How Not to Study the WWW

by Paul B. Hartzog (PaulBHartzog@panarchy.com)

Abstract

In his book Linked : How Everything is Connected to Everything Else and What it Means for Business, Science, and Everyday Life, Albert-Laszlo Barabasi gives us a detailed analysis of the topology of the WWW. In so doing, he makes many errors from which we can derive important lessons about ways *not* to study the WWW or complex networks in general. These lessons are crucial from the point of view of the philosophy of science, and suggest that more care and reflexivity is called for in pursuing WWW research. This paper is intended to provided impetus for meaningful thought and further discussion.

Contents

Introduction: Quality and Quantity

This is not a book review. Although I take Albert-Laszlo Barabasi's book Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life as my starting point, the kinds of errors that Barabasi makes in Linked are not atypical, and insofar as they are errors in general, and not in particular, they can be dangerous not only to studies of the WWW but also to studies of complex adaptive networks (CAS) as a whole. It is crucial to highlight these problems lest the study of the WWW and CAS become captivated by properties which are quantifiable and overlook those properties that are not.

The basic point that I will make is that **quantity** does not yield **quality**, and by extension quantitative analysis does not yield qualitative conclusions. **More is not better.** Put another way, **descriptive** findings do not yield **normative** conclusions. In philosophy-of-science terms: one should not confuse a **methodology** for an **ontology**, i.e. confuse the presence of something analytically for its value existentially. Henceforth, I will survey many of the claims made in <u>Linked</u> with an eye towards how they ignore the distinction between quantity and quality. Like complex networks, my argument is non-linear, it is a network of inter-related points. As a consequence, much of this paper can be re-read in essentially any order. Doing so should reinforce, clarify, and strengthen an understanding of the points made.

Also, in my view, it is precisely those properties that are not quantifiable that have real value. For this reason, I sincerely hope to spur further research on the WWW and CAS, but with a more holistic ontology. Research on complex networks is essential for the future of mankind, but this research must contain **both quantitative and qualitative dimensions**.

This paper consists of four sections. First, I will give some background on network analysis, both quantitative and qualitative, and discuss its relevance to the WWW. The second and third sections will contrast the kinds of "static quality" and "dynamic quality" that we find in networks, as well as how these qualities affect WWW and network analysis in general. The final section draws some general conclusions about appropriate network analysis.

Network Analysis:

Analytical Dimensions of Networks

Networks consist of **nodes** and **links**, alternately referred to as **agents** and **structure**. In a typical network analysis, links between nodes can be characterized as **undirected or directed** and **binary or valued**. (Scott, 2000, p. 47). In a social network an example of an **undirected** link would be a family connection: two persons are either both related to each other or they are both not. Friendship would be an example of a **directed** link: I may claim you as a friend, but you may not claim me as a friend. A **binary** link is one that either exists or it does not, i.e. it has no value other than yes or no. By contrast, a **valued** link is one that has weight: individuals may count each other as close friends or distant acquaintances, and valued links can range along a continuum from strong to weak. Although qualitative valued links can be approximated quantitatively, Barabasi chooses to characterize the hyperlinks connecting nodes on the WWW as binary rather than valued, and this strong quantitative approach overlooks some crucial features of qualitative networks.

Complex networks "can be analysed through a number of concepts that describe the quality of the relations involved" (Scott, 2000, p. 31). This goes beyond mere quantitative description. Analytical properties like Mitchell's "reciprocity, intensity, durability" (Scott, 2000, p. 31) and Homans' "direction, frequency, and intensity" (Scott, 2003, p. 22) suggest the possibility of a more complete typology of links on the WWW.

Following social network theorists like Mitchell and Homans, we can think of links on the WWW as having distinct qualities:

1. Direction

2. Persistence
3. Value

Direction is merely who is linking to whom; hyperlinks take you ***from*** one page *to* another. **Persistence** reflects whether the link is temporary or permanent. An example of a temporary link would be a headline or weblog entry which links to another page. **Value** refers to the weight that an incoming link carries for the destination page. It is this value that introduces a qualitative aspect to the WWW, and it is the various aspects of this value that we will explore further.

Robot Topology

Barabasi tells us a **robot** is "a piece of software that downloads any document, finds all the links on it, then visits and downloads the documents to which they point, continuing until all pages on the Web are captured…. This software is called a *robot* or *crawler* because it crawls through the Web without human supervision" (Barabasi, 2003, p. 32). The use of the robot methodology is significant, because for a robot links **are** binary, they either exist or they do not, and if they exist, they get used. What this reveals is a perfect congruence between Barabasi's *choice* of how to characterize links, and his *method* of measuring them. His **ontology** determined his **methodology** which in turn served only to reinforce his existing ontology.

