MANCHESTER 1824

The Social Statistics Discipline Area, School of Social Sciences

The University of Manchester

An introduction to exponential random graph models (**ERGM**)

Johan Koskinen

http://www.ccsr.ac.uk/staff/jk.htm

johan.koskinen@manchester.ac.uk

Mitchell Centre for Network Analysis

EURANDOM

Jan 29 2013, Statistical Analysis of Social Networks, YES Workshop on Statistics for Complex and High Dimensional Systems, Eindhoven

Exponential random graph models

ERGM: probability model for **adjacency matrices**

with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

Exponential random graph models

ERGM: probability model for **adjacency matrices** with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

Parameter vector: θ

Graph statistics: $z(x)$

Normalising constant: $\psi(\theta)$

Exponential random graph models

ERGM: probability model for **adjacency matrices**

with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

x **takes values in**

$$\aleph = \{0,1\}^{V^{(2)}}$$

Parameter vector: θ

Graph statistics: $z(x)$

Normalising constant: $\psi(\theta)$

Exponential random graph models

ERGM: probability model for **adjacency matrices** with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

Parameter vector: θ

Graph statistics: $z(x)$

Normalising constant: $\psi(\theta)$

***x* takes values in**

$$\aleph = \{0,1\}^{V^{(2)}}$$

***z*( ) can be chosen in many ways**

Exponential random graph models

ERGM: probability model for **adjacency matrices**

with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

***x* takes values in**

$$\aleph = \{0,1\}^{V^{(2)}}$$

Parameter vector: θ

Graph statistics: $z(x)$

Normalising constant: $\psi(\theta)$

***z*( ) can be chosen in many ways**

ψ*( ) difficult function of *θ

$$\psi(\theta) = \log \sum_x e^{\theta^T z(x)}$$

Exponential random graph models

ERGM: probability model for **adjacency matrices**

with pmf:

$$p(x) = \exp\{\theta^T z(x) - \psi(\theta)\}$$

Parameter vector: θ

Graph statistics: $z(x)$

Normalising constant: $\psi(\theta)$

x **takes values in**

$$\aleph = \{0,1\}^{V^{(2)}}$$

$z(\ )$ **can be chosen in many ways**

$\psi(\ )$ **difficult function of** θ

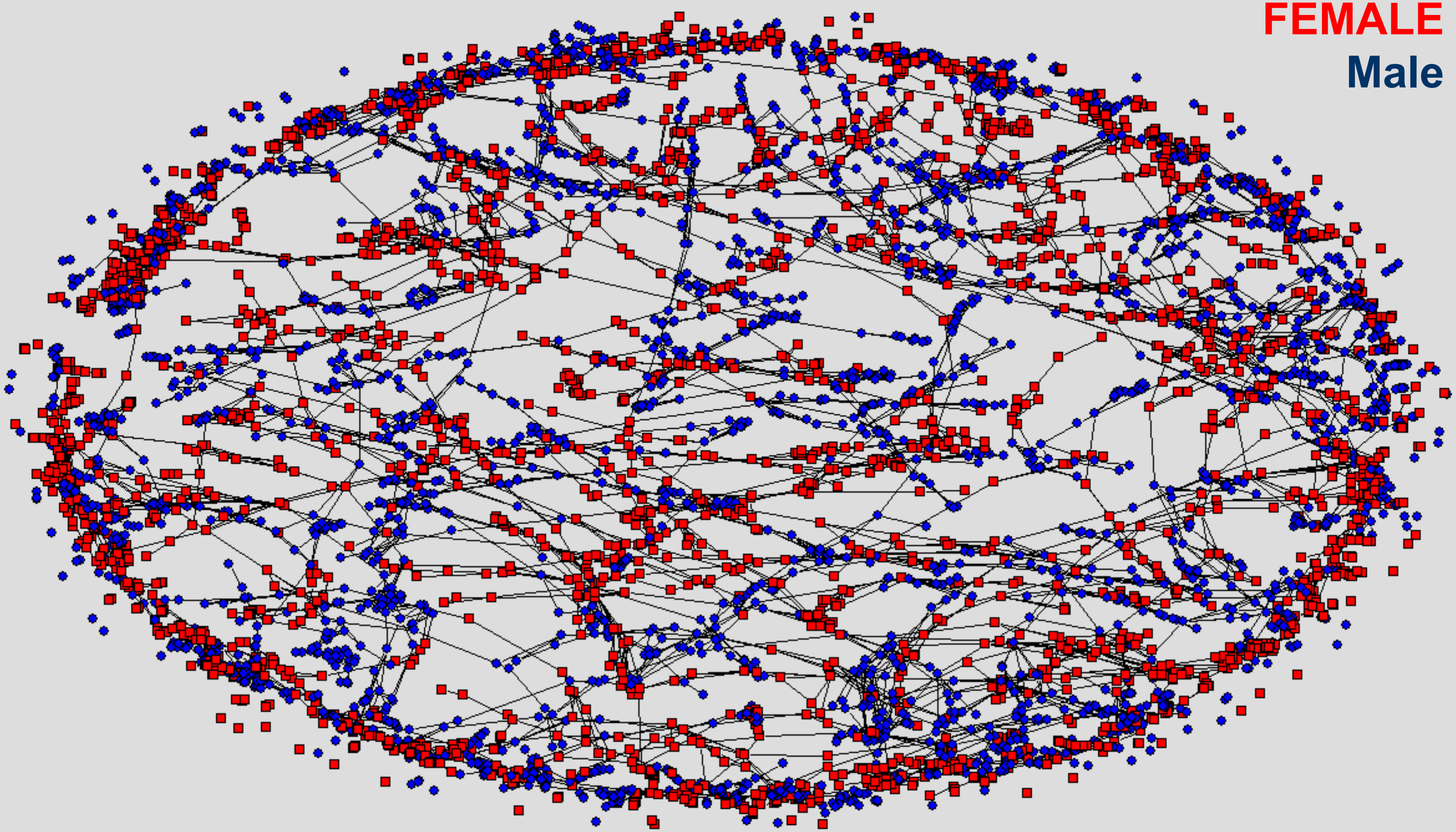
$$\psi(\theta) = \log \sum_x e^{\theta^T z(x)}$$

… it is an exponential family distribution (hence ERGM)

Part 1a

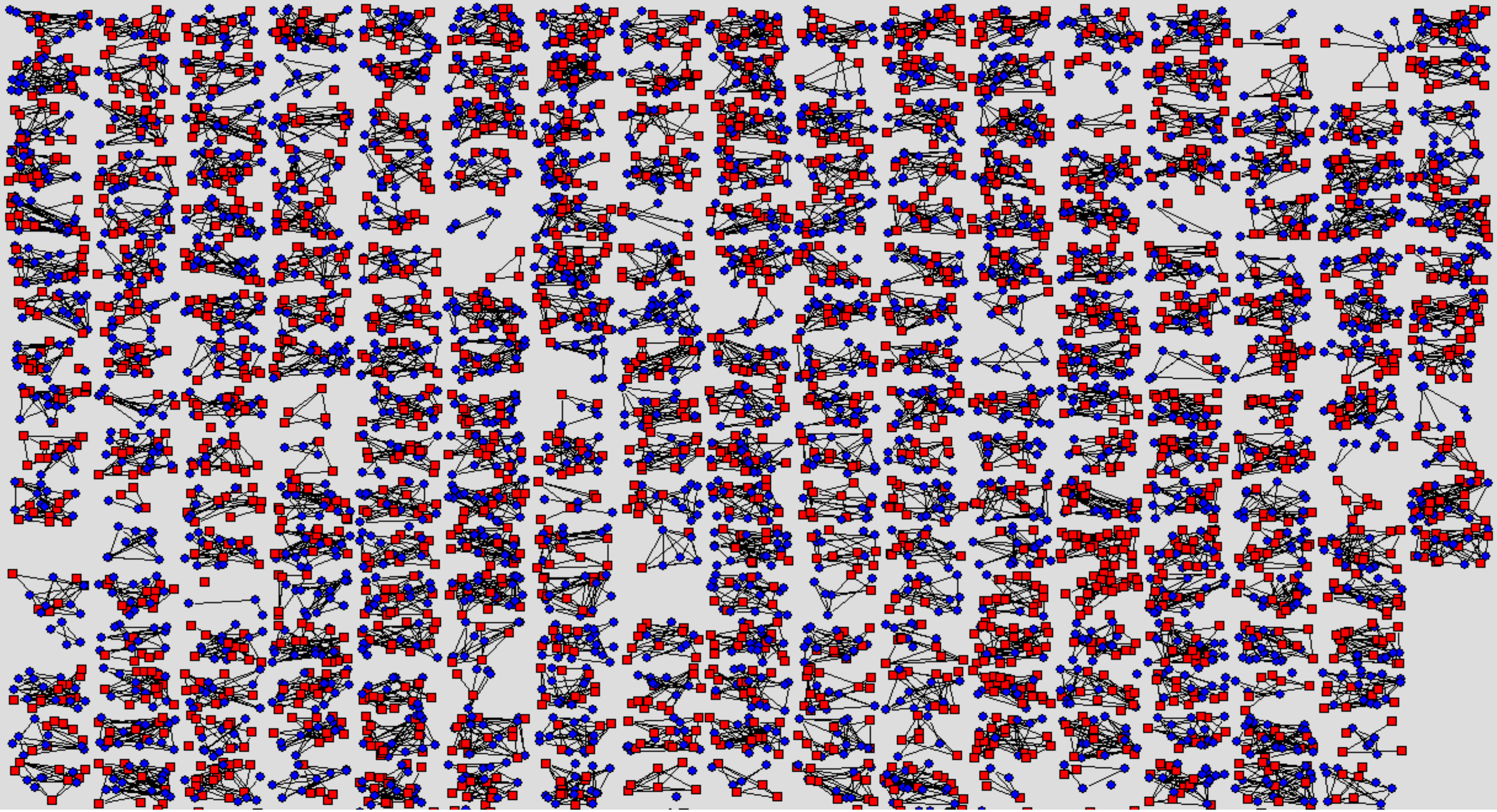
Why an ERGM

Networks matter – ERGMS matter



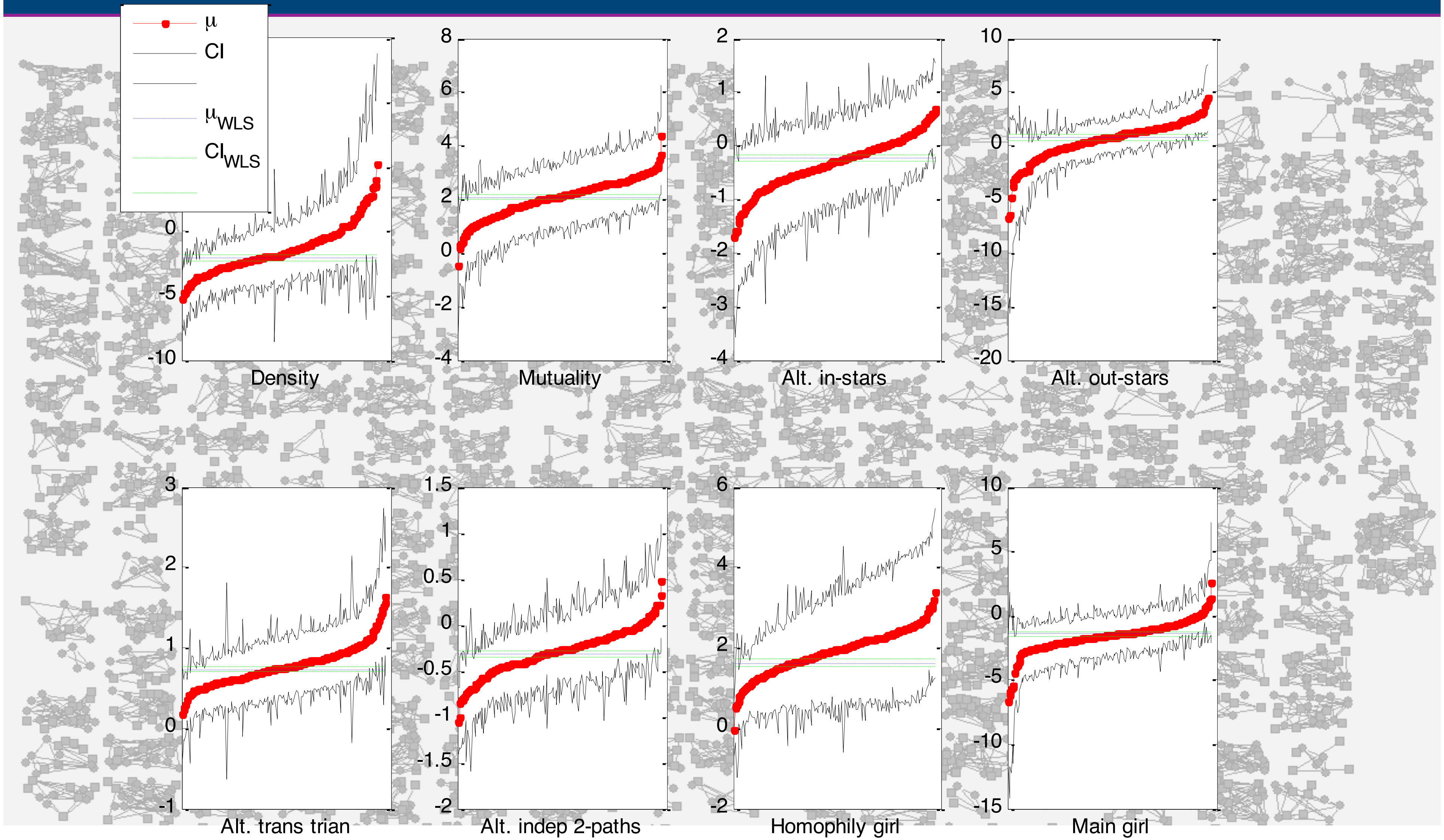
6018 grade 6 children 1966

Networks matter – ERGMS matter



6018 grade 6 children 1966 – 300 schools Stockholm

Networks matter – ERGMS matter



6018 grade 6 children 1966 – 200 schools Stockholm

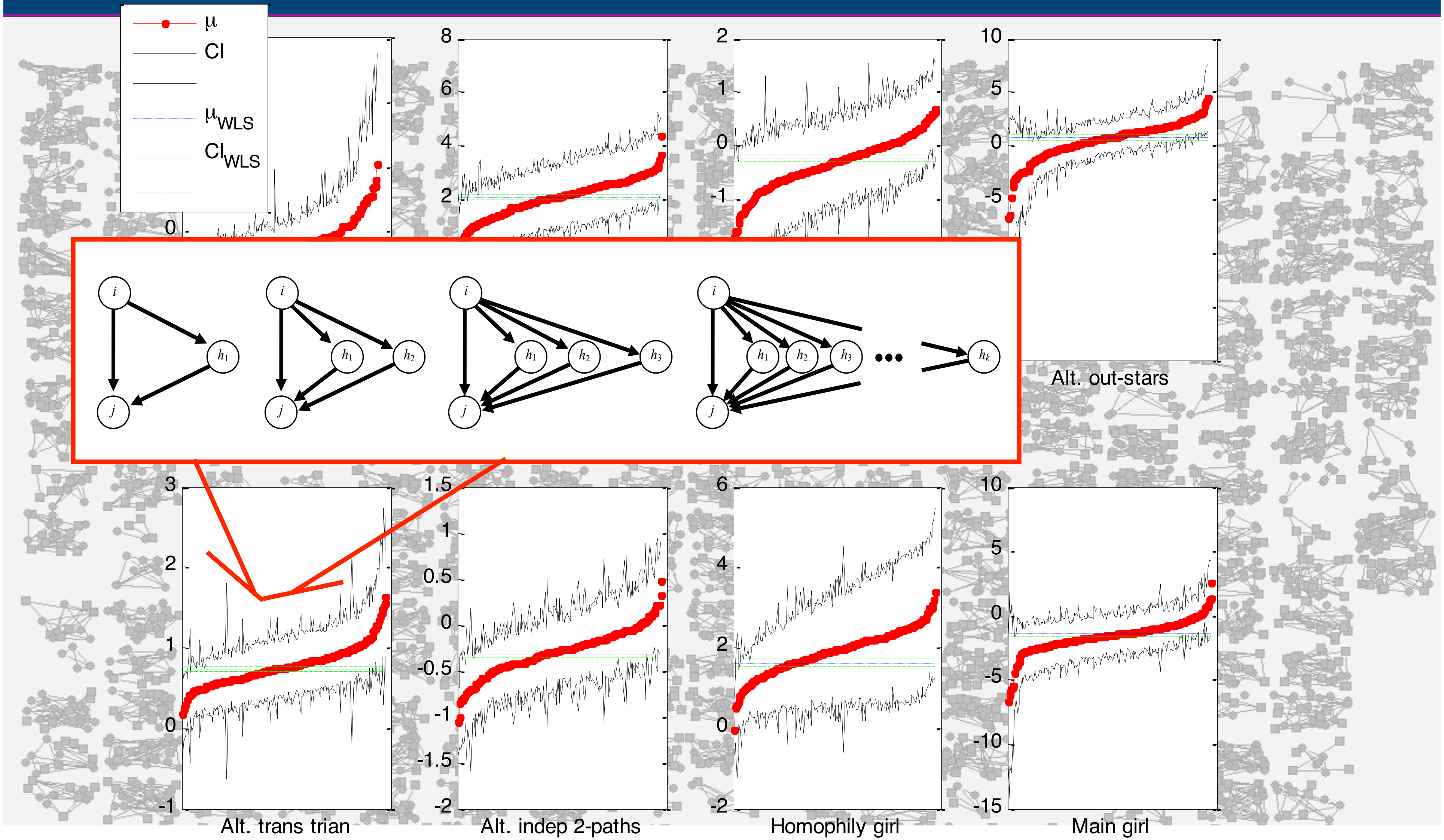
Koskinen and Stenberg (in press) JEBS

Networks matter – ERGMS matter

6018 grade 6 children 1966 – 200 schools Stockholm

Koskinen and Stenberg (in press) JEBS

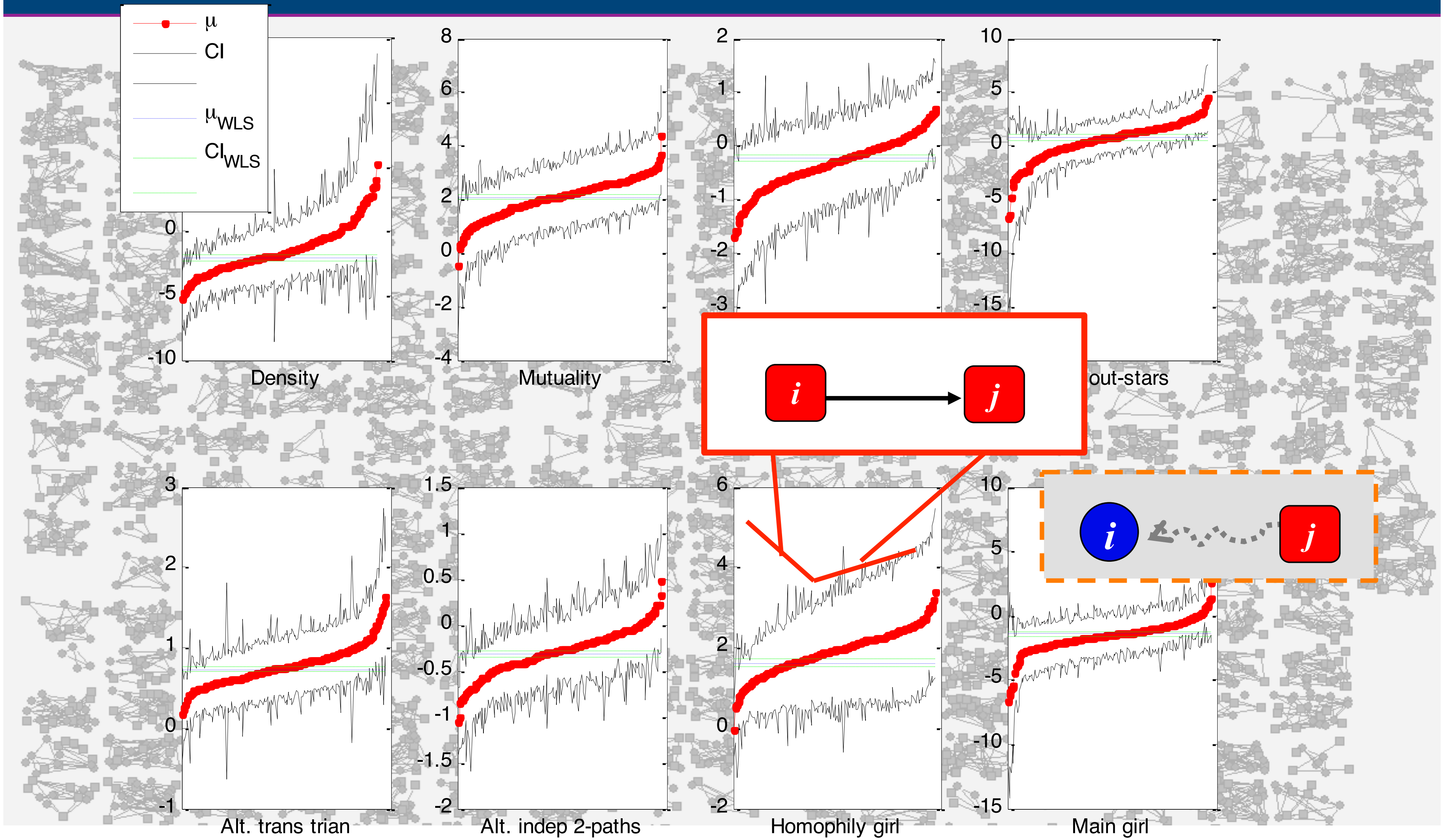
Networks matter – ERGMS matter



6018 grade 6 children 1966 – 200 schools Stockholm

Koskinen and Stenberg (in press) JEBS

Networks matter – ERGMS matter



6018 grade 6 children 1966 – 200 schools Stockholm

Koskinen and Stenberg (in press) JEBS

The University of Manchester

Part 1b

Minimum learning outcomes

- Get a working handle on what we are trying to model
- Familiarise ourselves with common model specifications
- Fit our first models in `statnet` and PNet
- Fitting procedure
  - ❑ Estimation
  - ❑ Convergence check
  - ❑ Goodness of fit
- Orientation about future developments

Part 1c

Modelling graphs

- Notational preliminaries
- Why and what is an ERGM
- Dependencies
- Estimation
  - Geyer-Thompson
  - Robins-Monro
  - Bayes (The issue, Moller et al, LISA, exchange algorithm)
- Interpretation of effects
- Convergence and GOF
- Further issues

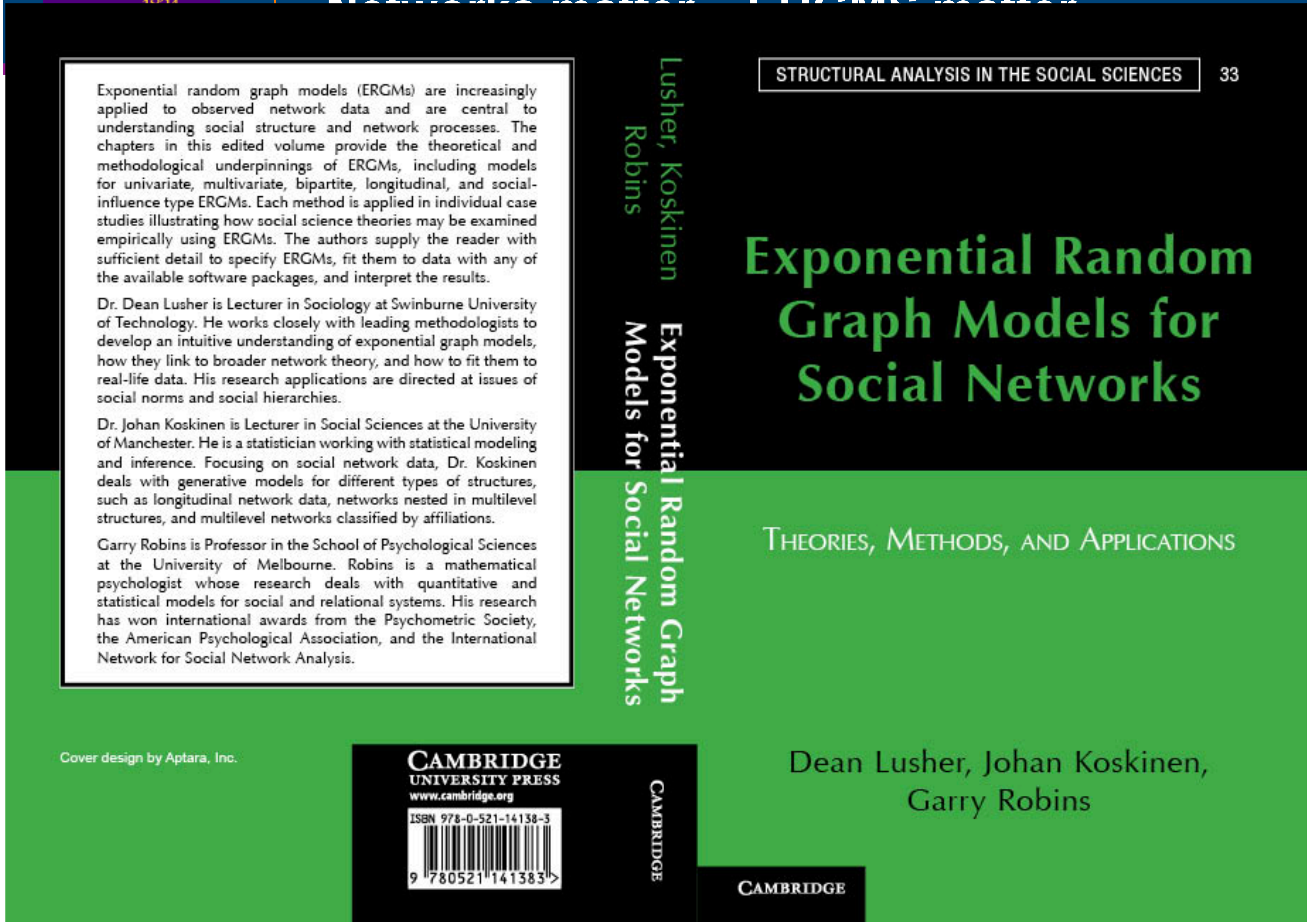
Networks matter – ERGMS matter

Numerous recent substantively driven studies

- Gondal, The local and global structure of knowledge production in an emergent research field, SOCNET 2011
- Lomi and Palotti, Relational collaboration among spatial multipoint competitors , SOCNET 2011
- Wimmer & Lewis, Beyond and Below Racial Homophily, AJS 2010
- Lusher, Masculinity, educational achievement and social status, GENDER & EDUCATION 2011
- Rank et al. (2010). Structural logic of intra-organizational networks, ORG SCI, 2010.

Book for applied researchers: Lusher, Koskinen, Robins ERGMs for SN, CUP, 2011

Networks matter – ERGMS matter



Exponential random graph models (ERGMs) are increasingly applied to observed network data and are central to understanding social structure and network processes. The chapters in this edited volume provide the theoretical and methodological underpinnings of ERGMs, including models for univariate, multivariate, bipartite, longitudinal, and social-influence type ERGMs. Each method is applied in individual case studies illustrating how social science theories may be examined empirically using ERGMs. The authors supply the reader with sufficient detail to specify ERGMs, fit them to data with any of the available software packages, and interpret the results.

Dr. Dean Lusher is Lecturer in Sociology at Swinburne University of Technology. He works closely with leading methodologists to develop an intuitive understanding of exponential graph models, how they link to broader network theory, and how to fit them to real-life data. His research applications are directed at issues of social norms and social hierarchies.

Dr. Johan Koskinen is Lecturer in Social Sciences at the University of Manchester. He is a statistician working with statistical modeling and inference. Focusing on social network data, Dr. Koskinen deals with generative models for different types of structures, such as longitudinal network data, networks nested in multilevel structures, and multilevel networks classified by affiliations.

Garry Robins is Professor in the School of Psychological Sciences at the University of Melbourne. Robins is a mathematical psychologist whose research deals with quantitative and statistical models for social and relational systems. His research has won international awards from the Psychometric Society, the American Psychological Association, and the International Network for Social Network Analysis.

Cover design by Aptara, Inc.

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org
ISBN 978-0-521-14138-3

Lusher, Koskinen Robins — Exponential Random Graph Models for Social Networks — CAMBRIDGE

STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES 33

Exponential Random Graph Models for Social Networks

THEORIES, METHODS, AND APPLICATIONS

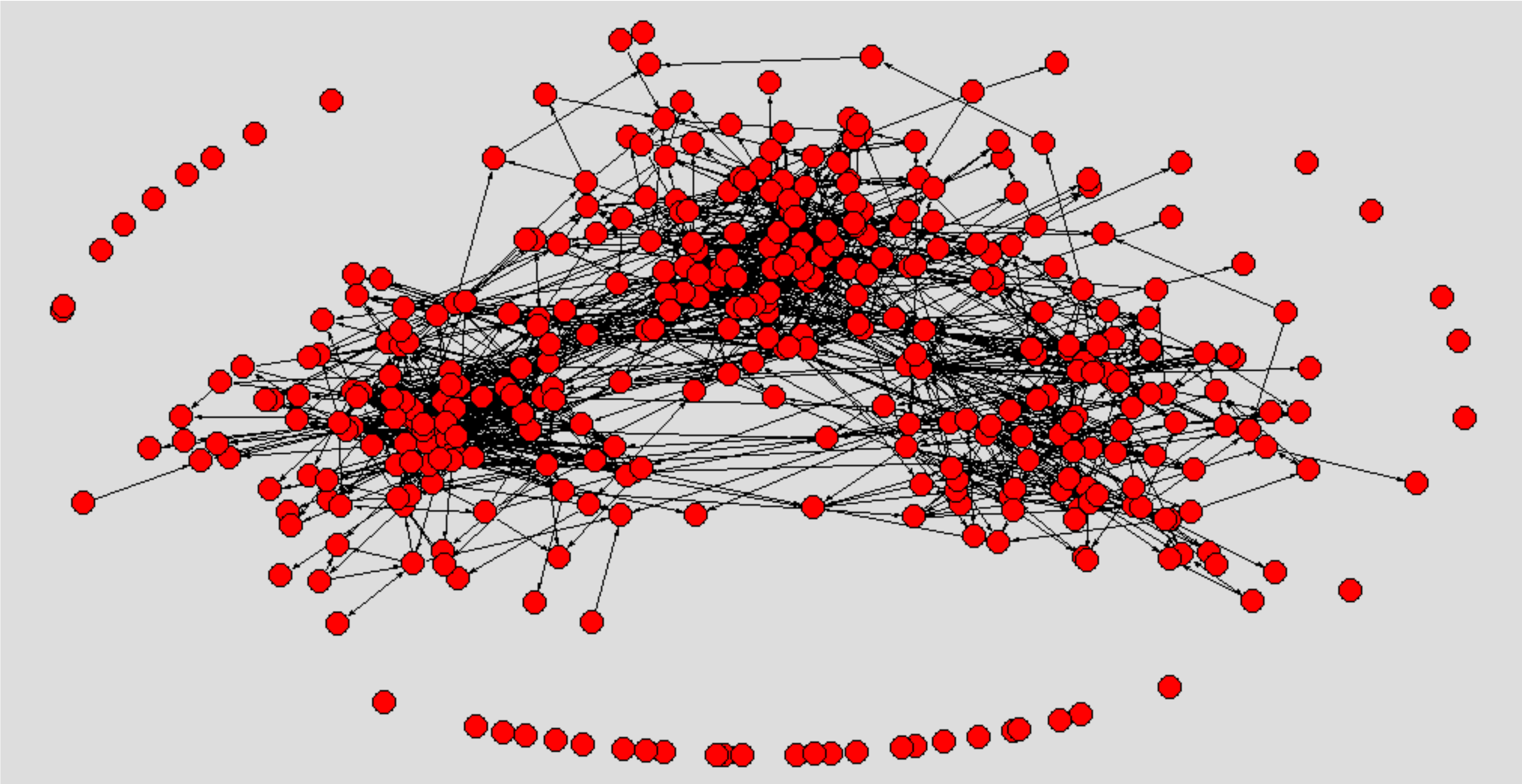
Dean Lusher, Johan Koskinen, Garry Robins

CAMBRIDGE

The exponential random graph model (p*) framework

- An ERGM (p*) model is a statistical model for the **ties** in a network
- Independent (pairs of) ties (p1, Holland and Leinhardt, 1981; Fienberg and Wasserman, 1979, 1981)
- Markov graphs (Frank and Strauss, 1986)
- Extensions (Pattison & Wasserman, 1999; Robins, Pattison & Wasserman, 1999; Wasserman & Pattison, 1996)
- New specifications (Snijders et al., 2006; Hunter & Handcock, 2006)

ERGMS – modelling graphs

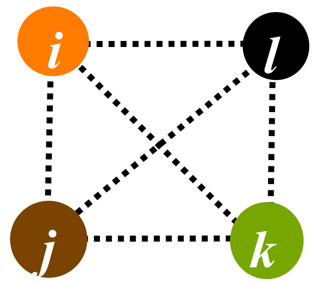


ERGMS – modelling graphs

- We want to model **tie variables**
- But **structure** – overall pattern – is evident
- What kind of **structural elements** can we incorporate in the model for the tie variables?

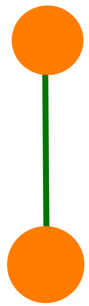
ERGMS – modelling graphs

If we believe that the frequency of interaction/**density** is an important aspect of the network



We should include

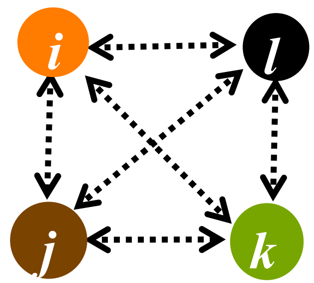
Counts of



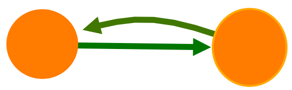
the number of ties in our model

ERGMS – modelling graphs

If we believe that the **reciprocity** is an important aspect of the (directed) network

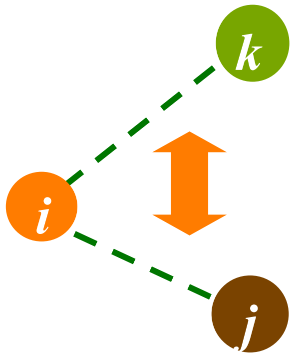


We should include

Counts of  the number of mutual ties in our model

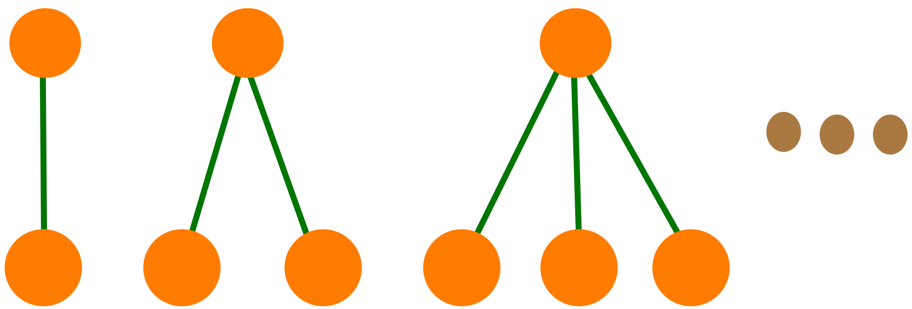
ERGMS – modelling graphs

If we believe that an important aspect of the network is that

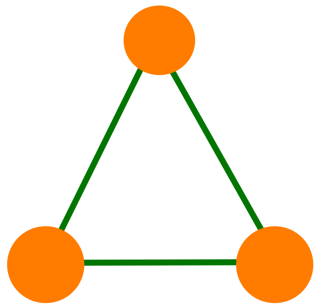


two edge indicators $\{i,j\}$ and $\{i',k\}$ are conditionally **dependent** if $\{i,j\} \cap \{i',k\} \neq \emptyset$

We should include counts of



degree distribution; preferential attachment, etc



friends meet through friends; clustering; etc

ERGMS – modelling graphs

If we believe that the **attributes** of the actors are important (selection effects, homophily, etc)

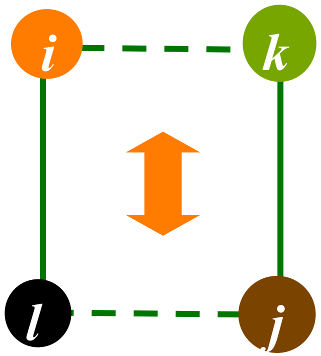
We should include counts of

Heterophily/homophily

Distance/similarity

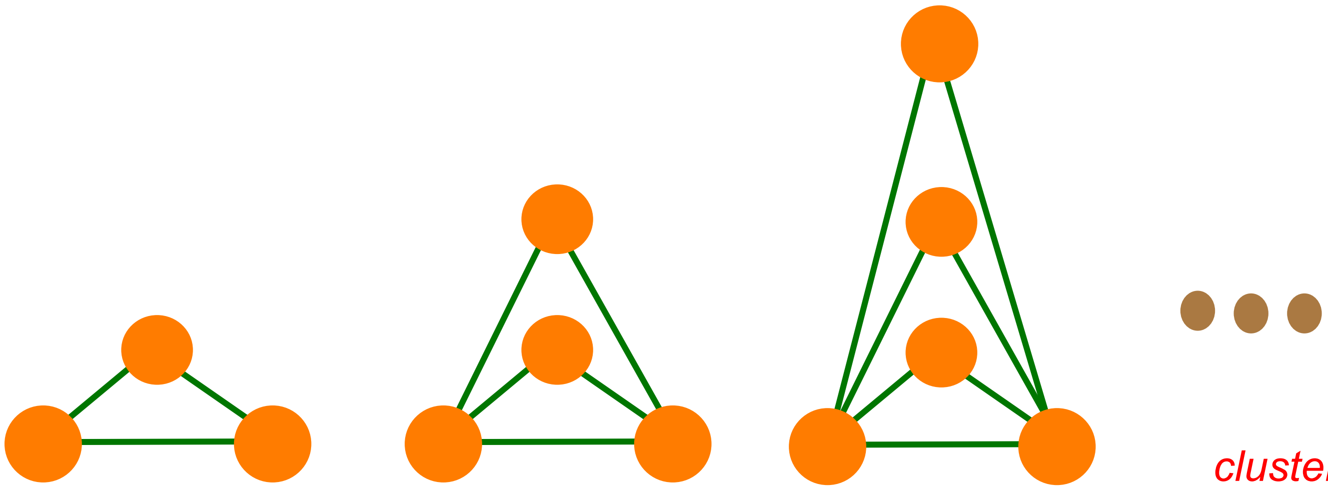
ERGMS – modelling graphs

If we believe that (Snijders, et al., 2006)



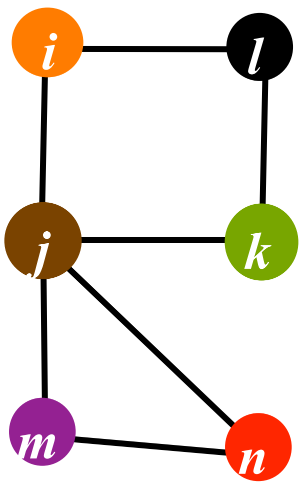
two edge indicators $\{i,k\}$ and $\{l,j\}$ are conditionally

dependent if $\{i,l\}$, $\{l,j\} \in E$

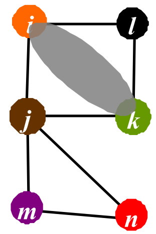


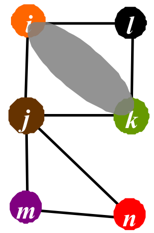
clustered regions

ERGMS – modelling graphs

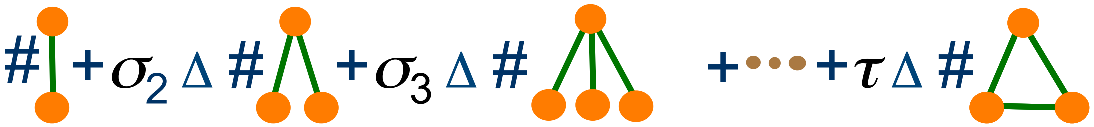


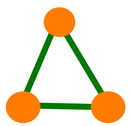
$$\log \frac{\Pr\ (i \text{ — } k) \text{ given the rest}}{\Pr\ (i \quad k) \text{ given the rest}}$$

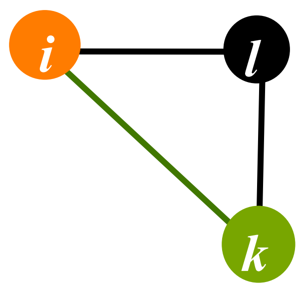


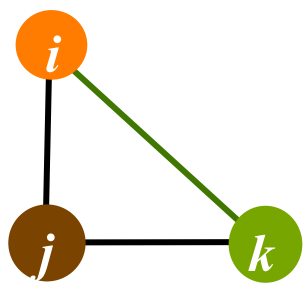


$$\approx \sigma_1 \Delta \# + \sigma_2 \Delta \# + \sigma_3 \Delta \# + \cdots + \tau \Delta \#$$



adding edge, e.g.: $+ \tau 2\times$  =





MANCHESTER 1824 The University of Manchester

ERGMS – modelling graphs

The conditional formulation

$$\log \frac{\Pr\left(i \text{ — } k \text{ given the rest}\right)}{\Pr\left(i \quad k \text{ given the rest}\right)}$$

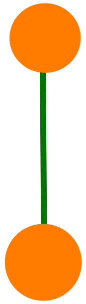
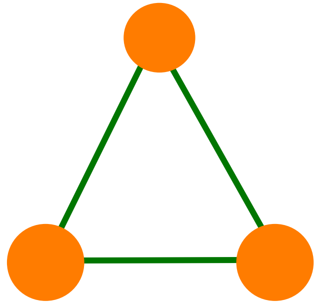
$$= \theta_1 \delta_{ik}^{1}(x) + \theta_2 \delta_{ik}^{2}(x) + \cdots + \theta_p \delta_{ik}^{p}(x)$$

where $\delta_{ik}^{r}(x) = z_r(\Delta_{ij} x) - z_r(x)$ Is the **difference** in **counts** of **structure type** k

May be ”aggregated” for all dyads so that the model for the entire adjacency matrix can be written

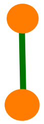
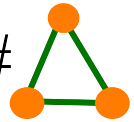
$$\log \Pr(X = x) = \theta_1 z_1(x) + \theta_2 z_2(x) + \cdots + \theta_p z_p(x) + \psi(\theta)$$

ERGMS – modelling graphs

For a model with edges  and triangles 

The model for the adjacency matrix X is a weighted sum

$$\log \Pr(X = x) = \sigma_1 L(x) + \tau T(x) + \psi(\theta)$$

where $L(x) = \#$  $T(x) = \#$ 

The parameters σ_1 and τ weight the relative importance of ties and triangles, respectively

- graphs with many triangles but not too dense are more probable than dense graphs with few triangles

ERGMS – modelling graphs: example

Padgett's Florentine families (Padgett and Ansell, 1993) network

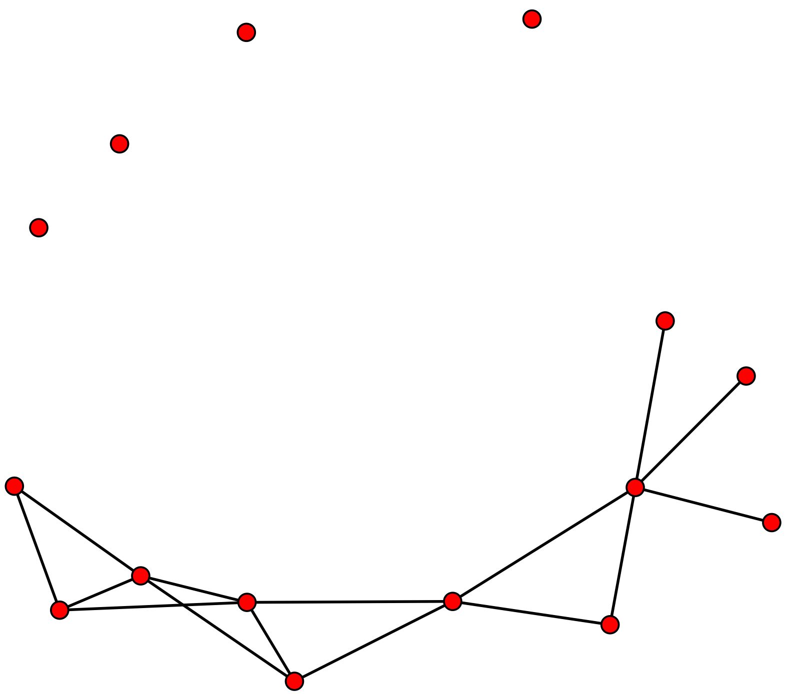
```
BusyNet <- as.matrix(read.table(
                   "PADGB.txt",header=FALSE))
```

```
> BusyNet
      V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16
 [1,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0
 [2,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0
 [3,]  0  0  0  0  1  1  0  0  1   0   1   0   0   0   0   0
 [4,]  0  0  0  0  0  0  1  1  0   0   1   0   0   0   0   0
 [5,]  0  0  1  0  0  0  0  1  0   0   1   0   0   0   0   0
 [6,]  0  0  1  0  0  0  0  0  1   0   0   0   0   0   0   0
 [7,]  0  0  0  1  0  0  0  1  0   0   0   0   0   0   0   0
 [8,]  0  0  0  1  1  0  1  0  0   0   1   0   0   0   0   0
 [9,]  0  0  1  0  0  1  0  0  0   1   0   0   0   1   0   1
[10,]  0  0  0  0  0  0  0  0  1   0   0   0   0   0   0   0
[11,]  0  0  1  1  1  0  0  1  0   0   0   0   0   0   0   0
[12,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0
[13,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0
[14,]  0  0  0  0  0  0  0  0  1   0   0   0   0   0   0   0
[15,]  0  0  0  0  0  0  0  0  0   0   0   0   0   0   0   0
[16,]  0  0  0  0  0  0  0  0  1   0   0   0   0   0   0   0
>
```

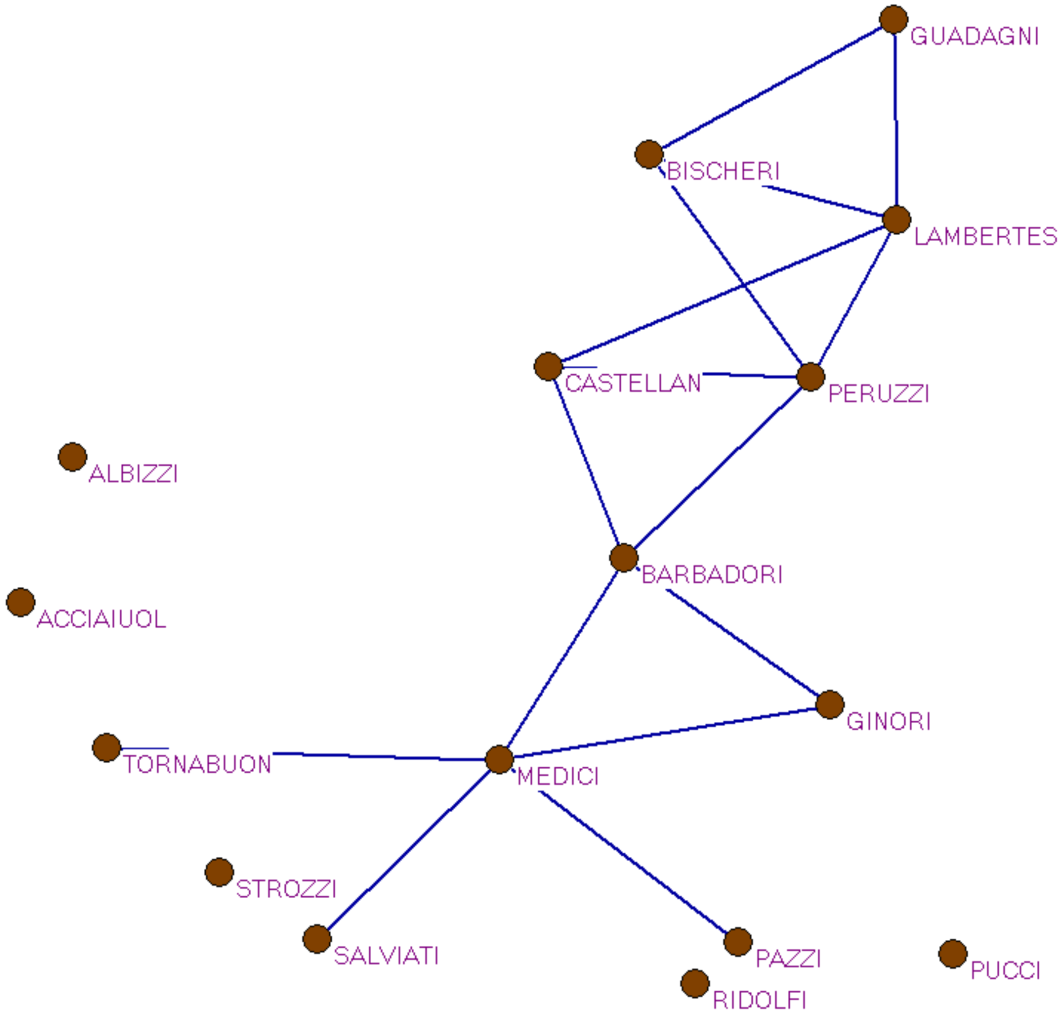
ERGMS – modelling graphs: example

```
BusyNetNet <- network(BusyNet, directed=FALSE)
plot(BusyNetNet)
```

