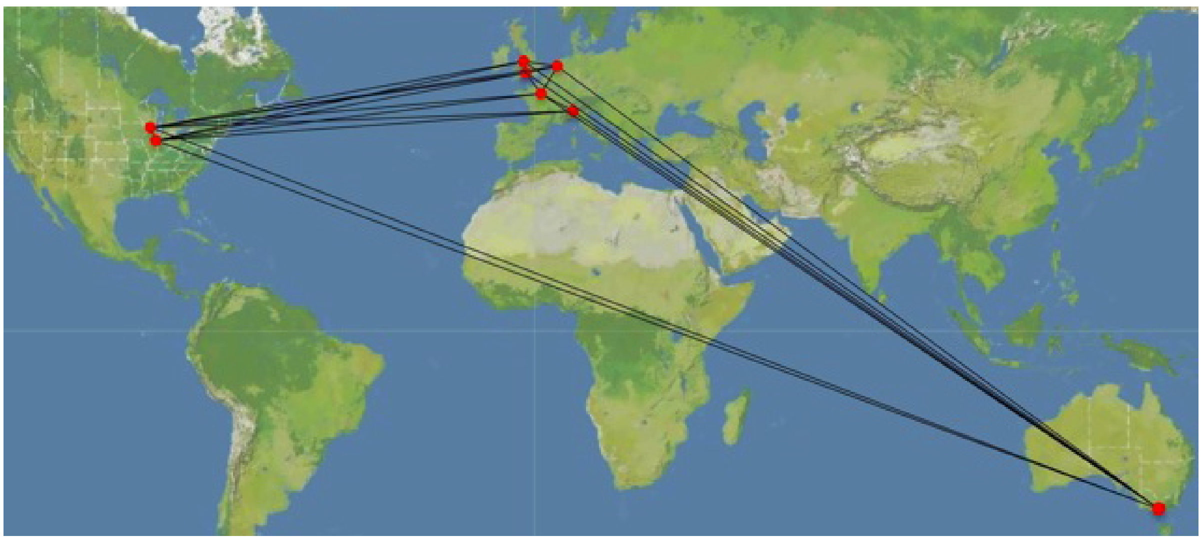
Multilevel Network Modelling Group (MNMG)

Project Guide, March 2010



The Leverhulme Trust

1 The Multilevel Network Modelling Group

- Mark Tranmer, University of Manchester, UK - *Principal investigator*
- Mark Elliot, University of Manchester, UK
- Tom Snijders, University of Oxford/Groningen, UK/NL
- Johan Koskinen, University of Oxford, UK
- Emmanuel Lazega, University of Paris-Dauphine, FR
- Rafael Wittek, University of Groningen, NL
- Alessandro Lomi, University of Lugano, CH
- Noshir Contractor, Northwestern University, USA
- Stanley Wasserman, Indiana University, USA
- Pip Pattison, University of Melbourne, AUS
- Garry Robins, University of Melbourne, AUS
- Margaret Martin, University of Manchester, UK - *Network Administrator*

2 Contact MNMG

Principal Investigator:

Dr Mark Tranmer
CCSR / Mitchell Centre for Social Network Analysis
Social Statistics
School of Social Sciences
Kantorovich Building
University of Manchester
Manchester
M13 9PL
UK
tel.: +44 (0) 161 275 4744
fax: +44 (0) 161 275 4722
mark tranmer “at” manchester.ac.uk
http://www.ccsr.ac.uk/staff/markhome1.htm

Network Administrator:

Margaret Martin
CCSR
Social Statistics
School of Social Sciences
Kantorovich Building
University of Manchester
Manchester
M13 9PL
UK
tel.: +44 (0) 161 275 4721
fax: +44 (0) 161 275 4722
margaret.martin “at” manchester.ac.uk

Project website: www.ccsr.ac.uk/mnmg

3 Introduction

There is at present a lack of consensus amongst different researchers regarding the modelling of large scale social systems, because multilevel dependencies are treated disparately by different disciplines: in general multilevel modelling in the social sciences focuses on aspects of the multilevel structure other than the social network, organisational theory focuses on organisations, whilst most social network analysis focuses on the network but not other aspects of multilevel social structure. This team of internationally recognised experts in social network analysis, multilevel modelling and organisational theory will develop new methods to bridge this divide.

