.[5] The value of this coefficient varies from zero to one. The value of zero implies that there is no asymmetry; that is, all Web sites

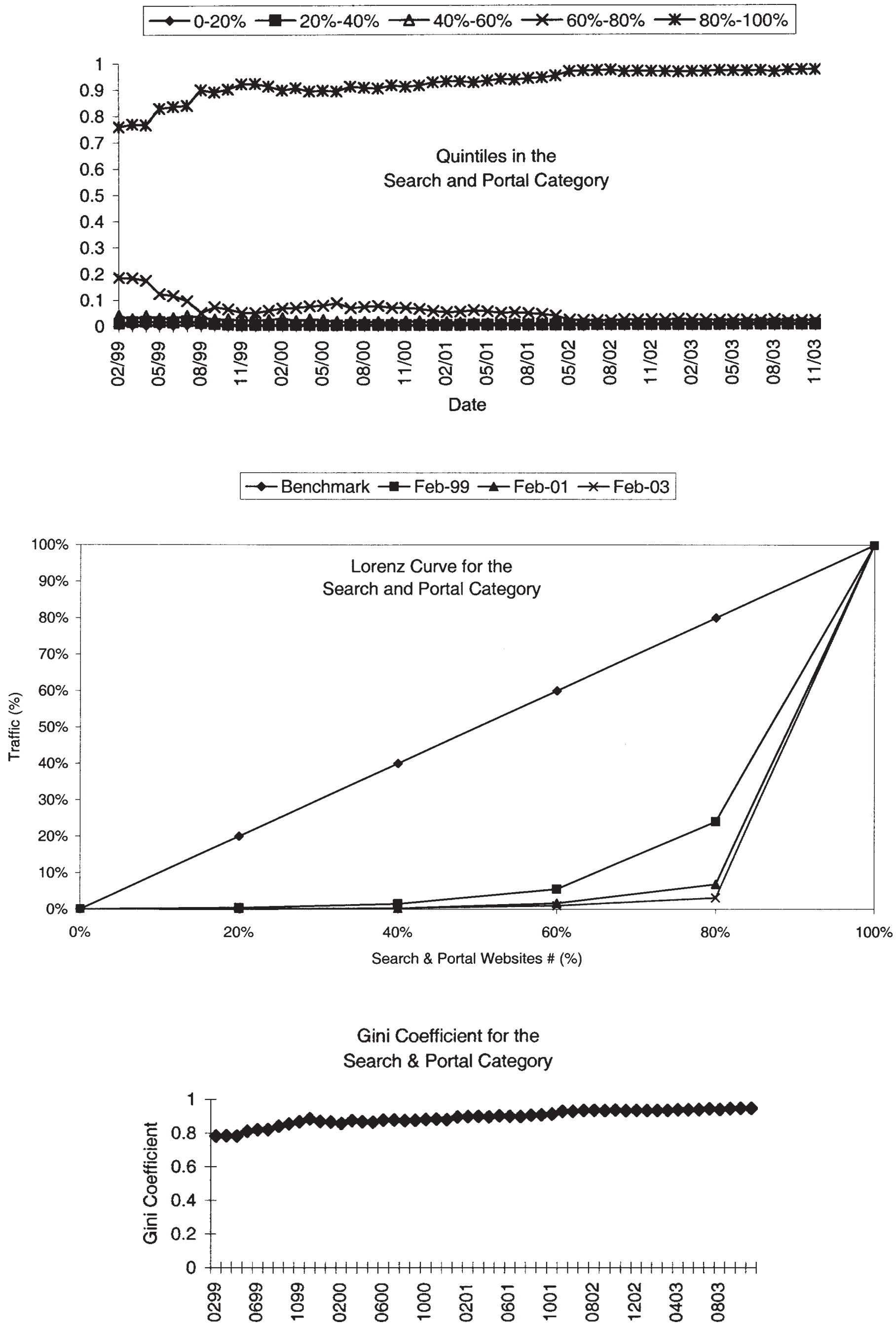


Figure 1. Media Concentration Among Search Portals as Indicated by the Quintiles, the Lorenz Curves, and the Gini Coefficient

have an equal number of page views. A value of one, on the other hand, indicates maximum asymmetry; that is, a single site gets all the page views.

