ABSTRACT

The Internet landscape is an increasingly crowded space where hundreds of thousands of companies are fighting for attention. Some of these companies are universally known (Yahoo!, Amazon) while others are more obscure (Literary Guild, Just Say Wow). Regardless of their level of Internet presence, all these companies are interested in improving their position, particularly now that managers are asked to justify their online activities. To help managers assess their online presence, and give them a way to compare their position relative to their competitors, this article aims to develop a measure of a company’s online visibility. Moreover, this study seeks to understand what factors drive online visibility, and how.

The visibility measure we develop captures the extent to which
a user is likely to come across an online reference to a company's Web site. It is based on data collected from multiple sources that include search engine results, Web-site contents, and online directory listings. It is calibrated using a large-scale telephone survey, and is validated using data obtained from Web crawlers and Internet consumer-panel sources.

In the latter part of the article, we show how managers can use the visibility measure to compare the online presence of their company to their competitors. We also show how they can use this information to perform scenario analysis aimed at finding the best way to improve their online position.

INTRODUCTION

In recent years, the usage of the Internet medium has grown at a staggering rate. A recent study of Internet use estimates the world-wide Internet population at 616 million people, with 185 million in the United States and Canada (European Telecom, 2003). In a review of the World Wide Web, it has been reported that the Web contains 7.1 million unique Web sites (Online Computer Library Center, Inc., 2000)—a 50% increase over the previous year's total of 4.7 million. In particular, e-commerce is expected to continue to grow at a dramatic pace. Forrester (2003) reported that U.S. online shopping reached $78 billion in 2002—a 52% increase over 2001.

Unfortunately, along with the dramatic growth in the number and presence of Web sites on the Internet, there is growing competition to draw the attention of Web surfers. With the proliferation of Web sites on the Internet, Web surfers are now faced with an ever-increasing number of Web-site alternatives, all of which compete for a portion of their limited viewing time. The availability and diffusion of faster means of Internet connection, such as DSL and cable modem, have furthermore been instrumental in speeding up, and greatly facilitating, the navigation process through the Internet. As more sophisticated search engines have evolved (e.g., Google), surfers are further being provided with more relevant guidelines to direct them to the Web sites and information they seek.

Internet surfers are becoming more sophisticated with respect to their ability to use Internet resources (e.g., links on Web sites, search engines, directories, and bookmarks) to more efficiently and swiftly navigate the Internet. However, in their search process, surfers may tend to increasingly avoid distractions from their search goals. For instance, it has been shown that click-through rates from banner ads have decreased markedly in recent years (e.g., Digitrends, 2000). Previous studies also have suggested that surfers may tend to exhibit a low level of patience while surfing the Internet and may experience irritation when subjected to impediments that lead to lengthy waiting times (e.g., Dellaert & Kahn, 1999; Drèze & Zufryden, 1997; Ducoffe, 1996).

Given the current proliferation of Web sites on the Internet and the tendency for surfers to exhibit a low level of patience, a key question that motivates the present study is “How can a Web site best draw the attention and patronage of surfers?”

The latter question is obviously of great interest to Internet companies that have been shown to spend increasingly large budgets in their efforts to enhance their visibility and presence on the Internet, and thereby to increase their site traffic. This has been accomplished by means of both online and offline promotions. For example, offline Web-site promotions include publicity through offline news reports as well as offline advertising for a Web site (e.g., through television, radio, newspapers, billboards, etc.). In addition, commonly used online means of promoting the visibility of a Web site include banner ads, links from search engine sites, and online directories or from other Web sites as well as from mentions in discussion or newsgroups, e-mail received, and online news reports.

Unfortunately, little is known about the relative effectiveness of alternative offline and on-
line promotion methods. In particular, to seek an answer to our key question of how a Web site can best draw the attention and patronage of surfers, this study focuses on the following corollary questions:

- How do surfers find Web sites?
- What is the relative significance, and impact, of alternative promotion methods on a Web site’s visibility?
- What is the relationship between a Web site’s visibility and its traffic?
- Are there differences in surfer search-behavior patterns across Web site industries’ classifications (e.g., automobiles vs. sports sites)?

Finding an answer to these questions is now more important than ever since the days of indiscriminate online advertising are over, and managers as well as shareholders are asking for more accountability.

In the following sections, we first describe an empirical study that was conducted, by means of a large-scale, national telephone survey, for the purpose of measuring the impact of alternative sources of Web site “visibility.” Next, we develop a model to characterize and test relationships of alternative Web site (online and offline) promotion variables to Web site visibility and traffic. We then propose a model approach for evaluating the relationship between alternative online promotions (e.g., placement in directories, results of searches, links from other sites, etc.) and the performance of a Web site (i.e., its online visibility and traffic generation) relative to competitors within given industry classifications. Finally, we conclude with a discussion of the managerial implications of our model and its limitations, and summarize the main results of our study.

**WHAT IS VISIBILITY?**

Visibility is defined as the extent to which a user is likely to come across a reference to a company’s Web site in his or her online or offline environment. Thus, one can view online visibility as a precursor to Web site traffic, in the same vein as awareness is a precursor to purchase. For Web sites, we need to distinguish between online and offline visibility. Many factors potentially contribute to a site’s visibility (see Figure 1). For instance, a Web site can be visible *online* through advertising, on other Web sites, or in newsgroups. A Web site also can be visible *offline* through advertising or public-relations efforts. In this study, we only consider public or broadcast sources of visibility. Word-of-mouth, for in-
stance, is not explicitly considered as a source of visibility in our study. Although we do consider the potential influences of e-mail, chat rooms, and newsgroups in our survey, these sources were not included in our model developments due to data restrictions. E-mail, aside from personal correspondence, can be viewed as direct marketing pieces.

**Offline Versus Online Visibility**

In this study, we consider two primary components of offline visibility: *advertising* and *news reports*. Thus, we consider advertising in the standard broadcast media of television and radio as well as in print media (i.e., newspapers and magazines). In addition, our study considers references to a Web site in any news reports that may be contained in offline media vehicles.

