Curved exponential family models for networks

David R. Hunter, Penn State University

Mark S. Handcock, University of Washington

February 18, 2005

Available online as Penn State Dept. of Statistics Technical report 04-02 from

`http://www.stat.psu.edu/reports/2004`

Overview

This talk will focus on the alternating k-star and alternating k-triangle statistics of Snijders et al.*

General goals:

1. Present equivalent (and preferable) formulations of these statistics

$$\text{alternating } k\text{-stars} \quad \longrightarrow \quad \mathrm{d}^G(\mathbf{y}, \theta)$$
$$\text{alternating } k\text{-triangles} \quad \longrightarrow \quad \mathrm{p}^G(\mathbf{y}, \theta)$$

2. Introduce the mathematical issues that make model-fitting particularly challenging when using these statistics

* *Snijders et al. is W.P. # 42 at* `www.csss.washington.edu/Papers`

Alternating k-star statistic

The alternating k-star statistic is defined as

$$\mathrm{s}_2(\mathbf{y}) - \frac{\mathrm{s}_3(\mathbf{y})}{\gamma} + \cdots + (-1)^n \frac{\mathrm{s}_{n-1}(\mathbf{y})}{\gamma^{n-2}},$$

where $\mathrm{s}_k(\mathbf{y})$ denotes the number of k-stars in the network $\mathbf{y}$.

Alternating k-star statistic

The alternating k-star statistic is defined as

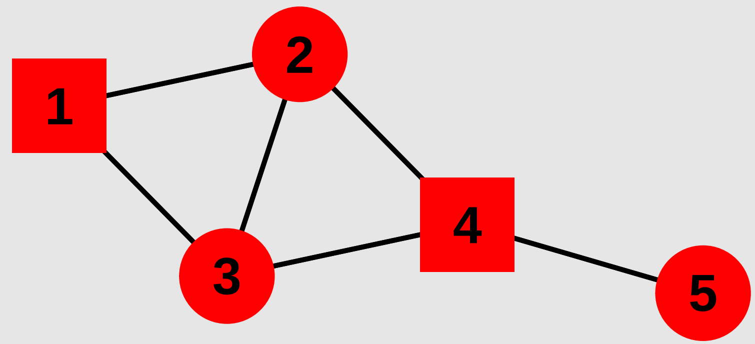
$$\mathrm{s}_2(\mathbf{y}) - \frac{\mathrm{s}_3(\mathbf{y})}{\gamma} + \cdots + (-1)^n \frac{\mathrm{s}_{n-1}(\mathbf{y})}{\gamma^{n-2}},$$

where $\mathrm{s}_k(\mathbf{y})$ denotes the number of k-stars in the network $\mathbf{y}$.

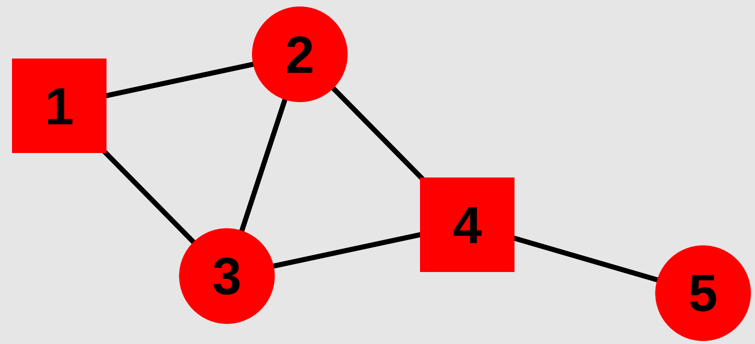
Consider the γ parameter. What does it do? How do we choose it?

Note in particular that if the alternating k-star statistic is used in a model, γ enters in a nonlinear way.

