

Signatures in Networks Generated from Agent-based Social Simulation Models

Shah Jamal Alam, Ruth Meyer, and Bruce Edmonds

Centre for Policy Modelling,
Manchester Metropolitan University Business School
Manchester M1 3GH, United Kingdom
{shah, ruth}@cfpm.org, bruce@edmonds.name

Abstract. Finding suitable analysis techniques for networks generated from social processes is a difficult task when the population changes over time. Traditional social network analysis measures may not work in such circumstances. It is argued that agent-based social networks should not be constrained by *a priori* assumptions about the evolved network and/or the analysis techniques. In most agent-based social simulation models, the number of agents remains fixed throughout the simulation; this paper considers the case when this does not hold. Thus the aim of this paper is to demonstrate how the network signatures change when the agents' population depends upon endogenous social processes. We argue for a much wider attention from the social simulation community in addressing this open research problem.

Keywords: agent-based social simulation, social network signatures, Kolmogorov-Smirnov statistic

1 Introduction

Many social simulation models assume a fixed number of agents, usually for simplicity's sake. As a result, only the degree distribution of the resulting social networks evolves over time, while the size of the network remains unchanged. Many agent-based social simulation (ABSS) models have shown that agents' local interactions can lead to structures such as *scale-free* and *small-world* networks, where the network stabilizes after a sufficient number of simulation runs (e.g. [6, 12]). However, it is usually an explicit algorithm and the network size that influences the simulation outcomes. Analyzing the networks where the size and structure of the agent population depends upon emergent social processes remains a largely open question for the social simulation community.

There are a variety of metrics measuring global properties of networks coming from two research communities. The 'complex network' community focuses mainly on size, density, maximum degree, average path length, connectivity, and degree distribution (e.g. fat-tailed) [10]. The 'social network analysis' (SNA) community have their own measures such as centrality, closeness, proximity, clustering,

homophily, and community structure [13]. However, it has been shown that networks sharing similar global properties can have significantly varying local structures [5].

ABSS modellers face an even bigger challenge: characterizing the co-evolution of a dynamic network as the simulation proceeds. If agents are heterogeneous and able to join or leave the modelled system during the simulation run, the network characteristics may change completely over the course of a simulation run. Typically, snapshots of the simulated network are taken at regular time intervals. Such snapshots of the network can then be compared based on the SNA measures and other complex network characteristics. However, as Edmonds and Chattoe [4] have argued, in many cases an agent's behaviour cannot just be reduced to the notion of a mere node. That is, any single network measure will be inadequate to capture the underlying social process then.

In the subsequent sections, we define the problem followed by a discussion on two agent-based models. We demonstrate how the use of some *a priori* assumptions can affect the evolved network's structure. In the end, we suggest two potential ways towards understanding changing agent-based social networks.

2 Networks from embedded social processes

Consider a simulated network $S(N, E)$ at the start of a simulation run, i.e. at time $t=0$. N is the set of nodes and E is the set of edges through which the nodes of the graph are connected. The agents are represented as nodes while the ties between them form the edges in the network. The network could be of any structure, e.g. unconnected ($E = \emptyset$), a small-world network [15], or a scale-free network [3]. At every time step, t : $t \in (0, \infty]$ new agents become a member of N , while n already present agents leave: $n \in (0, |N|]$.

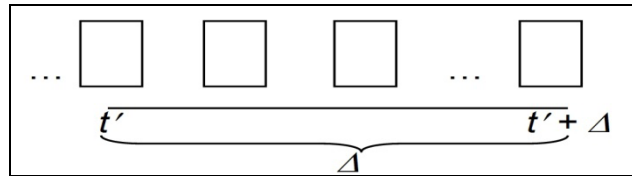


Figure 1: An agent-based simulation run over time, where Δ is the lag size for selecting simulation snapshots.

With each time step, the network is likely to change. Changes may happen to the size of the agent population (number of nodes in the network) and also to the formation and breaking of ties among the agents (number and position of edges). The degree distribution of the network alters depending upon the type of relations that the edges represent.

