NEW BOOKS IN REVIEW

EDITOR: *Naveen Donthu*

ASSOCIATE EDITORS: *Meryl P. Gardner*
Sandeep Krishnamurthy
Stephanie Noble

THE LAWS OF THE WEB: PATTERNS IN THE ECOLOGY OF INFORMATION, Bernardo Huberman, Cambridge, MA: Massachusetts Institute of Technology Press, 2003, 105 pages, $12.95.

SIX DEGREES: THE SCIENCE OF A CONNECTED AGE, Duncan J. Watts, New York: W.W. Norton, 2003, 368 pages, $27.95.

LINKED: THE NEW SCIENCE OF NETWORKS, Alberto-László Barábasi, Cambridge, MA: Perseus Book Group, 2003, 256 pages, $26.00.

AN ANALYSIS OF POWER-LAW PHENOMENA ON THE INTERNET AND WORLD WIDE WEB

Researchers in marketing have long been interested in social network theory (e.g., the word-of-mouth and new-product diffusion literature are predicated on a network view of the market). Therefore, in considering *The Laws of the Web*, *Six Degrees*, and *Linked*, my purpose is to make this work accessible to researchers in marketing and to excite them about the potential extensions and applications of this work in marketing.

All three books are written for the general intellectual audience and aim to increase the exposure of work published in reputable academic journals (e.g., *Nature*, *Science*). Readers who expect complicated statistical models will be disappointed and must turn to the original journal articles, which are extensively referenced. Even though Huberman, Watts, and Barábasi have different academic backgrounds (sociology, computer science, and physics), their discussion of the social network phenomenon and their results are remarkably similar. They are aware and knowledgeable about one another's work and exhibit a true interest in interdisciplinarity. There is also an interesting interplay among the authors' research teams; for example, Watts shared data with the Barábasi team at one point, in the interest of advancing the field.

Readers who want an introduction to this area may want to read Huberman's book first, which is the shortest. The Barábasi book is excellent general reading and has a great value-to-page-length ratio; the author also includes various examples, including a few business ones (e.g., Hotmail). Watts's book suffers from a disproportionate sense of its place in history. It is substantive, but it is written in a self-indulgent, awe-filled style that may put off some readers. Moreover, the orientation of the book is entirely sociological, whereas Barábasi and Huberman attempt to be interdisciplinary.

Borrowing from fields such as graph theory and building on famous results in sociology (e.g., Milgram's small-world experiments, Granovetter's theory of weak ties), all three books provide unique insights into how networks form and evolve. The results are general and equally applicable to various network forms, yet they are most applicable to the largest information network: the Internet. Indeed, the Internet and the World Wide Web have provided the authors with the measurement tools necessary to measure these phenomena quantitatively. As Huberman states (p. 4), "the Web becomes a gigantic informational ecosystem that can be used to quantitatively measure and test theories of social interaction."

POWER-LAW DISTRIBUTION

The Internet is an information network that contains at least three billion pages. If an expert were asked to estimate the distribution of traffic and users across this volume of pages, the best possible guess would be a normal distribution. Some Web pages receive less traffic, some receive a lot of traffic, and most receive traffic somewhere in between. However, what Huberman, Barábasi, and Watts show is that the Web displays extreme asymmetry, or a power-law distribution in its traffic (Adamic and Huberman 2000).

Watts describes (p. 104) the power-law distribution as follows: "[I]t starts at its maximum value and decreases relentlessly all the way to infinity." I refer readers who are interested in a practical description of the power-law distribution and how it is different from the related Zipf and Pareto distributions to Adamic (2000). An important property of the power-law distribution is that it is scale-free, which Huberman (p. 25) eloquently describes as follows:

> What this means is that if one were to look at the distribution of site sizes over one arbitrary range, say just sites that have been 10,000 and 20,000 pages, it would look the same as that for a different range, say from 10 to 100 pages. In other words, zooming in or out in the scale at which one studies the Web, one keeps obtaining the same result. It also means that if one can determine the distribution of pages per site for a range of pages, one can then predict what the distribution will be for another range of pages.

Specifically, Huberman and his colleagues studied user logs of 60,000 America Online users that covered 120,000 sites in December 1997. Their main findings were the following:

- The top .1% of sites accounted for 32.36% of unique visitors, the top 1% of sites accounted for 55.63% of unique visitors, and the top 5% of sites accounted for 74.81% of unique visitors (the top-ranked site was Yahoo!);
- In subcategories (e.g., education), the top 10% of Web sites accounted for 60% of unique visitors; and
- Few sites tended to have a high number of visitors, and most sites had a low number of unique users.

