Commentaries on Albert-László Barabási's books

Networks101Link3.1   ©frederick david abraham, 13 January 2013

So far we have seen the pioneering work or Erdős & Rényi and Rapaport that established how the bifurcation of clustering as a function of the number of links in a random network. This work highlighted not only the importance of the number of links, but also the importance of the difference in random and biased placements of the links. The next major development highlighted the importance of the lengths of paths in paths, and again the difference between random and biased paths. This chapter (Link 3 in Barabási, *Linked,* B.L.3 format from now on) includes the classic work of Milgram's small worlds/6-degrees studies, and Barabási's research on path lengths on the World Wide Web (Milgram, 1967; Travers & Milgram, 1969). Milgram's study is familiar to all, but remember that the critical part was to forward a letter to a target important person if known, or to a personal acquaintance who was more likely to know the target. Since all individuals from the whole population of the Unites States were available as nodes, it was expected that the paths the letters took to reach their target would be long. They were not. The results have reported that the median path length was 5.5 or 6 .

36 CHAPTER 2. GRAPHS

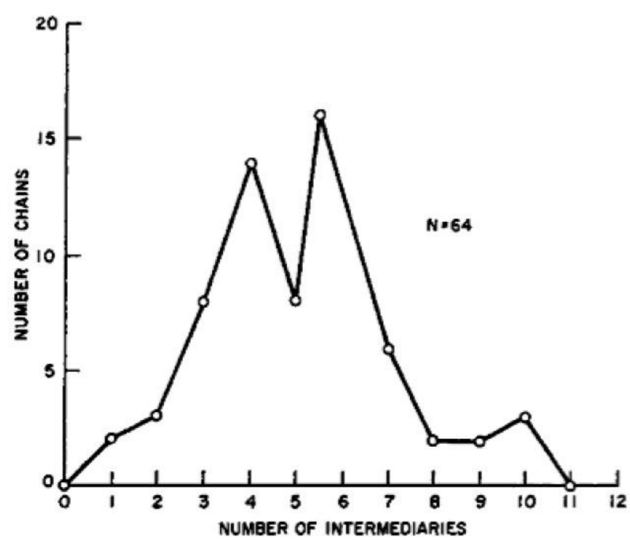

Figure 2.10: A histogram from Travers and Milgram's paper on their small-world experiment (391]. For each possible length (labeled "number of intermediaries" on the x-axis), the plot shows the number of successfully completed chains of that length. In total, 64 chains reached the target, with a median length of six.[i]

I was looking at their graph and from it I computed the median of 5, and a mean of 5.16. It seems that Karinthy was right! Barabási's guess that Karinthy's story "Chains" may have in part motivated the study, note their use of that term in the figure. Barabási credits Rapoport with providing the basis of the technical aspects for Milgram.[ii].[iii] Milgram heavily references Rapoport.

The phrase 'six degrees of separation' arose in popular culture and viraled there as meaning everyone on the planet can be connected to any other by 6 links or less, clearly a preposterous interpretation of the research. Obviously, any two individuals in large and densely populated networks can have paths of astronomical lengths, but the interesting thing about the research is that the shortest paths can be amazingly short. It is also clear that the path lengths provided by the research are but a few of the many, and are very difficult to interpret as to their ability to represent the population of paths between particular nodes. The important thing is that the work pointed to the need for measures of path length be added to degree of node to understand the flow of information through a network, and to understand the evolution of networks. And clearly there will be very different path lengths for random networks and those that have a bias, as provided by the instructions to subjects in Milgram's experiments.

The study of social and communication paths began in earnest with studies of the Web based URLs in a given document to another document, aided by crawlers, automated web searchers, which allowed analysis of size and path lengths of the Web. To do this with the limited computer capacity at Notre Dame, Barabási and his colleagues, Albert and Joeng, did a very clever thing, starting with a sample of 1000 nodes, then a sample of 10,000, and larger samples until they exhausted their computing capacity. Then they worked out the average (I assume shortest) path length between pairs of nodes at each sample size, and then extrapolated to the size of the entire net, which turned out to be ***d = 0.35 + 2logN***. That is, ". . . the average separation between nodes increased much more slowly than the number of document. With 800 million nodes on the 1998 Web, "Thus our expression predicted that the diameter of the Web was 18.59" or close to 19 clicks apart. I got 18.15, and if I understand it correctly, I think he carelessly slipped in calling it 'diameter' which is the maximum shortest path length of a network. With the network today at 12.37 billion nodes, d = 20.35 by my reckoning (check it for me, please).[iv]

