Random Effects Models for Social Network and Group Dependencies

Mark Tranmer
Mitchell Centre for Social Network Analysis
Social Statistics
University of Manchester

mark.tranmer@manchester.ac.uk

e-stat group meeting presentation, 8th April, 2011.

Contents.

- The Problem
- Models for Dependencies in a Single Social Network
- Models for Social Network and Group Dependencies
- Discussion / Conclusion
- References

The Problem.

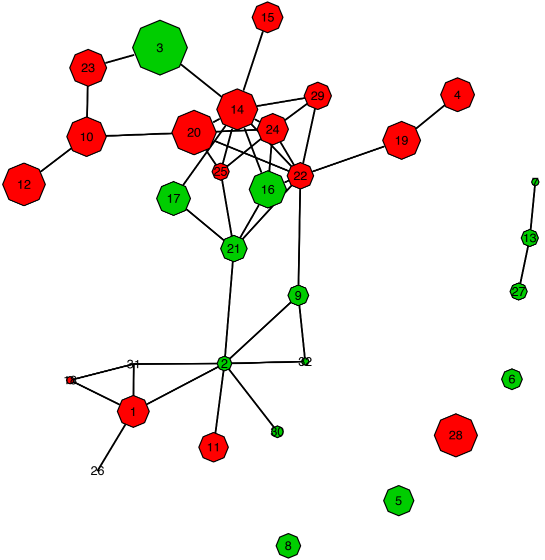
- We have a dependent variable y and we want to examine the nature and extent of variations in the values of y given the social networks and groups for a population of interest.
- The social network ties may be within groups only, or there may be ties between individuals in different groups.
- The number of groups may be large.
- We may want to relate y to a set of covariates, $\mathbf{x}$, at the individual, group, or network 'level'.
- How best to formulate a model for such a situation?
- Substantive implications of such a model?

Example 1: Dependencies for a Single Social Network. Freeman's EIES data.

- Let's start by considering just a network: no groups.
- 'friendship or met' ties of 32 academics
  (symmetric 0/1 matrix)
- y is log(#citiations+1)
- x is a 0/1 variable:

  1 if sociology discipline; 0 otherwise.

Freeman's EIES network, time 1.
red=sociology; green=other; label=ego number.



How to Model the Social Network Dependencies in this Dataset? Network Autocorrelation Model.

- Network Autocorrelation Model: *network effects* model or *network disturbances* model
- These models originate in spatial analysis (Cliff and Ord, 1975; Ord, 1975; Doreian, 1980).
- Later used in social network analysis.
- Leenders (2002) reviews network autocorrelation models, and discusses the effects of different specifications of the weight matrix $\mathbf{W}$ (see next two slides for model formulation).
- We can fit these models in `R` using using `lnam` in the `sna` package for social networks (Butts, 2008), and the `spdep` package for spatial analysis (Bivand, 2010).

Network Effects Model: Formulation.

The *network effects* model, also known as the *spatial effects* model (Doreian, 1980) is defined for a single network as:

$$\begin{gathered}\mathbf{Y_i} = \rho \mathbf{W} \mathbf{Y_i} + \mathbf{X_i} \beta + \epsilon_\mathbf{i} \\ E[\epsilon_\mathbf{i}] = 0, \; E[\epsilon_\mathbf{i} {\epsilon_\mathbf{i}}'] = \sigma_\epsilon^2 \mathbf{I}\end{gathered} \tag{1}$$

Network Disturbances Model: Formulation.

The *network disturbances* model, also known as the *spatial disturbances* model (Doreian, 1980) is defined for single network as:

$$\begin{aligned}
\mathbf{Y_i} &= \mathbf{X_i}\beta + \epsilon_\mathbf{i} \\
\epsilon_\mathbf{i} &= \rho\mathbf{W}\epsilon_\mathbf{i} + \nu_\mathbf{i} \\
E[\nu_\mathbf{i}] = 0,\ E[\nu_\mathbf{i}{\nu_\mathbf{i}}'] &= \sigma_\nu^2\mathbf{I}
\end{aligned} \tag{2}$$

Network Effects Model (1): Results.

