Physicists and sociological network modelling: New methodologies of social network analysis and theories of social structure

Malcolm Alexander
School of Arts, Media and Culture, Griffith University
M.Alexander@griffith.edu.au

Abstract
Physicists have suddenly found social networks. Triggered by the modelling of 'small world architecture' presented by Watts and Strogatz in *Nature* (June, 1998), physicists, computer scientists, and mathematicians have produce a flood of materials on social, ecological, biological and biochemical networks. They have generated working models of network dynamics and complex network simulations with many applications. Publicists in the field talk of a new, comprehensive 'science of networks' with the potential to supersede current social theories. In this paper I describe how social network analysts have, independently, developed simulation models of social networks. For the first time, these techniques allow us to generate the array of all possible networks of the size and density of an observed network. This allows empirical sociological researchers to identify features of an observed network that are unusual and, thus, make probabilistic assessments parallel to those of standard statistics. We illustrate this technique on an example network of interlocking company directorates in Australia. In conclusion we suggest how this methodology would develop for this example and reflect on the interests of sociology and complexity science in this area of work.

Introduction: Physicists find social networks

In June 1998 Duncan Watts and Steve Strogatz (1998) published an article on network theory in *Nature*, one of the two leading international journals of scientific research. They described a particular network configuration comprised almost completely of regular close connections (a lattice) that make a network highly clustered combined with a extremely small scattering of random links that join points anywhere in the network. They dubbed this configuration 'small world network architecture'.

Watts and Strogatz situate their theory in the subfield of physics known as statistical mechanics. This is the field of physics which explains changes in the state of matter (phase transitions) as the outcome of the vast number of interactions and collisions between its constituent atoms and molecules such as occurs when water boils or turns to ice or gases condense. Watts and Strogatz wanted to explain how actions of independent actors become synchronised and produce a ‘phase transition’ from chaotic to synchronised collective behaviour. Watts’ original problem was to explain the synchronisation of crickets when they chirp (Watts 1999). A social behavioural example is the spontaneous synchronisation of clapping in a crowd (Watts 1999). Small world network architecture, they argued, was the configuration that gave the right mix of communication and interaction to allow for such spontaneous synchronisation.

Scientific interest in small world networks is excited by their ability to explain how stable systems of biological, ecological and social organisation emerge spontaneously from micro-level interactions (Buchanan 2002). Theories of emergence are a basic agenda item in the growing field of ‘complexity science’ and complexity theory and have applications in the analysis of the internet and world wide web, computer science, power distribution networks, transportation systems and also in biology, bio-chemistry and ecology. Barabasi (2002), another physicist, proclaims complex network theory as having new answers for studies of the spread of infectious diseases, particularly AIDS, the growth of social differentiation and inequality (‘the rich get richer’) and collective behaviour (mobs and riots). In sociological terms, network architectures can be seen as mediating mechanisms by which multiple, unregulated micro-level agency produces stable, enduring social structure[1].

Watts and Strogatz found ‘small world’ networks in human activity and in biological structures. They adopted the name ‘small world networks’ from Watts’ initial, sociological inspiration for modelling network structures with random links. This inspiration came from the urban mythology of ‘six degrees of separation’, the belief that seemingly short chains of acquaintanceship (‘friends of a friend, of a friend…’) might stretch from any individual to any other individual in the world popularised by Stanley Milgram (1967) in the 1940s. Watts has now moved to sociology and implemented a large scale, web-based version of the Milgram experiment (Watts 1999; Watts 2003).

To explore the characteristics of small world networks[2] Watts and Strogatz used computer simulation techniques common in statistical mechanics but used also in many areas of the social sciences (Gilbert and Troitzsch 1999). These simulations, also known as agent-based models, model micro-level interactions, set them in train as random processes and then see what mixtures and intensity of the modelled interactions produce steady state outcomes, breakdowns of these steady states and transitions between them.