In contrast, there are various means through which a Web site can be visible online. The most ubiquitous and well known is, of course, *online advertising*. Companies have used banner ads for years to promote their sites (e.g., Yahoo.com, Netscape.com, and Zdnet.com). This is a proactive form of visibility where companies pay to be visible in the Internet landscape.

Another ubiquitous presence is through *search engines*. A common strategy for people who are looking for information on the Internet is to query a search engine or *online directory* for the location of the information they seek. According to the GVU's 10th WWW user survey, 85% of the Internet users use search engines to find out about Web pages or sites. Furthermore, 58% use online directories for this purpose. Indeed, search engines and online directories are such important sources of information that a number of companies now specialize in helping Web sites perform well in searches\(^1\) (i.e., insure that a company will be one of the first listed if a search relating to that company is made on a search engine). Given the potential importance of the order of appearance from search results, some search engines (e.g., Excite, Alta Vista) also will sell the top listings on keyword searches to the highest bidder.

In addition to search engines, companies also seek to be referred to from *other Web sites*. Indeed, many users will surf the Web, going from one link to another, while exhibiting a variety of exploratory behavior (see Hoffman & Novak, 1996; Novak, Hoffman, & Yung, 2000). The GVU's 10th WWW survey found that 88% of Internet users find Web sites through links from other Web sites. A link to another Web site may be seen as an endorsement of the value of the Web site that is linked to. Why would the site designer go to the expense of putting a link to another Web site on his or her own site if the site did not offer value? This is especially true since once a surfer has clicked away to the other Web site, he or she might never come back. Concerning sites providing links to other Web sites, news-related Web sites (e.g., CNN.com) are important in that they provide timely information. Indeed, a problem with the Internet is that it changes quickly, and consequently, many links become invalid or outdated.

*Newsgroups* and *chat rooms* are the final two potential drivers of online visibility that we consider in this study. These early pioneers of the Internet are used less now than in the early days of the Internet (GVU has reported a decline in news-group usage from 44% in 1996 to 30% in 1998.) Nevertheless, they are still widely used today and probably contribute significantly to a company's online visibility.

---

1 An online directory such as Yahoo! is different from a pure search engine in that it offers a classification of Web sites that resembles that of a phone book’s Yellow Pages.

Automotive, Shopping, Travel, Sports, Computer and Electronics, Health, Finance and Investment, News, and Internet Search and Service Providers (see Table 1). Furthermore, the Web sites within the industries were chosen to be representative of their respective industries by including Web sites over a wide range of visitor magnitudes (i.e., Web sites with low to medium to high levels of traffic) within each of the industry classifications. A list of the Web sites chosen for each industry is shown in Table 1.

The survey covered questions in three basic areas: Internet use behavior, Web-site-specific questions, and respondent demographics. In view of the significant data requirements on a total of 100 Web sites, and in the interest of maximizing response rate and minimizing respondent overload, 10 different survey versions were designed. This was accomplished by randomly allocating the 100 sites to 10 different survey versions so that each of the 10 resulting versions included 10 Web sites that were matched in terms of their respective mixes of unique visitor magnitudes. During the implementation of the survey, questions involving multiple-choice alternatives were rotated to minimize order bias.

To insure sample representativeness of the Internet user population, respondents were drawn on the basis of a national random-digit dialing (RDD) sample of Internet users. Furthermore, these respondents were qualified on the basis of age (at least 14 years of age) and Internet usage (used the Internet in the last 3 months). Our sample was based on about 150,000 initial dialings. Of these, about 43,000 resulted in respondent contacts (i.e., after the elimination of disconnects, fax tones, and answering machines). This led to 5,000 complete interviews (i.e., after eliminating refusals, mid-interview terminations, and ineligible respondents).

Since each respondent was surveyed on 10 distinct Web sites, we obtained 500 observations for each of the 100 Web sites in our study (i.e., a total of 50,000 observations). Thus, given the relatively high sample size that was ultimately achieved, random sampling error biases were kept at a reasonably low level. Following the survey design and pretesting phases, a commercial market-research company carried out the fieldwork on the survey in March of 2001.3

In view of the main focus of our study, Web-site visibility, the following highlights empirical results that relate to the Web-site-specific questions from our survey. As can be seen from Table 2, the major sources of site visibility for

---

3 The researchers would like to acknowledge Discovery Research Group for the administration of the telephone survey.
Web sites are (in order of importance): offline advertising, offline news report, Internet advertising, and links from other Web sites, links from a search site, and links from an online directory.

**ONLINE VISIBILITY**

Our empirical results suggest that both offline and online promotions may impact the visibility of a Web site. However, these results do not provide specific information about the potential effectiveness of these promotion types. Thus, to gain insights about the degree of association between online or offline promotions and a Web site’s “online visibility,” we measured a Web site’s “online visibility” by defining an “Internet Visibility Index” as follows:

For each Web site, a binary index (0 or 1) was assigned to a respondent depending on whether the respondent had seen reference to, or mention of, the Web site in any one or more of the following online sources:

- in Internet advertising/banner ads,
- in the results of a search done on a search site,
- in a listing in the directory section of a search site,
- on a link to the Web site from any other Web site(s),
- in a discussion group, newsgroup, or chat room,
- in an e-mail received from someone, or
- in an online news article.

Based on the individual responses, the “Internet Visibility Index” ($V$) for each Web site surveyed was obtained by averaging the individual binary index, corresponding to the Web site, over all the individual respondents in the survey. Thus, the index ($V$) for each Web site rep-
represents the proportion (defined over the range 0–1) of Internet users who have seen any online reference to or mention of a particular Web site. Although we evaluated alternative definitions of the latter index, we found the previous definition not only to be the most parsimonious but also to yield the best results in our subsequent empirical analyses.