Requires libraries
'sna','network'



ERGMS – modelling graphs: example



Observed

$$L(x) = \sum_{i<j} x_{ij} = 15$$

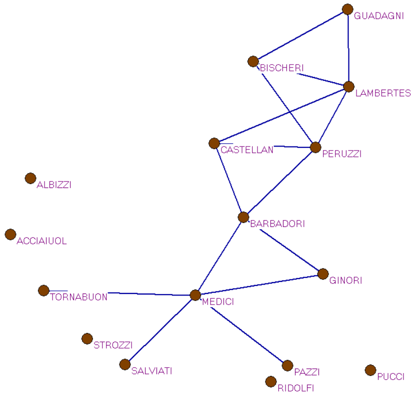
and

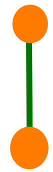
$$M = \binom{16}{2} = 120$$

Density:

$$d(x) = \frac{L(x)}{M} = \frac{15}{120} = \frac{1}{8} = 0.125$$

ERGMS – modelling graphs: example



For a model with only edges 

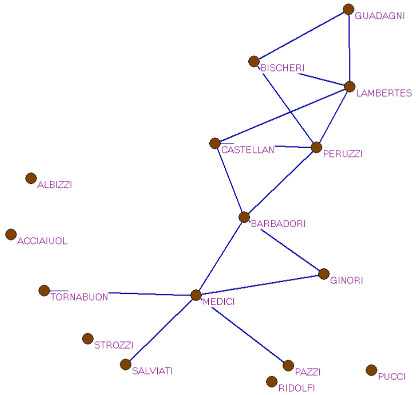
$$\log \Pr(X = x) = \sigma_1 L(x) + \psi(\sigma_1)$$

Equivalent with the model:

For each pair flip a p-coin

Where p is the probability coin comes up heads

ERGMS – modelling graphs: example



The (Maximul likelihood) estimate of p is the density

here:

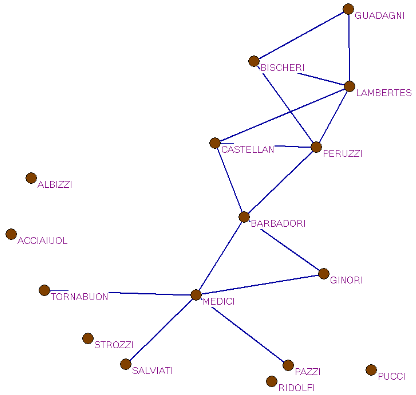
$$\hat{p}_{MLE} = d(x) = 0.125$$

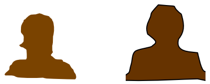
i.e., an estimated one in 8 pairs establish a tie

For an ERGM model with edges $\log \Pr(X = x) = \sigma_1 L(x) + \psi(\sigma_1)$

and hence $$p = \frac{e^{\sigma_1 L(x)}}{1 + e^{\sigma_1 L(x)}}$$

ERGMS – modelling graphs: example





$$\hat{p}_{MLE} = d(x) = 0.125$$

Solving

$$p = \frac{e^{\sigma_1 L(x)}}{1 + e^{\sigma_1 L(x)}}$$

for σ_1, we have that the density parameter

$$\sigma_1 = -\log(1/p - 1)$$

and for the MLE

$$\hat{\sigma}_{1,MLE} = -\log(1/\hat{p}_{MLE} - 1)$$

$$= -\log(8/1 - 1) = -1.945$$

Let's check in `stanet`

ERGMS – modelling graphs: example

```
Estim1 <- ergm(BusyNetNet ~ edges)
summary(Estim1)
```

```
==========================
Summary of model fit
==========================

Formula:   BusyNetNet ~ edges

Newton-Raphson iterations:  5

Maximum Likelihood Results:
      Estimate Std. Error MCMC s.e. p-value
edges   -1.946      0.276        NA  <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

For this model, the pseudolikelihood is the same as the likelihood.
```

approx. standard error of MLE of σ_1,

$\hat{p}_{MLE} = d(x) = 0.125$

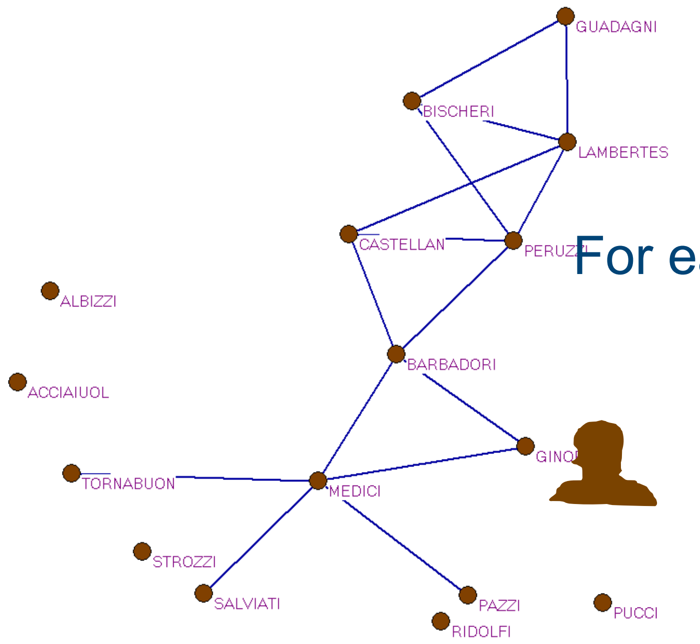
$\hat{\sigma}_{1,MLE}$

$\hat{\sigma}_{1,MLE} = -1.945$

Parameter corresponding to $L(x) = \#$

Success: $\boxed{\hat{\sigma}_{1,MLE} = -\log(1/\hat{p}_{MLE} - 1)}$

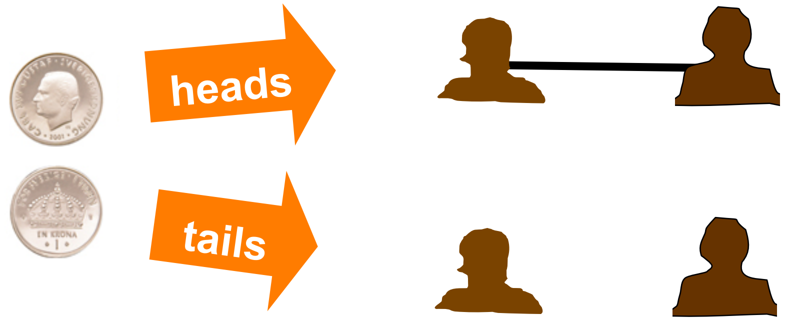
ERGMS – modelling graphs: example



Do we **believe** in the model:

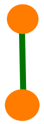
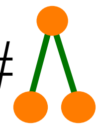
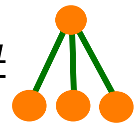
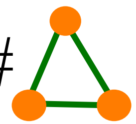
For each pair

flip a p-coin



Let’s fit a model that takes Markov dependencies into account

$$\log \Pr(X = x) = \sigma_1 L(x) + \sigma_2 S_2(x) + \sigma_3 S_3(x) + \tau T(x) + \psi(\theta)$$

where $L(x) = \#$  $S_2(x) = \#$  $S_3(x) = \#$  $T(x) = \#$ 

statnet

ERGMS – modelling graphs: example

$L(x) = \#$ (edge) $\quad S_2(x) = \#$ (2-star) $\quad S_3(x) = \#$ (3-star) $\quad T(x) = \#$ (triangle)

```
Estim2 <- ergm(BusyNetNet ~ kstar(1:3) + triangles)
summary(Estim2)
```

```
==========================
Summary of model fit
==========================

Formula:   BusyNetNet ~ kstar(1:3) + triangles

Newton-Raphson iterations:  42
MCMC sample of size 10000

Monte Carlo MLE Results:
         Estimate Std. Error MCMC s.e. p-value
kstar1    -1.6130     0.6699     0.462  0.0176 *
kstar2     0.7492     0.6407     0.455  0.2446
kstar3    -0.5408     0.3574     0.225  0.1330
triangle   1.4837     0.4592     0.138  0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

approx. standard error of MLE of τ,

$\hat{\tau}_{MLE}$

Part 2

Estimation

Likelihood equations for exponential fam

"Aggregated" to a joint model for **entire adjacency matrix** X

$$\log \Pr(X = x) = \theta_1 z_1(x) + \theta_2 z_2(x) + \cdots + \theta_p z_p(x) + \psi(\theta)$$

Sum over **all** $2^{n(n-1)/2}$ graphs

The MLE solves the equation (cf. Lehmann, 1983):

$$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

Likelihood equations for exponential fam

Solving $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

- Using the cumulant generating function (Corander, Dahmström, and Dahmström, 1998)
- Stochastic approximation (**Snijders, 2002**, based on **Robbins-Monro**, 1951)
- Importance sampling (**Handcock, 2003**; Hunter and Handcock, 2006, based on **Geyer-Thompson** 1992)

Robbins-Monro algorithm

Solving $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

Snijders, 2002, algorithm

- Initialisation phase
- Main estimation
- convergence check and cal. of standard errors

MAIN:

$$\theta^{(m+1)} = \theta^{(m)} - a_r D_0^{-1}\{z(x_{\theta^{(m)}}^{(m)}) - z(x_{\text{obs}})\}$$

Draw using MCMC

Robbins-Monro algorithm

Phase 1, **Initialisation phase**

Find good values of the initial parameter state

$$\theta^{(0)}$$

And the scaling matrix

$$D_0$$

(use the score-based method, Schweinberger & Snijders, 2006)

Robbins-Monro algorithm

Phase 2, Main estimation phase

Iteratively update θ

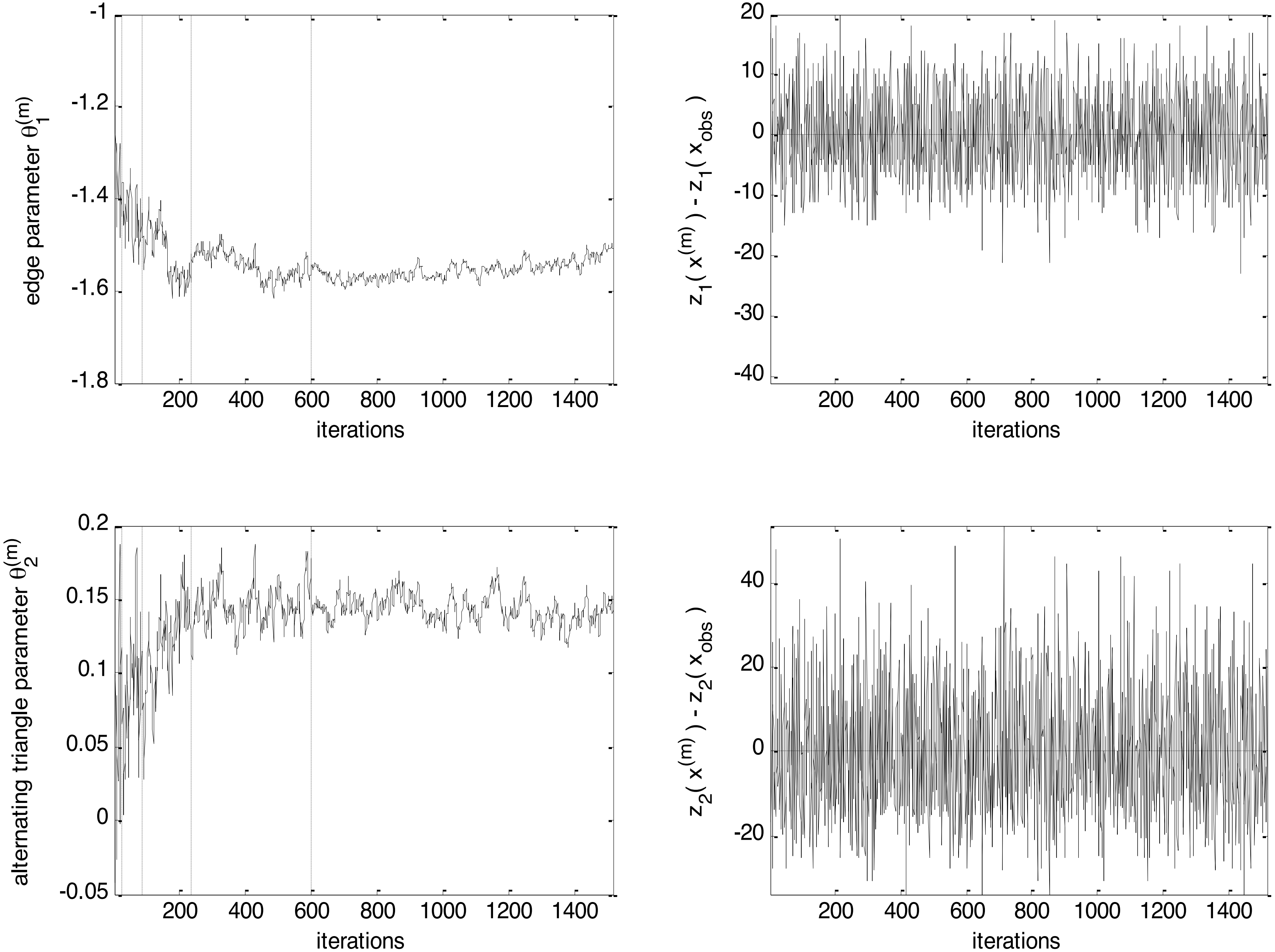
$$\theta^{(m+1)} = \theta^{(m)} - a_r D_0^{-1}\{z(x_{\theta^{(m)}}^{(m)}) - z(x_{\text{obs}})\}$$

by drawing one realisation

$$x_{\theta^{(m)}}^{(m)}$$

from the model defined by the current $\theta^{(m)}$

Repeated in **sub-phases** with fixed a_r



edge parameter $\theta_1^{(m)}$
-1
-1.2
-1.4
-1.6
-1.8
200 400 600 800 1000 1200 1400
iterations
$z_1(x^{(m)}) - z_1(x_{obs})$
20
10
0
-10
-20
-30
-40
200 400 600 800 1000 1200 1400
iterations
alternating triangle parameter $\theta_2^{(m)}$
0.2
0.15
0.1
0.05
0
-0.05
200 400 600 800 1000 1200 1400
iterations
$z_2(x^{(m)}) - z_2(x_{obs})$
40
20
0
-20
200 400 600 800 1000 1200 1400
iterations

Robbins-Monro algorithm

Phase 2, **Main estimation phase**

Relies on us being able to **draw** one realisation

$$x$$

from the **ERGM** defined by the current θ

We can **NOT** do this directly

We have to simulate x

More specifically use **M**arkov **c**hain **M**onte **C**arlo

Robbins-Monro algorithm

What do we need to know about MCMC?

Method:

Generate a sequence of graphs

$$x^{(0)}, x^{(1)}, x^{(2)}, x^{(3)}, \ldots$$

for arbitrary $x^{(0)}$, using an updating rule… so that

$$p(x^{(N)}) = \frac{e^{\theta^T z(x^{(N)})}}{\sum_y e^{\theta^T z(y)}} \quad \text{as} \quad N \to \infty$$

Robbins-Monro algorithm

What do we need to know about MCMC?

So if we generate an infinite number of graphs in the "right" way we have the ONE draw we need to update θ once?

Typically we can't wait an infinite amount of time so we settle for

$$N \rightarrow \text{very large}$$

multiplication factor

In Pnet very large is

$$\gamma \, \text{density}(x_{\text{obs}})[1 - \text{density}(x_{\text{obs}})]n^2$$

Robbins-Monro algorithm

Phase 3, Convergence check and calculating standard errors

At the end of phase 2 we always get a value

$$\hat{\theta}$$

But is it the MLE?

Does it satisfy

$$E_{\hat{\theta}}\{z(X)\} = z(x_{obs}) \qquad ?$$

Robbins-Monro algorithm

Phase 3, Convergence check and calculating standard errors

Phase 3 simulates a large number of graphs

And checks if

$$E_{\hat{\theta}}\{z(X)\} \approx \overline{E_{\hat{\theta}}\{z(X)\}} \approx z(x_{obs})$$

A minor discrepancy - due to numerical inaccuracy - is acceptable

Convergence statistics:

$$-.1 < \left| \frac{\overline{E_{\hat{\theta}}\{z(X)\}} - z(x_{obs})}{\overline{SD_{\hat{\theta}}\{z(X)\}}} \right| < .1$$

Geyer-Thompson

Solving $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

Handcock, 2003, approximate Fisher scoring

MAIN:

$$\theta^{(g)} = \theta^{(g-1)} - I(\theta^{(g-1)})^{-1}\left\{\sum_{m=1}^{M} w^{(m)} z(x^{(m)}) - z(x_{\text{obs}})\right\}$$

Approximated using importance sample from MCMC

Bayes: dealing with likelihood

The **normalising** constant **of the posterior** not essential for Bayesian inference, all we need is:

$$\pi(\theta \mid x) = \frac{\ell(\theta; x)\pi(\theta)}{\int \ell(\theta; x)\pi(\theta)\,\mathrm{d}\theta} \propto \ell(\theta; x)\pi(\theta)$$

… but

$$\ell(\theta; x) = \frac{\exp\{\sum_{k=1}^{p} \theta_k z_k(x)\}}{\sum_{y} \exp\{\sum_{k=1}^{p} \theta_k z_k(y)\}}$$

Sum over **all** $2^{n(n-1)/2}$ graphs

Bayes: MCMC?

Consequently,
in e.g. Metropolis-Hastings, acceptance probability of move to θ

$$\min\left\{1, \frac{\pi(\theta^* \mid x)}{\pi(\theta \mid x)} \frac{q_{prop}(\theta \mid \theta^*)}{q_{prop}(\theta^* \mid \theta)}\right\} = \min\left\{\frac{\ell(\theta^*; x)\pi(\theta^*)}{\ell(\theta; x)\pi(\theta)} \frac{q_{prop}(\theta \mid \theta^*)}{q_{prop}(\theta^* \mid \theta)}\right\}$$

… which contains

$$\frac{\sum_y \exp\{\sum_{k=1}^{p} \theta_k z_k(y)\}}{\sum_y \exp\{\sum_{k=1}^{p} \theta_k^* z_k(y)\}}$$

Bayes: Linked Importance Sampler Auxiliary Variable MCMC

LISA (Koskinen, 2008; Koskinen, Robins & Pattison, 2010): Based on Møller et al. (2006), we define an **auxiliary variable** ω

$$\omega \in \prod_{j=1}^{m} \mathscr{X}^{K} \times \{1,\ldots,K\} \times \{1,\ldots,K\}$$

And produce draws from the joint posterior

$$\pi(\omega,\theta \mid x_{obs}) \propto \frac{\exp\{\sum \theta_k z_k(x_{obs})\}}{\sum \exp\{\sum \theta_k z_k(y)\}} \frac{P^{B}_{\tau,\theta}(\omega)}{\sum \exp\{\sum \tau_k z_k(y)\}} \pi(\theta)$$

using the proposal distributions

$$\theta^{*} \mid \theta^{(t)} \sim N(\theta^{(t)},\Sigma) \qquad \text{and} \qquad \omega^{*} \mid \theta^{*} \sim \frac{P^{F}_{\theta^{*},\tau}(\omega^{*})}{\sum \exp\{\sum \theta^{*}_k z_k(y)\}}$$

Bayes: alternative auxiliary variable

LISA (Koskinen, 2008; Koskinen, Robins & Pattison, 2010): Based on Møller et al. (2006), we define an auxiliary variable ω

$$\omega \in \prod_{j=1}^{m} \mathscr{X}^{K} \times \{1,\ldots,K\} \times \{1,\ldots,K\}$$

Many linked chains:
- Computation time
- storage (memory and time issues)

Improvement: use **exchange algorithm** (Murray et al. 2006)

$$\theta^{*} \mid \theta^{(t)} \sim N(\theta^{(t)}, \Sigma) \qquad \text{and} \qquad x^{*} \mid \theta^{*} \sim \mathrm{ERGM}(\theta^{*})$$

Accept θ^* with log-probability: $\boxed{\min\{0, (\theta - \theta^*)^{\mathrm{T}} (z(x^*) - z(x_{\mathrm{obs}}))\}}$

Caimo & Friel, 2011

Bayes: Implications of using alternative auxiliary variable

- Storing only parameters
- No pre tuning – no need for good initial values
- Standard MCMC properties of sampler
- Less sensitive to near degeneracy in estimation
- Easier than anything else to implement

QUICK and **ROBUST**

Improvement: use **exchange algorithm** (Murray et al. 2006)

$$\theta^* \mid \theta^{(t)} \sim N(\theta^{(t)}, \Sigma) \qquad \text{and} \qquad x^* \mid \theta^* \sim \mathrm{ERGM}(\theta^*)$$

Accept θ^* with log-probability: $\boxed{\min\{0, (\theta - \theta^*)^{\mathrm{T}} (z(x^*) - z(x_{\mathrm{obs}}))\}}$

Caimo & Friel, 2011

Bayes: Implications of using alternative auxiliary variable

exchange algorithm (Murray et al. 2006)

auxiliary variables: $h(\theta^* \mid \theta)$

and $p(x^* \mid \theta^*) \sim \mathrm{ERGM}(\theta^*)$

Bayes: Implications of using alternative auxiliary variable

exchange algorithm (Murray et al. 2006)

auxiliary variables: $h(\theta^* \mid \theta)$

and $p(x^* \mid \theta^*) \sim \mathrm{ERGM}(\theta^*)$

To draw from joint posterior $\propto p(x^* \mid \theta^*) h(\theta^* \mid \theta) p(x \mid \theta) \pi(\theta) p(x \mid \theta)$

Bayes: Implications of using alternative auxiliary variable

exchange algorithm (Murray et al. 2006)

auxiliary variables: $h(\theta^* \mid \theta)$

and $p(x^* \mid \theta^*) \sim \mathrm{ERGM}(\theta^*)$

To draw from joint posterior $\propto p(x^* \mid \theta^*) h(\theta^* \mid \theta) p(x \mid \theta) \pi(\theta) p(x \mid \theta)$

Gibbs-draw: $(x^* \mid \theta^*) \sim p(x^* \mid \theta^*) h(\theta^* \mid \theta)$

Bayes: Implications of using alternative auxiliary variable

exchange algorithm (Murray et al. 2006)

auxiliary variables: $h(\theta^* \mid \theta)$

and $p(x^* \mid \theta^*) \sim \text{ERGM}(\theta^*)$

To draw from joint posterior $\propto p(x^* \mid \theta^*) h(\theta^* \mid \theta) p(x \mid \theta) \pi(\theta) p(x \mid \theta)$

Gibbs-draw: $(x^* \mid \theta^*) \sim p(x^* \mid \theta^*) h(\theta^* \mid \theta)$

then swap θ^* and θ with probability min{1,H}

$$\boxed{H = \frac{p(x_{\text{obs}} \mid \theta^*)}{p(x_{\text{obs}} \mid \theta)} \frac{\pi(\theta^*)}{\pi(\theta)} \frac{h(\theta \mid \theta^*) p(x^* \mid \theta)}{h(\theta^* \mid \theta) p(x^* \mid \theta^*)}}$$

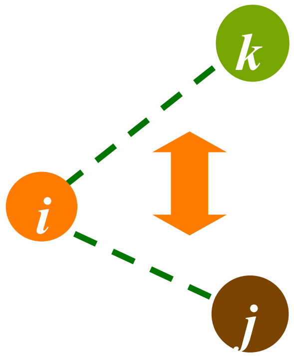
The University of Manchester

Part 3

Interpretation of effects

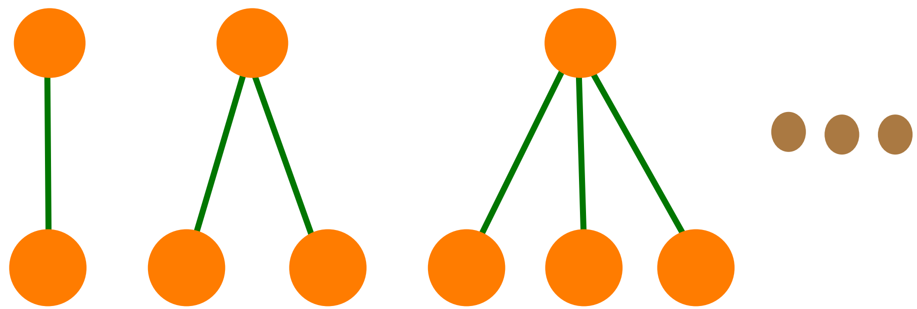
Problem with Markov models

Markov **dependence** assumption:

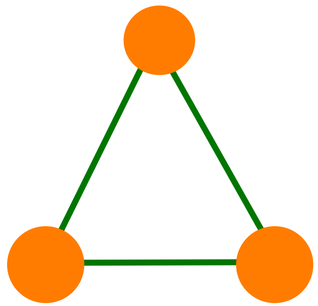


two edge indicators $\{i,j\}$ and $\{i',k\}$ are conditionally **dependent** if $\{i,j\} \cap \{i',k\} \neq \varnothing$

We have shown that the **only** effects are:



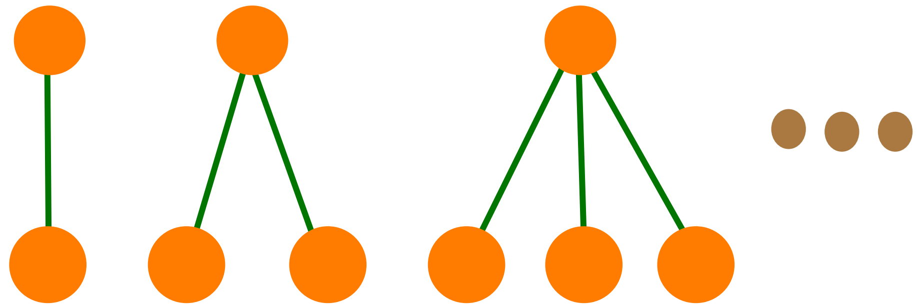
degree distribution; preferential attachment, etc



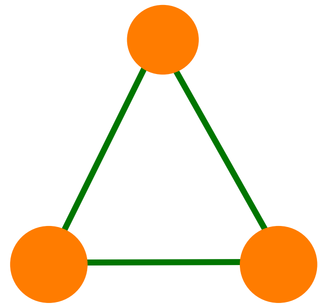
friends meet through friends; clustering; etc

Problem with Markov models

Often for **Markov model**



degree distribution; preferential attachment, etc



friends meet through friends; clustering; etc

Matching $$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

hard

or

Impossible!

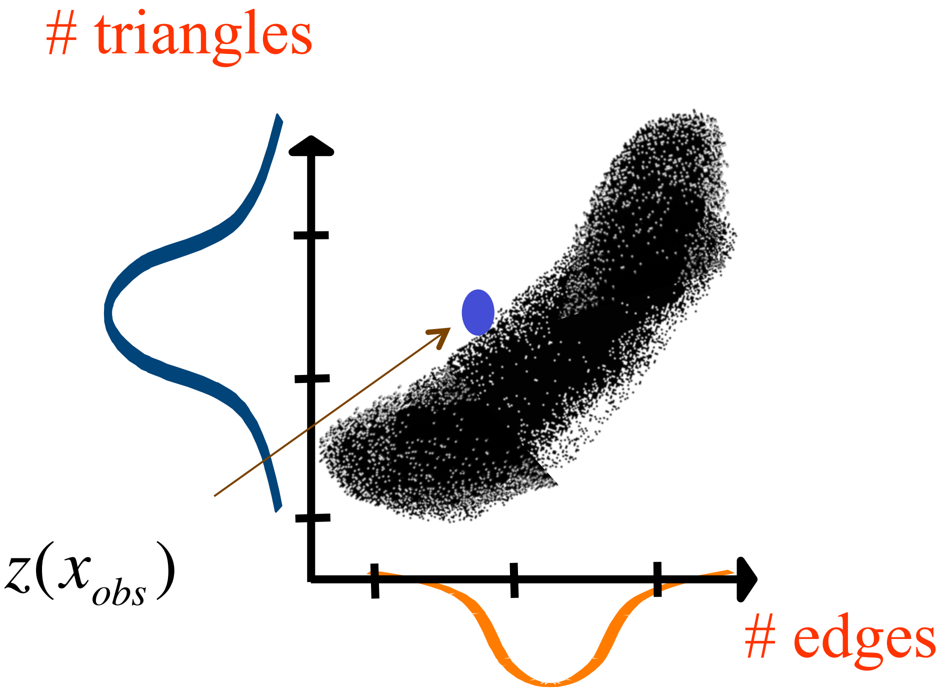
Problem with Markov models

Matching $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

hard

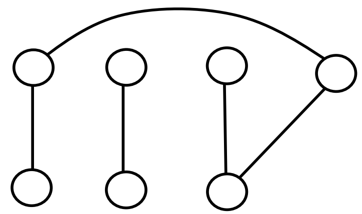
or

Impossible!



Some statistic k

$z_k(x_{obs}) = 0$ or $z_k(x_{obs}) = z_k^{\max}$

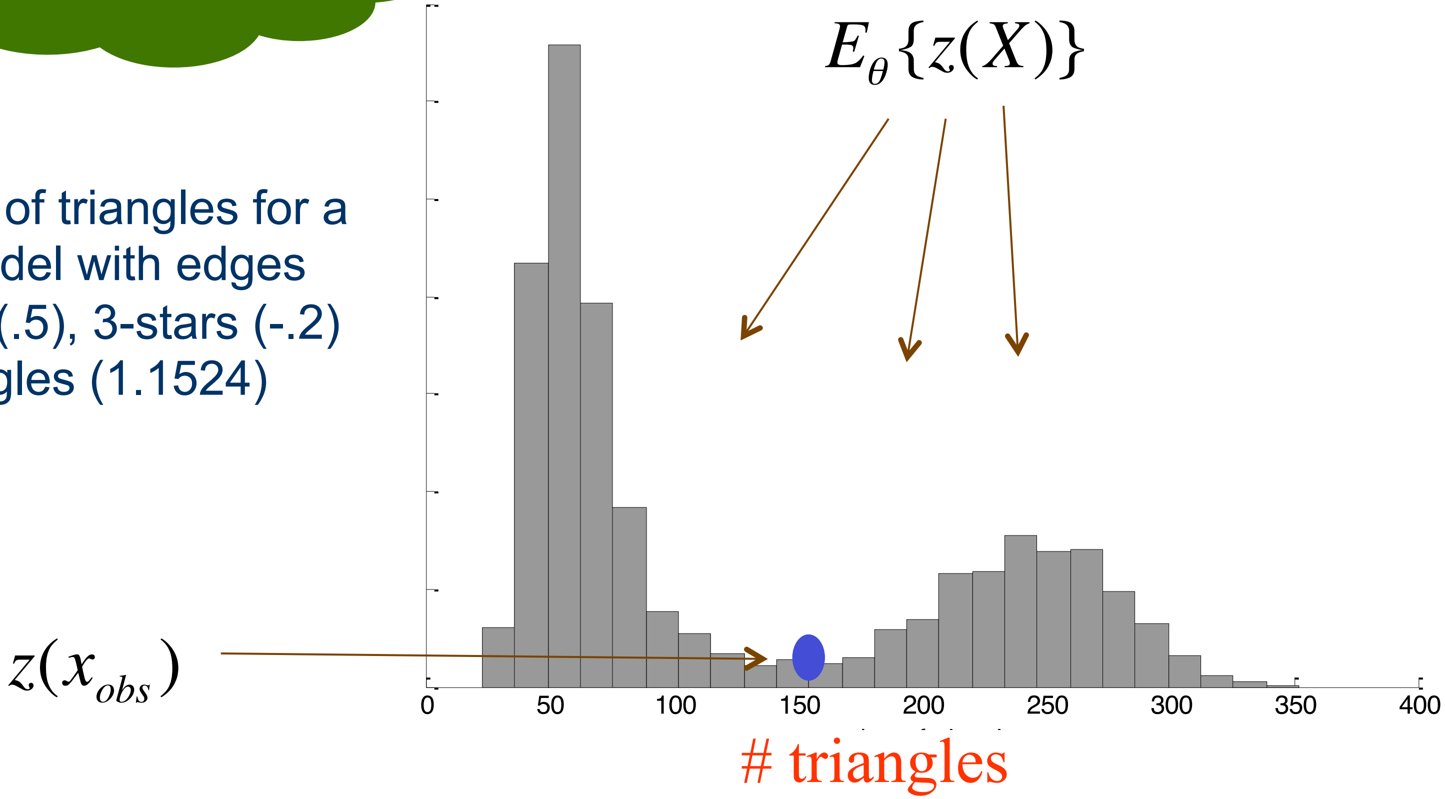


Problem with Markov models

Matching $$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

hard

The number of triangles for a Markov model with edges (-3), 2-stars (.5), 3-stars (-.2) and triangles (1.1524)



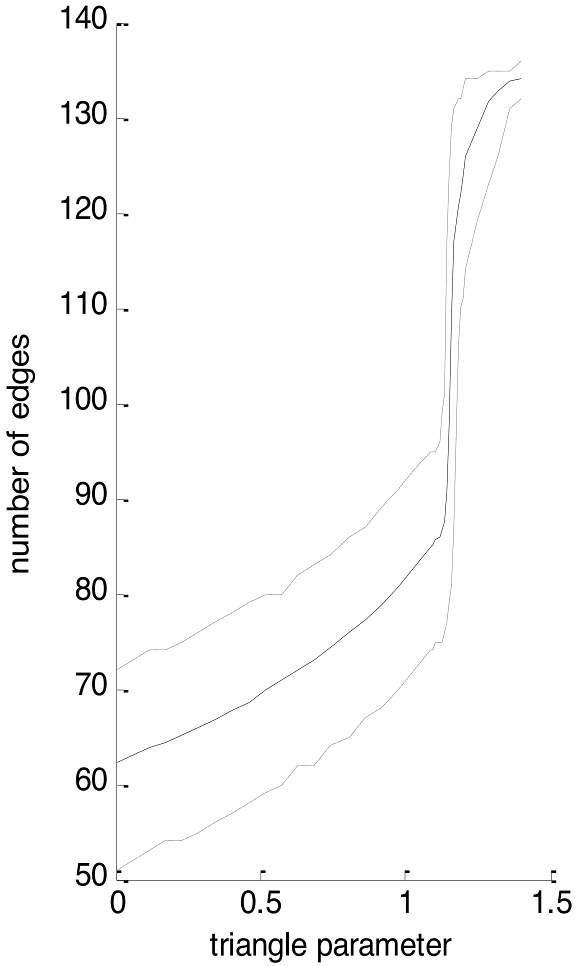
Problem with Markov models

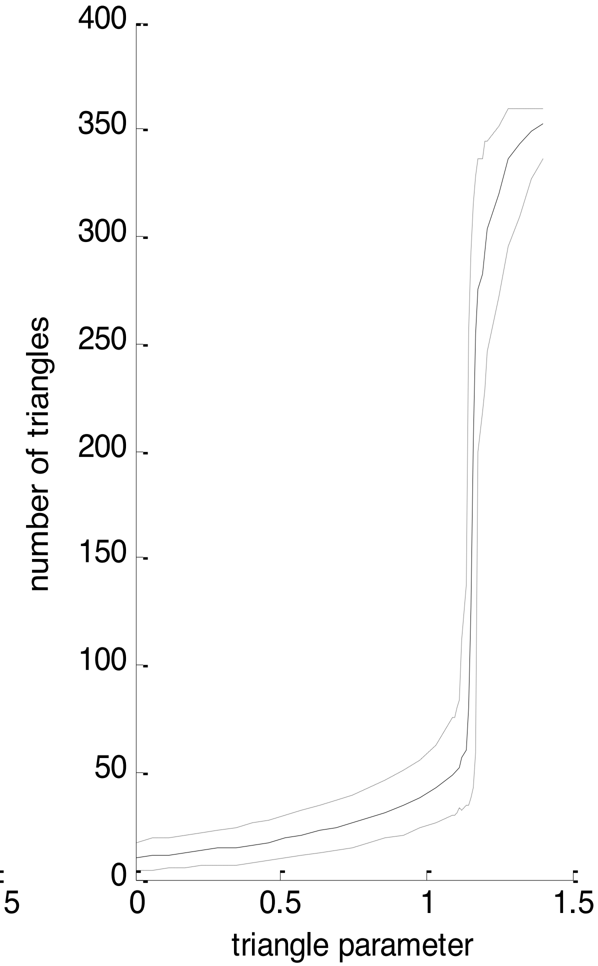
Matching $$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

hard

Exp. # triangles and edges for a Markov model with edges, 2-stars, 3-stars and triangles

$$E_{\theta}\{z(X)\}$$





Problem with Markov models

Impossible!

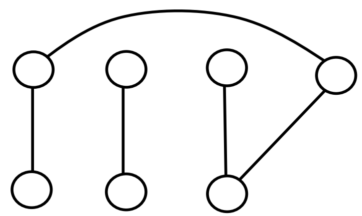
Matching $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

If for some statistic k $\quad z_k(x_{obs}) = 0$

$$E_\theta\{z(X)\} = \sum_y z_k(y)p(y \mid \theta) = z_k(x_{obs})$$

Implies: $p(y \mid \theta) = \begin{cases} 1 & \text{if } z_k(y) = z_k(x_{obs}) = 0 \\ 0 & \text{otherwise} \end{cases}$

Similarly for **max**
(or conditional min/max)
e.g. A graph on 7 vertices with 5 edges



\# edges

\# 2-stars

Problem with Markov models

Impossible!

Matching $E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$

Generally, let C be the convex hull of $\{z(x) : x \in \aleph\}$

Let rint(C) denote the relative interior of let C

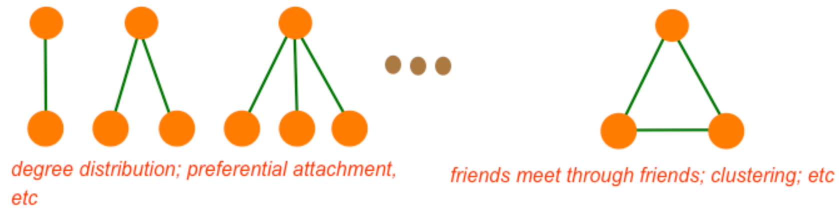
The MLE exists if and only if $z(x_{obs}) \in \text{rint}(C)$

With constant prior, the posterior exists only if $z(x_{obs}) \in \text{rint}(C)$

See Handcock (2003)

Problem with Markov models

First **solutions**
(Snijders et al., 2006)



Markov:

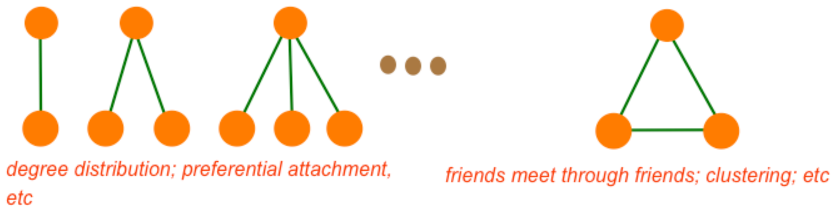
- adding k-star adds (k-1) stars
- alternating sign compensates but eventually $z_k(x)$=0

Alternating stars:

- Restriction on star parameters - Alternating sign
- prevents explosion, and
- models degree distribution

Problem with Markov models

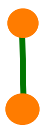
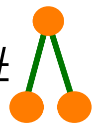
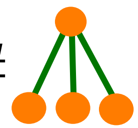
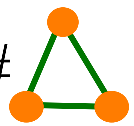
First **solutions**
(Snijders et al., 2006)



Markov:

- adding k-star adds $(k\text{-}1)$ stars
- alternating sign compensates but eventually $z_k(x)=0$

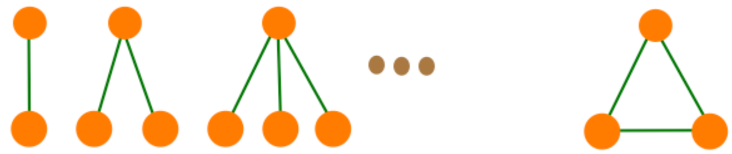
	Estimate	Std. Error	MCMC s.e.	p-value	
kstar1	-1.6130	0.6699	0.462	0.0176	*
kstar2	0.7492	0.6407	0.455	0.2446	
kstar3	-0.5408	0.3574	0.225	0.1330	
triangle	1.4837	0.4592	0.138	0.0016	**

$L(x) = \#$  $S_2(x) = \#$  $S_3(x) = \#$  $T(x) = \#$ 

MANCHESTER 1824
The University of Manchester

Problem with Markov models

First **solutions**
(Snijders et al., 2006)



degree distribution; preferential attachment, etc

friends meet through friends; clustering; etc

Include all stars but restrict parameter:

$$\sigma_3 = -\sigma_2 / \lambda \qquad \sigma_4 = -\sigma_3 / \lambda \qquad \sigma_5 = -\sigma_4 / \lambda \quad \cdots$$

new

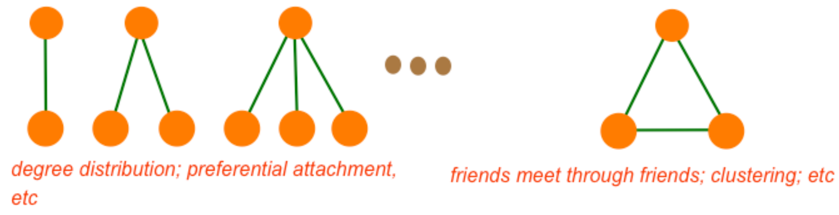
$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$

Alternating stars:

- Restriction on star parameters - Alternating sign
- prevents explosion, and
- models degree distribution

Problem with Markov models

First **solutions**
(Snijders et al., 2006)



Include all stars but restrict parameters:

$$\sigma_3 = -\sigma_2/\lambda \qquad \sigma_4 = -\sigma_3/\lambda \qquad \sigma_5 = -\sigma_4/\lambda \quad \cdots$$

new

$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$

Expressed in terms of degree distribution

$$= \left(\frac{1}{1-e^{-\alpha}}\right)^2 \sum_{j=0}^{n-1} d_j(x) e^{-\alpha j} + \frac{2L(x)}{1-e^{-\alpha}} - \frac{n}{(1-e^{-\alpha})^2}$$

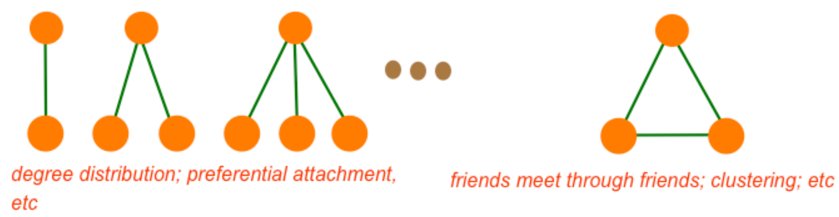
$$\boxed{d_j(x) = \#\{i : x_{i+} = j\}}$$

$$\lambda = e^{\alpha} / (e^{\alpha} - 1)$$

MANCHESTER 1824
The University of Manchester

Problem with Markov models

First **solutions**
(Snijders et al., 2006)



Include all stars but restrict parameters:

$$\sigma_3 = -\sigma_2 / \lambda \qquad \sigma_4 = -\sigma_3 / \lambda \qquad \sigma_5 = -\sigma_4 / \lambda \quad \cdots$$

$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$

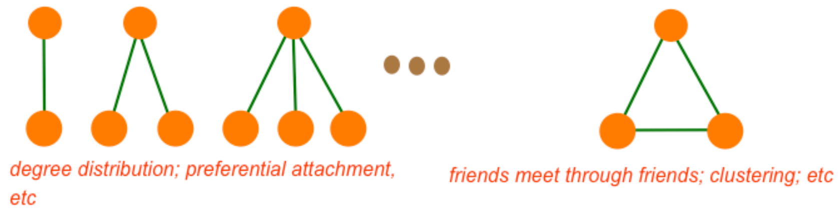
Interpretation:

Positive parameter ($\lambda \geq 1$) – graphs with some high degree nodes and larger degree variance more likely than graphs with more homogenous degree distribution

Negative parameter (λ ≥ 1) – the converse…

Problem with Markov models

First **solutions**
(Snijders et al., 2006)
Markov:



❖ triangles evenly spread out

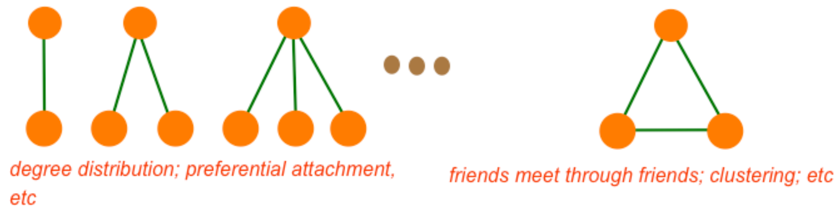
❖ but one edge can add many triangles…

Alternating triangles:

- Restrictions on different order triangles – alternating sign
- Prevents explosion, and
- Models multiply clustered regions
- **Social circuit dependence assumption**

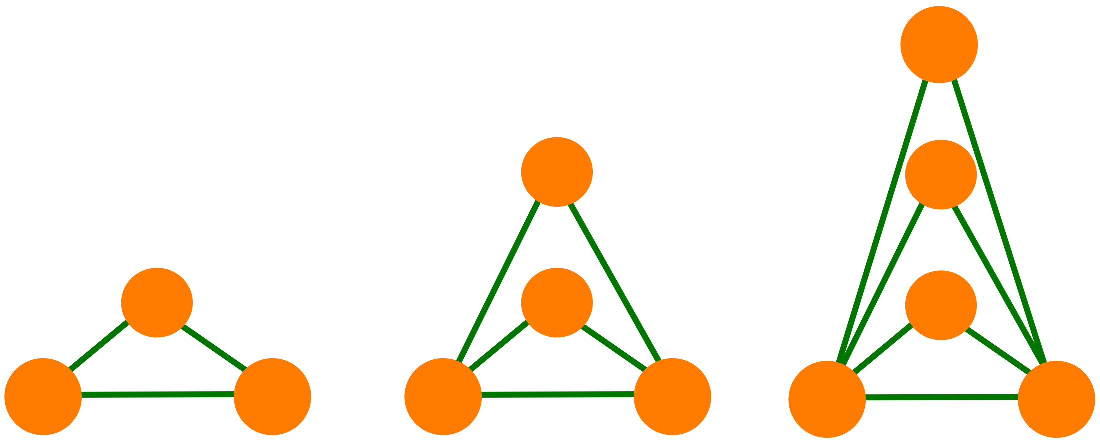
Problem with Markov models

First **solutions**
(Snijders et al., 2006)
Markov:



- ❖ triangles evenly spread out
- ❖ but one edge can add many triangles…

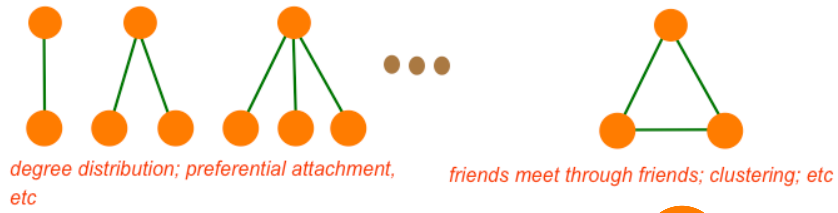
k-triangles:



1-triangles: 2-triangles: 3-triangles:

Problem with Markov models

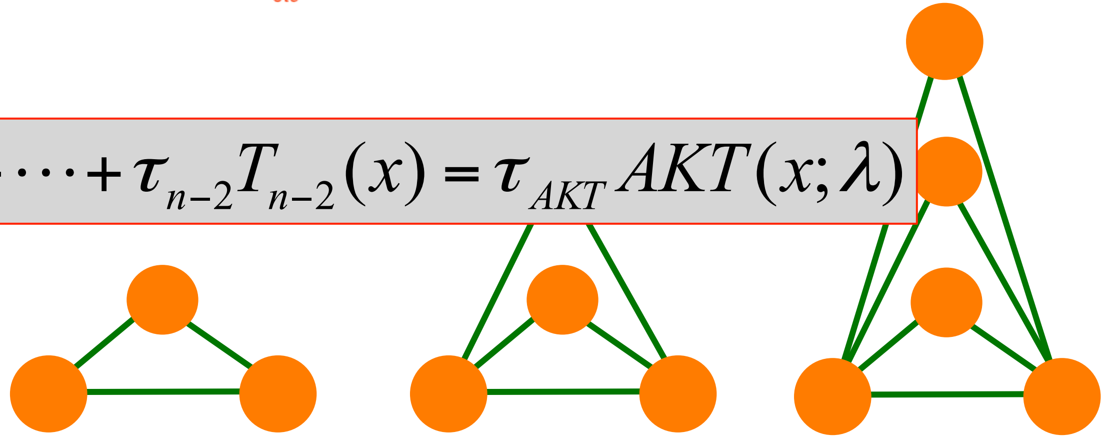
First **solutions**



Weigh the k-triangles:

$$\tau_2 T_2(x) + \tau_3 T_3(x) + \cdots + \tau_{n-2} T_{n-2}(x) = \tau_{AKT}\, AKT(x;\lambda)$$

Where: $\tau_k = -\tau_{k-1}/\lambda$



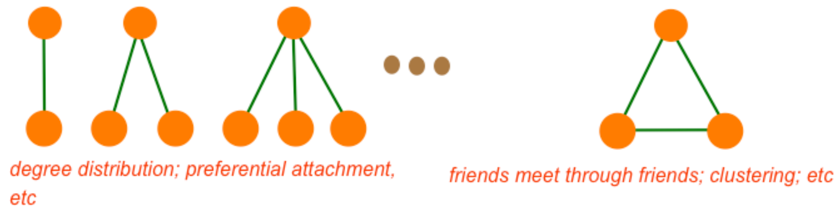
Alternating triangles:

- Restrictions on different order triangles – alternating sign
- Prevents explosion, and
- Models multiply clustered regions

MANCHESTER 1824
The University of Manchester

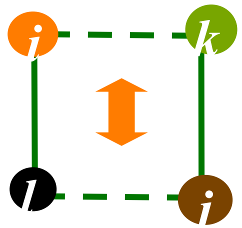
Problem with Markov models

First **solutions**



Underlying assumption:

Social circuit dependence assumption



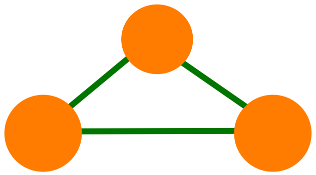
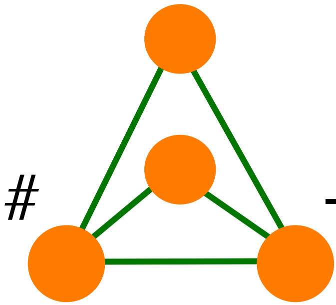
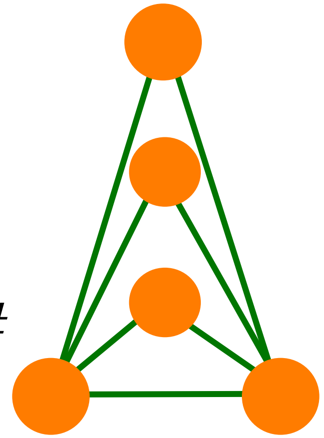
two edge indicators $\{i,k\}$ and $\{l,j\}$ are conditionally **dependent** if $\{i,l\}$, $\{l,j\} \in E$

Alternating triangles:

- Restrictions on different order triangles – alternating sign
- Prevents explosion, and
- Models multiply clustered regions

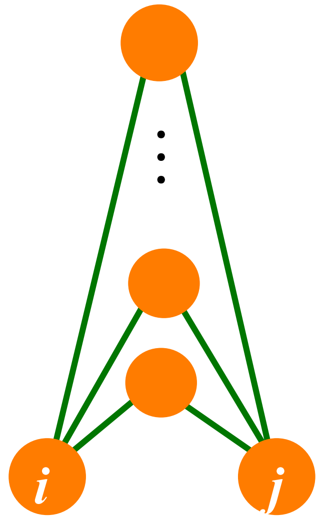
Problem with Markov models

Alternative interpretation

$\tau_1 \times \#$  $-\tau_1 / \lambda \times \#$  $-\tau_2 / \lambda \times \#$  $\cdots$

We may (geometrically) weight together :

$$z_T(x;\lambda) = \frac{e^{\alpha}}{e^{\alpha}-1}\left\{\sum_{i<j} x_{ij} - \sum_{i<j} x_{ij}\frac{1}{e^{\alpha S_{2ij}(x)}}\right\}$$

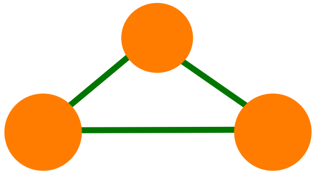
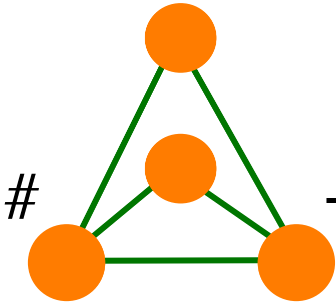
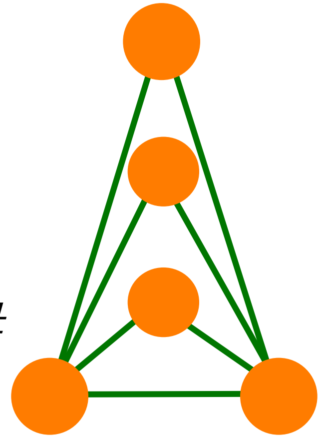


$$\lambda = e^{\alpha} / (e^{\alpha}-1)$$

$$S_{2ij} = \#\{k : i \to k, j \to k\}$$

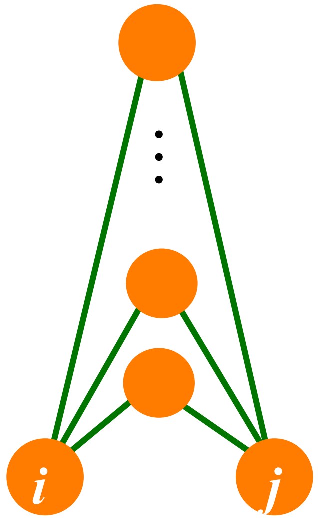
Problem with Markov models

Alternative intrepretation

$\tau_1 \times \#$  $-\tau_1 / \lambda \times \#$  $-\tau_2 / \lambda \times \#$  $\cdots$

We may also define the
Edgewise **S**hared **P**artner Statistic:

$$ESP_k = \#\{(i,j) : i \sim j, S_{2ij} = k\}$$



… and we can weigh together the **ESP** statistics using **G**eometrically decreasing weights: **GWESP**

Part 3b

Curved exponential family distributions for graphs

Curved ERGM

In an ERGM, alternating statistics

alternating stars

$$z_S(x;\alpha) = \left(\frac{1}{1-e^{-\alpha}}\right)^2 \sum_{j=0}^{n-1} d_j(x)e^{-\alpha j} + \frac{2L(x)}{1-e^{-\alpha}} - \frac{n}{(1-e^{-\alpha})^2}$$

alternating triangles

$$z_T(x;\alpha) = \frac{e^{\alpha}}{e^{\alpha}-1}\left\{\sum_{i<j} x_{ij} - \sum_{i<j} x_{ij}\frac{1}{e^{\alpha S_{2ij}(x)}}\right\}$$

… are “dampened” by a constant α

why not estimate α?

Curved ERGM

If we treat α as free parameters to be estimated

$$p(x) = \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}$$

Curved ERGM

If we treat α as free parameters to be estimated

$$p(x) = \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}$$

We have **more** statistics than parameters

… it is no longer an exponential family distribution

Curved ERGM

If we treat α as free parameters to be estimated

$$p(x) = \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}$$

We have **more** statistics than parameters

… it is no longer an exponential family distribution

For example, we no longer have the identity

$$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

$z(x;\alpha)$

Curved ERGM

If we treat α as free parameters to be estimated

$$p(x) = \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}$$

We have **more** statistics than parameters

… it is no longer an exponential family distribution

However, does not matter for Bayesian analysis

$$\pi(\theta,\alpha \mid x) \propto \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}\pi(\theta,\alpha)$$

Curved ERGM

If we treat α as free parameters to be estimated

$$p(x) = \exp\{\theta^T z(x;\alpha) - \psi(\theta,\alpha)\}$$

We have **more** statistics than parameters

… it is no longer an exponential family distribution

Formally it is a Curved exponential family distribution

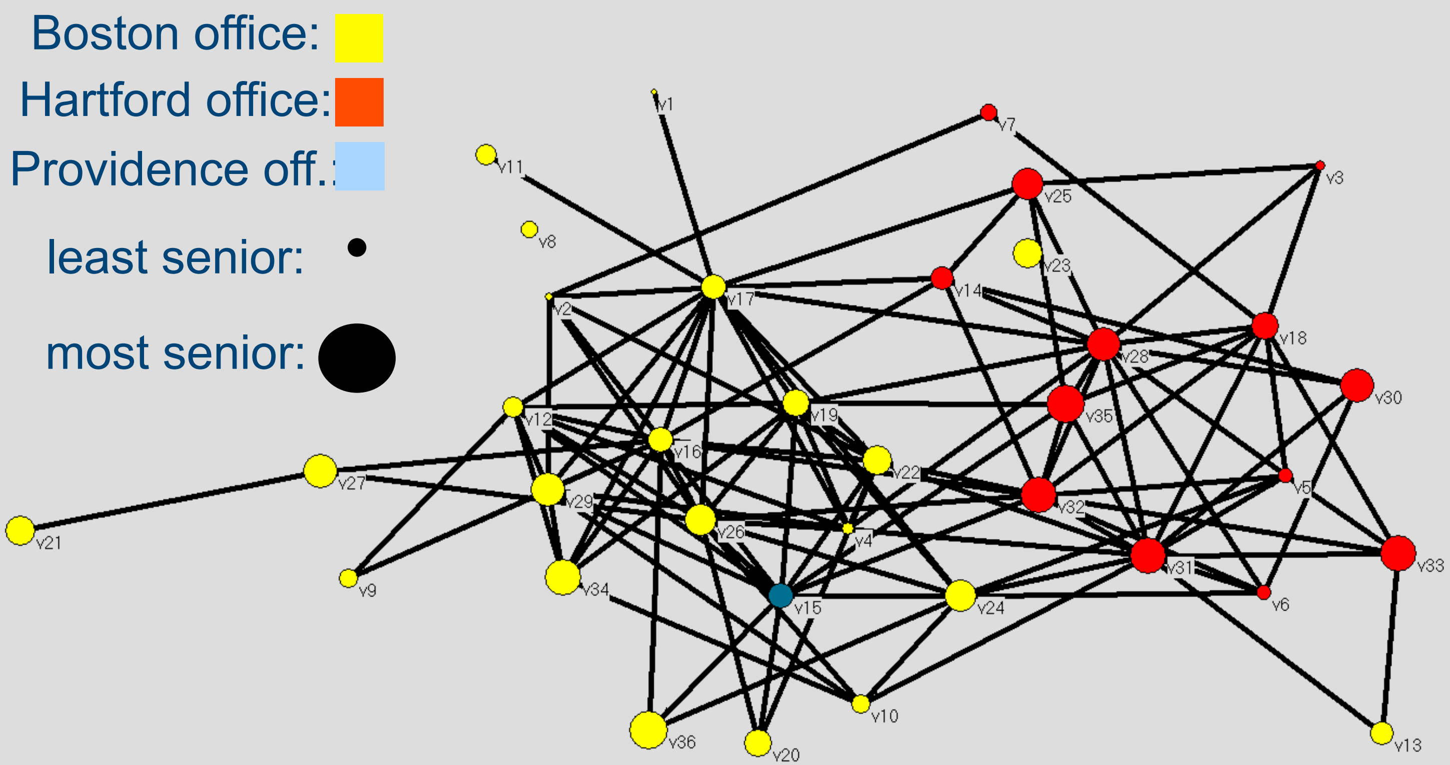
… and a Fisher scoring algoithm (using MCMC) can be applied (Hunter and Handcock, 2006)

Part 4a

Example Lazega's law firm partners

Lazega's (2001) Lawyers

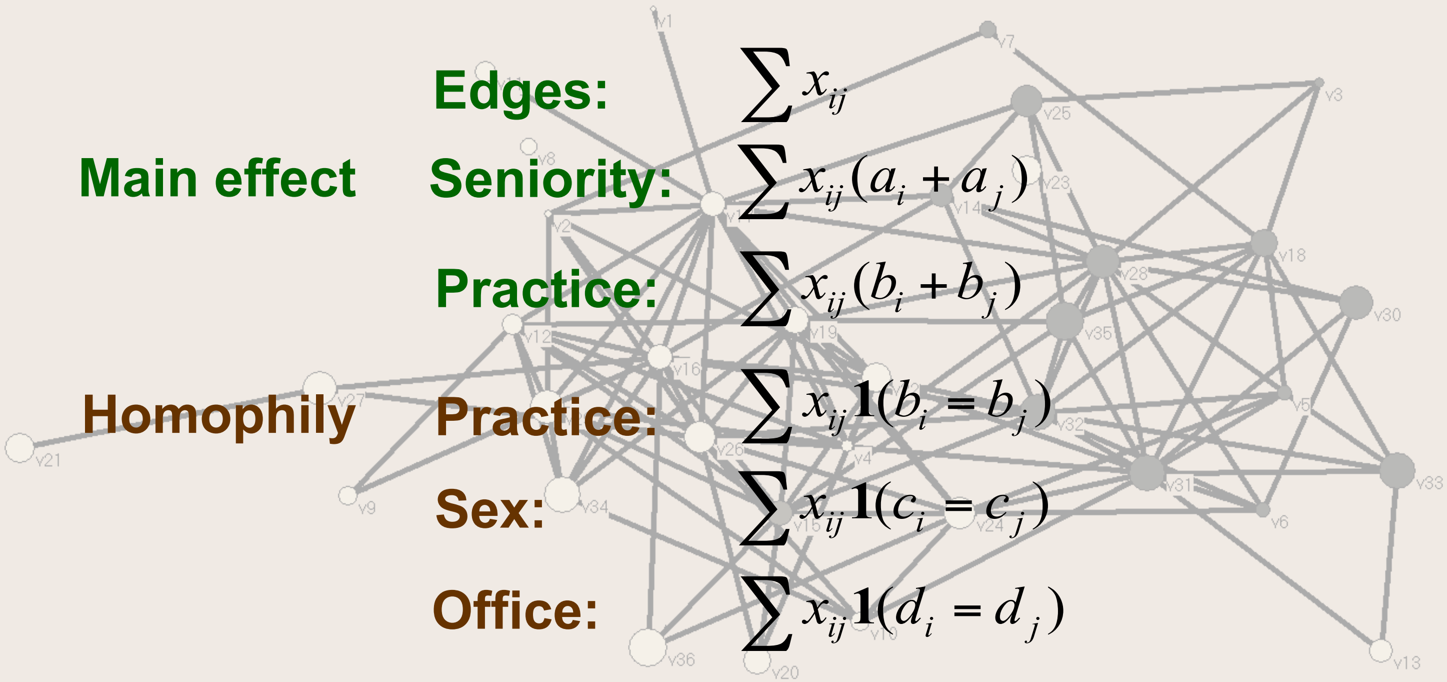
Collaboration network among 36 lawyers in a New England law firm (Lazega, 2001)



Lazega's (2001) Lawyers

Fit a model with "**new specifications**" and **covariates**

$$\log \Pr(X = x) = \theta_1 z_1(x) + \theta_2 z_2(x) + \cdots + \theta_p z_p(x) + \psi(\theta)$$



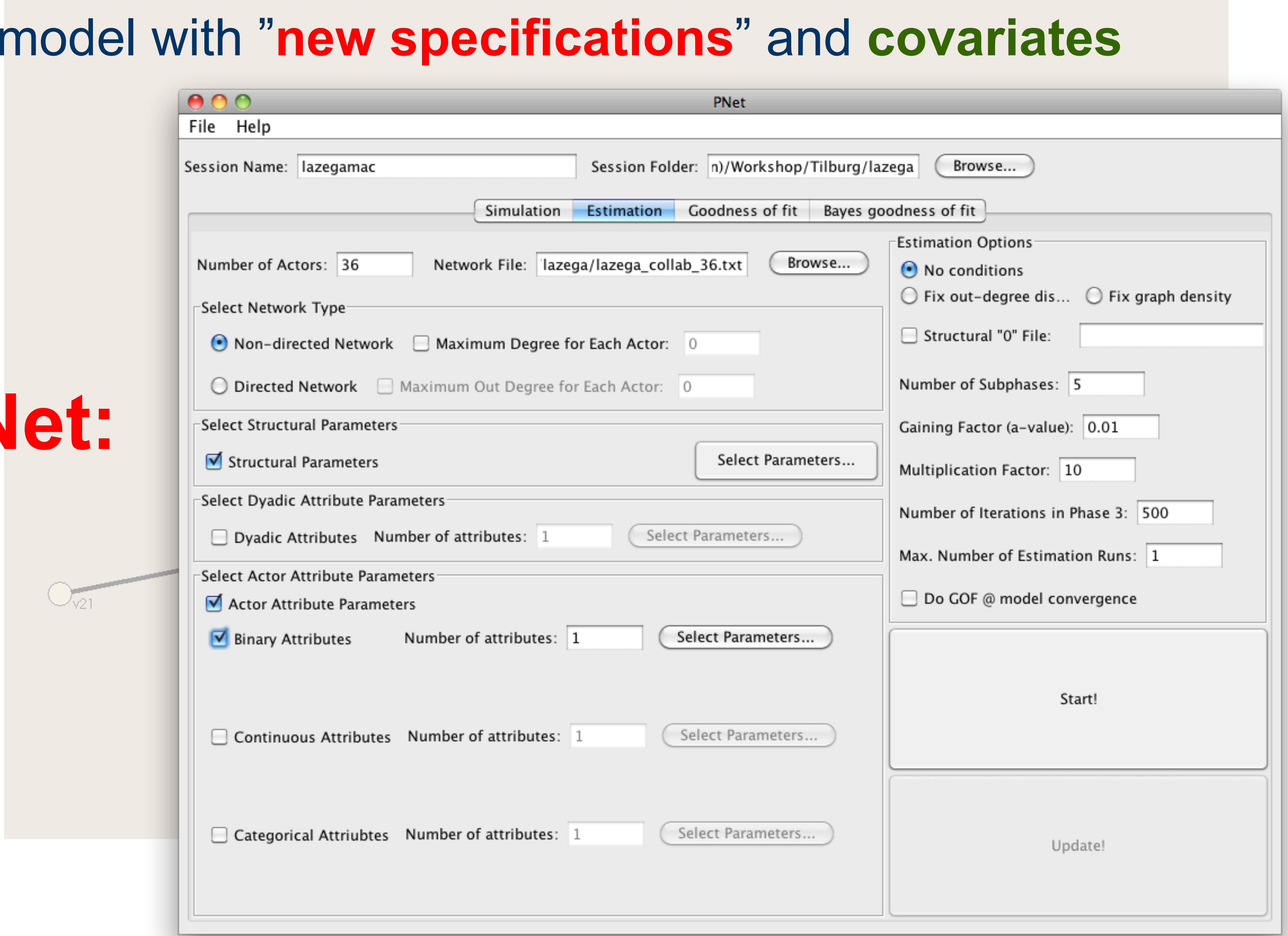
	Edges:	$\sum x_{ij}$
Main effect	**Seniority:**	$\sum x_{ij}(a_i + a_j)$
	Practice:	$\sum x_{ij}(b_i + b_j)$
Homophily	**Practice:**	$\sum x_{ij}\mathbf{1}(b_i = b_j)$
	Sex:	$\sum x_{ij}\mathbf{1}(c_i = c_j)$
	Office:	$\sum x_{ij}\mathbf{1}(d_i = d_j)$

GWESP: $$3t_1(x) - \frac{t_2(x)}{\lambda^1} + \cdots + (-1)^{n-3}\frac{t_{n-2}(x)}{\lambda^{n-3}}$$

Lazega's (2001) Lawyers

Fit a model with "**new specifications**" and **covariates**

PNet:



Lazega's (2001) Lawyers

Fit a model with "**new specifications**" and **covariates**

Main effect

Edges: $\sum x_{ij}$

Seniority: $\sum x_{ij}(a_i + a_j)$

Practice: $\sum x_{ij}(b_i + b_j)$

Homophily

Practice: $\sum x_{ij}\mathbf{1}(b_i = b_j)$

Sex: $\sum x_{ij}\mathbf{1}(c_i = c_j)$

Office: $\sum x_{ij}\mathbf{1}(d_i = d_j)$

estimation_lazega.txt