Parameter	Estimate	S.E.
cons (β_0)	1.871	0.330
sociology (β_1)	0.697	0.434
rho (feedback on y)	0.034	0.030
sigma (error s.d.)	1.163	0.021

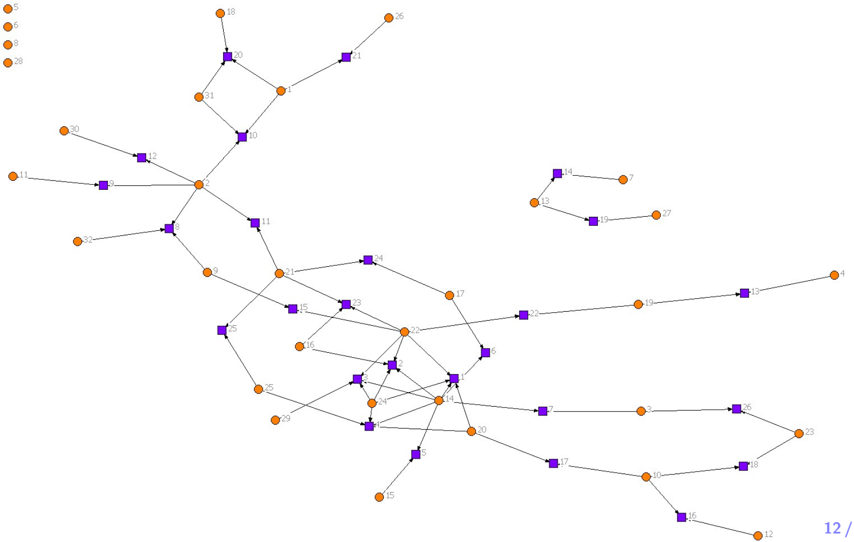
Network Distrubances Model (2): Results.

Parameter	Estimate	S.E.
cons (β_0)	2.135	0.340
sociology (β_1)	0.558	0.473
rho (feedback on ϵ)	0.101	0.059
sigma (error s.d.)	1.130	0.021

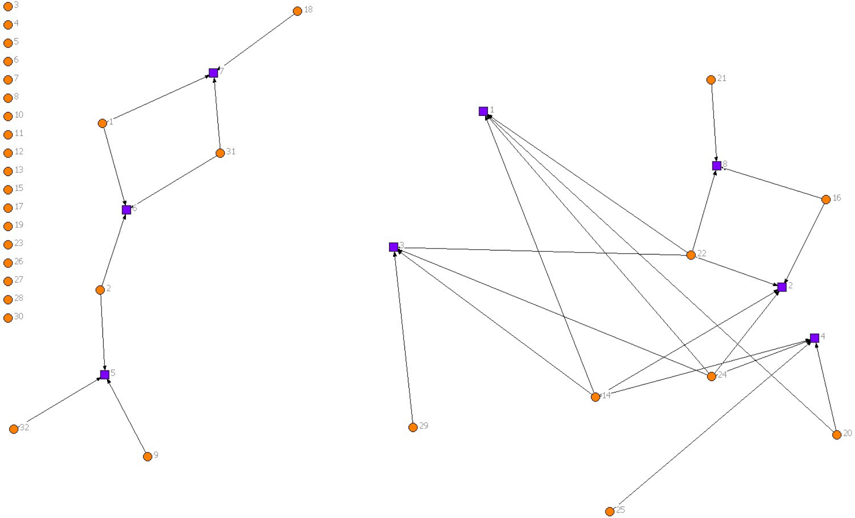
How to Model the Social Network Dependencies in this Dataset? Multiple Membership (MM) Model.

- Alternative approach: fit a ‘Multiple Membership’ model (Hill and Goldstein, 1998; Browne, 2009).
- Use this approach with ego as the group and alters as members. See formulation for Model (3) below.
- i.e. extract the 32 ego nets from the network and use these in analysis.
- Should ego be included in their own group, or not?
- Another MM alternative to ego-nets is to extract the clique sets, e.g. using UCINET, and use these as the groups. A lot of isolates in this example, but do not present problems in model.
- Understanding the extent of ego or clique level variation in y before and after controlling for covariates may be substantively useful.

EIES: Clique-Sets of Minimum Size 2.



EIES: Clique-Sets of Minimum Size 3.



