BY PETROS KAVASSALIS, STELIOS LELIS, MAHMOUD RAFEA, AND SEIF HARIDI

To ensure a constant increase in user interest, clicks, recommendations, loyalty, and market share, first understand the information flows and connection networks surrounding each Web site in cyberspace.

WHAT MAKES A WEB SITE POPULAR?

Human cities in the physical world are unevenly distributed centers of economic and social activity characterized by a variety of structural elements, including paths, edges, districts, nodes, and landmarks. They organize themselves in time and space as the result of the principles of convergence, conflict, randomness, and site planning. Likewise, information cities are envisioned as thriving over the Internet, inhabited by tens of thousands, even millions, of users (humans and software agents) exhibiting complex navigational behaviors and participating in cooperative online buying, selling, chatting, interaction, collaboration, and socializing. Evidence of this activity is apparent in such highly popular Web portals as www.ebay.com,, www.yahoo.com, www.aol.com,

ILLUSTRATION BY KARINE DAISAY

and www.amazon.com that represent information and/or commercial interests, business-to-business hubs, and aggregations of virtual communities, some routinely hosting millions of Internet users per day.

To explain the historical large-scale agglomerations of human activity in the global economy, including the concentration of social and economic activity into physical cities and the clustering of entire business sectors in specialized industrial and service districts like Silicon Valley, Hollywood, and the City of London, economists have developed a set of theories and analytical tools, synthesizing them into what they call the New Economic Geography [6]. At the heart of the responses it provides to the question "Where and why does economic activity occur?" is the critical role of "increasing returns." In line with [3], which reintroduced the concept into modern economics, increasing returns operate within industries, markets, and business as positive feedback reinforcing the elements that achieve success or aggravating the effect of loss.

Accordingly, proponents of the New Economic Geography argue that concentrations of human populations and their economic activity arise and are sustained by some form of increasing returns—backward and forward links associated with large local markets—initiating both circular and cumulative processes, with agglomeration being self-reinforcing [6]. It now appears the same strategy of increasing returns explains the intriguingly similar rules describing the Web economy's evolution and growth. Increasing returns contribute to directing disproportionately large numbers of Internet users to the largest, most compelling, most popular Web sites. But how do increasing returns operate and what are their implications for the Web-based economy?

Explaining Growth

The Web ecosystem includes an enormous variety of sites and online pages. Evidence suggests that the process of growth in this expanding business environment follows two notable patterns:

Few large sites. Many small Web sites, but few large ones, operate within the Web economy, so a relatively few sites host the majority of Internet users; or, surprisingly, as demonstrated in [1], the distribution of visitors per Internet site follows a universal power law, similar to the one found in the human population distribution of the world's larger cities.[1]

Quick growth. Popular Web sites grow quickly; for example, Yahoo!, AOL, Amazon, and eBay have built, as reported by consulting firm Morgan Stanley [10],

some of the fastest-growing, most valuable brands in history, achieving that status relatively inexpensively.

Any attempt to explain Web growth would seem to require a theory accounting for both patterns simultaneously. To begin with, a power law might successfully emerge from a stochastic model with the assumption that the expected growth in popularity of a Web site fluctuates in an uncorrelated fashion from time interval to time interval around a positive mean value and is independent of the site's size [1]. However, the problem with this approach is that growth rates selected randomly are exogenous, so they neither reflect the history of the system's evolution nor explain the quick growth rates of the most popular Web sites.[2] To abandon the idea of random growth, Web economists need a model in which Web site growth and visitor population agglomeration emerge endogenously from the behavior of individual agents—the sites themselves and their visitors—and their interactions. An increasing returns-based model is one in which the presence of some sort of agglomeration economies influence the decisions of Internet users concerning their visits to particular Web locations while allowing the final outcome to reflect the history of the decision process.
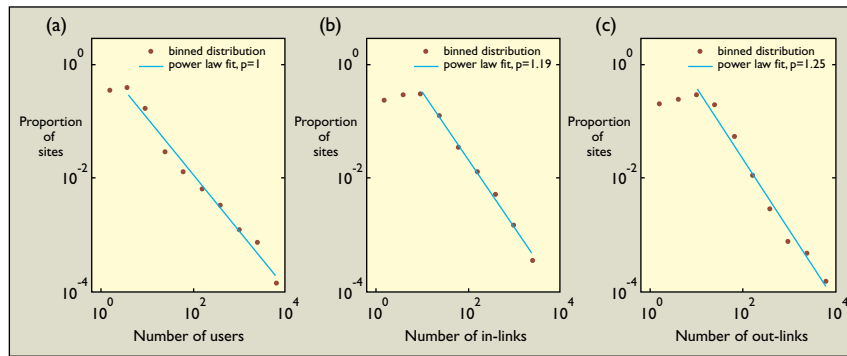
But how might increasing returns in cyberspace be modeled? Following a suggestion in [6], one can explore a modeling approach involving random networks of interaction among Internet users and Web sites, rather than one involving random growth of the size of Web sites. As [6] explains, the "randomness that creates the power law may not involve random growth but random 'connections' in space. For example, imagine port cities that serve the interior along a transport network formed with random connections among transport nodes, with the direction of the preferred connections reflecting accidents either of history or geography. Alternatively, we could suppose that the connections lie in some abstract space of industry linkages" as might exist between suppliers and manufacturers.