```
*************************************
mean statistics in phase3:114.362000    179.720562      42.846000
128.654000      129.287641      98.156000       84.984000

Estimation Result for Network SUMMARY (parameter, standard error, t-
statistics)
NOTE: t-statistics = (observation - sample mean)/standard error
effects                 estimates       stderr  t-ratio
edge                    -5.862515       0.56404 0.04105 *
AT(2.00)                1.011721        0.17095 0.05003 *
practice_interaction    1.499409        0.40322 0.02371 *
practice_activity       -0.331023       0.21995 0.02142
senior_sum              0.842661        0.23348 0.04943 *
sex_matching            0.702477        0.26389 0.05839 *
off_matching            1.145290        0.19749 0.00134 *

Estimated Covariance Matrix
```

GWESP: $3t_1(x) - \frac{t_2(x)}{\lambda^1} + \cdots + (-1)^{n-3}\frac{t_{n-2}(x)}{\lambda^{n-3}}$

Part 4b

Interpreting attribute-related effects

Fitting an ERGM in Pnet: a business communications network

For ”wwbusiness.txt” we have recorded wheather the employee works in the central office or is a traveling sales represenative

	office	sales	
office	interaction for attribute		activity for attribute
sales			

Fitting an ERGM in Pnet: a business communications network

Consider a dyad-independent model

$$\log \Pr(X_{ij} = x_{ij}) = \sigma_1 x_{ij} + \theta_R x_{ij}(OFF_i + OFF_j) + \theta_{Rb} x_{ij} OFF_i OFF_j + \psi(\theta)$$

	office	sales	
office	interaction for attribute		activity for attribute
sales			

Fitting an ERGM in Pnet: a business communications network

With log odds

$$\log \frac{\Pr(X_{ij}=1)}{\Pr(X_{ij}=0)} = \sigma_1 + \theta_R (OFF_i + OFF_j) + \theta_{Rb} OFF_i OFF_j$$

$\log \frac{\Pr(X_{ij}=1)}{\Pr(X_{ij}=0)}$	office $OFF_j = 1$	sales $OFF_j = 0$	
office $OFF_i = 1$	$\sigma_1 + \theta_R 2 + \theta_{Rb}$ interaction for attribute	$\sigma_1 + \theta_R$	activity for attribute
sales $OFF_i = 0$	$\sigma_1 + \theta_R$	σ_1	

Part 4c

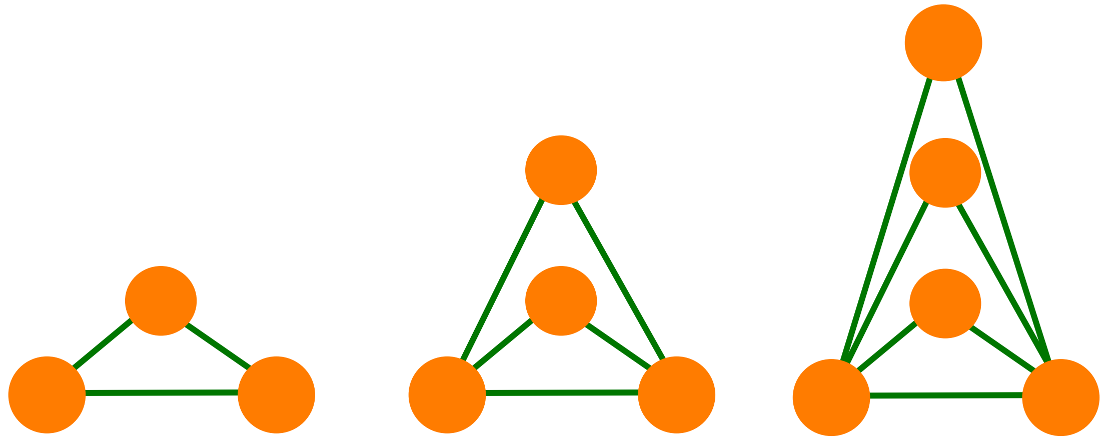
Interpreting higher order effects

Unpacking the alternating star effect

Alternating stars a way of

- "**fixing**" the Markov problems (models all degrees)
- **Controlling** for paths in clustering

$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$



1-triangles: 2-triangles: 3-triangles:

k-triangles measure clustering…

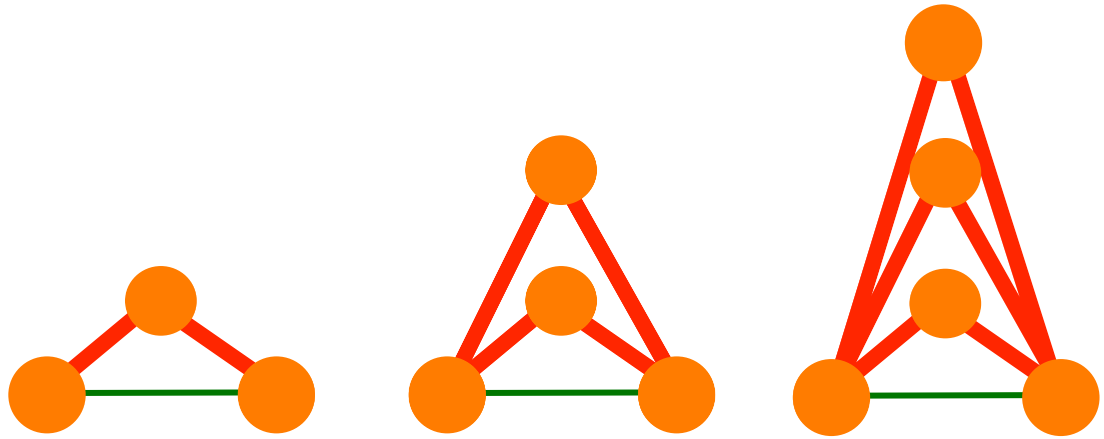
MANCHESTER 1824 The University of Manchester

Unpacking the alternating star effect

Alternating stars a way of

- "**fixing**" the Markov problems (models all degrees)
- **Controlling** for paths in clustering

$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$

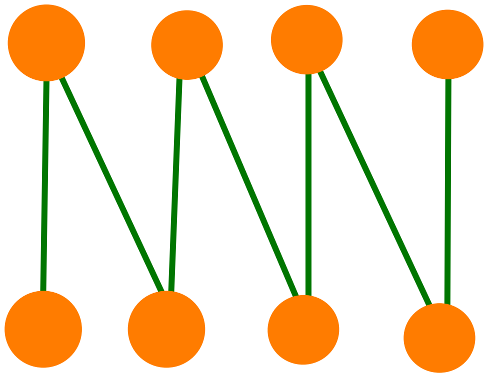


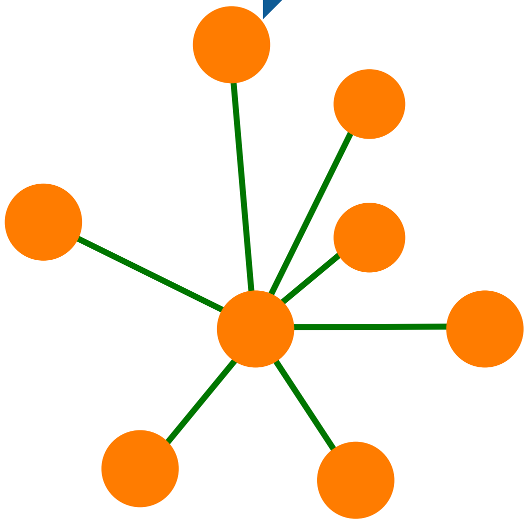
1-triangles: 2-triangles: 3-triangles:

Is it **closure** or an **artefact** of many stars/2-paths?

Unpacking the alternating star effect

Interpreting the alternating star **parameter**:





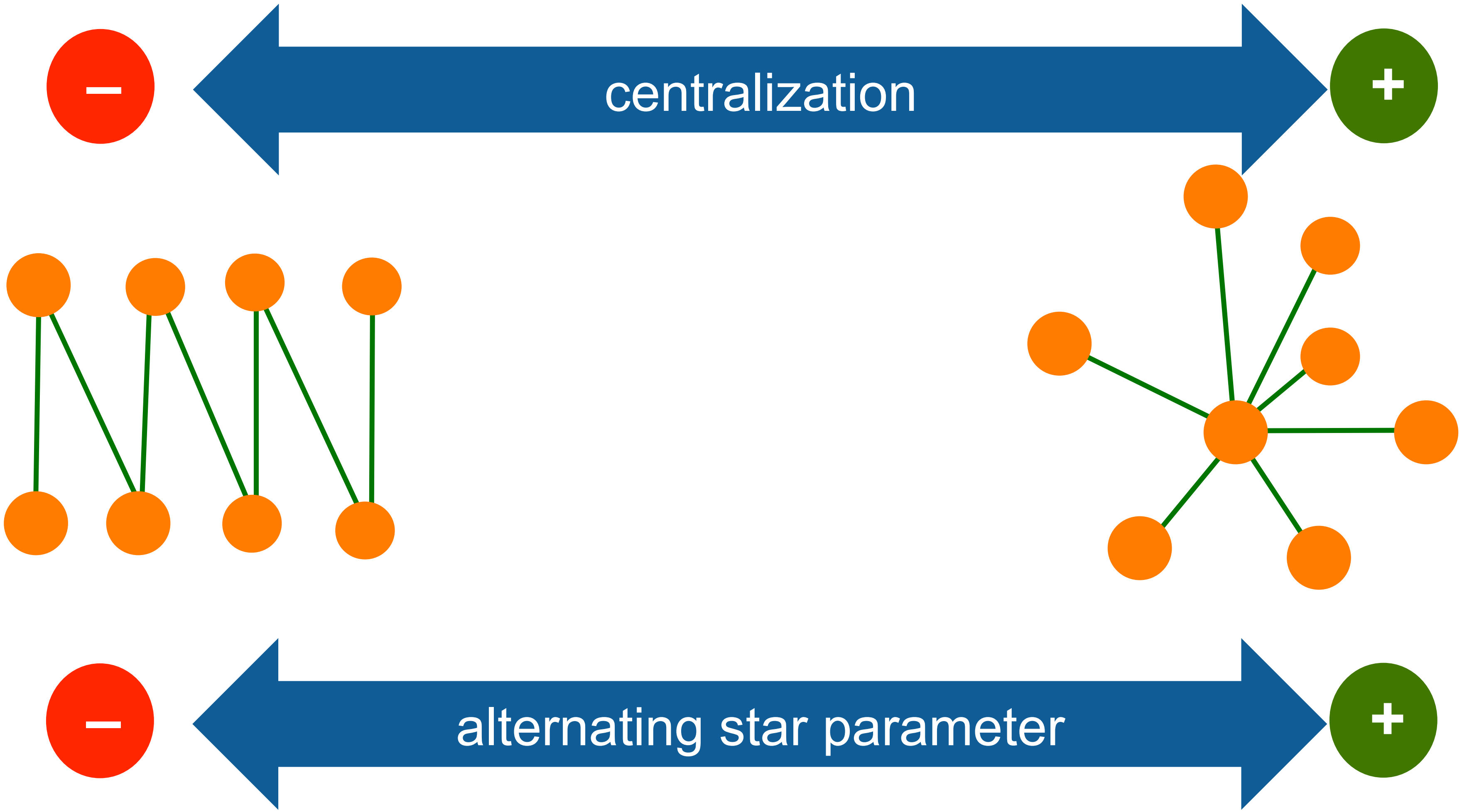
$$Var(x_{i+}) = \sum_i \frac{(x_{i+} - \bar{x})^2}{n-1} = .21$$

$$Var(x_{i+}) = 4.42$$

variance of degree measure centralization

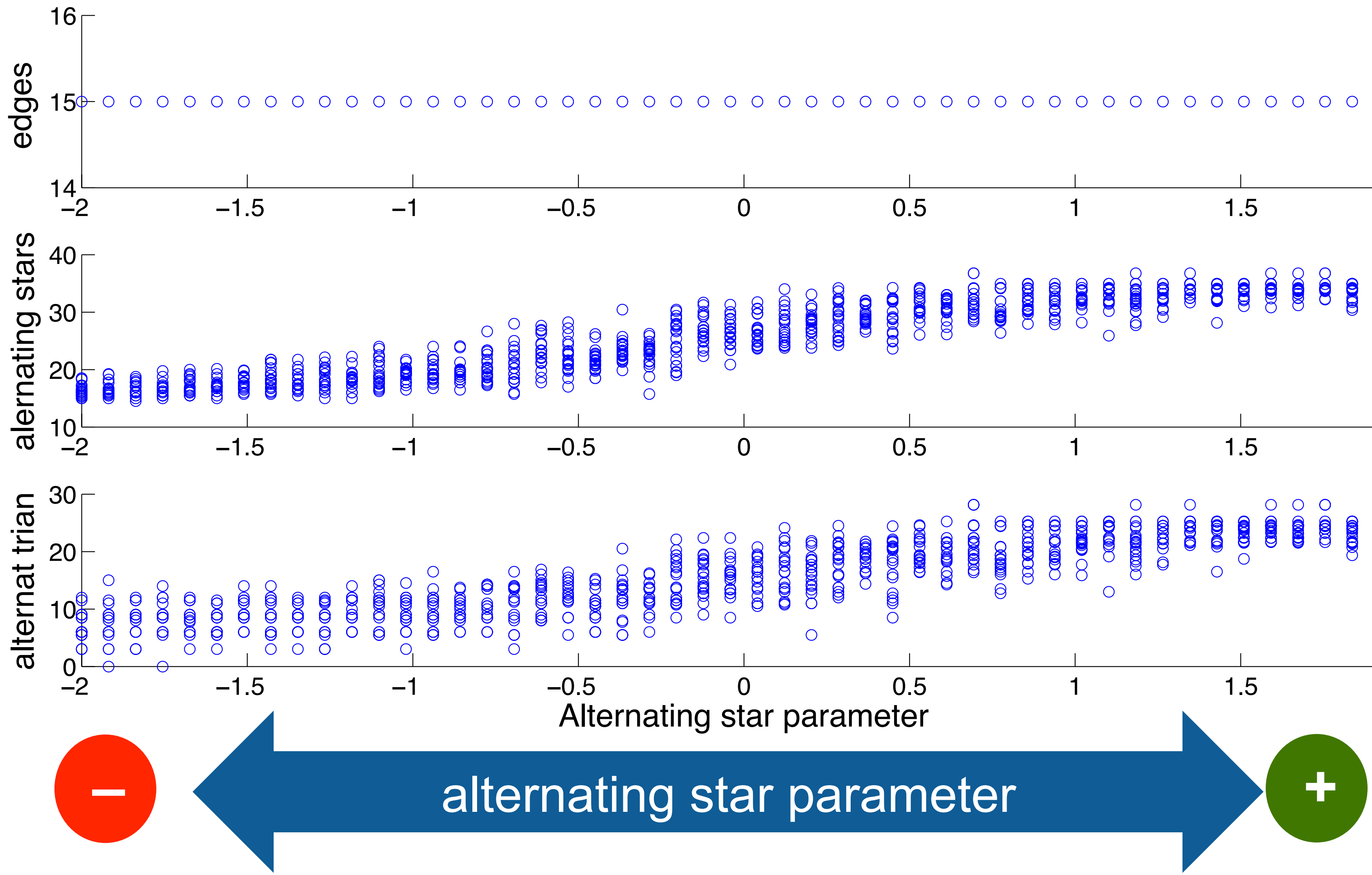
Unpacking the alternating star effect

Interpreting the alternating star **parameter**:



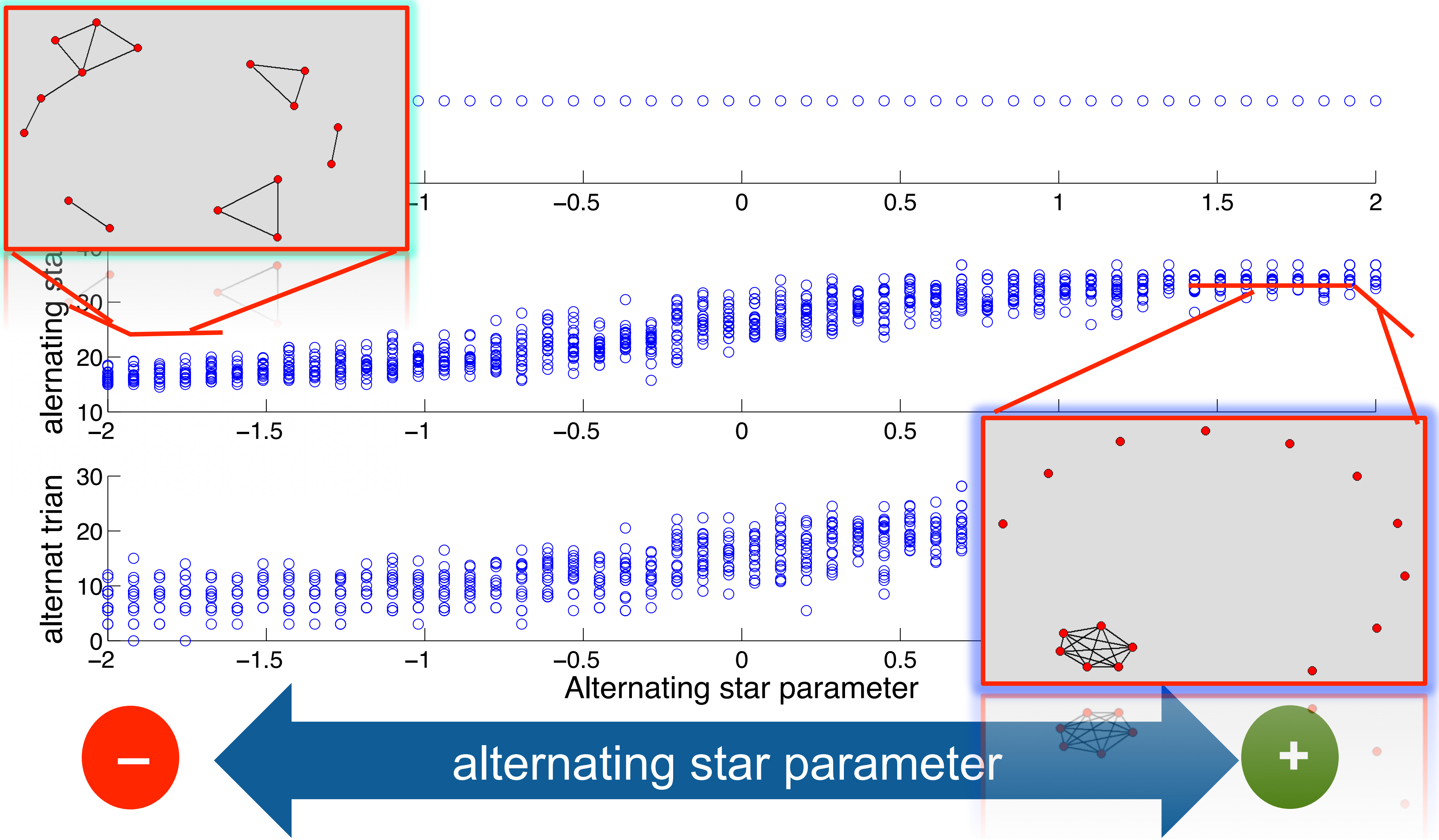
Unpacking the alternating star effect

Statistics for graph ($n = 16$); fixed density; alt trian: 1.17



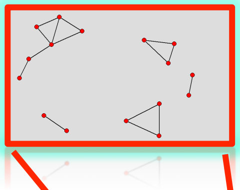
Unpacking the alternating star effect

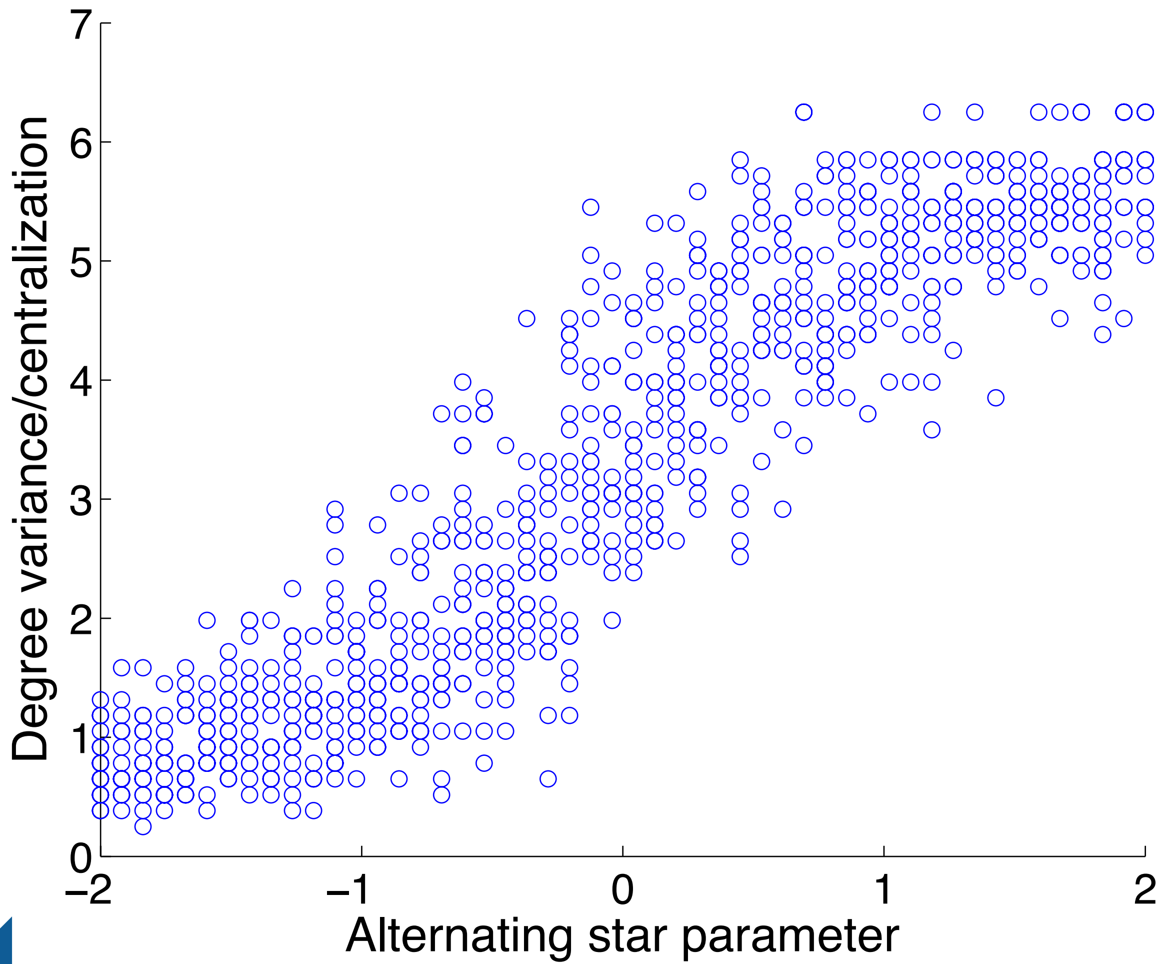
Statistics for graph ($n = 16$); fixed density; alt trian: 1.17

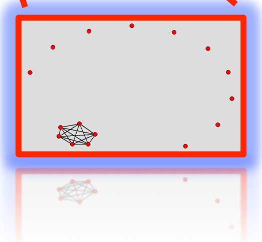


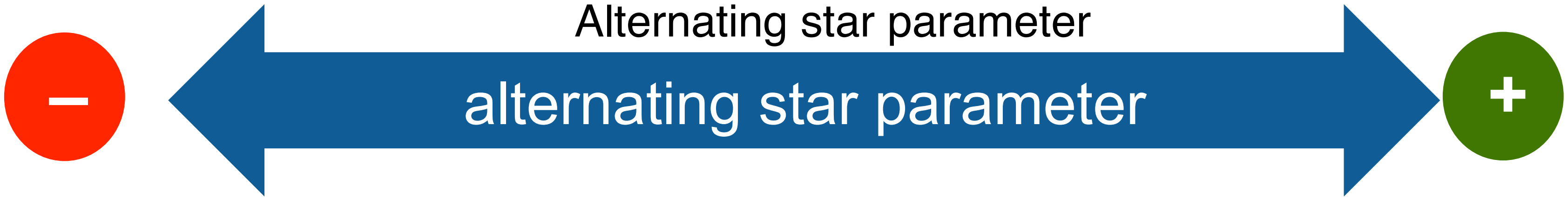
Unpacking the alternating star effect

Graphs ($n = 16$); fixed density; alt trian: 1.17









Problem with Markov models

Note also the influence of isolates:

Alt, stars in terms of degree distribution

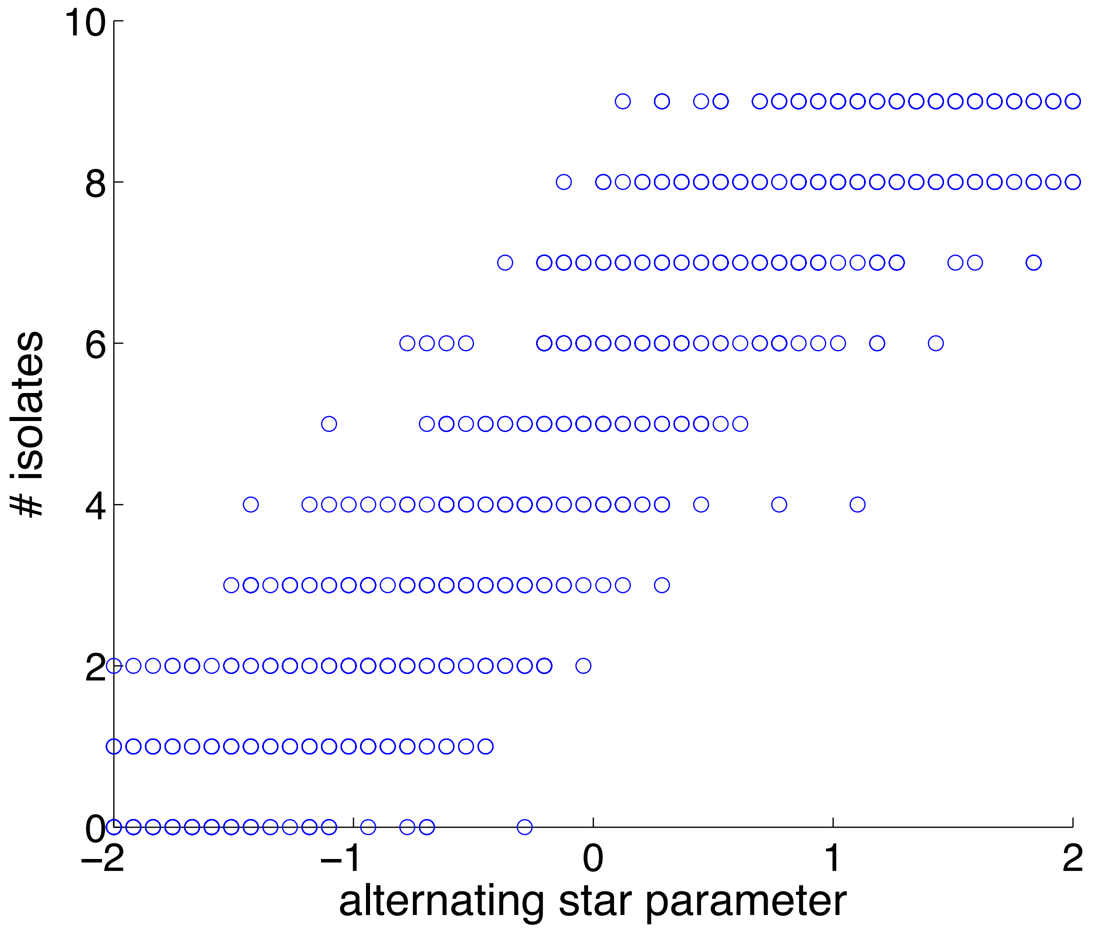
$$= \left(\frac{1}{1-e^{-\alpha}} \right)^2 \sum_{j=0}^{n-1} d_j(x) e^{-\alpha j} + \frac{2L(x)}{1-e^{-\alpha}} - \frac{n}{(1-e^{-\alpha})^2}$$

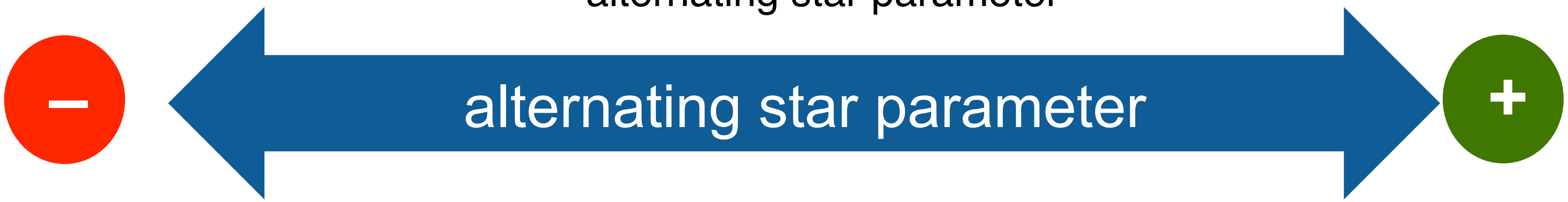
$$\boxed{d_j(x) = \#\{i : x_{i+} = j\}}$$

$$\lambda = e^{\alpha} / (e^{\alpha} - 1)$$

Problem with Markov models

Note also the influence of isolates:





Problem with Markov models

Note also the influence of isolates:

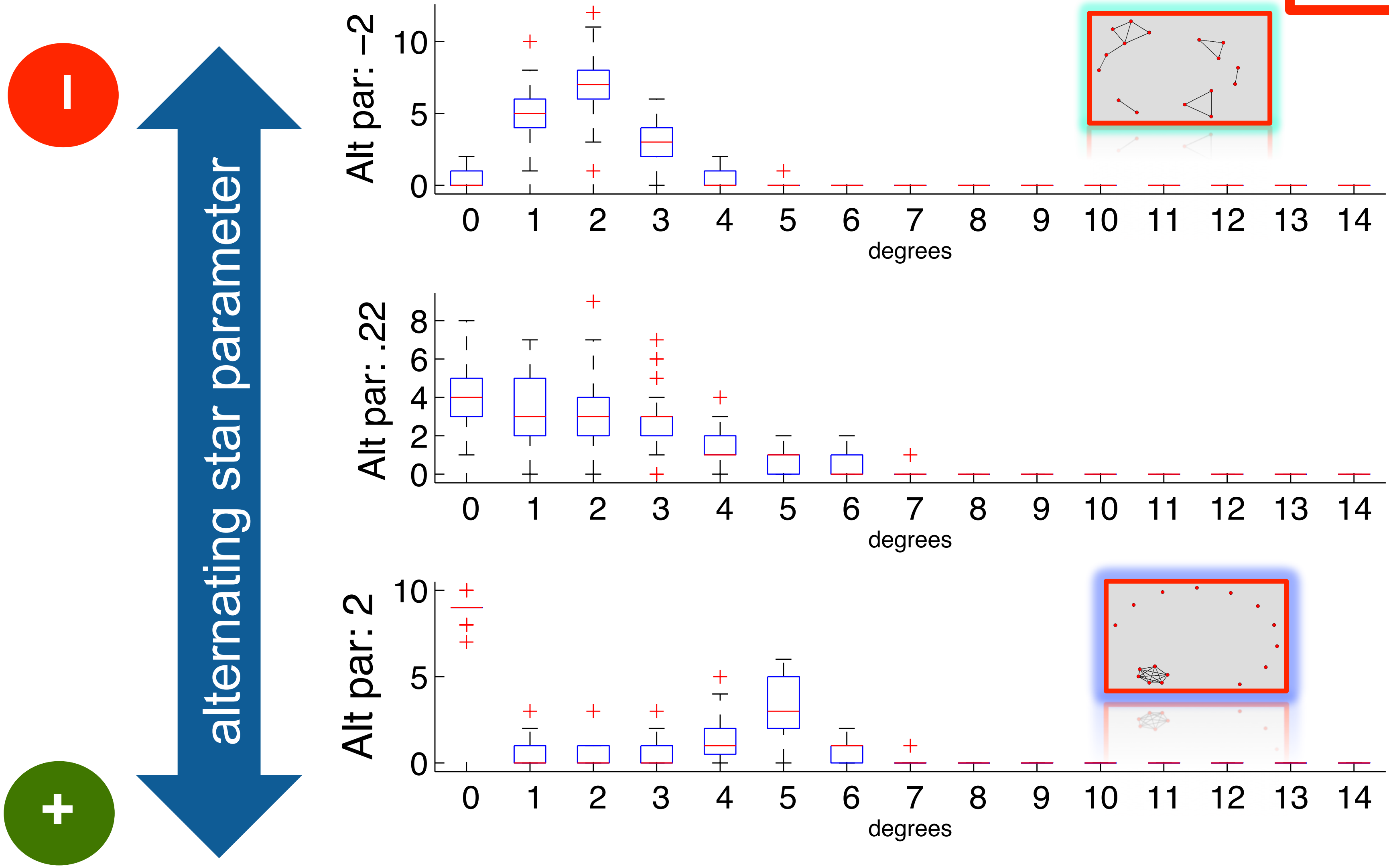
This is because we impose a particular shape on the degree distribution

$$\sigma_2 S_2(x) + \sigma_3 S_3(x) + \cdots + \sigma_{n-1} S_{n-1}(x) = \sigma_{AKS} AKS(x;\lambda)$$

$$= \left(\frac{1}{1-e^{-\alpha}}\right)^2 \sum_{j=0}^{n-1} d_j(x) e^{-\alpha j} + \frac{2L(x)}{1-e^{-\alpha}} - \frac{n}{(1-e^{-\alpha})^2}$$

Problem with Markov models

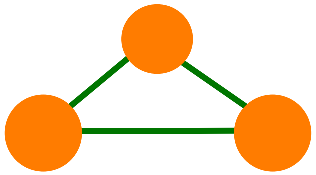
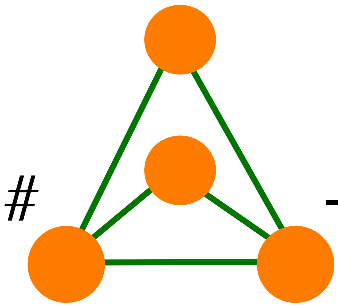
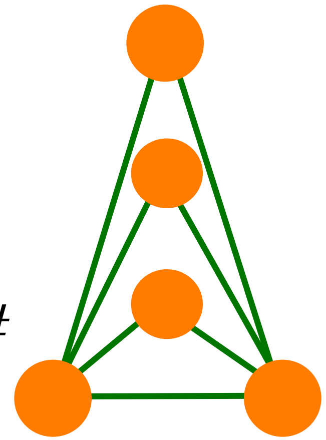
The degree distribution for different values of: $\boxed{\sigma_{AKS}}$



MANCHESTER 1824
The University of Manchester

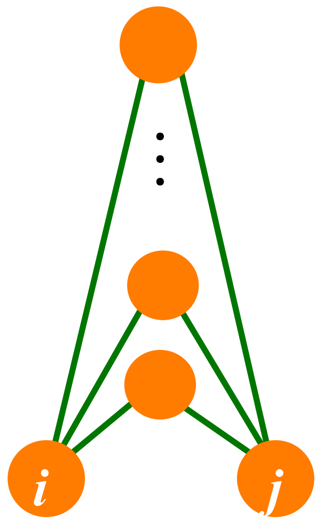
Unpacking the alternating triangle effect

Alternating triangles measure **clustering**

$\tau_1 \times \#$  $-\tau_1 / \lambda \times \#$  $-\tau_2 / \lambda \times \#$  $\cdots$

We may also define the
Edgewise **S**hared **P**artner Statistic:

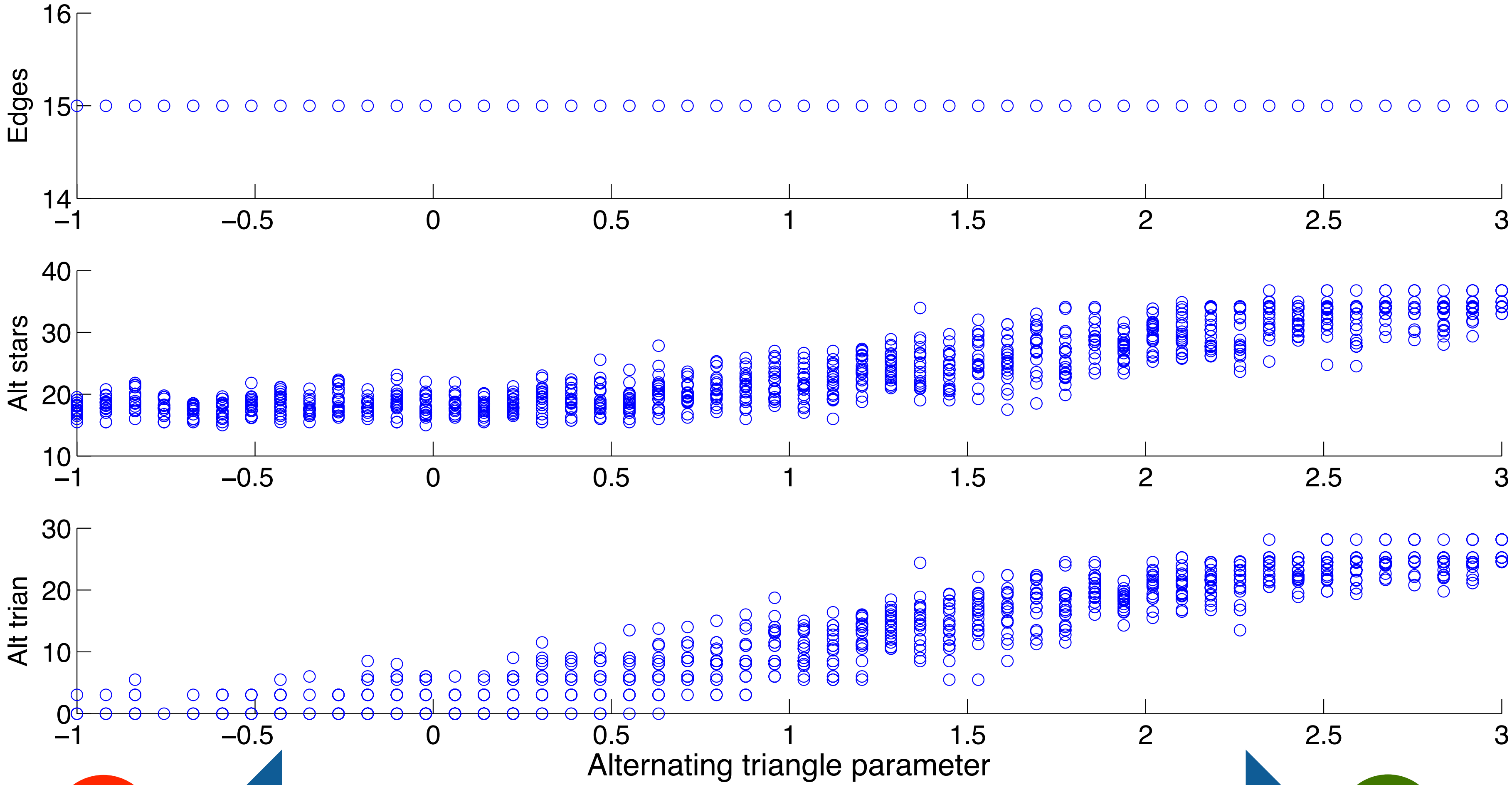
$$ESP_k = \#\{(i,j) : i \sim j, S_{2ij} = k\}$$



… and we can weigh together the **ESP** statistics
using **G**eometrically decreasing weights: **GWESP**

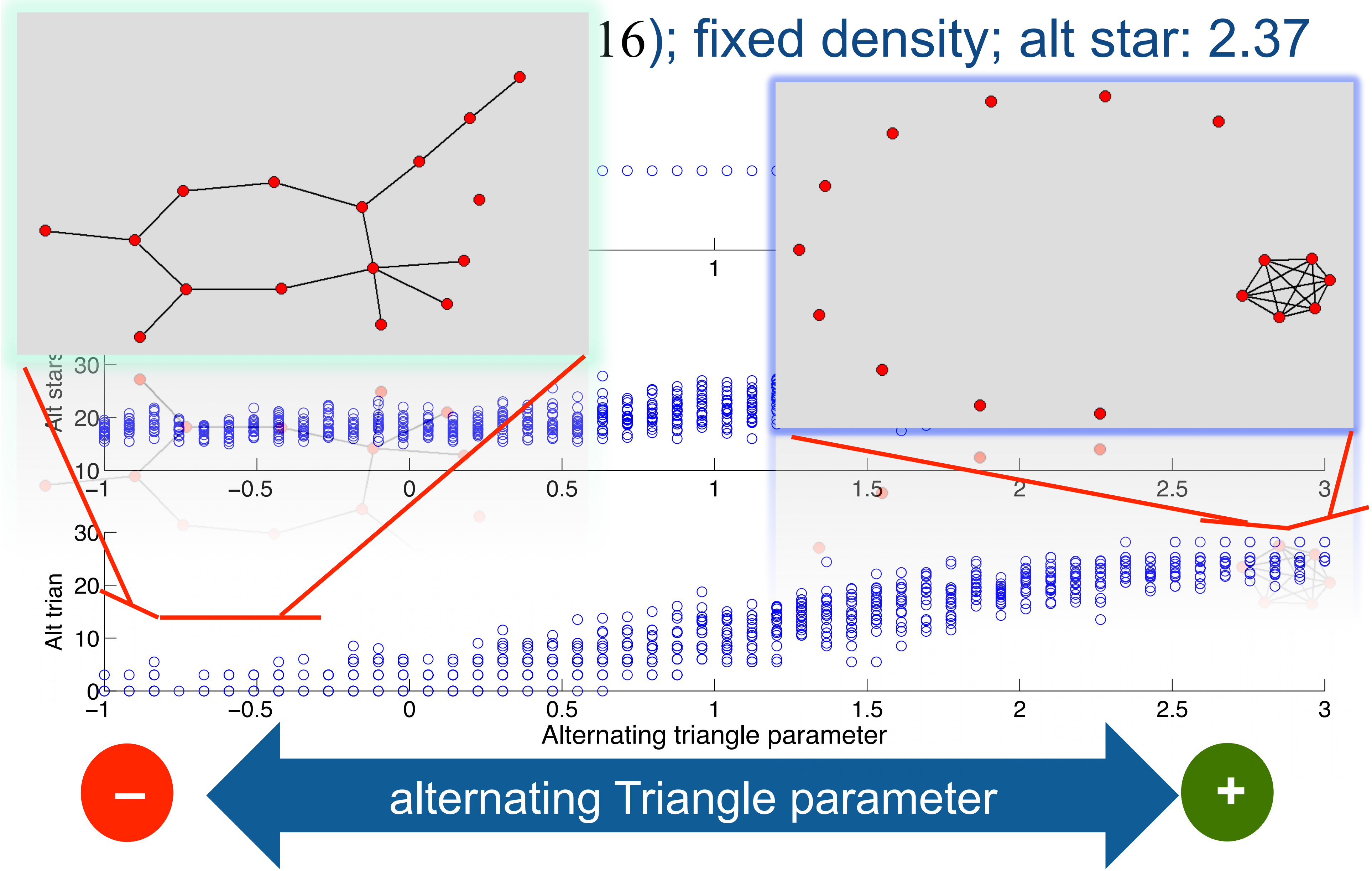
Unpacking the alternating triangle effect

Statistics for graph ($n = 16$); fixed density; alt star: 2.37



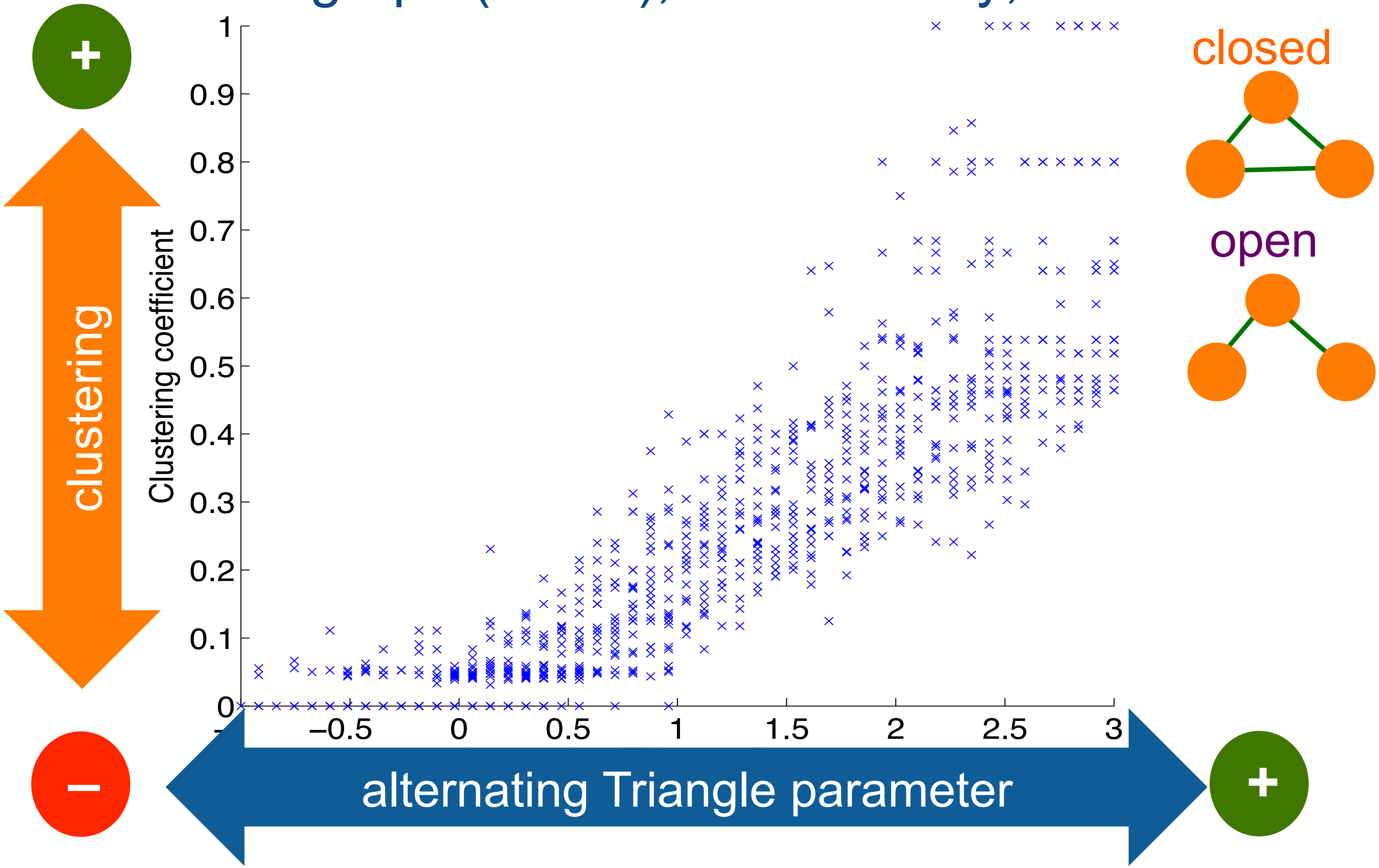
Unpacking the alternating triangle effect

16); fixed density; alt star: 2.37



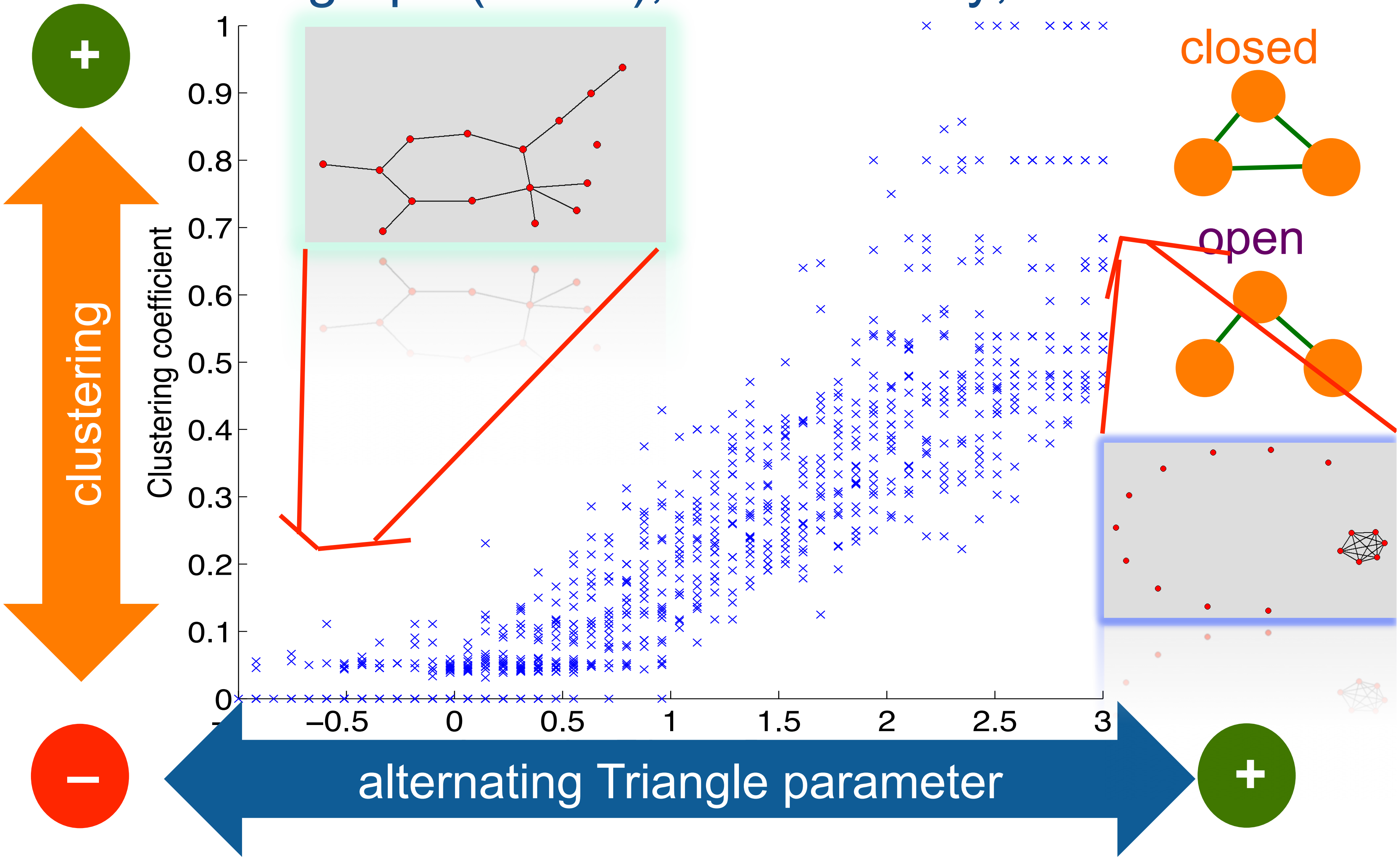
Unpacking the alternating triangle effect

Statistics for graph ($n = 16$); fixed density; alt star: 2.37



Unpacking the alternating triangle effect

Statistics for graph ($n = 16$); fixed density; alt star: 2.37



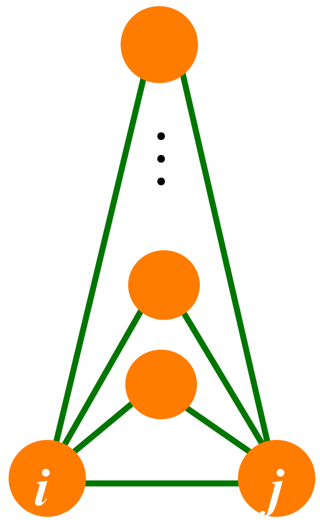
Unpacking the alternating triangle effect

Alternating triangles model **multiply** **clustered areas**

For multiply clustered areas triangles stick together
We model how many others tied actors have

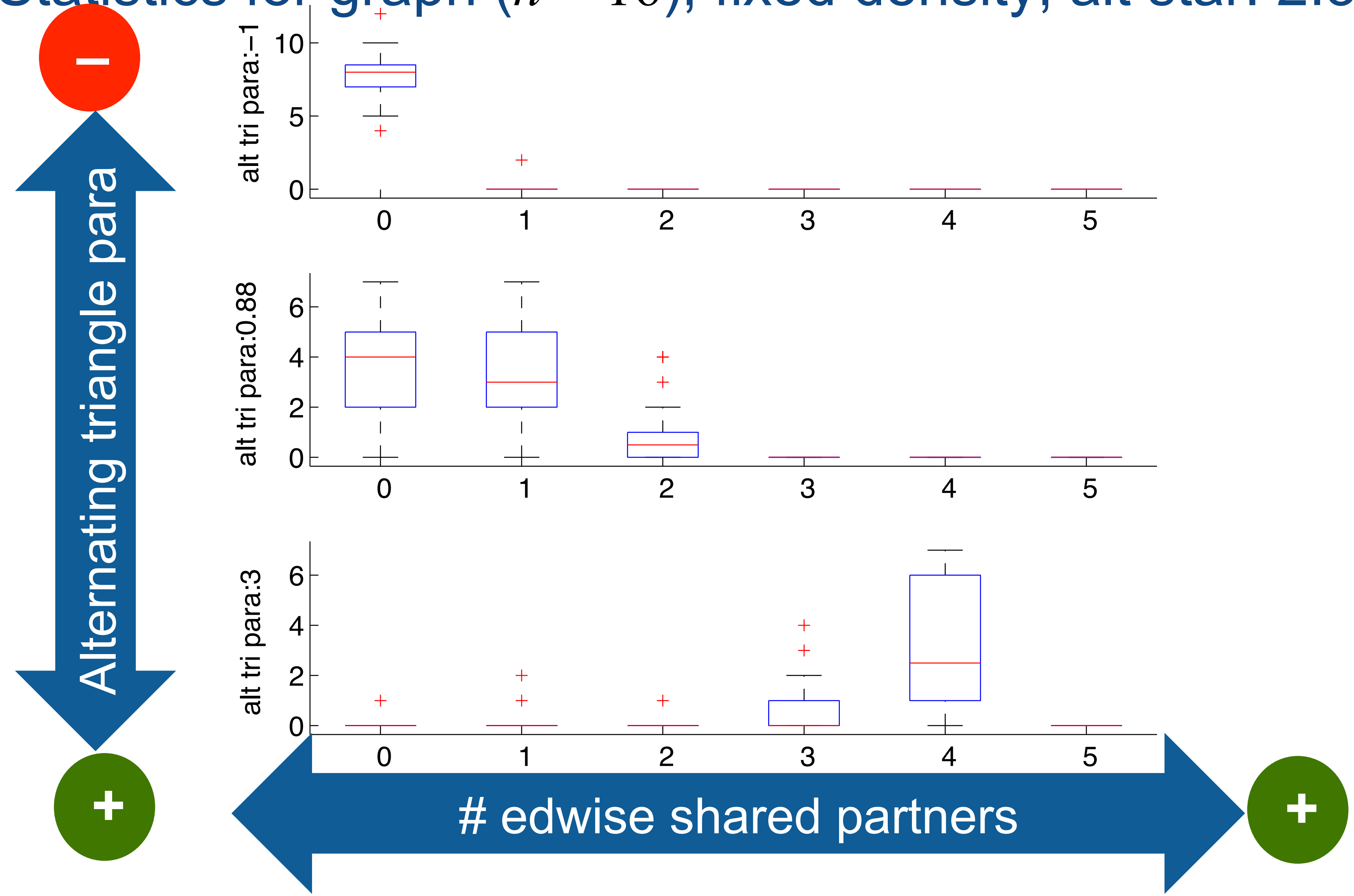
Edgewise **S**hared **P**artner Statistic:

$$ESP_k = \#\{(i,j) : i \sim j, S_{2ij} = k\}$$



Unpacking the alternating triangle effect

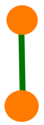
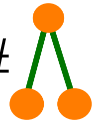
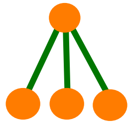
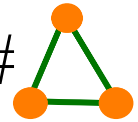
Statistics for graph ($n = 16$); fixed density; alt star: 2.37



Part 5

Convergence check and goodness of fit

Revisiting the Florentine families

$L(x) = \#$  $S_2(x) = \#$  $S_3(x) = \#$  $T(x) = \#$ 

