

## ERGMs 1.2

### RECENS / BCE PhD course

Lecturers: Takács Károly, Vörös András, Samu Flóra, Néray Bálint, Boda Zsófia

With thanks to Garry Robins, Paola Zappa, Vörös András and Boda Zsófia

MTA TK, Budapest 2016. április 27 – május 6.

A projekt az MTA TK „Lendület” RECENS Kutatócsoport támogatásával  
valósult meg | <http://recens.tk.mta.hu>

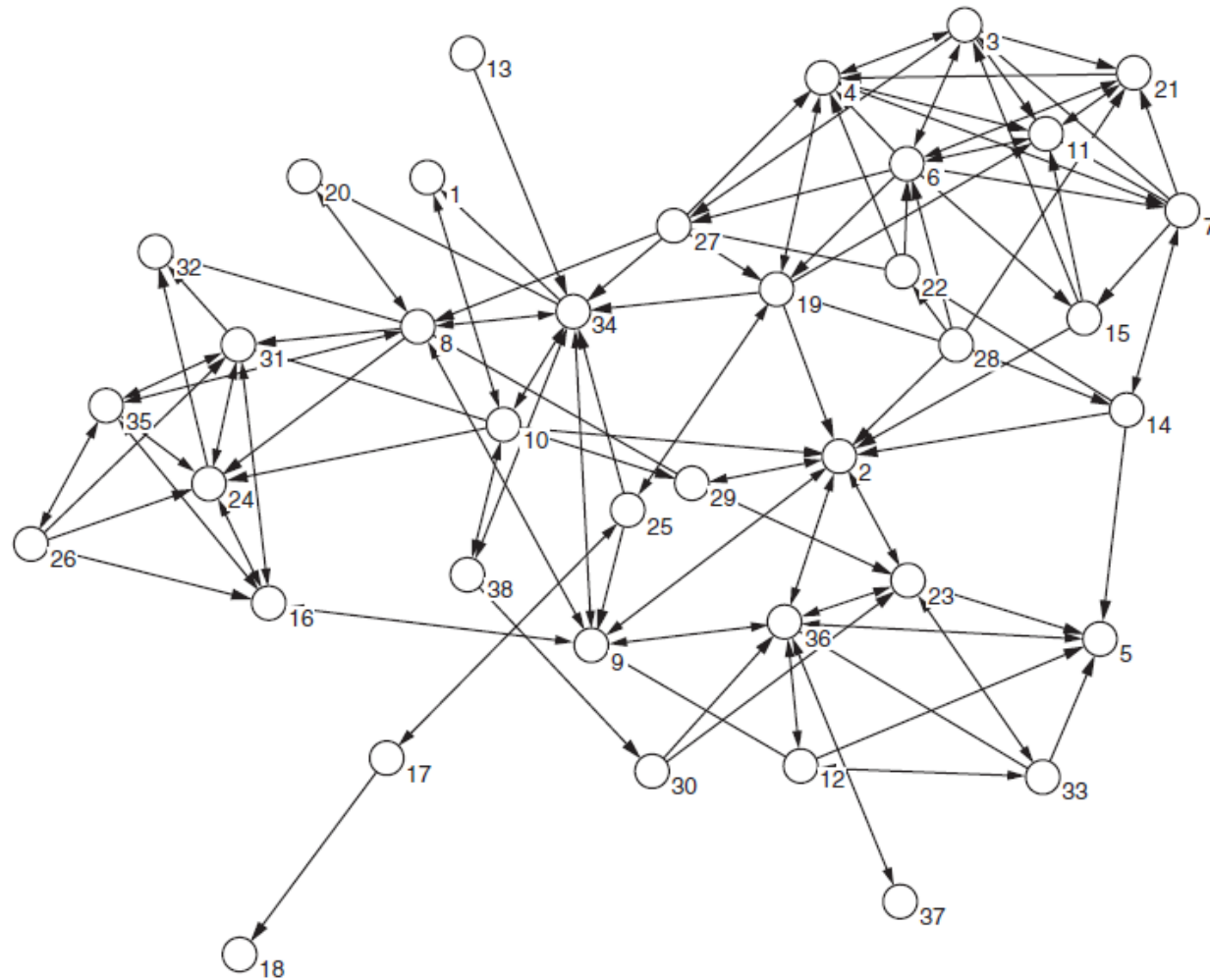


### 3. Understanding graph distribution

# Example: corporation data

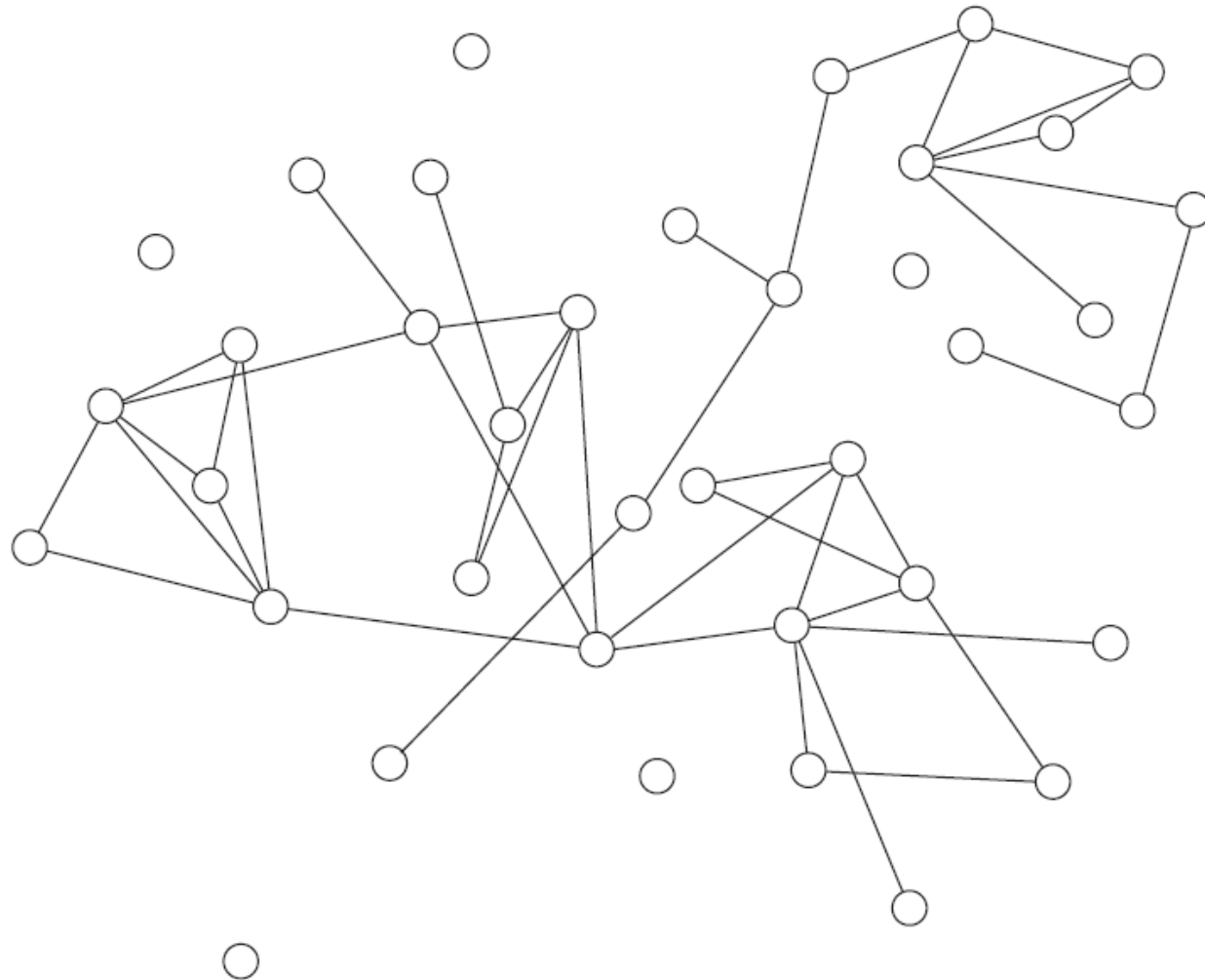
- context: a company (entertainment industry)
- nodes: 38 executives
- network: communication between nodes
- tie: another executive with whom it was important to communicate to get work completed effectively (binary, directed ties)
- individual attributes:
  - experience (number of projects actors have been involved with)
  - level of seniority
  - office membership
- dyadic attribute: advice (from whom actors received advice)

