Managing web sites for profitability: Balancing content and advertising

Rajiv Dewan, Marshall Freimer, Jie Zhang
{dewan, freimer, zhangji}@simon.rochester.edu

Abstract— **Balancing the amount of advertising and content on a web page affects the profitability of the web site and its attractiveness to potential visitors. This tradeoff is modelled as a control problem for a web site manager who is maximizing the net present value of cash flows by controlling the amount of advertising and content displayed on the web site over its life. The model is calibrated and justified using web site advertisement and audience data. A new measure of web site traffic is developed and it is statistically significant in explaining the market value of firms that own advertising supported web sites.**

I. Introduction

As the initial burst of enthusiasm for electronic commerce (ecommerce) and web related services wanes, profitability, if not profit, is the yardstick used to measure the success of ecommerce ventures. The primary means of managing profitability in most commerce settings is by adjusting the price of goods and services delivered. This tool is not available to most content ecommerce ventures as they do not charge for all or most of their content. While a few web sites, such as *Financial Times* and *Brittanica* have begun to charge for some content, the majority of content on the web is free. Most of these content web sites are supported by advertisements.

An advertisement impression, a click through, or an eventual purchase of an advertiser's offering may generate revenue for the web site that shows the advertisement. Even though prominent web sites are attempting to expand the revenue stream, advertisements remain the primary source of revenue. Consider the Yahoo family of sites. Apart from advertisements, it has created a number of new ways of charging for services. It offers annual membership for premium services at an annual fee, charges for listing in auctions, and for a fee and a transaction charge it provides a platform for over 13,000 retailers. All of these revenues totaled $1.1 Billion in the fiscal year 2000. Of this total, 90%, or $1.0 billion was from advertisements [4].

Date: December 13, 2001. We thank Nielsen Netratings for providing us audience data at a reduced rate.

Advertisements are not viewed benignly by site visitors. Advertisements clutter up the page, reduce the amount of space available for user valued content, grab attention using garish colors or animation, and in general reduce the usability of the web site. Further, they add to the time it takes to download the page. Some sites such as CNET stage the downloading of page resources to make sure that advertisements are downloaded before other content. In any case, a complex formatted page may not fully render in readable form until all the content is downloaded. This delay in page rendering imposes delay costs on users. While the visitor may find a particular advertisement of interest, advertisements, more likely than not, impose a cost on the visitor and reduce the attractiveness of the web site. The users respond to additional advertisements by visiting the site less often and/or visiting competing sites. In this sense the number of advertisements is analogous to price. This is examined further in the next section.

Setting the advertising policy for a site presents the web site owner with a tradeoff between revenue per customer and the number of customers who choose to visit the site. This decision directly affects the profitability of the site in each period. The number of visitors in any one period depends on the number of visitors in the prior period with some gain from growth of the number of users and loss from attrition. Consequently, the advertising policy cannot be made independently in each period but has to be managed over the life time of the site. We model this problem as an infinite horizon planning model with discounting. The objective is to maximize the market value of the firm which is the net present value or discounted sum of advertisement revenues and expenditures for site content. In section 3 we describe the model in greater detail.

Finally, we close the loop between the model and reality by a multi-step estimation procedure. Using time series data on visitors to a web site and results of the model, we estimate the number of visitors in the future. This is used to estimate the market value of the site. The relationship between the estimated market value and the current market value obtained from market data is examined to validate the model.

"Best Paper" in the Internet and Digital Economy Track at Hawaii International Conference on Systems Sciences 2002.

II. ARE ADVERTISEMENTS LIKE PRICE CHARGED?

We would like to make the case that the number of advertisements on a advertisement supported web site acts like price in other markets. Advertisements, on an ex-ante basis, are a cost to the users. They clutter up the page, reduce the amount of valuable content visible and increase download time. Advertisements provide revenue to the web site owner on a per impression basis. To make this analogy complete, we conducted an empirical examination to see if the number of advertisements is like price from an economics perspective. A well known fact about markets is that the price for goods is lower in more competitive markets. Does this hold for web advertisements?

To examine whether the number of advertisements is like price from an economics perspective, we looked at newspaper sites. *American Journalism Review,* a trade press publication for the journalism industry, classifies general news dailies as national or regional on a number of factors [1]. We examined all the national newspaper sites and the top regional newspaper sites listed at the site. Since the advertisements come in many different shapes, we used the percentage of screen devoted to advertisements as a measure of amount of advertisements. We used the median screen size of 800x600 to evaluate the pages. This data is shown in Table I.