```
Formula:   BusyNetNet ~ kstar(1:3) + triangles

Newton-Raphson iterations:  42
MCMC sample of size 10000

Monte Carlo MLE Results:
          Estimate Std. Error MCMC s.e. p-value
kstar1     -1.6130     0.6699     0.462  0.0176 *
kstar2      0.7492     0.6407     0.455  0.2446
kstar3     -0.5408     0.3574     0.225  0.1330
triangle    1.483[illegible] [illegible].45[illegible] [illegible] [illegible] *
---
```

statnet

Why difference?

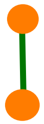
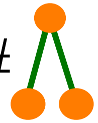
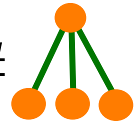
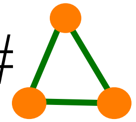
estimation-padgetestsunday.txt - Notepad

File Edit Format View Help

PNet

```
* num of iterations in each step = 280.000000
****************************************
mean statistics in phase3:15.246000 37.306000 25.856000 5.054000

*Estimation Result for Network SUMMARY (parameter, standard error, t-statistics)
NOTE: t-statistics = (observation - sample mean)/standard error
        Edge: -4.137319, 1.07210, -0.04370      *
        2-Star: 0.973517, 0.59180, -0.06408
        3-Star: -0.563624, 0.35420, -0.09745
        Triangle: 1.261550, 0.61588, -0.01524   *
```

Revisiting the Florentine families

$L(x) = \#$  $S_2(x) = \#$  $S_3(x) = \#$  $T(x) = \#$ 

Pnet checks convergence

$$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

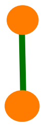
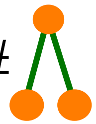
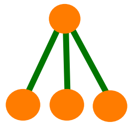
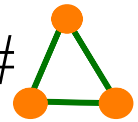
in 3rd phase

estimation-padgetestsunday.txt - Notepad

File Edit Format View Help

```
* num of iterations in each step = 280.000000
****************************************
mean statistics in phase3:15.246000 37.306000 25.856000 5.054000

*Estimation Result for Network SUMMARY (parameter, standard error, t-statistics)
NOTE: t-statistics = (observation - sample mean)/standard error
        Edge: -4.137319, 1.07210, -0.04370       *
        2-Star: 0.973517, 0.59180, -0.06408
        3-Star: -0.563624, 0.35420, -0.09745
        Triangle: 1.261550, 0.61588, -0.01524    *
```

Revisiting the Florentine families

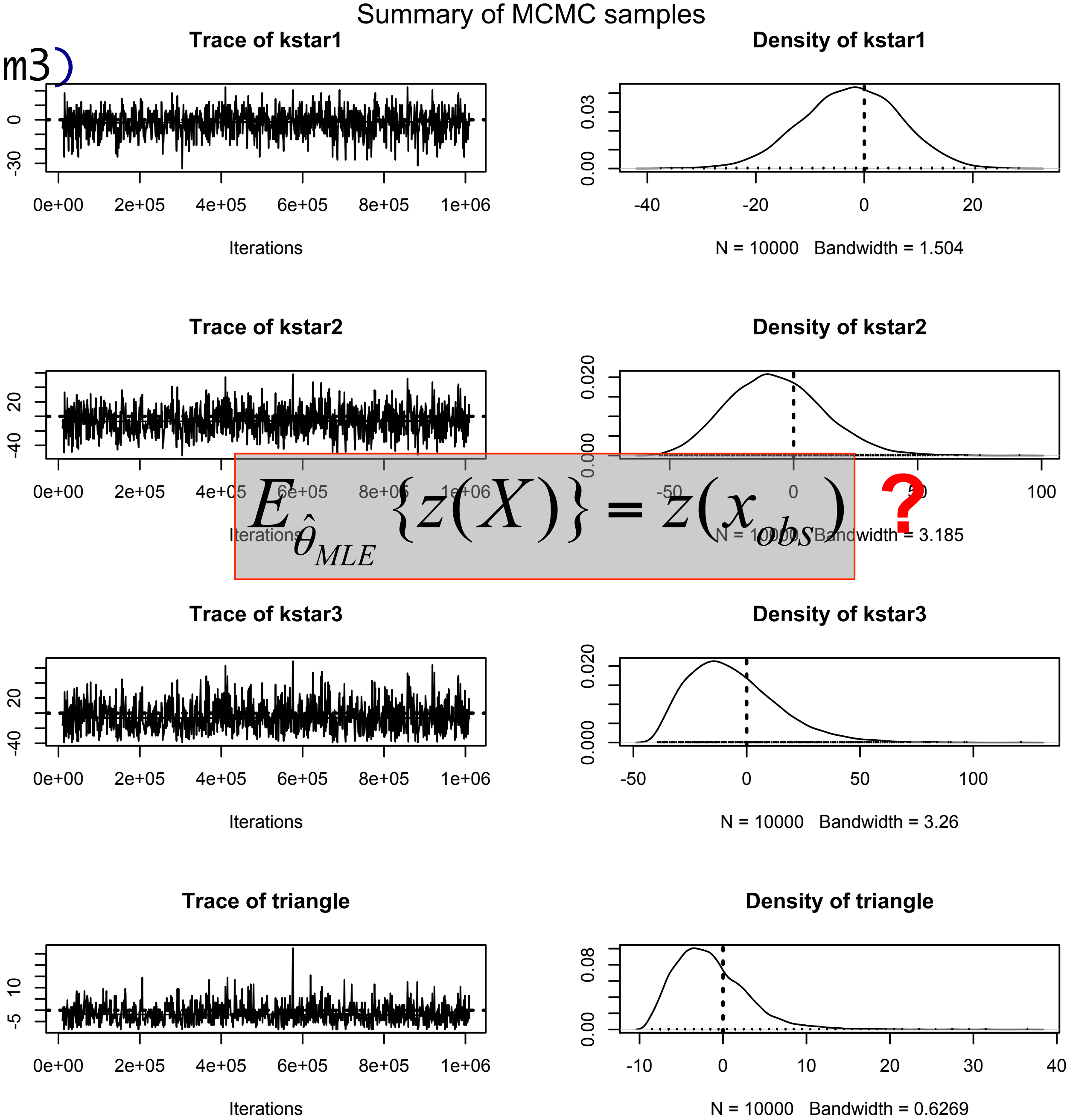
$L(x) = \#$  $S_2(x) = \#$  $S_3(x) = \#$  $T(x) = \#$ 

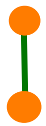
```
Formula:   BusyNetNet ~ kstar(1:3) + triangles

Newton-Raphson iterations:  42
MCMC sample of size 10000

Monte Carlo MLE Results:
         Estimate Std. Error MCMC s.e. p-value
kstar1    -1.6130     0.6699     0.462  0.0176 *
kstar2     0.7492     0.6407     0.455  0.2446
kstar3    -0.5408     0.3574     0.225  0.1330
triangle   1.4837     0.4592     0.138  0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

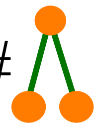
Lets check $$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs})$$

For statnet

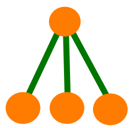
```r
Estim3 <- ergm(BusyNetNet ~ kstar(1:3) + triangles ,
verbose=TRUE)
mcmc.diagnostics(Estim3)
```

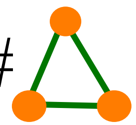


$L(x) = \#$ 

$S_2(x) = \#$ 

$$E_{\hat{\theta}_{MLE}}\{z(X)\} = z(x_{obs}) \; ?$$

$S_3(x) = \#$ 

$T(x) = \#$ 

Simulation, GOF, and Problems

Given a specific model we can simulate potential outcomes under that model. This is used for

- Estimation:
  (a) to match observed statistics and expected statistics
  (b) to check that we have "the solution"
- GOF: to check whether the model can replicated features of the data that we not explicitly modelled
- Investigate behaviour of model, e.g.: degeneracy and dependence on scale

Simulation, GOF, and Problems

- Standard goodness of fit procedures are not valid for ERGMs – no F-tests or Chi-square tests available
- If indeed the fitted model adequately describes the data generating process, then the graphs that the model produces should be similar to observed data
- For fitted effects this is true by construction
- For effects/structural feature that are not fitted this may not be true
- If it is true, the modelled effects are the only effects necessary to produce data – a "proof" of the concept or ERGMs

Simulation, GOF, and Problems

Example: for our fitted model for Lazega

```
NOTE: t-statistics = (observation - sample mean)/standard err(
effects                  estimates        stderr  t-ratio
edge                     -5.862515        0.56404 0.04105 *
AT(2.00)                 1.011721         0.17095 0.05003 *
practice_interaction     1.499409         0.40322 0.02371 *
practice_activity        -0.331023        0.21995 0.02142
senior_sum               0.842661         0.23348 0.04943 *
sex_matching             0.702477         0.26389 0.05839 *
off_matching             1.145290         0.19749 0.00134 *
```

We can look at how well the model reproduces

$$g(x_{obs})$$

For arbitrary function g

Simulation, GOF, and Problems

From the goodness-of-fit tab in Pnet we get

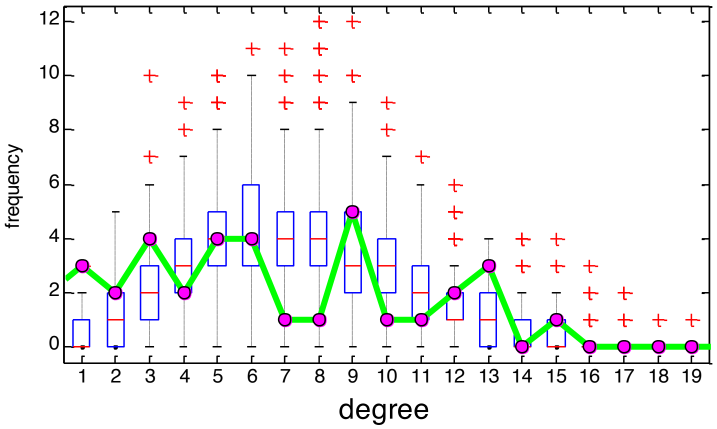
gof_lazega.txt

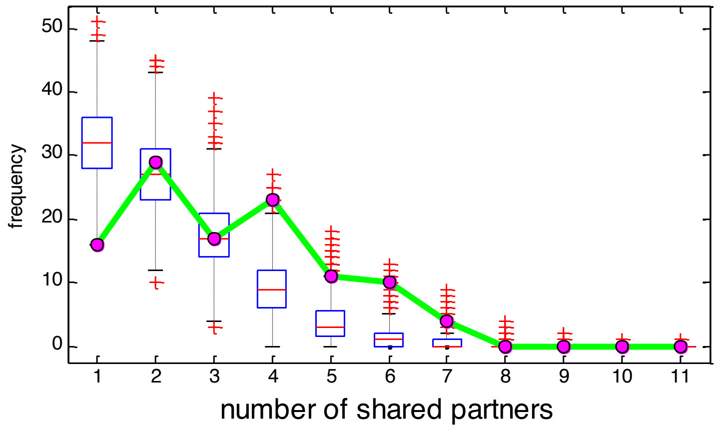
```
triangles       120     117.454 28.310  0.090
4-clique        0       0.000   0.000   -0.003
5-clique        0       -0.000  0.000   0.001
6-clique
0       -32995302646884151000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000
.000
2320676770410929200000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000.000
        0.001
7-clique        0       0.000   0.000   0.001
Isolates        2       3.008   1.798   -0.561
Triangle2       562     532.099 197.900 0.151
3Path   7202    7235.453        2314.026        -0.014
4Cycle  546     534.822 195.672 0.057
1ET     2884    2814.319        944.285 0.074
2ET     11930   11873.856       5210.503        0.011
0ET     81      51.891  29.776  0.978
ETNT    81      31.189  24.565  2.028
2_3Star -414    -392.237        214.970 -0.104
1_2Triangle     -161    -148.595        72.149  -0.172
1_2ET   -3081   -3122.609       1687.893        0.025
AS(2.00)        335.345 336.904 58.266  -0.027
AS(2.00)        335.345 336.904 58.266  -0.027
AT(2.00)        181.313 181.428 32.998  -0.003
AT(2.00)        181.313 181.428 32.998  -0.003
A2P(2.00)       557.648 551.518 100.656 0.061
AC(2.00)
```

Simulation, GOF, and Problems

Effect	MLE	s.e.
Density	-6.501	0.727
Main effect seniority	1.594	0.324
Main effect practice	0.902	0.163
Homophily effect practice	0.879	0.231
Homophily sex	1.129	0.349
Homophily office	1.654	0.254

Simulation, GOF, and Problems

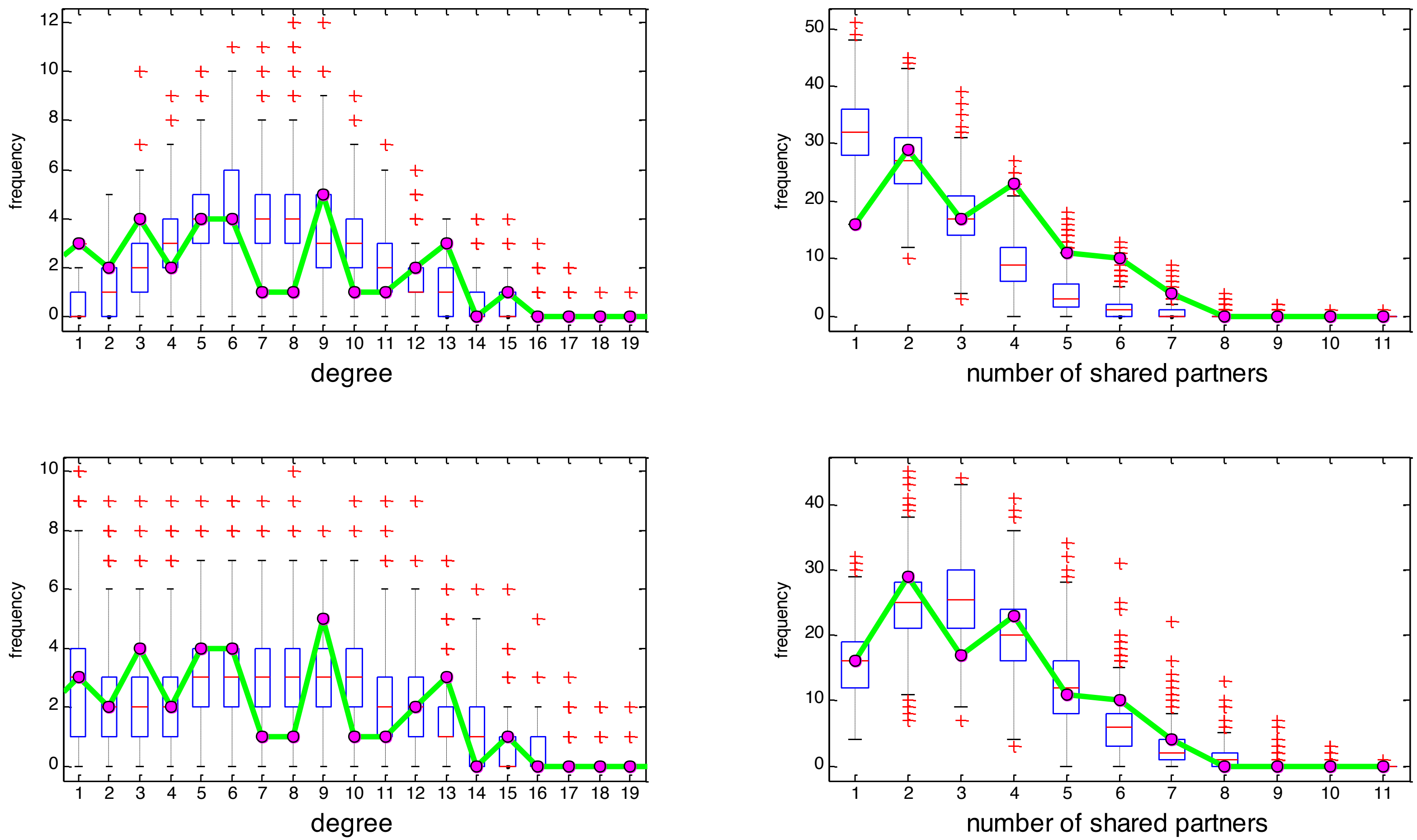




Simulation, GOF, and Problems

Effect	MLE	s.e.	MLE	s.e.
Density	-6.501	0.727	-6.510	0.637
Main effect seniority	1.594	0.324	0.855	0.235
Main effect practice	0.902	0.163	0.410	0.118
Homophily effect practice	0.879	0.231	0.759	0.194
Homophily sex	1.129	0.349	0.704	0.254
Homophily office	1.654	0.254	1.146	0.195
GWEPS			0.897	0.304
Log-lambda			0.778	0.215

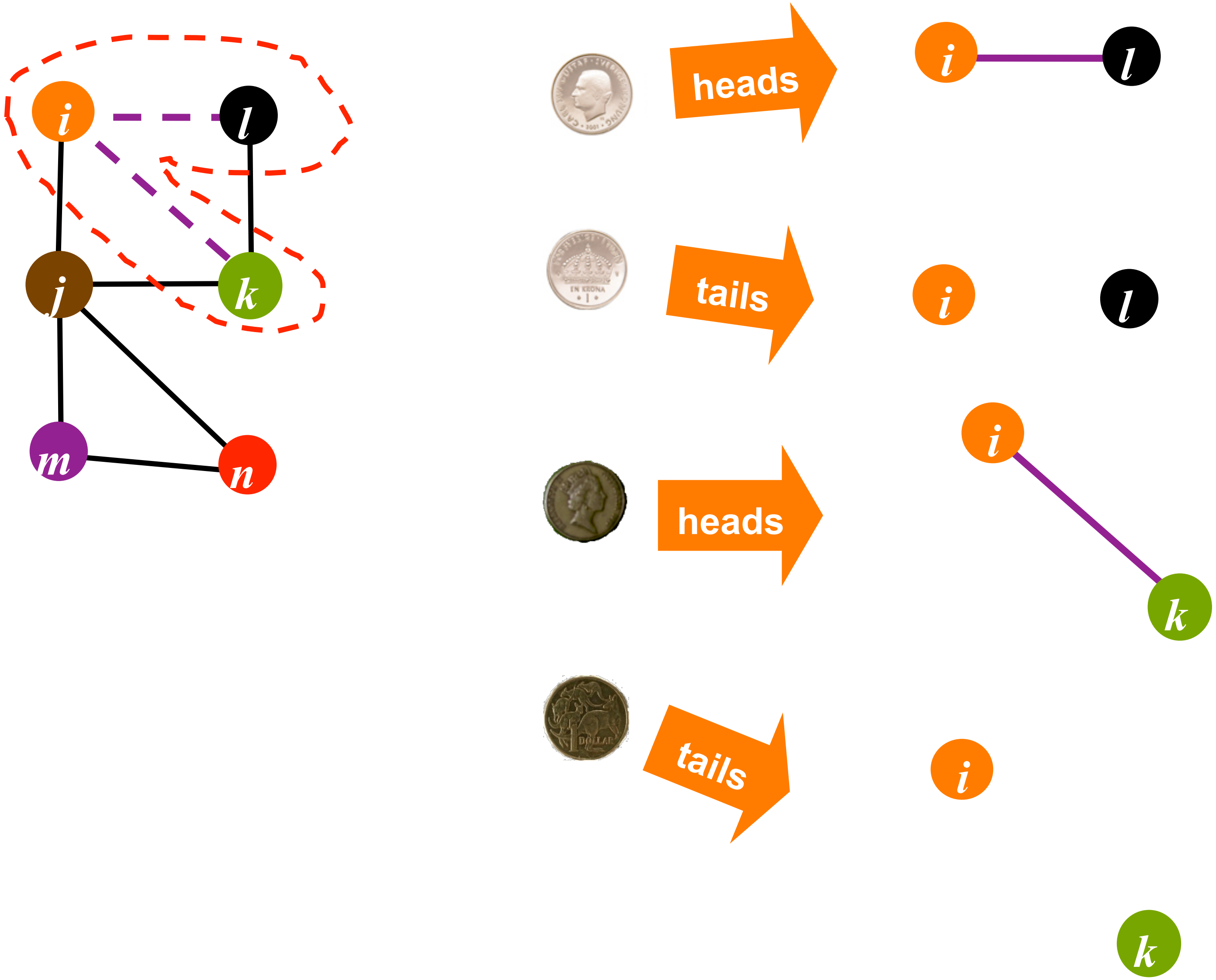
Simulation, GOF, and Problems



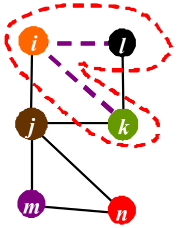
Part 6

Dependencies – Sufficient statistics - homogeneity

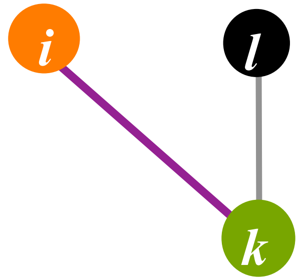
Independence - Deriving the ERGM

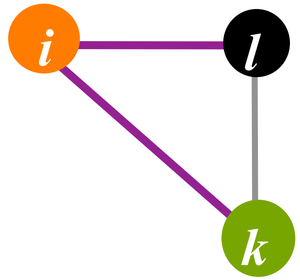


Independence - Deriving the ERGM



		AUD	
		 0.5	 0.5
SEK	 0.5	  0.25	  0.25
	 0.5	  0.25	  0.25





Knowledge of AUD, e.g.  does **not** help us predict SEK

e.g. whether  or 

Independence - Deriving the ERGM

Knowledge of AUD, e.g.  does **not** help us predict SEK

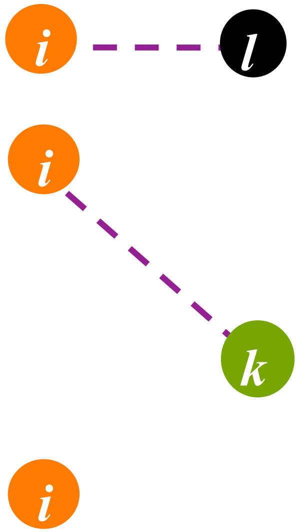
e.g. whether  or 

even though **dyad** $\{i,l\}$

and **dyad** $\{i,k\}$

have **vertex** i

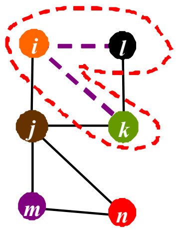
in **common**



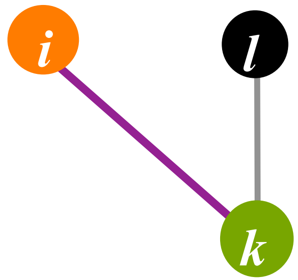
Independence - Deriving the ERGM

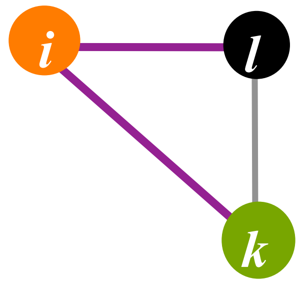
May we find model such that knowledge of AUD, e.g. **does** help us predict SEK

e.g. whether or ?