In order to fully understand the nature of multilevel social networks, a sophisticated methodology is needed that carefully integrates ideas from multilevel analysis with recent developments in statistical models for networks, such as Exponential Random Graph Models (ERGMs). Such a methodology will involve the development of: 1) an appropriate model framework that incorporates multilevel and network elements and 2) techniques for applying and interpreting this model framework. Such methods and approaches will have considerable substantive research value.

The project commenced on October 1st 2009 and runs until September 2012.

The research is funded by the Leverhulme Trust, under the *International Collaborative Networks* scheme.

4 What is Social Network Analysis (SNA)?

Social Network Research is broadly concerned with the way in which individuals are relationally tied to each other, and the consequences these relational patterns have for the individuals, and the social groups comprising these individuals. In social network research a *graph* is a commonly used analytical tool for describing a network. In such a graph, the nodes represent the individuals in the network, and ties between individuals that are connected to each other are represented by lines. These may be undirected *edges*, or directed *arcs*. Arcs include arrowheads to show the direction of the relationship. When the individuals are people, the ties may represent giving advice, receiving support, friendship, having sex, etc. Nodes may also represent different social units such as countries or boards of directors, with appropriately defined relations on these units. Since Moreno's foundational work in the 1930s, in what was then called *sociometry*, networks have proved to be of great use in explaining, for example: how innovations and opinions spread through social interaction; the consequences differential embeddedness (different position within the network) have on the power and well-being of individuals; how substance abuse and behaviour coevolve (Freeman, 2004; Borgatti et al., 2009). There has also been an explosion of interest in social networks in disciplines such as physics and biology.

5 What are Exponential Random Graph Models (ERGMs)?

Exponential (family) Random Graph Models (ERGMs, or p^* models) (Frank and Strauss, 1986) [www.sna.unimelb.edu.au/models/models.html] , are statistical models for the ties in a network that not only take exogenously defined characteristics of individuals into consideration (such as gender, organisation size, gross-domestic product), but also recognise the complicated interdependencies between tie variables, such as triangles, k-triangles, or k-stars. ERGMs are derived out of principled assumptions for the dependencies between tie variables and simple dependency assumptions give rise to a collection of configurations that are themselves interpretable in terms of theories of how ties self-organise. Much work has gone into model development, inference and the development of software for fitting ERGMs. For example, *PNet* [www.sna.unimelb.edu.au/pnet pnet.html], and also *statnet* [www.statnetproject.org], the latter implemented in *R*. The application of ERGMs to investigate social network structure is becoming increasingly popular in the social sciences.

References:

Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, Giuseppe Labianca, 2009, Network Analysis in the Social Sciences, Science, 323, 892-895.

Ove Frank and David Strauss, 1986, Markov Graphs, Journal of the American Statistical Association, 81, 832842.

Linton Freeman, 2004, The Development of Social Network Analysis: A Study in the Sociology of Science. Vancouver: Empirical Press.

6 What are Multilevel Models?

Multilevel Models began to be widely used in the 1980s. The models were first applied to study such issues as "school effectiveness" in two level structures of pupils in schools, or to study three level nested structures, such as pupils in classes in schools, where the response was interval - e.g a test score. Multilevel models have also been used to investigate variations in response variables for individuals in households, or individuals in areas of varying scales, and are useful for multivariate and longitudinal studies. Initially models for interval responses were developed, and more recently developments for generalised linear models have been made, including multilevel logistic models. Multilevel Models are now used for non-nested situations, including cross-classified models, multiple membership models, and multiple membership multiple classification models. Monte Carlo Markov Chain (MCMC) is often used for these more complex types of model. Multilevel models can sometimes be useful for social network analysis: especially for dyadic relationships, ego-networks, and to account for network influences on a response variable. Whether a multilevel model or another kind of statistical model is appropriate for a particular social network analysis depends on the targets of inference of the analysis, the available data, and the assumptions made about the network structure. A non-exhaustive list of software packages for fitting multilevel models includes MLwiN, lme and nlme (both implemented in R), and HLM.

