

# Exponential random graph models: new specifications

Snijders, Pattison, Robins & Handcock (2005)

©Department of Psychology, University of Melbourne

## Outline

### Modelling social networks

What do we know?

Possible “topologies”

Models for interactive systems and the Hammersley-Clifford theorem

### Exponential random graph models

Neighbourhoods

Homogeneous models

Related model parameters

$k$ -triangles and  $k$ -independent-2-paths

### Fitting models

Example: collaboration network

A few remarks on what we have learnt

©Department of Psychology, University of Melbourne

## What do we think we know about processes underlying network tie formation?

### *Exogenous effects:*

Homophily, shared affiliations, spatial propinquity all matter (e.g., McPherson, Smith-Lovin & Cook, 2001)

More generally: network tie formation is often set within *foci* (Feld, 1981) or *settings* (Pattison & Robins, 2002)

### *Endogenous network effects:*

*Clustering*: tie formation is often more likely when actors have network partners in common

More generally: “a social tie ... is subject to, and known to be subject to, the hegemonic pressures of others engaged in the social construction of that network.” (White, 1998)

### *Interactions between exogenous and endogenous effects:*

*General social selection* (Robins, Elliott & Pattison, 2001)

©Department of Psychology, University of Melbourne

## Modelling endogenous network processes

### Guiding principles:

*Network ties are the outcome of (unobserved) social processes that tend to be local and interactive*

*There are both regularities and irregularities in these local interactive processes*

### We hence construct statistical models in which:

local interactions are expected and assumptions about form of “local interactions” are explicit

regularities are represented by model parameters and estimated from data

consequences of local regularities for global network properties can be understood *and can also provide an exacting approach to model evaluation*

©Department of Psychology, University of Melbourne

## Local interactivity

We model *tie variables*:  $\mathbf{X} = [X_{ij}]$   $X_{ij} = 1$  if  $i$  has a tie to  $j$   
0 otherwise

*realisation* of  $\mathbf{X}$  is denoted by  $\mathbf{x} = [x_{ij}]$

We also incorporate node-level exogenous *attribute variables*:  $\mathbf{Y} = [Y_i]$

### Two modelling steps:

*Local interactions*: define two network tie variables to be *neighbours* if they are conditionally dependent, given the values of all other tie variables

*But*: what are appropriate *neighbour* assumptions?

©Department of Psychology, University of Melbourne

## Network topologies: which tie variables are neighbours?

Two tie variables are *neighbours* if:

they share an actor



*Markov model*

(Frank & Strauss, 1986)

they share connections  
with two existing ties  
(completing a social circuit)



*realisation-dependent model*

(Pattison & Robins, 2002;  
Snijders, Pattison, Robins &  
Handcock, 2004)

*There are other possibilities, but these two get us a long way*

©Department of Psychology, University of Melbourne

## Models for interactive systems of variables (Besag, 1974)

A *neighbourhood* is a set of mutually neighbouring variables and corresponds to a potential *network configuration*:

e.g.  $\{X_{12}, X_{13}, X_{23}\}$  corresponds to



**Hammersley-Clifford theorem** (Besag, 1974):

A model for  $\mathbf{X}$  has a form determined by its neighbourhoods

**This general approach leads to *exponential random graph* or *p\* models***  
(Frank & Strauss 1986; extended by Wasserman, Robins & Pattison)

©Department of Psychology, University of Melbourne

## Exponential random graph models

$$P(\mathbf{X} = \mathbf{x}) = (1/c) \exp\{\sum_Q \gamma_Q z_Q(\mathbf{x})\}$$

*normalizing quantity*

*parameter*

*network statistic*

the summation is over all neighbourhoods  $Q$

$z_Q(\mathbf{x}) = \prod_{x_{ij} \in Q} x_{ij}$  signifies whether  
all ties in  $Q$  are observed in  $\mathbf{x}$

$$c = \sum_{\mathbf{x}} \exp\{\sum_Q \gamma_Q z_Q(\mathbf{x})\}$$

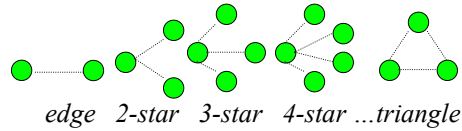
©Department of Psychology, University of Melbourne