Social network analysts had been developing computer simulations of complex social networks before networks became popular in the physics community. In this paper I describe a collaborative application of these social network simulations that allows me to directly examine the structure of the small world network of interlocking directorates[3] among Australian companies. Social network computer simulations provided by my collaborator allowed me to assess this observed network against all random networks of the same size and density providing a measure of probabilistic significance comparable to tests of significance in standard statistics. I find that the observed network tends toward extreme values in its clustering suggesting a tendency toward elite formation beyond that expected from the simulated networks.

In the concluding comments I suggest how collaboration between empirical social network investigations and modelling would proceed from this point and the complementary interests of sociologists and scientists in this type of research.

Computer simulation and statistics in social network analysis

The publicity created by Watts and Strogatz concept of small world networks spawned a flurry of interest and further work on complex networks by physicists. Because of their physics experience these researchers were comfortable working with very large and complex networks such as maps of the world wide web (Albert and Barabasi 2002; Barabasi 2002) or the complete databases of scientific and medical citations and co-citations (Newman 2001; Newman 2001). The scale of these investigations meant that existing computer programs for social network analysis were not adequate and they created their own. This meant also that they often overlooked the discussion and theorisation that had already occurred in the established field of social network analysis.

Within the community of social network analysts computer modelling of networks had advanced significantly through the development of *p** (p-star) models (Wasserman and Faust 1994; Robins, Pattison et al. 2005). This work can be understood as a specific application of the more general practice of 'agent-based-modelling'. There are, in fact, a wide range of simulation techniques used across the social sciences (Gilbert and Troitzsch 1999).

*P** models work with simple network models of points (nodes) and lines (edges) which in social networks are persons and two-by-two (dyadic) relations between persons. For a given network size and density (i.e. a given number of nodes and edges) computer simulations of the model subtract and add edges at random (stochastically) thus generating multiple examples of networks of this size and density. Social network analysis has theorised about a range of micro-level configurations such as (sociometric) 'stars', (persons with many direct connections), short geodesic paths and so forth. The *p** programs collect these statistics for each randomly generated network and compile a probability distribution of their occurrence across all networks in the family of networks of given size and density.

*P** models give social network analysts techniques and tools analogous to those used in statistical mechanics. However, researchers working in this tradition know much more about of micro-level interactions and formations researched and theorised by sociologists. By contrast, researchers coming from physics have a greater experience of manipulating and dealing with large-scale data sets and complex networks but lack this detailed knowledge that comes from studies of small scale complex networks.

P* analysis of the Australian network of interlocking directorates: Degree distributions

The simulation of families of networks of a given size and density gives the social network analyst for the first time the ability to compare an observed network to the array of all networks of that size and density. This possibility parallels the probabilistic tests of significance that other social scientists have had available for over a century. In standard statistics, measures of significance can be calculated from the imaginative construction of the probability space of events. Simple examples of this are tossing a coin or rolling a dice. Network structures have extraordinarily complex combinatorial characteristics however. It is not possible to calculate the range of outcomes in the same way. Computer simulations allow us to generate these outcomes. Actual probabilities are then taken from the array of simulated networks.

*P** models thus give to a researchers a statement about the probable frequency of occurrence of network features along with the standard deviation and other relevant probabilities. The researcher can look at the frequency of occurrence of a network feature in the observed network and judge how it falls relative to the mean frequency observed across the family of simulated networks. I will illustrate this process through an example social network dataset.

The observed dataset I use is the network of interlocking directors among Australia's largest 250 companies in 1996. In network terms this is a bipartite or '2-mode' dataset. I have discussed this dataset in an earlier paper (Alexander 2003) but the use of *p** analysis was done after that publication (Robins and Alexander 2004). The findings presented in this paper are based on the latter publication. For the *p** analysis we took the infrastructure of the largest component of 198 connected companies. Like nearly all networks of interlocking directors the 1996 Australian network had one very large component and scattering of separate, very small components. The *infrastructure* of the large component is simply all the directors who hold two or more board positions. These are the persons who create the intercorporate links.