References

Browne WJ, Goldstein H, Rasbash J (2001) Multiple Membership Multiple Classification Models. Statistical Modelling, 1, 103-124.

Fielding A and Goldstein H (2006) Cross-Classified and Multiple Membership Structures in Multilevel Models: an Introduction and review.

Pinheiro, Jose C. and Douglas M. Bates. (2000). "Mixed Effects Models in S and S-Plus" Springer-Verlag, New York.

Rasbash J, Steele F, Browne WJ, Goldstein H (2009) A user's guide to MLwiN

Snijders T. and Bosker R (1999) Multilevel Analysis. Sage.

See also: www.cmm.bristol.ac.uk

7 What are Multilevel Networks?

There are a variety of ways in which multilevel networks may be understood and defined. What multilevel networks are depends on whether terminology and methodology is borrowed from standard multilevel analysis, or is defined from a fully relational perspective. One way to classify multilevel networks, or multilevel approaches to investigate networks, is given below:

1. The scientific discussion about *peer effects* or *social effects* is about effects of social context on individual behaviour and performance. One tradition of studying this issue is via multilevel analysis - e.g., classroom effects or neighbourhood effects. Another tradition is in social network analysis, where the effect of the personal network on individual behavior and performance is studied. Examples of multilevel approaches to peer dependencies are given in, for example Manski (1993), on what he called *the reflection problem* (note here the distinction from models that do not explicitly introduce dependencies between outcomes, such as those of Brock and Durlauf, 2001). Peer groups that represent aggregate measures of actual network ties have also been used in a multilevel framework (Poteat et al., 2007). Various methods for modelling peer interdependence when the level of detail for peer-to-peer ties is at the level of pairs of respondents, have been proposed for several different types of response variables (Ebring & Young, 1979; Doreian, 1982; Marsden & Friedkin, 1994; Robins et al., 2001; Steglich et al., in press). There are also other theories on how the network ties relate to peer effects on outcomes (Mouw, 2006), as well as Network Autocorrelation Models (Leenders, 2002; Marsden and Friedkin 1993).

2. Multilevel techniques may also be used to model, or to take into account, peer dependencies. When modelling directed ties, the p2 model (van Duijn et al., 2004) has nodal random effects capturing the fact that ties are cross-classified with respect to sender and receiver nodes. Closely related is the mixed membership model of Airoldi et al. (2008) and the latent variable models of Hoff (2008) and Handcock et al. (2007). The idea that networks reflect latent categories (either emergent from the structure or exogeneous) has a relatively long tradition in social network analysis starting with block modelling (White et al., 1976), and consequently the statistically convenient latent variable models are mirrored by substantive theory (more straightforward statistical conceptualisations of block models and settings are found in Nowicki and Snijders, 2001, and Schweinberger and Snijders, 2003, respectively). For modelling nodal attributes, multilevel techniques may also be employed to capture the fact that each relational tie potentially induces dependencies for the two nodes of the tie. From first principles, the ties may then be considered memberships for the nodes in the ties, and a multiple membership approach may be used. The dyad is a level that should be taken into consideration, and the fact that the memberships are typically highly overlapping illustrates the uniquely complex dependence structure that networks produce. Conceptually, we may see individuals as nested in dyads, and dyads

as nested in triangles, etc, (Monge and Contractor, 2003). Pattison and Robins, 2002, make explicit the different levels of dependencies arising out of these partially overlapping contexts through defining local social neighbourhoods in an Exponential family Random Graph Modelling (ERGM) framework (Frank and Strauss, 1986). Multilevel models can also be used with ego-nets, where the unit of analysis is the tie between alter and ego. Given that ties between egos and alters exist, the strength or quality of these alter-ego ties can be modelled, and the multilevel approach recognises that particular alters have an ego in common. Snijders, Spreen, and Zwaagstra (1995) used such an approach to model reasons for cocaine use amongst networks of cocaine users in Rotterdam. De Miguel and Tranmer (2010) modelled the undirected ties between immigrants to Spain and their alters. They were interested in the probability of the ties being to Spaniards (the more settled population of Spain) as opposed to other recent immigrants, given the type of support role exchanged between alter and ego. Snijders and Kenny (1999) used a cross-classified multilevel model to fit the social relations model to family data, and Rasbash et al (2004) also applied complex mulitlevel models to family data.

