Realistic Social Networks for Simulation using Network Rewiring

Dekker, A.H.

Defence Science and Technology Organisation, Australia
Email: dekker@acm.org

Keywords: *Social network, scale-free network, small-world network, Kawachi process, network generation*

EXTENDED ABSTRACT

Simulation is a useful tool for exploring social and organisational questions. However, realistic social or organisational simulations must include realistic topologies for the social networks connecting people.

Real-world social networks are characterised by:

- a low average network distance, specifically one approximately equal to

$$\frac{\log n}{\log d},$$

  where n is the number of people, and d is the average degree;

- a moderate clustering coefficient (roughly in the range 0.2 to 0.7, depending on size and degree), reflecting the prevalence of triangles in social networks; and

- an approximately power-law distribution of node degrees.

The first part of this paper confirms these characteristics with an empirical study of nine social networks.

Three frequently-studied groups of networks are random networks, the preferential attachment networks of Barabási and Albert (1999), and the "small world" networks of Watts (2003). Each of these has some of the empirical characteristics of real social networks, but not all of them.

In the second part of this paper, we extend the rewiring process of Kawachi *et al.* (2004), which in turn extends that of Watts (2003). The extended process is controlled by two parameters, the original rewiring parameter p, and a new parameter π. Starting with a triangle-rich antiprism network, the combination of $p = 0.6$ and $\pi = 5$ generates networks which resemble real social networks.

The value of $p = 0.6$ means that approximately half of the original links are retained, leaving many triangles, and hence producing the required clustering coefficient. The value of $\pi = 5$ exaggerates the existing high-degree bias of the Kawachi process, therefore producing "scale free" behaviour, even when $p = 0.6$. Figure (i) shows a network produced by the extended rewiring process. Networks like this resemble real social networks, and are suitable for use in social and organisational simulations.

Even though the extended Kawachi rewiring process is not based on actual social processes, the resulting networks are realistic, and the simplicity of the rewiring process ensures that they can be integrated into simulation software with minimal effort.

We intend to use the rewiring process described here to produce networks for use in simulations of information flow through social networks.

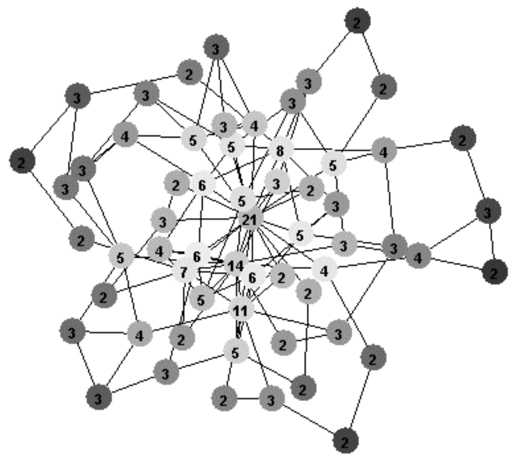

Figure (i). Example network produced by the extended Kawachi rewiring process.

1. INTRODUCTION

Simulation is a useful tool for exploring social and organisational questions (Perez and Batten 2006; Srbljinovic *et al*. 2003). Simulation can be used to explore the flow of ideas or information through an organisation or society (Dekker 2007b), to investigate variations in the performance of organisations (Dekker 2005; Dekker 2007a), or to study the development of social structures (Pujol *et al*. 2005). Simulation can answer many different questions about the impact of human characteristics and interactions on social and organisational behaviour.

However, realistic social or organisational simulations must include realistic topologies for the networks of relationships which connect human beings. Subtle differences in network topology can make for significant differences in behaviour (Pastor-Satorras and Vespignani 2001; Dorogovtsev *et al*. 2007).

In this paper, we examine some of the empirical characteristics of real-world social networks, and provide an extension to the network rewiring process of Kawachi *et al*. (2004), which produces networks having those characteristics. The extended rewiring process therefore provides a simple way of generating realistic social networks which can be incorporated into simulation studies of social and organisational systems.

2. EMPIRICAL CHARACTERISTICS OF SOCIAL NETWORKS

We first investigate some of the empirical characteristics of social networks, by reviewing the relevant literature, and examining nine real-world social networks:

- the two island voyaging networks of Hage and Harary (1995);
- the Florentine families network in Wasserman and Faust (1994);
- three Internet social networks, two from newsgroups, and one from a "blogging" community; and
- three work communication networks within the Australian Department of Defence.