Multiple Membership Model: Specification.

$$
\begin{aligned}
y_i = (X'\beta)_i + \sum_{j \in \text{network group}(i)} w_{i,j}^{(2)} u_j^{(2)} \quad &+ \quad e_i \\
u_j^{(2)} \quad &\sim \quad N(0, \sigma^2_{u(2)}) \\
e_i \quad &\sim \quad N(0, \sigma^2_e) \\
\text{Cov}(e_i, u_j) \quad &= \quad 0 \\
\text{network group}(i) \subset j \quad ; \quad j = 1, ..., J; \quad i \quad &= \quad 1, ..., n
\end{aligned} \qquad (3)
$$

Example: EIES Single Network.

- The maximum number of alters for any ego is 9.
- We can fit this model in `MLwiN`; see Rasbash et al. (2009) and (Browne, 2009) for software details.
- Use MCMC estimation, long chain.
- Wrote a script in `R` to organise W matrix from network connection matrix, then read this in to MLwiN.
- Used UCINET to generate **W** matrices in the clique-set approach
- Paper in progress by Tranmer, Browne and Goldstein.
- Multiple Membership models for networks being implemented in the next version of `MLwiN`, called "`e-stat / STAT-JR`"

Network Disturbances and Multiple membership Model Results: EIES.

	Network Disturbances	MM ego (γ)
$\widehat{\beta}_0$	2.135 (.339)	2.344 (.309)
$\widehat{\beta}_1$	.558 (.473)	.760 (.374)
$\widehat{\rho}$	.1005	
$\widehat{\sigma}_\nu$	1.13	
AIC	107.60	
$\widehat{\sigma}^2_{u_0}$		2.653 (1.78)
$\widehat{\sigma}^2_{e_0}$		.619 (.298)
DIC		84.81
pD		12.50

Multiple Membership Models for Clique-Sets: EIES.

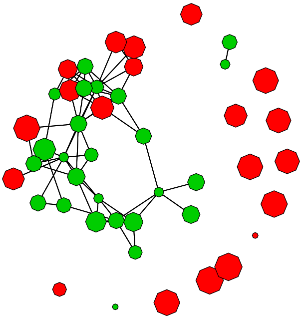
	clique-2 (γ)	clique-3 (γ)
$\widehat{\beta}_0$	2.226 (.332)	2.085 (.338)
$\widehat{\beta}_1$	.834 (.446)	.835 (.457)
$\widehat{\sigma}^2_{u_0}$	1.342 (1.051)	.807 (1.512)
$\widehat{\sigma}^2_{e_0}$	.928 (.454)	1.484 (.444)
DIC	95.47	107.41
pD	10.56	5.27

Example 2: Social Network and Group Dependencies. Delinquency in Dutch Schools.

- With kind permission from Chris Baerveldt (University of Utrecht), I have access to some data for pupils in 19 Dutch Schools.
- See Baerveldt and Rossem (2004); Snijders and Baerveldt (2003) for further details.
- Groups (schools), and social (friendship) networks within each group
- y is logged delinquency score
- x is gender ($0 = $ female ; $1 =$male)
- groups are different sizes

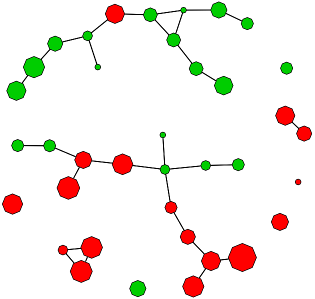
Examples: Dutch Delinquency Data.

school 1, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

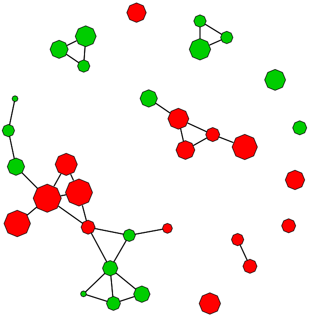
school 3, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

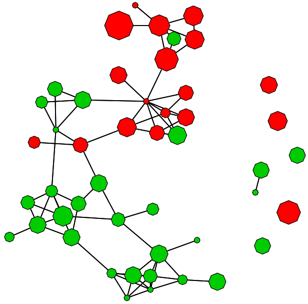
Examples: Dutch Delinquency Data.

school 4, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

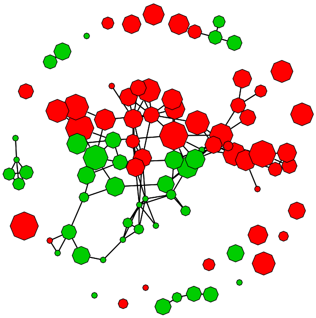
school 9, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