Not only does a robot lack human supervision, it also lacks human faculties. Remember, for a robot, if a link exists, it gets used. For a person, this is not the case. In a typical website that is a group of related webpages, there is a “Home” link to allow visitors to return to the “Home” page of the site, i.e. the main page. A robot would consider this link as valid and follow it from each and every page on the site, whereas a person surely would not. Furthermore, in many cases, there is also a “Navigation” bar or menu consisting of links to other areas of the same website. Persistent navigation gets used by human visitors, but any analysis of website traffic would reveal that some pages of the website, and thus links to those pages on the persistent navigation menu, get used more often. For a robot, this difference is nonexistent, as all links are treated as having equal weight. They either exist or they do not, and if they exist they will get followed. Additionally, a persistent navigation menu which puts the same set of links on each page, creates a clustering effect that is intended to keep website visitors at that website and to decrease the likelihood that a website visitor will go elsewhere. “It is tempting to believe that robots can avoid this popularity-driven trap. They could, but they don’t” (Barabasi, 2003, p. 174). People, on the other hand, are capable of escaping the popularity-driven trap quite easily, and usually do.

Finally, many pages on the WWW don’t even exist until visited. Because they don’t exist as pages per se, they can’t be downloaded and parsed for links in the same way as static webpages. For example, any webpage requiring input, such as a “Search Results” page, falls into this category. Since such a page will contain links to other pages, there is no way to determine how many links the robot is not taking into account.

Similarly, we must look deeper at what the robot is actually doing: following links. This is significant because Barabasi uses the link topology of the WWW to draw various conclusions about the WWW. "On the Web the measure of visibility is the number of links. The more *incoming links* pointing to your Webpage, the more visible it is" (Barabasi, 2003, p. 57). Visible to whom? A robot. Barabasi is making a qualitative statement drawn from quantitative analysis. This in and of itself is not necessarily troubling, but to accept such statements as if they were statements of brute fact when they are not, is dangerous. In fact, there are numerous ways to reach a webpage without clicking a link at all. I can:

1. type in a URL manually
2. use a bookmark from my browser software
3. click a saved link/shortcut on my computer
4. click in a software application on a link that takes us directly to a webpage

Thus, "visibility" is a more complex property of a page on the WWW, and a node in the network, that cannot be reduced to topology alone, and to robot topology in particular. By failing to take into account human traffic patterns on the WWW, which constitutes actual page visits, Barabasi gets precisely what he asks for, namely, a topology of the WWW *for a robot*, and not a topology for humans. It can be argued that geography presents a "real" topology regardless of how humans traverse it, i.e. where the roads are, but the distinction here is that one is a mechanistic topology and the other is a use topology. This is again a key issue in the philosophy of science, and which topology is more "real" than other is entirely up for grabs. Both may be useful, but for different reasons and within different contexts.

Finally, the robot has no way to distinguish the **persistence** of a link, and since by-rote continuous mapping is time-consuming, many of the links followed will have vanished by the time the map is completed. Because the WWW is dynamic, there is no sufficiently fast way at present to get a snapshot of its state at any given point in time. This may indeed be trivial, but we should not assume that it is.

Network Density

Network density provides us with a kind of measure of the connectedness of a given network. However, *maximum* connectedness does not equal *optimal* connectedness. It is usually used with binary graphs, and as I've mentioned the WWW is a valued graph. As Scott informs us, it is "possible to use the density measure with valued graphs, though there is very little agreement about how this should be done…. [Density] would need to be based on some assumption about the maximum possible value that could be taken by the multiplicity in the network in question…. There is, however, no particular reason why the highest multiplicity actually found should correspond to the theoretically possible maximum" (Scott, 2000, p. 73).

In WWW terms, this means that for any given webpage the number of outgoing links could range from 0 to some high value, but not infinity. In fact the upper limit is likely to be further constrained by the technical limitations of the average computer browsing the WWW. Pages that are very large require more time and memory to load, thus, beyond a reasonable size, large webpages actually drive away users rather than draw them. The qualitative lesson here is again "more" is not "better." If a page presents a visitor with too much information, such as a large number of outgoing links, this does not necessarily make this page a "better" node in the network, or even a more useful, visible, or visited one. The same principle applies to incoming links: if too many incoming links generate too much traffic, then the webserver crashes and the webpage becomes inaccessible. In hacking terms this is equivalent to a "packet flood" or a "denial-of-service attack," and is sometimes done on purpose to achieve precisely this result.

Assessing the Value of Hubs and Non-Hubs

Barabasi describes the topology of the WWW in terms of hubs, or highly central nodes, but centrality can be local or global. Hubs can be central with regard to most of their neighbors (local centrality), and they can be central with regard to the network as a whole (global centrality). These two concepts, however, are not always distinguishable because "the centrality of a particular point cannot be assessed in isolation from the centrality of all the other points to which it is connected. A point that is connected to central points has its own centrality boosted, and this in turn, boosts the centrality of the other points to which it is connected" (Scott, 2000, p. 87). On the WWW, this amounts some sites seeming to be more central than they actually are, i.e. more visible than they actually are, merely because their centrality has been unduly inflated. Human navigation on the WWW has an opportunity to detect such exaggerations, but again robotic navigation does not.