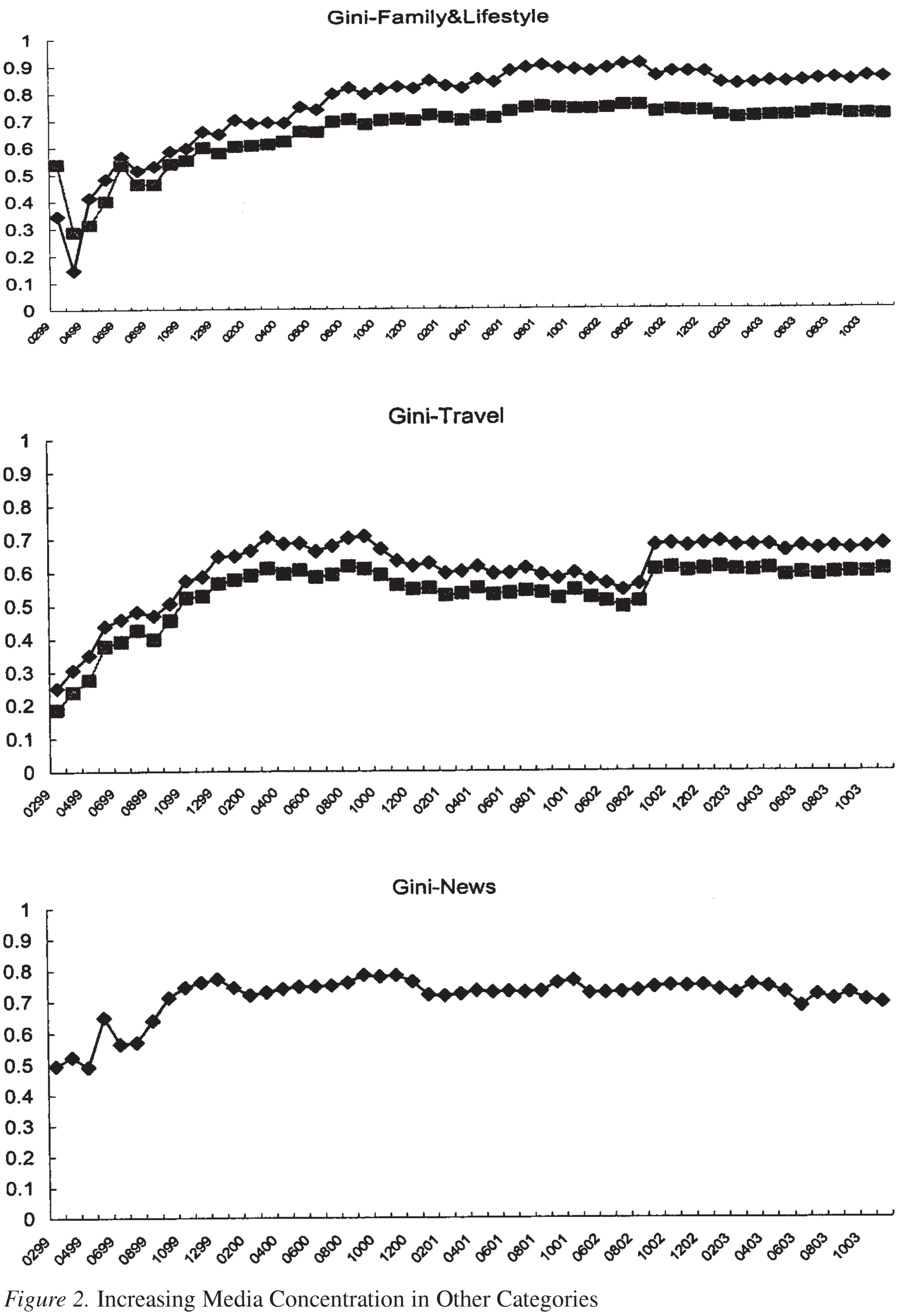
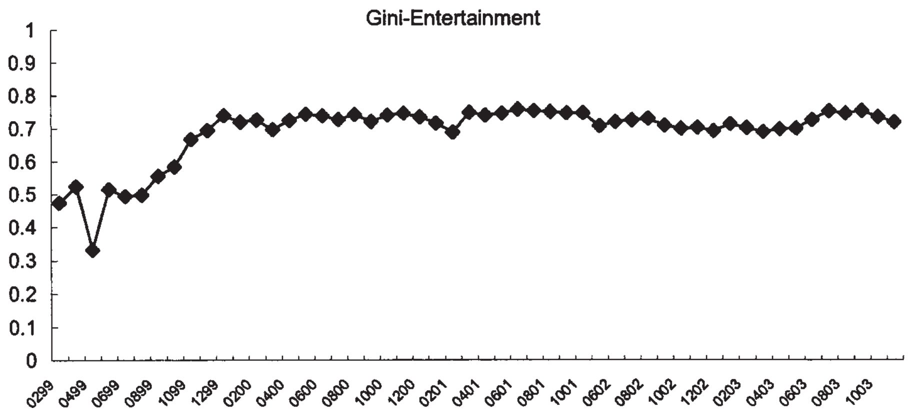


Figure 2. Increasing Media Concentration in Other Categories

The plot of the Gini coefficient over time clearly reveals the increasing disparity in page views between the top sites in the search category compared to others. The coefficient starts out at 0.78 and increases to 0.95 in 52 months.

This phenomenon of increasing disparity is present not just in the Search category, but in every category we examined. In Figure 2, we plot the Gini coefficient over



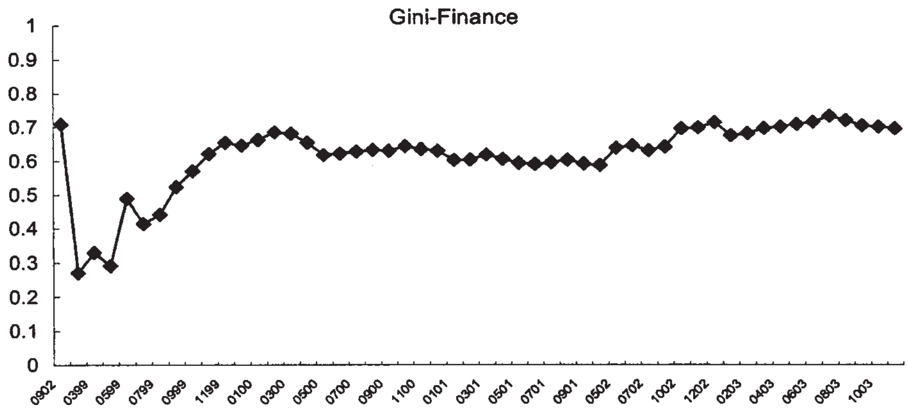


Figure 2. Continued

time for the family and lifestyle, travel, news, finance, and entertainment categories, respectively. Each category shows increasing disparity in viewership, with the top sites getting an increasingly disproportionate share of viewers.