A computational model involving two superimposed interaction networks with random connections in cyberspace might reproduce such randomness, thus creating a power law regularity on the Web. In this model, the first network (Web site connections) links the sites; included are nodes corresponding to Web sites and edges representing the links among them that transport users from one site to another. The second network (word of mouth) organizes social interactions among Internet users, allowing for word-of-mouth information propagation within a

---

[1]Many studies suggest that the distribution of larger cities worldwide follows a power law in which a city's size is inversely proportional to its rank in a list of cities ordered by population [6].

[2]It would be more reasonable to expect the magnitude of growth fluctuations for a particular Web site to decrease with its size. Such a decrease is empirically the case of the fluctuation in the growth rates of business firms [11].

structure consisting of local ties and long-range connections. Each network frames the choices made by Internet users about visiting sites in the sense that the way information propagates introduces an information-feedback mechanism into the process of competition among Web sites for market share (information-based increased returns). For example,

step, some percentage of Internet user-agents query other agents (friends and acquaintances) to recommend their own favorite sites. At the same time, user-agents explore the Web on their own, visiting new sites by following the out-links of the sites they've already visited. They might include these sites in their portfolios if they find them interesting or useful (compared with previously selected sites). However, users are relatively loyal to their portfolios, adding new sites as they perceive value in accessing and navigating them.

Utility function. To form and update their portfolios of sites, user-agents employ a utility function with two arguments: the "performance characteristic" of a site (sites are conceived as products, so a different performance characteristic is attributed to each one for determining its performance in terms of natural attractiveness, or intrinsic quality); and the "match" between user preferences and site offerings.

Web-site investment. Sites can deploy investment strategies to improve their performance (in terms of attractiveness), thus influencing the process of portfolio formation and update. Investment is either soft (to sustain performance) or aggressive (motivated by "animal spirit," or greedy self interest), hoping to capitalize on increased growth rates. Most such investments in the Web economy follow the mimic model in the sense that ambitious sites look to replicate successful competitors' investment strategies.[3]

Network structures. A small-world network, or a network structure in which the average shortest path between any two users is small while the clustering coefficient is large, can help describe the dynamics of social contacts among Internet users and mediate word-of-mouth information propagation [12]. Inversely, the Web-site connection network emerges from the stochastic decisions of Web sites to point-in to popular sites, with the quantity of outgoing links varying across sites according to an intrinsic preference for employing a directory strategy involving the categorization of links based on themes. (Some sites, especially large-scale Web directories, constantly add links, categorizing them to provide pointers to as many popular Web resources as possible [8].) The number of outgoing links also depends on a site's rate of growth; for example, popular sites, when employing

Internet users stochastically decide to visit Web sites with probabilities depending on numbers of links pointing-in to the site (in-links); conversely, sites attracting large numbers of visitors become more pointed-in than others (circular causation). On the other hand, sites that users learn about through word of mouth depend on which sites other users have already visited; Internet users are thus more likely to learn about popular sites than unpopular ones (information contagion).

Figure 1. Fitted power law distributions of numbers of site: (a) visitors; (b) in-links; and (c) out-links.

Computational Implementation

We implemented this model in a large-scale agent-based environment called the Web-Simulated Economy, developed at the Swedish Institute of Computer Science (with the collaboration of the Atlantis Group at the University of Crete) using Mozart, a distributed software architecture (www.mozart-oz.org); it allowed us to experiment and progressively produce global dynamic behavior. After t time steps, the model leads to a scale-free state with the distribution of visitors across Web locations following a power law (see Figure 1a). The model incorporates six assumptions:

Two populations. Two small populations of agents that increase exponentially over time represent Internet users and Web sites with diverse amounts of offerings.