Also, network nodes have a property called **betweenness** which "measures the extent to which a particular point lies 'between' the various other points in the graph: a point of relatively low degree may play an important 'intermediary' role and so be very central to the network" (Scott, 2000, p. 86). In other words, a point which is not linked to many other points and is therefore **quantitatively weak** may still occupy an important position in a network and thus be **qualitatively strong**. If this position is crucial enough, the network can exhibit what is called "local dependency." "A point is dependent upon another if the paths which connect it to the other points pass through this point" (Scott, 2000, p. 87). Another way of thinking about this is a called a "cut-point" "A cut-point is one whose removal would increase the number of components by dividing the sub-graph into two or more separate sub-sets between which there are no connections" (Scott, 2000, p. 107).

On the WWW, we see such local dependency often. Any redirect, such as a meta-tag redirect, a virtual directory or Unix symlink, or even DNS redirects are examples of local dependency. In addition, many websites have a "splash page" which shows something briefly and then forwards the visitor to the interior of the website. Some websites detect the parameters of the visiting browser software and then forward the visitor to a particular page that is customized for that visitor's system configuration. If the interior of a website is unavailable except through such redirects, then that portion of the network is locally dependent on the single incoming route, or cut-point.

But on the WWW, there is another way in which to understand dependency which is not even a function of centrality at all. The **quality** of automated verification and payment systems, i.e. the systems that allow you to pay using a credit card online (including systems like http://www.paypal.com) is absolutely vital to e-commerce, even though they may not possess a large **quantity** of hyperlinks. These verification and payment linkages are generally embedded in web-applications and/or forms, which are not accessible to robots in the same way as static hyperlinks. In fact, there are a number of Internet resources – XML and SOAP just to name a few – that are designed for data sharing and transmission, not for viewing. These kinds of "links" on the WWW are crucial for numerous web-applications, but are not part of the viewable website topology, the point being that although they have low degree they are of high value.

The Effect of Search Engines on Topology

Barabasi gives us an example of the difference between the highway network and the airport network: the road network is fairly spread out, whereas the airport network consists of clearly identifiable hubs (Barabasi, 2003, p. 69). What Barabasi unintentionally demonstrates is the difference between **far** and **near**. To get to somewhere far, the airport topology is effective, both for passengers as well as for the airline industry. By contrast, to get to somewhere near, the decentralization of the automobile and the system of roads it uses provide a better system. Search engines are the Star Trek transporter of the WWW, they collapse distance, turning **far** into **near**. The results of a search at http://www.google.com are inaccessible to a robot crawling the WWW's vast system of hyperlinks, because the page doesn't exist until I've both requested it and provided it with my search criteria. In effect, a search engine greatly increases the number of incoming links to my site (assuming that my site is indexed in the search engine in the first place). Search engines work by negating network distance, effectively flattening many path-lengths to 1. And if search engines flatten the hierarchy, then meta search engines like http://www.vivisimo or http://www.dogpile.com flatten it even more.

The link structure of the WWW clearly defines four "continents" of webpages: In, Out, Core, and Islands (Barabasi, 2003, p. 168). While it is true that these continents will emerge if you robotically crawl links, there are other ways to gather network information. Search engines often use robots to index websites, but they could do better. For example, programs can query servers directly and show you a hierarchy of all of the documents on that server without crawling any links at all (for example, the software Intellitamper at http://www.intellitamper.com). Also, there exist "hacker maps" of the Internet showing networks of sub-components, that appear as isolated island continents when using robots. In addition, a mapping approach or search engine that queried DNS would get a different starting point than linkage for indexing sites. (A search engine that lists only top-level domains would be useful tool for refined searching.)

Barabasi asks: "A physicists' Webpage might mix links to physics music, and mountain climbing, combining professional interests with hobbies. In which community should we place such a page?" (Barabasi, 2003, p. 171). More importantly, in which community should we place the physicist? Overlapping networks provides the value of the social fabric. If an internet friend knows me through my work, and I know a cook through my Aikido dojo, then my internet friend is only two steps from my cook friend, which is terribly useful to him if he happens to ask me for a good guacamole recipe. Attempting to identify the "right" group for everything is the bane of librarians: does a book on sports medicine go in "sports" or "medicine?" This mode of thinking is a legacy effect of dealing with real-world artifacts – the book has to be filed somewhere. Virtual artifacts, however, exist in multiple "places." This is one reason why hierarchical indexes (like the Librarians' Index to the Internet at http://www.lii.org or Yahoo's categories at http://dir.yahoo.com) are typically less useful, and less used, than search engines.

On the other hand, there's "google bombing:"

> (GOO.gul bawm.ing) *n*. Setting up a large number of Web pages with links that point to a specific Web site so that the site will appear near the top of a Google search when users enter the link text [1].

"Google bombing implies that Google can amplify hierarchy, rather than flattening it"[2]. For example, at one time "typing the phrase 'miserable failure' into the Google search box produced an unexpected result: the White House's official biography of President George W. Bush"[3].

Needless to say this amplified hierarchy will affect robots as well, and google bombing isn't the only way to do it. Many websites pay money to show up higher on the list in search engines, so their actual traffic may not be proportional to the number of incoming links that exist on the web. Plus, pornography sites have been using banner ads and tiny "feeder" sites to load the search engines and drive visitors to their websites since the WWW began. I would be interested to know how prominently pornography sites figure in quantitative WWW topologies.