2.1. Small Worlds

The first important characteristic of real-world social networks is a low average distance between people. This is often called the "small world" effect (Watts 2003; Szabó *et al*. 2003). For the wider human community, it is also described as "six degrees of separation," which reflects an average distance of about seven network links between randomly chosen human beings.

Theoretical analysis (Watts 1999; Bollobás 2001; Durrett 2007) suggests that the average distance between people will be approximately:

$$\frac{\log n}{\log d}$$

where n is the number of people, and d is the average degree, i.e. the average number of links per person. For our small sample, the average distance D between people is approximately:

$$D \approx 0.92 \frac{\log n}{\log d} + 0.41, \qquad (1)$$

(with a correlation of 0.94, statistically significant at the 0.0002 level). This is consistent with the theoretical prediction. For a network with n = 60 people and an average of d = 4 links per person, we would expect an average distance of about 3.1. Figure 1 illustrates the regression equation (1).

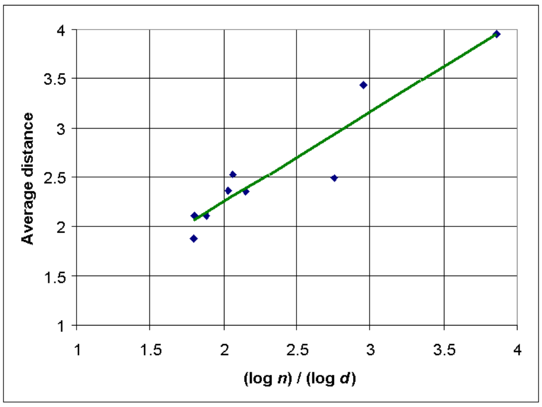

Figure 1. Prediction of average distance in social networks by (log n)/(log d).

This relationship does not only hold for social networks. For example, reviewing the ecological data in Williams *et al*. (2002) shows that a similar relationship holds for food webs, with:

$$D \approx 1.07 \frac{\log n}{\log d} + 0.24, \qquad (2)$$

(with a correlation of 0.98, statistically significant at the 0.0001 level). This explains the counterintuitive dependence of average distance on size noted in that work.

The "small world" effect also holds in the three commonly used classes of network generation processes – random networks (Bollobás 2001), preferential attachment (Barabási and Albert 1999), and small-world rewiring (Watts 2003).

2.2. Clustering Coefficient

Another important characteristic of real-world social networks is the formation of triangles (Kilduff and Tsai 2003; Watts 2003). Social factors mean that if John is a friend of (or works with) Peter, and Peter is a friend of (or works with) George, then John and George are more likely than average to also be friends (or work with each other). Mathematically, this is reflected in a moderate clustering coefficient, where the clustering coefficient is the probability that two nodes X and Y linked to Z will also be linked to each other. For our sample of networks, the clustering coefficient C ranged from 0.16 to 0.68, and roughly fitted the regression equation:

$$C \approx 1.7\frac{\log d}{\log n} - 0.3\,, \qquad (3)$$

(with a correlation of 0.84, statistically significant that the 0.005 level).

For a network with n = 60 and d = 4, we would therefore expect a clustering coefficient of about 0.29. Figure 2 shows a real-world social network with n = 60 and d = 4 (an informal communication network for an organisation, not including managers or formal links). Because of the way this network was constructed, it contains more nodes of degree 1 then is usual, but otherwise it is typical. It has a clustering coefficient of 0.37.

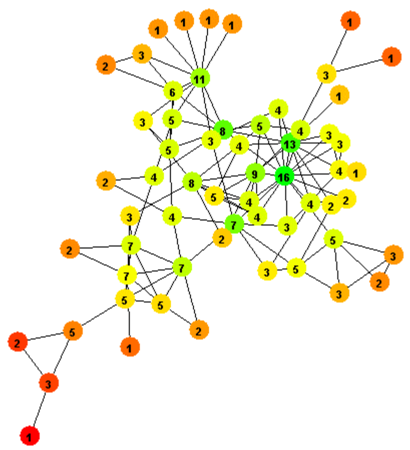

Figure 2. An informal communication network.

The "small world" rewiring process of Watts (2003) was invented to give a way of producing networks with moderate clustering coefficients. Random networks and the preferential-attachment networks of Barabási and Albert (1999) do not have this property (Dekker 2005).