We expect the market for viewers for web sites of national news dailies to be more competitive than that of regional news dailies. So we hypothesize:

H_a The average advertisement level at web sites of national dailies is is less than that at web sites of regional dailies.

H_0 There is no difference in mean amount of advertisements on web sites of national and regional dailies.

We tested these hypothesis on the data above. On web sites of national dailies, 4.7% of the screen was devoted to advertisements on average. This average for regional dailies was 9.6%.

To explore the differences further, define a dummy variable:

$$x_i = \begin{cases} 1, & \text{if site } i \text{ is regional;} \\ 0, & \text{otherwise.} \end{cases}$$

We regressed the percentage of advertisements against the dummy variable x_i. The results are shown in Table II. Note that the one-tail probability of the slope of the dummy variable being zero is 0.007. Hence, with great confidence we can conclude that web sites of regional dailies have a *greater* percentage of advertisements.

III. MODELING THE WEB SITE MANAGEMENT PROBLEM

Consider a web site manager who is managing an advertising supported web site that offers content of value to visitors. There are many design factors that affect the attractiveness of the page to a potential viewer. More well organized, frequently updated content attracts more viewers but is also expensive to acquire, format and place on the web on an ongoing basis. The content cost is modelled as kq_t^2 where q_t is the amount of content in period t and k is a constant[1].

In contrast, advertising reduces the attractiveness of the page — it adds clutter, increases download time and leaves less room on the page for the content that attracts a visitor. The web site owner, however, gets revenue from advertisements. More impressions (an instance of an advertisement viewed by a visitor) result in greater revenue. The relationship between impressions and revenue depends on the type of contract between the advertiser and the web site. The simplest kinds of contracts are written on "cost per thousand impressions" or CPM basis. More complex contracts may include click-throughs or actual sale performance. Even these can be simplified to a per impression rate. In this model we assume that the revenue in period t is cp_tu_t where c is the revenue per impression, p_t and u_t are the number of advertisements shown to each user and the number of visitors in period t, respectively.

Putting the revenue together, the web site owner's profit in period t is $cp_tu_t - kq_t^2$. In each period the manager has the option of varying the amount of content and advertisements. These are the variables he controls. We take the objective to be to maximize the net present value of the firm as this maximizes shareholder wealth. Firms may have operated with different objectives in the short run, but economics and finance dictate that the long term objective has to be the maximization of net present value of future cash flow streams. Let β be the rate of continuous discounting. The manager's objective function is to:

$$\max_{p_t, q_t} \int_0^\infty (cp_tu_t - kq_t^2)e^{-\beta t}dt \tag{1}$$

The formulation of the problem is not yet complete for we have not considered the effect of the manager's decisions on the potential viewers. Let u_t denote the number of viewers in time t. Starting at time 0 with u_0 viewers, we want to estimate the changes in viewership from period to period. For this purpose, we formulate an economy growth type model in which we relate the change in population to web site characteristics.

[1]This increasing marginal cost may reflect content choices by the manager — choosing to put the cheaper content on the web first.

National			
Christian Science Monitor	5.25%	Los Angeles Times	2.93%
New York Times	4.43%	Nando Times	7.35%
USA Today	2.93%	Washington Post	5.85%
Washington Times	4.50%		
Top Regional			
Boston Globe	18.03%	SF Chronicle	4.10%
Chicago Tribune	12.70%	Seattle Times	0.00%
Dallas Morning News	12.51%	Detroit Free Press	5.85%
Chicago Sun Times	15.29%	Houston Chronicle	2.25%
New York Post	12.35%	Star Tribune	6.88%
New Orleans T. Picayune	10.80%	Arizona Daily Star	5.85%
Miami Herald	8.80%	The Plain Dealer	12.85%
The Kansas City Star	15.8%	Denver Post	7.05%
The News & Observer	5.63%	Inland Empire	16.05%
Salt Lake Tribune	5.80%	Fort Worth Star Tele.	9.55%
Pittsburgh Post Gazette	6.18%	Orlando Sentinel	12.90%
Sun Sentinel	13.68%		

TABLE I
PERCENTAGE OF 800X600 SCREEN DEVOTED TO ADVERTISEMENTS

	Coeff.	**t-Stat**	**One Tailed P-value**
Intercept	0.0475	2.902	0.0037
Slope of x_i	0.0486	2.598	0.0074

TABLE II
REGRESSION ANALYSIS OF ADVERTISING

Let $\dot{u}_t$ be the change in number of visitors in period t. If the web site is perfectly sticky, and if we ignore the effect of advertising and content for the moment, then the only growth in viewership will come from growth of the Internet itself. Let a represent this rate. The web site, however, is not likely to be perfectly sticky. Let fu_t be the attrition in viewership. Putting these effects together, we get $\dot{u}_t = a - fu_t$.

