

Portal Kombat: The Battle between Web Pages to become the Point of Entry to the World Wide Web

Rajiv Dewan,
Simon School of Business
Dewan@ssb.rochester.edu

Marshall Freimer,
Simon School of Business
freimer@ssb.rochester.edu

Abraham Seidmann
Simon School of Business
seidmannav@ssb.rochester.edu

Abstract

Over the past few years the web has grown to several hundred million pages, while just a few of them get most of the visits. Such sites, called portals, attract visitors, advertisers and provide lots 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. In this paper we show the evolution of portals from a web in which all pages are equally accessible. In addition, we analyze the nature of the competition between portals and show the impact of technological developments and added services on their economic viability.

1 Introduction

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. While there is some disagreement on the exact numbers, recent reports indicate that there are over 1.2 million web sites [*Internet Timeline*, January 1998], with 320 million URLs [6] for about 55 million users online in US [4]. It is expected that the number of URLs will increase to over 1 Billion by end of the year 2000 [IDC].

This explosive growth of URLs on the Internet also creates a major problem for the users who would like to get directly at the set of web pages mostly relevant to their needs and interests without having to go through an excessive number of irrelevant pages. Over time, the population of Internet users embraces a broader spectrum of users – many of whom have limited expertise with the use of sophisticated search engines –

but they have very specific information needs that drive them to the Internet. For instance, consider a user looking for medical information who may use a search engine such as *Alta Vista* or *Yahoo*. A simple key word search is likely to produce hundreds or thousands of URLs, most of them irrelevant or of questionable quality. Recent research has raised some concerns about the effectiveness of general search engines for retrieving domain specific information [6], [8]. Clearly, as millions of pages are added to the Web each year, finding a specific piece of information will only get tougher. Instead of using a general search engine, the user looking for medical information can use a specialized web site *Medscape.com*. This site specializes in providing free and easy access to numerous sites with clinically oriented information.

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. AOL, for example, appeals to many individual users because it provides structured information on many popular categories. Similarly CNET appeals to users that have interest in information technology developments, investments, and products. Users find it easier to log on first to one of these sites (such as AOL 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 value added 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. For instance, 43% and 13% of all the users of the Web visit *AOL.COM* (which is accessible by all Internet users) and the newly developed *Disney* portal, respectively [4]. Having so many visitors going through the same few pages attracts advertisers who try to bring users to their own web sites.

This new phenomenon 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 among web sites heats up, the web sites compete for Internet visitors by offering more and more services. Providing greater services is costly for the web sites. This would eventually lead them to lose money. 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 lop-sided equilibrium is not unlike that between convenience stores and pharmacies and large grocery superstores. The portals are the superstores of the web.

2 Model

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 and technology will 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*: Persons surfing the web and looking for material of interest. Different “customers” like different content and value it differently. Each customer picks a web page amongst the ones available to maximize his net value. The customers get the highest value from content that exactly

matches their interests. A closer match, greater quantity of valuable content, and greater quality of content, possibly obtained from more frequent updating, increase the value of a web page. In contrast, 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*: 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. We assume that the advertisers use the industry standard method of paying for ads in proportion to the number of viewers.
- *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. Greater numbers of customers increases its advertising revenue. Each page owner is making decisions while taking others’ decisions into account so as to maximize its own profit.

Customers’ tastes vary widely. For instance, some customers might prefer to regularly 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, is variously referred to as *monopolistic competition* [2] or *spatial competition* along a line segment [5] or the circumference of a circle [7]. Such models have been used to model search behavior of goods by customers in general markets [9], and the Internet [1]. Following these papers, 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 interests are to the web page, intensity of preference, and the cost of advertisements on the page:

$$\text{Customer's Utility} = (y - f x)(q + r) - d p$$

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.
- r Quantity of static page content valued by the user but does not add to operating expenses of the page owner.
- 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, i.e., they take the same amount of extra time to download, add similar amount of clutter to the screen and have the same low probability of being useful to the customer.