A small undirected network

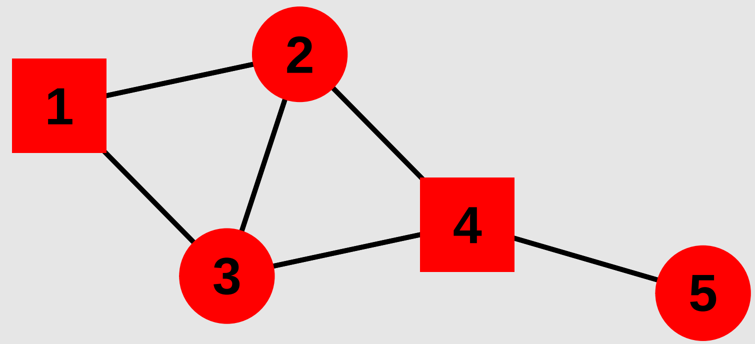


A small undirected network



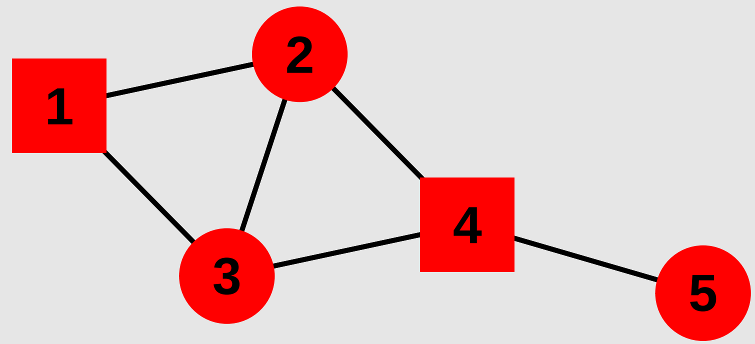
- Degree distribution: $(\mathrm{d}_0, \ldots, \mathrm{d}_{n-1}) = (0, 1, 1, 3, 0)$

A small undirected network



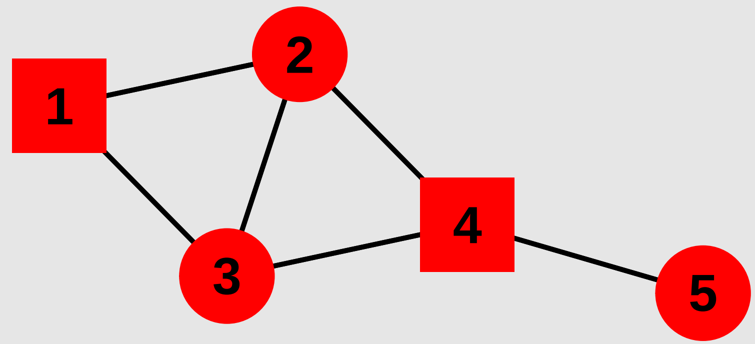
- Degree distribution: $(\mathrm{d}_0, \ldots, \mathrm{d}_{n-1}) = (0, 1, 1, 3, 0)$
- k-star distribution: $(\mathrm{s}_1, \ldots, \mathrm{s}_{n-1}) = (6, 10, 3, 0)$

A small undirected network



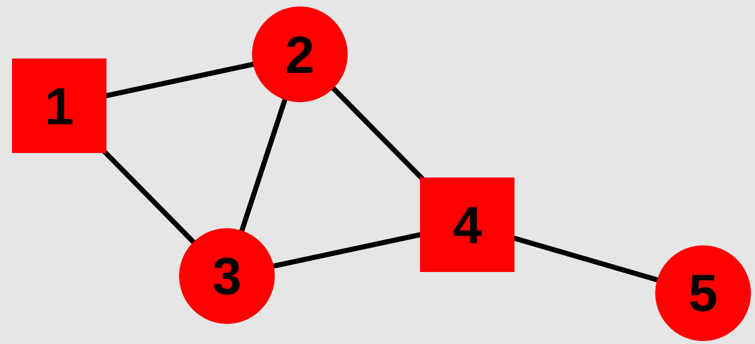
- Degree distribution: $(\mathrm{d}_0, \ldots, \mathrm{d}_{n-1}) = (0, 1, 1, 3, 0)$
- k-star distribution: $(\mathrm{s}_1, \ldots, \mathrm{s}_{n-1}) = (6, 10, 3, 0)$
- Edgewise shared partner distribution: $(\mathrm{p}_0, \ldots, \mathrm{p}_{n-2}) = (1, 4, 1, 0)$