## Neighbourhoods depend on proximity assumptions

Assumptions: two ties are neighbours:      Configurations for neighbourhoods

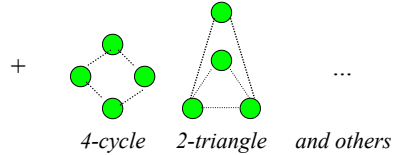
if they share an actor

*Markov*



if they complete a 4-cycle

*realisation-dependent*



©Department of Psychology, University of Melbourne

## Homogenous models

If we assume that *isomorphic neighbourhoods have equal parameters*, then:

There is one parameter for *each class* of network configurations

The corresponding statistic is the *number* of configurations in  $\mathbf{x}$

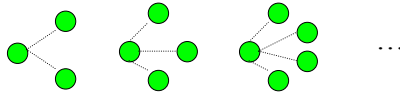
E.g. for a Markov model:

<i>Configurations</i>					...	
<i>Parameters</i>	$\theta$	$\sigma_2$	$\sigma_3$	$\sigma_4$	...	$\tau$
<i>Statistics</i>	$L(\mathbf{x})$	$S_2(\mathbf{x})$	$S_3(\mathbf{x})$	$S_4(\mathbf{x})$	...	$T(\mathbf{x})$

©Department of Psychology, University of Melbourne

## Related model parameters

*Star configurations*



*Parameters*

$\sigma_2$        $\sigma_3$        $\sigma_4$       ...

If we assume that  $\sigma_k = -\sigma_{k-1}/\lambda$ , for  $k > 1$  and  $\lambda \geq 1$  a (fixed) constant  
*alternating k-star hypothesis*

Then we obtain a single *star* parameter ( $\sigma_2$ ) with statistic:

$$S^{[\lambda]}(\mathbf{x}) = \sum_k (-1)^k S_k(\mathbf{x}) / \lambda^{k-2} \quad \textit{alternating k-star statistic}$$

*Note that if  $\lambda = 1$  and the edge parameter is included, the no. of isolated nodes is modelled separately*

## Geometrically weighted degree distribution

*An equivalent characterisation:*

Consider statistics  $d_k(\mathbf{x})$ , where  $d_k(\mathbf{x})$  is the number of nodes in  $\mathbf{x}$  of degree  $k$  (with corresponding parameters  $\theta_k$ )

Assuming that  $\theta_k = e^{-\alpha k \gamma}$  for  $k = 1, 2, \dots, n-1$        $\lambda = e^{\alpha / (e^{\alpha} - 1)}$

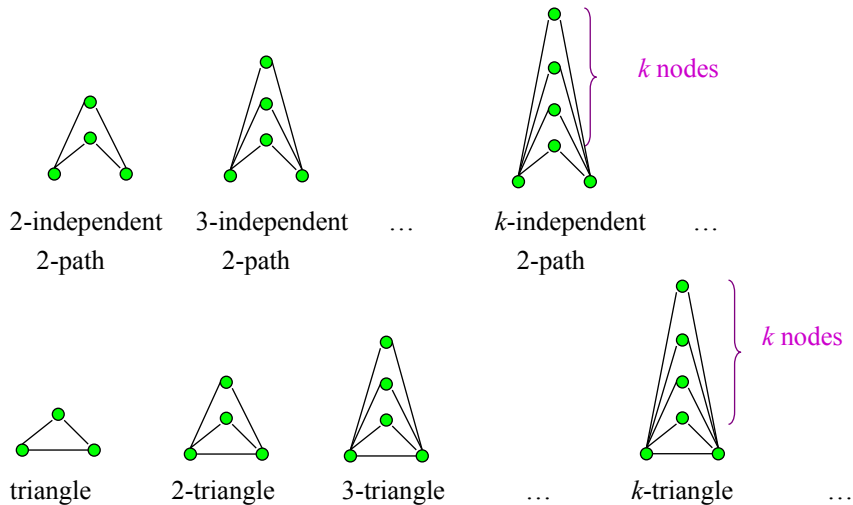
yields the statistic:

$$D^{[\alpha]}(\mathbf{x}) = \sum_k e^{-\alpha k} d_k(\mathbf{x}) \quad \textit{weighted degree distribution}$$