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. Note that, depending on the value of y , a customer may be willing to use a page of diametrically opposite interest if the page provides the largest utility. Or the customer may not use any web page on a regular basis. This is expressed as the following problem for the customer with value y and interest different by x from web page at 0:

$$\text{Max} \left\{ \begin{array}{l} 0, \\ (y - f x)(q_0 + r) - d p_0, \\ (y - f(1 - x))(q_1 + r) - d p_1 \end{array} \right\}$$

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

Web page owners decide on the quantity of dynamic content (also called features) and the number of advertisements to place on the page. We assume that any space left over on the page is filled with static content that is valued a constant amount $r(y - f x)$ by the users. The dynamic content is costly to maintain and yet it attracts customers. The web page owner finances this with advertising revenue that depends on the number of customers who view his site. This is modeled as:

$$\text{Web page profit} = c (\text{Number of Customers}) p - k q^2$$

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. While there are some other measures of advertising, CPM is the most common structure of advertisement contracts. The CPM for different media range from \$1 for outdoor billboards, \$5 for national radio, \$12 for prime time TV, \$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 selecting targeting of advertisements based on search terms, surfer 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 typically has a list of features of dynamic content that he could place on the page. He places then 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.

Web page owners make two decisions:

- How much costly dynamic content, q , to put on the page?
- How many advertisements, p , to put on the page?

To simplify the analysis, we assume that the web page is large enough and any remaining space is filled with static content. We next determine the profit of the web page owner as a function of decision made by all the participants.

Consider the case when owners of web page 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 1 for interests from 0 to 1. The other half of the circle is symmetric.

Y_{0n}, Y_{1n} and Y_{01} are the locus of consumers who are indifferent between using page 0 or nothing, using

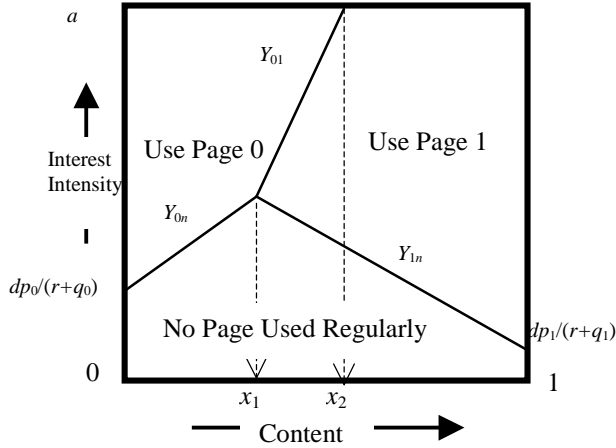


Figure 1: Web page usage by different customers

page 1 or nothing or between the two web pages. To wit, for the customer to be indifferent between getting page 0 and not, it must be the case that:

$$(y - f \cdot x)(q_0 + r) - d p_0 = 0.$$

Solving for y , provides $Y_{0n} = \frac{d p_0}{q_0 + r} + f \cdot x$. Other loci,

Y_{1n} and Y_{01} are obtained similarly. 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 1. With these figured out, the profit of web page 0 owner is:

$$\pi_0(p_0, q_0, p_1, q_1) =$$

$$p_0 c \frac{n}{a} \left(\int_0^{x_1} (a - Y_{0n}) dx + \int_{x_1}^{x_2} (a - Y_{01}) dx \right) - k q_0^2$$

where n is the total number of potential users of the web pages.

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. These have been determined, but for sake of brevity are not shown here.

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 while 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. With these assumptions, the decisions are made in two stages. This is shown in the timeline in Figure 2.

Working backwards in time, the content quantity decisions are known by both the page owners when they pick their advertisement policies. Hence for any (q_0, q_1) , the page owners pick the best number of advertisers, here onwards:

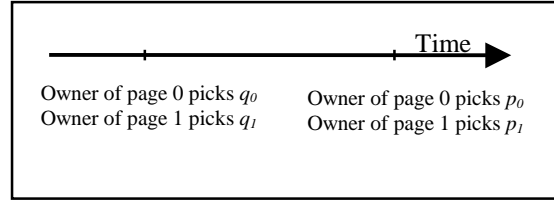


Figure 2: Time line of decisions

$$p_0^*(q_0, q_1) \in \text{Argmax}_p \pi_0(p, q_0, p_1^*, q_1) \text{ and}$$

$$p_1^*(q_0, q_1) \in \text{Argmax}_p \pi_1(p_0^*, q_0, p, q_1).$$

When picking the content quantity, each web page owner knows the other's advertisement policy reaction, and takes that into account. Thus, the two stage Nash equilibrium is described by the advertisement policy reactions described above, and the quantity decisions q_0^* , q_1^* , such that:

$$q_0^* \in \text{Argmax}_q \pi_0(p_0^*(q, q_1^*), q, p_1^*(q, q_1^*), q_1^*) \text{ and}$$

$$q_1^* \in \text{Argmax}_q \pi_1(p_0^*(q_0^*, q), q_0^*, p_1^*(q_0^*, q), q)$$

These equilibria were obtained analytically and numerically. These are described in the next section.

