Social Networks in Multilevel Structures

Mark Tranmer

Mitchell Centre for Social Network Analysis & CCSR
Social Statistics
School of Social Sciences
University of Manchester

mark.tranmer@manchester.ac.uk

University of Bristol
N-CESS e-stat node

First group meeting
February 26th, 2010

Contents

Population *dependencies*

- The multilevel model takes into account population dependencies with realistically complex models
- Dependencies may be of interest, or may be a nuisance
- Targets of inference of the analysis
- Are these group specific or marginal to the groups?

What about the *Social Network*?

- Alongside areas, households, organisations, time, another important population dependency is the social network.
- In a lot of social network analysis, the target of inference is the social network *structure*
- Statistical models for this: Exponential Random Graph Models (ERGMs) (Robins et al., 2007) - promising for cross sectional cases
- Multilevel models also useful for looking at network structure - particularly in terms of dyads and ego-nets (Snijders et al., 1995) ; (de Miguel Luken and Tranmer, 2010).
- Also family networks / social relations model (CCMs) (Snijders and Kenny, 1999) ; (Rasbash et al., 2004)
- Interest could also be change in network structure over time
- Stochastic Actor Based Models (Snijders and Steglich, 2010)

What about the *Social Network*?

- Say I am interested in the *association* of the social network with a dependent variable, or with a relationship such as a regression equation?
- Easy to think of substantive examples when it is important:
- School pupils that befriend each other or work together may have similar values in terms of educational performance measures
- Homophily might play a part here, but might not explain everything
- Peer effects models have been previously been used for this
- If we ignore a level in a multilevel analysis, the variation at that level doesn't go away! (Tranmer and Steel, 2001)
- So how should we reasonably and routinely include social network information information in a statistical analysis?
- Model based approaches - potential of multilevel models?

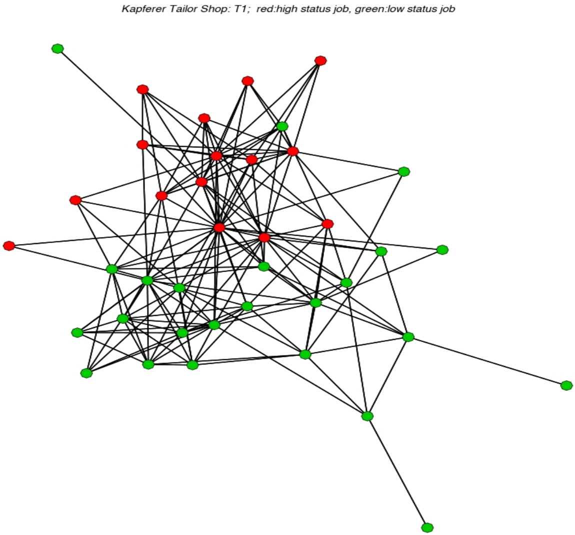
What I aim to do for this part of the *e*-stat node

- Timely to demonstrate potential of existing social science datasets and existing multilevel modelling software to investigate social networks, and their effects within other multilevel structures
- To show how social networks can be fitted as realistically complex multiple membership or cross-classified models - including visualisation.
- Exploit potential of existing datasets - DAMES / ESDS
- Even if we don't have all network info can we do something to account for networks in soc-stat analysis?
- Complements Bilateral australia grant "The role of households, geographical groups and social networks in social statistics" (ESRC/ARC funded; Tranmer, Elliot, Steel, Chambers, Clark, Suesse ; Aug 2008 - July 2011).
- How can we make MLwiN / e-stat more SNA friendly?

Real data / Simulation Study: Kapferer's tailor shop data

- Bruce Kapferer [from the 'Manchester School'] (1972) observed interactions in a tailor shop in Zambia over a period of 10 months. N=39 workers.
- Focus was the changing patterns of alliance among workers during extended negotiations for higher wages.
- He measured different types of interaction, recorded at two time points (seven months apart), over a period of one month: "instrumental" (work- and assistance-related) and "sociational" (friendship, socioemotional).
- Data are particularly interesting since an abortive strike occurred after the first set of observations, and a successful strike took place after the second.