To test the validity of this index, we regressed it on the site traffic at each of the 100 Web sites. Traffic data were obtained from MediaMetrix (a division of comScore, www.comscore.com) and represent 1 month of Web site activity, based on a sum of unique visitors, for each of the Web sites at the time the survey was being fielded. We also regressed site traffic individually on site awareness (derived from the questionnaire) and company total advertising spending (obtained from Advertising Age, www.adage.com). For purposes of comparison, the $R^2$ values for each of the three regressions are shown in Table 3. Our results show that the Visibility Index explains traffic better than either awareness or total advertising spending. Interestingly, we found total advertising spending to be uncorrelated with site traffic. However, it should be noted that our data did not permit us to evaluate the relationship of advertising with newly generated traffic.

The finding that online visibility, as measured by our survey, outperforms awareness and overall advertising spending, in terms of fitting site traffic, is a promising one. However, if we stopped our analysis here, this result would be merely an interesting fact. From a practical standpoint, both awareness and visibility are measured through survey instruments. However, the Visibility Index requires that respondents answer seven questions while awareness can be measured with only one question (i.e., to ascertain whether a respondent had ever heard of a given Web site). Hence, the visibility measure may be a better measure, but it also is a much a more expensive one to obtain.

For purposes of cost-effectiveness and practical implementation, this leads us to investigate several key questions: Can we predict the Visibility Index without having to conduct a survey? Can we take a snapshot of the Internet landscape and forecast the visibility of any company on that landscape without having to survey anyone? Indeed, if we were able to accomplish such tasks, would we now possess an accurate and inexpensive measure to describe a Web site’s position in the minds of consumers? With this goal in mind, this study provides a methodology that is specifically designed to achieve this task.

### PHYSICAL DRIVERS OF ONLINE VISIBILITY

Taking a snapshot of the Internet is not a novel idea. Indeed, that is in essence what search engines have been trying to do since their inception. Search engines work by crawling the Internet, via crawler software (i.e., bots) that constantly scan and track changes (e.g., new Web sites, new links, and new keywords) that are detected in the Internet environment. The process of taking snapshots that search engines perform is highly specialized. Search engines endeavor to build an index that can be used to search the Web. For the purposes of our study, the snapshot we want to take is just as specialized, but slightly different. Here, we are not attempting to build a searchable index but rather a list of all points of interactions between users and companies. Thus, to reflect how users find out about a company’s Web site, we need to capture data about factors that include the company’s online advertising, links from other Web sites to the company’s Web site, the listing of

---

**TABLE 3**

Explaining Site Traffic—Goodness of Fit Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>$R^2$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln (1 - V)$</td>
<td>68.4%</td>
<td>1</td>
</tr>
<tr>
<td>Awareness*</td>
<td>38.8%</td>
<td>1</td>
</tr>
<tr>
<td>$\ln (\text{AdSpend})$</td>
<td>11.4%</td>
<td>1</td>
</tr>
</tbody>
</table>

*We found that using the same transform for awareness as for visibility, in Table 3, does not improve the fit.

---

4 We regressed both $V$ and $\ln(1-V)$ on site traffic and obtained similar goodness of fit results. The results for $\ln(1-V)$, which relates to the form of the proposed model (3–4), to be discussed later, are reported in Table 3.
the company’s Web site in online directories, how the company will perform in an online search using various search engines, and any mention in online new reports, newsgroups, or e-mail.

**Links From Other Web Sites**

Compiling a list of all references to a Web site has become a straightforward task. In fact, this procedure is at the core of any search engine. The brute-force approach is simply to crawl the net, keeping track of all links contained in Web pages. If one does not want to spend the time and energy to crawl the Internet, one can take advantage of the fact that most search engines will actually provide users with a list of such links (Ilfeld & Winer, 2002). However, using the latter approach puts one at the mercy of the efficiency of search engines. Hence, one would be advised to repeat the procedure across multiple search engines to ensure maximum coverage (see Bradlow & Schmittlein, 2000, for details on search-engine overlap). For the purpose of this article, we chose the first approach. By using a proprietary software methodology to crawl the Web, we gathered 53,017 links to the 100 companies studied. This is an average of 530 links per company.

Once the list of links is collected, we compute a link visibility measure for each of our 100 companies. This can be done in numerous ways depending on how one believes people navigate the Internet. Indeed, Ansari and Mela (2000) showed that there is a serial position effect that relates to the effectiveness of links. In a series of experiments where link order was rotated, it was shown that the higher the rank order of a link, the higher the click-through rate on that link (i.e., first link on a page was clicked on more often than the second). With this relationship in mind, we can construct a link index, $\text{Link}_1$, as a summation over links ($l_i$) to a given Web site, which provides a greater weight to a link the higher its position on a page, as:

$$\text{Link}_1 = \sum_{l_i} \frac{1}{PP_{l_i}},$$  \hspace{1cm} (1a)

where $l_i$ is the number of links tracked, and is the link’s ordinal position in the HTML code of the page where it was found.

Previous studies of the Internet (Drèze & Zufryden, 1998) characterized Web sites as having treelike structures that one explores by going up and down branches. This view may be less applicable now that search engines are used so prevalently. Indeed, search engines serve to bypass the tree organization by sending users directly to the page that contains the information sought. Although the user might still perform a treelike search from there on, this will be done using the landing page as the root of the tree rather than the Web-site’s home page. Further, navigation bars and dynamic menus now allow users to traverse the tree horizontally rather than up and down. Nevertheless, if the tree-search argument holds, we should give less weight to links that belong to pages that are buried deep down a Web site, and more weight to links that are on the home page or close to it. Hence, to reflect such a relationship, we define our second link visibility measure as:

$$\text{Link}_2 = \sum_{l_i} \frac{1}{P_{l_i}},$$  \hspace{1cm} (1b)

where $P_{l_i}$ is the depth of the page on which the link was found (with the home page of a Web site having a depth of 1).

As a third measure, we can create a composite of $\text{Link}_1$ and $\text{Link}_2$ by penalizing links for both a poor page location as well as a poor location within a page:

6 Note that Hanssens and Weitz (1980) found that for magazines, the deeper in a magazine an ad is, the less likely it is to be seen or read.