The following tables provide the basic information about this (infrastructure) network. Table One lists the numbers of directors holding 2, 3 or more positions; Table Two lists the number of company boards having 1, 2, 3 or more interlocking directors on their boards.

Table 1: Interlocking Directors, Number of Positions held, Australia 1996 (Network Infrastructure only)

Number of Positions Held	Freq	Interlocks created*
2	150	150
3	67	201
4	24	144
5	8	80
6	4	60
7	2	42
Number of Directors	255	677

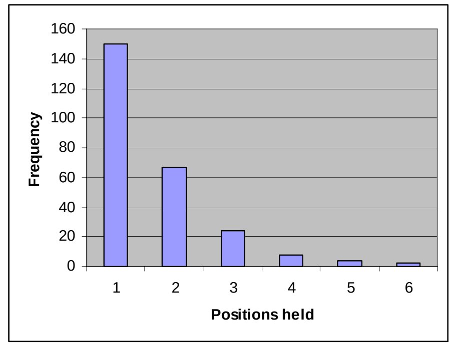
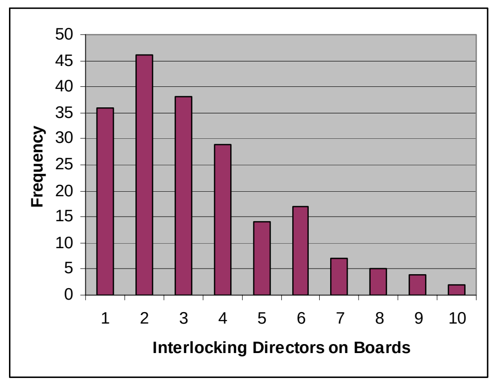


Table 2: Companies in Large Component; Number of interlocking directors on board.

Number of Interlocking Directors on Board	Freq	Contacts created*
1	36	0
2	46	46
3	38	106
4	29	174
5	14	104
6	17	255
7	7	147
8	5	140
9	4	144
10	2	90
Number of Company Boards	198	1,206



NB: In both tables, Column Three: 'Links/Contacts created' is the factorial expansion (n(n-1)/2) of column one, multiplied by the number of cases given in column 2. The histogram is the simple frequencies.

In network terms, these arrays are the dual *degree distributions* of directors' positions and board size (interlocking directors only) respectively. The factorial expansion of the array of positions held by the 255 directors yields the 677 interlocks between the

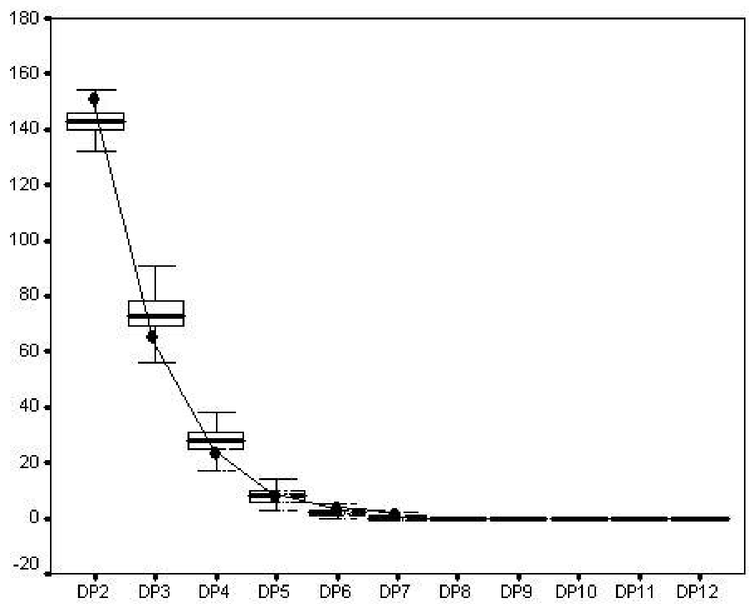
198 company boards (Alexander 1994). Conversely, the expansion of the array of directors on the boards will yields the 1,206 interpersonal connections among the 225 directors created by them sitting on a board together. These raw figures give the average density of the network of interlocks among companies and the network of personal connections among directors. They are raw arithmetic means however. We have no information about the distribution of these interlocks and connections.