3 Equilibria among web pages

3.1 Symmetric Equilibrium

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 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 aspect of spatial competition is well known [7] and is illustrated in the figure below.

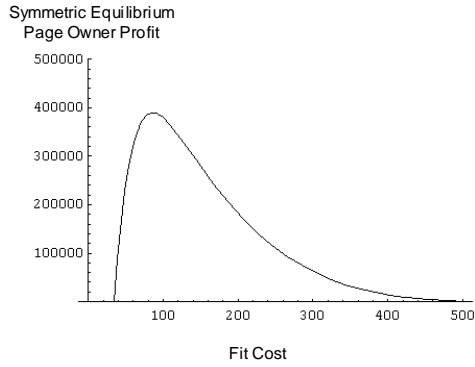


Figure 3: Page owner profit versus fit cost for symmetric equilibrium

The profit of a web page owner for $a = 150$, $c = 30$, $d = 10$, $k = 1,000$, $n = 1,000$ and $r = 1$, 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 Andreessen at University of Illinois released Mosaic and about 50 web sites provided pages. An estimated 320 million web pages were online in April 1998 [6]. Any one who has used a web search mechanism such as *Alta Vista* knows that not all web pages are equally or highly valuable in all circumstances. However, the web has become an important source of information for business to business and consumer marketing, part of strategy of product design for many firms and indeed the core of business process for some firms such as Amazon, Inc. We posit that the increasing proliferation of web pages in daily commerce has resulted in increasing value in some circumstances and that this will continue to increase. In our model we assume that the value of a web page feature that matches the taste of the user varies uniformly from 0 to a . We posit that a , the upper limit of distribution, is increasing over time as the web

proliferates and becomes a vital part of many businesses. The impact of this trend on the symmetric equilibrium is examined next.

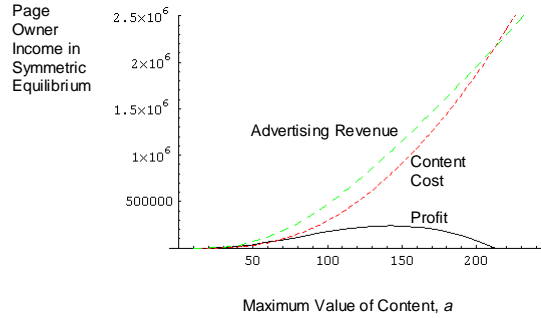


Figure 4: Impact of increasing value of the web on the page owner profit in symmetric equilibrium

Figure 4 reveals an unexpected result that the profit 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 value of a there exists no symmetric equilibrium in which the web page owners make a positive profit.*

From another perspective, the joint impact of web page value and fit cost is explicated in Figure 5.

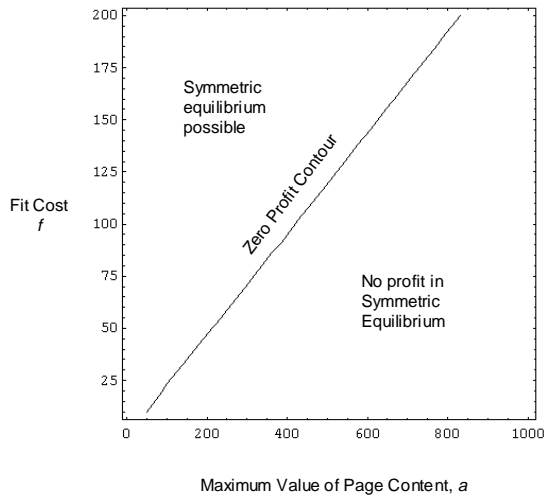


Figure 5: Zero Profit Contour in Symmetric Equilibrium

The zero contour for web page owner over maximum web page value, a , and fit cost, f , is plotted in the figure. For any fit cost, the web page owners make a positive profit in a symmetric equilibrium for smaller a . However, as a gets larger, the web page owners do not make any profit because of increasing competition. For instance, for $f = 50$, the web page owner makes no profit for a larger than 220. 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.

3.2 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 super markets. In this equilibrium, the two web sites differ in quantity of content and advertising: one web site has much more content, many more advertisers and many viewers. Such a web site is called a *portal*.

In the next figure we plot the web page owner profit versus a for $f = 50$. Recall from the previous figure that web page owners do not make a positive profit for a larger than 220 for this fit cost. However, as the following figure 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 non-portal web site increase with a . This is the only equilibrium possible for larger a .

