+Model
SON-521; No. of Pages 4
ARTICLE IN PRESS

Available online at www.sciencedirect.com
ScienceDirect

Social Networks xxx (2006) xxx–xxx

SOCIAL NETWORKS

www.elsevier.com/locate/socnet

Editorial

Advances in exponential random graph (p^*) models

Statistical approaches to social networks have a rather long history. Few now recall that Moreno and Jennings (1938) in the very first edition of *Sociometry* compared an observed network against a null distribution of graphs to investigate the prevalence of what they called social *configurations* (long before the much later terms *motif* or *subgraph count* were introduced). Such use of statistically based algorithms, concepts and arguments has been extremely helpful in understanding certain features of observed social networks.

But to specify a successful statistical *model* for a social network has been a more difficult endeavour. The requirements for a successful model are demanding, extending beyond the use of particular statistical techniques to identify given properties of a network. A good model needs to be both estimable from data and a reasonable representation of that data, to be theoretically plausible about the type of effects that might have produced the network, and to be amenable to examining which competing effects might be the best explanation of the data. Models that cannot be estimated from data, that cannot reproduce the data to some adequate extent, models with implausible or empirically unsound assumptions, and models that do not permit decisions between alternative hypotheses, cannot be considered complete. Yet they may have value as thought experiments, or as steps along the path to improved specification. This has been, and continues to be, the trajectory for the statistical modelling of social networks.

In this special edition of *Social Networks*, we argue that, due to very recent progress in the framework of exponential random graph models, we are now much closer to the goal of obtaining good statistical models for social networks than we have ever been before. With new specifications for these models, new estimation techniques and new software, we can now reliably estimate parameters that describe established human social processes (albeit sometimes in different forms from their classical representations). Using these parameter estimates as a basis for simulation, we can reproduce features of the observed network to an extent previously not possible. Finally, we can make informed inferences about competing processes that may be important in producing the network.

A major initial attempt at statistical modelling of social networks was the Bernoulli random graph distribution proposed independently by Erdös and Renyi (1959) and Rapoport (1953). Despite the importance of this model in modern random graph theory, its underlying assumption is that network edges are independent of each other, clearly implausible in almost all human social networks. In moving to dyad-independence with their p_1 model, Holland and Leinhardt (1981) set out what they saw as the program for future research. Dyad-independence was only a start, they recognized; the demand was for ways to represent more complex dependence in social networks.

0378-8733/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.socnet.2006.08.004

A crucial breakthrough was made by Ove Frank and his colleague, David Strauss, who realized that approaches from spatial statistics and statistical mechanics could be translated to social network contexts. They developed models that went beyond dyad-independence, with assumptions that for the first time could be viewed as empirically and theoretically plausible. Unfortunately, their paper on Markov random graphs (Frank and Strauss, 1986) was not given much initial attention by social network researchers. In the second half of the 1990s, Stanley Wasserman and Pip Pattison recognized the value of Frank and Strauss's work and introduced Markov random graphs and further generalizations to the social networks field as p^* models (Pattison and Wasserman, 1999; Robins et al., 1999; Wasserman and Pattison, 1996).

At the same time, there was growing interest in statistical models for other types of social network data, especially models for multiple observations of networks across time (Snijders, 2001). Nevertheless, for single network observations, p^* or exponential random graph models, as they came to be known, remained the major statistical approach, and indeed they have close links with certain of the longitudinal models.

But with more focussed attention, it became apparent that Markov random graph models presented serious problems. It is appealing to think of a network as a complex system—and a Markov random graph model in effect implies such complexity. But a complex system brings with it accompanying notions of feedback, attractors and degeneracy, and Markov graph models were not quite up to the task of dealing with real data. To a large extent, these problems were masked in the late 1990s by the then frequently used and often inadequate pseudo-likelihood estimation techniques. The development of more principled Monte Carlo maximum likelihood estimation procedures, and of effective means of model simulation, as well as the implementation of these in publicly available software, has meant that we now understand very much better how and when Markov graph models fail.

The four papers in this special edition derive from new specifications for exponential random graph models proposed by Snijders et al. (in press) and developed further by Hunter and Handcock (2006). These new specifications are intended to overcome the known problems of Markov random graph models and build on the idea proposed by Pattison and Robins (2002) that dependencies among network ties may be context-dependent. They include new formulations of classic network ideas relating to degree distributions, transitivity, and multiple connectivity. The goals of this special edition are to provide a simplified introduction to exponential random graph models in general, to the new specifications in particular, to describe some of the underlying statistical issues, and to illustrate how successful the new models are in practice. In short, we claim that these innovations provide a greatly improved scope for the statistical modelling of social networks.

The first paper in this special edition (Robins et al., 2006a) provides an introductory summary to the formulation and application of the general class of exponential random graph models. The authors begin by discussing motivations for the statistical modelling of social networks, asking why researchers might want to go beyond established analytic techniques and search for a well-fitting *model* of an observed social network, and in particular a *statistical* model. They explain how dependence assumptions determine the general form of the model and give examples of different dependence assumptions. They introduce Markov random graphs and discuss estimation of the models. This first paper is intended to review developments up to the point of the new specifications.