7 We also tried an alternate measure where the penalty was multiplicative rather than additive (i.e., $1/P_{l_i} + 1/PP_{l_i}$), but it did not improve model fit. We thank anonymous reviewers for suggesting this alternate penalty system.

---

5 The authors would like to acknowledge and thank Word of Net, Inc. for providing the data on Web-site links that were utilized in this study.
\[ \text{Link}_3 = \sum_{i_t} \left( \frac{1}{P_{l_t}} + \frac{1}{PP_{l_t}} \right). \] (1c)

**Listing in Online Directories**

Yahoo! pioneered the online directory business. Yahoo!’s philosophy is very different than that of most search engines. While search engines use computer agents (spiders) to index Web sites, Yahoo! uses actual people to perform this task. Sites are listed on Yahoo! only after a Yahoo! employee goes to the Web site, evaluates its appropriateness for inclusion into Yahoo!, and decides which categories the site belong to. For example, if a site sells stuffed toys, it likely will be placed in the Stuffed Toys Shopping category. Similarly, if a site provides information about the history of toy making, it would be listed in the Toy History category.

We gathered our directory information by crawling Yahoo!’s directory looking for references to any of our 100 companies. In the process, we found 49,491 links. These links fell into 18 different Yahoo! categories, as shown in Table 4. As with the links from other Web sites (Equations 1, 2, 3), we similarly computed three summary variables \( \text{Cat}_{lC} \) for online directory listings (e.g., \( \text{Cat}_{lC} = \sum_{i_t} 1/PP_{l_t} \) where \( C \) refers to categories, \( lC \) is the number of links tracked on Yahoo!’s directory, and \( PP_{l_t} \) is the link’s ordinal position in the HTML code of the page where it was found).

**Online Search**

Search engines index all the content they are exposed to indiscriminately. If one were to submit a page that contained made-up words (e.g., “gfalurp”) to a search engine, these words would be indiscriminately indexed by the search engine. This indexing would be pointless as nobody will make a search using a word he or she does not know. What this means is that to measure a company’s visibility in search engines, one does not merely need to make a search for that company, but rather one needs to see if this company gets listed when one does a search that a user is likely to do.

With this in mind, we compiled a list of the most used keywords (from “2pac” to “zoophia”), combined keywords (e.g., “Jennifer Lopez”), or search phrases (e.g., “How can I get an inexpensive hotel room?”). Once armed with the keyword list, we performed searches on five major search engines using each of the keywords and recorded the position of any links to the companies in our study. Table 5 shows the number of references returned by each of the search engines.

When computing our three visibility measures for keyword search, we slightly changed the formulas of our previous measures. Here, we took advantage of the fact that search engines return results in a well-formatted fashion to be more discriminate when penalizing a link for not being the first one on the page. In line with this fact, we assigned linearly decreasing weights to each link. That is, if a search engine returns 10 links per page, the first link would

---

**Table 4**

<table>
<thead>
<tr>
<th>Category</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts</td>
<td>3,930</td>
</tr>
<tr>
<td>Business and Economy</td>
<td>10,015</td>
</tr>
<tr>
<td>Computers and Internet</td>
<td>1,805</td>
</tr>
<tr>
<td>Education</td>
<td>320</td>
</tr>
<tr>
<td>Entertainment</td>
<td>8,175</td>
</tr>
<tr>
<td>Environment and Nature</td>
<td>70</td>
</tr>
<tr>
<td>Government</td>
<td>792</td>
</tr>
<tr>
<td>Health</td>
<td>484</td>
</tr>
<tr>
<td>Humanities</td>
<td>185</td>
</tr>
<tr>
<td>Law</td>
<td>55</td>
</tr>
<tr>
<td>News and Media</td>
<td>1,163</td>
</tr>
<tr>
<td>Politics</td>
<td>15</td>
</tr>
<tr>
<td>Recreation</td>
<td>5,157</td>
</tr>
<tr>
<td>Reference</td>
<td>28</td>
</tr>
<tr>
<td>Regional</td>
<td>13,724</td>
</tr>
<tr>
<td>Science</td>
<td>914</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>198</td>
</tr>
<tr>
<td>Society and Culture</td>
<td>2,461</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>49,491</td>
</tr>
</tbody>
</table>

---

*The authors would like to acknowledge Word of Net Inc. for the generation of these data as well as those on keyword use.*
have a weight of 1, the second one .9, the third one .8, and so on. Thus, our three measures are:

\[
Key_1 = \sum_{l_k} \left(1 - \frac{PP_{l_k} - 1}{N_{l_k}}\right), \quad (2a)
\]

\[
Key_2 = \sum_{l_k} \frac{1}{P_{l_k}}, \quad (2b)
\]

\[
Key_3 = \sum_{l_k} \frac{1}{P_{l_k} + \left(\frac{PP_{l_k} - 1}{N_{l_k}}\right)}. \quad (2c)
\]

where \(K\) indicates that the measurements were made in relation to keywords, and \(N_{l_k}\) is the number of links per page for the search engine that returned link \(l_k\).

**Online Advertising**

Gathering data about online advertising is much more difficult than gathering data on traditional links. This is due to two factors: First, a large proportion of online ads are dynamic ads (Drèze & Zufryden, 2000). That is, for a given page containing advertising, each request to the page may refer to a different banner ad. This means that whereas links to other Web sites contained in a Web page are fairly static (if the link is here today, it also will probably be there tomorrow), banner ads may change by the minute. Second, the ads are usually not served by the company that is advertising but rather by an ad-delivery company (e.g., DoubleClick). Hence, without actually looking at the banner itself, it is almost impossible to know which company the banner is for and where one will land if one clicks on the ad.

Faced with the shear impossibility of compiling a reasonable facsimile of the online-advertising landscape, we had to rely on aggregate measures for the time period of our study. Hence, in addition to the aggregate advertising data obtained from Advertising Age, we obtained overall banner-exposure numbers, for each company in our study, from MediaMetrix.