In models where the joining and leaving of agents is driven by socioeconomic conditions, agents' decision choices are constrained. The resulting network structures in such case are hard to predict in advance. The shape and size of the network at a certain time t' could be radically different at a later time $t'+\Delta$ (where Δ is the possible

time-lag of size ranging from 1, 2 ... onwards; see figure 1). It is even possible that the network dissolves after some time and recovers later on.

Unlike physical systems, social processes tend to be modelled descriptively and validated qualitatively [8]. The evidence is often gathered through fieldwork. An individual's relations and actions are driven by their position in society and other factors affecting the system. Social systems do not remain in a stable state and are dynamic in nature. Events changing the structure of the network may occur any time during the simulation which might be missed when using global measures.

Keeping the number of nodes fixed could possibly increase the risk of type-II error or a 'false-positive' effect. For instance, it is possible that a network's convergence to a certain topology may be rigidly determined by the mechanism through which the agents interact. Ideally, one would apply a scheme that helps in understanding the interplay of the processes and remains robust to the changing network up to certain degree. Of course, in case the network degenerates, any such scheme is likely to fail.

3 Examples from two agent-based social simulation models

Two different agent-based models are presented to support our proposition. The first is a model developed by Jin et al. [6] about friendship networks generated from three simple rules. The second is being developed by us based on evidence from a real case study. The two models are different with respect to their purpose and complexity and are certainly not compared in this paper. Our purpose is to highlight the likely effect of a fixed agent population and *a priori* assumptions about the agents' interaction that may bias an evolved network structure.

3.1 A model for growing social network

An example of simple rules resulting in the emergence of highly-clustered network structures is explained in Jin et al. [6]. The model is based on three rules: "(1) meetings take place between pairs of individuals at a rate which is high if a pair has one or more mutual friends and low otherwise; (2) acquaintances between pairs of individuals who rarely meet decay over time; (3) there is an upper limit on the number of friendships an individual can maintain." Since the number of agents, i.e. nodes in the network remain fixed at all time, the building and breaking up of ties are solely dependent upon the algorithm used in the model.

Simulating this model with 250 agents and a friendship upper limit set to 5 resulted in a quick emergence of a 'small-world' like structure. Once converged to a certain value, the clustering coefficient varies [13] within a small range (figure 2). We have used the implementation of the simplified model included in the Repast [11] demonstrations package.

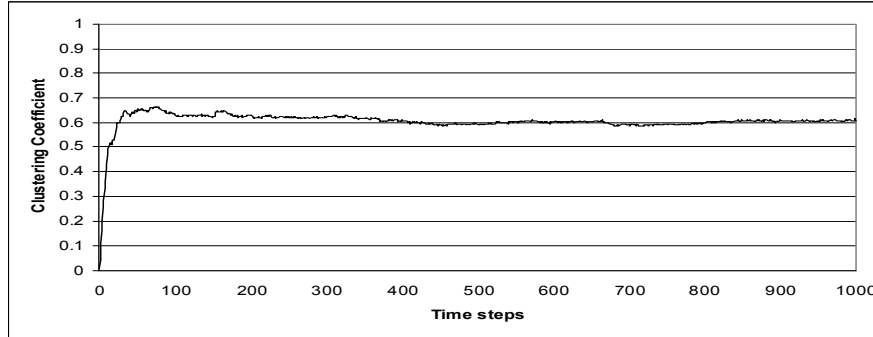


Figure 2: Clustering coefficient of the friendship network [6] over a simulation run.

