

Modelling the structure and dynamics of network-based social systems

P. E. Pattison^a, G. L. Robins^a

^a *University of Melbourne, Parkville, Victoria*
Email: pepatt@unimelb.edu.au

Abstract: The development of policy interventions to enhance the sustainability of our future often depends on the capacity to understand and model relevant social systems and to predict the consequences of policy change. To this end, we describe in this paper a suite of statistical models for the structure and dynamics of network-based social systems. In these models, global network structure is hypothesised to arise as the outcome of dynamic, interactive processes occurring within local neighbourhoods of a network. A hierarchy of such processes and model specifications are described. Models can be estimated from a number of possible data structures, including single or panel observations, and partial data structures obtained through certain types of network sampling schemes.

Using several illustrative applications, we describe how the models can be used to extend our understanding of social systems and build theoretically plausible and empirically-grounded simulation models. Advantages of the approach include a capacity to quantify uncertainty and its implications, and the facility to assess the extent to which a model captures features of the social system not parameterized in the model.

We conclude with a sketch of future work designed to improve our understanding of ways to enhance sustainability within network-based social systems.

Keywords: *Social networks, Network-based social systems, Statistical models, Exponential random graph models, Social processes, Network dynamics, Co-evolutionary processes*

1. INTRODUCTION

As many commentators are increasingly aware, the design of effective policy interventions can often depend on an understanding of the interactive, dynamic and socially-situated processes from which complex social systems derive their properties. By *dynamic* and *interactive*, we mean that the actions of one individual in a social system may both depend on and, in turn, influence the actions of others. By *socially-situated*, we mean that actions depend on a multi-layered complex of social entities that includes social relations, group affiliations, social settings and spatial neighbourhoods. These aspects of social “location” both constrain and provide opportunities for possible future actions. Further, an action by one individual may change the context for other individuals in neighbouring locations, so that a dynamic, interactive characterisation of social processes necessarily implicates an understanding of social location.

Social networks can play a particularly important role in these processes. Different types of networks serve as channels through which, for example, knowledge is diffused, opportunities are recognised, influence is exerted, support is offered, stereotypes are formed, and actions are coordinated. Yet this simple insight reflects the inherent complexity of social systems: the macro-level behaviour of a social system is critically dependent on the precise nature of self-organising network-based social processes occurring at local levels of interaction. In order to understand such social systems, we need a more precise appreciation of the nature of these local interactions and of the ways in which they cumulate to determine system-level properties. Indeed, a fuller recognition of the interdependent nature of social processes brings with it a *quantitative* imperative: relatively precise characterisations of local interactive processes are generally required in order to understand their implications at an aggregate or *global* system level.

This paper outlines a systematic suite of statistical models that allow us to explicate and understand these interactive local network-based processes and their system-level consequences. The models can be estimated from appropriate data on the system of interest, and provide theoretically principled and empirically defensible accounts of network structures, network-based social processes and network dynamics. They have been shown to be capable of reproducing a variety of important local and global features of such systems, including properties of social networks (e.g., Snijders et al., 2006; Robins et al., 2007; Goodreau, 2007), the distribution of individual characteristics across networks (Robins et al., 1999), and changes in network structure over time (Snijders and Koskinen, in press). In addition, it is possible to embed models for these network-based interactive processes within geographical space and within higher-level social settings (e.g. Daraganova et al., in press; Wang et al., 2011).

The development of this suite of models has been enabled by four important and interrelated developments:

- a principled approach to quantitative model specification for emergent higher order relational structures (Pattison and Robins, 2002; Snijders et al., 2006; Pattison et al., 2011a);
- compelling evidence that models for observed social networks typically *require* the inclusion of such emergent higher order effects (Robins et al., 2007; Snijders et al., 2006);
- a capacity to construe models for cross-sectional observations as equilibrium models for dynamic processes for the evolution and co-evolution of networks and network processes and to estimate dynamic models from longitudinal observations of networks and relevant node-level characteristics (e.g., Snijders, 2001; Snijders and Koskinen, in press; Steglich et al., in press);
- application of Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE) methods and their extension to adaptive sampling designs and designs with data missing at random (Snijders, 2002; Handcock and Gile, 2010; Hunter and Handcock, 2004; Koskinen et al., 2010; Pattison et al., 2011b).