A small undirected network



- Degree distribution: $(\mathrm{d}_0, \ldots, \mathrm{d}_{n-1}) = (0, 1, 1, 3, 0)$
- k-star distribution: $(\mathrm{s}_1, \ldots, \mathrm{s}_{n-1}) = (6, 10, 3, 0)$
- Edgewise shared partner distribution: $(\mathrm{p}_0, \ldots, \mathrm{p}_{n-2}) = (1, 4, 1, 0)$
- k-triangle distribution: $(\mathrm{t}_1, \ldots, \mathrm{t}_{n-2}) = (2, 1, 0)$

A small undirected network



- Degree distribution: $(\mathrm{d}_0, \ldots, \mathrm{d}_{n-1}) = (0, 1, 1, 3, 0)$
- k-star distribution: $(\mathrm{s}_1, \ldots, \mathrm{s}_{n-1}) = (6, 10, 3, 0)$
- Edgewise shared partner distribution: $(\mathrm{p}_0, \ldots, \mathrm{p}_{n-2}) = (1, 4, 1, 0)$
- k-triangle distribution: $(\mathrm{t}_1, \ldots, \mathrm{t}_{n-2}) = (2, 1, 0)$

Relationship between edgewise shared partners and k-triangles: analagous to the relationship between degrees and k-stars *(and also the relationship between dyadic shared partners and alternating independent 2-paths.)*

Rewriting alternating k-stars

The alternating k-star statistic

$$\mathrm{s}_2(\mathbf{y}) - \frac{\mathrm{s}_3(\mathbf{y})}{\gamma} + \cdots + (-1)^{n-1}\frac{\mathrm{s}_{n-1}(\mathbf{y})}{\gamma^{n-3}}$$

may be rewritten (brace yourself):

Rewriting alternating k-stars

The alternating k-star statistic

$$\mathrm{s}_2(\mathbf{y}) - \frac{\mathrm{s}_3(\mathbf{y})}{\gamma} + \cdots + (-1)^{n-1}\frac{\mathrm{s}_{n-1}(\mathbf{y})}{\gamma^{n-3}}$$

may be rewritten (brace yourself):

$$\mathrm{d}^G(\mathbf{y};\theta) \quad = \quad 2e^{\theta}\mathrm{s}_1(\mathbf{y}) - \sum_{i=1}^{n-1} e^{2\theta}\left[1 - \left(1 - e^{\theta}\right)^i\right]\mathrm{d}_i(\mathbf{y}),$$

where:

- γ is replaced by e^{θ} (to ensure $\gamma > 0$)
- $\mathrm{s}_k(\mathbf{y}) =$ # of k-stars in the graph $\mathbf{y}$. (In particular, $\mathrm{s}_1 =$ # of edges.)
- $\mathrm{d}_k(\mathbf{y}) =$ # of nodes of degree k in $\mathbf{y}$.

Alternating k-triangle statistic, rewritten

The alternating k-triangle statistic of Snijders et al. (2004) is

$$3\mathrm{t}_1(\mathbf{y}) - \frac{\mathrm{t}_2(\mathbf{y})}{\gamma} + \cdots + (-1)^{n-1}\frac{\mathrm{t}_{n-2}(\mathbf{y})}{\gamma^{n-3}}.$$

In analogy with the alternating k-star case, we rewrite:

Alternating k-triangle statistic, rewritten

The alternating k-triangle statistic of Snijders et al. (2004) is

$$3\mathrm{t}_1(\mathbf{y}) - \frac{\mathrm{t}_2(\mathbf{y})}{\gamma} + \cdots + (-1)^{n-1}\frac{\mathrm{t}_{n-2}(\mathbf{y})}{\gamma^{n-3}}.$$