*Relationship with alternating k-star statistic:*

$$S^{[\lambda]}(\mathbf{x}) = \lambda^2 D^{[\alpha]}(\mathbf{x}) + 2\lambda L(\mathbf{x}) - n \lambda^2$$

## Additional neighbourhoods for realisation-dependent model



©Department of Psychology, University of Melbourne

## Additional statistics

### *k-independent 2-paths*

$U_k(\mathbf{x})$  = no of  $k$ -independent 2-paths in  $\mathbf{x}$ , with parameter  $v_k$

Let  $v_{k+1} = -v_k/\lambda$

Statistic for  $v_2$ :

$$U^{[\lambda]}(\mathbf{x}) = \sum_k (-1)^k U_k(\mathbf{x}) / \lambda^{k-2}$$

*alternating independent 2-path statistic*

### *k-triangles*

$T_k(\mathbf{x})$  = no of  $k$ -triangles in  $\mathbf{x}$ , with parameter  $\tau_k$

Let  $\tau_{k+1} = -\tau_k/\lambda$

Statistic for  $\tau_1$ :

$$T^{[\lambda]}(\mathbf{x}) = \sum_k (-1)^k T_k(\mathbf{x}) / \lambda^{k-2}$$

*alternating k-triangle statistic*

### Equivalently (Hunter & Hancock, 2005):

Statistic: No of dyads with exactly  $k$  shared partners

Aggregate statistic: *geometrically weighted dyad-wise shared partner distribution DSP*

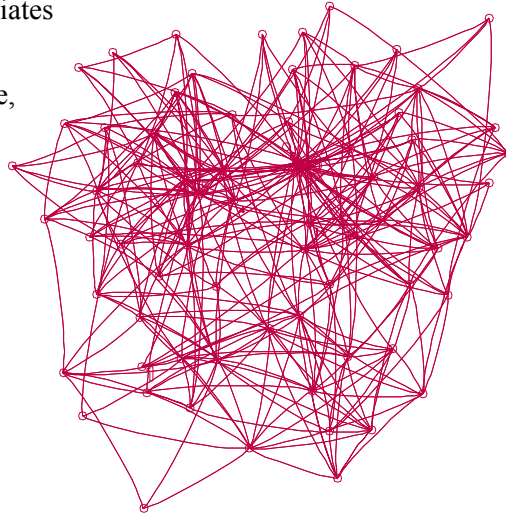
Statistic: No of dyads linked by an edge and having exactly  $k$  shared partners

Aggregate statistic: *Geometrically weighted edge-wise shared partner distribution ESP*

©Department of Psychology, University of Melbourne

## Collaboration network in a US law firm (Lazega)

36 partners, 35 associates  
3 offices  
2 practices (corporate,  
litigation)



©Department of Psychology, University of Melbourne

## Fitting exponential random graph models using MCMCMLE

### *Estimation approaches*

Snijders (2001): Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE)

Handcock, Hunter, et al (2004): implement Geyer & Thompson approach to MCMCMLE

### *Software:*

**SIENA** based on Snijders (2001)

**Statnet** (Handcock, Hunter et al, 2004)

**Pnet** (Wang, Robins & Pattison, 2005; also based on Snijders (2001)

See exercise for use of pnet

All programs still under development

©Department of Psychology, University of Melbourne

## Model for the collaboration network

	Effect	estimate (standard error)
<i>Exogenous</i>	Status main effect:	1.23 (.21)
	Same status effect	-1.55 (.31)
	Same gender effect:	0.25 (.12)
	Same practice effect:	1.27 (.16)
	Same office effect:	1.29 (.16)
<i>Endogenous</i>	Alternating <i>k</i> -stars:	-0.64 (.45)
	Alternating 2-paths:	0.03 (.03)
	Alternating <i>k</i> -triangles:	0.88 (.13)