These empirical observations raise a number of questions: What is the explanation for this growing disparity? How does it affect the profitability of the Web sites? In the next section, we address these questions using models of the Web site economy.

A Spatial Competition Model of the Internet Economy for Web Sites

IN THIS SECTION, WE EXPLAIN THE GROWING DISPARITY by exhibiting the indirect network externality that arises from investment in site content from advertising revenues.

Rather than start with a model in which we assume the existence of portals, Web sites that have many more visitors than others, we start with a Web of equals where all pages have the same opportunity to attract visitors. In later sections, we show that technological trends lead to portals. To start our analysis, we derive a symmetric equilibrium among Web pages in which the pages are identical in draw and scope although different in content. In this section, we model the basic economic forces that come about from the incentives and self-interest of all the entities in the Internet economy. In following sections, we will explore the symmetric equilibrium (no portals) and the asymmetric equilibrium in which one Web site gets many more visitors (portals).

There are three interested parties in the provision of Web pages for the public. These are:

- *Web surfers:* These are persons who are surfing the Web and looking for material of interest. Different "customers" like different content and value it differently. Each customer picks a Web page among the ones available to maximize his or her net value. The customers get the highest value from content that exactly matches their interests. Besides a closer match, a customer's value also increases with greater quantity and quality of valuable content. In contrast to content, advertisements on the page represent a cost for the customer as they have to wait for them to be downloaded and they add to the clutter on the screen.
- *Advertisers:* These are companies that want to get a message to the customers and are willing to pay for it. They contract with Web page owners to display their advertisements in return for compensation.
- *Web page owners:* Each Web page owner decides on the quality and quantity of content and the number of advertisements to put on its page. The content attracts surfers, but is costly to obtain and set up. A greater number of customers increases its advertising revenue. Each page owner makes his decisions while taking others' decisions into account so as to maximize his own profit.

The economy consisting of these three sets of players is modeled as a sequential game shown in Figure 3.

The Web site owners first pick the level of content on the Web site. We assume that the type and level of dynamic content is a longer-term decision than the number of advertisements to show on the Web page. The former gives rise to the brand identity of the Web page, whereas the latter is a tactical decision that may change in the short run. For instance, CNET, a popular technology-oriented Web site, has a system called DREAM that decides on which advertisements to place on the Web page based on user information. The customer responds to the content and advertisement policies by picking the page that results in a larger positive surplus. If no page results in a positive surplus, then the user decides not to pick a home page. We analyze this game by starting with the last decision, the decision by customers.

Usage Decision by Customers at Stage 3 of Game

Customers' tastes vary widely. For instance, some customers might prefer regularly to access a Web site with financial data while others may prefer sports. Web page owners take advantage of this while competing with each other. Rather than pick identical or similar choices for Web content, they tend to space themselves out and thus put themselves in mutually advantageous positions. This aspect of competition

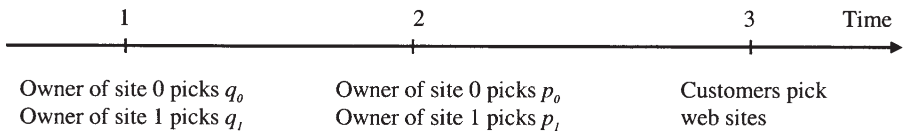


Figure 3. The Time Line of Decisions

is variously referred to as *monopolistic competition* [2] or *spatial competition* along a line segment [10] or the circumference of a circle [15]. Such models have been used to model search behavior of goods by customers in general markets [17] and the Internet [1]. Following the literature cited, we model the customers as having tastes that are arrayed along a circle of circumference 2. To simplify the analysis, we consider a situation in which there are two Web pages with content that are diametrically opposite on this circle. Continuing with the example earlier, one page, with content at 0 without loss of generality, may focus on financial content while another at 1 focuses on sports-related material. The value that a customer gets from a Web page depends on how close his or her interests are to the Web page, intensity of preference, and the cost of advertisements on the page:

$$\text{Customer's Utility} = \left(y - fx\right)q - dp,$$

where y = Net value of unit content that exactly matches the consumer's interest. The cost of time to download and other costs of using the Web site for valuable content, other than advertisements, are netted out of the value. Customers are heterogeneous in their valuation of content. We assume that y is uniformly distributed over the interval $[0, a]$.

x = Content difference between the page accessed and the customer's ideal, measured along the circle.

f = Unit fit cost reflecting the decrease in value from mismatch between customer's interests and the page.

q = Quantity of premium page content that is valued by the user but is costly for the page owner to keep updated and available to the customer.

p = Number of advertisements on the page.

d = Unit cost to customer of having an advertisement on the Web page. We assume that all the advertisements impose the same cost on the individual; that is, they take the same amount of extra time to download, add a similar amount of clutter to the screen and have the same low probability of being useful to the customer.