In analogy with the alternating k-star case, we rewrite:

$$\mathrm{p}^G(\mathbf{y};\theta) = \sum_{i=1}^{n-2} e^{\theta}\left\{1 - \left(1 - e^{-\theta}\right)^i\right\}\mathrm{p}_i(\mathbf{y}),$$

where

- γ is replaced by e^{θ} (to ensure $\gamma > 0$)
- $\mathrm{t}_k(\mathbf{y}) =$ # of k-triangles in the graph $\mathbf{y}$.
- $\mathrm{p}_k(\mathbf{y}) =$ # of nodes with k edgewise shared partners in $\mathbf{y}$.

An important question

We have shown that

- the alternating k-star statistic is the same as $\mathrm{d}^G(\mathbf{y}, \theta)$
- the alternating k-triangle statistic is the same as $\mathrm{p}^G(\mathbf{y}, \theta)$

where $\theta = \log \gamma$.

Suppose we wish to include $\mathrm{d}^G(\mathbf{y}, \theta_1)$ and/or $\mathrm{p}^G(\mathbf{y}, \theta_2)$ in an ERGM, but we wish to <u>estimate</u> θ_1 and θ_2.

How do we do it?

ERGM specification

For a (random, as-yet-unobserved) graph $\mathbf{Y}$, we assume

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\eta}^t \mathbf{g}(\mathbf{y})\}}{c(\boldsymbol{\eta})} = \frac{\exp\{\eta_1 g_1(\mathbf{y}) + \cdots + \eta_p g_p(\mathbf{y})\}}{c(\boldsymbol{\eta})}$$

for all possible realizations $\mathbf{y}$.

The name ERGM (exponential random graph model) arises because this model is based on a statistical exponential family.

ERGM specification

For a (random, as-yet-unobserved) graph $\mathbf{Y}$, we assume

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\eta}^t \mathbf{g}(\mathbf{y})\}}{c(\boldsymbol{\eta})} = \frac{\exp\{\eta_1 g_1(\mathbf{y}) + \cdots + \eta_p g_p(\mathbf{y})\}}{c(\boldsymbol{\eta})}$$

for all possible realizations $\mathbf{y}$.

The name ERGM (exponential random graph model) arises because this model is based on a statistical exponential family.

As usual, $\mathbf{g}(\mathbf{y})$ is a vector of statistics to be specified by the modeler (and p is the number of statistics).

The vector $\boldsymbol{\eta}$ is sometimes called the <u>canonical</u> parameter.

Not quite an ERGM?

The ERGM says

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1 g_1(\mathbf{y}) + \cdots + \eta_p g_p(\mathbf{y})\}}{c(\boldsymbol{\eta})}.$$

But suppose $\mathbf{g}(\mathbf{y})$ consists of only the statistic $\mathrm{p}^G(\mathbf{y}, \theta)$.

Thus, we wish to estimate the parameters θ_1 and θ_2 in the model

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\theta_1 \mathrm{p}^G(\mathbf{y}, \theta_2)\}}{c(\boldsymbol{\theta})}.$$

Not quite an ERGM?

The ERGM says

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1 g_1(\mathbf{y}) + \cdots + \eta_p g_p(\mathbf{y})\}}{c(\boldsymbol{\eta})}.$$

But suppose $\mathbf{g}(\mathbf{y})$ consists of only the statistic $\mathrm{p}^G(\mathbf{y}, \theta)$.

Thus, we wish to estimate the parameters θ_1 and θ_2 in the model

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\theta_1 \mathrm{p}^G(\mathbf{y}, \theta_2)\}}{c(\boldsymbol{\theta})}.$$

The second equation is not in ERGM form because the parameters are mixed up with the statistics!