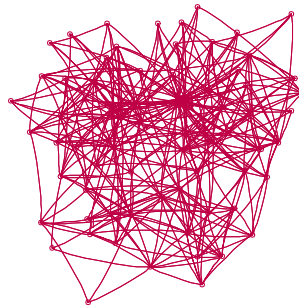
**NOTE:** *The model is fitted holding the number of edges constant*

©Department of Psychology, University of Melbourne

## Structural regularities

Ties are more likely when lawyers:

- are partners
- differ in status (partner/associate)
- work in the same office
- have the same practice
- are of the same gender



Structural “signature”:

- collaborators are likely to share multiple other collaborators
- the number of shared other collaborators varies, with relatively few pairs of collaborators having a very large number

©Department of Psychology, University of Melbourne

## Evaluating goodness of fit

Statistic	observed	fitted mean	t
2-stars	4818	4692	0.97
3-stars	24009	21733	1.34
4-stars	101266	82208	1.48
triangles	493	451	1.78
sd of degrees	5.74	5.42	0.97
skew of degrees	0.98	0.58	1.61
global clustering	0.31	0.29	1.72
mean local clustering	0.39	0.34	2.06
var of local clustering	0.03	0.02	0.10
geodesic distribution $Q_1$	2	2*	
geodesic distribution $Q_2$	2	2*	
geodesic distribution $Q_3$	3	3*	
geodesic distribution $Q_4$	infinite	4*	

\*Median reported rather than mean

©Department of Psychology, University of Melbourne

## Model specification: what have we learnt?

Relevant exogenous variables should be used

Realisation-dependent neighbourhoods appear to reflect social processes underlying network formation better than simple Markovian neighbourhoods

Hypotheses about relationships among the values of related parameters can provide practical and effective means of incorporating important higher-order effects without “death by parameter”: use

relevant *exogenous variables*

*edge, alternating 2-star, alternating k-independent 2-path and alternating k-triangle parameters (Siena, pnet);*

or, equivalently,

*edge, weighted degree distribution, DSP and ESP parameters (statnet)*

With well-specified models, MCMCMLE appears to work well

©Department of Psychology, University of Melbourne

## What can we use these social network models for?

Modelling allows us to:

- make inferences about hypothesised structural regularities in networks, eg
  - Kinship and generalized exchange**
  - Balance, clustering, transitivity, ranked cluster models**
  - Strength of weak ties**
- understand the relationship between *tie content* and *network structure*
- explore the *micro-macro gap*: can hypothesised structural regularities explain global network properties?
- investigate the *dynamics* of processes on networks

©Department of Psychology, University of Melbourne

## References

- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 36, 86-127.
- Feld, S. (1981). The focused organization of social ties. *American Journal of Sociology*, 86, 1015-1035.
- Frank, O., & Strauss, D. Markov graphs, *Journal of the American Statistical Association*, 81, 832-842.
- Handcock, Hunter et al. (2004). Statnet. Center for Statistics in the Social Sciences Working Paper, University of Washington.
- McPherson, M., Smith-Lovin, L., & Cook, (2001). Birds of a feather flock together. *Annual Review of Sociology*.
- Pattison, P. & Robins, G. (2002). Neighbourhood-based models for social networks. *Sociological Methodology*, 32, 301-337.
- Robins, G., Elliott, P., & Pattison, P. (2001). Network models for social selection processes. *Social Networks*, 23, 1-30.
- Snijders, T. A. B. (2001). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2).
- Snijders, T.A.B., Pattison, P., Robins, G., & Handcock, M. (2005). Exponential random graph models for social networks: new specifications.

©Department of Psychology, University of Melbourne