The second paper (Robins et al., 2006b) reviews the new specifications proposed by Snijders et al. (in press). The authors illustrate how simple Markov random graph models may be inadequate as models for empirical data. They fit models with the new specifications to a large number of classical small-scale network data sets, illustrating a dramatically improved performance over

Markov graph models. The authors review new software for obtaining Monte Carlo maximum likelihood estimates of model parameters and compare these maximum likelihood estimates with less reliable pseudo-likelihood estimates.

The third paper (Hunter, 2006) addresses some of the technical statistical issues for models with the new specifications. Hunter shows how the new specifications represent curved exponential models, and discusses methods for estimating parameters within this context. He introduces the shared partner distributions, which are new social network concepts relating to transitivity and to multiple connectivity. Analogous to degree distributions, these new distributions are important indicators of structural regularities. Hunter shows how the new specifications can be interpreted in relation to the degree and shared partner distributions.

The final paper (Goodreau, 2006) applies the advances described in the previous papers to fit models to a school network with over 1500 nodes. The adequacy of models is assessed through comparison of model predictions with the observed data on a variety of network statistics, interpreted as goodness of fit criteria. Markov models are inadequate; whereas the new specifications, especially when combined with attribute effects, provide good fit to the data. Using the final model estimates, Goodreau simulates to produce networks consistent with the observed network on many statistics, including the number of triangles, the size of the largest component, overall reachability, the distribution of geodesic distances, the degree distribution, and the shared partner distribution. It is a challenge to other modelling approaches to reproduce such an array of features of the observed network. For instance, degree-only models did a poor job of capturing observed network structure, and when added to other effects, a full parameterization of the degree distribution actually worsens fit, rather than improves it. The implication seems to be that degree-based effects for this data are best considered as outcomes of network processes, rather than the basis of network formation. Goodreau's conclusion – that this network can be best understood through a few simple localized processes, notably attribute effects of various types and higher order transitivity effects – is provocative and exciting. As he points out, if this result generalizes, then new possibilities for network sampling may be opened up.

We know from experience that these new specifications are not the answer for all network data. But they perform extremely well in many cases, and markedly better than earlier models. The task of model formulation does not stop here; rather, the new specifications illustrate how this class of models may be further generalized in the future. But even at this stage, it is easy to agree with Goodreau that the ability to fit such models to large datasets and to make inference about the underling processes generating the network represents a major advance in the field of statistical network analysis.

References

Erdös, P., Renyi, A., 1959. On random graphs I. Publicationes Mathematicae (Debrecen) 6, 290–297.

Frank, O., Strauss, D., 1986. Markov graphs. Journal of the American Statistical Association 81, 832–842.

Goodreau, S., 2006. Advances in exponential random graph (p^*) models applied to a large social network. Social Networks, doi:10.1016/j.socnet.2006.08.001.

Holland, P.W., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76, 33–65.

Hunter, D.R., 2006. Curved exponential family models for social networks. Social Networks, doi:10.1016/j.socnet.2006.08.005.

Hunter, D., Handcock, M., 2006. Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics 15, 565–583.

Moreno, J.L., Jennings, H.H., 1938. Statistics of social configurations. Sociometry 1, 342–374.

Please cite this article in press as: Robins, G., Morris, M., Advances in exponential random graph (p^*) models, Social Networks (2006), doi:10.1016/j.socnet.2006.08.004

Pattison, P.E., Robins, G.L., 2002. Neighbourhood-based models for social networks. Sociological Methodology 32, 301–337.

Pattison, P.E., Wasserman, S., 1999. Logit models and logistic regressions for social networks. II. Multivariate relations. British Journal of Mathematical and Statistical Psychology 52, 169–194.

Rapoport, A., 1953. Spread of information through a population with a socio-structural bias. I. assumption of transitivity. Bulletin of Mathematical Biophysics 15, 523–533.

Robins, G.L., Pattison, P.E., Wasserman, S., 1999. Logit models and logistic regressions for social networks. III. Valued relations. Psychometrika 64, 371–394.

Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2006. An introduction to exponential random graph (p^*) models for social networks. Social Networks, doi:10.1016/j.socnet.2006.08.002.

Robins, G.L., Snijders, T.A.B., Wang, P., Handcock, M., Pattison, P., 2006. Recent developments in exponential random graph (p^*) models for social networks. Social Networks, doi:10.1016/j.socnet.2006.08.003.

Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. In: Sobel, M.E., Becker, M.P. (Eds.),The statistical evaluation of social network dynamics. Sociological Methodology. Basil Blackwell, Boston, pp. 361–395.

Snijders, T.A.B., Pattison, P., Robins, G.L., Handock, M., in press. New specifications for exponential random graph models. Sociological Methodology.

Wasserman, S., Pattison, P.E., 1996. Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p^*. Psychometrika 61, 401–425.

Garry Robins *
School of Behavioural Science,
University of Melbourne, Australia

Martina Morris
Department of Sociology,
University of Washington, United States

* Corresponding author.
E-mail address: garrylr@unimelb.edu.au
(G. Robins)

Please cite this article in press as: Robins, G., Morris, M., Advances in exponential random graph (p^*) models, Social Networks (2006), doi:10.1016/j.socnet.2006.08.004