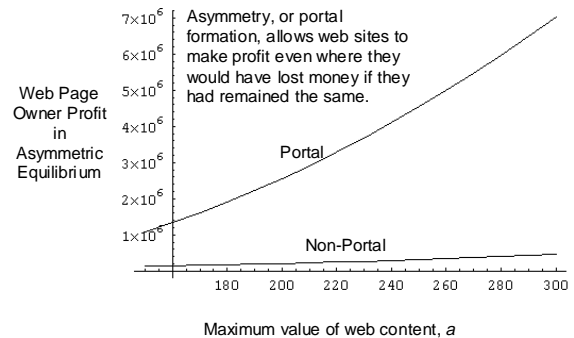


Figure 6: Profits in an asymmetric equilibrium

In the next section we explore the asymmetric equilibrium to assess the impact of information technology on Portal formation.

4 Impact of Technology on Portal Formation

4.1 Tools for Content Placement

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, k , might make it easier for the non-portal 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. This is illustrated in the Figure 7.

The profit for the web page owners in plotted against a , the maximum value of content, for different values of k , the cost coefficient. The solid lines are the profit for the

portal and non-portal web site owners for $k = 1,000$ and the dashed lines are for $k = 500$. Note that, as expected, the profits for both the owners are larger for lower cost coefficient. However, the asymmetry is larger for smaller k .

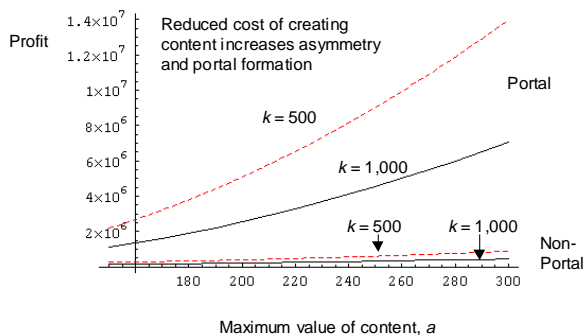


Figure 7: Impact of content creation technology on portal formation

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.

4.2 Impact of access technologies on portal formation

We next turn to the client side and study the impact of improvements in access technology on portal formation. Many web sites that use technologies for advertisement presentation and selection by companies such as CNET, Net Gravity, Accipiter and Bellcore, load advertisements before they load 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 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*, 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 "TV" or "VCR" 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.

As we see in Figure 8, reduced cost of advertisement to the user also enhances asymmetry and portal formation.

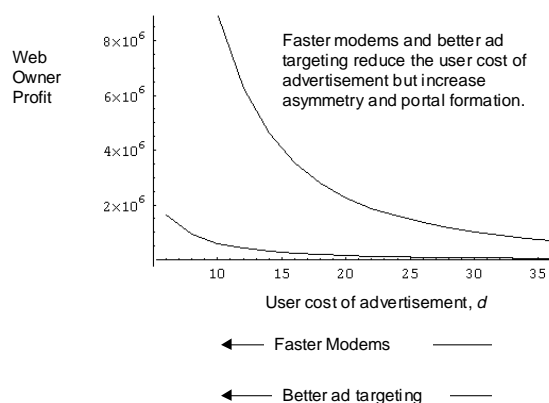


Figure 8: Impact of faster modems and better targeting of advertisements

5 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 non-portal 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 which 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.

6 References

1. Bakos, Y., “Reducing Buyer Search Costs,” *Management Science*, 1676 – 1692, Dec. 1997.
2. Chamberlain, E.H., *The theory of Monopolistic Competition*, 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*, pp 1-14, September 1990.
4. Green, H., Himelstein, L., and Judge, P.C., “Portal Combat comes to the net,” *Business Week*, March 2, 1998.
5. Hotelling, H., “Stability in Competition,” *The Economic Journal*, 39, 41-57, 1929.
6. Lawrence, S., and Giles, C. L., “Searching the World Wide Web,” *Science*, April 3, 1998, pp. 98 – 100.
7. Salop, S.C., “Monopolistic competition with outside goods,” *Bell Journal of Economics*, 141 – 156, 1979.
8. Weber, T.E., “Webs Vastness foils even best search engines,” *The Wall Street Journal*, April 3, 1998.
9. Wolinsky, A., “Product Differentiation with Imperfect Information,” *Review of Economic Studies*, pp 53 – 61, 1984.