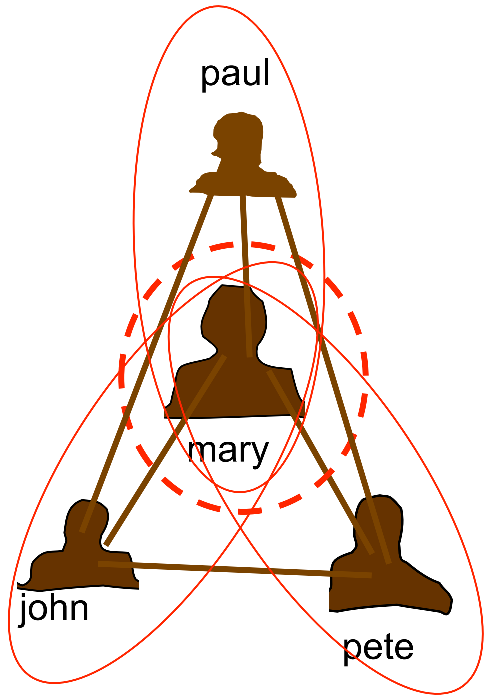
	AUD	
	0.5	0.5
SEK 0.5	0.4	0.1
SEK 0.5	0.1	0.4





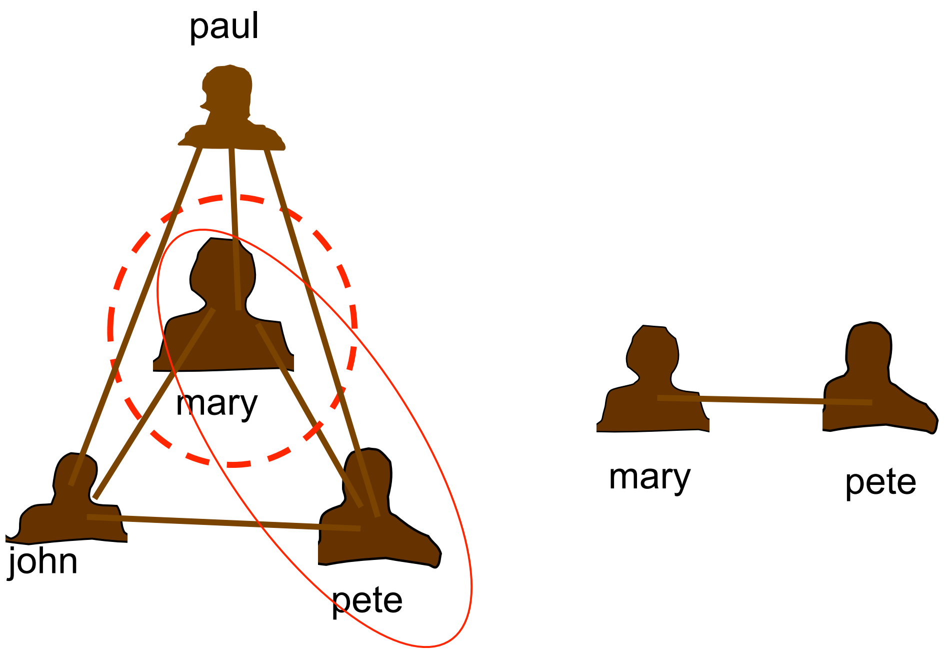
Deriving the ERGM: From Markov graph to Dependence graph

Consider the tie-variables that have **Mary** in **common**

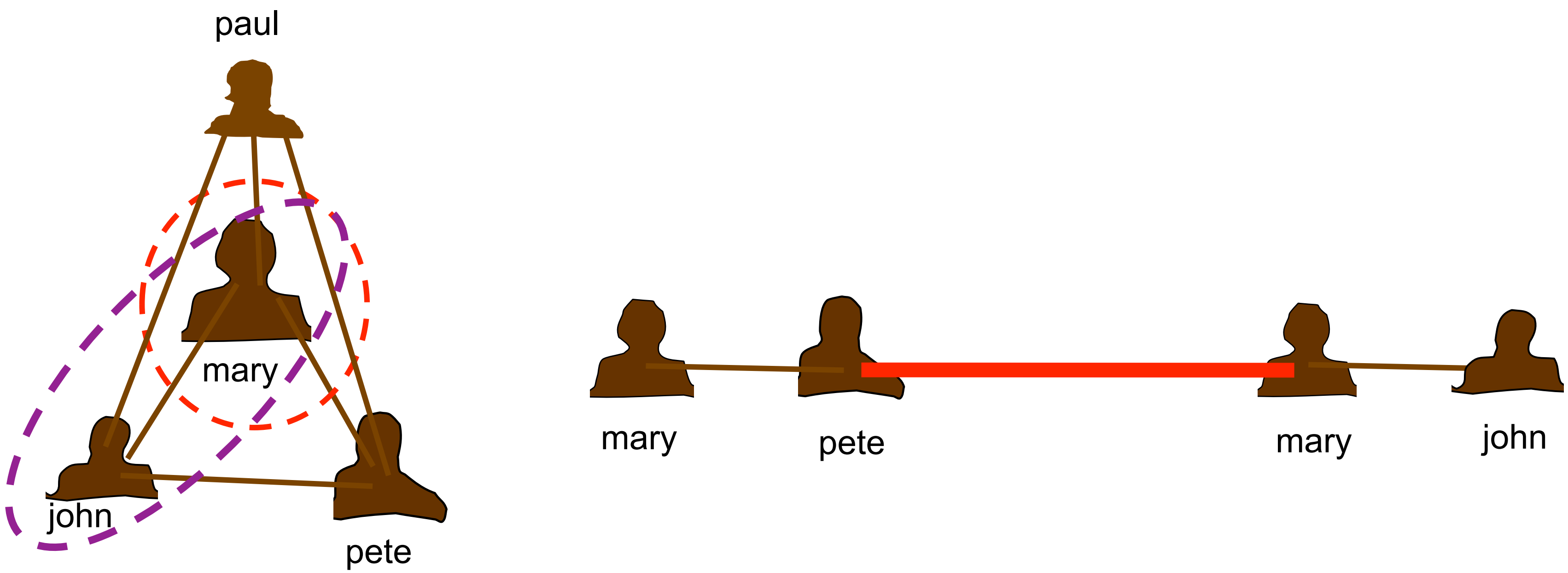


How may we make these "**dependent**"?

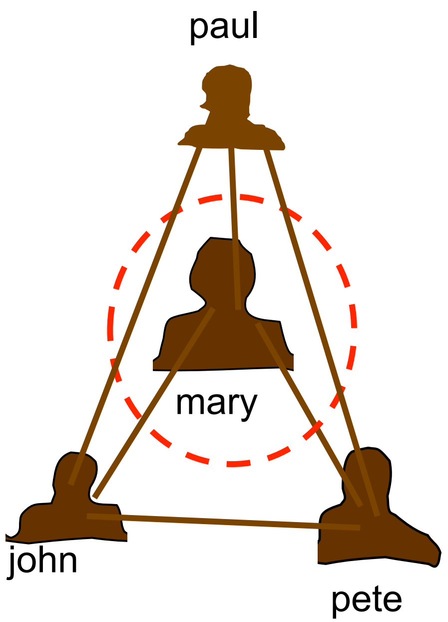
Deriving the ERGM: From Markov graph to Dependence graph

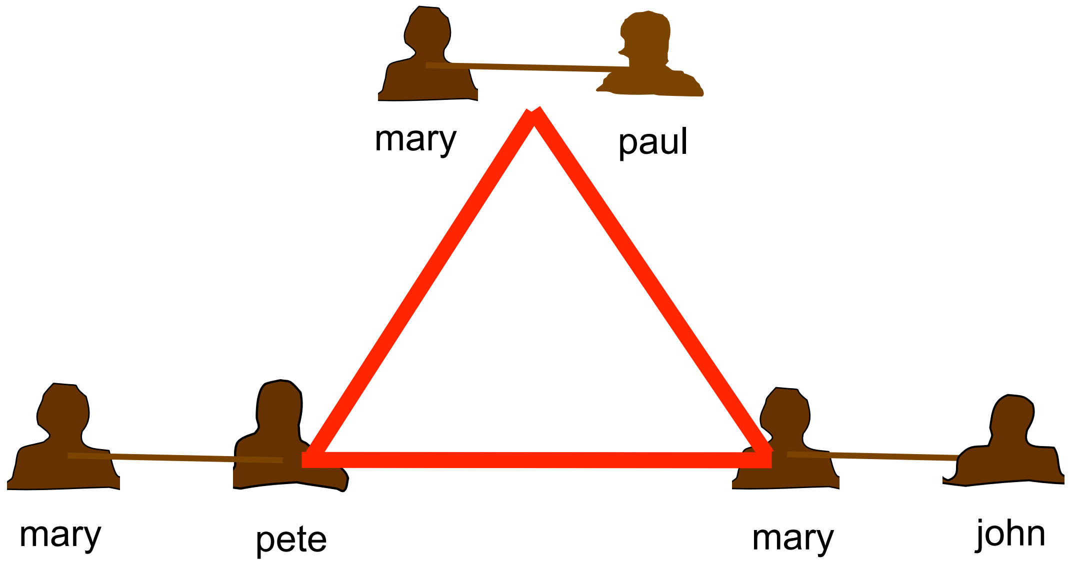


Deriving the ERGM: From Markov graph to Dependence graph

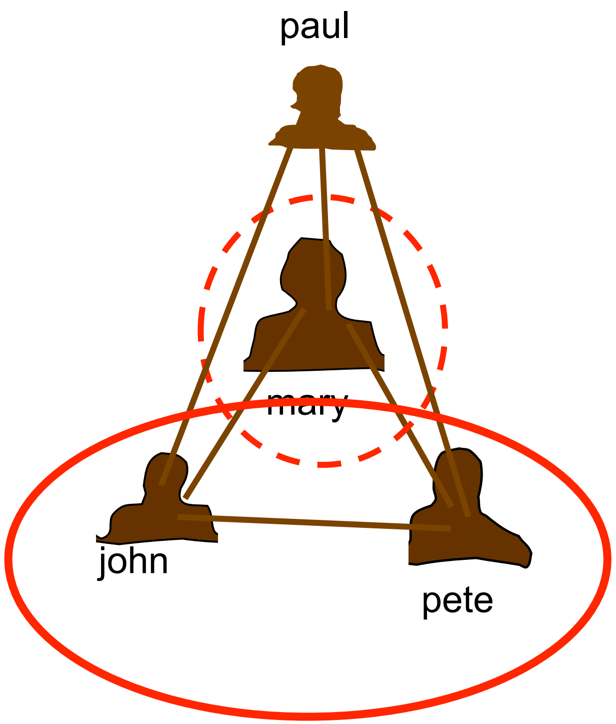


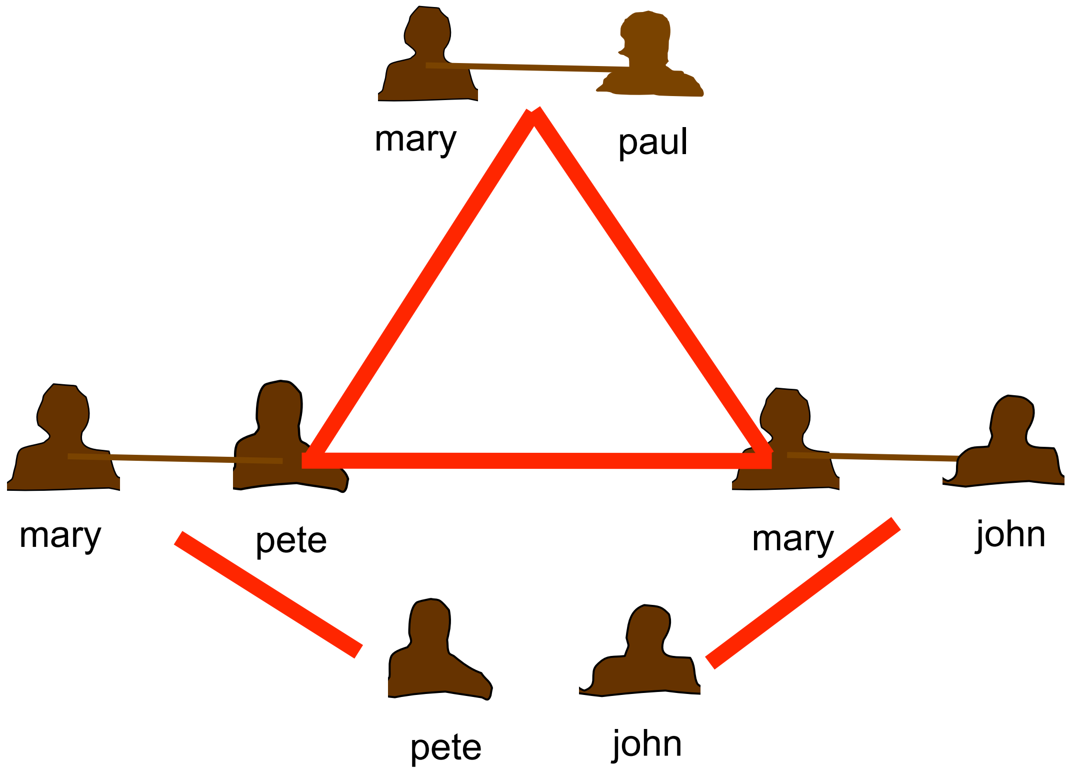
Deriving the ERGM: From Markov graph to Dependence graph



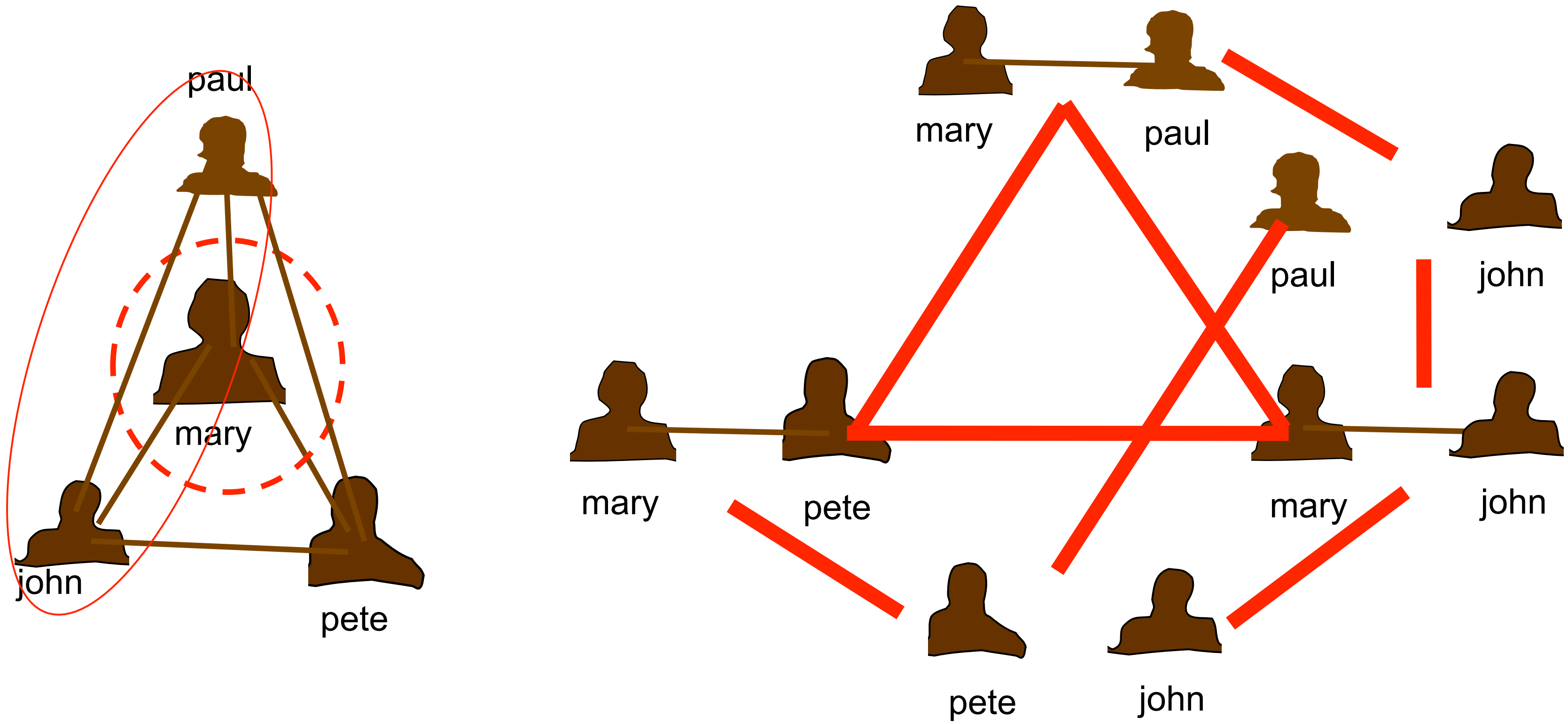


Deriving the ERGM: From Markov graph to Dependence graph

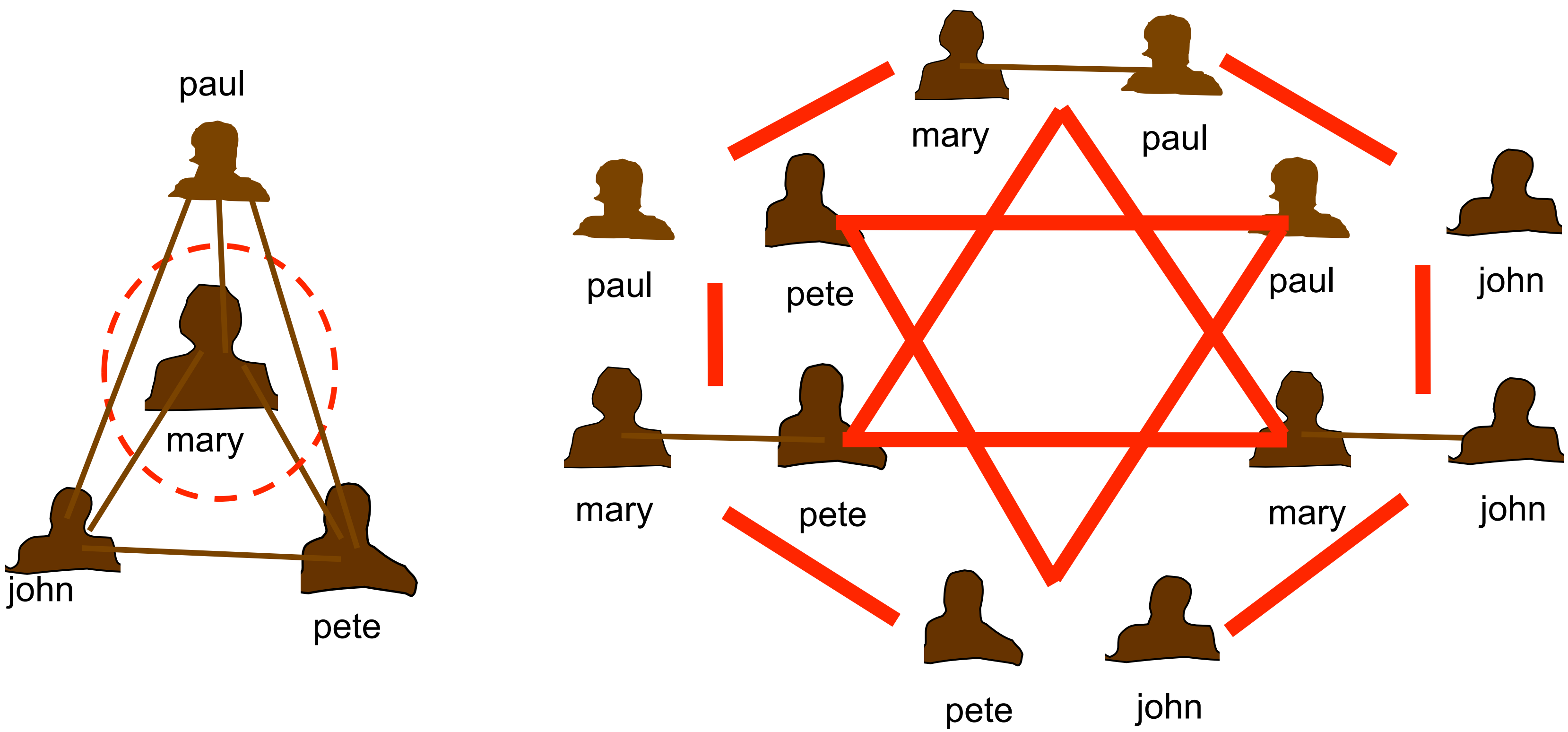




Deriving the ERGM: From Markov graph to Dependence graph

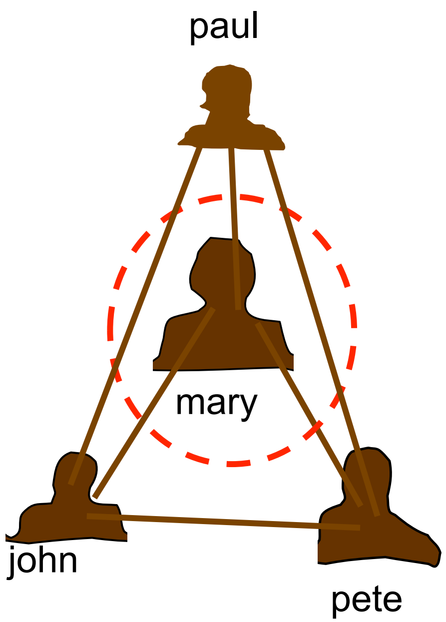


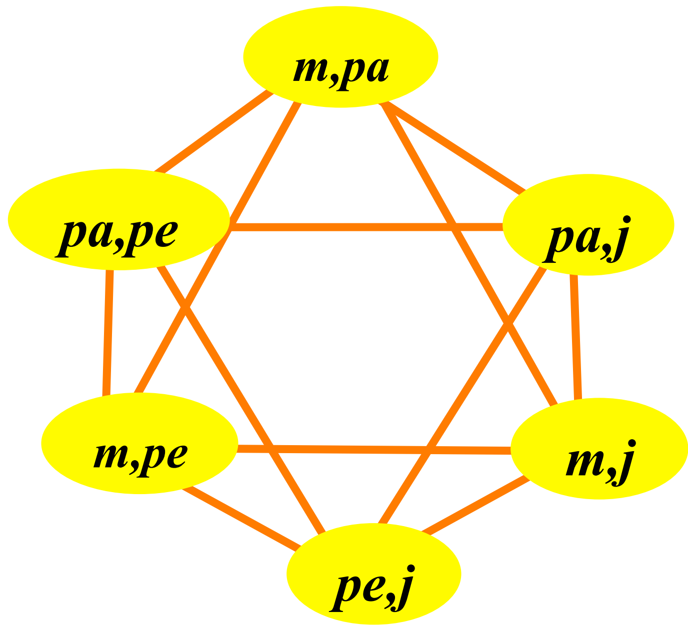
Deriving the ERGM: From Markov graph to Dependence graph



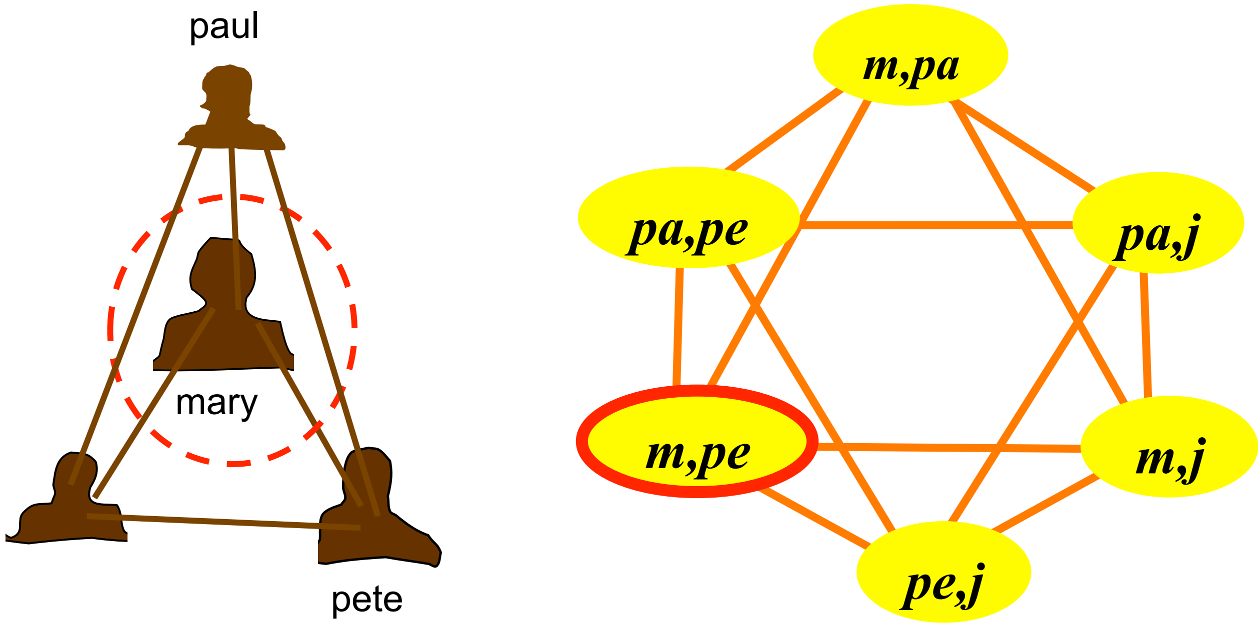
Deriving the ERGM: From Markov graph to Dependence graph

The "probability structure" of a Markov graph is described by **cliques** of the dependence graph (Hammersley-Clifford)….

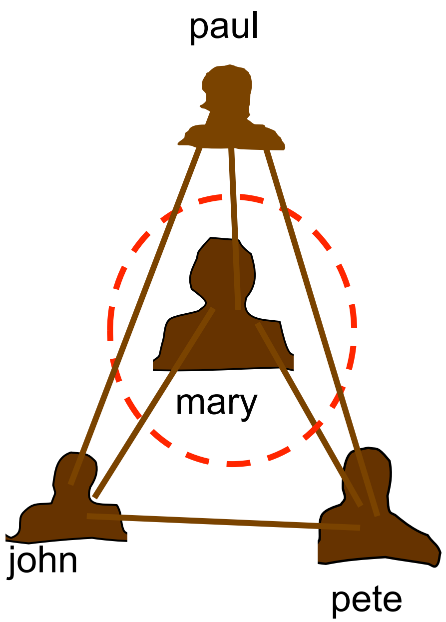


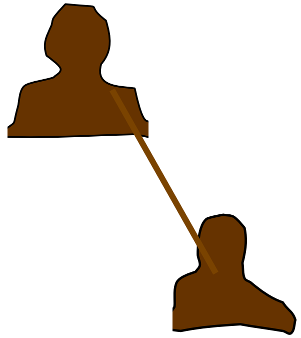


Deriving the ERGM: From Markov graph to Dependence graph

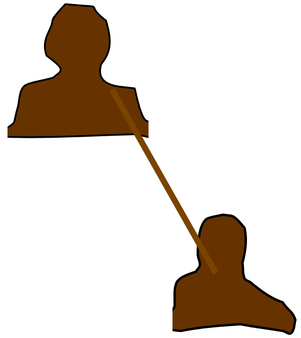


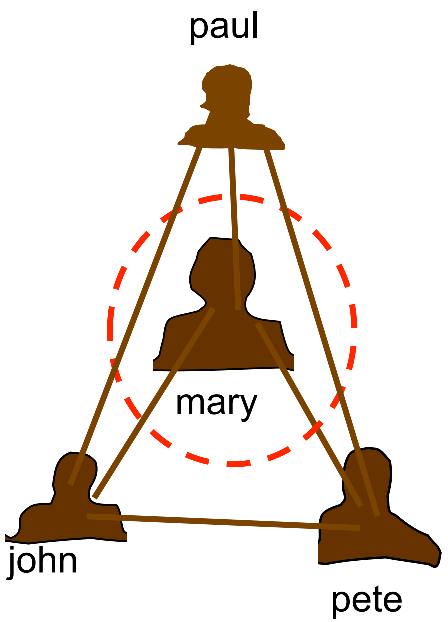
Deriving the ERGM: From Markov graph to Dependence graph

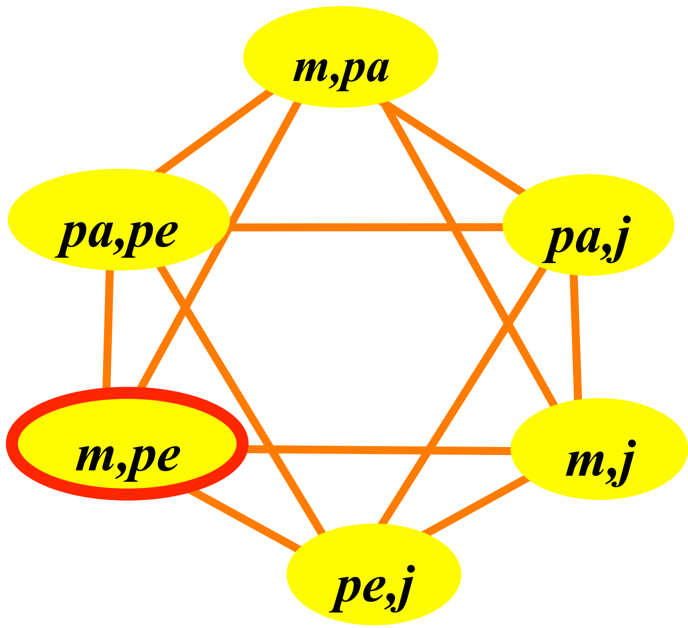




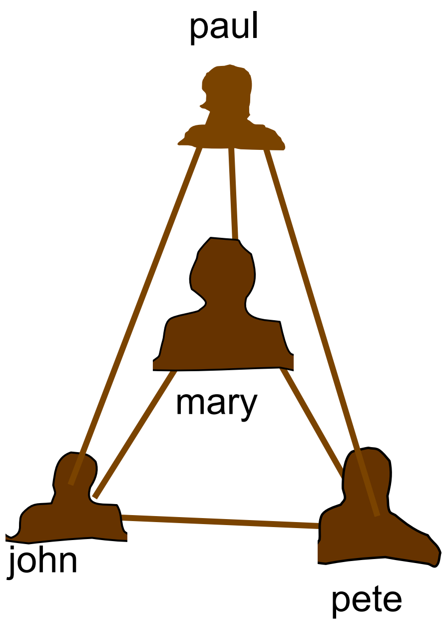
Deriving the ERGM: From Markov graph to Dependence graph

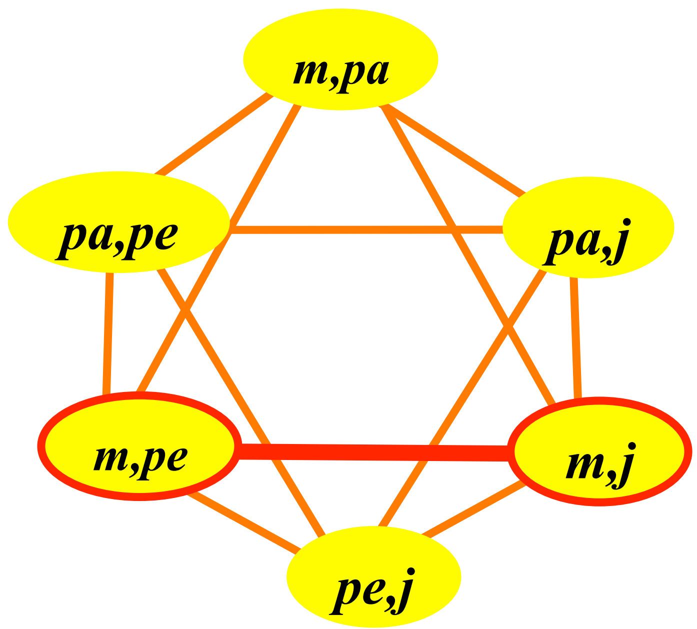


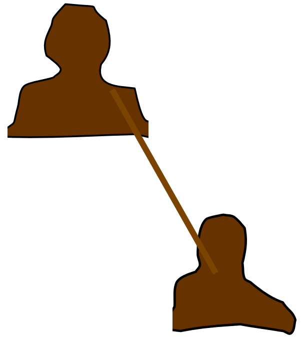




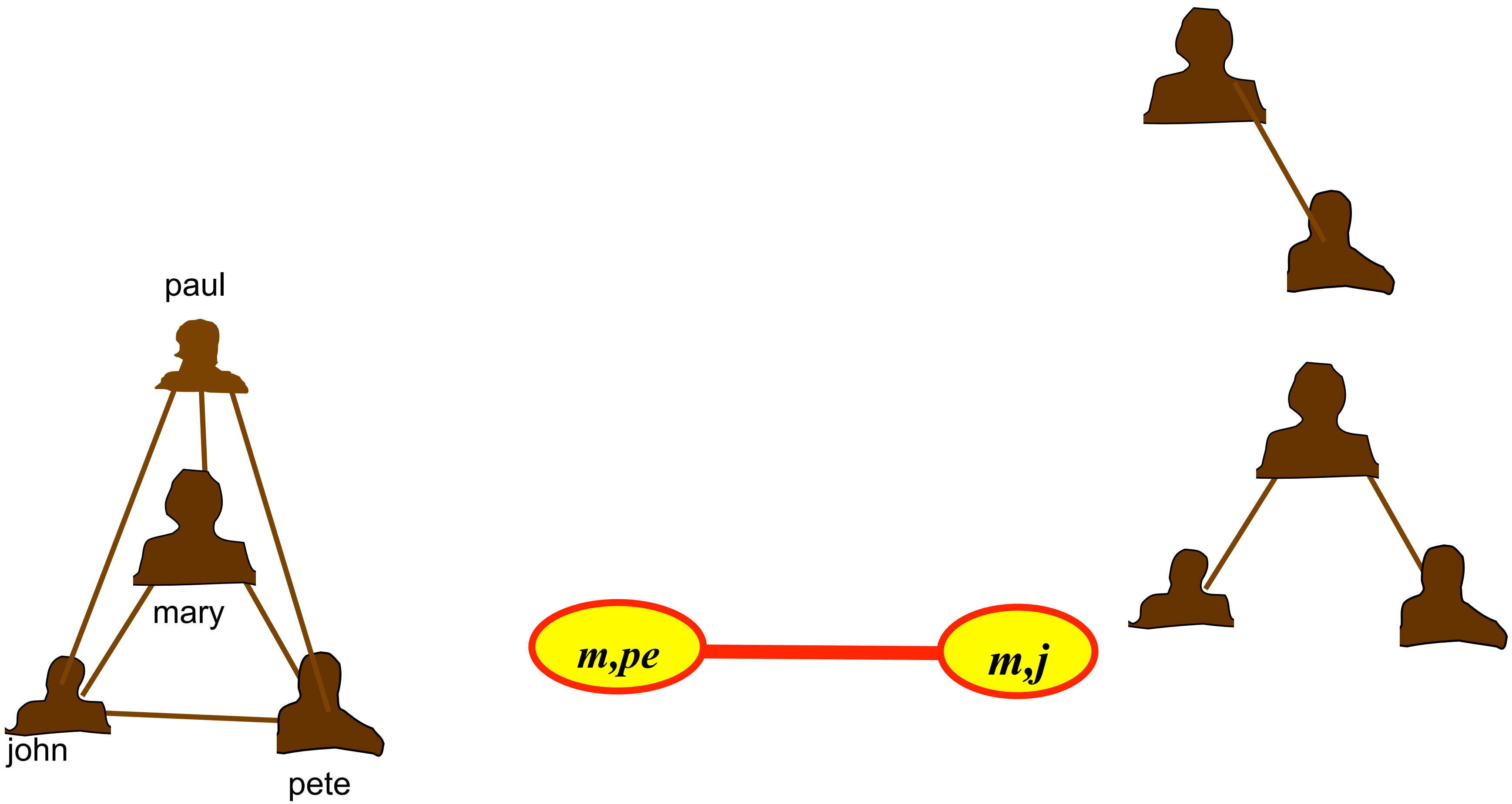
Deriving the ERGM: From Markov graph to Dependence graph



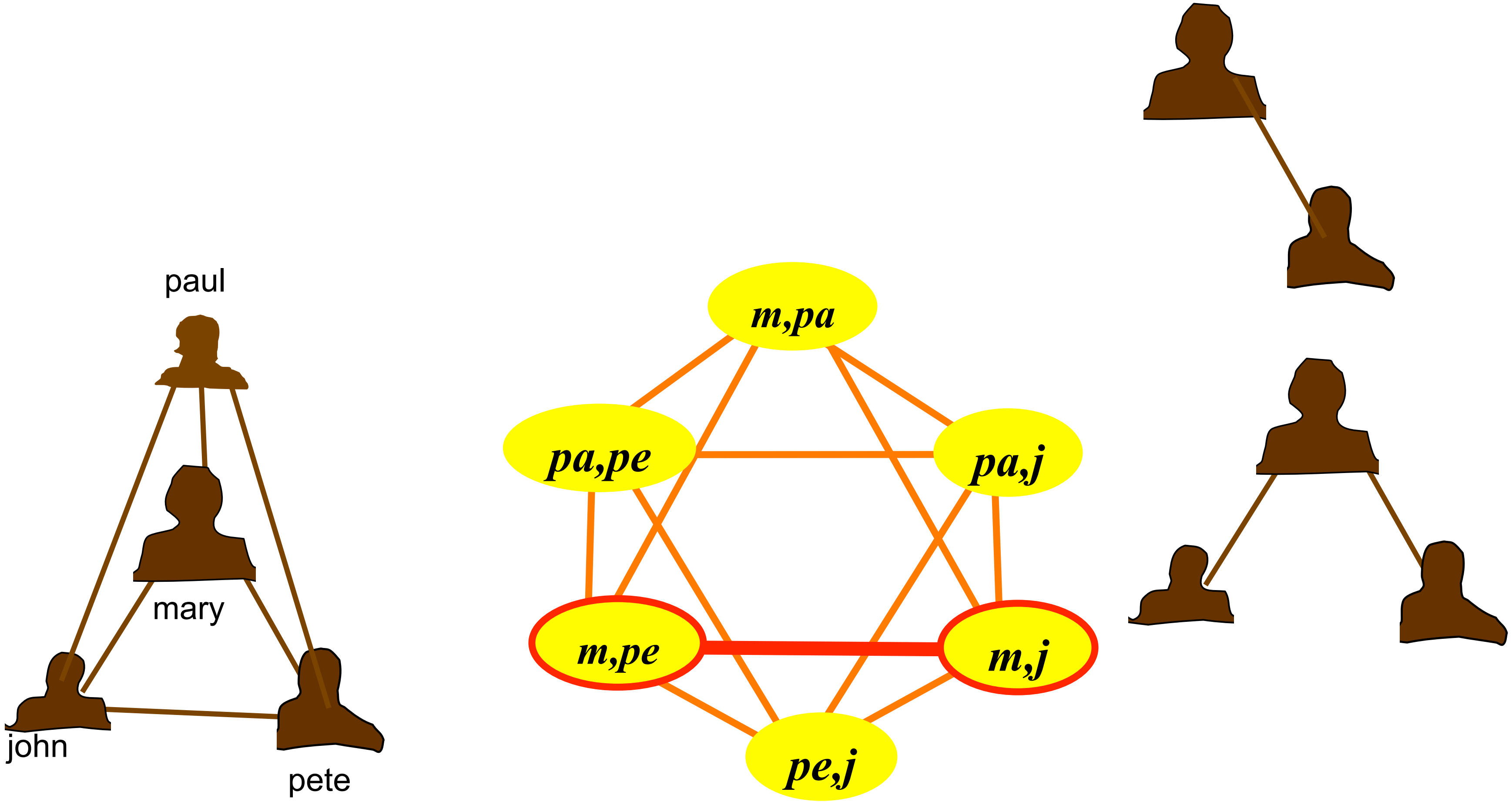




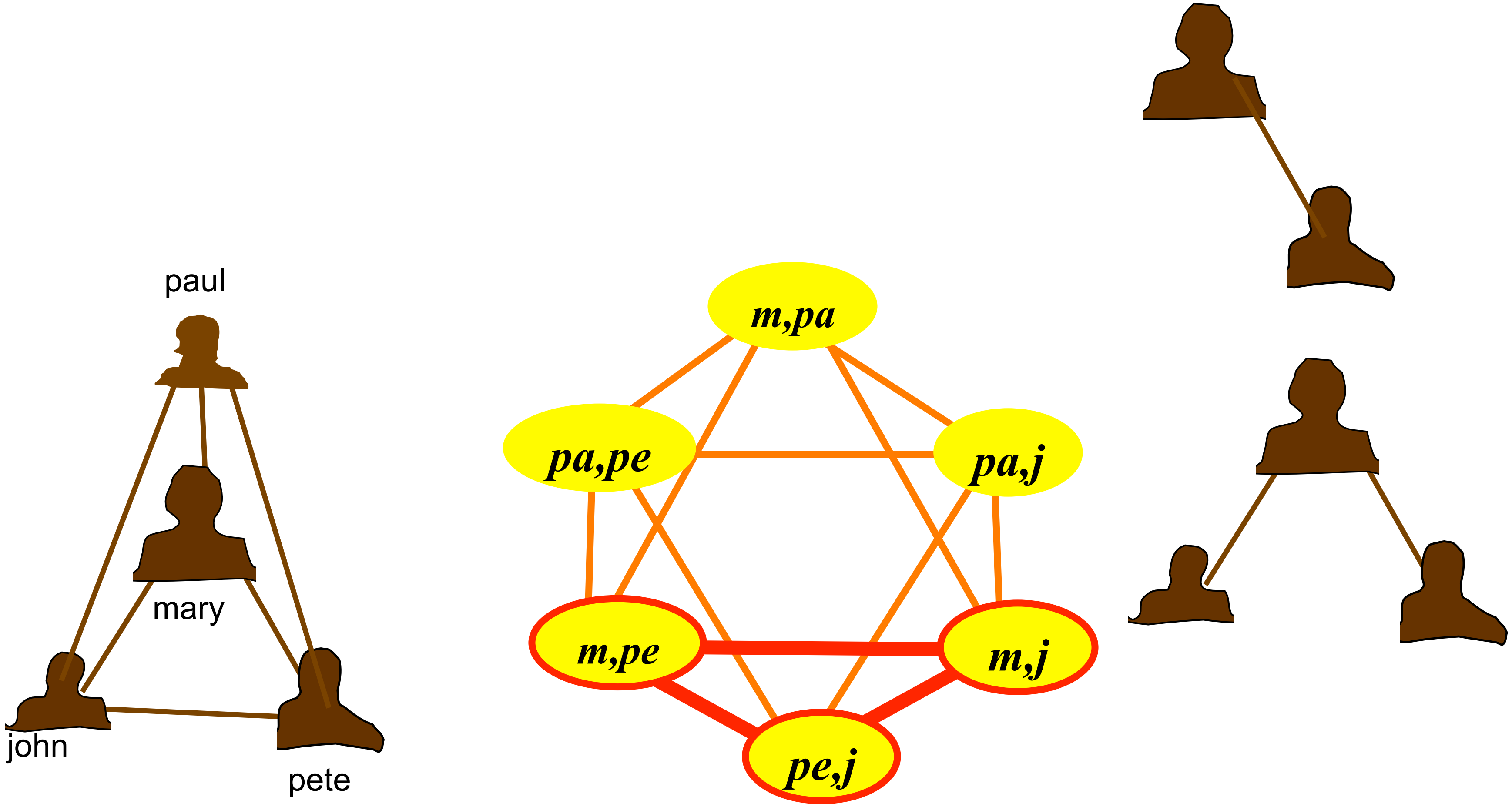
Deriving the ERGM: From Markov graph to Dependence graph



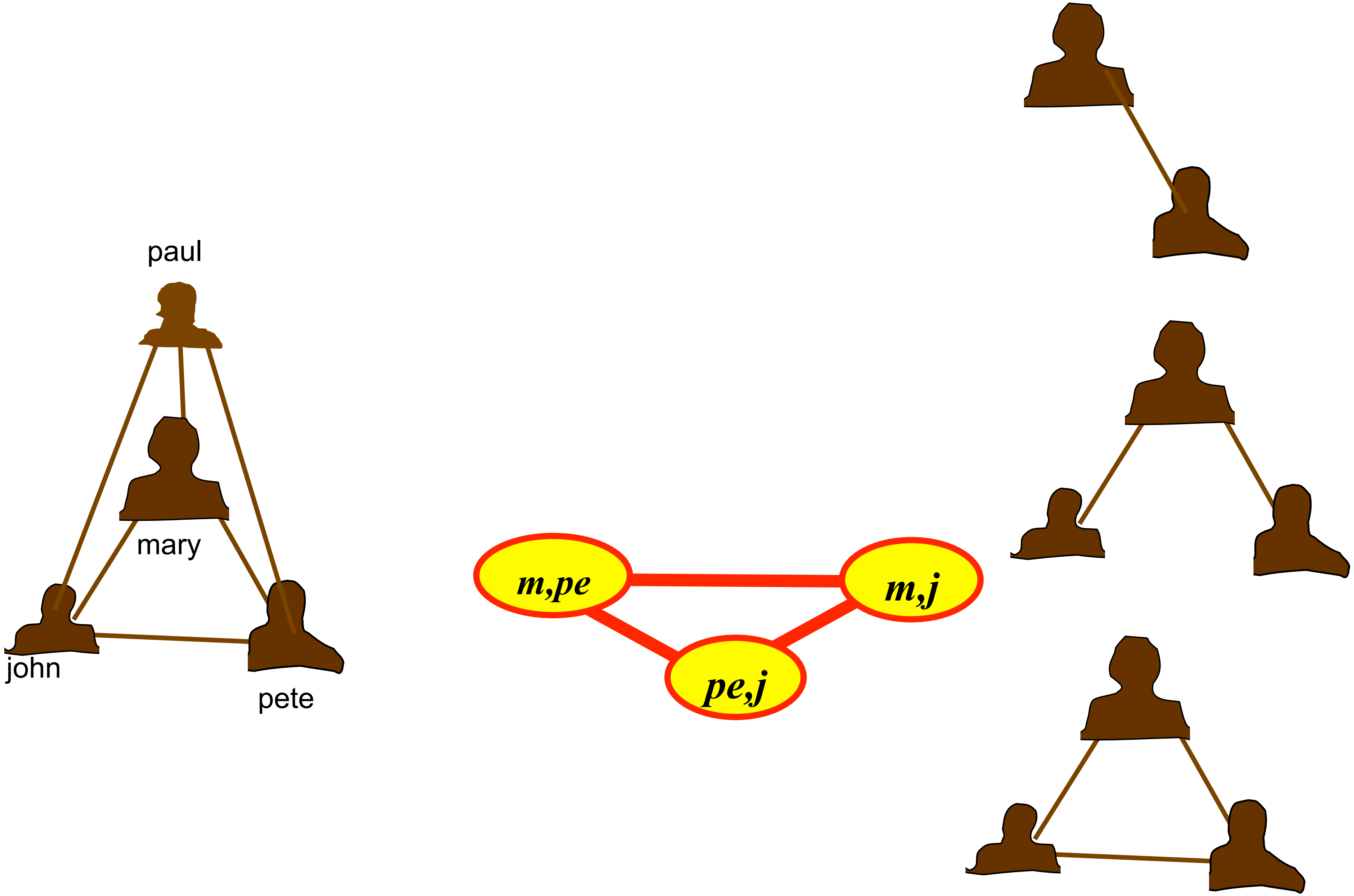
Deriving the ERGM: From Markov graph to Dependence graph



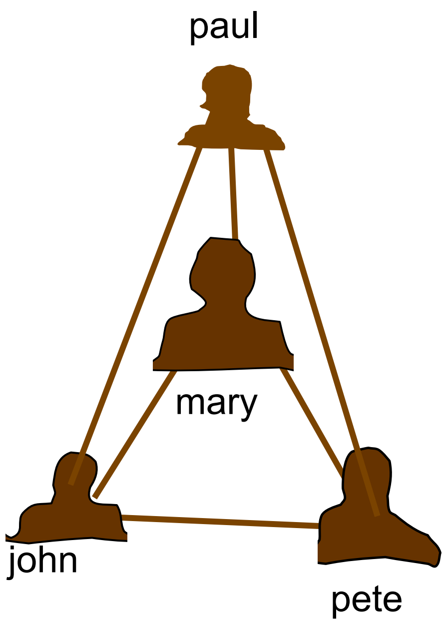
Deriving the ERGM: From Markov graph to Dependence graph