η vs. θ

Earlier, we showed

$$\mathrm{p}^G(\mathbf{y};\theta) = \sum_{i=1}^{n-2} e^{\theta}\left\{1-\left(1-e^{-\theta}\right)^i\right\}\mathrm{p}_i(\mathbf{y}).$$

Thus, the model we wish to fit turns into

$$\begin{aligned} P(\mathbf{Y}=\mathbf{y}) &= \frac{\exp\{\theta_1 \mathrm{p}^G(\mathbf{y},\theta_2)\}}{c} \\ &= \frac{\exp\left[\sum_{i=1}^{n-2}\theta_1 e^{\theta_2}\left\{1-\left(1-e^{-\theta_2}\right)^i\right\}\mathrm{p}_i(\mathbf{y})\right]}{c}. \end{aligned}$$

Thus, we can write η_i (the coefficient of the ith statistic) as a function of $\boldsymbol{\theta}$ (the parameter vector we want to estimate).

Curved exponential family models

The original model has turned into

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1(\boldsymbol{\theta})\mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2}(\boldsymbol{\theta})\mathrm{p}_{n-2}(\mathbf{y})\}}{c[\boldsymbol{\eta}(\boldsymbol{\theta})]}.$$

Curved exponential family models

The original model has turned into

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1(\boldsymbol{\theta})\mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2}(\boldsymbol{\theta})\mathrm{p}_{n-2}(\mathbf{y})\}}{c[\boldsymbol{\eta}(\boldsymbol{\theta})]}.$$

Thus, $\boldsymbol{\eta}$ is the vector of coefficients,
whereas $\boldsymbol{\theta}$ is the parameter vector to be estimated.

Often, the two vectors are the same so this distinction is ignored.

Curved exponential family models

The original model has turned into

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1(\boldsymbol{\theta})\mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2}(\boldsymbol{\theta})\mathrm{p}_{n-2}(\mathbf{y})\}}{c[\boldsymbol{\eta}(\boldsymbol{\theta})]}.$$

Thus, $\boldsymbol{\eta}$ is the vector of coefficients,
whereas $\boldsymbol{\theta}$ is the parameter vector to be estimated.

Often, the two vectors are the same so this distinction is ignored.

But sometimes, $\boldsymbol{\eta}(\boldsymbol{\theta})$ is a nonlinear function;
the equation above imposes nonlinear constraints on $\boldsymbol{\eta}$.
In that case, statisticians call the model a curved exponential family.

Staving off death

The complication of a nonlinear constraint on $\boldsymbol{\eta}$ can actually be a good thing:

What happens if we try to estimate the unconstrained $\boldsymbol{\eta}$ vector in

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1 \mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2}\mathrm{p}_{n-2}(\mathbf{y})\}}{c(\boldsymbol{\eta})}?$$

Staving off death

The complication of a nonlinear constraint on $\boldsymbol{\eta}$ can actually be a good thing:

What happens if we try to estimate the unconstrained $\boldsymbol{\eta}$ vector in

$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1 \mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2} \mathrm{p}_{n-2}(\mathbf{y})\}}{c(\boldsymbol{\eta})}?$$

Answer: In the words of Pip Pattison, “Death By Parameter”!

Staving off death

The complication of a nonlinear constraint on $\boldsymbol{\eta}$ can actually be a good thing:

What happens if we try to estimate the unconstrained $\boldsymbol{\eta}$ vector in