Note that we have the potential for both vertical and horizontal differentiation. The Web sites are vertically differentiated on the amount of content and horizontally on the type of content. The ratio a/f, which is the ratio of the maximum value of the Web (vertical differentiation) to the fit cost (horizontal differentiation), is a measure of the relative importance of the two forms of differentiation, with a higher value indicating greater vertical differentiation. For a given f, this measure also indicates increasing usefulness of the Web to the users. We call this ratio, *a*/*f*, the *value index.* As we shall see in following sections, the value index turns out to be critical in determining the profitability of the Web sites in different equilibria.

Continuing our analysis of customer reaction, the Web surfers pick a single page to use as their point of entry to the Web. They pick a page that gives them the largest utility. This page could be anywhere on the circle. Or the customer may not use any Web page on a regular basis. For a customer whose ideal is located at a distance x from the Web page at 0 and has a value y for the ideal Web page, the decision problem is

$$\text{Max}\begin{Bmatrix} 0, \\ (y - fx)q_0 - dp_0, \\ (y - f(1-x))q_1 - dp_1 \end{Bmatrix},$$

where q_0, p_0, q_1, and p_1 are the quantity of dynamic information and number of advertisements selected by Web pages at 0 and 1, respectively.

The Web Site Owners

At this stage, the Web page owners decide on the quantity of dynamic content (also called features) and the number of advertisements to place on the page. The dynamic content is costly to maintain and yet it attracts customers. The Web page owner pays for this with advertising revenue that depends on the number of customers who view his site. This is modeled as

$$9\ \text{CDRCI GRTQHK} = c(0\ \text{WODGTQHCWtfQOGTU})\,p - kq^2,$$

where c = Cost per thousand impressions (CPM) for advertisers. This is the rate at which advertisers pay the Web page owner for having their advertisement on his page. Each advertiser pays the Web page owner an amount that is CPM times the number of thousands of customers who view the advertisements. Although there are some other measures of advertising, CPM is the most common structure of advertisement contracts.[6] The CPM for different media range from $1 for outdoor billboards, $5 for national radio, $12 for prime-time Tt , $20 for generic search engine–type Web sites, $30 for specific information Web sites, to $35 for paid subscription magazines.

p = Number of different advertisers on the Web page. This is not necessarily the number of advertisers with whom the Web page owner has a contract but the number that is typically bundled on the page downloaded by the customer. The latter may be a smaller number because of selective targeting of advertisements based on search terms, surfer

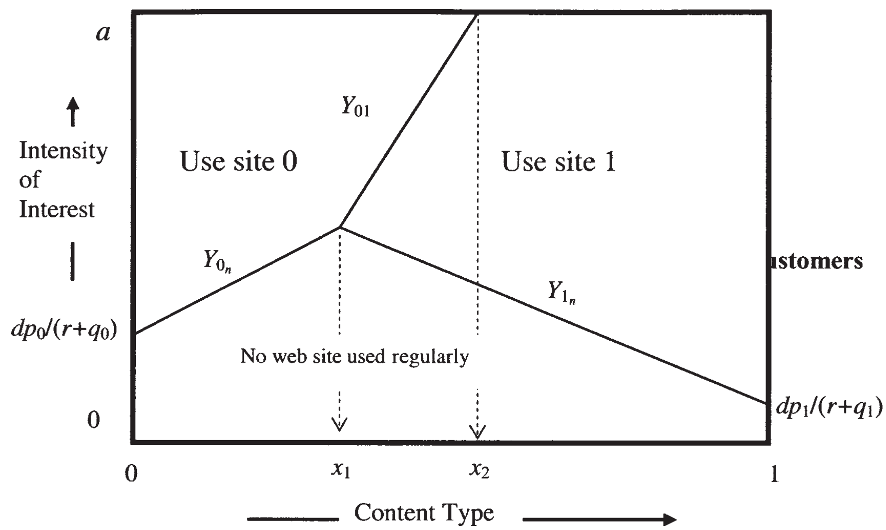


Figure 4. Web Site Usage by Different Customers

history (as gleaned from cookies or firms such as DoubleClick), and other demographic factors.

q = Number of dynamic content features incorporated on the Web page.

kq^2 = Cost of dynamic, updated content. The Web page owner may have a list of features of dynamic content that he could place on the page. The Web page owner places them on the Web page in increasing order of cost, with the cheaper ones being provided first. This makes the cost of dynamic content convex in the quantity of such content. To simplify the model, we have the simplest cost form, quadratic at the rate *k,* that satisfies this assumption.

We now turn to the decisions made by the Web site owner.

Consider the case when owners of Web pages at 0 and 1 have chosen q_0 and q_1 as quantity of dynamic content and p_0 and p_1 as the number of advertisements, respectively. Customers vary in their taste from 0 to 1 and the intensity of their interests, y, varies from 0 to a uniformly. The choice of the customers is shown in Figure 4 for interests from 0 to 1. The other half of the circle is symmetric.

Given the content and advertisement policy of each Web site, the users decide on which pages to use. A customer located at x and having value y will be indifferent between using page 0 and not if

$$\left(y - fx\right)q_0 - dp_0 = 0.$$

Solving for y provides $Y_{0n} = (dp_0/q_0) + fx$ as the locus of customers who are indifferent between Web site at 0 and choosing no site at all. Similarly, define Y_{1n} and Y_{01} as the loci of consumers who are indifferent between using page 1 and nothing, and between using page at 0 or 1. Further, let x_1 and x_2 be the content type at which $Y_{0n} = Y_{1n}$ and $Y_{01} = a$, respectively. These are marked on Figure 4. With these figured out, the profit of Web page 0 owner is

$$\pi_0\left(p_0, q_0, p_1, q_1\right) = p_0 c \frac{1}{a}\left(\int_0^{x_1}\left(a - Y_{0n}\right)dx + \int_{x_1}^{x_2}\left(a - Y_{01}\right)dx\right) - kq_0^2,$$

where the number of potential users of the Web pages is normalized to 1.

