VIEW FROM THE PENNINES: MAKING CONNECTIONS

PAUL GLENDINNING

Our village prides itself in its isolation: visitors can buy T-shirts showing the name of the village, a cartoon sheep, and the words *Back O'Beyond*. Even so I can find a connection from myself to my mother through the village involving only three steps (i.e. I know someone in the village who knows someone who knows my mother). Since my mother moves in London literary circles I find this remarkably short. Trying a different tack, I wondered how far I am from having played football with David Beckham. Again the answer (I have played football with someone who has played football with...) seems surprisingly low: four. These are examples of the 'small world' phenomenon that social scientists have been studying for over forty years: social connections lead to remarkably short average path lengths between individuals.

The last five years has seen mathematics beginning to make headway in explaining this empirical observation (see the recent books [6, 15]). Mathematically a social network, and many other networks, can be thought of as a graph: it has n vertices representing people, with edges connecting two people if they are friends. It is usual to ignore the unrequited angst of non-reflexive friendship and so the edges are not directed, although such complications can be added. Thus there are at most $M = \frac{1}{2}n(n-1)$ edges since each of n vertices can be connected to at most $n-1$ other vertices, and so there are $n(n-1)$ ordered pairs of vertices which could have edges, and the factor of $\frac{1}{2}$ reflects the fact that the edges are not directed. In practice, the number of edges (friendships) is much less than the theoretical maximum M and it is natural to associate with each vertex i the number of edges which end at i, which is called the degree of i, k_i. The average degree of a graph is denoted by k, so if the total number of edges in a graph is m then $m = \frac{1}{2}nk$, and the 'few friendships' assumption, $m \ll M$, implies that $k \ll n$. One final definition: the path length between two vertices is the number of edges on the shortest path of edges connecting the two vertices.

Models of such graphs are not new. In the late 1950s Erdös and Renyi showed that random graphs, where each of the M possible edges occurs with probability p, are small worlds in the sense that average path lengths are short. Extra levels of complexity have been added more recently to model real networks. Watts and Strogatz [15, 16] suggest that a small world network should have both short average path lengths and a clustering effect (the probability that two of my friends are friends of each other is higher than the random probability), whilst Barabási and others work with distributions of degrees which reflect empirical measurements of networks. These ideas have been used to model connections on the internet [4], calls between telephones [1], the interaction network of proteins in yeast cells shown in Figure 1 [10] and even terrorist organisations.

There are two main variants of the standard Erdös-Renyi random graph; one has a probabilistic flavour, the other is more combinatoric [7]. In the probabilistic model each of the M possible edges occurs with probability p and hence the expected number of edges is pM. In the combinatoric model the probability of a graph with n vertices and m edges

Date: February 18, 2003.

is assigned by considering the space of all possible graphs with n vertices and m edges, so every graph in this model has precisely m edges. The results for the two model are similar if $m = pM$, and the changes in the properties of the graphs as p (or $m = pM$) increases is truely remarkable. As $n \to \infty$ the following facts about the graph hold almost surely. If m is much smaller than $\sqrt{n}$ the graph has only isolated edges, but trees of size three (i.e. involving three vertices) start appearing if m is of the same order as $\sqrt{n}$. As m continues to increase there is a cascade of values of m at which trees of larger size start to exist, until m is of the order of $\frac{1}{2}n$. Here there is a phase transition which heralds an amazing condensation process. Above this transition region a 'giant' connected component with size of order n exists, and finally the graph becomes fully connected when $m \approx \frac{1}{2}n \log n$. It is tempting to speculate that this hierarchy of graph complexity might be mirrored in the creation of more and more complicated molecules as some excitation parameter is increased.

Average path lengths in the giant component of a random graph model seem to scale roughly as $\log_k n$, which is indeed small compared to the size of the graph, justifying the name small world. Unfortunately the local uniformity of the the graphs created in this process does not fit the data of most networks which are being studied at the moment. In a pioneering paper which appeared in 1998, Duncan Watts and his research supervisor at Cornell, Steve Strogatz, introduced some graphs which have the same small world property as random graphs but also have strong clustering effects [16, 15]. They define the clustering coefficient, C_i, of vertex i to be the proportion of the $\frac{1}{2}k_i(k_i - 1)$ possible edges between the k_i vertices connected to i which actually occur in the graph, and the clustering coefficient of the graph, C, to be the average of C_i over all the vertices of the graph. They show that some of their graphs have short ($\log n$) average path lengths and a clustering coefficient reasonably close to one. They go on to consider the dynamics of diseases on such networks.