The effect of content and advertising on viewership depends on the market conditions for web based content — the degree of differentiation of content, the level of competition, the disutility to consumers from advertisements, and other related factors. Overall, the viewership increases with the amount of content that is valued by users and decreases with the amount of advertising. Putting all of these factors together, we posit that:

$$\dot{u}_t = a - bp_t + eq_t - fu_t \tag{2}$$

The top left portion of the rendering of the web page on a viewers' screen is is the most important part as most users view this part. Let p_t and q_t represent the allocation of this valuable space to advertising and content, respectively. Normalizing, we get:

$$p_t + q_t \leq 1 \tag{3}$$

Combining (3) together with the objective (1) and constraint (2) we get the following control problem for the manager:

$$\max_{p_t,q_t} \quad \int_0^\infty (cp_t u_t - kq_t^2)e^{-\beta t}dt$$

Subject to:

$$\begin{aligned}\dot{u}_t &= a - bp_t + eq_t - fu_t \\ p_t + q_t &\leq 1 \\ p_t, q_t &\geq 0,\ u_0 \text{ given}\end{aligned}$$

This is a control problem with p_t and q_t as controls and u_t as the state variable. In the rest of the paper we consider the case when the constraint (3) is binding.[2]. Set $p_t = 1 - q_t$.

The *Hamiltonian* of the control problem is:

$$H(u_t, q_t, \lambda_t) = e^{-\beta t}\left(c(1-q_t)u_t - kq_t^2\right) + \lambda_t\ (a - b + bq_t + eq_t - fu_t)$$

The optimality condition is:

$$H_q = -e^{-\beta t}(cu_t + 2kq_t) + \lambda_t(b+e) = 0 \qquad (4)$$

The auxiliary or adjoint equation is:

$$\dot{\lambda}_t = -H_u = -e^{-\beta t}\, c(1-q_t) + f\lambda_t \qquad (5)$$

By suitable differentiation and substitution in equations (2), (4) and (5) we eliminate q_t, λ_t and $\dot{\lambda}_t$ to get a second order differential equation shown below.

$$ce^2 + b(ce - 2k(f+\beta)) + a(c(b+e) + 2k(f+\beta)) + (-2fk(f+\beta) - (bc+ce)(2f+\beta))u_t - 2k\beta\dot{u}_t + 2k\ddot{u}_t$$

A general solution for u_t that fits this system is

$$u_t = A_0 + A_1 e^{s_1 t} + A_1 e^{s_2 t}$$

with constants A_0, A_1, A_2, and

$$s_1, s_2 = \beta/2 \pm \sqrt{(f+\beta/2)(f+\beta/2+\frac{bc+ce}{k})}$$

The constants A_1, A_2, s_1 and s_2 describe the market for content and one may expect them to be the same within an industry/genre of web sites. Using the boundary condition of given u_0 and $\lim_{t\to\infty}\dot{u}_t = 0$, we obtain the constants. The specific solution obtained is shown below:

$$u_t = u_\infty - (u_\infty - u_0)e^{-rt} \qquad (6)$$

where

$$u_\infty = \frac{c(a+e)(b+e) + 2k(a-b)(f+\beta)}{(2f+\beta)(cb+ce+fk) + fk\beta}$$

$$r = \sqrt{(f+\beta/2)(f+\beta/2+\frac{bc+ce}{k})} - \beta/2$$

[2]We have solved the case when the page composition constraint is not binding. In this case we get u to be a constant. This case does not seem empirically relevant as we rarely see pages that fit well within a 800 by 600 pixel screen, a screen of median size, such that the web site owner can increase the advertisements and have no impact on the amount of content.

Note that u_t moves asymptotically from its initial value u_o to its asymptotic value u_∞ at t increases. This is shown in the figure 1 below for $\beta = 0.02$, $a = 2,000$, $b = 1,000$, $c = 20$, $e = 1,000$, $f = 0.2$, $k = 10,000,000$, and $u_0 = 20,000$ thousand page views.