This extreme asymmetry is notable, because the Internet originally was envisioned as a decentralized and fragmented network. However, many publishers' (i.e., Web site creators) and individuals' random actions have led to a structure in which only a few dominate. The astute reader will immediately note that Huberman conducted this study well before the dot-com meltdown. Thus, this extreme asymmetry was not due to a limited number of choices.

The power-law nature of the Internet has proved extremely robust. Other studies by Huberman and colleagues have uncovered a power-law distribution for the number of visits to a site (Adamic and Huberman 2000), the number of pages within a site (Huberman and Adamic 1999), and the number of links to a page (Albert, Jeoung, and Barábasi 1999). The power-law distribution is not limited to Web sites. For example, consider the data in Table 1 about people who upload information to Gnutella, a leading peer-to-peer file distribution system. Again, there is extreme asymmetry. Finding the same empirical regularity in so many diverse online contexts is not to be expected and should motivate careful development of the behavioral explanations that underlie the phenomenon.

Huberman and Adamic (1999) provide an explanation for the power-law distribution of customer traffic. This explanation is essentially based on the age of a Web site. Huberman and Adamic propose a stochastic model based on random multiplicative growth and different "arrival times" for Web sites. However, this is not a behavioral explanation.

Scholarly marketing readers will notice that this discussion of extreme asymmetry across Web sites is similar in some respects to the literature on double jeopardy in marketing (Uncles, Ehrenberg, and Hammond 1995). Uncles, Ehrenberg, and Hammond (1995) have repeatedly shown that strong brands have many advantages over weak brands. This double (or multiple) jeopardy is realized in many different product categories and even with attitudinal variables.

The comparison to the double jeopardy literature is especially relevant, because Donthu and Hershberger (2001) show that double jeopardy effects are visible on the Internet. In their study, they find that the top search engines have customers who use their services more frequently, are likely to use them more in the future, and spend more time on the Internet than the average customer. They show that leading online book and music retailers have similar advantages. This needs further exploration within the marketing literature.

Table 1
FREE-RIDING ON GNUTELLA

The Top	*Number of Files*	*Percentage of Content Provided*
333 hosts (1%)	1,142,645	37%
1667 hosts (5%)	2,182,087	70%
3334 hosts (10%)	2,692,082	87%
5000 hosts (15%)	2,928,905	94%
6667 hosts (20%)	3,037,232	98%
8333 hosts (25%)	3,082,572	99%

Source: Adar and Huberman (2000).

THE INTERNET AS A SMALL WORLD

Barábasi shows that the Internet is a small world, in the sense that Milgram intended. Motivated by the six-degrees-of-separation game known as the "Kevin Bacon game," Barábasi set out to find the number of links it takes to traverse from a random Web site to another random site. This also could be conceptualized as the diameter of the Web (i.e., the measure of the number of intervening steps between two points on the periphery). Barábasi discovers that there are 19 degrees of separation on the Web, which is a much higher number than the 4 degrees that Adamic (1999) suggests. In any case, this finding must be accepted with a critical eye because Barábasi was constrained by the search engines available at the time, and the Web is a moving target.

Dodds, Muhamed, and Watts (2003) recently conducted an online small-world experiment to prove that the online world is not as small as might be imagined. As Milgram does in his early work, Dodds, Muhamed, and Watts provide the name, location, profession, and educational background of 18 targets and encouraged participants to send a message to someone who was closer to their target. Approximately 61,000 people in 166 countries participated, for a total of 24,163 message chains. However, only 384 of the chains (or 1.59%) reached their designated targets. The authors note that of the chains that reached their destinations, the average chain length was four message relays. They try to finesse their results by noting that the true number of relays with perfect cooperation would have been five to seven, which is far more satisfactory.

Other studies have shown that the Internet is not necessarily as closely connected as the previously cited studies may claim. A particularly notable study, cosponsored by Compaq, IBM, and AltaVista (Broder et al. 2000), reviews 200 million Web pages and 1.5 billion hyperlinks. The results uncover a bow-tie structure of the Internet. Some sites have internal links only, some have external links only, a few have both internal and external links, and very few are completely unconnected. This challenges the notion of the diameter of the Web because this may be different from where you start.

EVOLUTION OF RANDOM NETWORKS

What are the sites at the top, and how did they get there? In other words, how does a site become a winner in a winner-take-all market? This is a nontrivial question. The earliest work in this area was by Paul Erdos, the founder of graph theory. Along with Albert Renyi, Erdos proposed the idea of a random graph as one in which all nodes have an equal chance of being linked. One of their earliest results was with sufficiently large networks: Almost all nodes will have approximately the same number of links. In other words, they predicted a world of symmetry.