**Online New Reports and Newsgroups**

Our study used data for online news report and newsgroups. These data were crawled as part of our main link searches.

**Chat Rooms and E-mail**

We could not collect data about e-mail and chat rooms. But, as shown by our survey results (Table 2), they contribute only in a small part to visibility. Consequently, the omission of these potential sources of visibility was felt to have little effect on the results of our study.

**HYPOTHESIS DEVELOPMENT**

In the preceding section, we have developed measures of five drivers of online visibility (V): links from other Web site(s) (\(Link_1-Link_3\)), listing in online directories (\(Cat_1-Cat_3\)), online search (\(Key_1-Key_3\)), total advertising (\(AdSpend\)), and online advertising (\(BannerSpend\)). We expect all these factors to influence a company’s online visibility. However, as we hinted to when we developed the measures, we expect each measure to affect visibility in a different way. For instance, in the case of links from other Web site(s), we expect \(Link_3\) to be the most significant driver. Indeed, the more buried into the site a link is, the less likely it is to be found by a surfer, and thus the less impact it is likely to have on online visibility. Further, as we expect links to be penalized for their lack of prominence on a page, we formulate our first hypothesis as:

---

**TABLE 5**

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Number of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOL</td>
<td>11,499</td>
</tr>
<tr>
<td>Altavista</td>
<td>6,333</td>
</tr>
<tr>
<td>Google</td>
<td>10,909</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>6,469</td>
</tr>
<tr>
<td>Dmoz</td>
<td>6,806</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>42,016</strong></td>
</tr>
</tbody>
</table>
**H1:** An outside link’s contribution to a Website’s online visibility (V) will be a function of both the page location (P, the earlier the page the better) and the position on the page (PP, the higher up on a page the better).

When returning their results, search engines try to order results in terms of relevance. The most relevant link is listed first, followed by the second most relevant link, and so on. When too many links are found, they are listed on separate pages. It logically follows that users should pay more attention to the first links as opposed to the others that follow. The further down the list of pages and the further down a page a link is, the less likely it is to be seen and thus to contribute to online visibility. Furthermore, as we expect traffic-weighted link measures to outperform their nonweighted counterparts, we similarly expect usage-weighted keyword measures to outperform the nonweighted measures. Hence, we formulate the following:

**H2:** A search result’s contribution to a Website’s online visibility (V) will be a function of both the position of the result link on the result page (the higher the better) and the result page’s ordinal position (the lower the better).

When going through an online directory, users usually drill down to the level of detail they are interested in and then search through the page for the specific information they need. Given this two-step process, we expect Cat₁ and Cat₂ to be more significant in explaining online visibility than the composite Cat₃ measure. This is reflected by the following hypothesis:

**H3:** The contribution of a Website’s listing in an online directory to online visibility (V) will be a function of both the link’s position on the page in which it is listed and the depth of search that has to be performed to reach that page.

The first three hypotheses deal with the impact of the various measures on online visibility. These measures account for differences in how Web sites are referenced on the Internet. A site that has a great many links pointing to it or that scores high on a keyword search will fare better than a site that does not have any links pointing to it or that is not indexed by search engines. However, these measures do not account for any differences across industries (e.g., Sports vs. Travel). Indeed, one would intuitively anticipate differences both in how Web sites are organized and how users search for information about sites across industries. One only has to compare etrade.com and espn.com to see that the two sites are organized very differently. E*Trade only has a few different pages (e.g., portfolio management, stock quotes, etc.) whose content is dynamically generated and changes by the minute. ESPN has a much vaster Web site whose content is (relatively) more static. A search for a sports-related item (e.g., Lakers) is likely to return a link to ESPN while a search for a stock-related item (e.g., MSFT) is unlikely to return a link to E*Trade.

How will industry differences affect our analysis? Our measures are built from two components: the position of a link on a page and that page’s position on a Web site. The reason why link position is important is that humans tend to process printed information in a linear fashion (from top to bottom and left to right in Western countries). Hence, the first link on a page is more likely to be processed than the last one. This effect is a consumer-level effect that depends on internal mental processes and not on external industry-specific factors. Hence, we do not expect any of the measures that deal only with the links’ positions of their respective pages to be industry specific.

In contrast, the effect of the page position on a Web site could vary widely across industries. Some industries might favor a ‘flat’ Web site organization (à la E*Trade) while other industries might favor “deep” Web sites (à la ESPN). If this were true, we would expect the impact of page depth (as measured by Link₂ and Link₃) to vary from one industry to another. Similarly, a complex industry might require online directories to create many subcategories (and hence more levels of organizations) while a simple industry could be captured using only a few
levels of depths. Hence, we also can expect the effect of page depth to vary from industry to industry for the online directory measures (\(Cat_2\) and \(Cat_3\)). Finally, in reflection of the differences in sites’ structure and number across industries, we can expect keyword searches to produce different types of results for different industries. This would yield different industry-level effects for \(K_9\) and \(K_{10}\) as well.

The previous discussion suggests the formulation of the following two hypotheses:

**H4:** Constructs that are solely related to page processing (i.e., \(Link_1\), \(Key_1\), and \(Cat_1\)), which significantly explain online visibility (\(V\)), will do so in the same way across industries.

**H5:** For constructs related to page position (i.e., \(Link_2\), \(Link_3\), \(Key_2\), \(Key_3\), \(Cat_2\), and \(Cat_3\)) that significantly explain online visibility (\(V\)), the associations will be industry specific.

### EMPIRICAL ANALYSIS

To model the online visibility measure that was developed earlier and test our hypotheses, we built a series of regression models. However, it should be noted that standard linear regression analysis, with visibility as the dependent variable, is not appropriate in this case in that this can lead to logically inconsistent results (i.e., online visibility index cannot be properly range-constrained between 0 and 1). Consequently, to seek logical consistency, we specified regressions derived from the following modified exponential form:

\[
V = 1 - e^{-\beta X},
\]

where \(\beta = \text{vector of model parameters (}\beta_0 \ldots \beta_M\text{)}\) and \(X = \text{vector of potential independent variables (i.e., } Link_1, \text{ } Cat_1, \text{ } Key_1 \ldots \).