Kapferer's tailor shop data: time 1; Socialisation ties: red = high status workers ; green = low status workers



Network Statistics: from UCINET

- Density overall: 0.2132
- Ties: 158
- Mean degree: 4.05
- Suppose there is a response variable for each network member, as well as the explanatory variable i.e. working in high status jobs: (0/1).

Kapferer's tailor shop data

- How can we meaningfully summarise the network structure here, with a view to accounting for it in a model?
- You might say "Don't summarise it, use all of it!"
- Could be good idea, but (i think) if we look at each dyad separately, we don't really get an idea of the nature and extent of clustering in the network. Though that might not be true with CCMs and MMMs
- Break data into ego-nets for each of the 39 workers?
- Find cohesive subgroups? Dense clusters in the network.
- Cohesive subgroups: cliques, cores, clans, *k*-plexes, Markov Clustering (UCINET), Latent clustering (R) etc ...
- Control for network information in a single-level model; e.g. ego-network size

3-cliques - there are 58 groups - but n=39. What about 5-cliques?

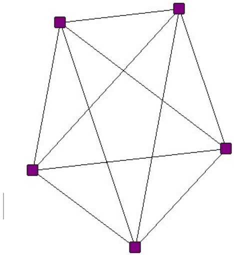


Figure: A 5-clique: maximally connected, each node is degree 4

Kapferer's tailor shop data: 5 cliques

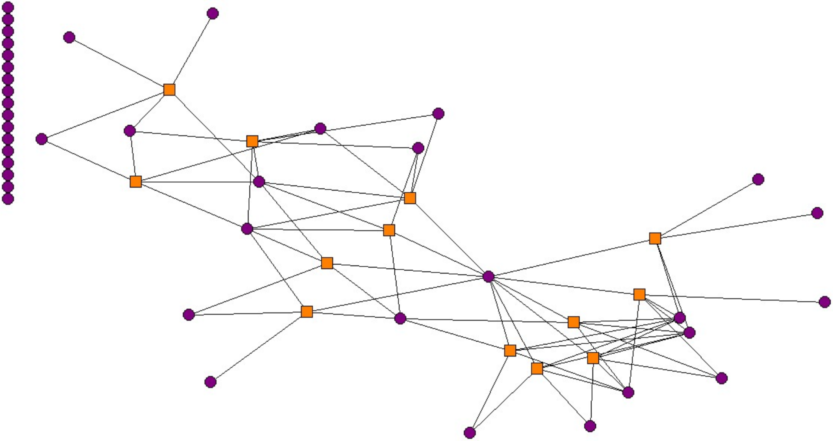


Figure: Kapferer's tailor shop data: 3 cliques; M=13, N=39

Multiple membership structure

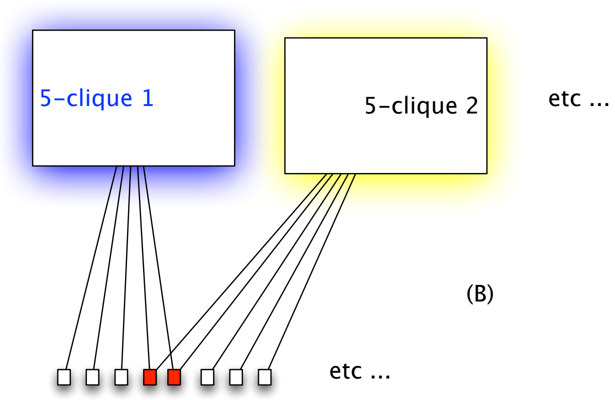


Figure: Multiple membership structure

Generating a simulated response variable, given network structure

- I used a *network effects* model:
  (Leenders, 2002) ; Marsden and Friedkin (1993)
- $y = \rho \mathbf{W}_1 y + \beta' \mathbf{x} + \epsilon$
- $\mathbf{W}_1$ represents the network structure: adjacency matrix in my case
- ρ decides the influence of connected individuals (0=Straightforward (OLS) regression)
- A special case of $\mathbf{W}_1$ leads to peer effects model
- $\mathbf{W}_2$ could be 'included in' ϵ term - Spatially Auto-Regressive (SAR) Model.
- $\mathbf{W}_1$ and $\mathbf{W}_2$ can both be in model, and $\mathbf{W}_1 = \mathbf{W}_2$ is possible
- Further extensions for multiple networks and to allow exogenous variables to be influenced by $\mathbf{W}$