The profit for the owner of Web page 1, $\pi_1(p_0, q_0, p_1, q_1)$, is determined similarly. The diagrams and profit functions are somewhat different when $q_0 - q_1$ or a is large. In this case, the line y_{01} intersects the boundary at $x = 1$; that is, $x_2 = 1$. The profit functions for this case have been determined, but for sake of brevity are not shown here. Point $x_1 = 0$ for even larger a. This case is analyzed in the fifth section.

Working backward in time, the content quantity decisions are known by both the page owners when they simultaneously pick their advertisement policies. And, when picking the content quantity, each Web page owner knows the other's advertisement policy reaction, and takes that into account. Thus, there is a two-stage Nash game. The equilibrium in this game depends on the value index a/f.

Symmetric Equilibrium Among Web Sites

WE FIRST EXAMINE THE SYMMETRIC EQUILIBRIUM among Web pages where the Web page owners find it in their interest to pick similar content scope and advertising. A simple symmetric equilibrium is obtainable under certain circumstances. In this equilibrium, the two Web page owners make similar decisions about quantities of content and advertising. The equilibrium depends on the value index, a/f, which is a dimensionless ratio of the value of the Web to the fit cost. The market is not covered, that is, there are some categories for which no customer finds any of the sites worth visiting, when the value index is very small ($a/f < 1.5$). In this case, the Web sites are not competing directly. Setting this uninteresting case aside, a nontrivial and a competitive symmetric equilibrium is obtained for intermediate values of the value index. This is exhibited in the next theorem.

Theorem 1: For $a/f \geq 1.5$, a symmetric equilibrium between Web site owners with the following content and advertisement policies is obtained:

$$q^* = \frac{c}{32adk}\frac{16a^3 - 12a^2 f + 40af^2 + 148f^3 + \left(8a^2 - 41f^2 - 18af\right)\sqrt{4a^2 - 4af + 13f^2}}{2a - 13f + 5\sqrt{4a^2 - 4af + 13f^2}}$$

$$p^* = \frac{\left(2a + 3f - \sqrt{4a^2 - 4af + 13f^2}\right)}{4d}q^*.$$

The profit functions are continuous and locally concave. The proof of the theorem follows from standard first-order and second-order conditions for optimality that are

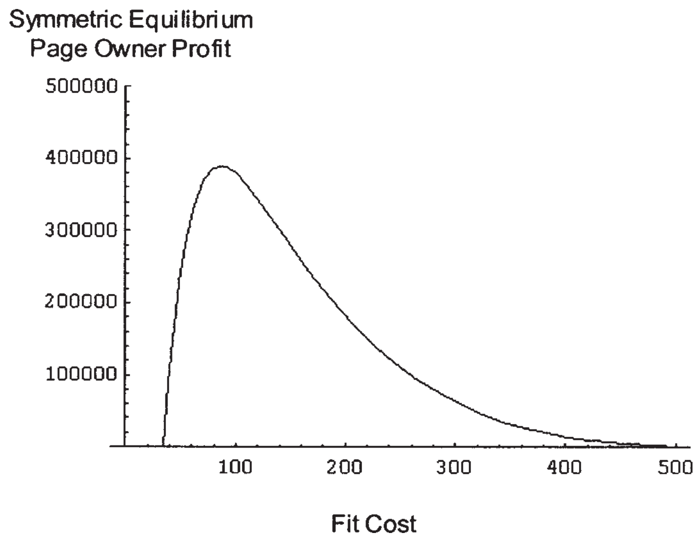


Figure 5. Page Owner Profit Versus Fit Cost for Symmetric Equilibrium

applied successively to advertisement and content policies, respectively. The proof is tedious but straightforward and is omitted here. The equilibrium obtained has many of the properties of other spatial equilibria [15]. For instance, the Web pages are competing for customers, but are separated by content type. This allows them to have "local monopolies" and have more advertisements than if they were more similar in content. This is illustrated in Figure 5.

The profit of a Web page owner for $a = 150$, $c = 30$, $d = 10$, $k = 1{,}000$, and $n = 1{,}000$ in a symmetric equilibrium is plotted versus f, the fit cost—the disutility to the user of the Web page from a complete mismatch in taste of page content. Note that the profit tends to zero for very small and very large fit costs. When the fit cost is very small, the pages are very similar and owners get into a very competitive situation where they try to attract customers by reducing ads. This reduces the profit for the Web page owners. The "spatial monopoly" enjoyed by the Web pages because of user taste differentiation breaks down into Bertrand competition and Web page owners do not make a profit. In contrast, when the fit cost gets large, the "spatial monopoly" gets strengthened, but the increasing fit cost reduces the value of any single Web page to users. Fewer and fewer users find the pages attractive and profit dwindles.