For purposes of estimation, (3) can be transformed into a linear form whose parameters may be readily estimated by OLS:

\[
\ln(1 - V) = -\beta X.
\]

Provided that \(\beta X < 0\), it can be shown that (3) will yield logically consistent results.

Due to the number of candidate independent variables, we could not run one single model to test all hypotheses in one step. Indeed, we have three sets of constructs (links, keywords, and categories); each construct is operationalized in three different ways (location on page, page location, and composite of both); and each measure can be used as a main effect and or as an interaction with the 10 industries. Hence, there are potentially \((2 \times 3 + 1 \times 3) \times 10 = 90\) parameters, and we have only 100 data points (one for each company). Further, even if we had enough data points, we would probably suffer from collinearity problems if all variables were used at once.

To address our dataset size problem, we started by running stepwise regressions using only the main effects (15 potential independent variables) and the advertising spending variable (\(AdSpend\) and \(BannerSpend\)). This first step picked up \(Link_1\), \(Cat_1\), and \(AdSpend\) (see Table 6).

In a second step, we defined industry classifications by using dummy variables and thus added interaction terms between the main effects and industries. This was done manually one variable at a time, keeping only the interactions that were found significant. This second step added the \(Key_3 \times Industry\) interaction (10 parameters; see Table 7, top). Finally, we fitted a last model that did not incorporate advertising spending (Table 7, bottom). This was done to obtain a model that only takes into account the data that can be captured automatically by a spider program.

**Hypothesis Testing**

Earlier in this article, we developed five hypotheses related to the various drivers of online
visibility. We can now test these hypotheses by looking at which constructs are statistically significant in our final model. For instance, Hypothesis 1 states that: An outside link’s contribution to a Web site’s online visibility ($V$) will be a function of both the page location and the position on the page. Full support for this hypothesis would be found if \textit{Link} 3 were significant or if both \textit{Link} 1 and \textit{Link} 2 were significant. Partial support would be found if either \textit{Link} 1 or \textit{Link} 2 were significant. No support would be found if none of \textit{Link} 1, \textit{Link} 2, or \textit{Link} 3 were significant. Accordingly, we found the following results:

**H1: Partially supported.** We only found \textit{Link} 1 to be statistically significant. This result suggests that only the position of a link on a page is important while the page location itself is not. This may suggest that people do not surf a site in a treelike fashion (since position on page provides a more significant explanation of visibility). This is probably due to the use of search engines that can send a Web surfer anywhere within a site without necessarily starting at the top and digging down.

**H2: Supported.** \textit{Key} 3 is significant. It performs better than a combination of \textit{Key} 1 and \textit{Key} 2. This suggests that both position of a link on a result page and the position on the page affect visibility (i.e., the higher up the link appears on a page, the better).

**H3: Partially Supported.** As with the link construct, we find that all that matters with online directories is where on a page the link appears (\textit{Cat} 1), and not where the page is located (\textit{Cat} 2 or \textit{Cat} 3).

**H4: Supported.** We found no significant interactions between \textit{Industry} and \textit{Cat} 1 or \textit{Link} 1.

**H5: Supported.** The interaction between \textit{Key} 3 and \textit{Industry} is significant.

### TABLE 6
Visibility Model–Stepwise Regression Results

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable Entered</th>
<th>Partial R²</th>
<th>Model R²</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\textit{Link} 1</td>
<td>65.65%</td>
<td>65.65%</td>
<td>187.27</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2</td>
<td>\textit{Cat} 1</td>
<td>5.16%</td>
<td>70.81%</td>
<td>17.14</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3</td>
<td>\textit{AdSpend}</td>
<td>2.54%</td>
<td>73.34%</td>
<td>9.13</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

### TABLE 7
Visibility Model–Interaction Results

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Link} 1</td>
<td>1</td>
<td>0.62</td>
<td>24.45</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>\textit{Cat} 1</td>
<td>1</td>
<td>0.68</td>
<td>26.82</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>\textit{Key} 3 \times \textit{Industry}</td>
<td>10</td>
<td>0.62</td>
<td>2.44</td>
<td>0.0130</td>
</tr>
<tr>
<td>\textit{AdSpend}</td>
<td>1</td>
<td>0.22</td>
<td>8.70</td>
<td>0.0041</td>
</tr>
<tr>
<td>Model R²</td>
<td>79.23%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{N}</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| \textit{Link} 1 | 1    | 0.71        | 25.82   | <0.0001|
| \textit{Cat} 1  | 1    | 0.71        | 25.76   | <0.0001|
| \textit{Key} 3 \times \textit{Industry} | 10 | 0.66        | 2.41    | 0.0142 |
| Model R²       | 77.13% |            |         |        |
| \textit{N}     | 100  |             |         |        |

### MANAGERIAL IMPLICATIONS

The results of our hypothesis testing give us some insights on how people process online information. First, our findings support the importance of the position of a link on a Web page (Ansari & Mela, 2000). Second, it seems that Drèze and Zufryden’s (1998) view that Web sites are treelike in their structure might be correct from a technical standpoint, but not from a user-experience standpoint. It appears that a page’s depth on site has little significance. This
can be due, in part, to the ability of search engines (both search engines such as AltaVista and a Web site’s own search engines) to direct users directly to the page they need, virtually compressing any Web site’s depth to three pages (a search page, a result page, and the pages found). Another reason for the lack of significance of page depth is the wide use of navigation bars that allow users to traverse a Web site transversally without having to continuously go up and down the tree.

Our research has significant managerial implications in that it can help managers define their positions vis-à-vis their competitors, and give them insights on what their strengths and weaknesses are. Moreover, it suggests what can be done to improve one’s position. We discuss these implications in greater detail in the next two sections.