Barabási next asks the question, "How do networks achieve such a uniformly short path despite consisting of billions of nodes?" For the answer he turns to the critical parameter responsible for the bifurcation to giant cluster of Erdős and Rényi's random networks, namely, the average degree of the nodes, the number of links to nodes. Again the collapse-bifurcation depends on logarithmic relations. The dependence of the average separation (shortest path length) upon the size and average degree is given by ***d = log N/log k***. So now we have two parameters that determine major properties of networks, which suggest the complex of possibilities for the analysis of such networks. To this we add another he mentioned without definition, diameter, the length of the longest of the shortest paths between two nodes, and summarize a few for networks to which I had easy access.

Table of basic statistics of a few networks. Computed using Gephi (except WWW).

Network	$\bar{k}$	$\bar{d}$	D	nodes	edges
c elegans (d)	7.663	3.992	14	306	2345
c. elegans (u)	7.663	2.455	5	306	2345
val (d)	11.619	2.523	8	307	3567
val (u)	11.619	3.487	9	307	3567
fred (d)	1.438	1.497	9	80	115
fred (u)	1.438	2.907	6	80	115
lesmiserables (d)	3.299	2.400	5	77	254
lesmiserables (u)	3.299	2.641	5	77	254
WWW Jan2013		20.35		1.35E+10	

$\bar{k}$ = average degree   d = directed network interpretation
$\bar{d}$ = average path length   u = undirected network interpretation
D = diameter

C. elegans	nervous system, parasitic nematode
Val, Fred	FaceBook networks
Les Miserables	Characters in the 19th Century novel by Victor Hugo

Note that for *C. elegans* the value reported by Barabási *ibid*. p. 34 seems to be what Gephi reports for me as the diameter for the directed network interpretation. Please confirm or correct me.

End Notes

[i] From: Easly & Kleinberg, 2010, p. 36 and Travers & Milgram, 1969). Note that the paths the each letter took was not necessarily the shortest available. In their 1969 article, An Experimental Study of the Small World Problem", Travers & Milgram, *Sociometry, 32(4),* pp. 425-443), comment on the possibility of a dual distribution accounting for the bimodal appearance of the figure "The mean of the distribution is 5.2 links. . . Figure 1 concealed two underlying distributions: when the completed chains were divided into those which approached the target through his hometown and those which approached him via Boston business contracts, two distinguishable distributions emerged. The mean of the Sharon distribution is 6.1 links, and that of the Boston distribution is 4.6." pp. 431-432. This footnote updated 28 November 2015.

[ii] Barabási, ibid, endnote to p. 29, p. 247. With mine or their computations, this is five, not 6 degrees of separation.

[iii] Rapoport, A. (1957). Contribution to the Theory of Random and Biased Nets. *Bulletin of Mathematical Biology* 19, 257-77. "Anatol Rapoport was an early developer of social network analysis. His original work showed that one can measure large networks by profiling traces of flows through them. This enables learning about the speed of the distribution of resources, including information, and what speeds or impedes these flows—such as race, gender, socioeconomic status, proximity andkinship.[4] This work linked social networks to the diffusion of innovation, and by extension, to epidemiology. Rapoport's empirical work traced the spread of information within a school. It prefigured the study of Six degrees of separation, by showing the rapid spread of information in a population to almost all—but not all—school members (see references below). From http://en.wikipedia.org/wiki/Anatol_Rapoport, 14 January 2013. Rapoport was an activist and theorist on social issues of conflict and human rights, and a cofounder of the Society for General systems Research. It is a pity that in our shared time at the University of Michigan we were not as involved in these issues and were unaware that Rapoport was their at the same time.

[iv] Barabsiá quotes and calculations, ibid, pp. 33-34. The current WWW size I got by browsing the web; some sites give daily or near daily results. I also found that Googling the log x, x any number, show the result immediately under the query. Nifty. Log x of course being base 10.