It is instructive to examine the symmetric equilibrium as value of material placed on the Web increases. The Web has come a long way since January 1993 when Marc Andreesen at the University of Illinois released Mosaic and about 50 Web sites provided pages. With billions of pages indexed by search engines such as Google [8], the Web has become an important source of information for business-to-business and consumer marketing, part of the strategy of product design for many firms and indeed the core of business process for some firms such as Amazon, Inc. Correspondingly, we posit that the Web sites are becoming increasingly valuable. In our model, we assume that a, the upper limit of distribution of value of the Web page, is increas-

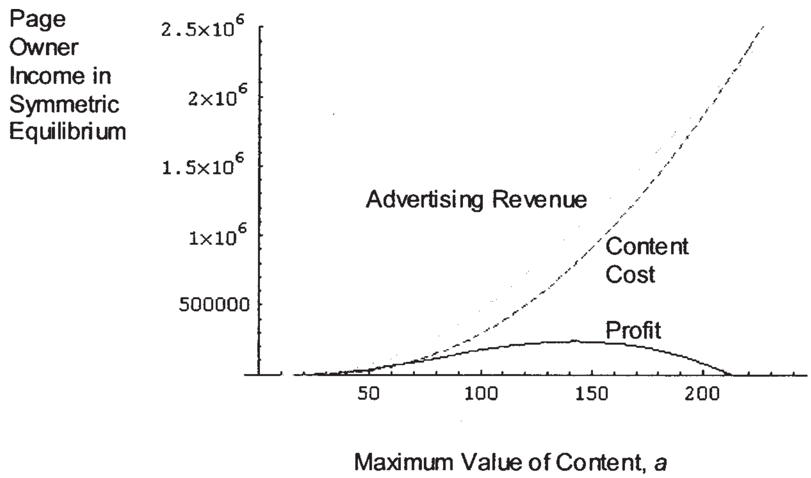


Figure 6. Impact of Increasing Value of the Web on the Site Owner Profit in Symmetric Equilibrium

ing over time. The impact of this trend on the symmetric equilibrium is examined next.

Theorem 2: The Web site owners do not make a positive profit in the symmetric equilibrium when the value index, a/f, is greater than 4.12311.

The proof is simply obtained by substituting the optimal q^* and p^* obtained in Theorem 1 into the profit functions. The theorem is illustrated below.

Figure 6 reveals that the profits of the Web page owners decrease as the Web content becomes more valuable. This is because the impact of fit cost is mitigated and the spatial competition begins to break down. The Web page owners then begin to compete more actively. The quantity of valuable content increases and this raises the cost of content to Web page owners. The Web page owners compensate by increasing the number of advertisements, but the increasing competition prevents them from increasing advertising revenue to match the cost of content. This reduces profit when a gets larger, so much so that *for even greater values of a/f there exists no symmetric equilibrium in which the Web page owners make a positive profit.*

So what does the future hold for Web page owners? Unequal size with one Web site having many more visits, more content, and advertising—portals.

Asymmetric Equilibrium: Formation of Portals

As competition for Web pages to get advertising dollars and viewers gets fierce, the Web page owners can do better by not copying each other's strategy but instead by differentiating on quantity of content and advertising. This is similar to the market for groceries: small convenience stores can coexist with larger supermarkets. In this equilibrium, the two Web sites differ in quantity of content and advertising: one Web

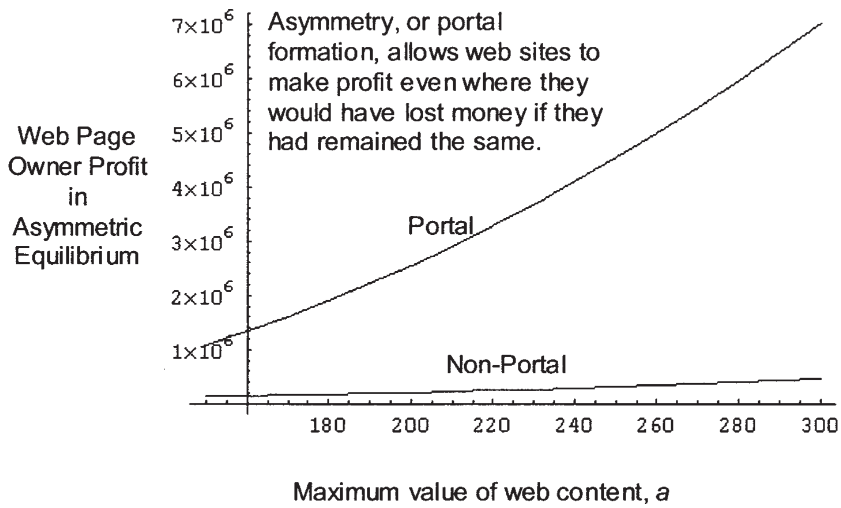


Figure 7. Profits in an Asymmetric Equilibrium

site has much more content, many more advertisers, and many viewers. This is the portal.

In Figure 7, we plot the Web page owner profit versus a for $f = 50$. Recall from Figure 6 that Web page owners do not make a positive profit for a larger than 220 for this fit cost. However, as Figure 7 indicates, both Web page owners make a positive profit in the asymmetric equilibrium in which one Web site is much larger than the other. The profit for both the portal and the nonportal Web site increases with a. This is the only equilibrium with positive profits for larger a.

In fact, as the Web becomes even more valuable, the asymmetric equilibrium continues to obtain with increasing profits for the Web sites. This asymmetric equilibrium is characterized in the next two theorems. In this equilibrium, each portal services some customers along the whole line. This would correspond to a negative x_1 in Figure 4. The limits of the integration in profit functions have to modified to reflect this geometry. This is developed in a fashion similar to the symmetric case described earlier and the details are omitted here.