2. MODELLING FRAMEWORK

We outline the modelling framework by describing three types of data structure and a general model form associated with each. These three approaches can then be combined to model more complex structures.

2.1. Models for Social Networks

The first data structure is a (social) network, defined on a fixed set $N = \{1, 2, \dots, n\}$ of social actors, or *nodes*, who are assumed to be linked by various forms of relational *tie*. A potential tie between two nodes, i and j , is regarded as a discrete random variable ($X_{ij} = 1$ if the tie is observed, $X_{ij} = 0$, otherwise). If X_{ij} and X_{ji} are (not) distinguished, the ties are (*non-*)*directed*. A *random network* (or *random graph*) is represented by a two-way

array $\mathbf{X} = [X_{ij}]$ of these random variables; $\mathbf{x} = [x_{ij}]$ denotes a particular set of realised values. The models can be generalised to discrete-valued observations and multiple types of tie (Frank and Strauss; 1986; Pattison and Wasserman, 1999; Wasserman and Pattison, 1996; Robins et al., 1999; Robins et al, 2009).

Two variables are termed *neighbours* if they are assumed to be conditionally dependent given the values of all other variables. Neighbourhood relations can be represented by a *dependence graph* \mathbf{D} whose nodes are the random variables X_{ij} and whose edges link pairs of neighbouring variables. The Hammersley-Clifford theorem (Besag, 1974) can then be applied to any hypothesis about the neighbourhood structure to determine the general parametric form for a probability model, $\Pr(\mathbf{X} = \mathbf{x})$. Non-zero parameters pertain to subsets A of variables for which every pair of variables are neighbours; these subsets are termed *local neighbourhoods* (Pattison and Robins, 2002). Specifically:

$$\Pr(\mathbf{X} = \mathbf{x}) = \exp(\sum_A \lambda_A z_A(\mathbf{x})) / \kappa \quad (1)$$

where:

- A is a local neighbourhood defining a *configuration* of possible ties;
- λ_A is a parameter associated with the neighbourhood A ;
- the quantity $z_A(\mathbf{x})$ is a *relational statistic* indicating whether all ties in A are present in \mathbf{x} ;
- and κ is a normalizing quantity.

Given many overlapping neighbourhoods A , the model expresses the probability of a network as a function of self-organising processes in local regions of the network. In *homogeneous* models, parameters for isomorphic configurations are equated; a positive value of λ_A then indicates that networks with more configurations of type A are more probable.

Equation (1) is the general form of an *exponential random graph model* (ERGM) for social networks. The attraction of models of this form is that, unlike many alternatives, they can incorporate important features of observed social networks such as *clustering* (the tendency for two actors sharing a tie to a third party to themselves be tied) and dispersion in the degree distribution (the distribution of the number of network partners across nodes in the network). A critical step is to specify the dependence structure, since the neighbourhood relation determines the *form* of network configurations in the model.

The simplest topology assumes that possible ties are independent and occur with constant probability, as in the *Bernoulli graph* model (Erdős and Renyi, 1959). The implausibility of this assumption for social systems led Frank and Strauss (1986) to propose *Markov random graph models*, whereby X_{ij} and X_{lm} are assumed to be neighbours when they share a node ($\{i,j\} \cap \{l,m\} \neq \emptyset$). Pattison and Robins (2002) argued that Markov models were simplistic, and this has since been confirmed by the difficulty in obtaining convergent MCMCMLE estimates for actual network data (Handcock, 2003; Snijders, 2002; Snijders et al., 2006). Pattison and Robins (2002) introduced higher order processes hypothesised to influence network topologies. They argued that neighbourhoods may recursively emerge from the network processes themselves, with new neighbourhoods created as network ties are generated. For instance, X_{ij} and X_{kl} might become neighbours if there is an observed tie between j and k or between l and i ; such assumptions lead to a class of *realisation-dependent* network models (Baddeley and Möller, 1989). Snijders et al. (2006) utilised a realisation-dependent topology in which X_{ij} and X_{kl} become neighbours only if there is an observed tie between j and k and between l and i . This assumption gives rise to various higher-order neighbourhoods, the most important being k -triangles (a configuration of k triangles each sharing a common base). The resulting higher order models fit dramatically better than Markov models and suggest the importance of realisation-dependent model forms.