2.3. Power Laws

The third important property of real-world social networks is an approximately power-law distribution of node degrees. This has been confirmed in many empirical studies (Albert and Barabási 2002), and holds also for our sample. In particular, if d_i is the degree of person i in the network, and R_i is the rank of that person's degree (1 for the highest, etc.), then there will be an approximately linear relationship between log d_i and log R_i, except possibly at the limits of very high or very low degree. Networks with this property are described as "scale free," and this property has an important impact on, for example, the spread of viruses (Pastor-Satorras and Vespignani 2001).

Figure 3 shows the log-log plot of node degree against node degree rank for the network in Figure 2. The points approximately fit a straight line (with a correlation of –0.91), indicating that the network is approximately "scale free." There is some deviation at high degrees: the degrees of the most-connected people are somewhat lower than a power law would predict. There are also fewer nodes of degree 1 than a power law would predict.

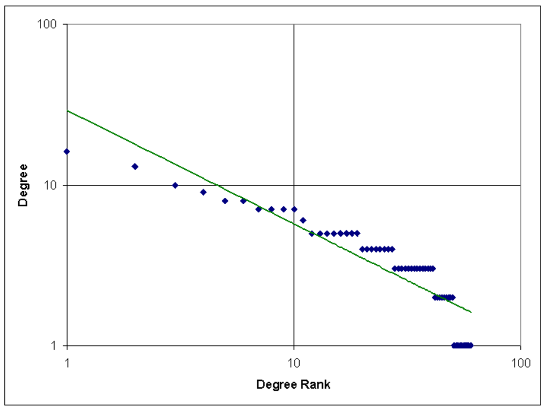

Figure 3. Log-log plot of degree against degree rank for the network in Figure 2.

The preferential-attachment process of Barabási and Albert (1999) was invented to give a way of producing "scale free" networks. In contrast, random networks have a Poisson degree distribution (Bollobás 2001), while the "small world" networks of Watts (2003) have almost identical degrees for every node.

2.4. Summary

Table 1 summarises the characteristics of social networks which we have described. The three commonly used network generation processes – random networks (Bollobás 2001), preferential attachment (Barabási and Albert 1999), and small-world rewiring (Watts 2003) – each produce networks with some of these characteristics, but not others.

There is therefore a need for a simple network generation process which will produce networks having all these properties. Such networks would be suitable for use in social and organisational simulations.

Table 1. Properties of social networks, compared to three network generation models.

Property	Social Network	Random Net	Preferential Attachment	Small World
Low average distance	✓	✓	✓	✓
Moderate clustering coefficient	✓			✓
Power laws for degrees	✓		✓	

3. THE KAWACHI REWIRING PROCESS

Kawachi *et al*. (2004) introduced an extension to the "small world" rewiring process of Watts (2003). Starting with a regular network, rewiring occurs in several phases (usually three), and is biased so as to preferentially move links to be adjacent to highly linked nodes (nodes of high degree).

We have improved the Kawachi process to avoid disconnection of the network during rewiring (Dekker 2007b). In the improved Kawachi process, the link *X*–*Y* (where *X* has higher degree) is replaced by *X*–*Z*, where *Z* is chosen to be an isolated node from a previous rewiring (if there are any), or chosen randomly (with probability proportional to the degree of nodes plus one) otherwise.

Links are rewired with probability $p/3$, where p is a rewiring parameter, and the rewiring is repeated three times, so that the probability of an edge being rewired is $1 - (1 - p/3)^3$. If the resulting network is disconnected, the entire process is repeated from the beginning.

For small values of p, the Kawachi process has almost exactly the same behaviour as the Watts process, and produces similar "small world" networks, but for higher values of p (above 2), "scale free" networks are produced (Kawachi *et al.* 2004). Figure 4 shows the result of applying the Kawachi process with p = 0.05 to a 30-node antiprism network.

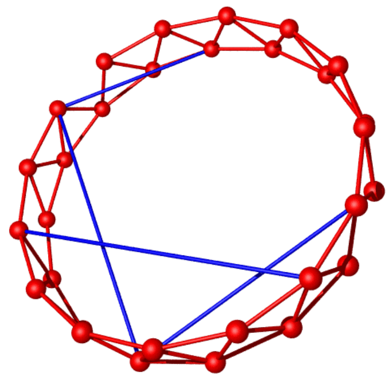

Figure 4. Kawachi rewiring of an antiprism.