Theorem 3: For value index $a/f \geq 5.2057$, an asymmetric equilibrium is obtained with:

$$q_0^* = \frac{c(2a-f)^2\left(4-3h+6h^2\right)}{2adk(4-h)^3}$$

$$q_1^* = hq_0^*$$

$$p_0^* = \frac{(2a-f)q_0^*\left(q_0^*-q_1^*\right)}{d\left(4q_0^*-q_1^*\right)}$$

$$p_1^* = \frac{(2a-f)q_1^*\left(q_0^* - q_1^*\right)}{2d\left(4q_0^* - q_1^*\right)},$$

where the constant

$$h = \frac{1}{2} + \left(\frac{-99+\sqrt{39727}}{676}\right)^{1/3} - \frac{17}{2\left(-594+6\sqrt{39727}\right)^{1/3}} \approx 0.1904.$$

The proof follows straightforward optimization principles and is omitted here.

Theorem 4: For value index, a/f, greater than or equal to 5.2057, the profits of the Web sites are positive and increasing in a. In particular, the profits are: $\pi_0^* = 0.00038156((c^2(2a-f)^4)/(a^2d^2k))$ *and* $\pi_1^* = 0.000047732((c^2(2a-f)^4)/(a^2d^2k))$.

Theorem 4 follows directly from the Theorem 3 by substitution and simplification. A somewhat different asymmetric equilibrium for intermediate value index—that is, for $4.12311 \leq a/f \leq 5.2057$, is shown in Dewan et al. [4].

Juxtaposition of Theorems 2 and 4 reveals the most important result of the paper: Symmetric equilibria cease to be profitable for large value index, $a/f > 4.12311$, whereas even the smaller Web site continues to make a positive profit in the asymmetric equilibrium. Hence, even from the small Web site's perspective, they would rather be in the asymmetric equilibrium than in the symmetric one for larger a.

From Theorem 4, we also note that the profits are decreasing in parameters d and k. The economic significance of this is explored next. Starting in 1991—with Tim Berners-Lee's basic tools for creation of documents in the then-new markup language, HTML—to today when a plethora of tools for Web content capture, setup, and serving are available, information technology for the Web has come a long way. This trend continues today with new technologies for large database interface, streaming technologies for real time exposition of video and audio, to XML—the new specification for markup language that has much greater customizability. These tools make it easier to serve up content to users. In our model, we measure the impact of Web content creation and publishing tools with parameter k, which is the coefficient for the quadratic term in the cost for content. One might think that a reduction in the cost of putting content out on the Web, k, might make it easier for the nonportal Web site to offer greater content and reduce the asymmetry. But this facile conclusion ignores the fact that the portal is similarly advantaged by the change in technology and the new equilibrium that emerges is even more asymmetric. Based on Theorem 4, we note that the profits for both the Web site owners increases as k decreases. The difference in the profits is also larger for smaller k. This is a significant result in that it points to the stability of portals in the face of changing technology for content creation and provision. As technology makes it increasingly cheap to place content on the Web and to provide new types of content, we expect this result to hold in the future.

We next turn to the client side and study the impact of improvements in access technology on portal formation. Many Web sites, such as ones allied with CNET and Accipiter, use staging technologies to load advertisements before content on the client machine. This makes the user wait for the advertisements to finish before any valuable content is displayed. Even if such technology is not used, advertisements and content compete for time to download through a user’s modem. This will also make the user wait, although not as much as with the advertisement technology described above.

Faster modems reduce the time taken to download the desired content and advertisements on a page. This reduces the cost of advertisements to the user, as the user has to wait less time for the page to load. This would reduce the net cost of each advertisement, d, to the user.

Advertisement presentation software, such as the ones described above, also aid the Web page owners to target advertisements to users based on search criteria, cookie-embedded surfing information, information from Web sites that collect surfing information such as DoubleClick.com, and other demographic information. This improves the value of the advertisement to the advertiser and the user alike. Searching on any of the popular Web search sites for items such as “Tt ” or “t CR” will provide an example of advertisement presentation software at work. The search results might be accompanied by advertisements relating to items searched for. This improvement in advertisement targeting reduces the net cost of an advertisement, d, to the user.

From Theorem 4, it is immediately apparent that not only do the profits for both the Web sites increase with decreasing d, the difference in the profits also increases.

Conclusions

MANY TECHNOLOGY TRENDS REINFORCE the formation of portals. In particular, improving technology for creating Web content enhances the formation of portals. New technologies, such as streaming technology and XML, make it easier to create and publish content. The portal and the nonportal Web sites take advantage of it with the portal site having the upper hand because of its larger budget for content. This provides a higher return for investments in content for portals that share the cost among a larger number of visitors.

Many Web sites get their income from advertising, which the users view as a cost. Web page managers carefully balance the income-producing and the customer-deterring ability of advertisements. Improving access technology, such as faster modems, reduces the downloading delay cost that advertisements place on the user. This allows the Web sites to sell more advertisements on the page. The impact is greater on the Web site with more advertisements—the portal. Consequently, improvements in access technology also accentuate the formation of portals.

Finally, our analysis explains the positive feedback loop created by the economics of the Web. Portals attract many more visitors, which in turn attracts more advertisers. This increases the advertisement revenue that the Web page owners can use to improve the quality of free information available to Web surfers. This in turn attracts more visitors and the cycle goes on. It is this basic mechanism that gives rise to portals. We show that in the long run, the portals cannot sustain a competition where each one of them competes with similar offerings. To survive, they have to differentiate from each other on quantity and quality of information.