Portfolio of Web sites. Internet users organize their site preferences in portfolios of choices, including their most frequently visited sites. As the process evolves, a portfolio may be updated with new sites identified via word-of-mouth information propagation; at each time
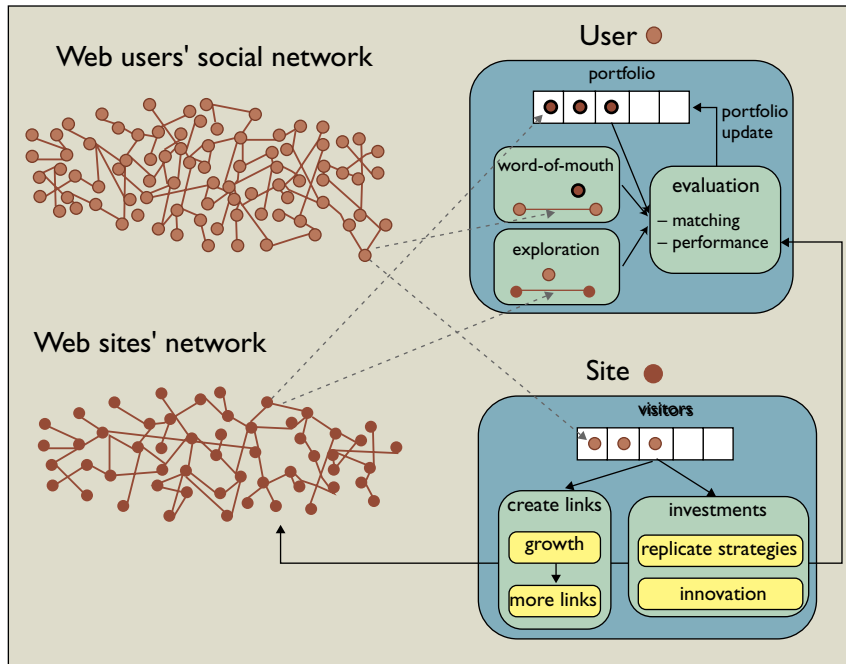

Figure 2. Graphical representation of the Web-simulated economy model.

a directory strategy, naturally increase their outflow of links more rapidly than their less popular counterparts.

Entry strategies. New sites enter the Web economy at different time steps using entry strategies of relatively high initial investments that grow larger and larger as the overall Web economy grows larger.

Beyond its capacity to accurately reproduce the power law regularity, the overall model achieves interesting results in terms of Web market efficiency and Web economy organization, explaining, in the language of organizational economics, how the scale-free nature of the Web emerges in practice (see Figure 2).

The model produces several notably interesting results. First, sites are rewarded for bringing in more and more users by way of relative performance, or their ability to compete in the marketplace, rather than absolute performance, or their natural attractiveness or product quality. In many cases, sites with relatively equal performance differ significantly in the numbers of users they are able to attract and keep. We verified that the effect of word-of-mouth information propagation, combined with a hierarchical exploration pattern privileging the best-connected Web sites, is a powerful mechanism for promoting sites that establish themselves quickly and become well known quickly, while possibly excluding sites with relatively good performance in terms of attractiveness.

Second, newly established sites may achieve a top-ranked position, indicating weak correlation between site age and number of user visits. The reason some sites manage to top the charts so quickly is that once the pos-

sibility of economic behavior or strategic investment is accorded to newcomers, newly established sites can, with a positive probability, quickly accumulate large numbers of incoming links (in-links), thus surpassing older sites. Accordingly, the model obtains only a limited correlation between a site's age and the number of incoming links the site acquires.

Third, visitor distribution across Web sites is not the only factor following a power law; factors that decay as a power law include the number of incoming links a Web site receives during the course of the model (in-links), as in Figure 1b, and the number of outgoing links sites point-out (out-links), as in Figure 1c. Such behavior is another validation-against-reality test for model results, the first being visitor distribution.

Comparative Advantages

The scale-free nature of the Web is explained elsewhere in the context of preferential attachment-based models [5], which assume a network underpinned by two structural mechanisms: continuous network expansion through addition of new vertices and the preferential attachment of new vertices to sites already well connected. An increasing-returns-based model may obtain the same results by modeling behaviors that induce positive feedback into the process of Web sites competing for market share. The more a site is visited, the more users are aware of it and the more additional links it receives (any given Web site generally wants to and does point-in to popular sites). The more users learn about (via word-of-mouth) and discover the site (via user navigation paths reflecting the direction of the links), the more visits it receives. In addition, since the model also involves investment by Web site owners hoping to improve site performance (in terms of attractiveness), economic variables enter it directly. Such investment generates potentially diverse Web site performance. The model suggests that the interplay between a large variation in the landscape of Web-site performance and the complex structure of the networks in which agents are embedded can produce a power law.[4]

---

[4]See [2, 9] for other explanations of power law, also based on the interplay between variation in quality and the complexity of the structure hosting individual agent interaction.

Moreover, in analyzing these increasing returns, one makes an interesting observation about the sources of growth in the Web economy. The model reveals a specific growth process on the Web relating to particular institutional structures, that is, the networks within which individual navigation and site behavior are associated. Information economies in general, and the Web economy in particular, generate specific institutional structures consisting of random information flows involving networks of interaction among individual agents; as these structures are specific to information economies, they propel the growth process for particular sites, quickly claiming market share.