Static Quality:

Proportional Linkage

Consider a webpage with only two incoming links: one is from a referrer with only two outgoing links at all, and the other is from a referrer with a long list of outgoing links. A visitor to the page with two outgoing links may click them both and thus reach my page. By contrast, a visitor to the page with many outgoing links may feel overwhelmed and click none of them. The value of an incoming link for my page is a function of many factors. Barabasi's quantitative approach treats all incoming links as having identical weight, or **value**.

Also, as the number of links to my page increases, each one represents an ever decreasing proportion of my incoming linkage. How much traffic they each bring in is an entirely different variable. Even if my total traffic increases, the proportional value of each new link decreases, and popular sites like http://www.google.com/ benefit almost nothing from new links to them. If each incoming link is thought of as a "friend" then the situation is analogous to social networks where "the more neighbors there are, the less impact any one of them has, in a fractional sense" (Strogatz, 2003, p. 267).

Certainly, a website with too few links is alienated, but a website with too many links experiences a roll-off in the value of each new link, probably asymptotic. I suspect there is a parallel here with the upper and lower boundaries constraining network cascades: "either it is not well connected enough or (and this is the surprising part) it is *too well* connected" (Watts, 2003, p. 237-40).

Website Design

Another important factor affecting link quality can be the website design of the page that links to my page. If the link to my site is featured prominently, such as "Cool Website of the Day," then I could potentially get a lot of incoming traffic. Conversely, if the link is buried in the small print at the bottom of the referring website then I'm unlikely to get much traffic at all.

The fact that humans use visual information to navigate online is expressly acknowledged by Barabasi. "The trick, of course, is that we do not follow all links. Rather, we use clues…. By *interpreting* the links, we avoid having to check all the pages" (Barabasi, 2003, p. 37). Oddly, he fails to appreciate the importance of this for his research methods. I am referring to the fact that Barabasi and his team used robots to search links: robots don't **see** the page design, they just follow the paths as if they all had equal "presence" on the page (i.e. in the source code, or HTML). Since there is no perception, there is no visual effect of the page's content.

Valuable Referrers

A binary characterization of links treats all links as qualitatively identical. If I link my page to the New York Times, and they link to my page, then both of those links are equivalent. Anyone who has put up a webpage knows this is patently absurd. *Qualitatively* those links are not equivalent: the link to my page from the New York

Times is more valuable to me than my link to them is to them. Another way of putting it is that I am likely to get more out their link to me, than they will get out of my link to them. This is the same in social networks: presumably I would benefit more from knowing Bill Clinton, than he would from knowing me.

"If each document on the Web had a link to your Webpage, in a very short time everyone would know what you had to say." (Barabasi, 2003, p. 57) This is in effect like using a bulldozer to weed your garden. Much more effective than many links is *one* link from a **valuable referrer**.

In fact, I may have only one link at all – from the New York Times – but, they get a lot of traffic, and if I have a prominent link, then I'll get a lot of value from it. Perhaps this would result in people linking directly to my page after that, but for the moment lets assume that they don't: either my URL is too long, or it has an annoying "~" in it, etc. If my URL is a sub-page of a website, then people may link to the main page and not necessarily to my sub-page. Not only does this protect the link from breaking if the site gets rearranged, but it also means that visitors will be able to follow the link to the main page and then use whatever tools are available there – search, navigation, sitemap, etc. – to find the actual page they are looking for. A referrer who is savvy enough to do just that is another kind of valuable referrer because they are helping me to insure that incoming visitors to my site can reach their destination appropriately.

The Effect of Closeness

The “closeness” of a referrer can also affect the value of any incoming links it provides. Again drawing an analogy from social network theory, the influence of my close friends differs from the influence of those more socially distant from me. This influence is not always greater.

This difference arises from something in network parlance called clustering: when a group of nodes is highly interconnected. Thus we can speak of nodes as being within my cluster or outside of it. In his seminal paper entitled “The Strength of Weak Ties” (Granovetter, 1973, in Scott, 2000, p. 35) Mark Granovetter studied how people get jobs. Granovetter found that because my friends are likely to know each other and have access to the same information, it is uncommon for them to be the source of new job leads. “It was the short, weak chains of connection that were of greatest significance in the receipt of useful job information” (Scott, 2000, p. 35). Similar findings exist in “The Search for an Abortionist” which found that women tended to acquire information about where to find an abortionist from persons outside of her typical circle of friends and family (Lee, 1969, in Scott, 2000, p. 35).

On the WWW, clusters often exist between websites that have similar goals, such as environmental activists, and just as often those activists are already aware of each other’s activities and information sources. In such clusters, a link from outside of the cluster is a much better source of incoming traffic and further links, than a link from inside the cluster.