P* **analysis: The observed network compared with the simulated networks**

The basic dimensions of this network, the number of boards, the number of directors and the density of connections between them were used as the base parameters of the *p** simulation model. The simulation program generated some 400,000 networks of these dimensions from which a random sample of 400 were used to generate the statistical profile of this family of networks. The model we used simulated the bipartite network in its original form.

Once the simulated networks are generated and sampled we chose some network features about which to collect statistics. For this paper I will discuss the basic features of the networks, the two degree distributions. Figure One below presents the information from the simulated networks about the distribution of positions held among the 255 directors.

Figure 1: Frequency count of persons holding 1, 2, 3 etc. positions: Distributions from simulated networks, observed values overlaid.



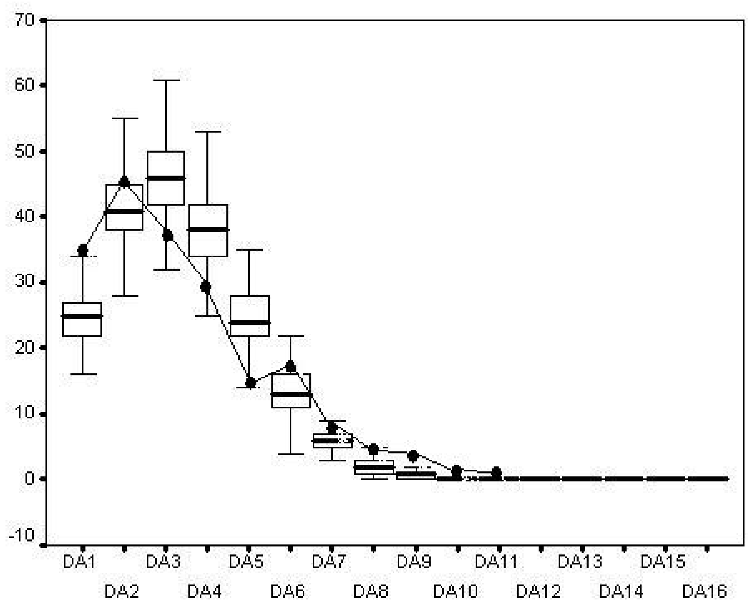
The box component of the box-and-whisker plot shows the outer values of the first and third quartiles and the middle line is the median value. The whiskers extend to 1.5 times the interquartile range and, conventionally, indicate the range of values that are not outliers. Thus the median frequency of persons holding just two board seats in the simulated networks was around 142. Networks with more than 151 or less than 131 directors holding two seats would be outliers. The observed frequency was 150, just within the range of acceptable values on the upper side.

The number of persons holding 3 and 4 positions sees the observed value under the boxes, indicating less people in these categories than would be expected from the simulated networks. Those holding 5 positions are on the median of the simulated networks while those holding 6 and 7 are above the level in the simulated networks.

Compared to the expectations of the random, simulated networks, we thus find a more highly differentiated distribution of board positions among persons in the observed data set. The numbers of persons at both extremes i.e. those holding 2 positions and those holding 6 and 7 positions is greater than expected while those in the middle range (3, 4 and 5 positions) are around or below average. There is a tendency for those with many positions to be holding more positions than expected thus creating a larger than expected frequency of persons holding just 2 positions. Compared to the expected distributions we see evidence of a small elite who gather more than their proportional share of positions making it harder for people with just 2 positions to gather positions and move into the 3-5 position holding categories.

Figure Two provides the information about the degree distribution of the interlocking directors across boards.