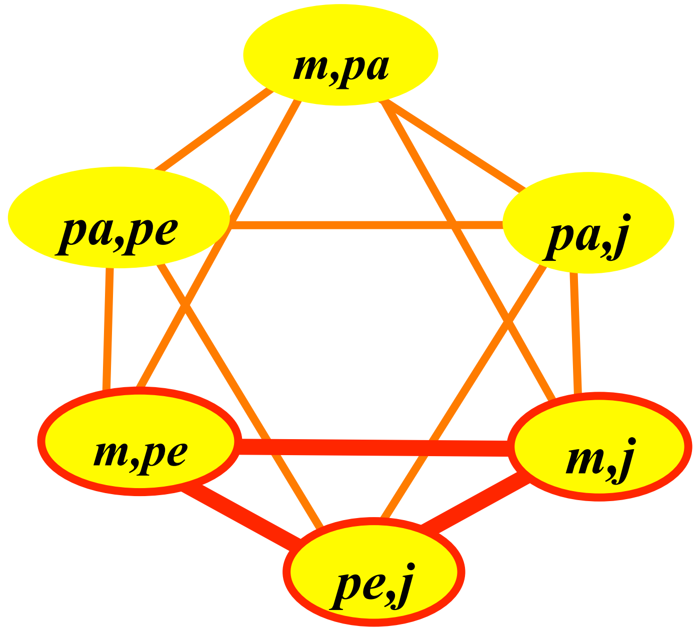


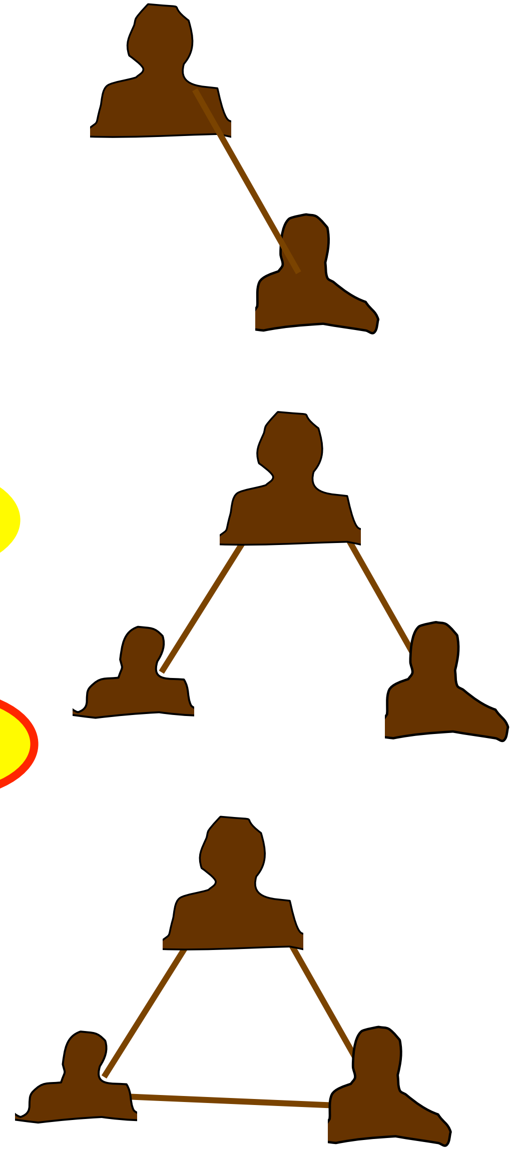
Deriving the ERGM: From Markov graph to Dependence graph



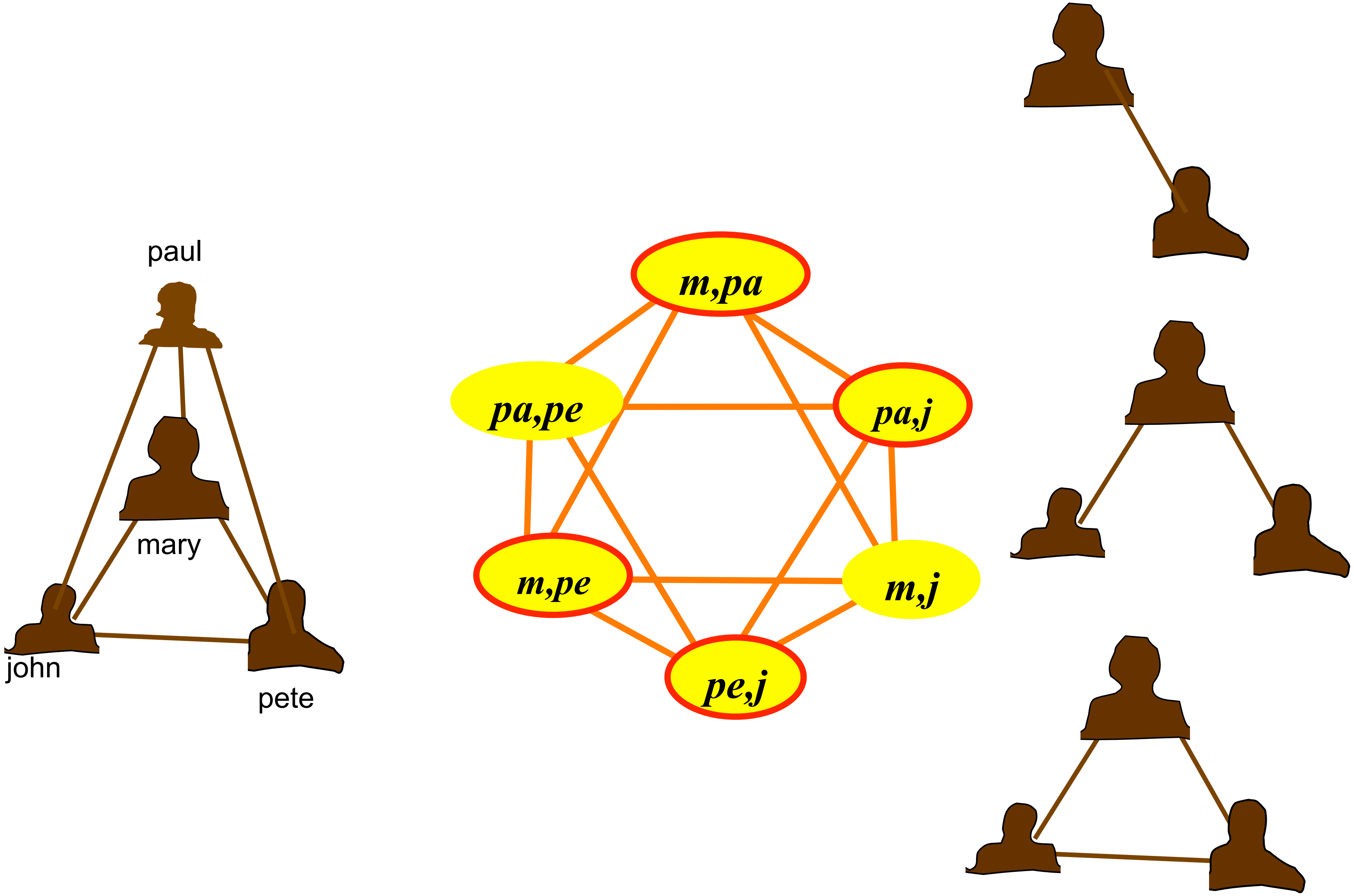
Deriving the ERGM: From Markov graph to Dependence graph



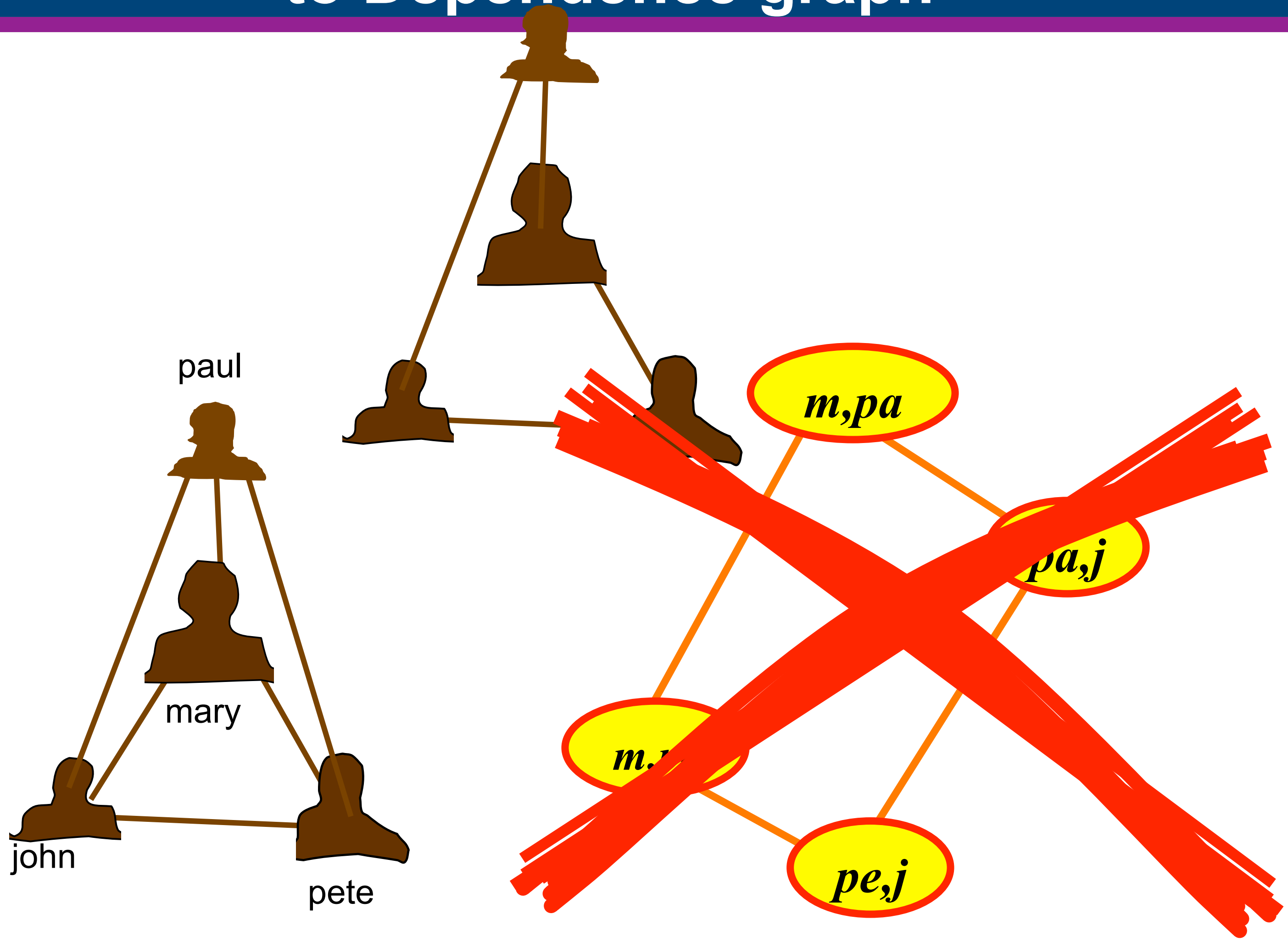




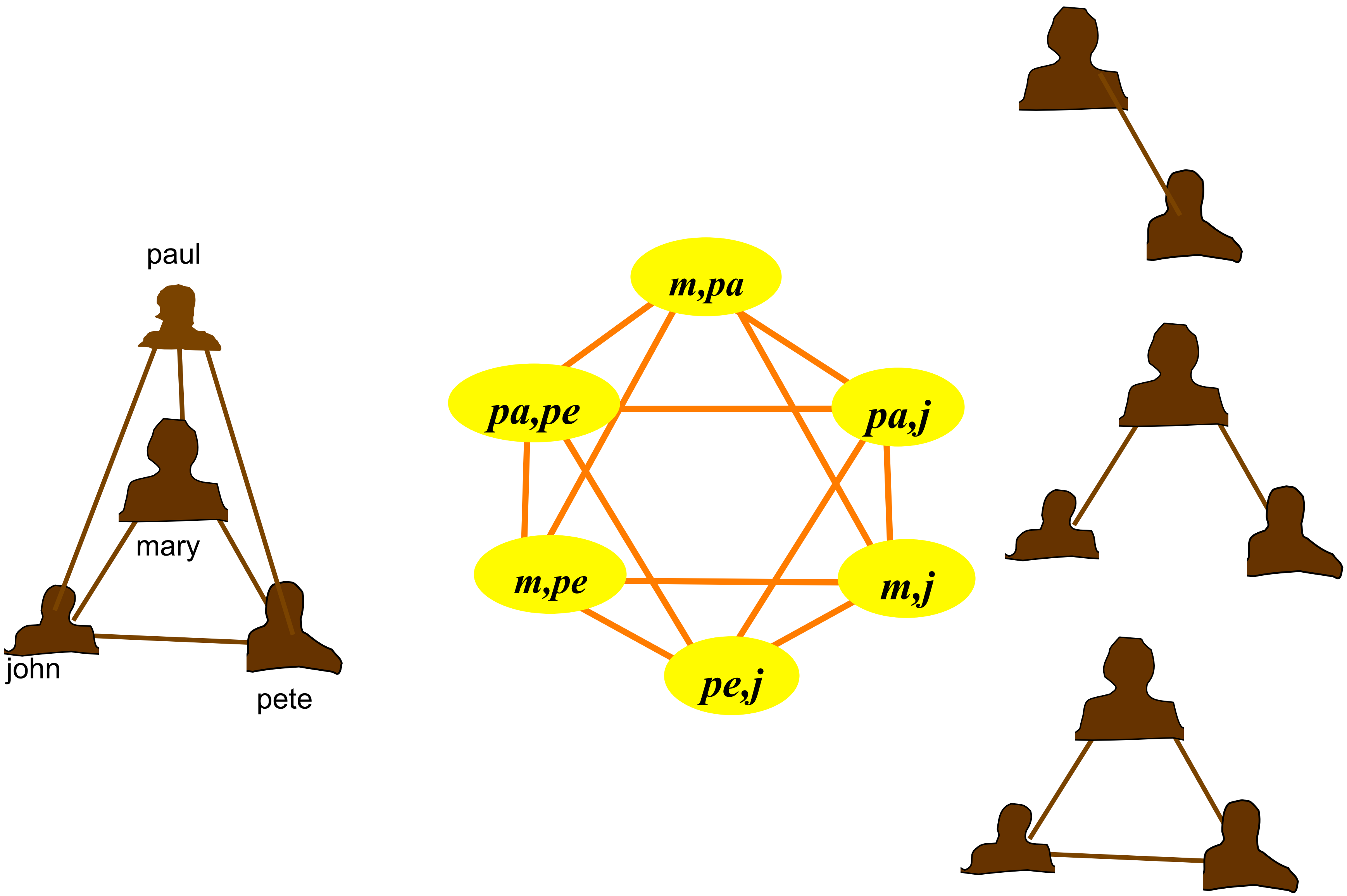
Deriving the ERGM: From Markov graph to Dependence graph



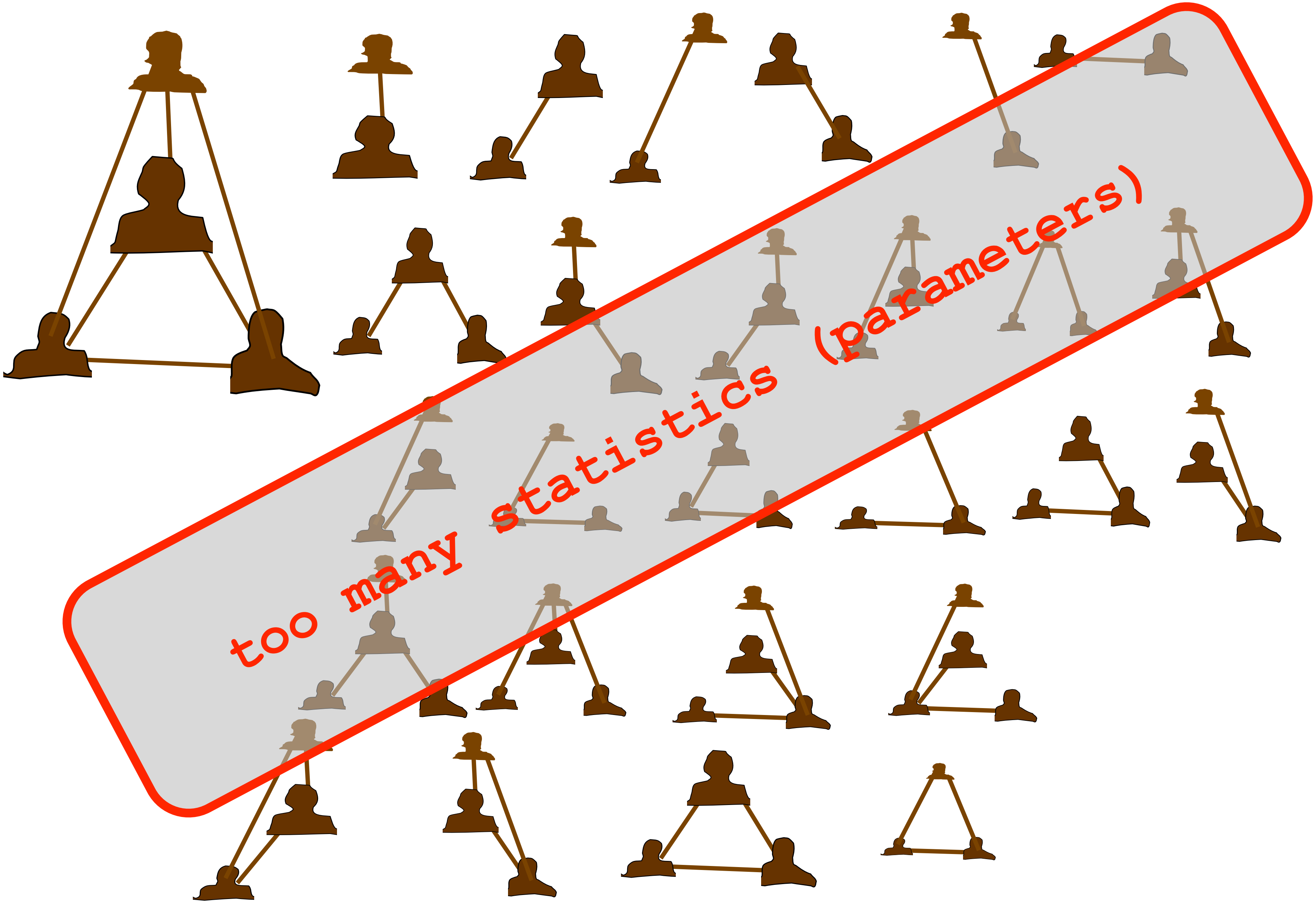
Deriving the ERGM: From Markov graph to Dependence graph



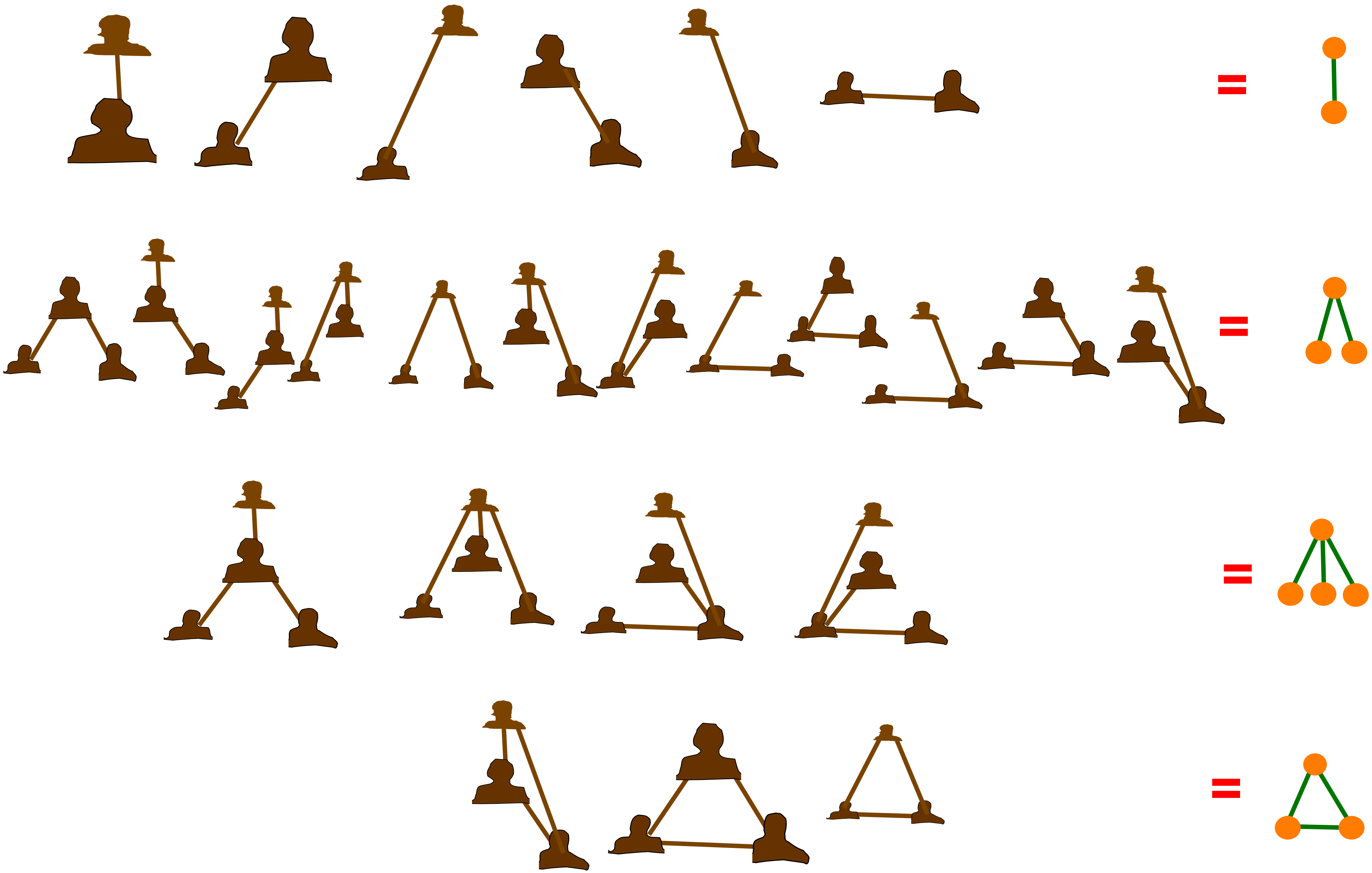
Deriving the ERGM: From Markov graph to Dependence graph



From Markov graph to Dependence graph – distinct subgraphs?

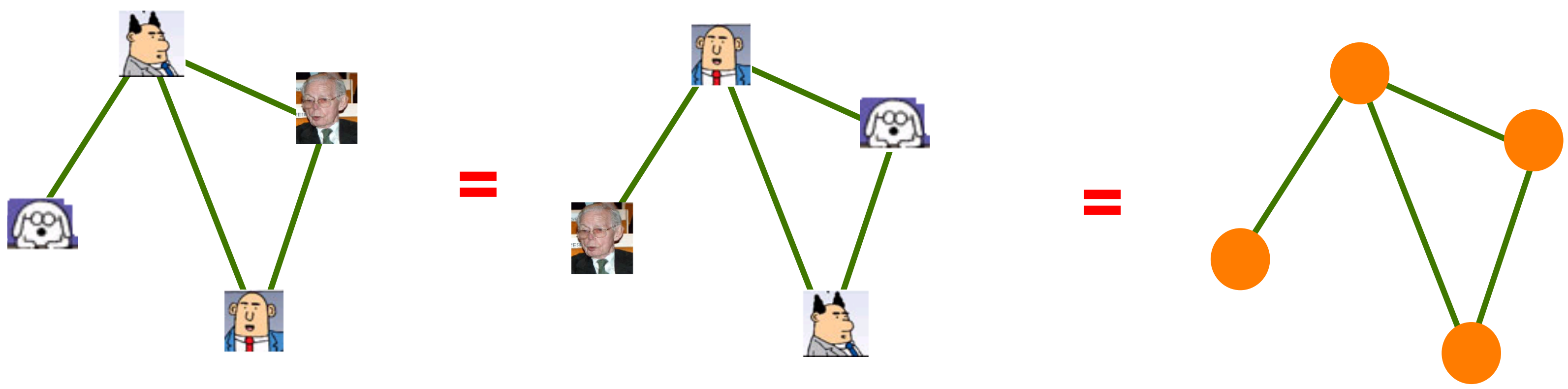


The homogeneity assumption



The homogeneity assumption

Interpretation: the probability of a graph depends only on the structure of the graph



A log-linear model (ERGM) for ties

”Aggregated” to a joint model for **entire adjacency matrix**

$$\log \Pr(X = x) = \theta_1 z_1(x) + \theta_2 z_2(x) + \cdots + \theta_p z_p(x) + \psi(\theta)$$

Interaction terms in log-linear model of types

$$X_{ij} \qquad X_{ij}X_{ik} \qquad \cdots \qquad X_{ij}X_{ik}X_{jk}$$

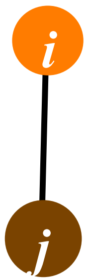
A log-linear model (ERGM) for ties

By definition of (in-) dependence

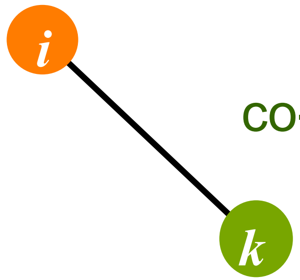
$$\Pr(X_{ij} = x_{ij}, X_{ik} = x_{ik}) \neq \Pr(X_{ij} = x_{ij}) \Pr(X_{ik} = x_{ik})$$

More than is explained by margins

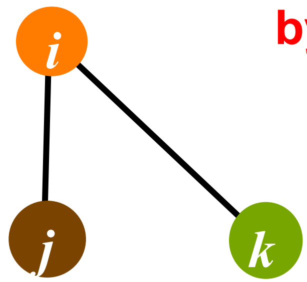
E.g.



and



co-occuring



X_{ij} X_{ik} $X_{ij}X_{ik}$

Main effects **interaction term**

Part 7

Summary of fitting routine

The steps of fitting an ERGM

- fit base-line model
- check convergence
- rerun if model not converged
- include more parameters? GOF
- candidate models

Part 8

Bipartite data

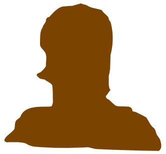
Bi-partite networks: cocitation (Small 1973)

MANCHESTER
1824
The University of Manchester

Bi-partite networks: cooffending (. Sarnecki, 2001)

Bi-partite networks

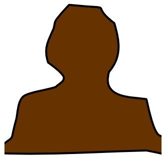
people



participating in

Social events
(Breiger, 1974)

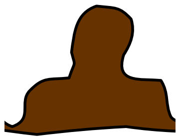
people



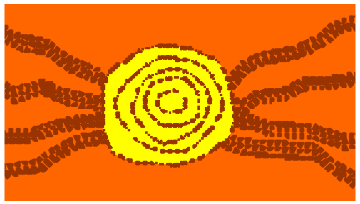
belonging to

voluntary organisations
(Bonachich, 1978)

directors



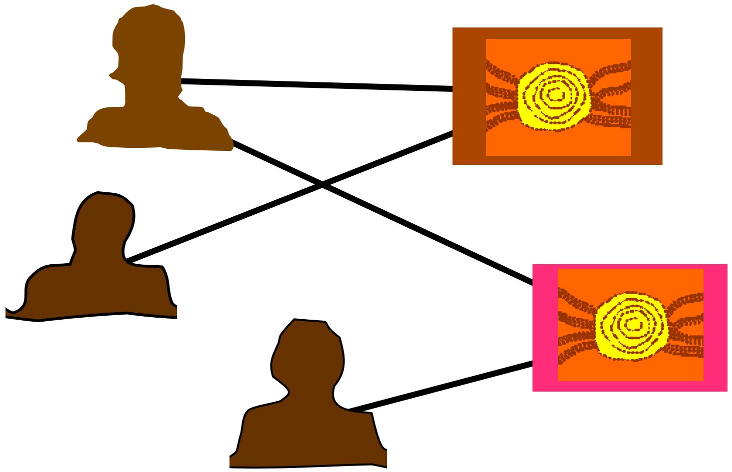
sitting on



corporate boards
(eg. Mizruchi, 1982

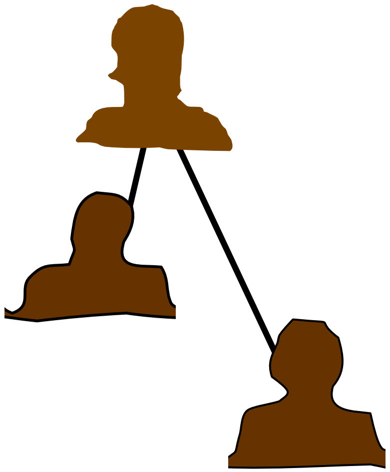
One-mode projection

Two-mode



Tie: If two directors share a board

one-mode

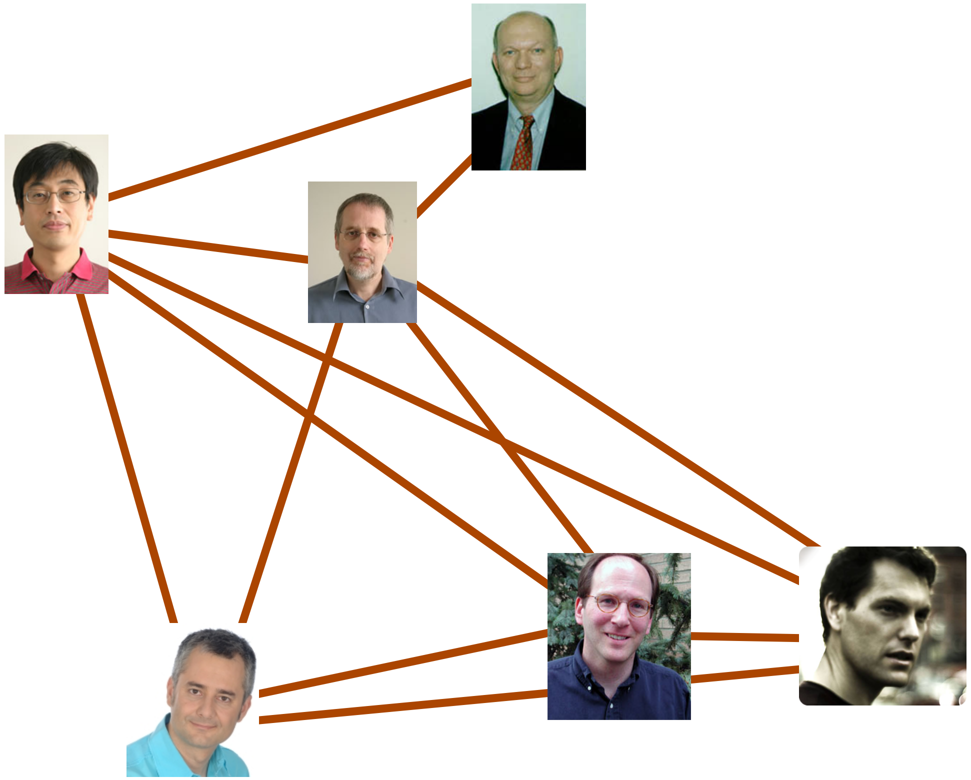


MANCHESTER 1824
The University of Manchester

Bi-partite networks

Bi-partite networks

Researchers

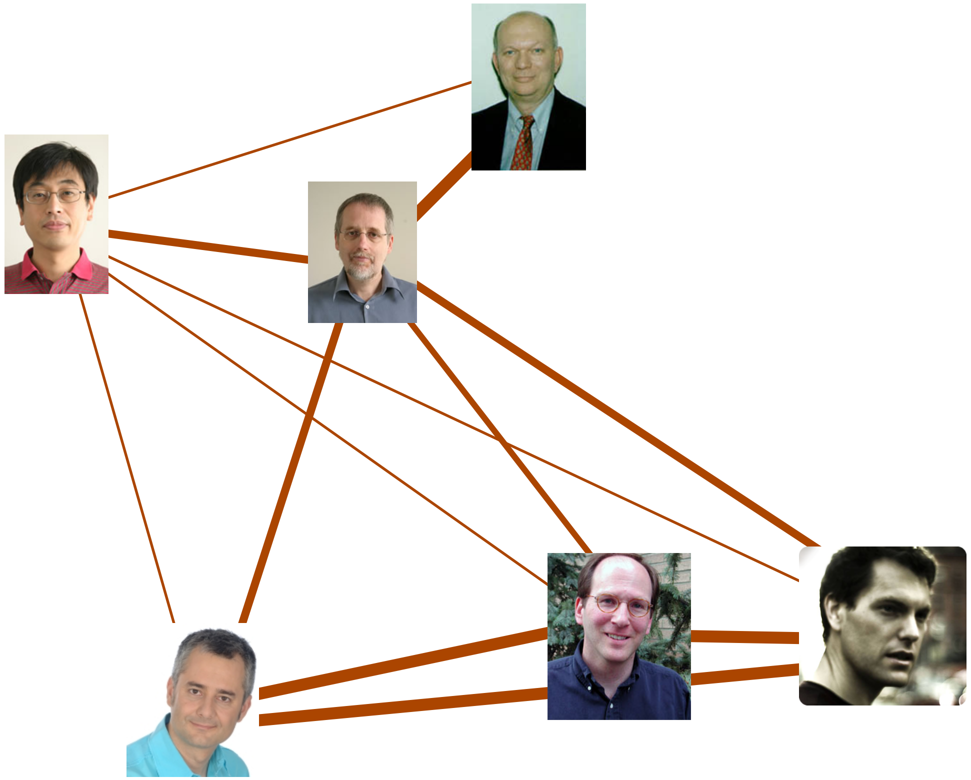


texts

cite

Bi-partite networks

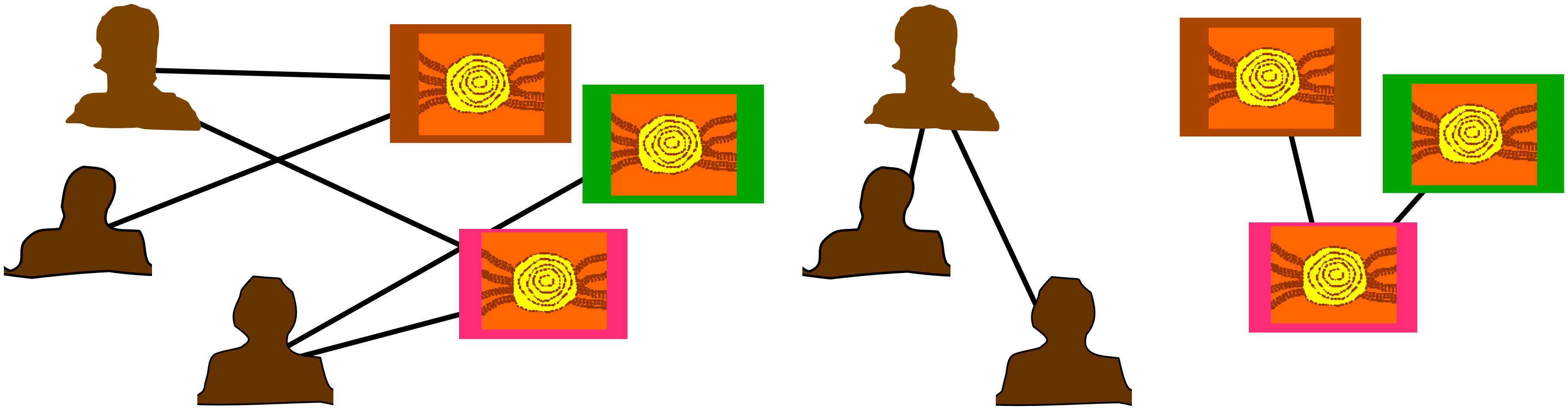
Researchers



texts

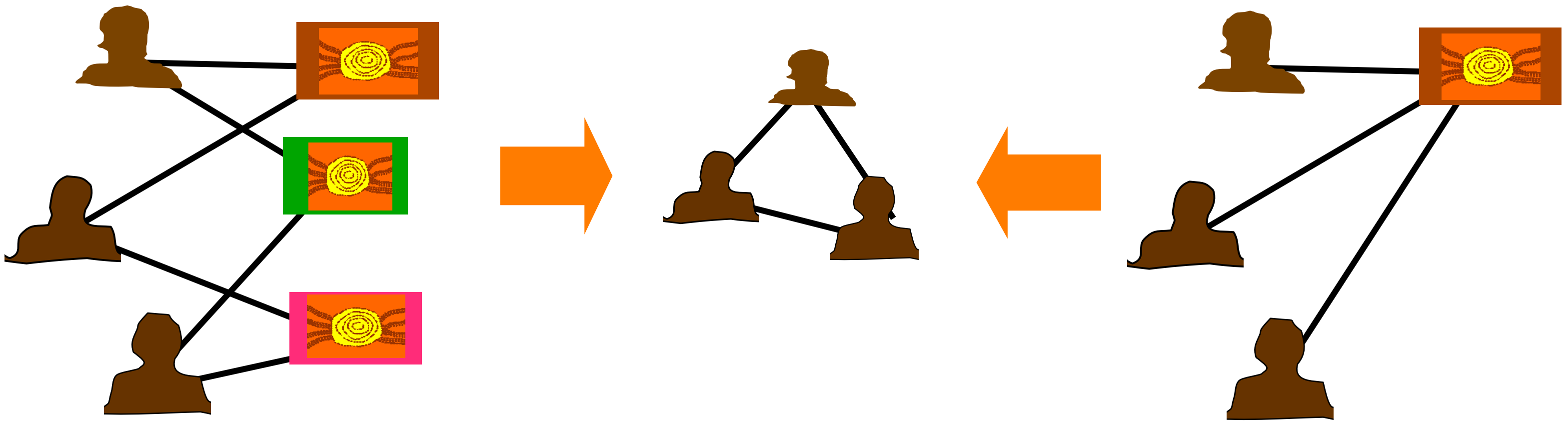
cite

One-mode projection



What mode given priority? (Duality of social actors and social groups; e.g. Breiger 1974; Breiger and Pattison, 1986)

Loss of information

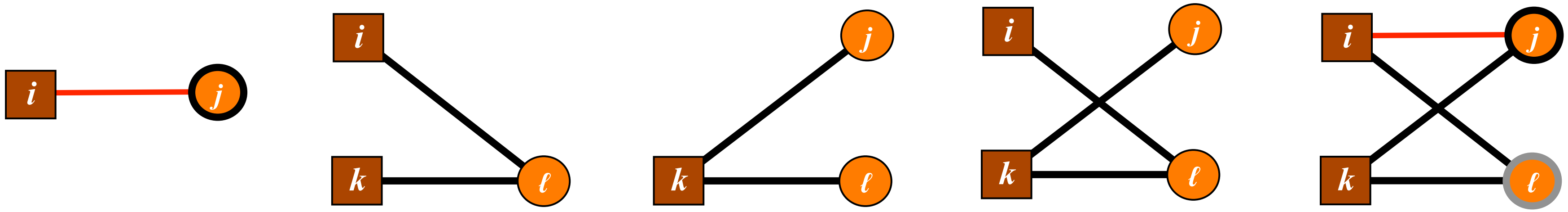


MANCHESTER
1824
The University of Manchester

ERGM for bipartite networks

The model is the same as for one-mode networks (Wang et al., 2007)

$$\log \Pr(X = x) = \theta_1 z_1(x) + \theta_2 z_2(x) + \cdots + \theta_p z_p(x) + \psi(\theta)$$



Edges

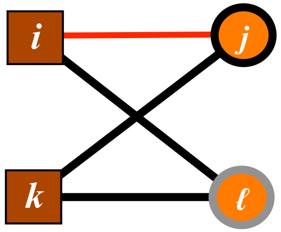
"people" 2-stars

"affiliation" 2-stars

3-paths

4-cycles

ERGM for bipartite networks



4-cycles

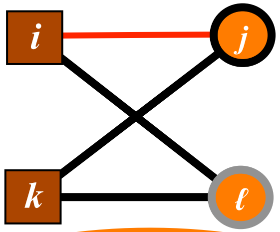
A form of bipartite clustering

Also known as Bi-cliques

One possible interpretation:

MANCHESTER 1824
The University of Manchester

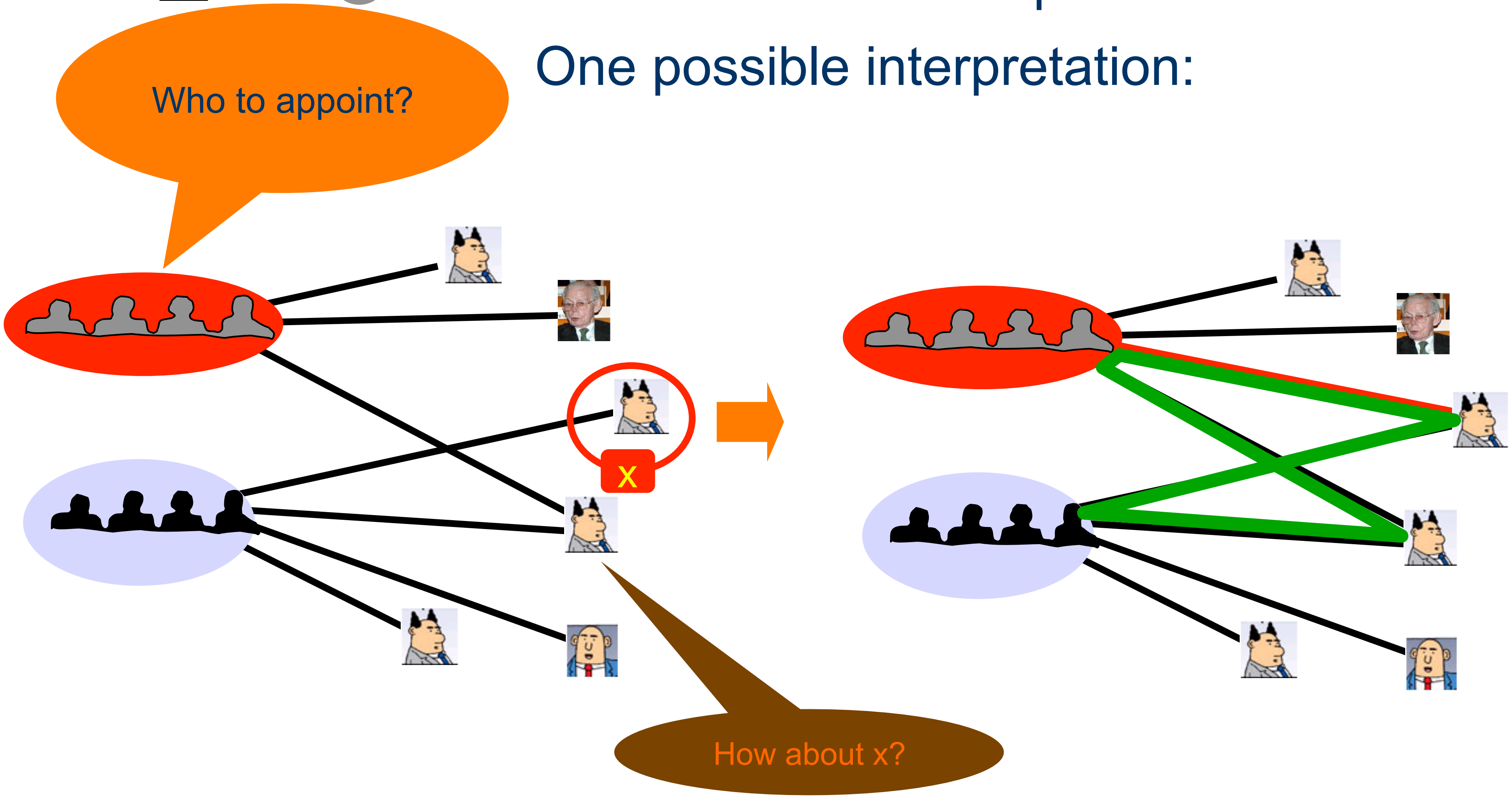
ERGM for bipartite networks



A form of bipartite clustering

Also known as Bi-cliques

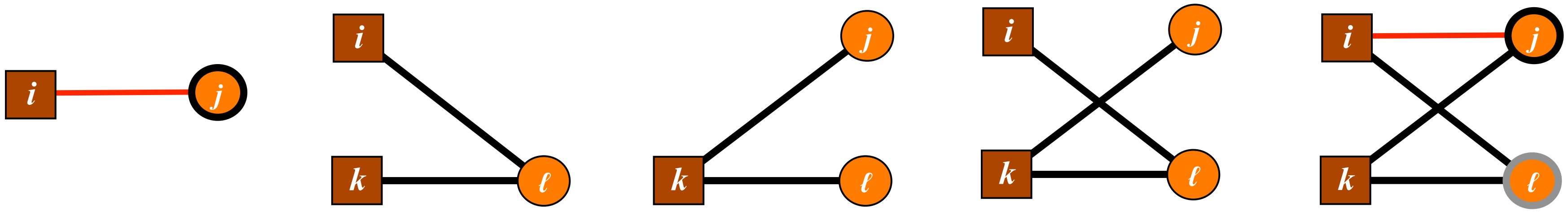
One possible interpretation:



ERGM for bipartite networks

Fitting the model in (B)Pnet straightforward extension

These statistics are **all Markov**:



Edges

”people” 2-stars

”affiliation” 2-stars

3-paths

4-cycles

<BPNet>

Part 9

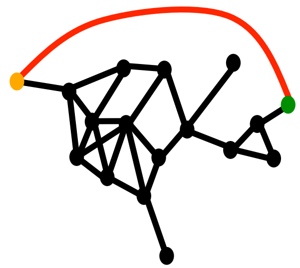
Missing data

Methods for missing network data

Effects of missingness

Perils

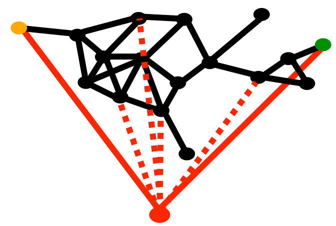
Some investigations on the effects on indices of structural properties (Kossinets, 2006; Costenbader & Valente, 2003; Huisman, 2007)



Problems with the “boundary specification issue”

Few remedies

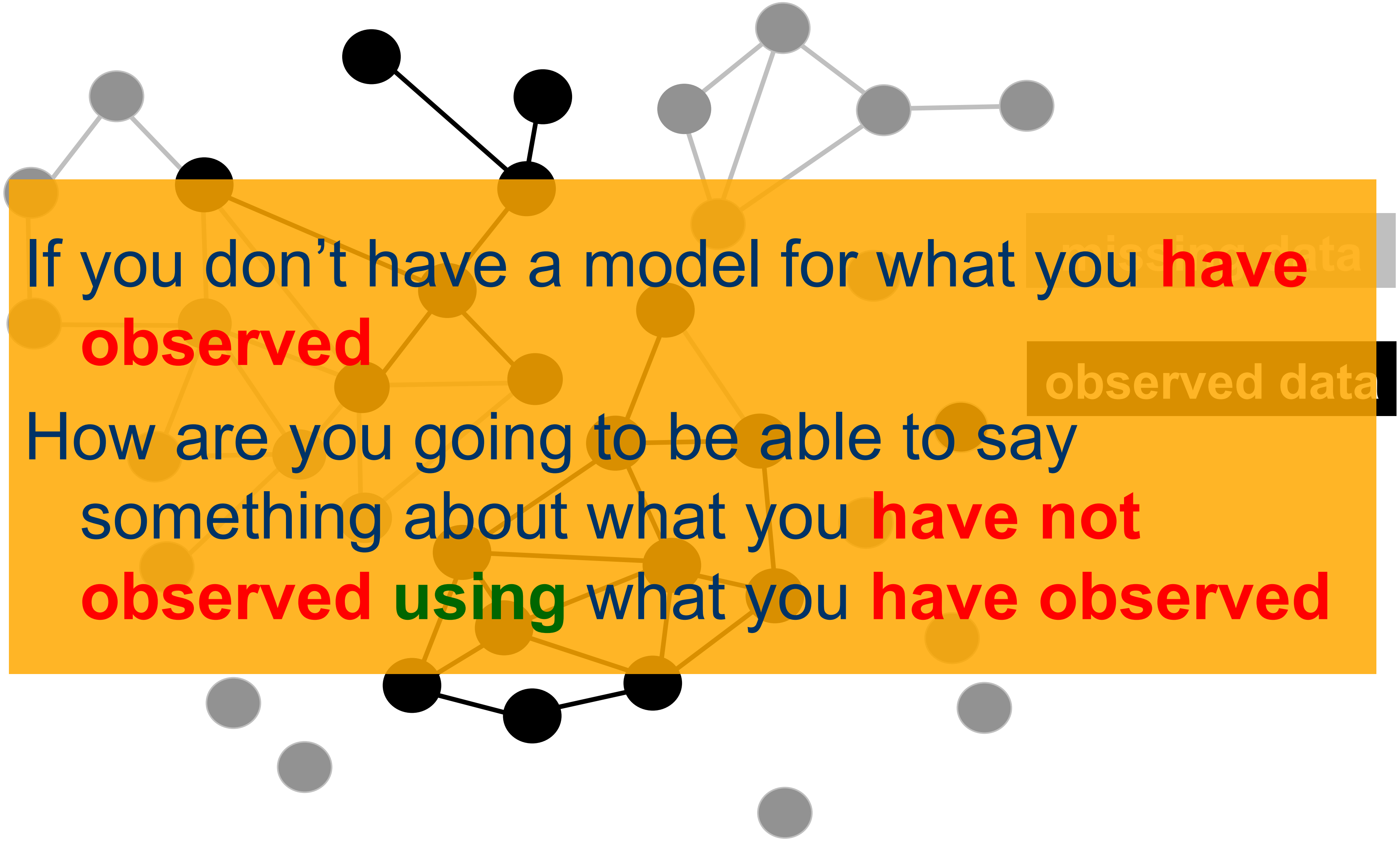
Deterministically “complement” data (Stork & Richards, 1992)



