BOOK REVIEW
By
Sandeep Krishnamurthy

1. Huberman, Bernardo (2003), "The Laws of the Web: Patterns in the Ecology of Information", MIT Press.
2. Watts, Duncan J (2003), "Six Degrees: The Science of a Connected Age", WW Norton and Co.
3. Barabasi, Alberto-Laszlo(2002), "Linked: The New Science of Networks", Perseus Book Group.

<u>Introduction</u>

This is a review of three authoritative books that provide us with an understanding of the latest developments in social network theory. Researchers in marketing have long been interested in networks of people- e.g. the Word-of-Mouth and the new product diffusion literatures are predicated on a network view of the market. Therefore, the purpose of this review is to make this work accessible to researchers in marketing and get them excited about extending and applying this work within marketing.

All three books are written for the general intellectual and aim to make work written in reputable academic journals (such as *Nature* and *Science*) accessible to a larger audience. Readers expecting complicated statistical models will be disappointed and must turn to the original journal papers that are extensively referenced. Even though the authors have different backgrounds (sociology, computer science and physics), their discussion of this phenomenon shows remarkable similarity. They uncover similar results and make similar remarks. They are aware and knowledgeable about each others' work and exhibit a true interest in inter-disciplinarity. The interplay between the teams is also interesting. For instance, Watts shared some data with the Barabasi team at one point in the interest of advancing the field.

Readers who want to prioritize may want to read Huberman's book first which is the shortest. The Barabasi book is excellent general reading and has a great value/page length ratio. The author draws from a variety of examples including a few business ones (e.g. Hotmail) and writes in an engaging style. The Duncan Watts book suffers from a disproportionate sense of its place in history. It is substantive. But, it is written in a self-indulgent style that may put off some readers. Moreover, the orientation of the book is entirely sociological whereas the Barabasi and Huberman books try to be interdisciplinary.

These three books offer a unique look into the state of art in networks. 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), these books provide us with unique insights into how networks form and evolve. The results provided by these researchers are very general and are equally applicable to a wide variety of network forms. Yet, they are most applicable to the biggest information network we have ever known- the Internet. Indeed, the Internet and the World Wide Web have provided these researchers the measurement tools necessary to quantitatively measure these phenomena. As Huberman put it on pg 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 with at least three billion pages. Without knowing too much, if an expert were to be asked to estimate the distribution of traffic and users across these pages, the best guess would be a normal distribution. Some web pages would be expected to receive less traffic, some a lot and most would be expected to be in between. Yet, what these researchers have shown is that the Web displays extreme asymmetry or a power law distribution in its traffic(Adamic and Huberman 2000).

Bernardo Huberman's team has shown that the distribution of Internet traffic follows a power-law distribution. This distribution is characterized by exceptional asymmetry. Watts describes the power law distribution in this way- "it starts at its maximum value and decreases relentlessly all the way to infinity". Readers interested in a practical description of the power-law distribution and how it is different from the related Zipf and Pareto distributions are referred to Adamic(2000). An important property of the power-law distribution is that it is scale-free. Huberman describes this eloquently on pg 25 as-

> 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(AOL) users covering 120,000 sites in December 1997. Their main findings were that-

- Top 0.1% of sites accounted for 32.36% of unique visitors, 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 # 1 site was Yahoo!
- Similar results were found in sub-categories- e.g. education- top 10% of web sites accounted for 60% of unique visitors.
- Very few sites tended to have very high number of visitors and the vast majority of sites ended up with very low number of unique users.

This extreme asymmetry is ironic since the Internet was envisioned as a decentralized and fragmented network. However, random actions by many publishers and individuals have led to a structure where only a few dominate. The astute reader will immediately note that that this study was conducted well before the dot-com meltdown. Thus, this extreme asymmetry was not due to limited number of choices.

The power-law nature of the Internet has proven to be 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 Barabasi 1999). Similarly, consider this data about those who *upload* information to Gnutella, a leading peer-to-peer file distribution system-

Table 1
Free-riding on Gnutella
(Source: Adar and Huberman 2000)

The top	Number of files	Provided this percent of content
333 hosts (1%)	1,142,645	37%
1,667 hosts (5%)	2,182,087	70%
3,334 hosts (10%)	2,692,082	87%
5,000 hosts (15%)	2,928,905	94%
6,667 hosts (20%)	3,037,232	98%
8,333 hosts (25%)	3,082,572	99%

An explanation for the power law distribution of customer traffic is provided in Huberman and Adamic (1999). This explanation is essentially based on the age of a web site. The authors propose a stochastic model based on random multiplicative growth and different "arrival times" for web sites. This is clearly not a behavioral explanation, however.

The scholarly marketing reader will notice the similarity of this discussion to the Double Jeopardy phenomenon in marketing(Uncles, Ehrenberg and Hammond 1995). Ehrenberg and his colleagues have repeatedly shown that strong brands have multiple advantages over weak brands. This double (or multiple) jeopardy is realized in many different product categories and even with attitudinal variables.