Figure 2: Frequency count of numbers of interlocking directors on company boards: Distributions from simulated networks, observed values overlaid.



Here we see a similar pattern to figure one. The observed values for this variable are more differentiated in the observed data than would be expected from the simulated networks. The boards having 6 or more interlocking directors among their members are more numerous than the simulations would predict at the expense of a larger than expected number of boards with just 2 interlockers and lower than expected numbers of boards with 3-5 interlockers on them.

A third feature of this network we considered was the average (median) geodesic path length. This is the feature of networks most prominent in the physicist's conception of small world theory. What we found was the median path length between company boards in our observed data set was longer than the expected length suggested by the simulated networks, 4 ‘degrees of separation’ rather than 3. The median is, however, a single measure. It is likely that the differentiation and elite formation we have observed in the degree distributions is also evident in the distribution of path lengths. Thus the median path length in the top quartile of path lengths in the observed network would be shorter than the expected value from the simulated networks but the observed network’s path lengths would increase quickly in the second quartile while the expected lengths would not. Longer path lengths for the majority of directors in the observed network may be the cost carried for the shorter path lengths among the elite.

Interpretation of the observed network

The simulated networks provide a template of network characteristics. By collecting statistics across the range of simulated networks we have an expected value for that feature as well as estimates of its likely range distribution and other probabilities. We can thus say whether the features of an observed network (of the same size and density) are those we would expect normally or are outside the range of normal expectations.

In the observed network of interlocking directors in Australia we have found that both the degree distribution of board seats to persons and the distribution of interlockers across boards are more extreme than the template from the simulated networks. In both distributions there is a bunching at the extremes, more people and boards in the very top range of the distribution balanced by a greater number of people and boards at the lowest, entry point of the distribution, and an under-representation of cases in the middle of the distributions. It thus appears that to those that have, more is given, and the ‘rich get richer’. There also appears to be an effect on median path length in the observed network.

We conclude that there is a tendency to elite formation in the observed network as compared to the conditions that would occur if there were no patterning of activity other than randomness. Note however that there is no a priori reason to see the tendency to elite formation in the two dimensions of activity as interrelated. It is quite possible that a tendency to elite formation in the personal ‘scorecards’ of directors is compatible with a spread of directors across boards characteristic of the random, simulated networks. Conversely, a tendency for directors to cluster together need not generate a differentiation among personal holdings of board seats.

Comments and future directions

This paper exemplifies new methodologies for probabilistic statistical analysis of social networks. Techniques of computer modelling and simulations combined with statistical analysis of micro-level configurations are analogous to the methods in the field of statistical mechanics which stimulated physicists’ interest in social networks. *P** modelling is however more tailored to social network analysis.

The example in this paper was an observed small world network of interlocking directorates. The innovative aspect of the Robins and Alexander study (Robins and Alexander 2004) was the direct simulation of the dual, bipartite network. The bipartite simulations of that study allowed me to assess the degree to which the observed network was unusual in the family of networks of the same size and density. I have argued that the observed network sits toward the extremes of this family of networks. There is evidence of bias away from randomness.

The next step in the development of this approach would be to refine the model and undertake another round of simulations. The additions to the simulation model could be built on attributes of individuals (e.g. age or experience) or attributes of companies (size, location, industry etc.). We would also model interaction effects between the two degree distributions of the networks. If the model fits the observed case more closely, we can argue that the factors we have chosen to add to the model are indeed important. However it may be that we cannot improve the model.

Watts and Strogatz’ championing of small world networks has linked social network analysis firmly to the work of complexity science and studies of synchronicity and emergence. *P** models developed by social network analysts provide customised tools for applying computer simulations to real world networks. Developed within social network analysis, computer models of this kind can contribute directly to the 'science of networks' proposed by the physicists. A 'science of networks' produces general models however. The worth of the model is its ability to reveal the same underlying processes and structure in many different situations and contexts. Enthusiasm for Watts and Strogatz' 'small world architecture' was stimulated by its discovery in many physical, biological and social networks.

