

The Web as human network

Eric Blair

14 May 2005

I'd like to discuss the question of how technology has changed personal relations. That'll come next time. For now, let's look at a specific, vaguely related question: does the link structure of the Net mirror the link structure of human networks?

Back when Alta Vista was the highest view in Internet search, a few IBM and Alta Vista researchers did a rather detailed study of the Web's structure [Broder et al.(2000)Broder, Kumar, Maghoul, Raghavan, Rajagopalan, Stata, Tomkins, and Wiener]. They, as with many others, found that the distribution of links on the Net looked a lot like the distribution of human links. There is a power law distribution where there are a few sites that are linked endlessly, and a long tail of sites that only have a few links.

To give an example of a power law, here is a graph based on data from junior high classes. The most popular student is on the X-axis at the far left (at $X=0$), and was nominated as a best friend by a mean of 9.75 other students (over 88 classrooms in the sample). Over on the other end of the X axis, the 25th through 35th ranked student in the classroom was nominated as a best friend by a mean of less than one other student. So you've got a few very well-connected students and a lot of students who have no connections at all.

We see this pattern in social networks of all scales, and among Web pages. The nomination count graph is typically a little more curvy than this one, with even more of a steep slope down from the most popular members of the group and a longer tail at the other end.

It sounds like the WWW as interpersonal network metaphor is working OK, but two caveats: first, there is much debate as to whether the best fit for the link distribution of the Web is a Negative Exponential, a Gamma, a Zipf, or a variety of other distributions that all look identical to a non-expert. Unless you hope to study this stuff seriously, you don't have to care about this caveat and can just call it a power law. The best fit to the student data is a Gamma distribution, by the way.

Second, human networks are pretty symmetric, in that there are few face-to-face contacts where one party is ignorant of the other. This is true of celebrities, whom we know but don't know us, but we can throw those out and have a reasonably symmetric set of acquaintance links. The popular kids may not want to hang out with the unpopular ones, but they know them nonetheless.

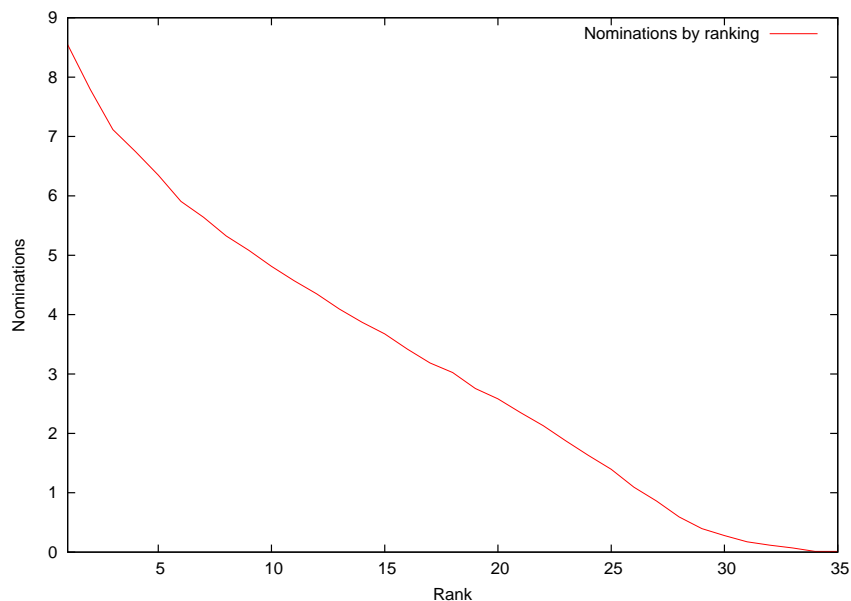


Figure 1: Junior high class photo. That's me on the far right.

But with Web pages, it happens all the time that a page makes no indication of what other pages are linking to it.

Broder et al found that this asymmetry occurs on a grand scale. They divide the Web into a giant Strongly Connected Component (SCC) comprising about a quarter of the Web; these are sites that interlink with each other. Then there's a quarter that only links in to the SCC but does not receive links. That would be blogs from losers like me. Then there's a quarter that is linked from the SCC but does not link to anything in particular, comprising corporate sites that just go in internal circles and things like online books and manual pages that are informative but not filled with links. The final quarter, they called *airq*, indicating a trail of limited links that doesn't readily fall into the first three categories. Thus, because a web page is not a person, the symmetry of human networks does not map to web links.

Another important distinction is that the whole small world game, where we try to find a chain of people from a guy in Katmandu to a guy in Omaha, does not work for the Net, because if you start on the right side of the bowtie, you can not get to the left side. For humans, you can almost certainly find a chain, and it'll be well under ten people in almost all cases; for the Net, you only have about a 25% chance of being able to form a chain from any randomly selected site to any other randomly selected site. E.g., try getting from your favorite online manual to your favorite friend's blog. When you can form a chain, say from the in-feeding region to the SCC region, then it can still be hundreds of nodes long

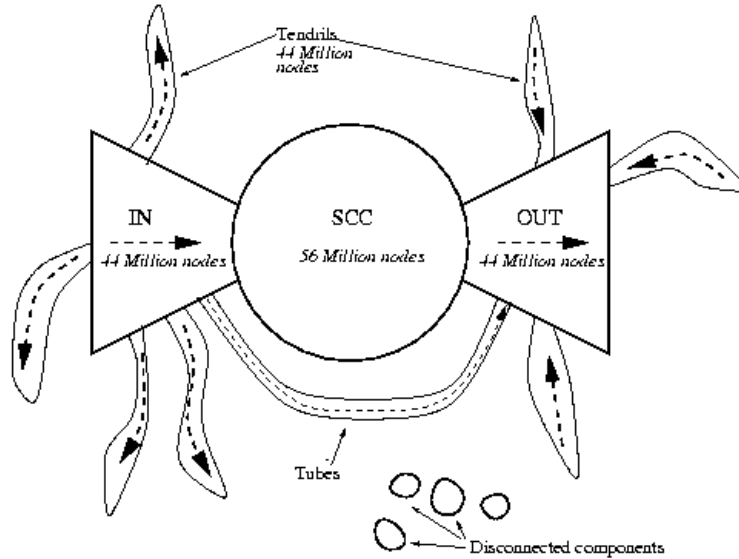


Figure 2: The Insidious Bowtie of Nyrothænim, aka The Internet.

if one element is well-buried in a subculture.

Now, with human networks, we can distinguish between acquaintance, which is almost by definition symmetric, and friends, which is depressingly unidirectional, typically from low-status to high-status. I don't believe this metaphor is particularly well-studied, but it doesn't work very well. The net receivers of links for the Net are not high-status pages, but pages that just provide information (corporate, technical, whatever).

But getting back to the part of the metaphor that does work, there are two characteristics to both networks. First, there's a cost to linking both socially and online, because you need to find the subject of your interest and know them. Second, there is a cost to searching for new links. An immediate corollary to expensive search is a principle that the rich get richer: the easiest way to find new links for your own personal address book is to ask others for their contacts, so well-linked people/sites are more likely to get more links.

References

- [Broder et al.(2000)Broder, Kumar, Maghoul, Raghavan, Rajagopalan, Stata, Tomkins, and Wiener]
 Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer Networks*, 33:309–320, 2000.