Naïve single level model at ego level

A naïve modelling approach, that ignores the population network dependencies, is to fit a single level regression model, as follows:

$$y_j = \beta_0 + \beta_1 x_{1j} + \epsilon_j$$

$$\epsilon_j \sim N(0, \sigma^2) \tag{1}$$

Model does not include information about network structure. $x_1 i$ denotes high job status for worker i (0=low ; 1=high)

Multiple membership model specification: egos are members of cohesive subgroups

For each individual, i, w_{ij_2} is their weight for membership of each of the network 5-cliques, where $\sum_{j=1}^{J} w_{ij}^{(2)} = 1$, and if the regression parameters β_1 and β_2 are fixed, the model is specified as:

$$y_{i(j)} = \beta_0 + \beta_1 x_{1i(j)} + \sum_{j \in \text{5-clique}(i)} w_{ij}^2 u_j^{(2)} + \epsilon_i$$

$$i = 1, ..., N \quad \text{5-clique}(i) \subset (1, ..., J)$$

$$u_j^{(2)} \sim N(0, \sigma^2_{u(j)}) \quad \epsilon_i \sim N(0, \sigma^2_\epsilon) \tag{2}$$

Model includes information about network structure as random effects. Could extend to a random coefficient for 'highstatus', x_1

Ego-net model; 2 level - alters nested in egos

Break network into ego nets. ego j is level 2 and alters, indexed by i level 1. Assume (wrongly) alters of one ego independent of next. Also could be double counting: eg is worker 36 in the ego net of worker 39 or is 39 in the ego net of 36, or both?

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + U_{0j} + e_{0ij}$$
$$i = 1, ..., n_j \quad j = 1, ..., 39$$

$$u_{0j}^{(2)} \sim N(0, \sigma^2_{u(j)}) \quad \epsilon_{0ij} \sim N(0, \sigma^2_{\epsilon}) \tag{3}$$

Model includes information about network structure as a random effect. could also have a random coefficient for x_1

Single level model at ego level; control for ego net size

For each individual, i, w_{ij_2} is their weight for membership of each of the network 5-cliques, where $\sum_{j=1}^{J} w_{ij}^{(2)} = 1$, and if the regression parameters β_1 and β_2 are fixed, the model is specified as:

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 n_j + \epsilon_j$$

$$\epsilon_i \sim N(0, \sigma^2) \tag{4}$$

Model does includes aggregate information about network structure (n_j) as a fixed effect- could be extended to interact with x_1.

Initial simulation results: starting values $\beta_0 = 1 \;\; \beta_1 = 2 \;\; \sigma^2_\epsilon = .1 \;\; \rho = .01$

param.	neff	ols	MM	ego	n_j
$\hat{\beta}_0$	1.024	1.025	1.022	1.020	1.042
s.e. $(\hat{\beta}_0)$	.026	.018	.018	.007	.032
$\hat{\beta}_1$	2.404	1.988	1.990	1.986	1.993
s.e. $(\hat{\beta}_0)$	.002	.030	.029	.011	.031

Initial simulation results: starting values $\beta_0 = 1 \;\; \beta_1 = 2 \;\; \sigma^2_\epsilon = .1 \;\; \rho = .1$

param.	neff	ols	MM	ego	n_j
$\hat{\beta}_0$	1.025	1.143	1.102	1.168	1.013
s.e. $(\hat{\beta}_0)$	.026	.026	.023	.010	.038
$\hat{\beta}_1$	1.988	2.088	2.067	2.125	2.053
s.e. $(\hat{\beta}_0)$	.033	.043	.036	.015	.036