Next, we compared network snapshots $P = \{P_0, P_{0+\Delta}, P_{0+2\Delta}, \dots\}$ where each element P_i represents a distinct population of agents at a particular time in the run, P_0 being the population at $t=0$. We performed the Kolmogorov-Smirnov (K-S) test [9] for each consecutive pair $(P_t, P_{t+\Delta})$ [1]. The K-S test belongs to the family of nonparametric statistics and is ‘distribution-free’. Unlike classical statistical techniques such as e.g. the t -test, there is no prior assumption about the distribution of the population from which the data is sampled. A two-population K-S-test indicates the likelihood that two datasets come from the same distribution. A comparison of the nodes degree distribution based on the K-S statistic in figure 3 shows the network stabilizing after some time. The p-score (Y-axis) that can be calculated from the KS-test can be roughly interpreted as the probability that the two data sets have the same distribution. In our context, a value 1.0 (maximum) implies that there is no change in the degree distribution of the two consecutive snapshots of the agents’ network. On the other hand, a low p-score is an indication that the network structure has changed from the previous snapshot even though, for example, the clustering coefficient remains the same. Figure 3 (left) illustrates the case when the snapshots were taken at every 100th time step. The network composition does change initially, but around the 500th time step, the degree distribution of the network stabilizes and there is no difference. Figure 3 (right) shows the comparison of networks snapshots taken at every 25th time step. As can be seen, the network starts to take shape increasingly and finally stabilizes in the distribution.

Choice of a lag does influence the outcome of the statistic when comparing consecutive snapshots of the degree distribution. With a smaller lag size in figure 3 (right), one may conclude that the network’s composition stabilizes around the around 100th time step. Figure 4 shows the p-score for the K-S test statistic where we compared consecutive snapshots at the smallest possible lag size, i.e. every 5 time steps. This is due to the reason that the three basic actions ‘random meeting’, ‘neighbour meeting’ and ‘random removal’ occur at every 5th step in this model. The agents’ network grew in the initial steps and the similarity in the network configuration continued to increase thereon. In all three different granularity cases (figures 3 and 4), it is clear that once the network stabilizes, its structure remains the same. One may conclude this to be a consequence of a fixed population size, where only the ties are changed.

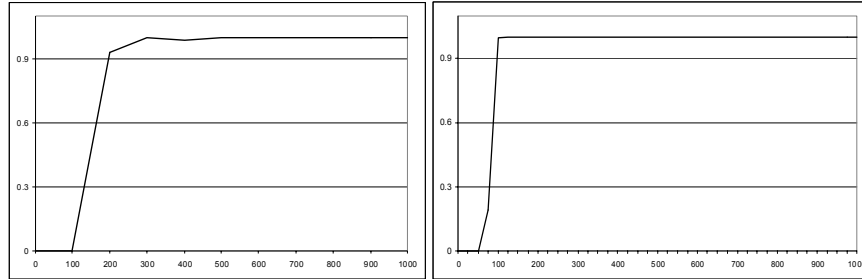


Figure 3: Y-axis: P-score of the K-S test for the degree distribution of the friendship network compared for consecutive snapshots taken at every (left) 100th tick and (right) 25th time step.

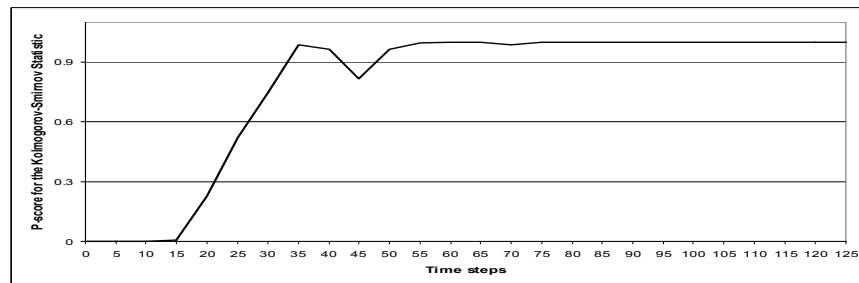


Figure 4: P-score of the K-S test for the degree distribution of the friendship network compared for consecutive snapshots taken at every 5th time step for the first 125 time steps of the simulation.