The Kawachi process transitions between four important classes of network: regular, "small world," random, and "scale free." This makes it a valuable tool for social simulation (Dekker 2005, Dekker 2007b). However, none of these four classes of network have the characteristics of real-world social networks.

4. EXTENDED KAWACHI REWIRING

In order to produce more realistic networks, we extend the Kawachi process quite simply by adding a parameter π (the power), and modifying the preferential attachment in the rewiring process to use as a probability the degree of nodes plus one raised to that power, i.e. $(d_i + 1)^{\pi}$.

We investigated the range of rewiring parameters p from 0.4 to 1.0, and the range of powers π from 1 to 6. We began with a 60-node antiprism (i.e. n = 60 and d = 4), and generated ten networks for each combination of parameters.

Our first consideration was to identify the combination of parameters producing approximately "scale free" degree distributions. For this purpose, we restricted attention to networks where the (negative) correlation between the logarithm of node degree and the logarithm of node degree rank was (on average) at most –0.95.

This restriction was satisfied by the subset of networks with $p \geq 0.5$ and $\pi \geq 6.5 - 5p$. Figure 5 shows the average values of the correlation.

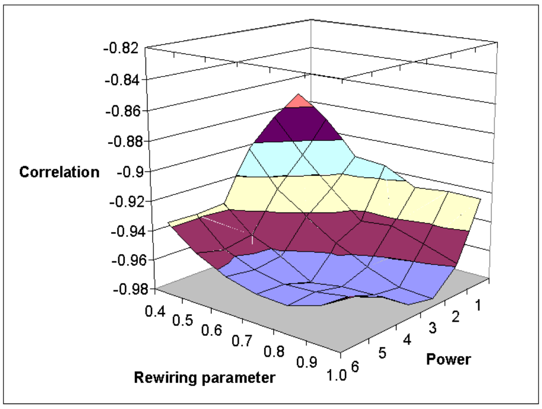

Figure 5. Correlation between log of node degree and log of node degree rank for different parameter combinations.

The average distance for the different generated networks ranged from 2.3 to 3.2. However, the smaller of these distances corresponded to "star-like" networks with a single node of very high degree. Such networks are not realistic representations of social networks. With an expected average distance of 3.1 for social networks with $n = 60$ and $d = 4$, we introduced a cut-off of 2.75, corresponding to $\pi \leq 11 - 10p$. Figure 6 shows the average distance values.

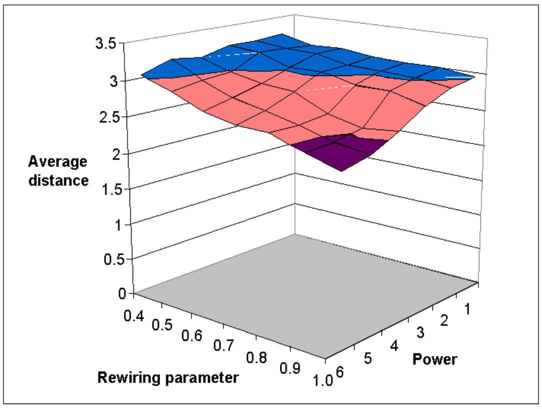

Figure 6. Average distance values for different parameter combinations.

With a predicted clustering coefficient of 0.29 for networks of this size, we took a threshold of 0.25 for the clustering coefficient. This corresponded to $\pi \geq 8 - 5p$ and $\pi \geq 4$, a tighter constraint than for a "scale free" degree distribution. Figure 7 shows the clustering coefficient values.

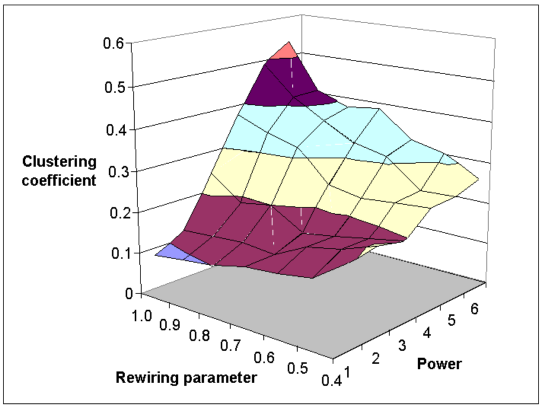

Figure 7. Clustering coefficient values for different parameter combinations.