As it turns out, Granovetter also studied why crowds riot, and created the “threshold model” for collective behavior. Extending Granovetter’s work, Duncan Watts has found that in the case of rioting crowds, the influence of those that are “close” may have a greater effect than the influence of those that are “distant.” “As in Granovetter's model, the decision to adopt, riot, or sign is determined by how many other nodes have already chosen to do so, except that now each node only pays attention to its specific set of 'neighbors' -- the nodes whose decisions influence it” (Strogatz, 2003, p. 264).

On the WWW, this kind of qualitative influence happens among academia, for example. A new phenomena like “cold fusion” may get links from the media, but if it fails to garner links from the scientific community, i.e. from within its valued cluster, then the incoming links that it does retain won’t add up to significant value. Echoing Watts, the decision to accept cold fusion depends on the number of academic nodes who do so. Value clusters occur among friends’ sites, graphics sites, conspiracy sites, advocacy sites, etc.

Dynamic Quality:

Networks are not static; they evolve over time, and Barabasi sums up the key network dynamics as growth and preferential attachment. He makes no mention of node or link “death,” at this point, but he does note later that now “we understand that internal links, rewiring, removal of nodes and links, aging, nonlinear effects, and many other processes affecting network topology can be seamlessly incorporated into an amazing theoretical construct of evolving networks” (Barabasi, 2003, p. 90). Unfortunately, if the “amazing theoretical construct" is anything like Barabasi's, it is predicated on some jarring ontological assumptions, as we shall see.

The Myth of Fitness

"Fitness is in the driver's seat, making or breaking the hubs" (Barabasi, 2003, p. 97). Barabasi conjures up the shibboleth of "fitness" to explain why some nodes gather more links over time than others. He claims that "independent of when a node joins the network, a fit node will soon leave behind all nodes with smaller fitness. Google is the best proof of this: A latecomer with great search technology it acquired links much faster than its competitors, eventually outshining all of them" (Barabasi, 2003, p. 97).

However, this is a tautology. Which nodes gather links? The ones that are fit. Which ones are fit? The ones that gather links. Because there is no independent criteria for fitness, the definition is circular.

"We have an additional preferred disposition to assume that the actual outcome was somehow *preferred* over all other possibilities… And not simply because greatness [fitness]… is hard to judge, or often misunderstood, but because it is almost never solely an intrinsic property in the first place" (Watts, 2003, p. 245).

The same observation holds for the growth of cities. "Large cities like New York, therefore, are more likely to attract new arrivals than small cities like Ithaca, thereby amplifying initial differences in size and generating a power-law distribution in which a few ‘winners’ account for a disproportionately large share of the overall populate. In reality, there is nothing random about New York being larger than Ithaca – New York lies at the mouth of one of the major rivers on the eastern seaboard, whereas Ithaca nestles in the midst of a sleep farming community…. The point is that once an individual city, business, or Web site becomes large, then *regardless of how it did so,* it is more likely than its smaller counterparts to grow larger still.” (Watts, 2003, p. 110)

Thus, it is external factors that lead to the growth of nodes in a network, not an intrinsic “fitness.” As one scholar has noted, insofar as google benefited from word-of-mouth advertising, the “fitness” of Google may have been how catchy it is when you tell your friends[4].

The New York Times has an even more obvious source of fitness that has nothing to do with its linkage: it is arguably the most highly visible newspaper in the real world. People go to the New York Times *online*, because people go to the New York Times *offline*. Once a website has high value and gains high presence it doesn't *need* incoming links. Even if the New York Times had no links to it at all, it would still get visited, particularly if it continued to appear in search engines. Furthermore, one only needs to locate it once to be able to visit it again. I would not be surprised to find any number of highly trafficked sites that do *not* have a large number of incoming links, just because, as I mentioned earlier, people usually have those sites bookmarked or just type in the URL.

Competition is Cooperation

"The fitness model allows us to describe networks as *competitive* systems in which nodes fight fiercely for links" (Barabasi, 2003, p. 106). The idea that competition allows the fittest to survive has largely been abandoned in the biological sciences because it fails to accurately describe the dynamics of evolution. It is no less inaccurate on the WWW.

"Although individuals do themselves frequently compete, the process of speciation (which begins at the level of populations) frequently involves greater specialization and is thus itself a mechanism for *avoiding* competition"(Seegert, 1998, p. 32).

So, for example, businesses – who do not see themselves as collectively supplying a variety of widgets to mankind – do not link to their competitors' websites, but academics – who see themselves as a population with common goals – do. It all depends on *whom* you think you are competing with. For instance, I run a site on the web about Panarchy (http://www.panarchy.com), but the few other "panarchy" sites that are out there are not in competition with me per se because they discuss different definitions of the word than panarchy.com does. In effect, my site is unique, and as such, is in competition with no one else in its niche, because there are *qualitative* differences between our sites.

As I noted earlier, many communities online see themselves engaged in a cooperative effort, and not in competition with each other at all. As in biology, in these cases, fitness is a group attribute and not an individual one. Even the connection between individual fitness and group fitness is not enough, however, to explain the growth of hubs. For that we must question fitness itself, and if fitness is not an intrinsic property, then where does fitness fit in?