This is especially relevant since Donthu and Hershberger(2001) have shown that Double Jeopary effects are clearly visible on the Internet. In their study, the top search engines were found to have customers who used their services more frequently, were likely to use it more in the future and spent more time on the Internet than the average customer. Leading online Book/Music retailers were also shown to have similar advantages. This needs further exploration within the marketing literature.

<u>The Internet as a Small World</u>

Barabasi's first contribution was to show that *the Internet was a small world in the Milgram sense.* Motivated by the six-degrees-of-Kevin-Bacon game, Barabasi set out to find the number of links it takes for a random site on the Web to get to another random site. One could also think of this as the diameter of the Web- i.e., measure the number of intervening steps between two points on the periphery. Their research uncovered *the degrees of separation on the Web to be nineteen*. This is much higher than an earlier study published by a colleague of Huberman, which suggested that this number was 4. In any case, this finding must be accepted with a critical eye because Barabasi was constrained by the search engines available at the time and the Web is a moving target.

A very recent online small world experiment conducted by Duncan Watts and colleagues proves that the online world is not as small as one might imagine(Dodd, Muhamad and Watts 2003). Much like the early work by Milgram, Watts and colleagues provided the name, location, profession and some educational background of 18 targets and encouraged participants to send a message to someone who was closer to their target. At least 61,000 individuals from 166 countries participated. A grand total of 24,163 message chains were created. However, *only 384 of the chains (or 1.59%) reached their designated targets*. The authors note that, *of those chains that reached their destinations*, four relays of the message was the average chain length. They try to finesse their results by saying that the true number of relays with perfect cooperation would have been five to seven with an average of six. This is far more satisfactory, however.

Other studies have shown that the Internet is not necessarily as closely connected as these studies may claim. A particularly interesting study reviewed 200 million web pages and 1.5 billion hyperlinks and was co-sponsored by Compaq, IBM and Alta Vista(Broder, Kumar, Maghoul et. al. 2000). Their main contribution is that they uncover a bowtie structure of the Internet. Some sites have links to them only, some have links emanating from them only and only a few have links to and from. A very small number are completely unconnected. This challenges the very notion of the diameter of the Web because this may be different from where you start!

It is clear from these results that these researchers are on the cutting-edge of this phenomenon. As more empirical evidence is uncovered, it is likely that our understanding will evolve as well leading to more robust theories.

Evolution of Random Networks

The obvious question that has not been answered so far in this review is this-"What are the sites at the top and how did they get there?". In other words, *how does one become a winner in a winner-take-all market*?

This is, clearly, a non-trivial question. The earliest work in this area was by the founder of graph theory- Erdos. Along with Renyi, he proposed the idea of a random graph as one where all nodes have an equal chance of getting a link. One of their earliest results was with sufficiently large networks, almost all nodes will approximately have the same number of links. In other words, they were predicting a world of symmetry.

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

In the Barabasi and Albert(1999) model, early entry is preferable- but is not sufficient to sit atop the pyramid. Late entrants may not have anywhere to link. The authors showed that this model structure directly led to a power law distribution.

This was a very influential piece of work. It 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. Duncan Watts gushes on page 111-

> The sheer generality of Barabasi 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

All three authors have their own ideas on where to go in the future. But, before we get to their ideas, I must note that the biggest gap in this literature is that we do not understand the behavior of individuals that leads to a scale-free network. This is where marketing researchers can bring their expertise to the table. We need empirical studies that use experiments, online observation and secondary data to help us understand the mechanisms that lead to this robust phenomenon. It will also be worthwhile to understand what it is about the Internet that leads to such a network structure. How much of this is being driven by the unique properties of the Internet? Are we likely to observe similar effects with mCommerce, for instance?

Barabasi's vision of the future is this(pg. 225)-

> 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.

Duncan Watts calls for an interdisciplinary push to achieve a deeper study of networks. He calls for "the mathematical sophistication of the physicist, the insight of the sociologist and the experience of the entrepreneur" (pg. 304). Huberman muses about how the information uncovered by these studies 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 individuals, thus, bringing down the network itself. Huberman, perhaps, had the most apt finishing sentence- "Oh brave new world!".

REFERENCES

Adamic, Lada(2000), Zipf, Power-laws, and Pareto - a ranking tutorial, Available at- http://www.hpl.hp.com/shl/papers/ranking/ranking.html, Accessed on August 11, 2003.

Adamic, Lada 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*, Vol5(10), Available at- <http://www.firstmonday.org/issues/issue5_10/adar/index.html.

Albert, R. , H. Jeoung, A-L Barabasi(1999), "The Diameter of the World Wide Web", *Nature*, 401:130.

Barabasi, 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, Janet Wiener(2000), "Graph structure in the web", IBM Working Paper, 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), 827-829.

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, Ehrenberg, Andrew, Hammond, Kathy(1995), "Patterns of buyer behavior: Regularities, models, and extensions", *Marketing Science*, 14(3), Part 2. pg. G71-G79.