# Communication network



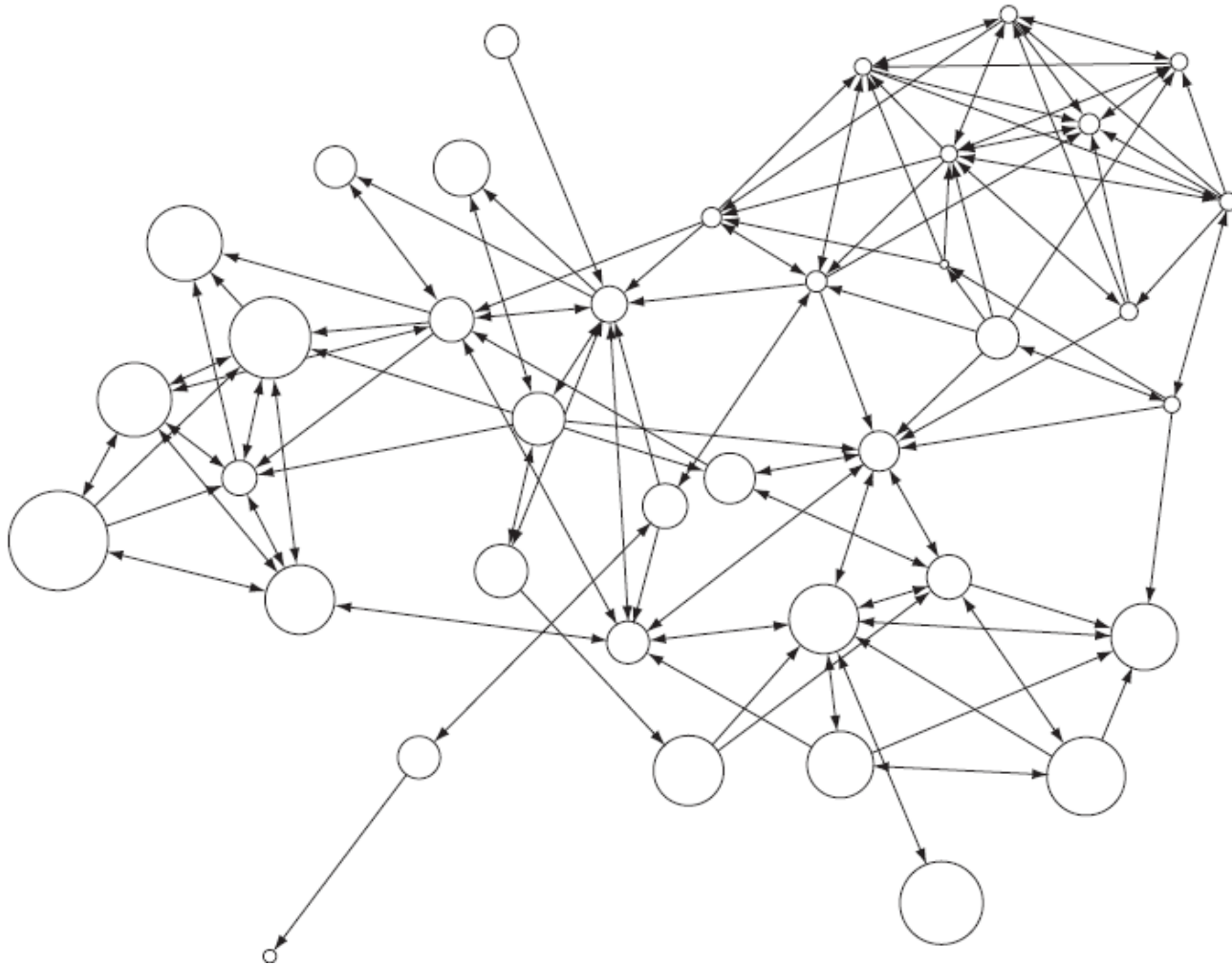
(Figure: Robins&Lusher 2012, p. 38)