3.2 A model of socioeconomic impact of HIV/AIDS

The second model is being developed based on a case study in the CAVES project¹. It investigates the impact of HIV/AIDS on households and the overall community structure. The case study area is located in the Limpopo region in South Africa and is one of the most vulnerable areas lacking water, food security, jobs and other social infrastructure. Many households have female heads because the men are often living away from the house as migrant workers. State grants are the primary source of income of which a high percentage is spent on food, health and funeral costs. People try to cope with stressors via mutual help amongst neighbours, friends and extended family. A major concern is the number of orphans in the community that has increased mostly due to HIV/AIDS related adult deaths. The extended family by and large accommodates orphans of a dissolved household [16].

The model focuses on the behaviour of individual agents as well as that of households and thus attempts to take into account both the individual interactions and the decisions taken at the household level. We have adopted a multi-layer network approach to model the social networks. Individuals are represented as agents with a network of friends. Each individual is member of a household, with one of the

¹ CAVES (Complexity, Agents, Volatility, Evidence, Scale): <http://www.cfpm.org/caves>

household members acting as the household head. Households have a network of social neighbours with whom they interact. In addition, agents may join different groups like funeral clubs or savings clubs.

Agents and households are created based on the available demographic data. Both endogenous and exogenous factors influence the dynamics of agent interactions. As a result, the size of the generated networks changes over time. At creation agents are assigned some random friends. With the agents joining the clubs, the size of the friendship network remains dynamic. A high prevalence of HIV/AIDS affects the health of those who get infected resulting in a high increase death. Orthogonally to the network layers of friends and the household social neighbours is the extended family structure. As stated above, each individual agent is a member of a household. Households in turn form clusters which represent the extended family. This comes into play when a household dissolves due to the death of all care-providing adults, leaving the dependants (orphan children and possibly any seniors without income) behind. If this happens, an accommodating household has to be found. This search uses the family hierarchy to determine the nearest living relative who is able to accommodate the surviving dependants. If there is none in the extended family, the search is expanded to the networks of neighbours and friends. Alam et al. [2] describes the model in detail.

The simulation runs evaluated in this paper were initialized with 30 households and about 150 adult and senior agents. These numbers were chosen because the online tool used in calculating the Kolmogorov-Smirnov p-score allowed only a limited data set size [7]. For future work, we will apply standard statistical software for a larger number of agents.

In this setup, only the friendship network of adult and senior agents has been considered. We added the three rules for growing the friendship network from the model by Jin et al. [6]. In contrast to the latter model, the size of the population is not fixed – agents die and are born over the course of the simulation. Agents' expected life was sampled from a normal distribution with a mean of 75 years and standard deviation 10 years. However, the actual lifespan of an agent may be shorter due to declining health. This is influenced both by the socioeconomic status of a household, which affects the food intake of its members, and the prevalence of HIV/AIDS. The birth rate was set to 15 children per 1000 individuals of a population. The size and the network composition resulting from this model cannot be assumed in advance.

To investigate the impact of HIV/AIDS, or in other words the impact of a rapidly decreasing agent population on the evolution of the friendship network we performed two batches of simulation runs, one without prevalence of HIV/AIDS and one with a prevalence of 14%. Agents were randomly assigned 1 to 5 friends thus using the maximum limit as in the model described in the previous section.

The first batch comprised of three simulation runs with slightly different settings regarding the rules governing the formation and breaking of friendship links. Figure 5 shows the clustering coefficient of the agents' friendship network for the three different cases.

In the first case (Case-I), friendships are formed by two of the three rules from the Jin et al. model [6], the neighbour meeting and the random meeting rule. Agents also make friends when they join the same club. Friendship ties only break when agents die. Case-II incorporated the random removal rule in addition, whereas for Case-III, only the three rules from [6] were used and the same club membership was ignored

In all the three cases, the network remained sparse with a small clustering coefficient. Adding the random removal rule did influence the abundance of links initially, but was not significant in the long run. The relative changes in the friendship network's degree distribution were more evident when the Kolmogorov-Smirnov test was applied to all the three cases, as in the previous model. While a global property of a network, here the clustering coefficient, showed a similar pattern, networks actually changed in different ways for each case (see figure 6).