Watts and Strogatz originally avoid the problem of the creation of giant components in graphs by modifying connected graphs. More recent work returns to this problem using techniques from percolation theory, statistical mechanics and combinatorics [3, 11, 14]. The starting point in these new models is to assume that there is a given distribution of the degrees of vertices in a graph. That is, that the probability that a vertex i has degree $k_i = s$ is P_s, a function of s. In the classical random model the edges are generated independently with probability p, leading to a binomial distribution of the degrees of vertices. Measurements of a variety of systems suggest that a more complicated degree distribution is more appropriate. Bell Labs monitored the calls made and received by 53 million phone numbers called during a day and Aiello et al [1] show that the data is consistent with a power law distribution. Barabási has argued that this 'scale free' power law distribution should arise in a wide variety of contexts. Newman et al [14] use $P_s = Cs^{-\tau}\mathrm{e}^{-s/\kappa}$ to model data from scientific collaborations and the Kevin Bacon game – a network based on co-starring in films [9].

A striking feature of almost all these models is that there is a phase transition between fragmented graphs and graphs with a 'giant' component as the number of connections increases. A particularly clean example is provided by Molloy and Reed [11]. Roughly speaking they show that if $P_s \approx \lambda_s$, with $\sum \lambda_s = 1$, then (almost surely as $n \to \infty$) there is a phase

transition as the quantity Q defined by

$$Q = \sum_{s \geq 1} s(s-2)\lambda_s$$

passes through zero, with a giant component if $Q > 0$. So a giant component exists if not too many of the vertices has degree one.

At the moment the role and properties of graphs in self-organization, the spread of disease, strategies in game theory, biochemical interactions and communication networks are being actively pursued. The results look interesting. For example, the power law distribution of degrees in the graph of the world wide web (or the yeast cell protein interactions) suggest that these models are usually very robust to the removal (or attack) of vertices, but that there are a few vertices which carry a much more significant load, making the networks vulnerable to selective attack [5, 8, 10, 13]. I would love to have a picture of the evolution of the properties of graphs under dynamics: are there natural maps which can be defined on graphs which allow small world networks to evolve from other types of networks? As an intermediary question, is it possible to define non-trivial maps on graphs which preserve the degree distribution? Some steps have been taken in these directions [2], but almost all the current models involve dynamics in which the number of vertices increases with time. Given the scientific and technological importance of the applications it should not be long before a clearer picture emerges.

References

[1] W. Aiello, F. Chung & L. Lu (2001) A random graph model for power law graphs, *Experimental Mathematics* **10** 53–66

[2] W. Aiello, F. Chung & L. Lu (2002) Random evolution in massive graphs, in *Handbook of Massive Data Sets* eds. J. Abello, P.M. Pardalos, and M.G.C. Resende, Kluwer, Dordrecht, pages 97–122

[3] R. Albert & A.-L. Barabási (2002) Statistical mechanics of complex networks, *Reviews of Modern Physics* **74** 47–97

[4] R. Albert, H. Jeong & A.-L. Barabási (1999) Diameter of the World-Wide Web, *Nature* **401** 130–131

[5] R. Albert, H. Jeong & A.-L. Barabási (2000) Error and attack tolerance of complex networks, *Nature* **406** 378–381

[6] A.-L.Barabási, *Linked: the new science of networks*, Perseus Books, 2002.

[7] B. Bollobás, *Random Graphs*, Academic Publishing, London, 1985

[8] D.S. Callaway, M.E.J. Newman, D.J. Watts & S. Strogatz (2000) Network robustness and fragility: percolation on random graphs, *Physics Review Letters* **85** 5468–5471

[9] Kevin Bacon Game website (Oracle of Bacon): `http://www.cs.virginia.edu/oracle/`

[10] H. Jeong, S.P. Mason, A-L. Barabási & Z.N. Oltvai (2001) Lethality and centrality in protein networks, *Nature* **411** 41–42 (see also `http://www.nd.edu/~networks`)

[11] M. Molloy & B. Reed (1995) A critical point for random graphs with a given degree sequence, *Random Structures and Algorithms* **6** 161–180

[12] M.E.J. Newman (2002) Spread of epidemic disease on networks, *Physics Review E* **66** 016128

[13] M.E.J. Newman, S. Forrest & J. Balthrop (2002) Email networks and the spread of computer viruses, *Physics Review E* **66** 035101(R)

[14] M.E.J. Newman, S. Strogatz & D.J. Watts (2001) Random graphs with arbitrary degree distributions and their applications, *Physics Review E* **64** 026118

[15] D.J. Watts, *Small Worlds*, Princeton University Press, Princeton, 1999.

[16] D.J. Watts & S. Strogatz (1998) Collective dynamics of 'small world' networks, *Nature* **393** 440–442

Department of Mathematics, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom. E-mail: P.A.Glendinning@umist.ac.uk