More recently, Pattison et al. (2011b) have described a systematic framework for characterising the dependence assumptions that may underlie ERGM model specifications and hence for also classifying their associated network effects. They observed that the various dependence assumptions for a pair of tie variables X_{ij} and X_{kl} that have been proposed to date may be differentiated according to the *form* of proximity among the two node pairs and the *distance* in terms of which proximity is judged. These two features serve to characterise the graph-theoretical proximity of the two node pairs $\{i,j\}$ and $\{k,l\}$ for which the tie variables X_{ij} and X_{kl} are assumed to be conditionally dependent. Pattison et al. (2011b) show that the resulting hierarchy of models can be used to capture network effects that reflect the degree to which networks exhibit structural patterns associated with cohesion, closure or exchange, brokerage and connectivity. As a result, ERGMs provides a very broad framework for building effective models for social networks.

A further generalization of the model arises when relational variables are conceptualized as links or contacts from a set of social actors to a set of organisations. In this case, a potential contact tie X_{ik} between an individual i and an organisation k is regarded as a discrete random variable, with value 1 if the individual is in the organisation and 0 otherwise. $\mathbf{X} = [X_{ik}]$ is then a *random bipartite network*. Wang et al. (2009) described models for a bipartite version of the social circuit hypothesis; and Wang et al. (2011) have extended these models to more complex data and model forms in ways associated with the dependence hierarchy described above.

2.2. Exogenous Covariates

The second data structure of interest is one in which observations of a network are accompanied by observed exogenous characteristics of the nodes (or possibly pairs of nodes) of the network. Exogenous variables may be assumed to affect network variables and so interact with network effects in ways determined from an extended dependence graph incorporating directed dependencies. As Robins et al. (2001) demonstrated, the addition of exogenous covariates leads to models of the form:

$$Pr(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y}) = \exp(\sum_Q \sum_P \lambda_{P,Q} z_P(\mathbf{x}) w_Q(\mathbf{y})) / \kappa \quad (2)$$

where \mathbf{Y} denotes a collection of (often node-level) covariate variables, and \mathbf{y} a corresponding array of observations; the summation is over all subsets P of interdependent tie variables, and over all subsets Q of covariate variables hypothesised to affect each variable in P ; $\lambda_{P,Q}$ is a parameter associated with the subsets P and Q ; $z_P(\mathbf{x})$ and $w_Q(\mathbf{y})$ are *statistics* associated with P and Q (taken to have the value 1 if P or Q are empty); and κ is a normalizing quantity.

Reversing the role of \mathbf{Y} and \mathbf{X} in model (2) leads to models for the outcome of network-based processes, such as *social influence* and network-based *diffusion* processes (Robins et al., 2001). These models then reflect processes in which the course and outcome of interpersonal interactions depend on network structures. For example, they can capture the outcome of a variety of interpersonal encounters in which a social product of some form (e.g., knowledge, attitude, commodity or pathogen) is transmitted from one person to another.

2.3. Dynamic Models

The third important data structure comprises longitudinal, or panel, observations of a network. The basic framework of equation (1) can then be generalised to model network evolution. Suppose the relational variables $X_{ij}(t)$ are now time-dependent and that, at any time t , there is a possible change in status for some randomly chosen tie, X_{ij} (i.e. from 0 to 1, or 1 to 0). Suppose also that the network changes with transition rate $\rho \text{logistic}(\sum_P \lambda_P (z_P(\mathbf{x}_{ij}^*(t)) - z_P(\mathbf{x}(t))))$, where: $\mathbf{x}(t)$ denotes the state of the network at time t ; $\mathbf{x}_{ij}^*(t)$ equals $\mathbf{x}(t)$ but with the value of $X_{ij}(t)$ changed from $x_{ij}(t)$ to $1-x_{ij}(t)$; ρ is a rate parameter and the logistic function is defined by $\text{logistic}(z) = \exp(z) / (1 + \exp(z))$.