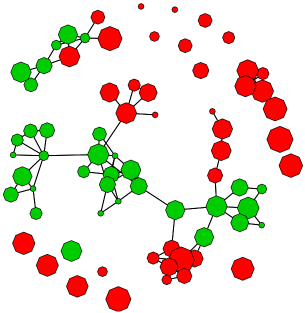
Examples: Dutch Delinquency Data.

school 14, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

school 23, time 1



vertex size=rnd(log(delinq+1)+1; green = girl, red=boy

Multiple Membership Model for Multiple Networks: Specification.

$$y_i = (X'\beta)_i + v_k + \sum_{j \in \text{network group}(i)} w_{i,j}^{(2)} u_j^{(2)} + e_i$$

$$v_k \sim N(0, \sigma_v^2) \quad ; \quad u_j^{(2)} \sim N(0, \sigma_{u(2)}^2)$$

$$e_i \sim N(0, \sigma_e^2) \quad ; \quad \text{Cov}(e_i, u_j) = 0$$

$$\text{network group}(i) \subset j, k$$

$$j = 1, ..., J \quad ; \quad k = 1, ..., K \quad ; \quad i = 1, ..., n \tag{4}$$

Dutch Delinquency Data: Multilevel Models [ego]. (Gamma Priors, MLwiN, MCMC).

	ignore net	**ignore sch**	**ego+sch**	**ego+sch rnd: male**
cons	1.472 (.053)	1.462 (.050)	1.457 (.057)	1.457 (.061)
male	.592 (.060)	.594 (.062)	.596 (.063)	.573 (.073)
school:				
cons/cons	.016 (.012)		.015 (.012)	.023 (.017)
ego:				
cons/cons		.054 (.043)	.048 (.041)	.047 (.028)
cons/male				.018 (.031)
male/male				.111 (.085)
indiv:				
cons	.910 (.041)	.905 (.043)	.893 (.042)	.874 (.041)
DIC	2726	2729	2723	2717

Clique-Set Analysis (Gamma Priors, MLwiN, MCMC).

	cliques-2 (γ)		cliques-3 (γ)	
	est	s.e.	est	s.e.
Fixed Part:				
$\hat{\beta}_0$ (cons)	1.468	.054	1.469	.053
$\hat{\beta}_1$ (male)	.595	.061	.594	.060
Random Part:				
Level: school				
$\hat{\sigma}^2_{v00}$	.017	.012	.017	.012
Level: clique				
$\hat{\sigma}^2_{u00}$	.056	.050	.091	.084
Level: individual				
$\hat{\sigma}^2_{e00}$	.848	.043	.860	.040
DIC	2681		2682	

Multilevel Network Disturbances Model: (could also extend Network Effects Model in this way)

$$
\begin{aligned}
\mathbf{Y_{ij}} &= \mathbf{X_{ij}}\beta + \mathbf{u_j} + \epsilon_{\mathbf{ij}} \\
\epsilon_{\mathbf{ij}} &= \rho\mathbf{W}\epsilon_{\mathbf{ij}} + \nu_{\mathbf{ij}} \\
E[\nu_{\mathbf{ij}}] = 0,\ E[\nu_{\mathbf{ij}}{\nu_{\mathbf{ij}}}'] &= \sigma^2\mathbf{I} \\
u_j &\sim N(0, \sigma_u^2)
\end{aligned} \tag{5}
$$

Results for Multilevel Network Disturbances Model. MCMC, Based on 50,000 Iterations.

estimate	$\hat{\beta}_0$ (cons)	$\hat{\beta}_1$ (male)	$\hat{\rho}$	$\hat{\sigma}^2_{U0}$ (sch.)	$\hat{\sigma}^2_{e0}$ (ind.)
median	1.795		.369	.380	.957
median	1.467	.595	.272	.385	.923

Table: Thanks to Johan Koskinen (Manchester) for the fast R code, to Pete Neal (Manchester) for the initial R and fortran code, and to Malcolm Fairbrother (Bristol) for additional code tweaks.

Extensions to the Multilevel Network Disturbances Model: I

- Consider a two level situation, e.g. individuals in areas.
- Here, the areas could be networked - e.g. contiguity, migration.
- And the individuals could be networked - e.g. friendship, support.
- Moreover, there could be ties between individuals in the same areas as well as ties between individuals in differerent areas.