3. A well-established strain of social network analysis is directed at studies of one complete network. Recently we have seen this extended to studies of populations of networks. When considered in this way, we can think of a population that contains a number of complete networks, but we can consider these networks replications of each other. For example, friendship networks of pupils in several classes in the same school; each class is a separate complete network, but the features of friendship structure or the effects of friendship networks may be replicated in the different classes in the school. This type of multilevel study was proposed in Snijders & Baerveldt (2003); data sets such as Add Health, those collected by the group of Laurence Moore (Cardiff University) and now being collected in the Dynet project of John Light (Oregon) are examples on the data side. Some other papers in this tradition are Baerveldt, van Duijn, Vermeij and van Hemert (2004) and Lubbers & Snijders (2007). Studying multiple replications of networks allows us to investigate how the network processes interact with contexts and settings. Studies of complete networks typically rely on the implicit assumption that the network is located in a closed relational system. Hence, if we were able to parse systematic interaction tendencies from those patterns that are unique to a particular context, this would make for a strong case for the roles and functions of networks. By the same token, however, this raises issues of how to define boundaries (Laumann et al., 1983), how to accommodate networks of different sizes, and how to allow for heterogeneity in the model. A miss-specified boundary may result from not taking boundary crossing ties (of same or different type as the type studied) into consideration or leaving out important nodes. In terms of dependencies of ties, we may, for example, have the case that a lot of ties may be left unexplained if many people in a room know each other through a person, and that person then leaves the room. Networks of different sizes are notoriously hard to compare, and the way in which network models scale is not fully understood. From the perspective of the multiple membership analogy, the

problem with homogeneity may be understood in terms of the fact that the memberships not are neatly nested and consequently it is hard to define homogeneous regions of a graph (cf e.g. the case of school classes in schools). Many conventional complete network designs are de facto multilevel in the sense that nodes are classified according to geography, affiliations, organisational levels, etc. A multiple replicates approach may benefit from these approaches, recognising the fact that just because there may exist ties between different level units, this may still be a multilevel structure.

4. As network research represents a relational perspective, it is natural to also view a multilevel structure in terms of relations as do Brass et al. (2004). We may, for example, have ties between organisations and ties between people in these organisations but we may also have ties between people and organisations. The organisations may be purely constitutive, in the sense that they are no more than collections of individuals in which case the organisation has ties to the people that make up the organisation or the organisations may be defined with relative independence. In the case where only ties between people and organisations are studied, the network simplifies to bipartite network analysis for which many methods have already been developed (see Wang et al., 2009, and the references therein). Methods for the case where people to organisation ties allow ties between people are currently being developed. Examples where all within- and between-level ties are analysed jointly are thus far relatively rare, with Lazega et al. (2008) being a notable exception. Further studies are underway in which this type of data are collected. From a modelling perspective, this data collection paradigm, while potentially being the most realistic, requires careful consideration of the different properties of different types of ties and how they are interrelated. As an example, the four-cycle that is created when two people in two different organisations get to know each other while at the same time their respective organisations form a ties, is different from a four cycle only consisting of people-to-people ties. The relational perspective on multilevel structures promises to offer rich descriptions and it is more general than networks in multilevel structures (i.e., the multiple replications).

There is plenty of scope for combining different aspects of the four characterisations above. In addition, there are particular issues associated with implementing these ideas for either network structure, response variables with respect to the network, or both. Furthermore, there are other dimensions for these issues, depending on whether cross-sectional data or longitudinal data are used.