### The Visibility Index as a Predictor of Traffic

Our premise is that online visibility is a precursor to Web site traffic, in the same vein as awareness is a precursor to purchase. To test this, we evaluated the ability of our predicted visibility measure ($\hat{V}$) to explain a Web site’s traffic. Using the predicted values generated by the models from Table 7 to predict site traffic, we obtained $R^2$ values of 70.8 and 73.0% for the model that incorporates advertising and the model that does not, respectively (see Table 8). If we use interaction terms with our independent variables and Industry dummy variables, the $R^2$ was found to climb to 84%.

Further, if we build a model of site visit that incorporates our fitted measures of online visibility, awareness, and advertising spending, we see (Table 9) that only online visibility is significant. The other two variables do not contribute significantly to the model fit. As illustrated in the next section, our results have potentially important consequences from a managerial perspective. Indeed they can suggest ways in which a Web-site’s visibility can be enhanced (e.g., placement of links on other sites, use of keywords, and position of a link on a search result page and within the result page). Clearly, our empirical results support the need for a company to focus on enhancing its online visibility to favorably affect the level of its Web-site traffic.

### The Visibility Index As A Benchmark

The ability to predict site traffic using the visibility index is a good way to validate our measure, but it is of limited use given the availability of traffic reports from companies such as MediaMetrix or comScore. These reports have been heavily criticized in the past for being grossly inaccurate and showing wide variances.

---

10 There is a potential endogeneity issue in that advertising might drive traffic and that as companies increase their traffic they might increase their sales, which in turn might result in larger advertising budgets. We thank anonymous reviewers for pointing this out to us and discuss this issue further in the study limitations.

---

**TABLE 8**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Without Industry Interaction</th>
<th>With Industry Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$df$</td>
</tr>
<tr>
<td>$\ln (1 - V)$</td>
<td>68.4%</td>
<td>1</td>
</tr>
<tr>
<td>Awareness</td>
<td>38.8%</td>
<td>1</td>
</tr>
<tr>
<td>$\ln (\text{AdSpend})$</td>
<td>11.4%</td>
<td>1</td>
</tr>
<tr>
<td>Fitted $\ln (1 - V)$</td>
<td>70.8%</td>
<td>1</td>
</tr>
<tr>
<td>Fitted $\ln (1 - V)$ No Ads</td>
<td>73.0%</td>
<td>1</td>
</tr>
</tbody>
</table>
across the reported Web site performance statistics from various research suppliers (Business Week, 1998; InternetWorld, 1998a, 1998b, 1998c). A more managerially relevant use for the Visibility Index is to benchmark competing companies, compare where they stand from a visibility standpoint, and see on which visibility driver they lag or dominate.

To illustrate the use of the Visibility Index for purposes of competitive analysis, we show the predicted visibility of companies that we grouped in our Shopping category (see Figure 2). In this case, we see widely different levels of predicted visibility. At the top, we have Amazon (79%) and Ebay (44%); at the bottom, we have the Literary Guild with a predicted visibility of 17%. What can the second best, Ebay, do to improve its visibility relative to Amazon?

Looking at the first two rows of Table 10, we see that Ebay trails Amazon on every dimension. Its link index ($\text{Link}_1$), category index ($\text{Cat}_1$), and keyword index ($\text{Key}_3$) are all lower than those of Amazon. There are many possible reasons why Amazon has a leg up on Ebay. On the one hand, Amazon has an extensive affiliate program (Wang, 2001) in which it rewards financially Web sites that drive traffic to Amazon through links (positive impact on Amazon’s $\text{Link}_1$). On the other hand, Ebay has millions of pages, and the contents of these pages change by the minute. This would make it difficult for a search engine to accurately index Ebay’s Web site (negative impact on Ebay’s $\text{Key}_3$).

To illustrate the uses of our model, we examined a series of scenarios to see how Ebay can improve its position.\footnote{Similar studies can be done with respect to other companies. However, a complete study of the strengths and weaknesses of each company’s online position is beyond the scope of this article.} First, we looked at what would happen if Ebay were able to match Amazon on one of the three visibility drivers (e.g., increase its $\text{Link}_1$ index from 31.68 to 168.29) while keeping the other two constant. Second, we looked at what would happen if Ebay were able to increase one of the three drivers by 10%. Finally, we looked at what would happen if Ebay were able to double one of the three drivers. The outcomes of these scenarios are shown in Table 10. The first column describes the scenario; the fifth column shows the resulting predicted Visibility ($\hat{V}$); the last column shows the computed arc-elasticity for that scenario. The arc-elasticity, in this case, is the percent change in Visibility divided by the percent change in

\begin{table}[htb]
\centering
\begin{tabular}{lllrrr}
\hline
Source                              & $df$ & Type III SS & $F$ Value & $Pr > F$ \\
\hline
Fitted $\ln (1 - V)$                & 1    & 1254141625  & 106.23    & <0.0001  \\
$\text{Awareness}$                 & 1    & 30230651    & 2.56      & 0.1128   \\
$\ln (\text{Adspend})$             & 1    & 4028618     & 0.34      & 0.5605   \\
Model $R^2$                         &      & 71.58\%    &           &          \\
$N$                                 &      & 100        &           &          \\
\hline
Fitted $\ln (1 - V)$ (No Ads)      & 1    & 136522486   & 128.21    & <0.0001  \\
$\text{Awareness}$                 & 1    & 28617492    & 2.69      & 0.1044   \\
$\ln (\text{AdSpend})$             & 1    & 14271431    & 1.34      & 0.2499   \\
Model $R^2$                         &      & 74.37\%    &           &          \\
$N$                                 &      & 100        &           &          \\
\hline
\end{tabular}
\caption{Traffic Model}
\end{table}
the decision variable. One should note that since we used a log-linear specification for our model, the underlining elasticities are not constant and typically decrease as the level of the decision variable increases. One also should note that this analysis is done without regard to the actual cost that would be incurred by Ebay to improve its position on any of these drivers. In our illustration, the sensitivity analyses show that if it is desired to match Amazon on the single most important of the three dimensions, it should be on the link index. However, the keyword index has the highest arc-elasticity. Finally, the category index has by far the lowest arc-elasticity.\(^\text{13}\)

The strategies used to increase a company's position on one of the three drivers will vary depending on the driver considered. Improving one's Link\(_1\) index is a matter of convincing other Web sites to link to one's own. Improving on Key\(_3\) is a matter of altering one's Web site so that it indexed well by search engines. Improving on Cat\(_1\) means convincing online directories to give better representation to one's site. In short, each measure is associated with different factors and will represent different levels of difficulties and costs.