Initial simulation results: starting values $\beta_0 = 1 \;\; \beta_1 = 2 \;\; \sigma^2_\epsilon = .1 \;\; \rho = .2$

param.	neff	ols	MM	ego	n_j
$\hat{\beta}_0$	1.025	1.311	1.211	1.378	.980
s.e. $(\hat{\beta}_0)$	.026	.048	.035	.018	.048
$\hat{\beta}_1$	1.997	2.222	2.128	2.310	2.133
s.e. $(\hat{\beta}_0)$	.033	.077	.060	.026	.047

Initial simulation results: starting values $\beta_0 = 1 \;\; \beta_1 = 2 \;\; \sigma^2_\epsilon = .1 \;\; \rho = .5$

param.	neff	ols	MM	ego	n_j
$\hat{\beta}_0$	1.028	2.314	1.854	2.641	.858
s.e. $(\hat{\beta}_0)$	.028	.184	.134	.067	.111
$\hat{\beta}_1$	1.988	2.892	2.609	3.235	2.500
s.e. $(\hat{\beta}_0)$	.033	.296	.218	.099	.107

Initial simulation results: starting values $\beta_0 = 1 \;\; \beta_1 = 2 \;\; \sigma^2_\epsilon = .1 \;\; \rho = .8$

param.	neff	ols	MM	ego	n_j
$\hat{\beta}_0$	1.032	8.030	5.304	9.815	.579
s.e. $(\hat{\beta}_0)$	.029	.907	.674	.332	.384
$\hat{\beta}_1$	1.988	5.456	5.124	6.898	3.455
s.e. $(\hat{\beta}_0)$	.033	1.462	1.024	.486	.371

Discussion

- Some people don't belong to any cliques.
- Why cliques? Why 5-cliques?
- Results will vary according to size of clique *c.f.* scale effects in geographical analysis
- Same methodology could be applied to other ways of grouping the network
- cross-classified models for ego-nets would be better i think.
- why wouldn't we use network effects model?
- extend it and make elements of it multilevel?
- more simulations; more networks; random graphs with same density.