# Communication network



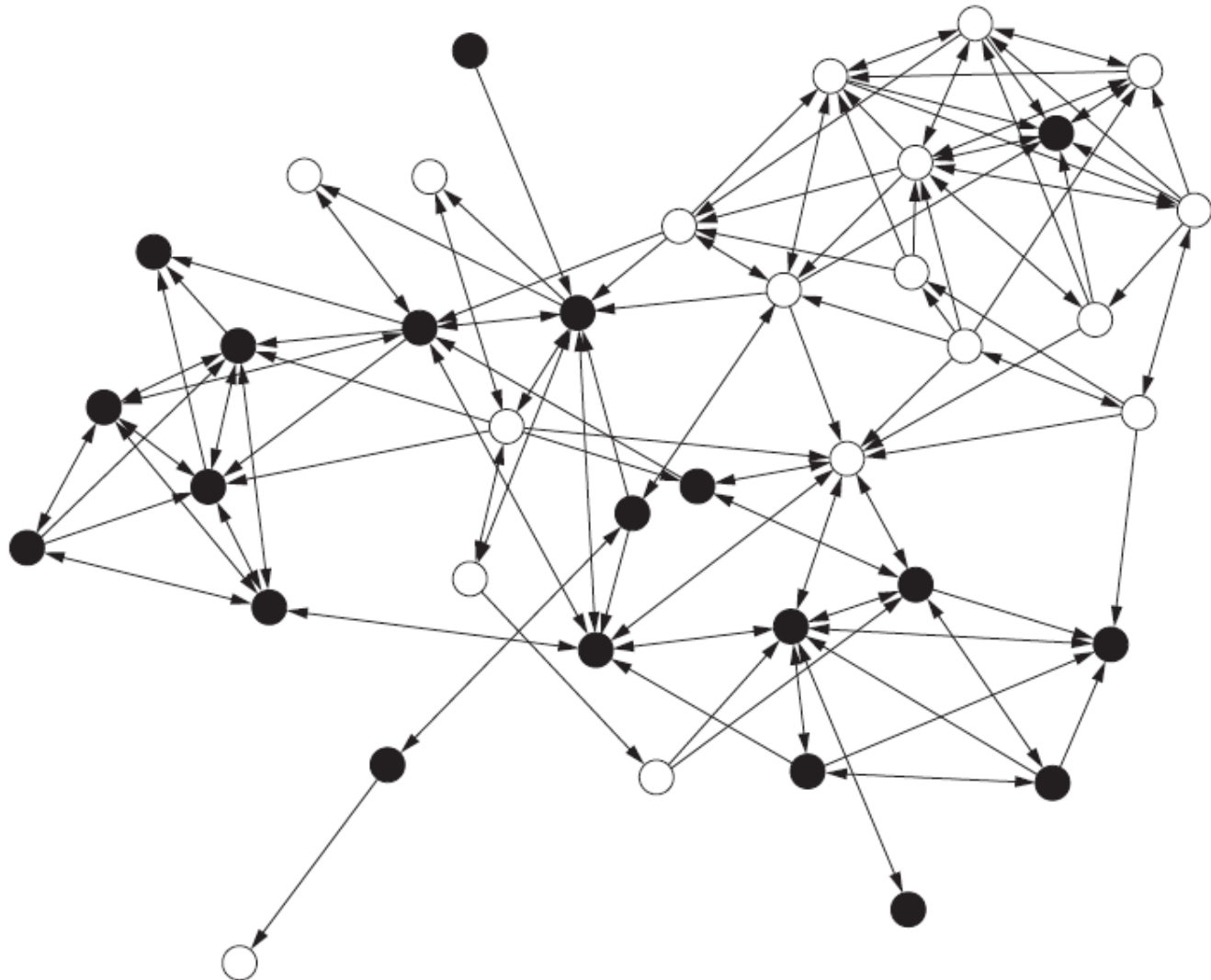
mutual ties only (Figure: Robins&Lusher 2012, p. 39)