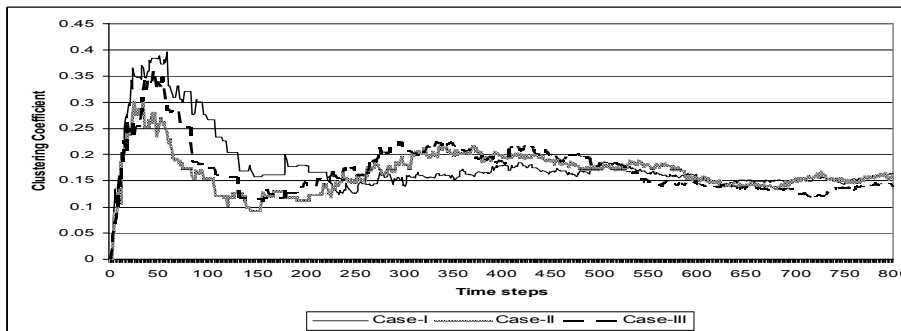


Figure 5: Clustering coefficient of the agents' friendship network for the three different simulation settings.

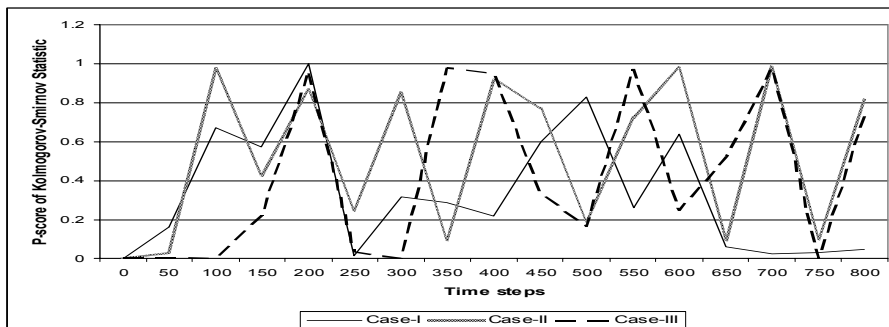


Figure 6: P-score of the K-S test for the degree distribution of the friendship network compared for consecutive snapshots taken at every 50th time step for the three different simulation settings.

In the next step, the simulation as in Case-III was run, but this time with an HIV/AIDS prevalence of 14%. Since this results in a rapidly declining agent population the clustering coefficient shown in figure 7 (left) declines much faster compared to the three previous cases in figure 5. Most agents die around the 600th time step and so the clustering coefficient becomes zero.

The p-score of the K-S statistic drops between time periods 350-450 indicating significant change occurring in the network. This happens due to increasing deaths in the system, which affects the degree distribution of the agents' friendship network. As the network becomes sparse after the 500th step, the corresponding p-value shows no change at all, followed by a sharp decline when the entire community structure collapses.

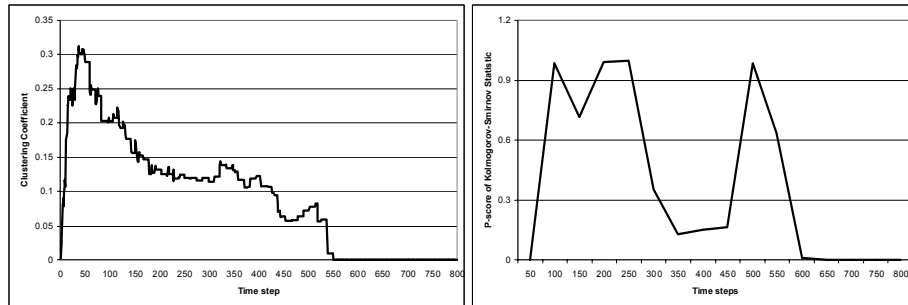


Figure 7: (left) clustering coefficient of the network; (right) P-score of the K-S test for the degree distribution of agents' friendship network taken at every 50th time step, with HIV/AIDS prevalence.