Survival of the Fitters

"Being fit often means fitting in"(Seegert, 1998, p. 26).

"*The seed alone is not enough*….only one in very many will grow to fruition, and not because that one bears some special, unique quality, but because *lands in the right place*"(Watts, 2003, p. 248-9).

So, fitness is not a property ascribed to an individual agent, but rather, complex webs of agents not only "fit" together, but also "fit" with their environment. Although the two contextual variables are still usually aggregated, together they constitute a **fitness landscape** into which individuals adapt. The fitness landscape itself also changes over time. This model is descriptively accurate whether the system is an economy of interacting firms or an ecology of interrelated species (Kauffman, 1997). In focusing on individual nodes on the WWW, Barabasi has missed the constitutive effect of their interrelations. Duncan Watts sheds some light on why, when he recounts that the study of some networks requires us "to think simultaneously about two distinct kinds of structure – social structure and network structure – rather than just one. Of course, this view is entirely natural for sociologists… Sociologists have thought long and hard about the relation between social and network structure. But it is not at all natural for physicists or mathematicians, to whom the idea of a network node possessing an identity sounds vaguely ridiculous"(Watts, 2003, p. 116).

But websites *do* have identity. Some are content, some are tools. Comparing http://www.w3.org to http://www.amazon.com is misleading, like comparing books to bookstores. There are few bookstores but many books. The W3C site is a source of information; Amazon is a tool, a bookstore, its focus is not content but *pointers* to content. Quantitative analysis conflates different kinds of websites because it does not distinguish the qualitative differences between them.

The **co-evolution** of agents with their ever-evolving fitness landscape is a process called **structural coupling**, and can ultimately work *against* the agents' fitness. Agents that are too tightly coupled to their fitness landscape become vulnerable: if their network goes down, they go down with it. Online, if the only way to get to my page is from a collection of links called Bob's List, and Bob's List goes down, then suddenly my page is stranded. My webpage is too tightly coupled and becomes extinct. Per Bak has suggested that this is a natural process and that over time structural coupling must necessarily result in disequilibrium, with entire webs of agents becoming extinct periodically via punctuated equilibrium (Per Bak, 1996).

Moreover, fitness landscapes themselves are embedded within other fitness landscapes. We can characterize fitness landscapes along a continuum: local, regional, global, etc. In addition we can speak of the selection pressures acting on agents and groups as coming from inside (endogenous) or outside (exogenous) the system. As a result, the term "fitness" is revealed as a complex of interacting constraints acting to assist agents and groups to more effectively couple with their landscapes. This may result in "local fitness," but not always "global fitness."

Innovation Changes the Landscape

But landscapes change. In fact, in any ecology, agents are always simultaneously adapting to and altering their fitness landscape (Kaufmann, 1997). For most agents in a network, "fitness" is "fitting to" the landscape, and innovators do no such thing. Innovative websites, like Google, do not fit with the landscape; they redefine it. As a result, you cannot explain their successes by relying on the "fitness" tautology.

Furthermore, even so extreme a success as a winner-take-all success, still does not constitute grounds for claiming "fitness." As Brian Arthur has noted, survival in a network often contains an element of "lock-in" that enhances a single agent for no particularly good reason (Arthur, 1997). The typical example of lock-in is the QWERTY keyboard, an arbitrary configuration of keys that succeeded only because *it became the standard*, after which change becomes exceedingly difficult. Other examples would be the VHS vs. Betamax war over VCR standards, and even the effect of the invention of the crossbow on armored knights and the medieval world. The adoption of the VHS standard made Betamax extinct; the invention of the crossbow (because it could penetrate armor and be fired by a peasant with no knightly training) made armored knights and the social system that relied on them extinct. By changing the landscape, *innovators alter the criteria by which "fitness" is measured*.

Many innovative technologies have been adopted not because they are fit, but because no alternatives are available or allowed. For example, "in 1926, General Motors began the systematic purchase and destruction of trolley lines across the country, and by 1950 it had replaced street cars with its own buses in more than 100 cities" (Oppenheimer and Boyle, 1990, p. 117). Also, it is no secret that the oil industry has maintained a long battle against the adoption of alternative energy sources. For innovations to succeed they *require* a parallel change in the fitness landscape, either by disruption from within, or by manipulation from without. So we must return to Watts' point: we *want* to believe that fitness is at work, but such a proposition cannot resist scrutiny.

This is not to say that the landscape is infinitely malleable however. After the eruption of the volcanic isle of Krakatoa, only certain species had a chance of acquiring a foothold there. If polar bears had arrived, they would have had no hope of out-competing local fauna. However, continental drift and ice ages both have had substantial impacts on local ecologies. Landscapes have their own constraints: trolleys and busses are alike enough. Nonetheless, network contexts, even online, are impermanent.

Limits To Growth

"Network ties in Barabasi and Albert's model are treated as costless, so you can have as many of them as you are able to accumulate, without regard to the difficulty of making them or maintaing them"(Watts, 2003, p. 113).