This stationary continuous-time Markov process converges to the distribution of equation (1) and is an adaptation of the approach of Snijders (2001), who was the first to develop continuous-time Markov chain models for network evolution that relax the implausible assumption of *dyad-independence*. Snijders and Koskinen (in press) describe how to estimate parameters for this model from longitudinal investigations on a network.

Longitudinal models can also be developed for cross-level interactions. When cross-level interactions involve individual and tie variables, these models instantiate both social selection and social influence processes in which, respectively, changes in tie status depend on characteristics of the actors involved, and actor characteristics change in response to characteristics of their tie partners. As noted in 2.2 above, separate selection and influence models have been developed from equation (1), but in reality influence and selection processes are likely to occur simultaneously. The co-evolution model of Snijders and colleagues (e.g. see Steglich et al., in press) provides a means of disentangling selection and influence effects through analysis of longitudinal observations.

2.4. Network Sampling and Missing Data

Where social systems are large, the task of obtaining data on whether each pair of actors is linked by a tie can be a daunting and expensive one. It is therefore of interest to develop means of estimating models such as equation (1) from suitable sampling schemes. For example, the technique of snowball sampling was devised in the 1940s by the sociologist Paul Lazarsfeld to observe parts of a network that might be important to social influence processes (Barton, 2001). Recent methodological developments have included a likelihood based method for estimating the model of equation (1) in the case of network sampling schemes and data that is

missing at random (Handcock and Gile, 2010; Koskinen et al., 2010); a conditional estimation approach has also been developed specifically for snowball sampling designs (Pattison et al., 2011b).

2.5. Modelling Complex Social Systems

Putting together the approaches just described yields a capacity to build models for social systems with multiple types of social actors (e.g. individuals, groups, organisations of different kinds), each of which might have a variety of different types of characteristics, and amongst whom there are a number of potential connections of different types. Models can be built that are not only theoretically informed but also described by parameters estimated from data. In some cases, models can even be estimated where only partial observations have been made of the set of variables that constitute the system. Moreover, if we observe the system on multiple occasions and so have access to longitudinal observations, we can specify a continuous-time Markov process with parameters that can also be estimated from data, and whose equilibrium form has the same structure as models estimated from cross-sectional observations.

Together, these capacities allow us to use data to build models for social systems that have a number of advantages, many of which accrue from the general statistical framework within which the modelling approach is set. In particular:

- we can estimate the uncertainty of parameters, and use this information to simulate possible states of the system in a way that takes account of uncertainty (e.g. Snijders et al., 2006; Wang et al., 2009);
- we can simulate the social system, and compare the properties of simulated states of the system with observed data, providing the means to assess how well the model captures features of the system that are not directly parameterised in the model (e.g., Hunter et al., 2008);
- if we make longitudinal observations of the system, we can model the evolutionary and co-evolutionary processes more directly, and hence assess the degree to which such processes are captured by cross-sectional models (Snijders and Koskinen, in press);
- provided that the model does a sufficiently good job of capturing aspects of the system which we believe to be important, we can use it as a baseline representation of the system for various forms of scenario modeling.

3. ILLUSTRATIVE APPLICATIONS

In this section, we briefly describe several case studies which illustrate the ways in which the models and their extensions described above can be applied to understand social systems, and to serve as scenario testing models.

3.1. Environmental Governance

Robins et al. (in press) analysed networks of ties among organisations involved in management of the Swan River in Western Australia. The relevant data comprised relations among 21 organisations charged with primary collective responsibility for river management through a 2004 framework for collective management known as Riverplan (W. A. Government, 2004). The relational data were reported by key organisational informants who judged the importance and ease of collaboration of their ties to other organisations that were relevant to a specified set of responsibilities in relation to river management. Robins et al. (in press) argued that effective network governance required a number of structural preconditions, including network structures that can facilitate effective coordination of action, such as relational and structural embeddedness and agreement among network actors about goals and actions. They demonstrated through their analysis of the observed governance network for the Swan River that these preconditions were not in fact met and that, in particular, the system did not exhibit relational and structural embeddedness among constituent organisations, and relations judged to be crucial involved significant levels of contestation. They concluded that organisations making up the governance system appeared to be pursuing different goals. Legislative changes for environmental governance of the river have since been made, the need for which was supported by the empirical analysis reported in Robins et al. (in press).