To further investigate whether the changes in the number of agents does really make an impact on the network signatures, we ran simulations for three additional cases without prevalence of HIV/AIDS. In Case-I, no birth of an agent was possible, so the agent population slowly declined. Case-II used the opposite scenario with only births but no deaths happening during the simulation run. For Case-III we kept the initial population of agents constant throughout the run.

Figure 8 shows the clustering coefficients for these three cases. For the first case we get a similar pattern as in figure 7 (left). Whereas in Case-I no birth was allowed, in the latter case agents' fertility had declined due to the HIV/AIDS prevalence, thus equally constraining births. Case-II follows a similar pattern as the simulations displayed in figure 5, where the population is equally growing due to the high birth rate. Case-III however, shows a high clustering coefficient, well-above 0.5 after some initial steps. This is very similar to the clustering coefficient obtained for the Jin et al. model in figure 2. The reason for achieving a so-called 'small-world' effect is that the agent population remained the same in this case.

The p-score for the Kolmogorov-Smirnov (K-S) statistic for the respective three cases is shown in figure 9. Similar network configurations may be observed for Case-I and the HIV/AIDS prevalence scenario shown in figure 7 (right). The K-S statistic for Case-II keeps changing over time which is due to the birth of agents during the simulation. For Case-III, the signature is again almost identical to the Jin et al. model run as shown in figure 3. Once the agents' friendship network stabilizes it remains the same for the rest of the simulation run.

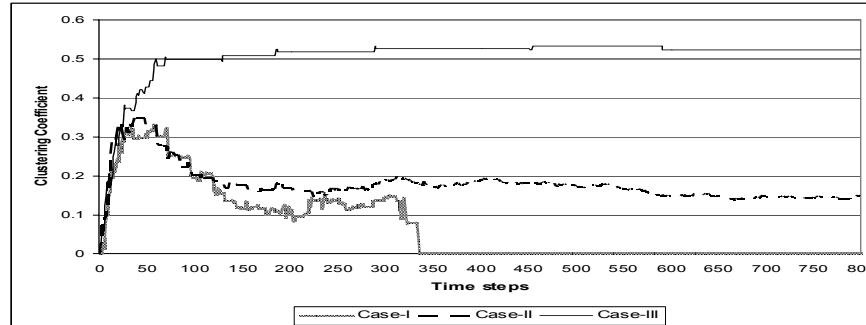


Figure 8: Clustering coefficient of the agents' friendship network for the three different simulation settings.

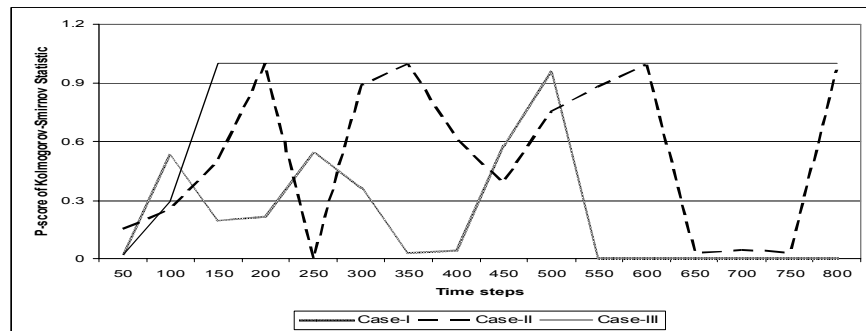


Figure 9: P-score of the K-S test for the degree distribution of the friendship network compared for consecutive snapshots taken at every 50th time step for the three different simulation settings.

4 Understanding Dynamic Agent-based Social Networks

Agent-based social simulations are usually analysed based on a set of hypotheses. One way of testing the hypotheses is observing the time-series charts of 'key' performance measures. An important issue is the sensitivity of parameters as it requires excessive computational workload to explore the parameter space. Another issue is representing the results as aggregates (arithmetic mean, standard deviation, etc.) where the knowledge about the varying trajectories generated at different simulation runs is lost.