Appendix 1: Kapferer's tailor shop data: 3-cliques

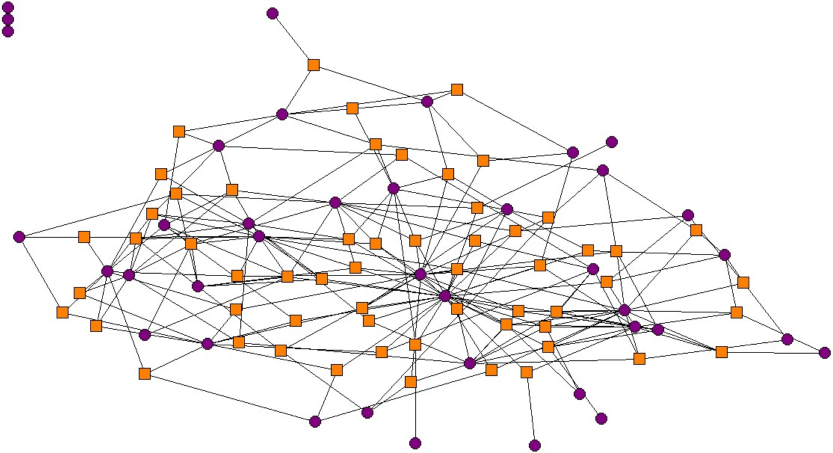


Figure: Kapferer's tailor shop data: 3-cliques; M=58, N=39

Appendix 2: Weights matrix w_{ij}: first 20 rows only; '.' indicates a zero

i	$j1$	$j2$	$j3$	$j4$	$j5$	$j6$	$j7$	$j8$	$j9$	$j10$	$j11$	$j12$	$j13$	n_gp
1	.	.	.	.	.	.	.	.	.	.	.	.	.	0
2	.	.	.	.	.	.	.	.	**1**	.	.	.	.	1
3	**.17**	**.17**	.	.	.	**.17**	**.17**	**.17**	**.17**	.	.	.	.	6
4	**.33**	.	.	.	.	.	.	**.33**	**.33**	.	.	.	.	3
5	.	.	.	.	.	.	.	.	.	.	.	.	.	0
6	.	**.50**	.	.	.	.	**.50**	.	.	.	.	.	.	2
7	.	.	.	.	.	**1**	.	.	.	.	.	.	.	1
8	.	.	.	.	.	.	.	.	.	.	.	.	.	0
9	.	.	.	.	.	**1**	.	.	.	.	.	.	.	1
10	.	.	.	.	.	.	.	.	.	.	.	.	.	0
11	**.17**	**.17**	.	.	.	**.17**	**.17**	**.17**	**.17**	.	.	.	.	6
12	**.20**	**.20**	.	.	.	.	**.20**	**.20**	**.20**	.	.	.	.	5
13	.	.	.	.	.	.	**.50**	**.50**	.	.	.	.	.	2
14	.	.	.	.	.	.	.	.	.	.	.	.	.	0
15	.	.	.	.	.	.	.	.	.	.	.	.	.	0
16	**.10**	**.10**	**.10**	**.10**	**.10**	**.10**	**.10**	**.10**	**.10**	**.10**	.	.	.	10
17	.	.	.	.	.	.	.	.	.	.	.	.	.	0
18	.	.	.	.	.	.	.	.	.	.	.	.	.	0
19	**.20**	**.20**	**.20**	**.20**	**.20**	.	.	.	.	.	.	.	.	5
20	.	.	.	.	.	.	.	.	.	.	.	.	.	0
etc.	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Appendix 3: Network Degrees of all workers

row	name	degree	row	name	degree
1	CHISOKONE	24	21	LWANGA	8
2	MUKUBWA	17	22	BEN	7
3	LYASHI	15	23	PAULOS	7
4	HENRY	14	24	NKOLOYA	6
5	ZULU	14	25	DONALD	6
6	MUBANGA	14	26	ANGEL	6
7	ABRAHAM	13	27	NYIRENDA	5
8	IBRAHIM	11	28	CHIPATA	5
9	WILLIAM	10	29	MABANGE	5
10	CHOBE	10	30	KALUNDWE	5
11	HASTINGS	10	31	NKUMBULA	5
12	KALONGA	10	32	KAMWEFU	4
13	JOSEPH	10	33	MESHAK	4
14	JOHN	9	34	MATEO	3
15	CHILUFYA	9	35	ADRIAN	2
16	SEAMS	9	36	ENOCH	2
17	CHILWA	9	37	ZAKEYO	1
18	MPUNDU	9	38	CHIPALO	1
19	KALAMBA	8	39	SIGN	1
20	CHRISTIAN	8			

References I

de Miguel Luken, V., Tranmer, M., 2010. Personal support networks of immigrants to spain: a multilevel analysis. Submitted to Social Networks, Feb 2010.

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

Marsden, P. V., Friedkin, N. E., 1993. Network Studies of Social Influence. Sociological Methods and Research 22 (1), 127–151.

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

References II

Robins, G. L., Pattison, P. E., Kalish, Y., Lusher, D., 2007. An introduction to exponential random graph (p *) models for social networks. Social Networks 29, 173–191.

Snijders, T., Spreen, M., Zwaagstra, R., 1995. The use of multilevel modelling for analysing personal networks: networks of cocaine users in an urban area. Journal of Quantitative Anthropology 5, 85–105.

Snijders, T. A. B., Kenny, D. A., 1999. The social relations model for family data: a multilevel approach. Personal Relationships 6, 471–486.

Snijders, T. A. B., Steglich, C. E. G., 2010. Introduction to stochastic actor-based models for network dynamics. Social Networks.

References III

Tranmer, M., Steel, D. G., 2001. Ignoring a level in a multilevel model: evidence from UK census data. Environment and Planning 33, 941–948.