How could this work be adapted to the world of asymmetry? Barábasi and Albert (1999) extend Erdos and Renyi's work by proposing a "rich-get-richer" model based on two rules: growth and preferential attachment. They assume that networks are assembled one node at a time; however, the key departure is in the preferential attachment assumption. They assume that the probability of a new node choosing a given node in the network is proportional to the number of links of the chosen node. In other words, if a node has several links, it is likely to receive even more links because new entrants to the network are likely to attach to it.

In Barábasi and Albert's (1999) model, early entry is preferable but is not sufficient for the site with the most number of inbound links (i.e., links to it). Late entrants may not have anywhere to link. The authors show that this model structure directly leads to a power-law distribution.

Barábasi and Albert's work was quite influential, but was later considered too simple (e.g., adding links was costless). However, it inspired a whole new style of models and enabled researchers to move forward. In *Six Degrees*, Watts states (p. 111): "The sheer generality of Barábasi and Albert's model promised a new way to understand the structure of networks as dynamically evolving systems. It didn't matter whether the networks were of people, Internet routers, Web pages, or genes. As long as the system obeyed the two principles of growth and preferential attachment, the resulting network would be scale-free."

THE FUTURE

Barábasi, Huberman, and Watts have their own ideas on where to go in the future. I must note, though, that the biggest gap in this literature is that there is not an understanding of the behavior of individuals that leads to a scale-free network, which is an area in which marketing researchers can bring their expertise to the table. Empirical studies that use experiments, online observation, and secondary data are required to help provide an understanding of the mechanisms that lead to this robust phenomenon. It would also be worthwhile to understand what it is about the Internet that leads to such a network structure. How much of this is driven by the unique properties of the Internet? Are similar effects likely to be observed with mobile commerce, for example?

Barábasi's (p. 225) vision of the future is as follows: "We must remove the wrapping. The goal before us is to understand complexity. To achieve that, we must move beyond structure and topology and start focusing on the dynamics that take place between the links. Networks are only the skeletons of complexity, the highways for the various processes that make our world hum. To describe society, we must dress the links of the social network with actual dynamical interactions between people."

Watts calls for an interdisciplinary push to achieve a deeper study of networks; he (p. 304) calls for "the mathematical sophistication of the physicist, the insight of the sociologist and the experience of the entrepreneur." Huberman muses about how the cumulative knowledge produced in this area will be put to use. For example, the asymmetric nature of uploaders to peer-to-peer networks may lead to selective law enforcement of these people, thus bringing down the network itself. Huberman (p. 99), perhaps, has the most apt finishing sentence: "Oh brave new world!"

Sandeep Krishnamurthy
University of Washington, Bothell

REFERENCES

Adamic, Lada (1999), "The Small World Wide Web," in *Proceedings of the Third European Conference on Digital Libraries: Lecture Notes in Computer Science*. New York: Springer, 443–52.

——— (2000), "Zipf, Power-Laws, and Pareto—A Ranking Tutorial, (accessed August 11, 2003), [available at http://www.hpl.hp.com/shl/papers/ranking/ranking.html].

——— and Bernardo Huberman (2000), "The Nature of Markets in the World Wide Web," *Quarterly Journal of Electronic Commerce*, 1 (1), 5–12.

Adar, Eytan and Bernardo Huberman (2000), "Free Riding on Gnutella," *First Monday*, (accessed October 2, 2003), [available at http://www.firstmonday.org/issues/issue5_10/adar/index.html].

Albert, R., H. Jeoung, and Alberto-László Barábasi (1999), "The Diameter of the World Wide Web," *Nature*, 401, 130.

Barábasi, A. and R. Albert (1999), "Emergence of Scaling in Random Networks," *Science*, 286, 509–512.

Broder, Andrei, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener (2000), "Graph Structure in the Web," IBM Working Paper, (accessed October 2, 2003), [available at http://www.almaden.ibm.com/cs/k53/www9.final].

Dodds, Peter Sheridan, Roby Muhamad, and Duncan J. Watts (2003), "An Experimental Study of Search in Global Social Networks," *Nature*, 301 (8), 827–29.

Donthu, Naveen and Edmund Hershberger (2001), "Double Jeopardy in Internet Site Choice," *Quarterly Journal of Electronic Commerce*, 2 (3), 199–204.

Huberman, Bernardo and Lada Adamic (1999), "Growth Dynamics of the World Wide Web," *Nature*, 401, 131.

Uncles, Mark, Andrew Ehrenberg, and Kathy Hammond (1995), "Patterns of Buyer Behavior: Regularities, Models, and Extensions," *Marketing Science*, 14 (3), G71–G79.