3.2. Diffusion of Sustainable Farming Practices

Lusher et al. (2010) used a snowball sampling design in rural regions in New South Wales and Western Australia to identify farmers' networks of advice and approaches to sustainable farming practices. Beginning with a random sample of farmers in each selected district, they asked farmers to identify sources of advice,

and then recruited the identified individuals into the sample. This process was repeated several times. The resulting data constitutes a partial observation of the system of interest, and Lusher et al. (2010) used the conditional estimation approach described earlier to model the distribution of sustainable farming practices as a function of the measured advice network (e.g. Daraganova and Pattison, 2011). Lusher et al. (2010) were able to identify interactions between the distribution of sustainable practices and the advice network, and hence formulate valuable hypotheses for exploration in future longitudinal studies.

4. DISCUSSION AND CONCLUSIONS

The two examples in Section 3 illustrate the use of cross-sectional designs to study components of a larger social system that are relevant to sustainability. Such designs are valuable as they inform our detailed understanding of the behavior of important social sub-systems. Yet, ideally we need to explore the processes by which various components of such sub-systems interact, and verify the interactions through multiple observations of the system. As a result, longitudinal, multi-level designs are likely to yield particularly valuable insights.

4.1. Climate Adaptation in Multi-level Governance Environments

With such goals in mind, Robins et al. (2010) have analysed research designs that are likely to be most valuable in enhancing our understanding of the acceptance, adoption and diffusion of adaptive behaviours. They propose as key components of such a study: regional organisations, including local government; the adaptive capacity of these organisations; the organisational affiliations of individuals within regions; and attitudes and behavioural adaptation at the individual level. Although the studies they recommend are yet to be conducted, they hold substantial promise for the development of scenario models that can then be used to predict the impact of policy interventions proposed to enhance sustainability.

ACKNOWLEDGMENTS

The authors are grateful to the Australian Research Council under grants DP0665261 and DP0987730, the University of Melbourne-CSIRP Collaboration Fund, the PUTTI project and the CSIRO Flagship Collaboration Fund for support of the work described here and to Mr Peng Wang, Dr Galina Daraganova, Dr Johan Koskinen and Dr Dean Lusher for their contributions to it.

REFERENCES

- Baddeley, A., and Möller, J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, 57, 89-121.
- Barton, A. H. (2001). Paul Lazarsfeld as institutional inventor. *International Journal of Public Opinion Research*, 13, 245-269.
- Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36, 96-127.
- Daraganova, G., and Pattison, P. (in press). An ALAAM analysis of unemployment: the importance of who you know and where you live. In Lusher, D., Koskinen, J., and Robins, G. (Eds.), *Exponential random graph models for social networks: Theory, method and applications*. Cambridge Uni. Press.
- Daraganova, G., Pattison, P., Robins, G., Koskinen, J., Mitchell, B., Bill, A., Watts, M., and Baum, S. (in press). Networks and geography: modelling community network structures as the outcome of both spatial and network processes. *Social Networks*.
- Erdős, P., and Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6, 290-297.
- Frank, O., and Strauss, D. (1986). Markov graphs. *Journal of the Amer. Statistical Association*, 81, 832-842.
- Goodreau, S. (2007). Advances in exponential random graph (p^*) models applied to a large social network. *Social Networks*, 29, 231-248.
- Handcock, M. S. (2003). *Assessing degeneracy in statistical models of social networks*. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 39.
- Handcock, M. S., and Gile, K. J. (2010). Modelling networks from sampled data. *Annals of Applied Statistics*, 4, 5-25.
- Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008) Goodness-of-fit of social network models. *Journal of the Amer. Statistical Association*, 103, 248-258.
- Hunter, D., and Handcock, M. S. (2006). Inference in curved exponential families for networks. *Journal of Computational and Graphical Statistics*, 15, 565-583.