Stochastically Impute missing data (Huisman, 2007)

Ad-hoc “likelihood” (score) for missing ties (Liben-Nowell and Kleinberg, 2007)

Model assisted treatment of missing network data



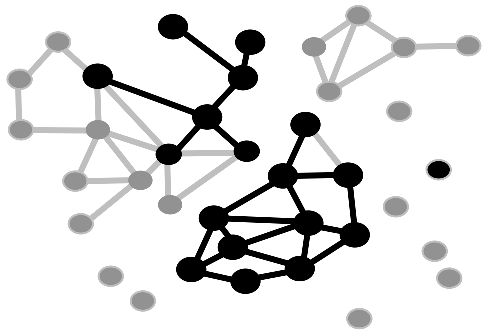
If you don’t have a model for what you **have observed**

How are you going to be able to say something about what you **have not observed using** what you **have observed**

Model assisted treatment of missing network data

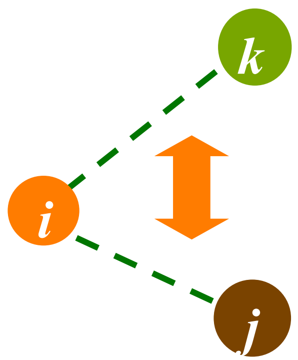
- Importance sampling (**Handcock & Gile 2010**; **Koskinen, Robins & Pattison, 2010**)
- Stochastic approximation and the missing data principle (**Orchard & Woodbury,1972**) (**Koskinen & Snijders, forthcoming**)
- Bayesian data augmentation (**Koskinen, Robins & Pattison, 2010**)

Marginalisation (Snijders, 2010; Koskinen et al, 2010)

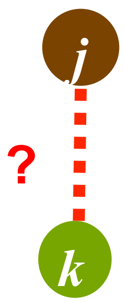


Subgraph of ERGM not ERGM

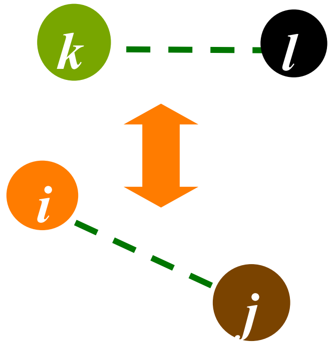
Dependence in ERGM



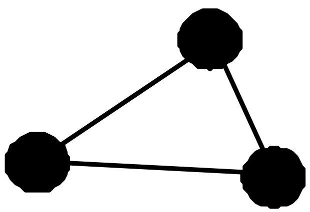
But if

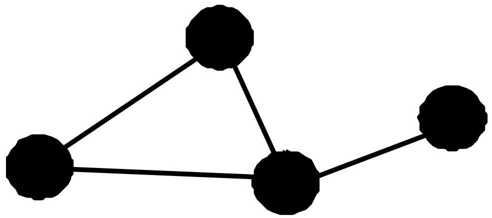


We may also have dependence



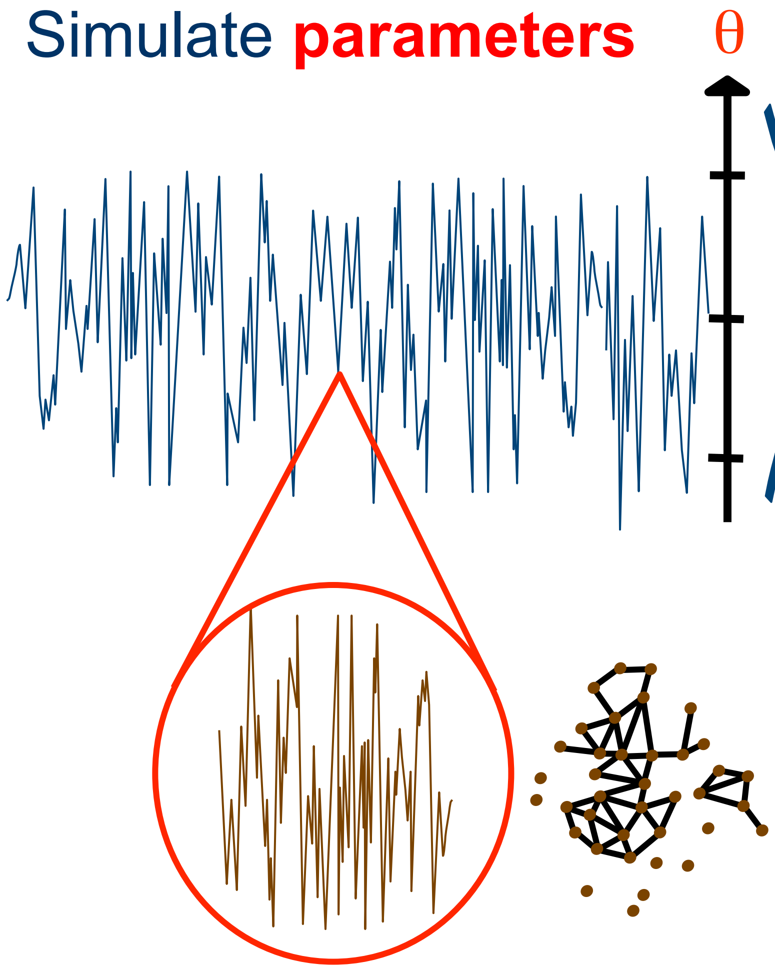
We should include counts of:



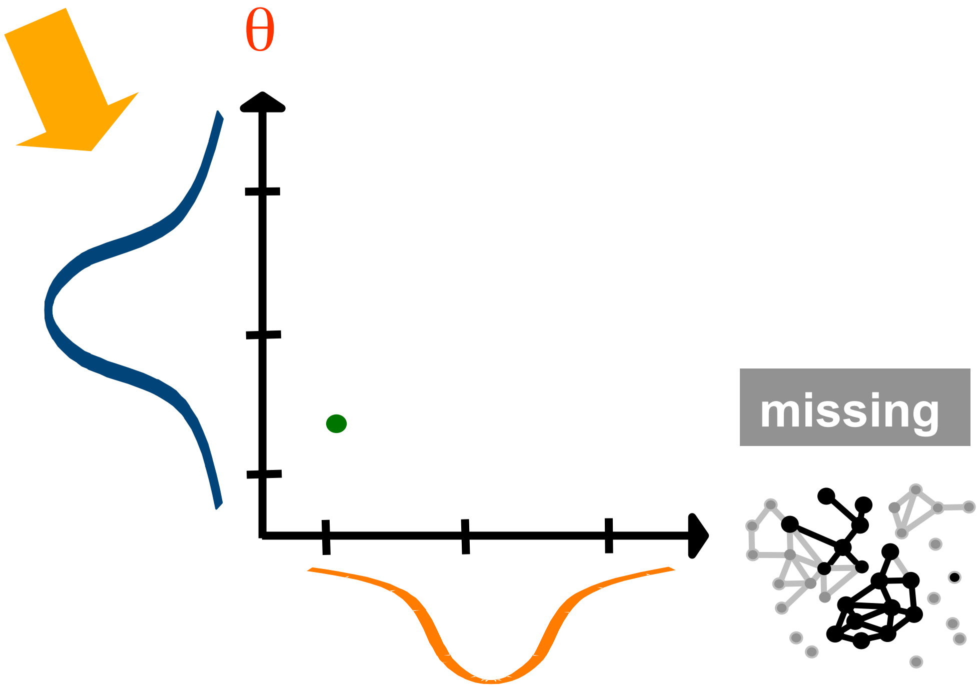


Bayesian Data Augmentation

Simulate **parameters**



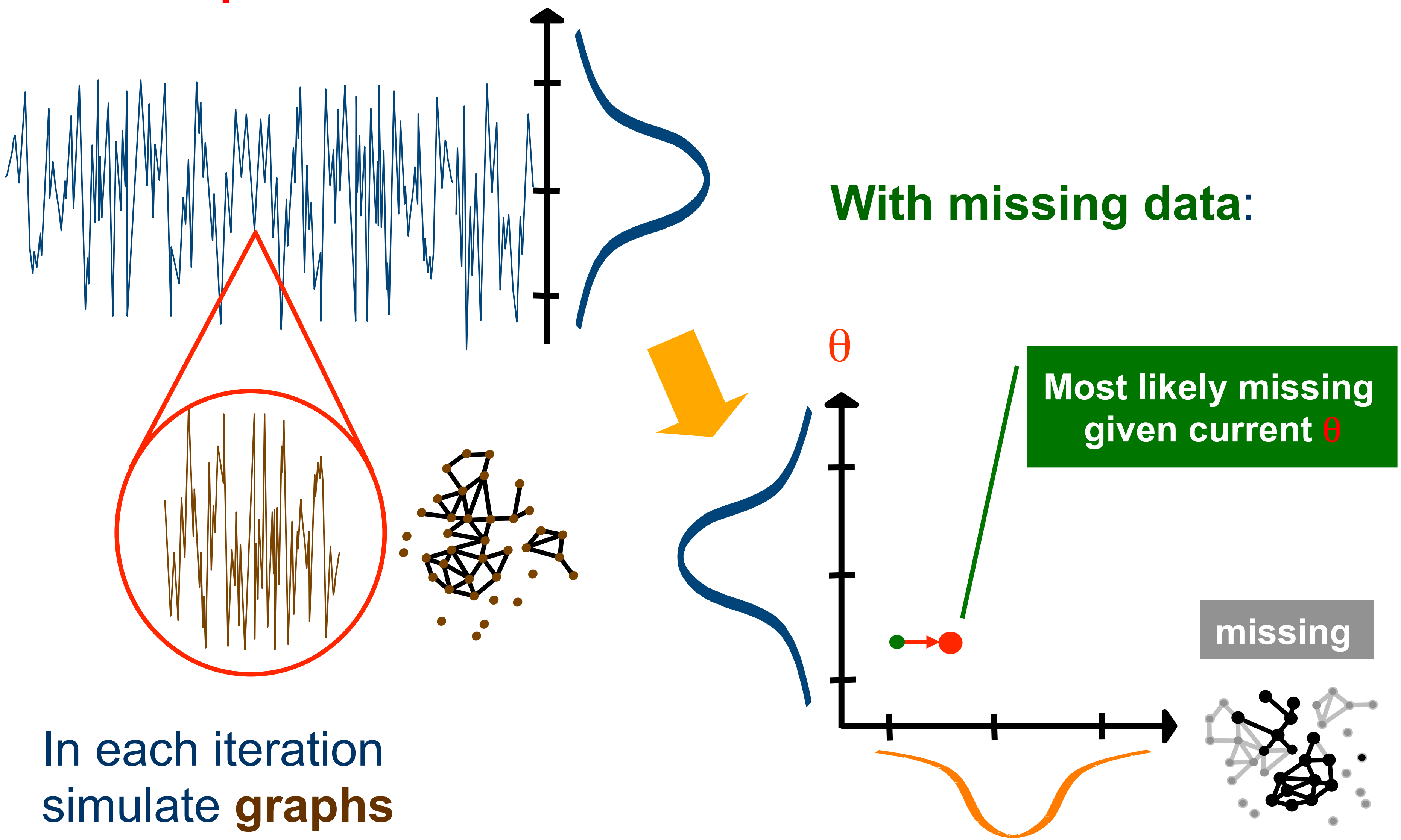
With missing data:



In each iteration
simulate **graphs**

Bayesian Data Augmentation

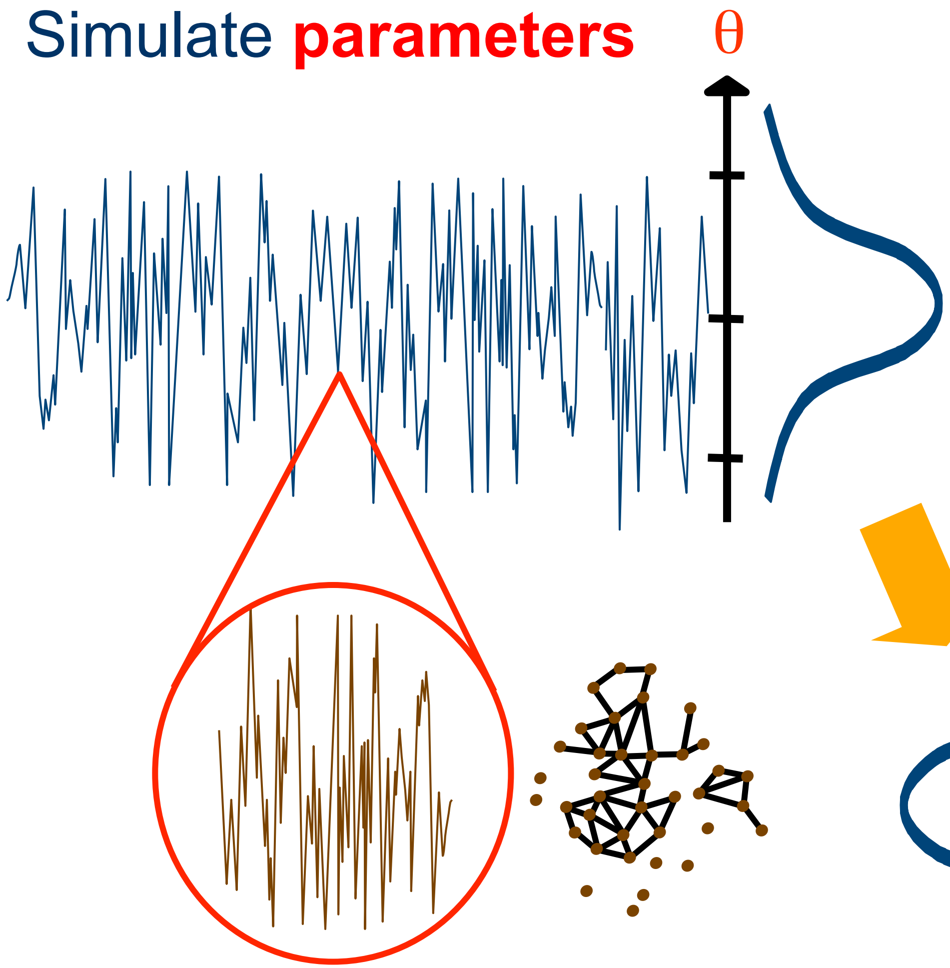
Simulate **parameters** θ



In each iteration simulate **graphs**

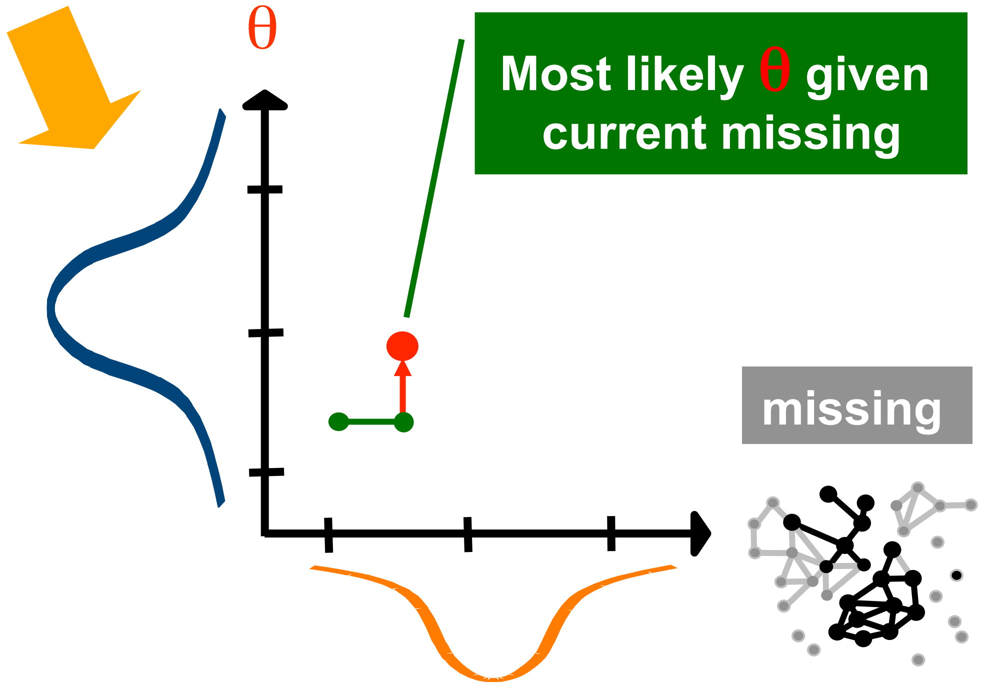
Bayesian Data Augmentation

Simulate **parameters**



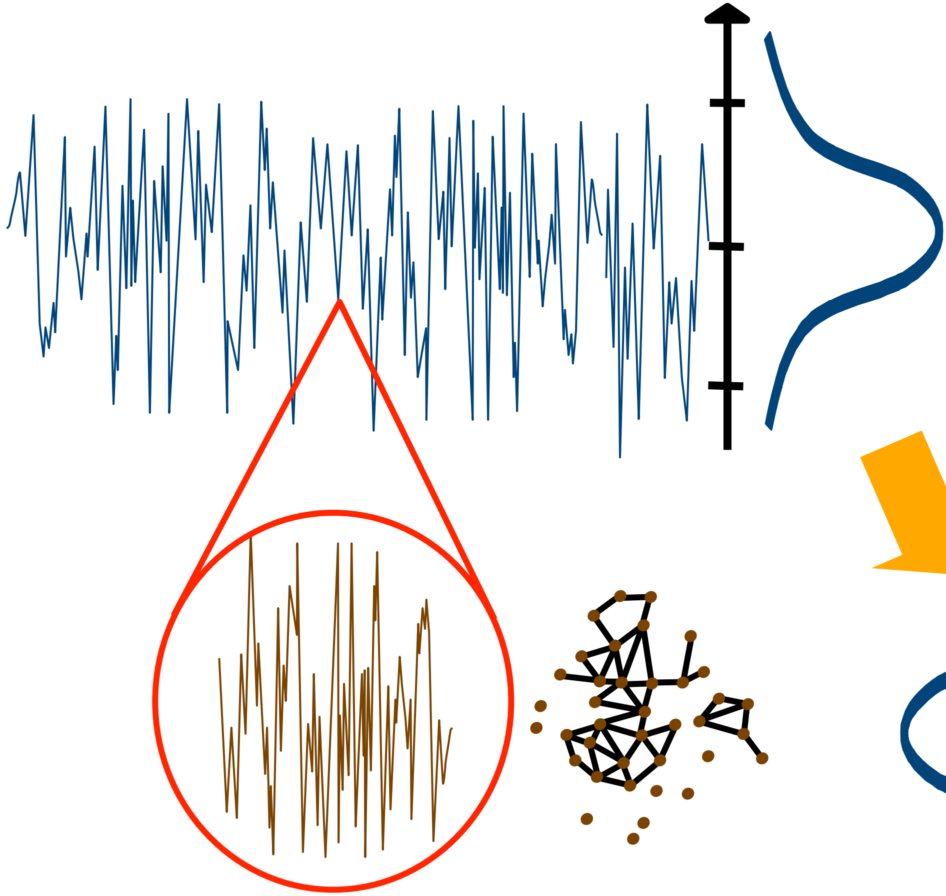
In each iteration
simulate **graphs**

With missing data:

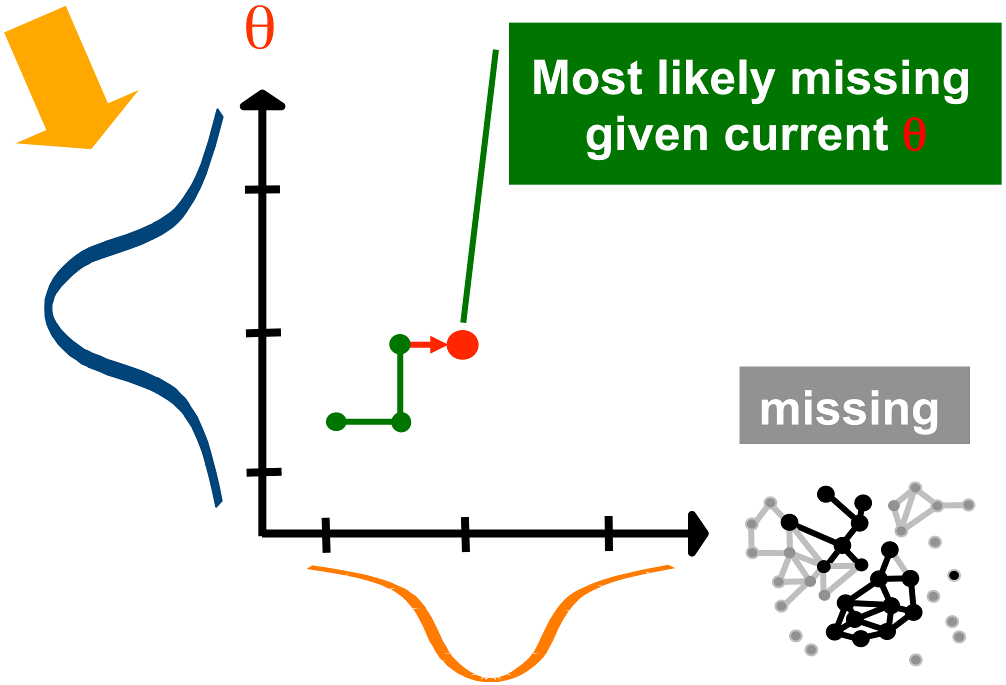


Bayesian Data Augmentation

Simulate **parameters** θ



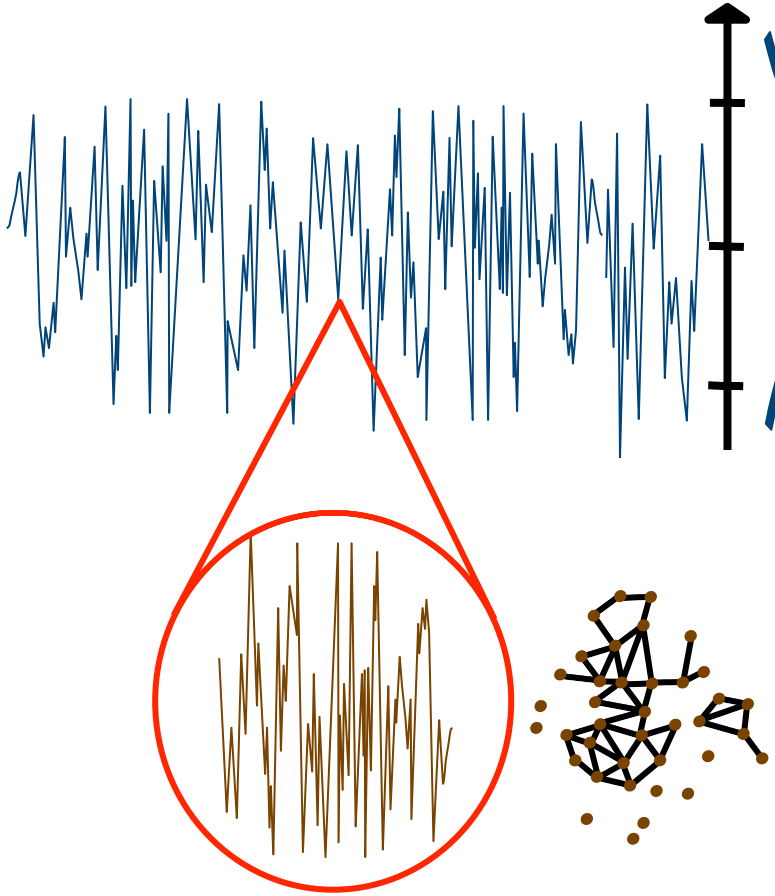
With missing data:



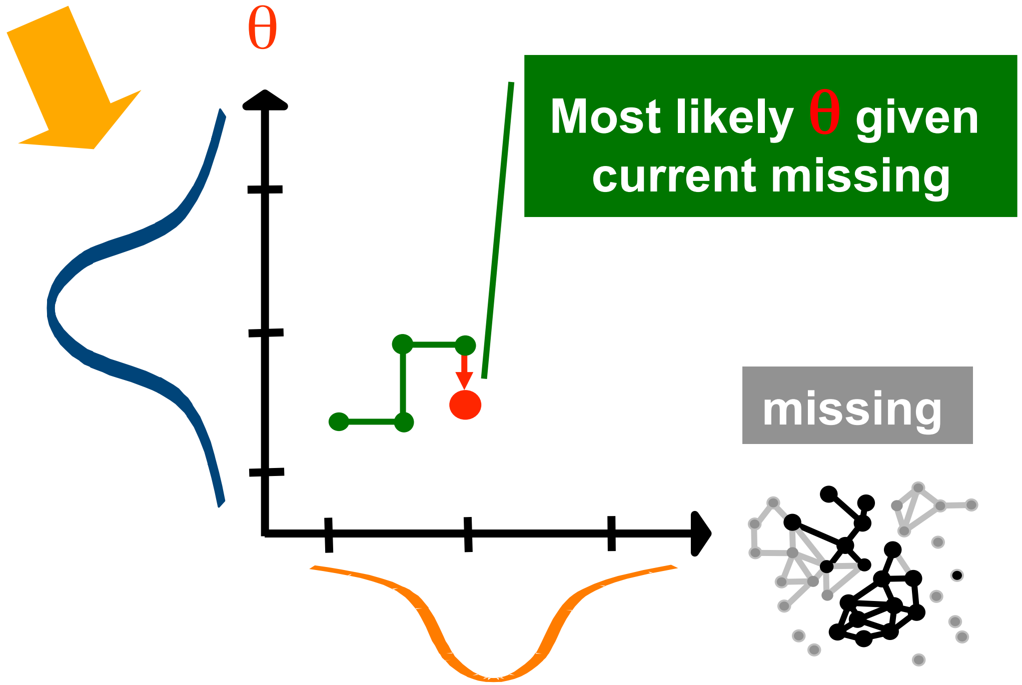
In each iteration simulate **graphs**

Bayesian Data Augmentation

Simulate **parameters** θ



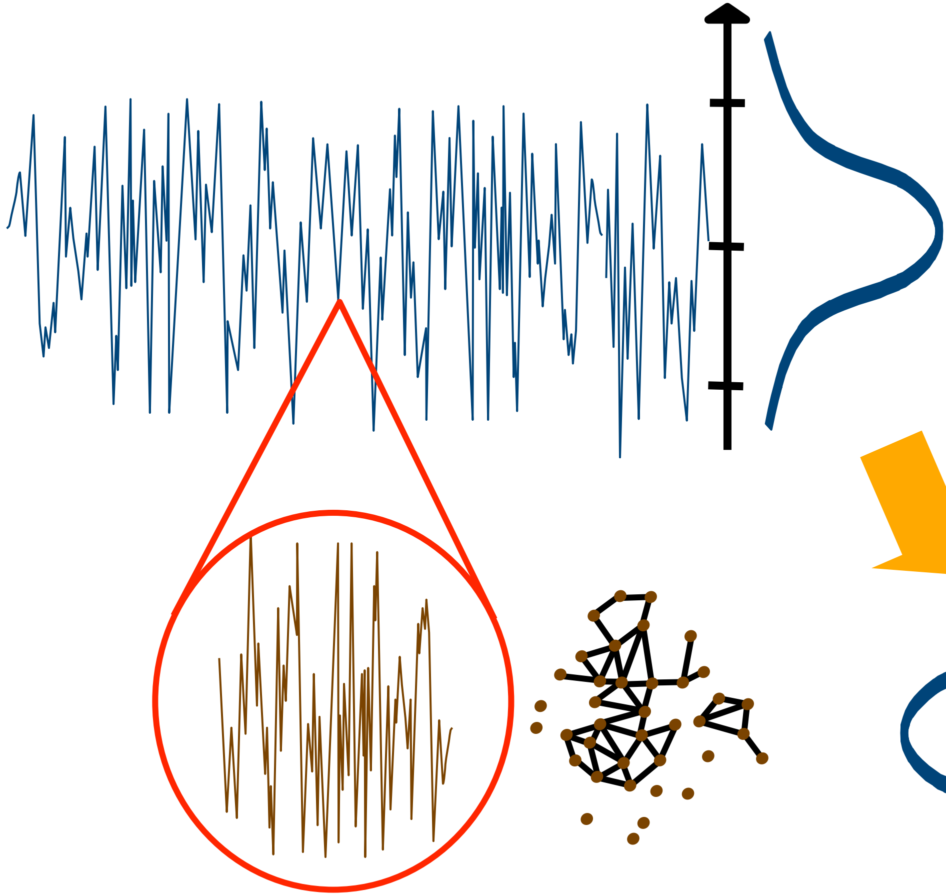
With missing data:



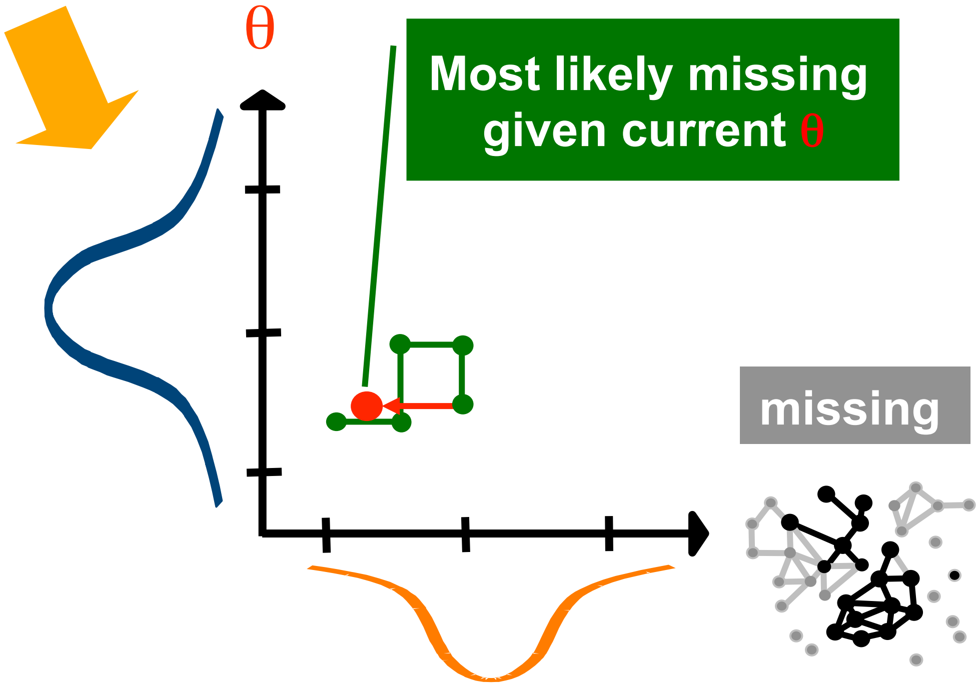
In each iteration simulate **graphs**

Bayesian Data Augmentation

Simulate **parameters** θ



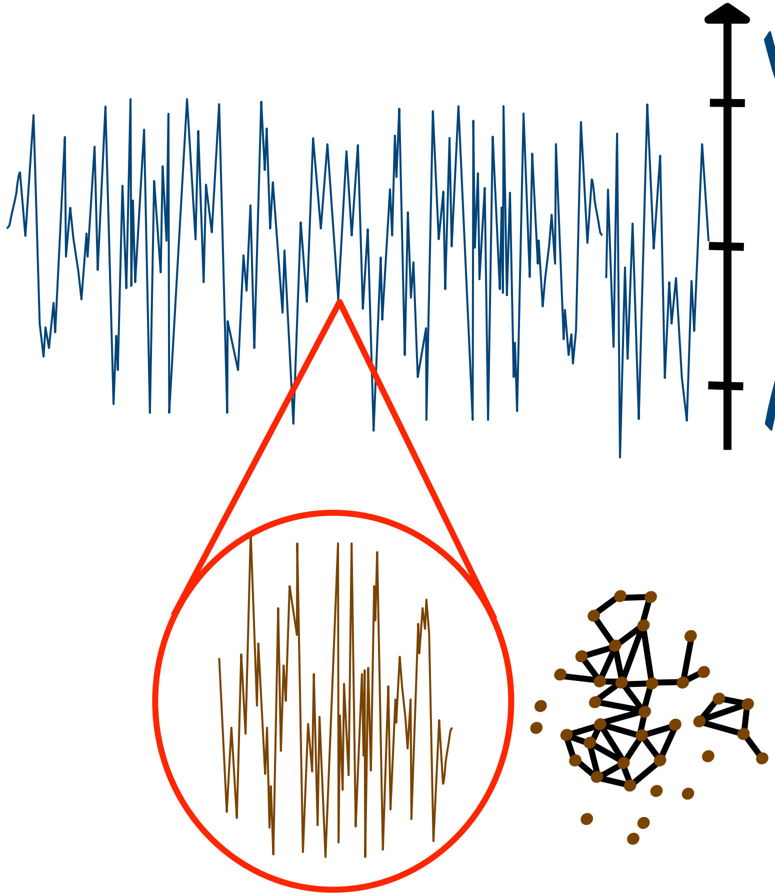
With missing data:



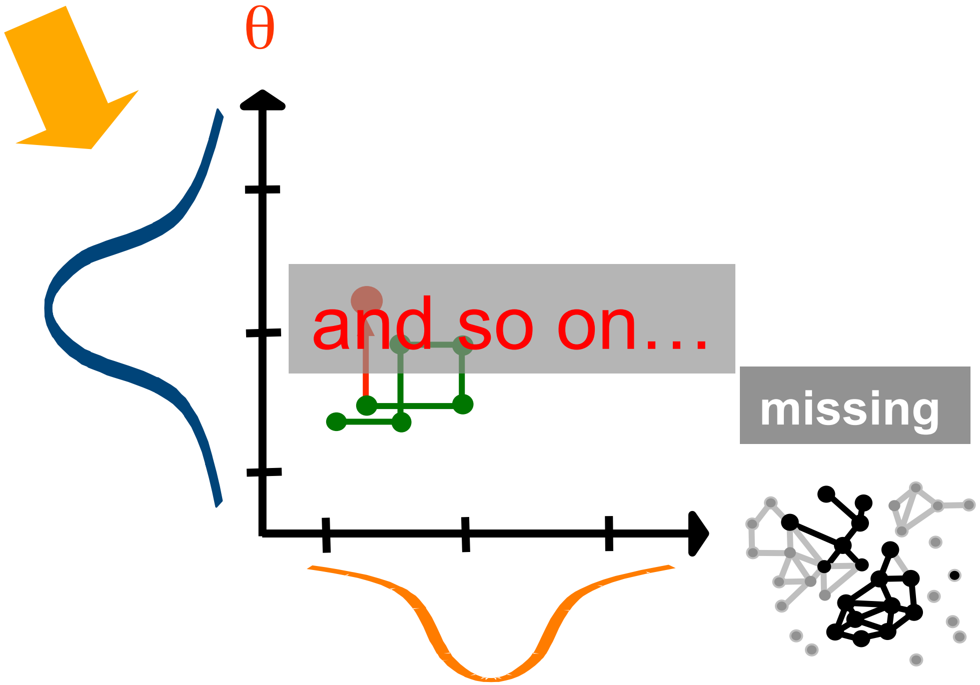
In each iteration
simulate **graphs**

Bayesian Data Augmentation

Simulate **parameters** θ



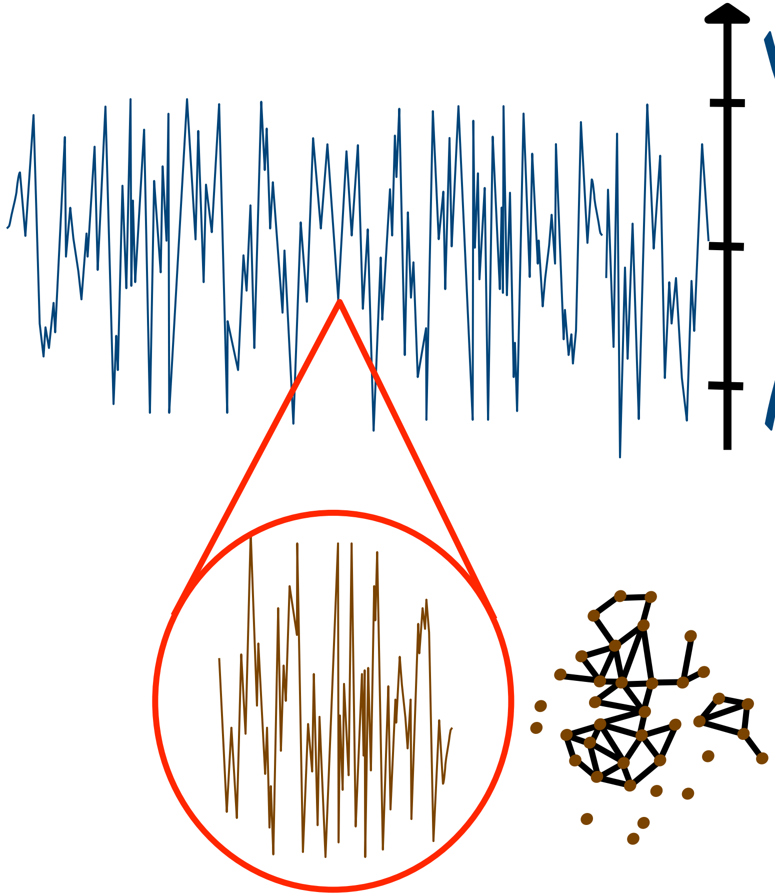
With missing data:



In each iteration
simulate **graphs**

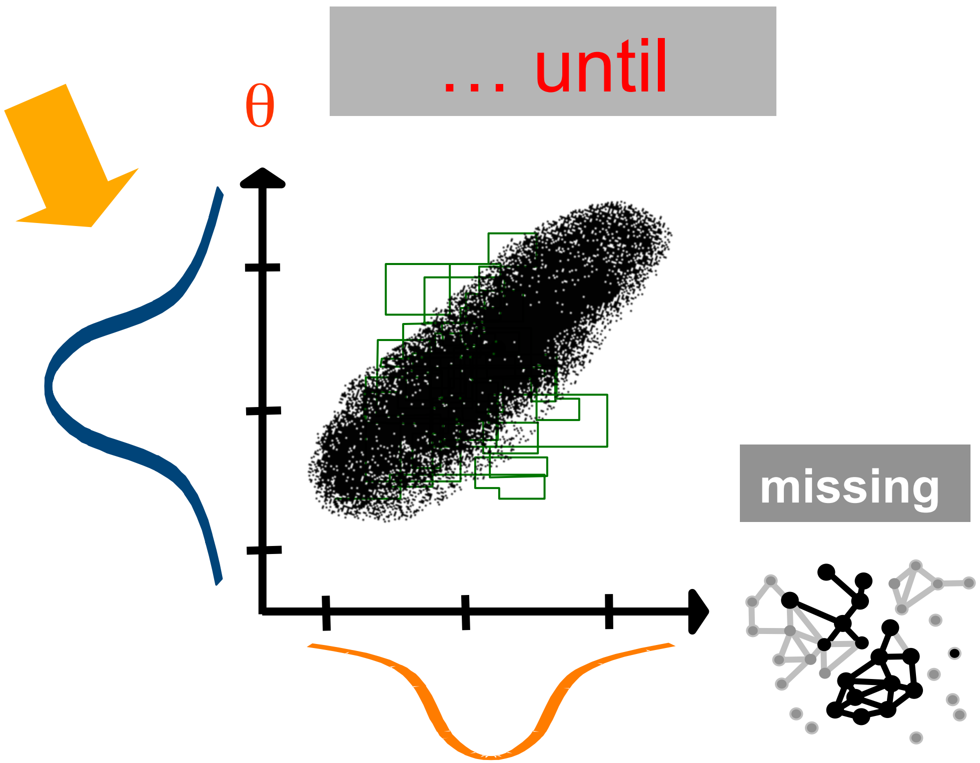
Bayesian Data Augmentation

Simulate **parameters** θ

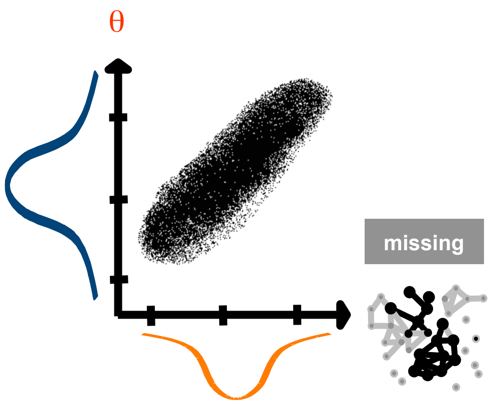


In each iteration
simulate **graphs**

With missing data:



Bayesian Data Augmentation



What does it give us?

Distribution of parameters

Distribution of missing data

Subtle point

Missing data does **not** depend on the parameters (we don’t have to choose parameters to simulate missing)

Lazega's (2001) Lawyers

Collaboration network among 36 lawyers in a New England law firm (Lazega, 2001)

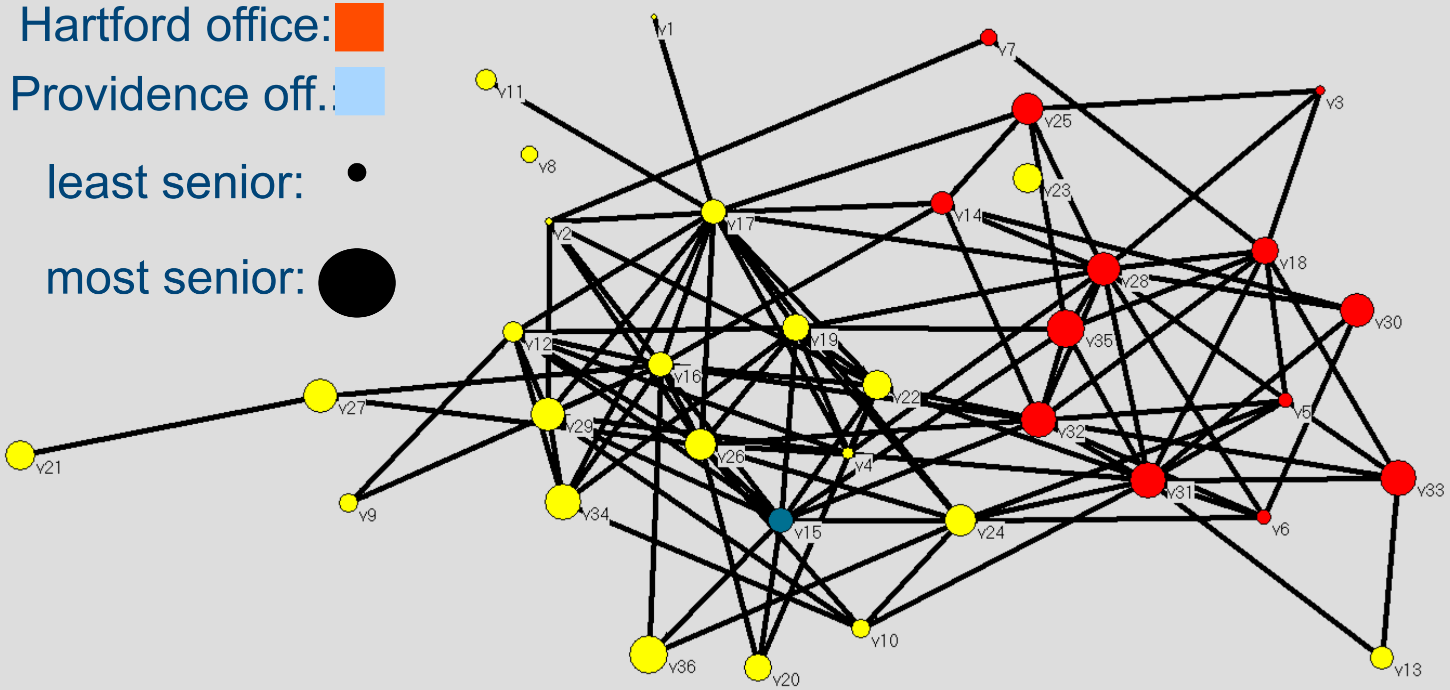
Boston office:

Hartford office:

Providence off.:

least senior:

most senior:



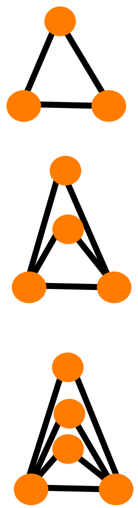
Lazega's (2001) Lawyers

(b_i = 1, if i corporate, 0 litigation)

t_1 :

t_2 :

t_3 :



etc.