This paper has used the generality of network simulation for sociological rather than scientific investigation. We used the general, randomised model to see what features of an observed network differed significantly from the simulated models. We have been able, therefore, to specify the specific divergences of the observed data from the general model. As well as validating the applicability of the general model we have begin to specify where local or non-predicted ('human') elements of this situation are producing macro-level effects. In the empirical network we examine we suggest there is a tendency to clustering at the extremes that is greater than what the model suggests.

The science of networks does not, therefore, produce new theory about social networks. Rather it allows us to investigate which aspects of social behaviour are general and reappear in the networks we investigate empirically. Conversely, it allows us to specify the exact elements of a particular network that show evidence of influences beyond the micro-level interactions used by the computer model. Physicists and sociologists have much to learn from one another in this new area of work and the dialogue can be complementary as long as we recognise the different ambitions of each discipline.

Footnotes

[1] Of particular interest to sociologists is the work of complexity theory on self-organising systems. Ecologists have done the most work in these seeing how the complexities of food chains organise into steady states.

[2] Watts and Strogatz determined a very specific mix of connectedness and randomness required to produce a small world network. Less than 1 per cent links being random connects a very large 'small world' network at six degrees of separation. The balance between clustering and randomness was very narrow; an increase of the random links beyond 1 or 2 percent breaks down the structure of clustering. The wonder of their discovery was that small world networks occupy a very tiny proportion of the range between complete local clustering and randomly linked networks yet appear to be very common in social life and in nature.

[3] The clusters are the boards themselves, while the random links are the interlocking directors.

References

Albert, R. and A.-L. Barabasi (2002). "Statistical mechanics of complex networks" *Reviews of Modern Physics* Vol. 74 (1): 47-97.

Alexander, M. (1994). "Business Power in Australia: The Concentration of Company Directorship Holding Among the Top 250 Corporates " *Australian Journal of Political Science* Vol. 29 (1): 40-61.

Alexander, M. (2003). "Boardroom networks among Australian company directors, 1976 and 1996: The impact of investor capitalism" *Journal of Sociology* Vol. 39 (3): 231-251.

Barabasi, A.-L. (2002). *Linked : the new science of networks* Cambridge, Mass., Perseus Pub.

Buchanan, M. (2002). *Nexus : small worlds and the groundbreaking science of networks* New York, W.W. Norton.

Gilbert, G. N. and K. G. Troitzsch (1999). *Simulation for the social scientist* Buckingham ; Philadelphia, Pa., Open University Press.

Milgram, S. (1967). "The small world problem" *Psychology Today*, (no. 2): 60-67.

Newman, M. E. J. (2001). "Scientific Collaboration Networks. I. Network construction and fundamental results" *Physical Review E* 64 (016131): 1-8.

Newman, M. E. J. (2001). "Scientific collaboration networks. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality" *Physical Review E* 64 (016132).

Robins, G. and M. Alexander (2004). "Small worlds among interlocking directors: Network structure and distance in bipartite graphs" *Computational and Mathematical Organization Theory* 10 (1): 69-94.

Robins, G., P. Pattison, et al. (2005). "Small and other worlds: Global network structures from local processes" *American Journal of Sociology*, Vol.110 (4): 894-936.

Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications* New York, Cambridge University Press.

Watts, D. J. (1999). "Networks, dynamics and the small-world phenomena" *American Journal of Sociology* Vol. 105 (2): 493-527.

Watts, D. J. (1999). *Small worlds : the dynamics of networks between order and randomness* Princeton, N.J., Princeton University Press.

Watts, D. J. (2003). *Six degrees : the science of a connected age* New York, Norton.

Watts, D. J. and S. H. Strogatz (1998). "Collective dynamics of 'small-world' networks" *Nature* No. 393: 440-442.