The model proves that the exceptionally high growth rates achieved by the most popular Web sites can be explained by information-based increasing returns, information-feedback mechanisms in the competition for market share, and links among sites generated through point-in conventions. In this regard, one can see (by observing a number of simulations) the effect of parameter d, which represents the number of sites Internet users visit through their personal explorations. When one eliminates this assumption ($d = 0$), thus deactivating the Web-site-connection network (a powerful generator of increasing returns), a very different picture emerges, with no particularly popular sites and relatively slow growth rates. As the parameter d increases (implying a fluid transport network like the Internet itself), the number of sites with quick growth rates increases considerably. The density of the information propagation network, whereby many users adopt word-of-mouth information attitudes, seems to have similar influence, though weaker, on the emergence of the fastest growing sites.

Finally, this broad perspective, which combines behavioral and economic assumptions, explains the scale-free nature of the Web in two ways: in terms of information flowing within the social networks of Internet users and in terms of connections among sites via links associated with pointing-in to popular Web sites. Web sites employ directory strategies to satisfy existing users and attract new ones, but how intensively a site employs them depends on its own growth and popularity. In this context, growth and Web competition are emerging phenomena on top of information network structures.

Conclusion

E-marketers should thus investigate and leverage the long-term ramifications of these structures to help predict the behavior of Internet users toward their organizations' Web sites and to identify the best Web-based ways to promote information about their products. An increasing-returns approach has the notable advantage of being able to identify the sources of population agglomeration and growth on the Web and, modeling them, provide useful insights on how to diffuse marketing information across the multiple networks in which user preferences are embedded. Software like TouchGraph (www.touchgraph.com), for visualizing networks of interrelated information, may be available within a few years to Webmasters and e-marketers who want to observe and analyze the networks involved in Web site positioning and evolution. 🄲

REFERENCES

1. Adamic, L. and Huberman, B. The Web's hidden order. *Commun. ACM 44,* 9 (Sept. 2001), 1–4.
2. Amaral, L., Buldyrev, V., Halvin, S., Salinger, A., and Stanley, H. Power law scaling for a system of interacting units with complex internal structure. *Physical Rev. Let. 80,* 7 (Feb. 1998), 1385–1388.
3. Arthur, B. *Increasing Returns and Path Dependence in the Economy.* University of Michigan Press, Ann Arbor, 1994.
4. Arthur, B. and Lane, D. Information contagion. *Structural Change and Economic Dynamics 4,* 1 (1993), 81–103.
5. Barabasi, A. and Albert, R. Emergence of scaling in random networks. *Science 286* (Oct. 1999), 509–512.
6. Fujita, M., Krugman, P., and Venables, A. *The Spatial Economy: Cities, Regions, and International Trade.* MIT Press, Cambridge, MA, 1999.
7. Gordon, J. Does the 'New Economy' measure up to the great inventions of the past? *J. Econom. Perspect. 4,* 14 (Fall 2000), 49–74.
8. Kleinberg, J. and Lawrence, S. The structure of the Web. *Science 294* (Nov. 2001), 1849–1850.
9. Krugman, P. Confronting the mystery of urban hierarchy. *J. Japanese and Int. Econ. 10* (1996), 399–418.
10. Meeker, M., Mahaney, M., Joseph, D., Trowbridge, M., Cascianelli, F., and Brown, M. Morgan Stanley Dean Witter. *Global IU3, Brand Value, and Customer Monetization for AOL, Yahoo, eBay, Amazon.* Morgan Stanley, 2001.
11. Stanley, H., Amaral, A., Buldyrev, V., Halvin, S., Leschhorn, H., Maass, P., Salinger, A., and Stanley, H. Scaling behavior in the growth of companies. *Nature 379* (Feb. 1996) 804–806.
12. Watts, D. and Strogatz, S. Collective dynamics of 'small-world' networks. *Nature 393* (June 1998), 440–442.

PETROS KAVASSALIS (petros@itc.mit.edu) is the director of the Atlantis Group at the University of Crete in Greece.
STELIOS LELIS (slelis@csd.uoc.gr) is a Ph.D. candidate in the Department of Computer and Communication Engineering at the University of Thessaly, Greece, and a research fellow in the Atlantis Group at the University of Crete in Greece.
MAHMOUD RAFEA (mahmoud@mail.claes.sci.eg) is the director of the Department of Knowledge Engineering and Expert System Building Tools at the Central Laboratory for Agricultural Expert Systems in Egypt. The work described here was done while he was a senior developer at the Swedish Institute of Computer Science in Kista, Sweden.
SEIF HARIDI (seif@sics.se) is scientific leader of the Distributed Systems Laboratory and Chief Scientific Advisor at the Swedish Institute of Computer Science in Kista, Sweden. He is also a professor of computer systems in the Department of Microelectronics and Information Technology at the Royal Institute of Technology in Stockholm, Sweden.