Using the specific solution (6) and equation (2), we obtain q_t as:

$$q_t = q_\infty + \frac{r-f}{b+e}(u_\infty - u_t) \qquad (7)$$

where q_∞, the asymptotic value of q_t, is

$$q_\infty = \frac{b-a}{b+e} + \frac{f}{b+e}u_\infty$$

Examining equation (7) we note that q_∞ is a constant and $u_\infty - u_t$ decreases over time if $u_0 < u_\infty$. Further, from definition of r in equation (6) we note that $r > f$. Hence q_t decreases with time if $u_0 < u_\infty$. This is illustrated in Figure 2. Using the same parameters as in Figure 1, the proportion of content, q_t is plotted. The proportion of advertisements is $1 - q_t$. We note that the firms make an initial investment in the site by having extra content and fewer advertisements. The proportion of content falls as the web site matures. Higher content and lower advertising may imply that the net cash flow is negative initially. This is shown in Figure 3 and fits with the notion of investing in visitors.

IV. Market value of firms

In this section we close the loop between the model and reality by looking at the association between the model predicted market value of the firm the stock price based market capitalization. We do it in three steps:

1) Using monthly audience data obtained from Nielsen Netratings, we estimate the asymptotic traffic, u_∞, and the growth rate, r at a number of content sites.
2) Define and compute a special statistic of traffic at a web site that we call the *discounted total traffic (DTT)* — the discounted sum of predicted future traffic at a site.
3) Examine the association between the discounted total traffic and market capitalization that is based on stock prices.

A. *Estimating u_∞ and r*

The solution to the control problem predicts that the number of visitors will increase asymptotically from a starting value to an asymptotic value. The final value and the rate of change depend on the firm - the attractiveness of the content, degree of competition and response of customers to advertising. We do not have time series data on

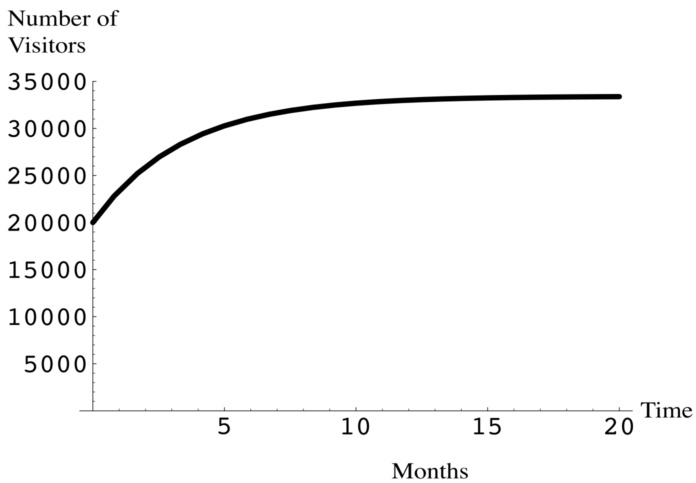

Fig. 1. Number of visitors over time.

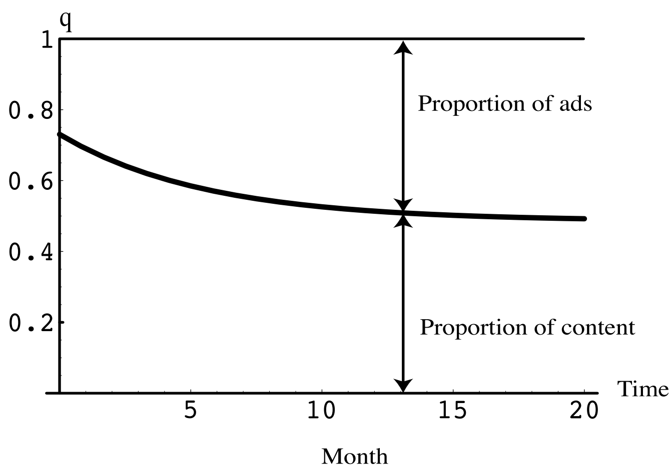

Fig. 2. Proportion of content and advertisements.