**LIMITATIONS**

The models we presented here were built on an extensive and unique database that was assembled by merging together data from many different sources (search engines, panel data, survey data, etc.). They do, however, suffer from some limitations. First, due to the high cost of telephone surveys, we limited ourselves to the analysis of 100 companies. These companies were chosen to represent a wide range of industries and sizes. Nevertheless, being limited to only 100 companies means that we have only limited power in our analysis. For instance, this required us to use a stepwise approach in our model-building steps. It also prevents us from using a holdout sample to test our predictive performance. Hence, ideally, this study should be extended to include more companies and industry classifications.

Aside from the limitations in the scope of our

---

\(^\text{13}\) Remember that these arc-elasticities are company specific. Thus, different companies, with different base values for Link\(_1\), Cat\(_1\), and Key\(_3\), would exhibit different arc-elasticities and thus could face a different order.
work, another major limitation is that the model outputs (parameter estimates used to predict visibility) have a limited shelf life. Because the Internet landscape changes quickly, one would need to recrawl the net on a regular basis (e.g., every month or so), and update the survey on a regular basis (perhaps every 6 months or so) to keep the data current.

A limitation of our model approach is that it does not explicitly account for the possibility that traffic, in and of itself, may have a feedback effect on visibility. For example, when Web site traffic is high, a Web site may be able to spend more on visibility-producing activities (e.g., banner ads, better placement on search engine results, more links on other sites, etc.). A specific consideration of such endogenous effects would require a more complex model framework, and is left as a venue for future research.

In terms of the managerial implications of our work, one shortcoming is that we do not take the costs of the means for improving visibility into account. To make an informed decision regarding which of the three drivers (links, directories, or search engines) a particular company should try to improve on, the company needs to look not only at its current position but also at the cost of implementing any changes. Only through a cost–benefit analysis can normative decisions be made.

We feel, however, that these limitations are outweighed by the quality of the data gathered. The models were built using real-world Internet data and a large-scale survey conducted through random digit dialing. Our study results suggest that model outputs could be used to build a production level system that can be used to help companies evaluate as well as manage their online presence.

**CONCLUSION**

Our study suggests that online visibility is an important concept. A key finding is that it strongly relates to Web site traffic. Moreover, it was shown to have a more significant association with traffic generation than either advertising spending or awareness. From a managerial perspective, our study has focused on the identification of key control variables that have a potentially significant impact on online visibility. The study has illustrated how the evaluation of a Web site’s online visibility can provide a useful tool that can be used to effectively gauge a company’s position in the Internet landscape, relative to competitors, and to diagnose its strengths and

---

### Table 10

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\text{Link}_1$</th>
<th>$\text{Cat}_1$</th>
<th>$\text{Key}_3$</th>
<th>$\hat{V}$</th>
<th>Arc-Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMAZON.com</td>
<td>168.29</td>
<td>8.43</td>
<td>418.14</td>
<td>0.7943</td>
<td>–</td>
</tr>
<tr>
<td>EBAY.com</td>
<td>31.68</td>
<td>4.76</td>
<td>165.66</td>
<td>0.4438</td>
<td>–</td>
</tr>
<tr>
<td>Same $\text{Link}_1$</td>
<td><strong>168.29</strong></td>
<td>4.76</td>
<td>165.66</td>
<td>0.6911</td>
<td>0.129</td>
</tr>
<tr>
<td>Same $\text{Cat}_1$</td>
<td>31.68</td>
<td><strong>8.43</strong></td>
<td>165.66</td>
<td>0.4443</td>
<td>0.001</td>
</tr>
<tr>
<td>Same $\text{Key}_3$</td>
<td>31.68</td>
<td>4.76</td>
<td><strong>418.14</strong></td>
<td>0.6293</td>
<td>0.274</td>
</tr>
<tr>
<td>+10% $\text{Link}_1$</td>
<td>34.85</td>
<td>4.76</td>
<td>165.66</td>
<td>0.4514</td>
<td>0.170</td>
</tr>
<tr>
<td>+10% $\text{Cat}_1$</td>
<td>31.68</td>
<td><strong>5.24</strong></td>
<td>165.66</td>
<td>0.4439</td>
<td>0.001</td>
</tr>
<tr>
<td>+10% $\text{Key}_3$</td>
<td>31.68</td>
<td>4.76</td>
<td>182.23</td>
<td>0.4584</td>
<td>0.329</td>
</tr>
<tr>
<td>Double $\text{Link}_1$</td>
<td>63.37</td>
<td>4.76</td>
<td>165.66</td>
<td>0.5148</td>
<td>0.160</td>
</tr>
<tr>
<td>Double $\text{Cat}_1$</td>
<td>31.68</td>
<td><strong>9.52</strong></td>
<td>165.66</td>
<td>0.4445</td>
<td>0.001</td>
</tr>
<tr>
<td>Double $\text{Key}_3$</td>
<td>31.68</td>
<td>4.76</td>
<td><strong>331.33</strong></td>
<td>0.5738</td>
<td>0.293</td>
</tr>
</tbody>
</table>

* Bold numbers indicate a change relative to the base case.
shortcomings. In particular, the approach described in our study suggests specific ways in which a Web-site’s visibility can be improved (e.g., through the placement of links on other sites, use of keywords, position of links on a search result page and within the result page) so that the Web site’s visitor traffic can be ultimately enhanced.

REFERENCES