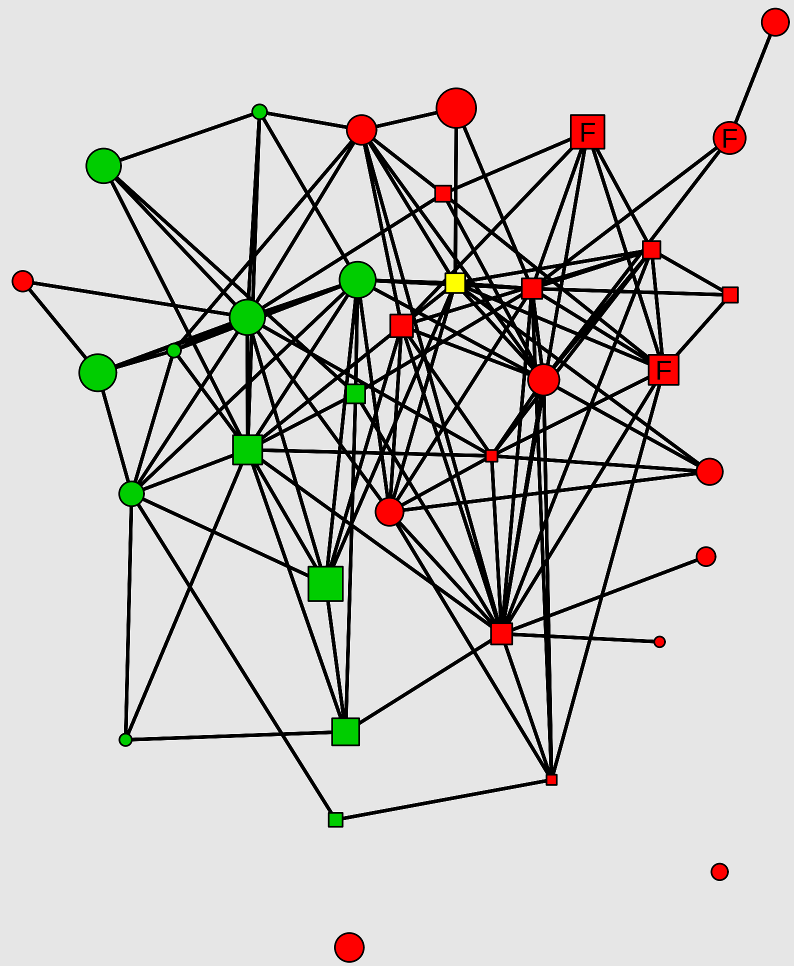
$$P(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\eta_1 \mathrm{p}_1(\mathbf{y}) + \cdots + \eta_{n-2}\mathrm{p}_{n-2}(\mathbf{y})\}}{c(\boldsymbol{\eta})}?$$

Answer: In the words of Pip Pattison, “Death By Parameter”!

Boiling the entire $\boldsymbol{\eta}$ vector down into a function of just the (θ_1, θ_2) is actually healthy.

But nothing that’s good for you is ever fun, so there’s more work to be done at the estimation step. *See paper for details.*

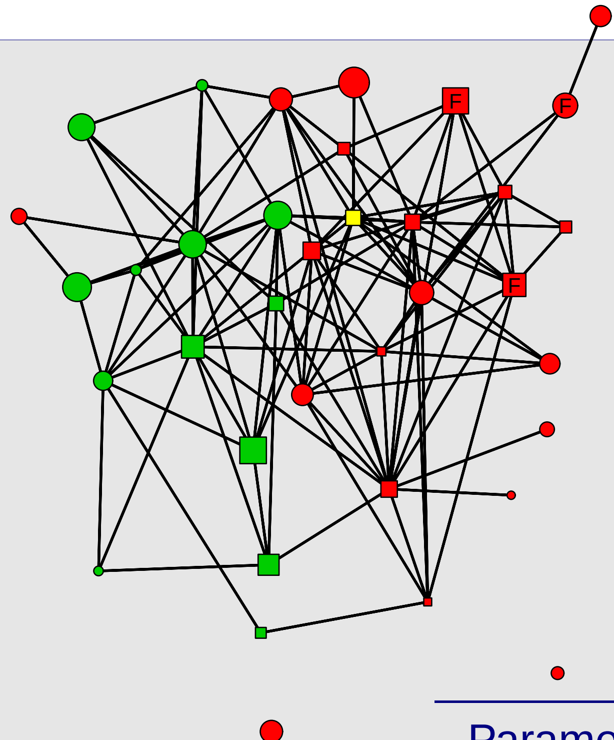
Lazega's lawyer collaboration data



Sizes indicate seniority (larger=more recent); colors indicate office location;