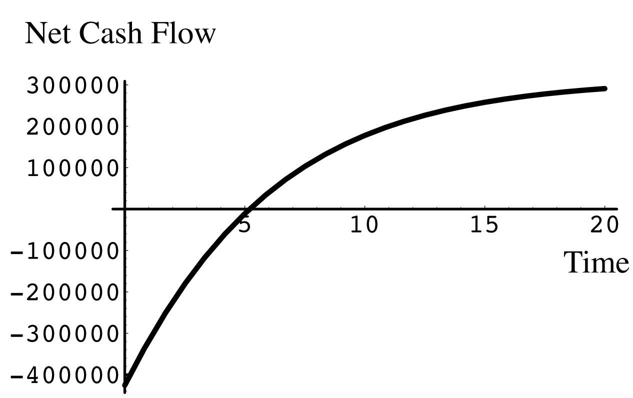

Fig. 3. Net cash flow over time.

Firm	Market Cap US$ Millions	Asy. Traffic Millions (u_∞)	Growth Rate Per Month (r)	Fit (R^2)	DTT Millions
Yahoo	8,871.5	11.44	0.0307	0.95	491.5
Terra	3,606.2	1.08	0.2071	0.51	53.7
HomeStore	3,005.5	42.50	0.00023	0.79	34.8
CNET	1,360.1	9,572.84	0.86E-6	0.82	34.6
Goto	1,047.4	9,376.94	0.2E-6	0.82	9.75
Infospace	984.5	2,783.15	1.24E-6	0.36	19.0
MP3	333.7	0.04	0.0708	0.79	1.81
HotJobs	313.2	1,273.28	2.0E-6	0.84	7.83
Starmedia	138.7	1,646.07	1.56E-6	0.88	8.92
NetZero	121.7	6,109.47	0.73E-6	0.70	19.1
Switchboard	113.6	0.04	0.0485	0.71	1.87
LookSmart	105.6	23.74	0.00025	0.41	19.41
SportsLine	75.9	17.93	0.00039	0.21	25.6
Ask Jeeves	72.9	0.29	0.0442	0.85	12.9
iVillage	41.5	0.17	0.0447	0.82	7.62
ArtistDirect	27.2	7.95	0.000227	0.16	5.89
Women	19.4	0.19	0.0638	0.62	8.43
InfoNautincs	10.8	1.67	0.00019	0.07	1.31
LaunchMedia	9.5	6.04	0.00023	0.28	4.67
Salon	3.6	4.07	0.00026	0.28	3.05

TABLE III
ESTIMATES OF ASYMPTOTIC TRAFFIC AND GROWTH RATE

proportion of content and advertising, but we do have time series data on number of visitors. We obtained this data from *Nielsen Netratings*. They estimate the global viewership of top 500 sites using panels. Using this data we estimated u_∞ and r for a number of sites that are primarily advertising supported. In particular, for site i, let

$$u_{it} = u_{i0}e^{-r_i t} + (1 - e^{-r_i t})u_{i\infty} + \epsilon \qquad (8)$$

where $u_{i\infty}$ and r_i are the asymptotic traffic and growth rate for the site.

The first order conditions for minimizing the sum of square errors are:

$$\sum_{t=0}^{n}(u_{it} - u_{i\infty} + (u_{i\infty} - u_{i0})e^{-r_{it}})(1 - e^{-r_{it}}) = 0$$

$$\sum_{t=0}^{n}(u_{it} - u_{i\infty} + (u_{i\infty} - u_{i0})e^{-r_{it}t})(u_{i\infty} - u_{i0})e^{-r_i t}t = 0$$

We wrote out the second order conditions and found the sum of squared errors to be convex in the estimators. While the first order conditions are necessary and sufficient, they are hard to solve analytically. Consequently, we used the numerical Newton-Raphson method to obtain the estimates. These estimates and the quality of fit for a set of advertisement supported content sites are exhibited in Table III. We used monthly audience measurement data for these sites from February 1999 to May 2001.

B. *The Discounted Total Traffic at a site*

In Table III note that there is a wide variation in growth rates. This reflects firms at different levels of maturity. It is hard to interpret and compare firms with such large variations in asymptotic traffic and growth rates. To provide greater insight, we define *Discounted Total Traffic* (DTT) as the area under the u_t curve discounted continuously at the rate β:

$$\text{DTT} = \int_0^\infty u_t e^{-\beta t} dt$$

Substituting for u_t from (6), we get:

$$\text{DTT} = \frac{u_{i\infty}}{\beta} - \frac{u_{i\infty} - u_{i0}}{r_i + \beta}$$

To illustrate the economic significance of DTT, consider the number of visitors at time t, u_t, from equation (6) for

$r_i = 0$. It reduces to u_{i0}. The total traffic in this case, with continuous discounting at the rate of β, is u_{i0}/β which is equal to DTT with $r_i = 0$. On the other extreme, if r_i is much larger than β then from equation (6) we see that u_t approaches $u_{i\infty}$ rapidly and the total traffic discounted at rate β would approach $u_{i\infty}/\beta$ from below. This is approximated by DTT for large r_i. Essentially, the DTT is a weighted measure of u_{i0} and $u_{i\infty}$ with the weights depending on the specific growth and discount rates.