There are limits to growth as well. Cities that grow too large begin to attract fewer people. And in social networks, "there are limits to the amount of time that people can invest in making and maintaining relations… Agents will, therefore, decide to stop making new relations… when the rewards decline and it becomes too costly" (Scott, 2000, p. 75). "As the network grows, the rich get richer for a time, and hubs do emerge. But eventually the most highly linked elements begin to lose their advantage in gathering new links" (Buchanan, 2002, p. 124). Moreover, "whenever limitations or costs eventually come into play to impede the richest getting still richer, then a small-world network becomes more egalitarian, as seems to be the case with the airports and a number of other real-world networks." (Buchanan, 2002, p. 125). Barabasi treats the rich-get-richer phenomena as an end-state in a network, i.e. the emergence of a power law is constantly in operation with no homeostatic mechanisms at all. He describes a network with positive feedback, but no negative feedback. In the real word however, negative feedback keeps power laws from destroying the very systems they help to create.

Alternative Norms to "Preferential Attachment"

"When choosing between two pages, one with twice as many links as the other, about twice as many people link to the more connected page" (Barabasi, 2003, p. 85). Barabasi never questions *why* this is the case. More importantly he probably doesn't think it is relevant. In fact, it is crucial.

Preferential attachment is a norm, a rule that influences how people behave. *People* are creating links on the WWW, and those same people *believe* that the more connected page is the one they should link to. They believe that "more" is "better." The complex web of norms – the norm landscape – that people follow *offline* affects what they do *online*. In the presence of an alternative normative landscape, offline behavior is different, and online behavior would be as well.

“Gift culture” is a term describing various societies where giving things away is an operative cultural norm. In many such societies, if you visit someone and admire something, they will likely give it to you. In addition, many gift cultures have ritualized ceremonies and practices wherein individuals give away some or all of their possessions. These are typically deeply held beliefs and the norms they engender carry great weight. When one elder of a tribal gift culture was asked why they behave as they do, he replied, “A poor man shames us all” (Maybury-Lewis, 1992 , p. 85). Gift-cultures are not uncommon and range from small-scale tribal societies to global religions. In contrast to more-is-better thinking, gift-culture norms could be summed up as more-is-unfair. Oddly enough this norm shows up even in American society where it emerges as “rooting for the underdog.”

A WWW fully permeated by gift-culture norms would look a lot different than the one we have now. We might find a tendency to link to the “underdog,” or the page with the least links, just to increase their visibility. I use this example because we do in fact find precisely this kind of culture online. It manifests when websites agree to “mirror” information. Members of a sharing community will create servers around the world to provide access to information through multiple nodes, thus avoiding overload on a single centralized server. Since *all* of these links get followed by a robot, they introduce distortion into the resultant topology. In addition, when a new but valuable site comes online, it must, by definition, have very few incoming links. In order to gain links, *someone* must be linking to it while it is still relatively invisible. It is only *after* a visibility threshold is crossed that the page begins to seem highly linked. If it weren't for people linking to unlinked pages, new pages would never gain any links, and the WWW would have an absolute barrier-to-entry. Thus, we would never see any new pages at all.

Conclusion: Getting it Right

My point is not that Barabasi didn't measure everything, or even that he didn't measure the *right* things. He could hardly be expected to produce a complete analysis of the WWW in one fell swoop. The point is that *what* is measured and *how* it is measured are decisions that are value-laden. Barabasi would have us believe the WWW *is* x, when in fact it is only x *in relation to* the way it was studied. The WWW does not exhibit a power law, but only certain measurements of it exhibit a power law. "The method is very impressive and produces lots of elegant results, but it doesn't solve the actual problem *because it's not about the actual problem*"(Watts, 2003, p. 67).

Furthermore, The Web does not exist in isolation. It piggybacks on a pre-existing social, political, and economic structure called the "real world." All too often, scientists forget this. "For any complex system, there are many simple models we can invent to understand its behavior. The trick is to pick the right one. And that requires us to think carefully – *to know something* – about the essence of the real thing"(Watts, 2003, p. 304). This leads Watts to call for a "taxonomy" of networks to helps us to understand which properties are relevant under what circumstances. If quantitative analysis has limits on the WWW, it fails with other networks as well. Local dependency and betweenness reveal that something need not be *well-connected* to be *crucial*. There are other measures of value besides a simple connectedness.

We find many statements like "documents with only one incoming link have less than a 10 percent change of being noticed by any search engine" (Barabasi, 2003, p. 174). It is not surprising however that Barabasi makes such statements. He fully admits his belief that "most of the Web's truly important features and emerging properties derive from its large-scale self-organized topology" (Barabasi, 2003, p. 174). But topology isn't everything, and even defining topology is problematic. There are many different kinds of topologies differently measured. A link topology is merely one kind. I've suggested a "use" topology as an alternative. This alone is not sufficient however, for a use topology would have further subdivisions and characteristics based on gender for example, or age. It is well known that pornography sites target young males. We need to be careful with *what* and *how* we study networks online. "Researchers must take great care over the nature of their relational data. They must, in particular, be sure that the level of measurement used is sociologically appropriate… Mathematics itself cannot provide an answer for the researcher"(Scott, 2000, p. 49).