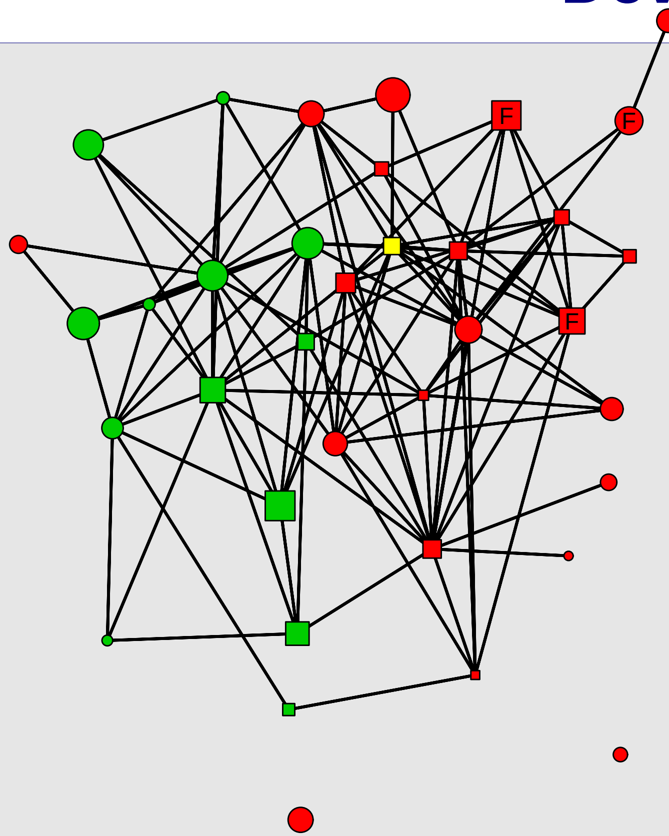
F indicates female; shapes indicate practice (circle=litigation, square=corporate)

Coefficient estimates (generated by `statnet`)



	Model 1		Model 2	
Parameter	est.	s.e.	est.	s.e.
Alternating k-triangles	0.612	0.091	0.878	0.279
Rate of transitivity	1.099	—	0.814	0.196
Seniority main effect	0.024	0.006	0.023	0.006
Practice main effect	0.352	0.113	0.390	0.117
Same practice	0.708	0.194	0.757	0.194
Same gender	0.621	0.257	0.688	0.248
Same office	1.151	0.195	1.123	0.194

Deviance analysis



Model	Residual Deviance	Deviance	Residual d.f.	p-value
NULL	598.78	–	–	–
Covariates	501.80	96.98	5	0.000
Full model	456.21	45.59	2	0.000

“Covariates”: The model with only the covariate terms

“Full model”: The model with covariate terms plus $\mathrm{p}^G(\mathbf{y}, \theta)$

Conclusion

Other things covered in our paper:

Conclusion

Other things covered in our paper:

• A general formulation of the problem of fitting curved exponential family models

Conclusion

Other things covered in our paper:

- A general formulation of the problem of fitting curved exponential family models

- Numerical algorithms for estimating curved EF parameters and their standard errors

Conclusion

Other things covered in our paper:

- A general formulation of the problem of fitting curved exponential family models
- Numerical algorithms for estimating curved EF parameters and their standard errors
- How to estimate likelihood ratio statistics and loglikelihoods using MCMC

Conclusion

Other things covered in our paper:

- A general formulation of the problem of fitting curved exponential family models
- Numerical algorithms for estimating curved EF parameters and their standard errors
- How to estimate likelihood ratio statistics and loglikelihoods using MCMC

Huge thanks to:

Tom Snijders for extremely helpful suggestions about the manuscript;
Steve Goodreau for blackboard brainstorming sessions;

Martina Morris, Garry Robins, and Pip Pattison for insightful comments.