References:

Airoldi, Blei, Fienberg, Xing (2008) Mixed Membership Stochastic Blockmodels, The Journal of Machine Learning Research, 9, 1981-2014.

Baerveldt, C., van Duijn, M.A.J., Vermeij, L., & D. A. van Hemert (2004). Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils networks. Social Networks 26, 55-74.

Brass, D.J., Galaskiewicz, J., Greve, H.R., and Tsui, W. (2004). Taking stock of networks and organizations: A multilevel perspective. Academy of Management Journal, 47, 795-819.

Brock, W. A. and Durlauf, S. N. (2001) Interactions-based Models, in Handbook of Econometrics, eds. J.J. Heckman & E.E. Leamer, ed 1, 5, ch 54, Amsterdam: North-Holland, pp. 32973380.

de Miguel Luken and Tranmer M (2010; in press) Personal Support Networks of Immigrants to Spain: a Multilevel Analysis. Social Networks.

Doreian, P. (1982). Maximum likelihood methods for linear models. Sociological Methods and Research, 10 , 243269.

Ebring, L., & Young, A. (1979). Individuals and social structure: Contextual effects as endogeneous feedback. Sociological Methods and Research, 7 , 396430.

Ove Frank and David Strauss, (1986). Markov Graphs. Journal of the American Statistical Association, 81, 832842.

Handcock, M.S.,Tantrum, J.M., Raftery, A.E. (2007) Model-Based Clustering for Social Networks, JRSS A, 170, 301-354.

Hoff, P. (2008), Multiplicative latent factor models for description and prediction of social networks, Computational & Mathematical Organization Theory, (doi: 10.1007/s10588-008-9040-4)

Lazega, E. Jourda, M., Mounier, L., Stofer, R. (2008) Catching up with big fish in the big pond? Multi-level network analysis through linked design. Social Networks Vol. 30, No. 2. pp. 159-176.

Laumann, E. O., Marsden, P. V., Prensky, D., 1983. The boundary specification problem in network analysis. In:Burt, R. S., Minor, M. J. (Eds.), Applied Network Analysis. Sage Publications, London, pp. 18-34.

Lubbers, M. and Snijders, T.A.B (2007) A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. Social

Networks Vol 29 489-507.

Manski, C. (1993) Identification of Endogenous Social Effects: The Reflection Problem, Review of Economic Studies, Vol. 60, No. 3, pp. 531-542.

Marsden, P., & Friedkin, N. (1993). Network studies of social influence. In W. S. & Galaskiewicz (Eds.), Advance in social network analysis (pp. 325). Thousand Oaks, CA : Sage.

Monge, P. R., & Contractor, N. S. (2003). Theories of communication networks. New York: Oxford University Press.

Mouw, T. (2006) Estimating the Causal Effect of Social Capital: A Review of Recent Research Annual Review of Sociology, Vol. 32: 79-102

Nowicki, K. & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 96, 10771087.

Pattison, P., & Robins, G. (2002). Neighborhood-based models for social networks, Sociological Methodology, 32, 301-337.

Poteat, V. P., Espelage, D. L., & Green Jr., H. D. (2007). The socialization of dominance: Peer group contextual effects on homophobic and dominance attitudes. Journal of Personality and Social Psychology, 92, 10401050.

Rasbash, J. O'Connor T & Jenkins J (2004). Multilevel models for family data. http://www.cmm.bristol.ac.uk/team/jon.shtml

Robins, G., Pattison, P., & Elliot, P. (2001). Network models for social influence processes. Psychometrika, 66 , 161-190.

Schweinberger & Snijders (2003) Settings in social networks: A measurement model, Soc Met, 33, 307-341.

Snijders T. A. B. and Baerveldt, C. (2003), A Multilevel Network Study of the Effects of Delinquent Behavior on Friendship Evolution, Journal of Mathematical Sociology, 27, 123-151.

Snijders T.A.B and Kenny D (1999) The Social Relations Model for Family Data: a Multilevel Approach. Personal Relationships, 6, pp 471-486.