C. *Market Capitalization*

The Discounted Total Traffic is the discounted area beneath the u_t curve. If each visitor generates the same revenue then the market value of the firm would be expected to be proportional to DTT. To test this hypothesis, we obtained the market capitalization of web site owning firms classified into search and portal category by *Wall Street Research Net* [3]. Taking the discount rate, β, to be 2% (per month) we estimated the coefficients in the following structural equation[3]:

$$\text{Market Capitalization}_i = A + B\ \ \text{DTT}_i + \epsilon \qquad (9)$$

The result of the linear regression is shown in Table IV.

We note that the adjusted R^2 is 84%, the F statistic is 101.1, and the slope is almost surely not zero. This is very re-assuring. However, examining the data more carefully, we note that Yahoo is much larger than others and has a strong influence on the fit. The leverage of this observation is 0.98, which is well above the acceptable limit of $2p/n = 0.2$ where $p = 2$ is the number of coefficients estimated and $n = 20$ is the number of observations. The next highest leverage for any observation is 0.051. The Cook's D statistic, a measure of influence, for Yahoo, is 315 which is far greater than 0.703, the fiftieth percentile of the F distribution with 2 and 18 degrees of freedom, respectively. The next highest Cook's D statistic for any observation is 0.22. Given these statistics, we re-estimated the coefficients for all but Yahoo. The results are shown in Table V. The adjusted R^2 drops to 68%, the F statistic drops to 39.7 (still a very good fit).

[3]A different β would not change the structural equation. Only the coefficients would be different.

R^2	0.85
Adjusted R^2	0.84
F	101.1

	Coeff.	**t-Stat**	**P-value**
Intercept (A)	315.7	1.57	0.133
Slope (B)	18.1	10.05	< 0.0001

TABLE IV
REGRESSION OF MARKET CAPITALIZATION VS DISCOUNTED TOTAL TRAFFIC RATE

R^2	0.70
Adjusted R^2	0.68
F	39.7

	Coeff.	**t-Stat**	**P-value**
Intercept (A)	-316.8	-1.6	0.13
Slope (B)	62.1	6.3	< 0.0001

TABLE V
REGRESSION OF MARKET CAPITALIZATION VS DISCOUNTED TOTAL TRAFFIC RATE **for smaller firms**

V. MANAGERIAL AND INVESTMENT IMPLICATIONS

The most reassuring aspect of the empirical and analytical model examined in this paper is that it finds a significant link between the operational decisions of web site managers, the growth rate of number of visitors to a web site and the market capitalization of the firms. This has important implications for the manager of the web site and investors in evaluating web site performance.

Examining Figures 2 and 3 we see that one can expect the web site manager to "over" invest in content and have few advertisements initially. As time passes, the number of advertisements increase and investment in content decrease. This implies that initially the web sites may not be profitable an only approach profitability over time. This may be a shareholder value maximizing strategy. What matters is not the current profits but the discounted total profit.

Further examining Figure 2 we see that there are positive network externalities, from a consumers perspective, in the web market for content. Notice that a web site with fewer visitors at time 0 has less content and more advertisements than a site with more visitors. Consequently, the web site with more visitors is more attractive to new visitors. This gives rise to portal behavior, where a few web sites get a disproportionately large share of the visitors – a phenomenon that is also explained by competition among web sites for visitors [2].

REFERENCES

[1] `http://ajr.newslink.org,` web site of *American Journalism Review*.

[2] Dewan, R.M., Freimer, M., and Seidmann, A ., "Portal Kombat: The fight to become the common entry point for consumers to the web," in *Proceedings, HICSS 31,* Hawaii, 1998.

[3] *The Wall Street Research Net,* at `http://www.wsrn.com.`

[4] *Yahoo Annual Report 2000,* available at `www.yahoo.com,` 2001.