The "best" book is not necessarily a bestseller. The "best" actor is not necessarily the most popular. In fact, often the inverse is true. The best computer or the best car is often less popular due to high prices or other factors. Moreover, the use of prizes and awards, academic certifications, and wealth as a reward for success, all attest to the social value we recognize in elites, a value that is represented and reproduced by the social roles we entrust to those elites, everything from Nobel prize winners to CEOs to political leaders.

To be sure, popularity has its place, but so does other kinds of quality. More is not automatically better.

"Barabasi analyzes the web as an observer, not as a participant"[5].

The notion that inquiry can best be pursued by an "objective observer" has been at the heart of philosophy of science debates for some time. The dogma of Science insists that a positivist methodology is the only path to Truth, and in particular that quantitative analysis is the only valid mode of inquiry. By contrast, almost all of the insights in this paper stem from my own subjective experiences gathered while participating in Internet industry for many years. While this may not yield Geertz's "thick description," it does provide a thicker description than a reductionist quantitative analysis or a robot topology (Geertz, 1973). And as Geertz and others assert, the value of the inside view is that it helps us not merely to "explain" phenomena, but to "understand" them (Hollis, 1994).

Barabasi, like most network researchers, gets more things right than he gets wrong. But as I've stated time and again more is not better. The *quality* of his misleading or erroneous statements is significant. But sometimes "close" is worse than flat wrong. Descriptive conclusions get used to yield normative advocacy, we use what *is* to decide what we *should* do. Incomplete or misleading information can be dangerous, particularly when popularly consumed. When civilization follows a "pretty close" idea, it often ceases to pursue the "really good" idea. "We have ninety-eight years to succeed at this, and make the twenty-first the century of complexity." (Barabasi, 2003, p. 226) Yes, lets, but lets do it right.

About the Author

Paul B. Hartzog is a futurist and instructor at the University of Utah. His primary research interests are post-modernism, Complexity, networks (ecology, internet, social, etc.), and the philosophy of science.

Acknowledgements

It is essential first and foremost to thank Albert-Laszlo Barabasi for his riveting and engaging work. It provided both the substantive and motivational impetus for this paper. I would also like to acknowledge the members of the Howard Rheingold's Brainstorms community for their feedback, in particular Burl Grey and Jim Lai. I would like to thank Alf Seegert for our many fruitful discussions over the years. And a final note of appreciation goes to Dr. Mark Button at the University of Utah for his unwavering commitment to his students' critical self-reflexivity.

Notes

1. UrbanDictionary entry for "google bombing" at http://www.urbandictionary.com/define.php?term=google+bombing&f=1, accessed 23 April 2004.
2. Lai, Jim, personal communication, April 2004. This is entirely Jim's point, and I am indebted to him for mentioning it.
3. McNichol, Tom, "Engineering Google Results to Make a Point," New York Times, Published: January 22, 2004, online at

http://www.nytimes.com/2004/01/22/technology/circuits/22goog.html?ex=1081141200&en=06775aba6a9568ed&ei=5070, accessed 23 April 2004.

4. Seegert, Alf, personal conversation, April 2004.

5. Seegert, Alf, personal conversation, April 2004.

References

Brian Arthur, 1988. "Self-Reinforcing Mechanisms in Economics," In P.W. Anderson, Kenneth Arrow, David Pines (editors). *The Economy as an Evolving Complex System*. Redwood City, CA.: Addison-Wesley, pp. 9-32.

Per Bak, 1996. *How Nature Works: The Science of Self-organized Criticality*. New York.: Copernicus.

Albert-Laszlo Barabasi, 2003. *Linked : How Everything is Connected to Everything Else and What it Means for Business, Science, and Everyday Life*. New York.: Plume.

Mark Buchanan, 2002. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York.: W.W. Norton.

Clifford Geertz, 1973. *The Interpretation of Cultures*. New York.: Basic Books.

Martin Hollis, 1994. *The Philosophy of Social Science*. Cambridge.: Cambridge University.

Stuart Kauffman, 1997. "The Evolution of Economic Webs," In P.W. Anderson, Kenneth Arrow, David Pines (editors). *The Economy as an Evolving Complex System*. Redwood City, CA.: Addison-Wesley, pp. 125-146.

David Maybury-Lewis, 1992. *Millennium: Tribal Wisdom and the Modern World*. New York.: Viking.

Michael Oppenheimer and Robert Boyle, 1990. *Dead Heat: The Race Against the Greenhouse Effect*. New York.: Basic Books.

Alf Seegert, 1998. *Ontology Recapitulates Ecology: The Relational Real in Evolution and Ecophilosophy*. Master's Thesis, University of Utah.

Steven Strogatz, 2003. *Sync: The Emerging Science of Spontaneous Order*. New York.: Theia.

M. Mitchell Waldrop, 1992. *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York.: Simon & Schuster.

Duncan Watts, 2003. *Six Degrees: The Science of a Connected Age*. New York.: W.W. Norton.