Generating time series charts for global measures may not be sufficient to understand the underlying process. On the other hand, models of descriptive social processes driven from evidence are difficult to analyse. In such cases, the population is dependent upon the social processes and no longer remains an invariant. Observing the distributions generated from multiple runs of the simulation may help in guessing the system's behaviour in general. In the case of the second model, the network signatures differed considerably when the number of agents was kept fixed. It would therefore be imperative to understand the factors that lead to the emergence of

networks during a simulation. If one was lucky, one could then identify measures and independent variables that remain valid in most runs.

Nonparametric statistical techniques, and specifically in the context of this paper the Kolmogorov-Smirnov test, are potentially suited in helping to analyse dynamic agent-based networks. These techniques do not assume any prior distribution of the generated data. The classical statistical tests such as the *t-test* inherently assume that the data comes from the normal distribution. However, the condition for normality does not hold in a number of agent-based models of social processes, c.f. [8].

The Kolmogorov-Smirnov (K-S) test is used to test if there is a difference in the distribution of two populations. We have, in this paper, considered different snapshots of the same simulation run as separate populations and compared them using the K-S test. Our main motivation has been to look for techniques that can help keeping track of the network structure over the course of a simulation run. Observing the p-score for consecutive snapshots can indicate whether the network structure has changed or not. A high p-score is an indicator that there is no evidence from the distributions that the network has changed – a difference is possible but this is unlikely. Moreover, significant differences in the p-score could be one possible indicator in selecting a set of snapshots for further analysis.

The general scheme presented in this paper (see appendix) raises the question of choosing the ‘right’ lag size. As we have found previously [1], choosing different lags influences the outcome of the K-S statistic. Over longer periods, one may expect less a greater change in the network’s configuration, while in some cases, zooming in to a small period may reveal more or less subtle changes occurring in the network structure that are otherwise overlooked. As future work, we will look into the prospects of automating this process and determining heuristics for choosing lags.

Unlike other measures such as average degree, clustering coefficient, or mean path length, the K-S statistic can indicate when structural changes have occurred in the network during simulation. It takes into account the network as a whole in many different dimensions at once whereas traditional techniques typically provide an 1-dimensional aggregate measure. The downside is that it does not tell you how they differ if they do, even when there is a low p-score the networks might be differing in what we might consider to be an insignificant way

5 Discussion and Outlook

Choice of suitable analysis techniques for networks changing in terms of number of nodes and links is an interesting question for social simulation modellers. We discussed examples from two agent-based models in which agents’ interactions result in the evolution of social networks. The first by Jin et al. [6] presented a mechanism for the emergence of a small-world network by means of agents’ interaction rules in a population of fixed size. The second model is driven by evidence and intended to understand the impact of HIV/AIDS based on a real case study. Here the population size is dynamic with agents dying and being born during the simulation depending upon social processes. In this paper, we have reported the resulting effects on two network characteristics: the clustering coefficient and the networks’ degree distribution.

The point we want to make here is that for real case study problems, it might be a good idea to check whether the evolved network is due to certain procedures that specifically lead to a particular topology. Networks resulting from social processes may or may not produce topologies such as the *scale-free* or *small-world* topology. For instance, in this paper, we have focused on the evolution of *small-world* networks in agent-based models which are determined in terms of the clustering coefficient; mean shortest path length and degree distribution. We provide a comparison of the clustering coefficient and the application of the Kolmogorov-Smirnov statistic on the degree distribution of the simulated network. For the distance measure (mean shortest path length) similar signatures as those shown in the case of clustering coefficient are obtained. These results have not been included in this paper due to lack of space. The work presented in this paper is in its early stages and we expect to do a detailed investigation of the approach to support our claim.

It seems evident that, in many cases, the size and structure of social networks is contingently constructed by intricate social processes, rather than determined in a rigid and explicit manner. To capture and understand such phenomena it would seem sensible to develop models that embedded such a contingent development of social networks. However this makes the task of choosing suitable techniques for understanding the emergence of networks in such models much more difficult. Scale-independent measures such as nonparametric statistics or motif analysis may be of help in indicating the presence of change in dynamic social networks, which then may be the focus for further investigation

Acknowledgments. We would like to thank our colleagues at the Centre for Policy Modelling, the participants of the 2nd Social Network Analysis Forum (Leeds) and our partners at the CAVES Project for their comments and feedback. This work is supported under the EU FP6 Project CAVES.