This phenomenon is observed in many other realms of economic activity where a more featured product is available in the marketplace along with a simpler one. Coexistence of convenience stores and grocery superstores is an example. Portals are the emerging superstores of the Web.

Acknowledgments: The authors thank Robert Kauffman and anonymous reviewers for their useful comments and suggestions. The paper benefited from comments and suggestions from the participants of the Thirty-First Hawaii International Conference on System Sciences, Hawaii, 1998, where an earlier version of this work, [4], was presented.

Notes

1. Whereas we have focused on number of visitors to a Web site, Pennock et al. [14] examine media concentration in terms of number of links that point to a site. They find that the distribution of incoming links has a unimodal rather than a power law distribution, indicating that there is less of "concentration" effect on the incoming links. So, if there is media concentration, it is more in the realm of number of visitors rather than in number of links.
2. According to Nielsen/NetRating, a Web property meets the cutoff in any given month if a sufficient number of Nielsen/NetRatings' approximately 50,000 panel members visit the site such that extrapolation to the population of Web surfers as a whole can be reliably performed.
3. In other words, if a sample of n sites is ordered by the number of page views, with x_i being the number of page views for ith least-visited Web site, and $x_1 \leq x_2 \leq \ldots \leq x_n$, then the sample Lorenz curve is the polygon joining points (i, L_i) where $L_i = \Sigma^{i}_{j=1} x_j / \Sigma^{n}_{j=1} x_j$.
4. The Gini coefficient is dimensionless. It is not affected by scale, mean, or total number of page views. This invariance makes it a reliable measure of disparity.
5. Using the same notation as in the definition of the Lorenz curve above, the Gini coefficient, G, can simplified to: $G = \Sigma^{n}_{j=1}(2j - n - 1)x_j / n^2\mu$, where μ is the mean of the sample.
6. According to the *Internet Advertising Revenue Report for 2003,* published by the Internet Advertising Bureau, 63 percent of revenue contracts in 2003 used CPM. Performance-based contracts accounted for the rest. For the purposes of our model, even the latter can be accounted for using CPM as a revenue driver, as we are not focusing on advertisement performance and placement.

References

1. Bakos, Y. Reducing buyer search costs. *Management Science, 43,* 12 (December 1997), 1676–1692.
2. Chamberlain, E.H. *The Theory of Monopolistic Competition.* Cambridge, MA: Harvard University Press, 1931.
3. Dertouzos, J.N., and Trautman, W.B. Economic effects of media concentration: Estimates from a model of a newspaper firm. *Journal of Industrial Economics, 39,* 1 (September 1990), 1–14.

4. Dewan, R.M.; Freimer, M.; and Seidmann, A. Portal kombat: The fight to become the common entry point for consumers to the Web. R.H. Sprague, Jr. (ed.), *Proceedings of the Thirty-First Hawaii International Conference on System Sciences.* Los Alamitos, CA: IEEE Computer Society Press, 1998, <<page range or URL>>.

5. Dewan, R.M.; Freimer, M.; and Zhang, J. Managing Web sites for profitability: Balancing advertising and content. In R.H. Sprague, Jr. (ed.), *Proceedings of the Thirty-Fifth Hawaii International Conference on System Sciences.* Los Alamitos, CA: IEEE Computer Society Press, 2002 <<URL>>.

6. Dewan, R.M.; Freimer, M.; and Zhang, J. Management and valuation of advertisement-supported Web sites. *Journal of Management Information Systems, 19,* 3 (Winter 2002–3), 87–98.

7. Gini, C. Variabilitá e mutabilita, 1912, <<translate title>>. In E. Pizetti and T. Salvemini (eds.), *Memorie de metodologia statistica* <<translate title>>. Rome: Libreria Eredi Virgilio Veschi, 1955, <<page range>>.

8. Google. Web page www.google.com viewed on May 24, 2004.

9. Green, H.; Himelstein, L.; and Judge, P.C. Portal combat comes to the Net. *Business Week,* March 2, 1998. <<volume / issue / page range>>

10. Hotelling, H. Stability in competition. *Economic Journal, 39* <<issue>> (1929), 41–57.

11. Lawrence, S., and Giles, C.L. Searching the World Wide Web. *Science, 280* <<issue>> (April 3, 1998), 98–100.

12. Lorenz, M.O. Methods for measuring the concentration of wealth. *Journal of the American Statistical Association, 9* <<issue>> (1905), 209–219.

13. Nielsen/NetRatings. <<what is specifically being referred to here? / location of organization>>, 2004 (available at www.netratings.com).

14. Pennock, D.M.; Flake, G.W.; Lawrence, S.; Glover, E.J.; and Giles, C.L. Winners don't take all: Characterizing the competition for links on the Web. *Proceedings of the NAS,* <<spell out title / periodical?>> *99,* 8 (April 16, 2002), 5207–5211.

15. Salop, S.C. Monopolistic competition with outside goods. *Bell Journal of Economics, 10,* 1 (Spring 1979), 141–156.

16. Weber, T.E. Webs vastness foils even best search engines. *Wall Street Journal,* April 3, 1998, <<page>>.

17. Wolinsky, A. Product differentiation with imperfect information. *Review of Economic Studies, 51,* 1 (1984), 53–61.