Main effect

Edges: $\sum x_{ij}$

Seniority: $\sum x_{ij}(a_i + a_j)$

Practice: $\sum x_{ij}(b_i + b_j)$

Homophily

Practice: $\sum x_{ij}\mathbf{1}(b_i = b_j)$

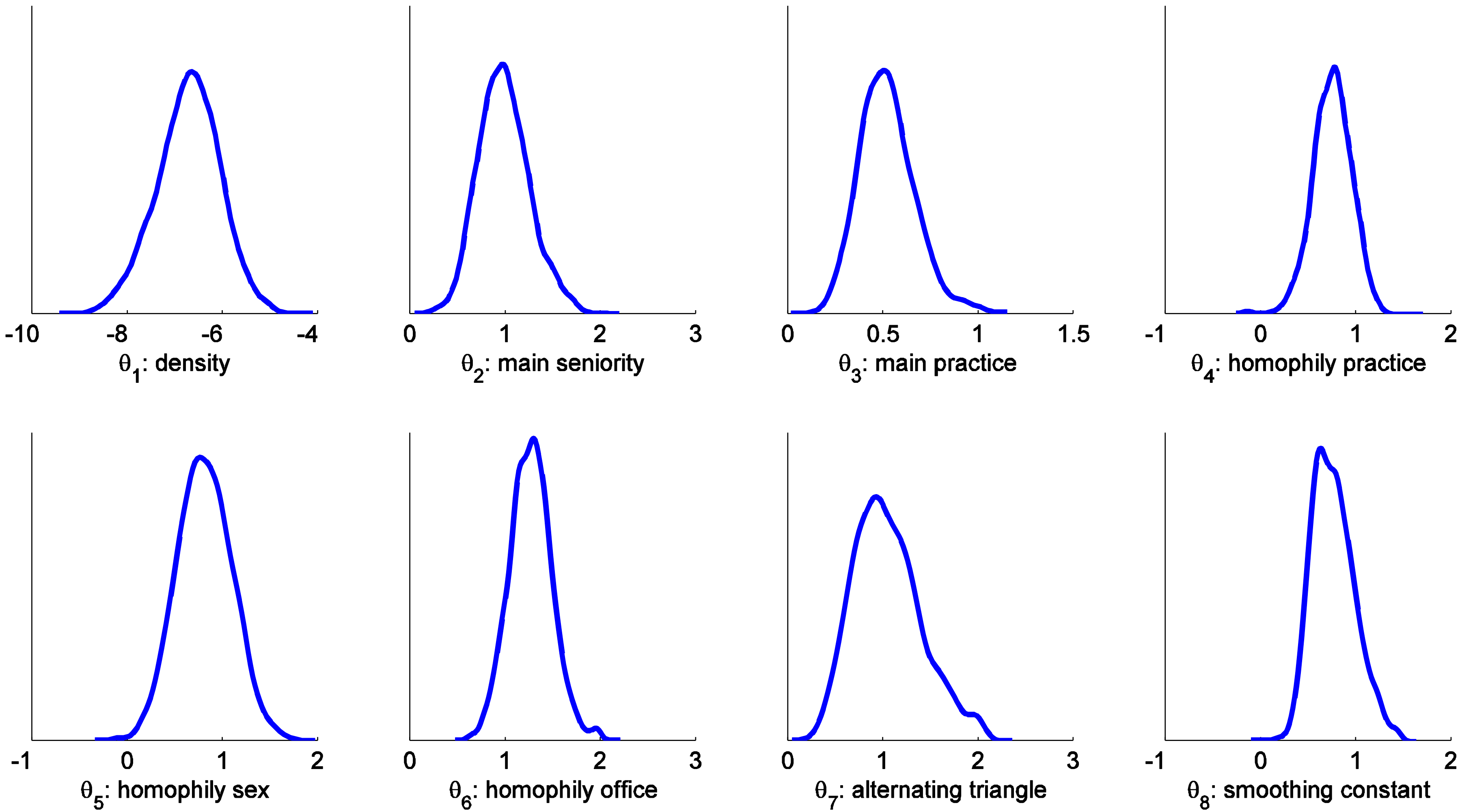
Sex: $\sum x_{ij}\mathbf{1}(c_i = c_j)$

Office: $\sum x_{ij}\mathbf{1}(d_i = d_j)$

GWESP: $$3t_1(x) - \frac{t_2(x)}{\lambda^1} + \cdots + (-1)^{n-3}\frac{t_{n-2}(x)}{\lambda^{n-3}}$$

with $\theta_8 = \log(\lambda)$

Lazega's (2001) Lawyers – ERGM posteriors (Koskinen, 2008)



Cross validation (Koskinen, Robins & Pattison, 2010)

Remove 200 of the 630 dyads at random

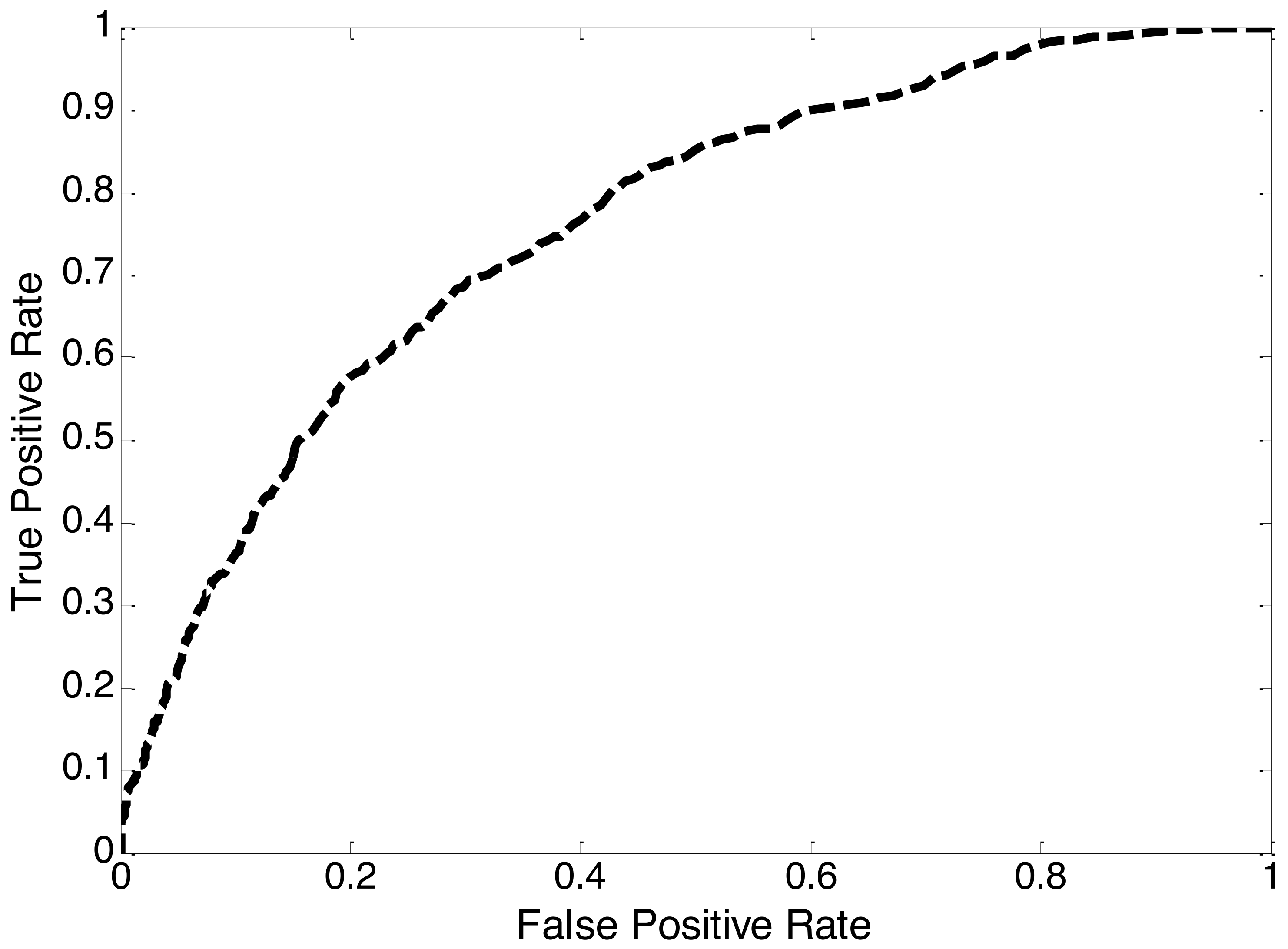
Fit inhomogeneous Bernoulli model obtain the posterior predictive tie-probabilities for the missing tie-variables

Fit ERGM and obtain the posterior predictive tie-probabilities for the missing tie-variables (Koskinen et al., in press)

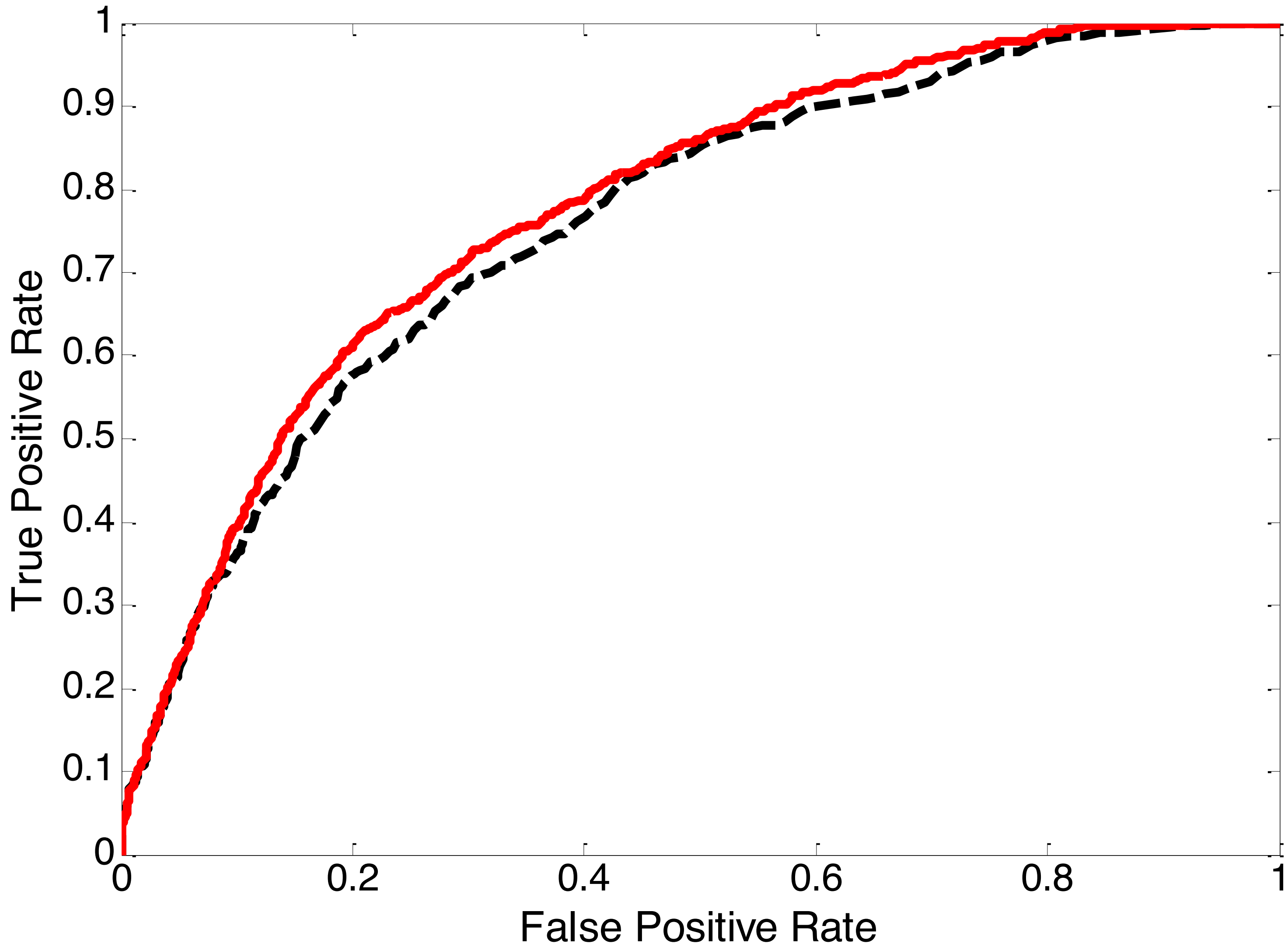
Fit Hoff's (2008) latent variable probit model with linear predictor $\theta^{T}z(x_{ij}) + w_i \Lambda w_j^{T}$

Repeat many times

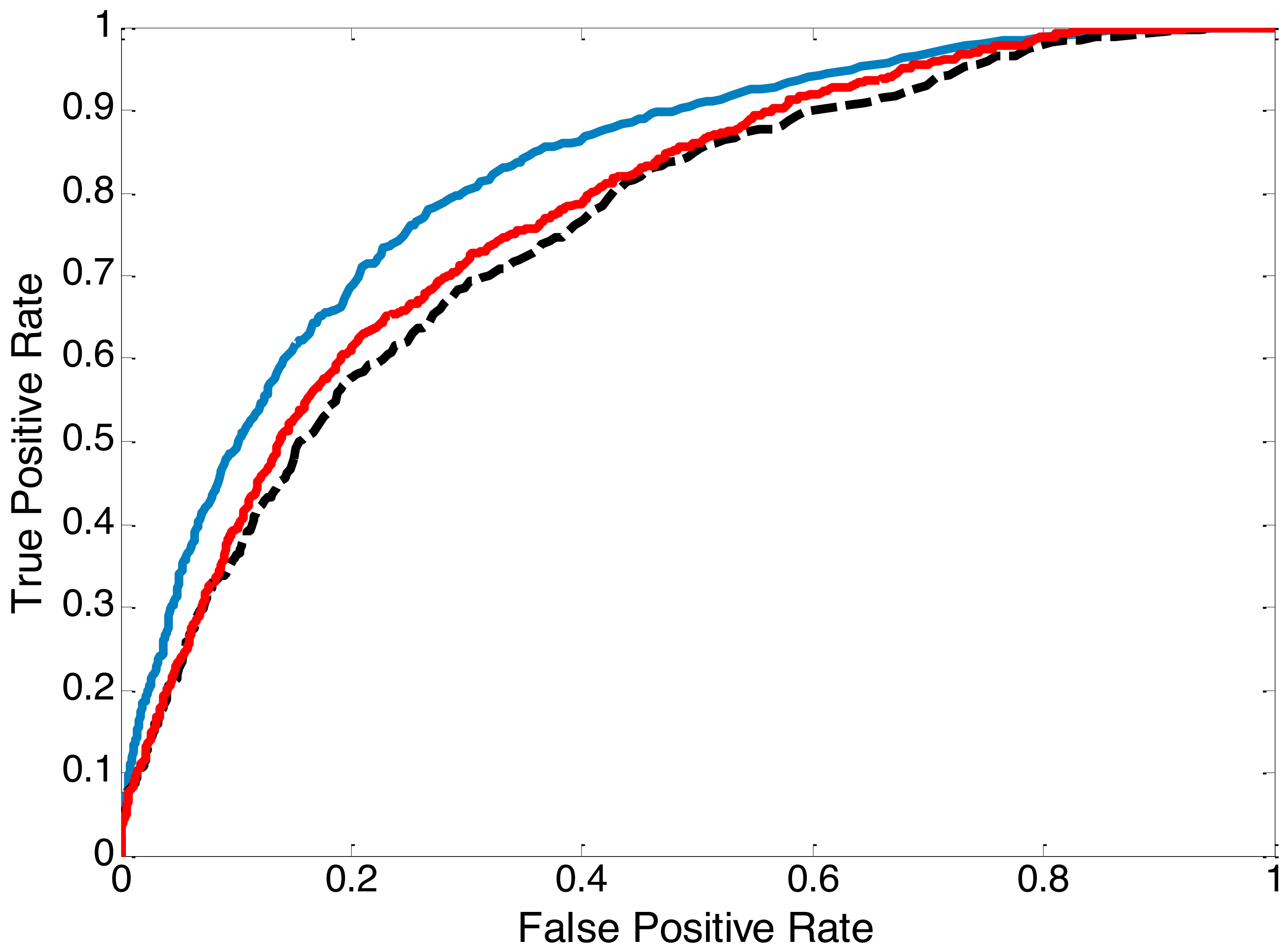
ROC curve for predictive probabilities combined over 20 replications (Koskinen et al. 2010)



ROC curve for predictive probabilities combined over 20 replications (Koskinen et al. 2010)



ROC curve for predictive probabilities combined over 20 replications (Koskinen et al. 2010)



Snowball sampling

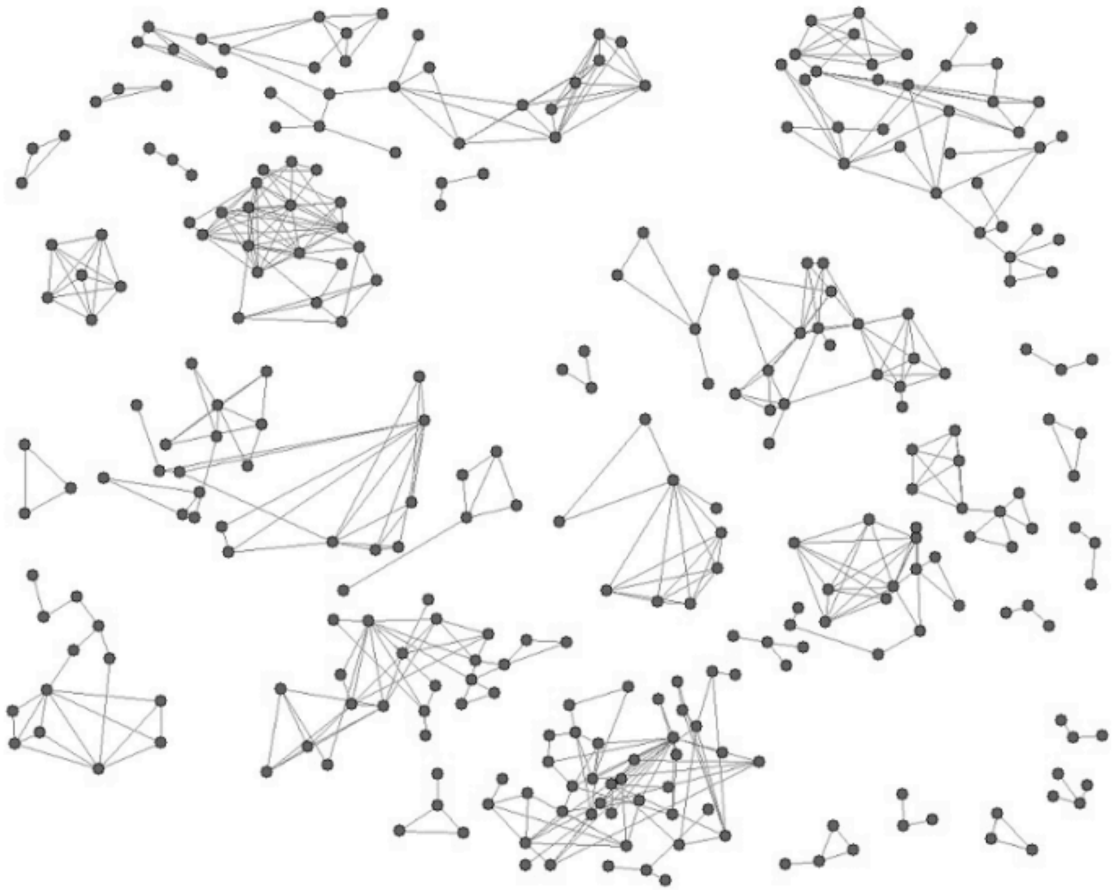
- Snowball sampling design ignorable for ERGM (**Thompson and Frank, 2000**, **Handcock & Gile 2010**; **Koskinen, Robins & Pattison, 2010**)
- ... **but** snowball sampling rarely used when population size is known...
- Using the Sageman (2004) “clandestine” network as test-bed for unknown N

Part 10

Spatial embedding

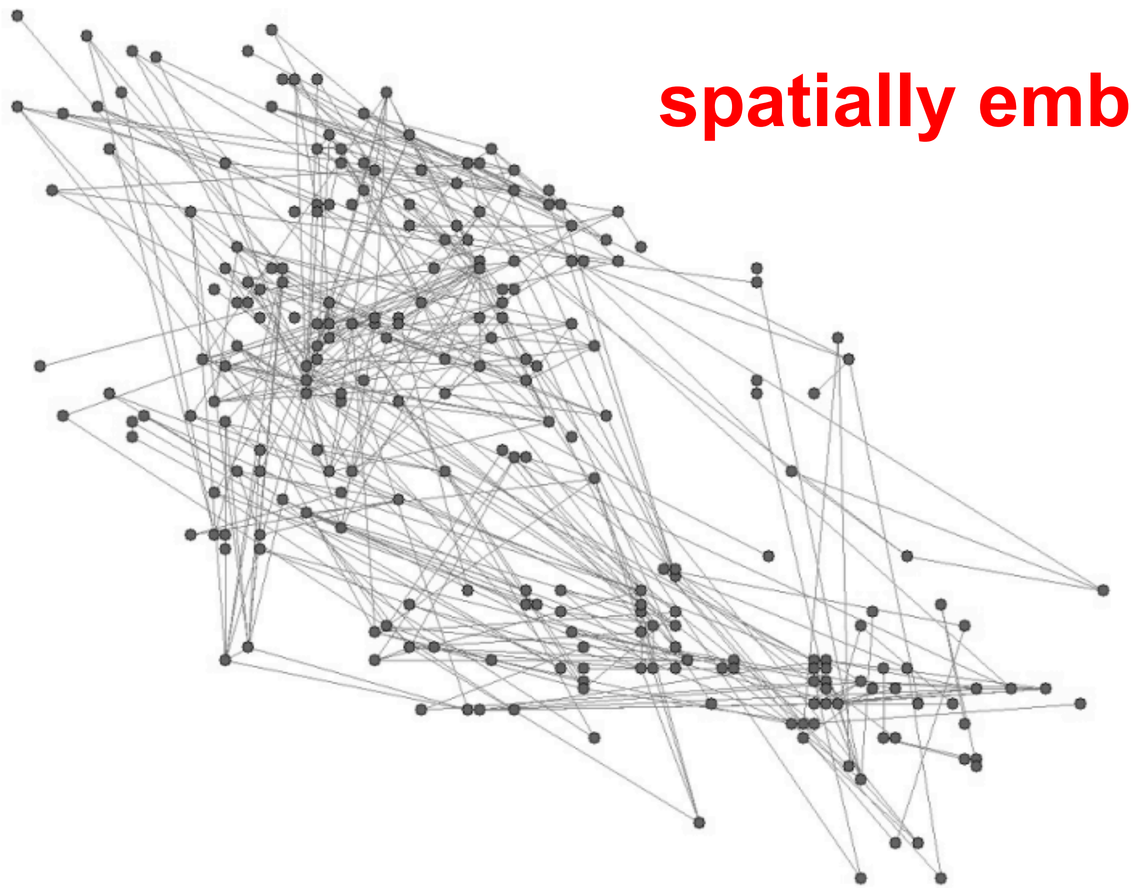
Spatial embedding – Daraganova et al. 2011 SOCNET

306 actors in Victoria, Australia



Spatial embedding – Daraganova et al. 2011 SOCNET

306 actors in Victoria, Australia

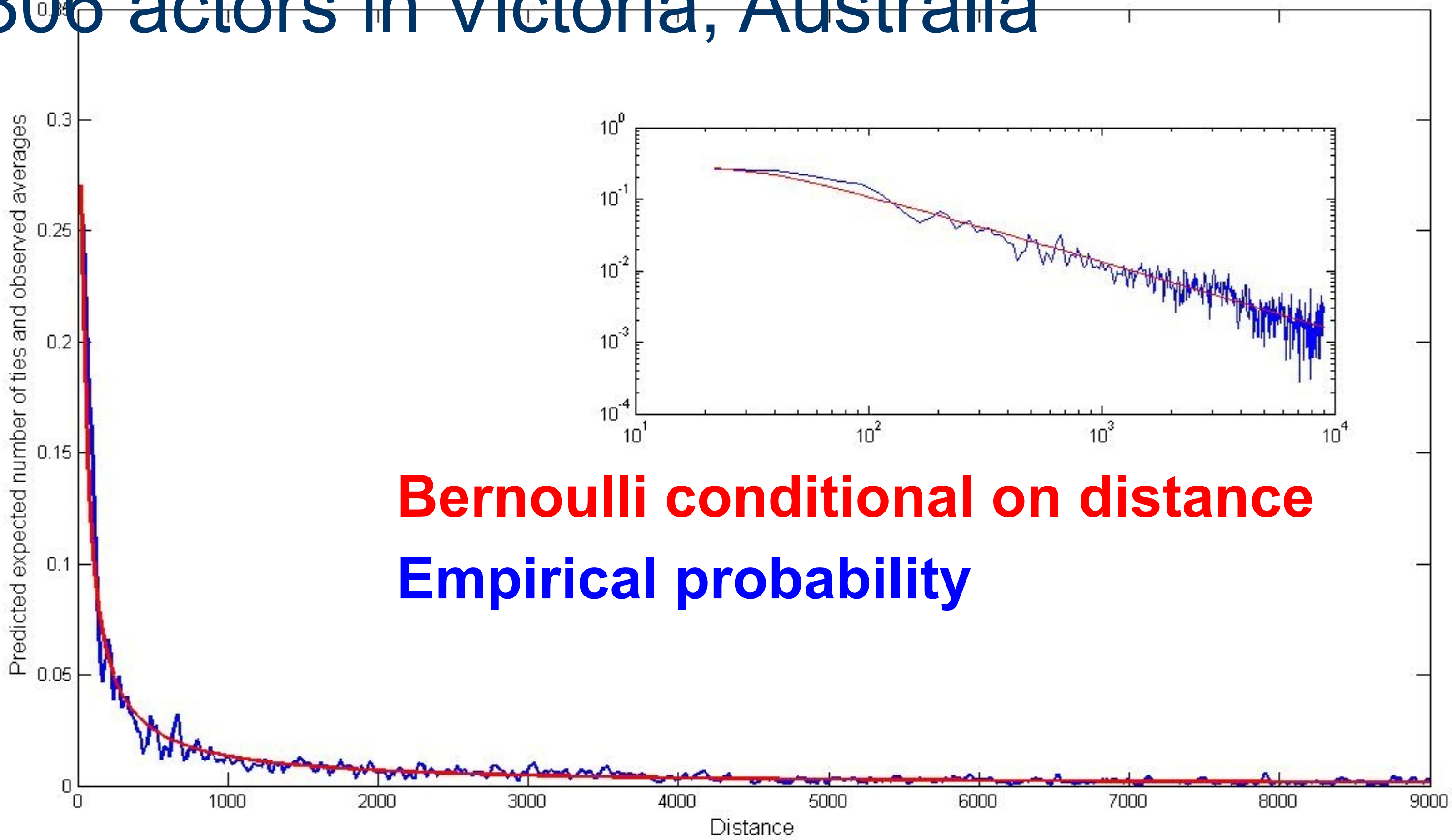


spatially embedded

... all living within 14 kilometres of each other

Spatial embedding – Daraganova et al. 2011 SOCNET

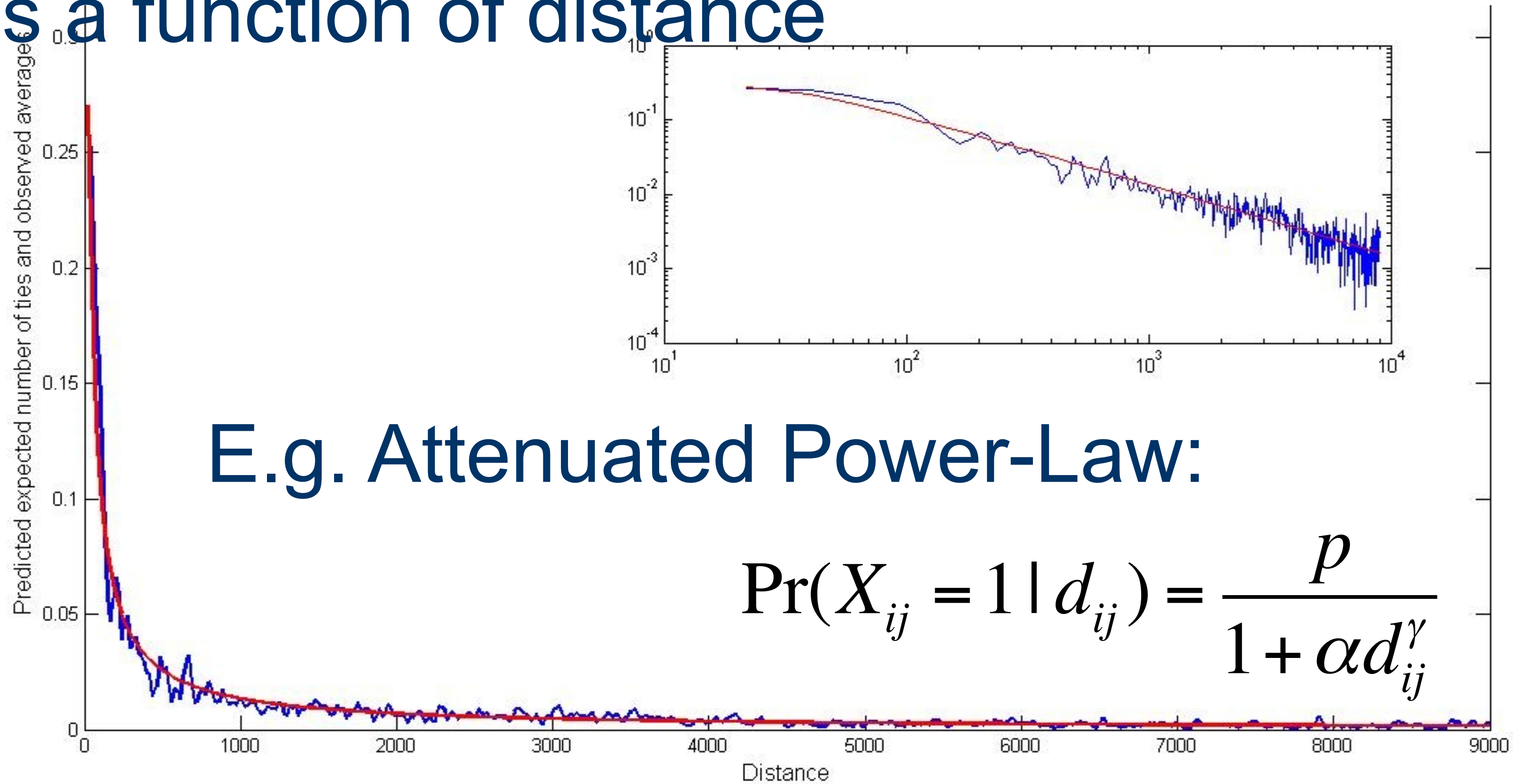
306 actors in Victoria, Australia



... all living within 14 kilometres of each other

Spatial embedding – Daraganova et al. 2011 SOCNET

Spatial interaction function: Tie probability as a function of distance

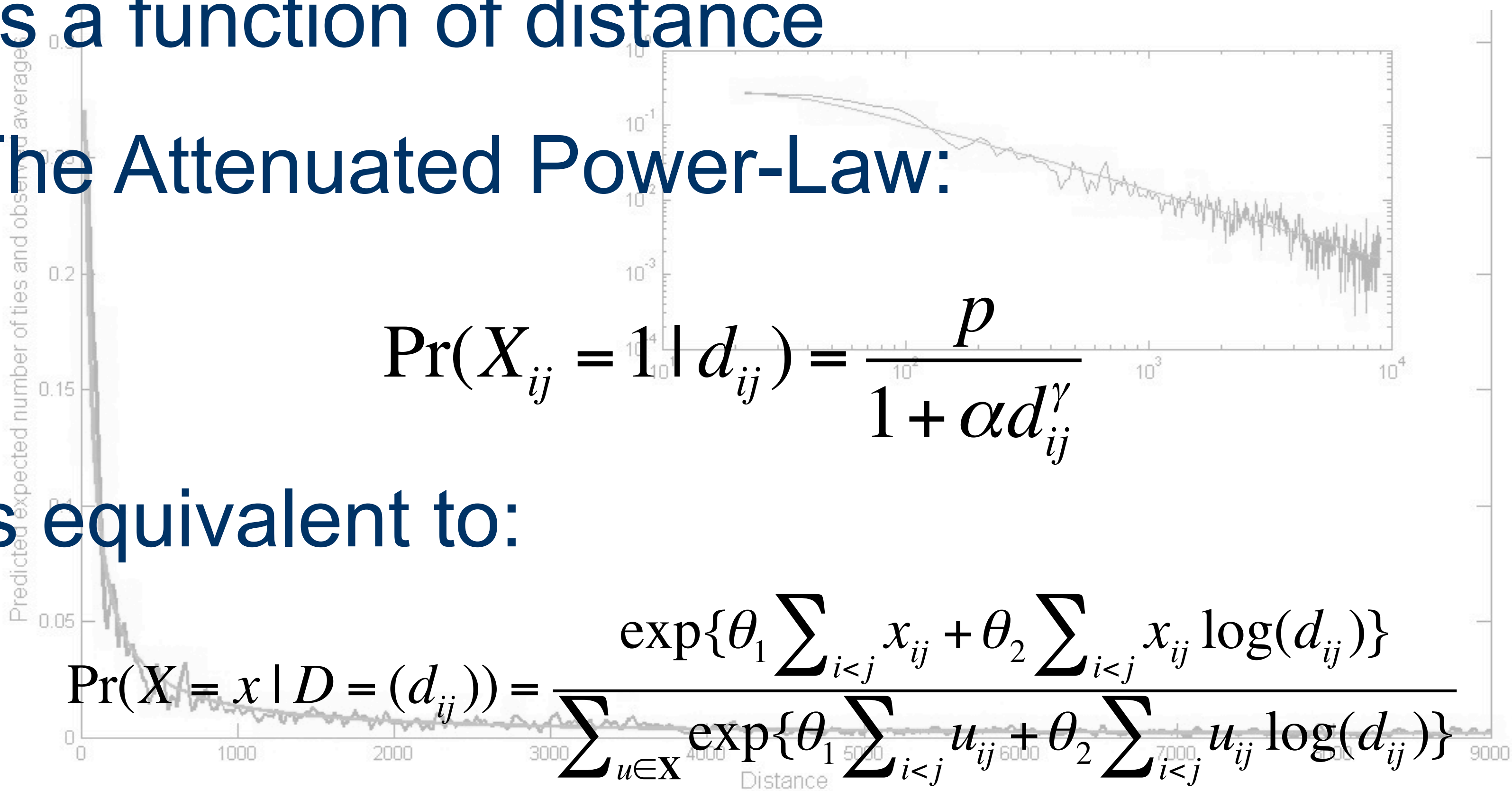


E.g. Attenuated Power-Law:

$$\Pr(X_{ij}=1 \mid d_{ij}) = \frac{p}{1+\alpha d_{ij}^{\gamma}}$$

Spatial embedding – Daraganova et al. 2011 SOCNET

Spatial interaction function: Tie probability as a function of distance



The Attenuated Power-Law:

$$\Pr(X_{ij} = 1 \mid d_{ij}) = \frac{p}{1 + \alpha d_{ij}^{\gamma}}$$

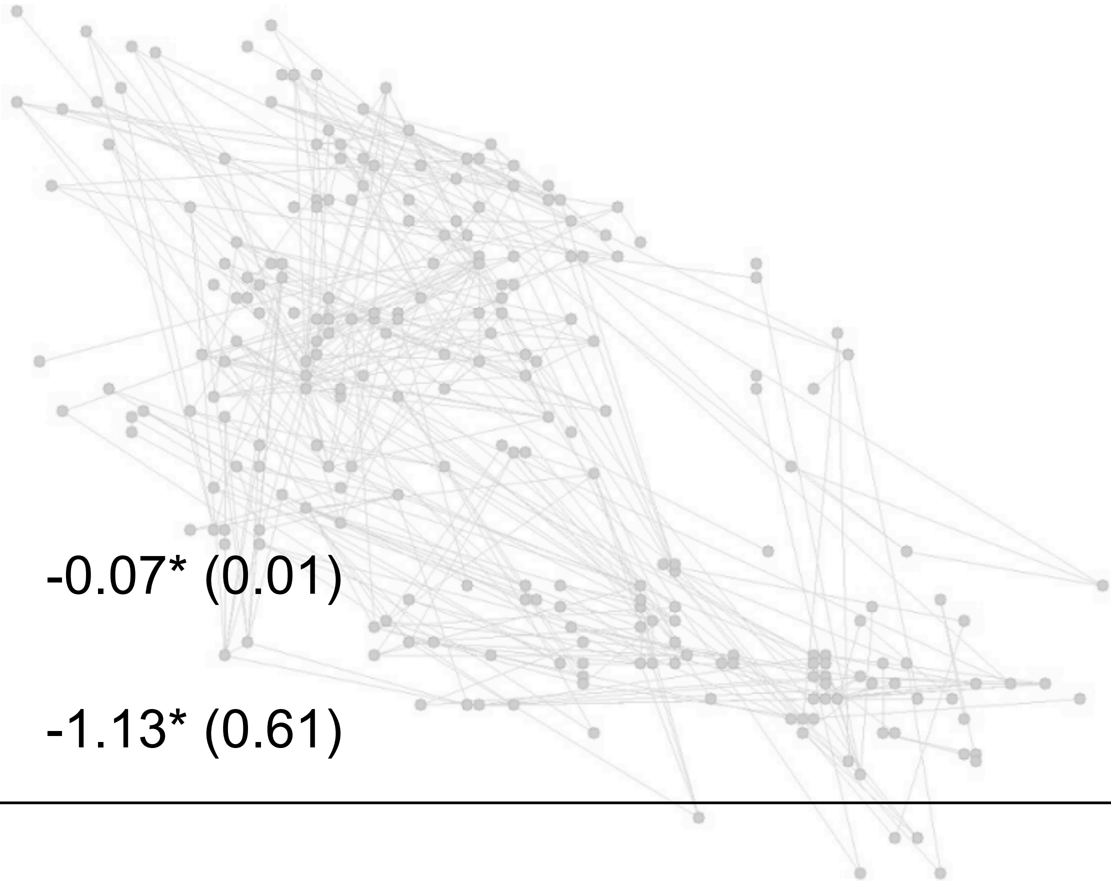
Is equivalent to:

$$\Pr(X = x \mid D = (d_{ij})) = \frac{\exp\{\theta_1 \sum_{i<j} x_{ij} + \theta_2 \sum_{i<j} x_{ij} \log(d_{ij})\}}{\sum_{u \in \mathbf{X}} \exp\{\theta_1 \sum_{i<j} u_{ij} + \theta_2 \sum_{i<j} u_{ij} \log(d_{ij})\}}$$

with: $p = 1 \quad \alpha = e^{-\theta_1} \quad \gamma = -\theta_2$ AND: $\log(d_{ij})$

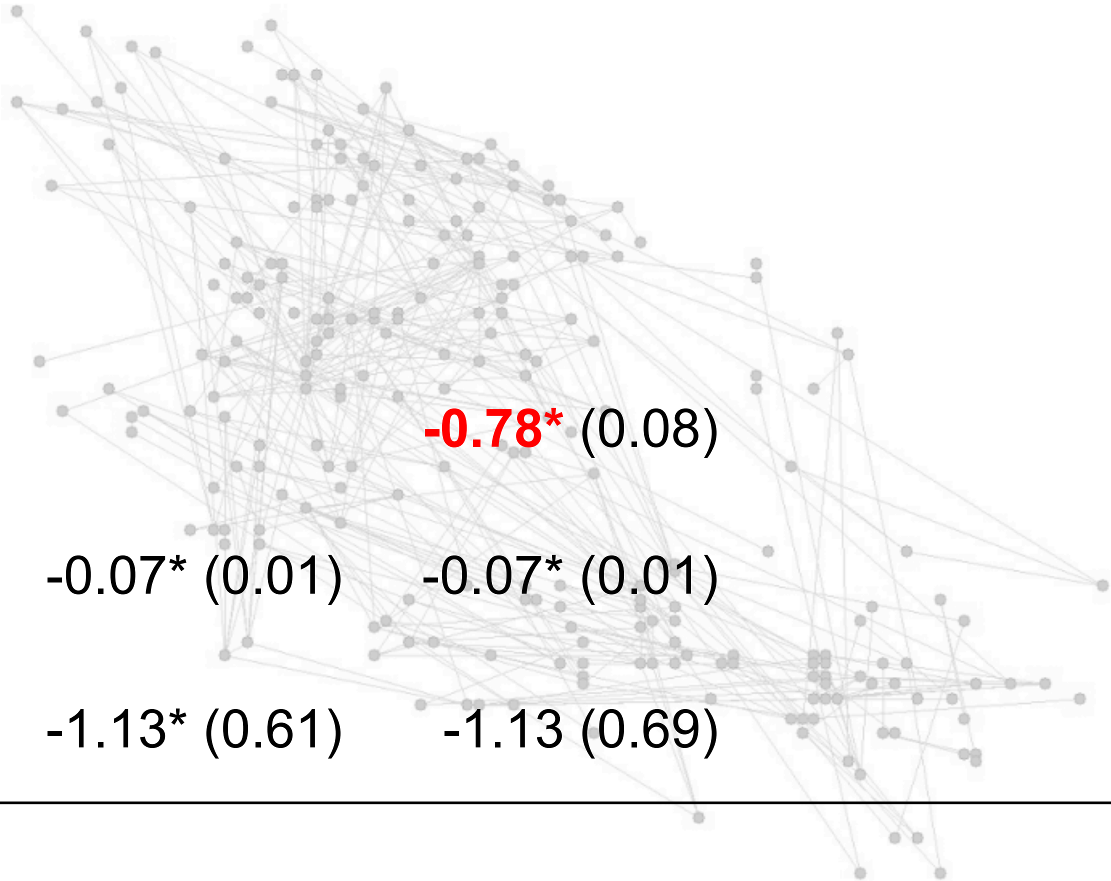
Spatial embedding – Daraganova et al. 2011 SOCNET

Edges	-4.87* (0.13)
Alt. star	
Alt. triangel	
Log distance	
Age heterophily	-0.07* (0.01)
Gender homophily	-1.13* (0.61)



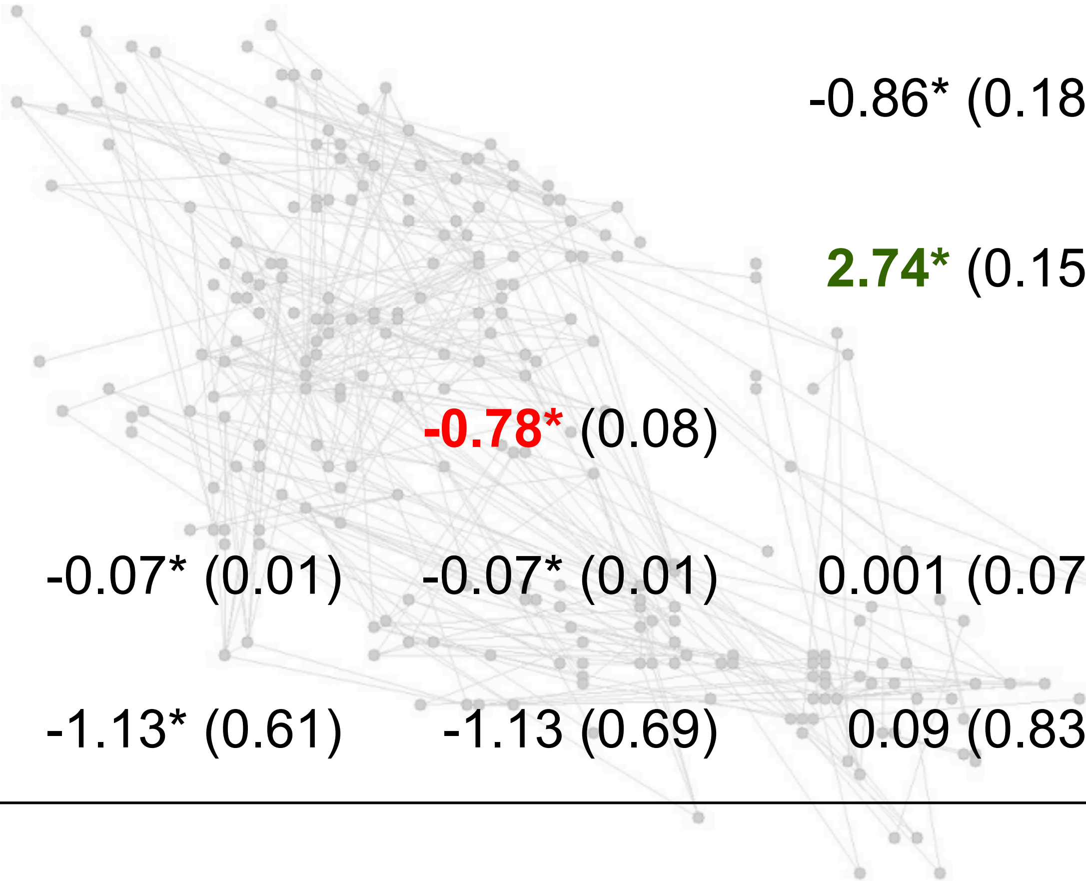
Spatial embedding – Daraganova et al. 2011 SOCNET

Edges	-4.87* (0.13)	1.56* (0.65)
Alt. star		
Alt. triangel		
Log distance		**-0.78*** (0.08)
Age heterophily	-0.07* (0.01)	-0.07* (0.01)
Gender homophily	-1.13* (0.61)	-1.13 (0.69)



Spatial embedding – Daraganova et al. 2011 SOCNET

Edges	-4.87* (0.13)	1.56* (0.65)	-4.79* (0.66)
Alt. star			-0.86* (0.18)
Alt. triangel			**2.74*** (0.15)
Log distance		**-0.78*** (0.08)	
Age heterophily	-0.07* (0.01)	-0.07* (0.01)	0.001 (0.07)
Gender homophily	-1.13* (0.61)	-1.13 (0.69)	0.09 (0.83)



Spatial embedding – Daraganova et al. 2011 SOCNET

Edges	-4.87* (0.13)	1.56* (0.65)	-4.79* (0.66)	-0.20 (0.87)
Alt. star			-0.86* (0.18)	-0.86* (0.2)
Alt. triangel			**2.74*** (0.15)	**2.69*** (0.14)
Log distance		**-0.78*** (0.08)		**-0.56*** (0.07)
Age heterophily	-0.07* (0.01)	-0.07* (0.01)	0.001 (0.07)	0.002 (0.06)
Gender homophily	-1.13* (0.61)	-1.13 (0.69)	0.09 (0.83)	0.07 (0.47)

ERGM: distance and **endogenous** dependence explain different things

Part 8

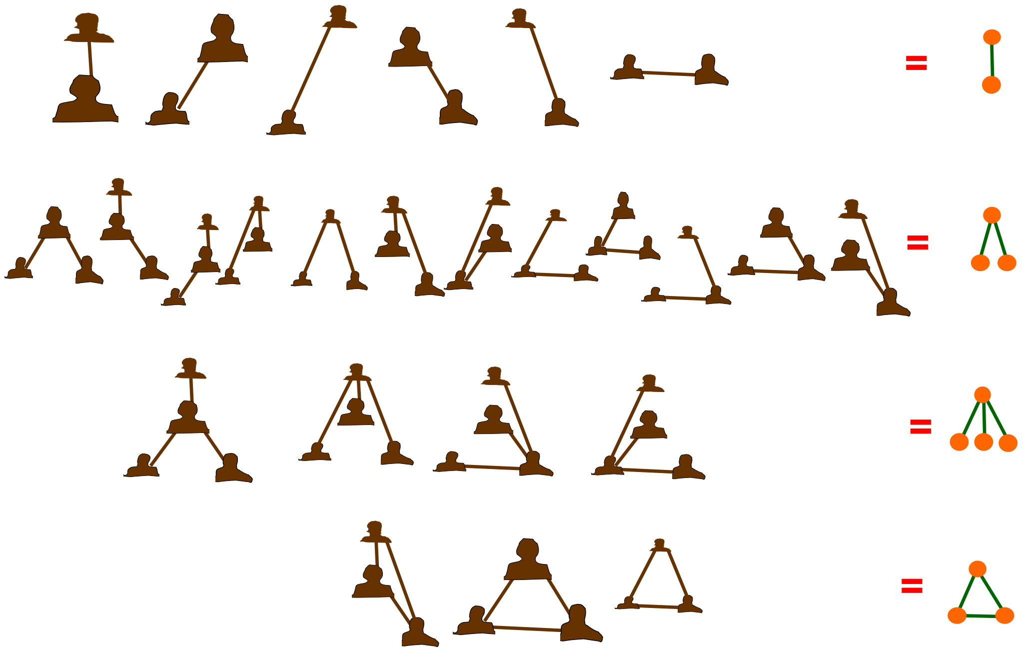
Further issues

Other extensions

There are **ERGMs** (ERGM-like models) for
- directed data
- valued data
- bipartite data
- multiplex data
- longitudinal data
- modelling actor autoregressive attributes

Further issues – relaxing homogeneity assumption

ERGMs typically assume homogeneity



(A) Block modelling and ERGM (Koskinen, 2009)
(B) Latent class ERGM (Schweingberger & Handcock)

Further issues – model selection

Assessing Goodness of Fit:

- Posterior predictive distributions (Koskinen, 2008; Koskinen, Robins & Pattison, 2010; Caimo & Friel, 2011)
- Non-Bayesian heuristic GOF (Robins et al., 2007; Hunter et al., 2008; Robins et al., 2009; Wang et al., 2009)

Model selection

-Path sampling for AIC (Hunter & Handcock, 2006); Conceptual **caveat**: model complexity when variables dependent?

- Bayes factors (Wang & Handcock…?)

Wrap-up

ERGMs

- Increasingly being used
- Increasingly being understood
- Increasingly being able to handle imperfect data (also missing link prediction)

Methods

-Plenty of open issues

-Bayes is the way of the future

Legitimacy and dissemination

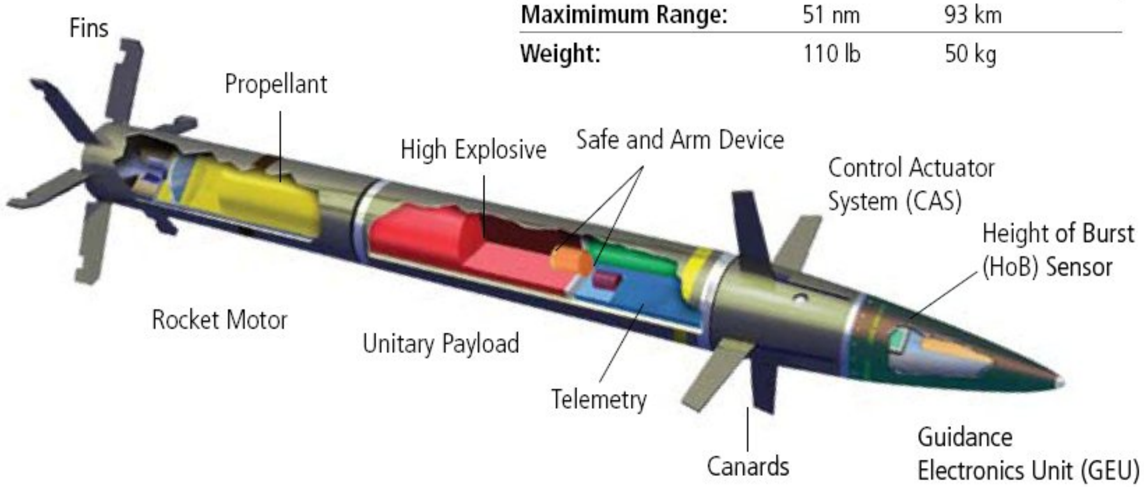
- e.g. Lusher, Koskinen, Robins ERGMs for SN, CUP, 2011

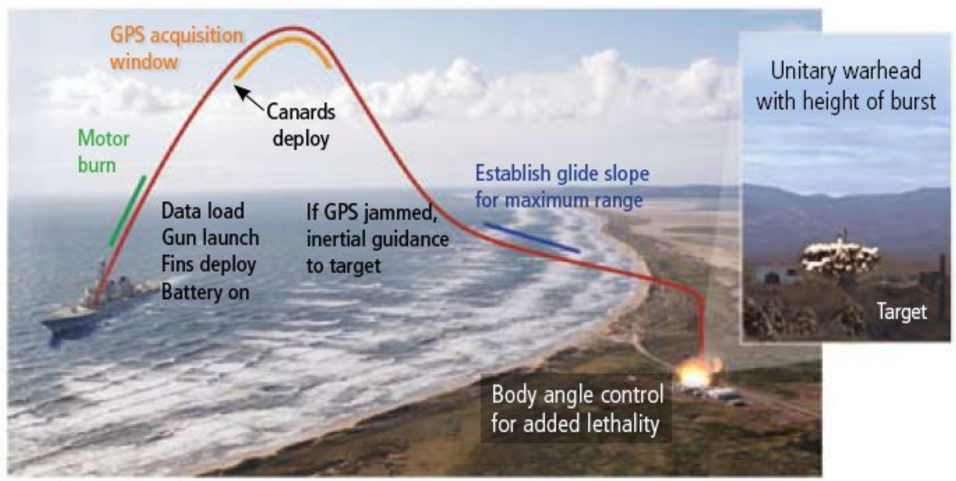
Remaining question: used to be p*…

... why ERGM?

ERGM Specifications

Length:	5.1 ft	1.55 m
Diameter:	5.0 in	127 mm
Maximimum Range:	51 nm	93 km
Weight:	110 lb	50 kg





Operational Concept