- Koskinen, J. H., Robins, G. L., and Pattison, P. E. (2010). Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation. *Statistical Methodology*, 7, 366-384.
- Lusher, D., Tucker, D., Green, M., Dzidic, P., Bates, L., Leviston, Z., Robins, G., and Pattison, P. (2010). *Social influence and environmentally sustainable land management behaviour in Australia*. Paper presented at the XXX International Sunbelt Social Network Conference. Riva del Garda, Italy.
- Pattison, P. E., and Robins, G. L. (2002). Neighbourhood-based models for social networks. *Sociological Methodology*, 32, 301-337.
- Pattison, P. E., Robins, G. L., Snijders, T. A. B., and Wang, P. (2011a). *A hierarchy of exponential random graph models for social networks*. Working Paper, Psychological Sciences, Uni. of Melbourne.
- Pattison, P. E., Robins, G. L., Snijders, T. A. B., and Wang, P. (2011b). *Conditional estimation of exponential random graph models from snowball sampling designs*. Working Paper, Psychological Sciences, Uni. of Melbourne.
- Pattison, P., and Wasserman, S. (1999) Logit models and logistic regressions for social networks, II. Multivariate relationships. *British Journal of Mathematical and Statistical Psychology* 52: 169-193.
- Robins, G., Bates, L., and Pattison, P. (in press). Network governance and environmental management: conflict and cooperation. *Public Administration*.
- Robins, G., Bates, L., Pattison, P., and Lusher, D. (2010). Climate adaptation behaviours in multi-level governance environments: a proposed network-based research design. Report to Climate Adaption Flagship.
- Robins, G., Elliott, P., and Pattison, P. (2001). Network models for social selection processes. *Social Networks*, 23, 1-30.
- Robins, G., Pattison, P., and Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, 66, 161-190.
- Robins, G. L., Pattison, P. E., and Wang, P. (2009). Closure, connectivity and degrees: new specifications for exponential random graph models for directed social networks. *Social Networks*, 31, 105-117.
- Robins, G., Pattison, P., and Wasserman, S. (1999) Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika*, 64, 371-394.
- Robins, G.L., Snijders T. A. B., Wang, P., Handcock, M. S., and Pattison, P. E. (2007). Recent developments in exponential random graph (p*) models. *Social Networks*, 29, 192-215.
- Snijders, T. A. B. (2001). The Statistical Evaluation of Social Network Dynamics. Pp. 361-395 in Sobel, M. E. and Becker, P. (Eds.), *Sociological Methodology 2001*. Boston and London: Basil Blackwell.
- Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2). <http://www.cmu.edu/joss/content/articles/volindex.html>.
- Snijders, T. A. B., and Koskinen, J. (in press). Longitudinal models. In Lusher, D., Koskinen, J., and Robins, G. (Eds.), *Exponential random graph models for social networks: Theory, method and applications*. Cambridge University Press.
- Snijders, T.A.B., Pattison, P., Robins, G.L., and Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36, 99-153.
- Steglich, C. E. G., Snijders, T. A. B., and Pearson, M. (in press). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*.
- W. A. Government (2004). *Riverplan : an environmental management framework for the Swan and Canning Rivers: comprehensive management plan and implementation strategy for the Environmental Protection (Swan and Canning Rivers) Policy 1998*.
- Wang, P., Robins, G., and Pattison, P. (2009). *PNet: Program for the estimation and simulation of p*exponential random graph models, User Manual*. Dept. of Psychology, University of Melbourne.
- Wang, P., Robins, G., and Pattison, P. (2011). *Exponential random graph models for multi-level networks*. ASNA Conference, Zurich, September, 2011.
- Wang, P., Sharpe, K., Robins, G., and Pattison, P. (2009). Exponential random graph (p*) models for affiliation networks. *Social Networks*, 31, 12-25.
- Wasserman, S., and Pattison, P. E. (1996). Logit models and logistic regressions for social networks, I. An introduction to Markov graphs and p^* . *Psychometrika*, 61, 401-425.