Extensions to the Multilevel Network Disturbances Model: II

- We are developing a general model formulation for this
- Also allows for different ρ in different groups.
- Pupils in schools example only has networks at level 1, and all observed networks are within group (i.e. within school).
- Hence for this example, we could exploit the block diagonal structure of $\mathbf{W}$ to speed up the `R` code.
- Paper in progress by Tranmer, Koskinen, Neal and Fairbrother.

Discussion and Conclusion: I

- Which approach is best?
- Depends on substantive standpoint and targets of inference.
- Friendship may co-evolve with behaviour and a stochastic actor based model (see, for example Snijders et al. (2010)) might then be appropriate to test social theories if have longitudinal data.

Discussion and Conclusion: II

- However, there are some useful substantive and descriptive implications for the random effects models presented here (and also those still being developed!).
- For example: how strong is the feedback on y or ϵ when network and group dependencies taken into account? Does this feedback change in strength when covariates are added to the model?
- How much variation in y is there at the individual, network and group levels?
- What happens if a level is ignored?
- Is some of the variation in y at each level explained by covariates?

Further ideas

1. If had data for single network over three or more time points, we could extend the ideas for the multiple membership model to include a *time* level.
2. For particular social network relation, could include several $\mathbf{W}$ matrices e.g. best friend in class, etc, etc, in the multiple membership model to assess the relative importance of a particular kind of network relation on an outcome of interest y
3. Could include several definitions of social network structure in the multiple membership model e.g. cliques-2, cliques-3, ego, Girvan-Newman, to assess the relative importance of a particular network structure on an outcome of interest y

How Much Variance is Explained at Each Level, When the 'Male' Covariate is Added to the Model?

	null	**+ male**	**%**
2 cliques			
variation:			
school	0.016	0.017	106
clique	0.178	0.056	31
indiv.	0.886	0.848	96
3 cliques			
variation:			
school	0.016	0.017	106
clique	0.237	0.091	38
indiv.	0.930	0.860	92

Discussion and Conclusion: III

- Ability to combine network and other data at the group level within the model framework.
- Flexibility within Multilevel Model framework to make covariates random at the group or network level.
- Ability to control for individual, group or network levels or to aggregate variables to these levels.
- Some of these methods only need ego-nets or clique-sets, rather than the full network, though it is generally better to have full network information where possible.

www.ccsr.ac.uk/mitchell

- Thank you for listening!
- This work partially funded by ESRC / ARC
- Thanks to Peter Neal, Johan Koskinen, Malcolm Fairbrother, Bill Browne, Harvey Goldstein, and Chris Baerveldt.

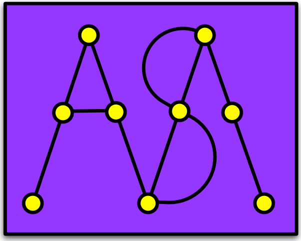


Figure: www.ccsr.ac.uk/mitchell

References : I

Baerveldt, C., Rossem, R. V., 2004. Students' delinquency and correlates with strong and weaker ties: A study of students' networks in Dutch high schools. connections 26 (1), 11–28.

Bivand, R., 2010. Package spdep in R.

Browne, W. J., 2009. MCMC estimation in MLwiN.

Butts, C. T., 2008. Social Network Analysis with sna. Journal Of Statistical Software 24 (6).

Cliff, A. D., Ord, J. K., 1975. Model Building and the Analysis of Spatial Pattern in Human Geography. Journal of the Royal Statistical Society, Series B 37 (3), 297–348.

Doreian, P., Aug. 1980. Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects? Sociological Methods & Research 9 (1), 29–60.

References : II

Hill, P. W., Goldstein, H., Jan. 1998. Multilevel Modeling of Educational Data With Cross-Classification and Missing Identification for Units. Journal of Educational and Behavioral Statistics 23 (2), 117–128.

Leenders, R., 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. Social Networks 24, 21–47.

Ord, K., 1975. Estimation Methods for Models of Spatial Interaction. Journal of the American Statistical Association 70 (349), 120– 126.

Rasbash, J., Steele, F., Browne, W. J., Goldstein, H., 2009. A Users Guide to MLwiN.

References : III

Snijders, T. A. B., 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., van de Bunt, G. G., Steglich, C. E. G., 2010. Introduction to stochastic actor-based models for network dynamics. Social Networks Volume 32 (Issue 1, January 2010), Pages 44–60.