References

1. Alam, S.J., R. Meyer and E. Norling. Agent-based Model of Impact of Socioeconomic Stressors: A Dynamic Network Perspective. Sixth Int'l Joint Conf. on Autonomous Agents and Multiagent Systems. Hawaii, USA (forthcoming)
2. Alam, S.J., R. Meyer and E. Norling. Using Agent-Based Modeling to Understand the Impact of HIV/AIDS in the Context of Socio-Economic Stressors. Centre for Policy Modelling Report CPM-06-167 (2006) <<http://cfpm.org/cpmrep167.html>>
3. Albert, R. and A. Barabasi. "Statistical Mechanics of Complex Networks." Rev. Mod. Phys. 74 (2002)
4. Edmonds, B. and E. Chattoe. When Simple Measures Fail: Characterising Social Networks Using Simulation. Social Network Analysis: Advances and Empirical Applications Forum. Oxford, UK (2005) <<http://cfpm.org/cpmrep158.html>>
5. Faust, K. Comparing Social Networks: Size, Density, and Local Structure. Metodološki zveski 3 (2) (2006): 185-216
6. Jin, E.M., M. Girvan and M.E.J. Newman. Structure of growing social networks. Phys. Rev. E. 64.4 (2001)
7. Kirkman, T.W. Statistics to Use (1998) <<http://www.physics.csbsju.edu/stats/>> (accessed last on 14/4/2007).

8. Moss, S. and B. Edmonds. *Sociology and Simulation: Statistical and Qualitative Cross-Validation*. American Journal of Sociology 110 (4). The University of Chicago Press (2005): 1095-1131
9. Neave, H and K. McConwa. *Distribution free methods*. Open University Press (1987)
10. Newman, M.E.J. The structure and function of complex networks. *SIAM Review* 45 (2004): 167-256
11. North, M.J., N.T. Collier and J.R. Vos. Experiences Creating Three Implementations of the Repast Agent Modeling Toolkit." *ACM Transactions on Modeling and Computer Simulations*. 16 (1). The ACM Press (2006): 1-25
12. Pujol, J.M., et al. How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges. *Journal of Artificial Societies and Social Simulation* 8 (4) (2005): 12
<<http://jasss.soc.surrey.ac.uk/8/4/12.html>>
13. Wasserman, S. and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press (1994)
14. Watts, D.J. and S. Strogatz. Collective Dynamics of 'small world' networks. *Nature* 393 (1998): 440-442
15. Watts, D.J. *Six Degrees: The Science of a Connected Age*. W.W. Norton & Company (2003)
16. Ziervogel, G., et al. "Adapting to climate, water and health stresses: insights from Sekhukhune, South Africa." *Stokholm Environment Institute (Oxford)* (2006)

Appendix: A scheme for comparing changes in the network structure during a simulation run

- I. Choose a lag of size Δ for the comparison of network snapshots at different time steps.
- II. For all time lags t_i ($i \leftarrow 0, \Delta, 2\Delta, \dots$), obtain a corresponding series of the node degrees, P_i . (We call this the profile)
- III. For all consecutive pairs $(P_i, P_{i+\Delta})$ ($i \leftarrow 0, \Delta, 2\Delta, \dots$), compute the p-score for the Kolmogorov-Smirnov (K-S) test for 2-populations.

The K-S test gives an indication of the likelihood that the two series are from the same distribution. A low p-score indicates that it is unlikely that the two series are from the same distribution, and hence that the degree distributions and hence the network type differs in some respect. A high p-score would mean that there is no evidence that the distributions are different and hence is a fairly tough test of network similarity. However, it is still possible that the networks do differ, albeit in subtle ways. The web-page we used to do this test is at <<http://www.physics.csbsju.edu/stats/>> (accessed last on 14/4/2007).