The three constraints left only two possibilities, namely $p = 0.5$ and $\pi = 6$, or $p = 0.6$ and $\pi = 5$. However, the first of these combinations produce networks which were too star-like, and so a Kawachi parameter of $p = 0.6$ and a power of $\pi = 5$ is the combination we recommend (at least for networks with n around 60 and d around 4).

Figure 8 shows an example network produced with this combination of parameters. The log-log plot of node degree against node degree rank in Figure 9 shows that it is indeed "scale free" (the correlation is –0.98), and the clustering coefficient of 0.26 and average distance of 3.0 is consistent with social networks of this size and average degree.

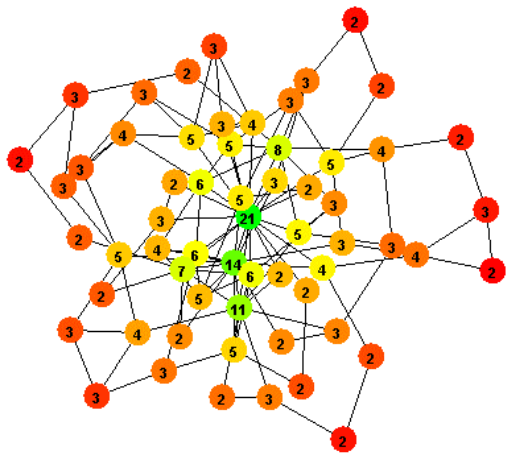

Figure 8. Network produced by extended Kawachi rewiring with $p = 0.6$ and $\pi = 5$.

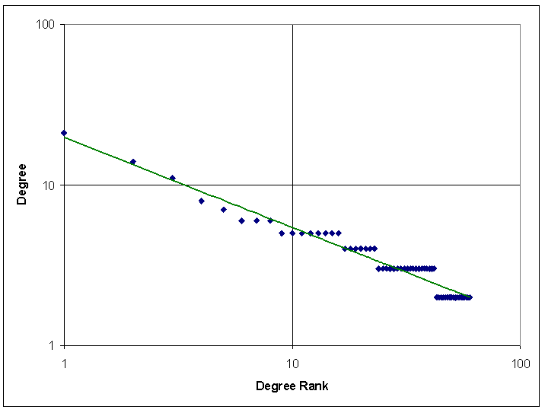

Figure 9. Log-log plot of degree against degree rank for the network in Figure 8.

The rewiring parameter of p = 0.6 means that approximately half of the original links are retained, since $(1 - p/3)^3 = 0.51$. This leaves many triangles intact, and hence produces the clustering coefficient of 0.26.

The power π = 5 exaggerates the existing high-degree bias of the Kawachi process, therefore producing "scale free" behaviour, even when p = 0.6. Taking a degree of 4 as a base, a node with half that degree is 1/13 as likely to be chosen during the rewiring process, while a node with double that degree is 19 times as likely to be chosen. This introduces sufficient high-degree bias to produce "scale free" behaviour.

In other words, the rewiring process we have described produces all three of the properties of real-world social networks described in Table 1, even though the rewiring process does not model the social activities by which such networks are produced in practice.

5. RELATED WORK

An alternate method for producing realistic social networks uses the p^* family of random graph models (Carrington *et al*. 2005). These models associate inter-dependent probabilities of links existing to each pair of nodes. This can take account of clustering coefficients, the distribution of node degrees, and other more subtle topological characteristics.

However, this approach requires substantial amounts of empirical data in order to select appropriate model coefficients, and the mathematical sophistication of the technique makes it suitable largely for expert practitioners. The method which we have proposed, although it produces somewhat less realistic social networks, is more easily incorporated into simulation models.

Davidsen *et al*. (2002) give a technique for creating social networks by allowing people to "introduce" acquaintances to each other. This generates networks with "scale free" degree distributions and moderate clustering coefficients. However, the average distance between people is significantly lower than for real social networks or for random networks, suggesting that the networks produced are too star-like to be realistic.

Holme and Kim (2002) extend the preferential-attachment process of Barabási and Albert (1999) to add a "triad formation" step, generating any desired clustering coefficient. This also produces networks with a lower average distance between people, and so it is not clear that these networks adequately model real social networks.

The spatial growth mechanism of Kaiser and Hilgetag (2004) can also generate networks with "scale free" degree distributions and moderate clustering coefficients. However, this is done by imposing a spatial structure on network growth, which results in average distances significantly larger than would be expected in social networks.

6. DISCUSSION

We have demonstrated through an empirical study three important characteristics of real-world social networks:

- a low average network distance;
- a moderate clustering coefficient; and
- an approximately power-law distribution of node degrees.

We have introduced an extension to the network rewiring process of Kawachi *et al*. (2004). The extended rewiring process produces networks having all three of these real-world characteristics, and which is superior to previous methods. It therefore provides a simple way of generating realistic social networks which can be incorporated into social and organisational simulations, and we intend to so use it in future experimental work.

7. REFERENCES

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

Barabási, A.-L. (2002), *Linked: The New Science of Networks*, Perseus Publishing.

Barabási, A.-L. and Albert, R. (1999), "Emergence of scaling in random networks," *Science*, Vol. 286, pp 509–512.

Bollobás, B. (2001), *Random Graphs*, 2nd edition, Cambridge University Press.

Carrington, P., Scott, J., and Wasserman, S. (2005), *Models and Methods in Social Network Analysis*, Cambridge University Press.

Davidsen, J., Ebel, H., and Bornholdt, S. (2002), "Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks," *Physical Review Letters*, Vol. 88, No. 12, 128701.

Dekker, AH (2005), "Network Topology and Military Performance," in Zerger, A. and Argent, R.M. (eds), *MODSIM 2005 International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, December, pp 2174–2180, ISBN: 0-9758400-2-9.

Dekker, A.H. (2007a), "Using Tree Rewiring to Study 'Edge' Organisations for C2," *Proceedings of SimTecT 2007*, Simulation Industry Association of Australia, June, pp 83–88, ISBN: 0-9775257-2-4.

Dekker, A.H. (2007b). "Studying Organisational Topology with Simple Computational Models," *Journal of Artificial Societies and Social Simulation*, to appear.

Dorogovtsev S.N., Goltsev A.V., and Mendes J.F.F. (2007), "Critical phenomena in complex networks," arXiv:0705.0010v2 [cond-mat.stat-mech], 1 May.

Durrett, R. (2007), *Random Graph Dynamics*, Cambridge University Press.

Hage, P. and Harary, F. (1995), "Eccentricity and centrality in networks," *Social Networks*, Vol. 17, pp 57-63.

Holme, P. and Kim, B.-J. (2002), "Growing scale-free networks with tunable clustering," *Physical Review E*, Vol. 65, 026107.

Kaiser, M. and Hilgetag, C. (2004), "Spatial growth of real-world networks," *Physical Review E*, Vol. 69, 036103.

Kawachi, Y., Murata, K., Yoshii, S., and Kakazu, Y. (2004), "The structural phase transition among fixed cardinal networks," *Proceedings of the 7th Asia-Pacific Conference on Complex Systems*, 6–10 December, Cairns, Australia, pp 247–255.

Kilduff, M. and Tsai, W. (2003), *Social Networks and Organizations*, SAGE Publications.

Pastor-Satorras, R. and Vespignani, A. (2001), "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, Vol. 86, No. 14, pp. 3200–3203.

Perez, P. and Batten, D. (2006), *Complex Science for a Complex World: Exploring Human Ecosystems with Agents*. Canberra: ANU E Press: epress.anu.edu.au

Pujol J.M., Flache A., Delgado J., and Sangüesa R. (2005), "How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges," *Journal of Artificial Societies and Social Simulation*, Vol. 8, No. 4.

Srbljinovic A., Penzar D., Rodik P., and Kardov K. (2003), "An Agent-Based Model of Ethnic Mobilisation," *Journal of Artificial Societies and Social Simulation*, Vol. 6, No. 1: jasss.soc.surrey.ac.uk/6/1/1.html

Szabó, G., Alava, M., and Kertész, J. (2003), "Structural transitions in scale-free networks," *Physical Review E*, Vol. 67, 056102.

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

Watts, D.J. (1999), "Networks, Dynamics, and the Small-World Phenomenon," *American Journal of Sociology*, Vol. 105, No. 2, September, pp 493–527.

Watts, D.J. (2003), *Six Degrees: The Science of a Connected Age*, William Heinemann.

Williams, R.J., Berlow, E.L., Dunne, J.A., Barabási, A.-L., and Martinez, N.D. (2002), "Two degrees of separation in complex food webs," *PNAS*, Vol. 99, No. 20, October 1, pp. 12913–12916.