# Communication network



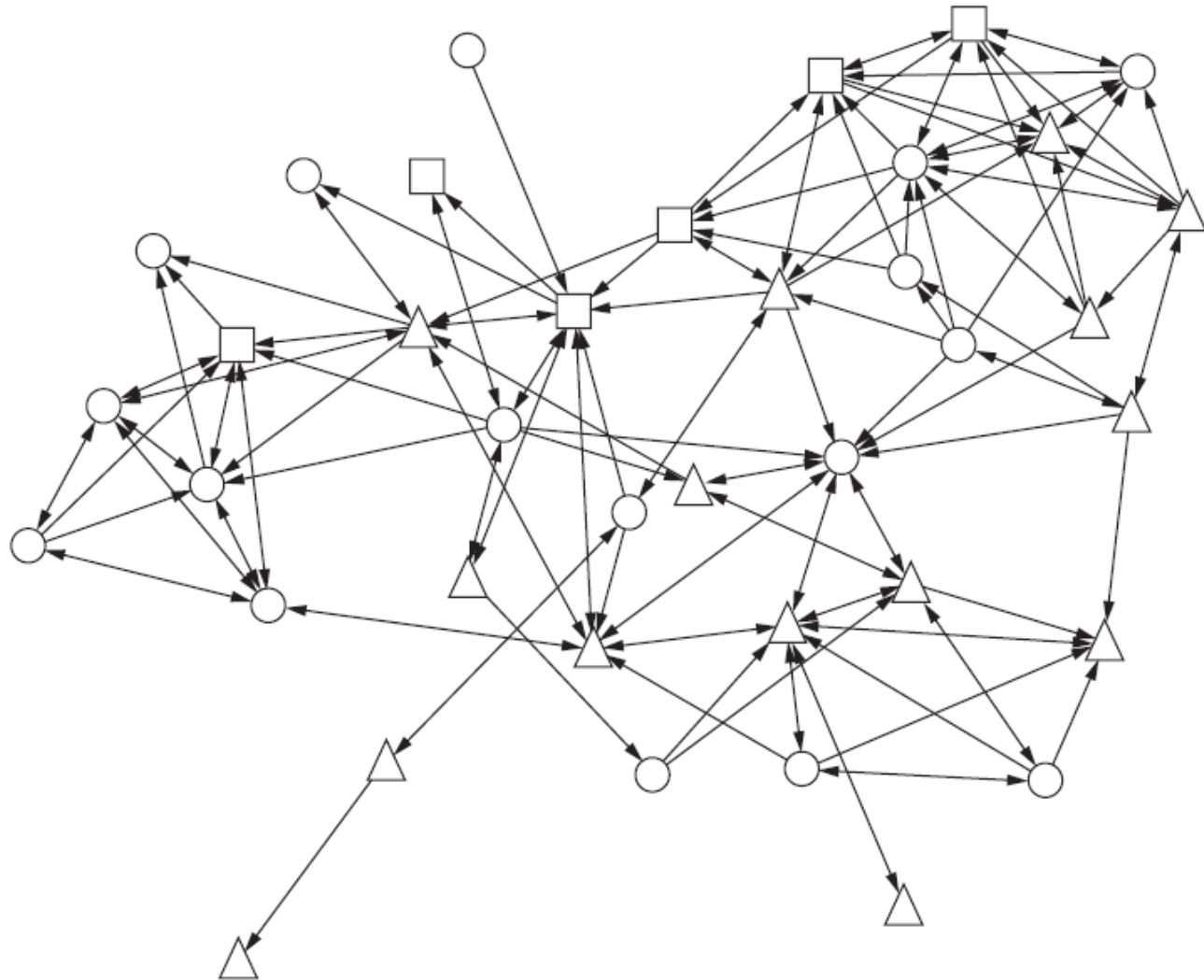
by employee experience (Figure: Robins&Lusher 2012, p. 40)

# Communication network



by seniority (black: senior) (Figure: Robins&Lusher 2012, p. 40)

# Communication network

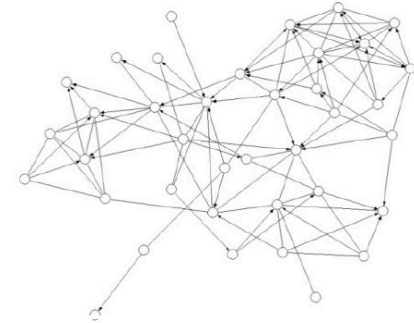
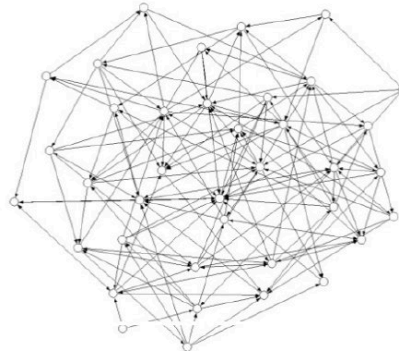


by office membership (Figure: Robins&Lusher 2012, p. 41)



# Simulated vs observed network

Lusher et al,  
2013



	Random network	Communication network
actors	38	38
arcs	146	146
reciprocated arcs	<b>6</b>	<b>44</b>
transitive triads	<b>53</b>	<b>212</b>
in-2stars	292	313
out-2stars	254	283

# Distribution of simple random graph

- A graph distribution is the set of all possible graphs (in this case on 38 nodes) with a probability assigned to each graph
- Uniform distribution of graphs with 44 edges: each graph has equal probability if it has exactly 44 edges
  - If it does not have exactly 44 edges the probability is zero
- $UIL = 44$ 
  - U = Uniform distribution of graph
  - I = “given that” or “conditional on”
  - L = number of edges

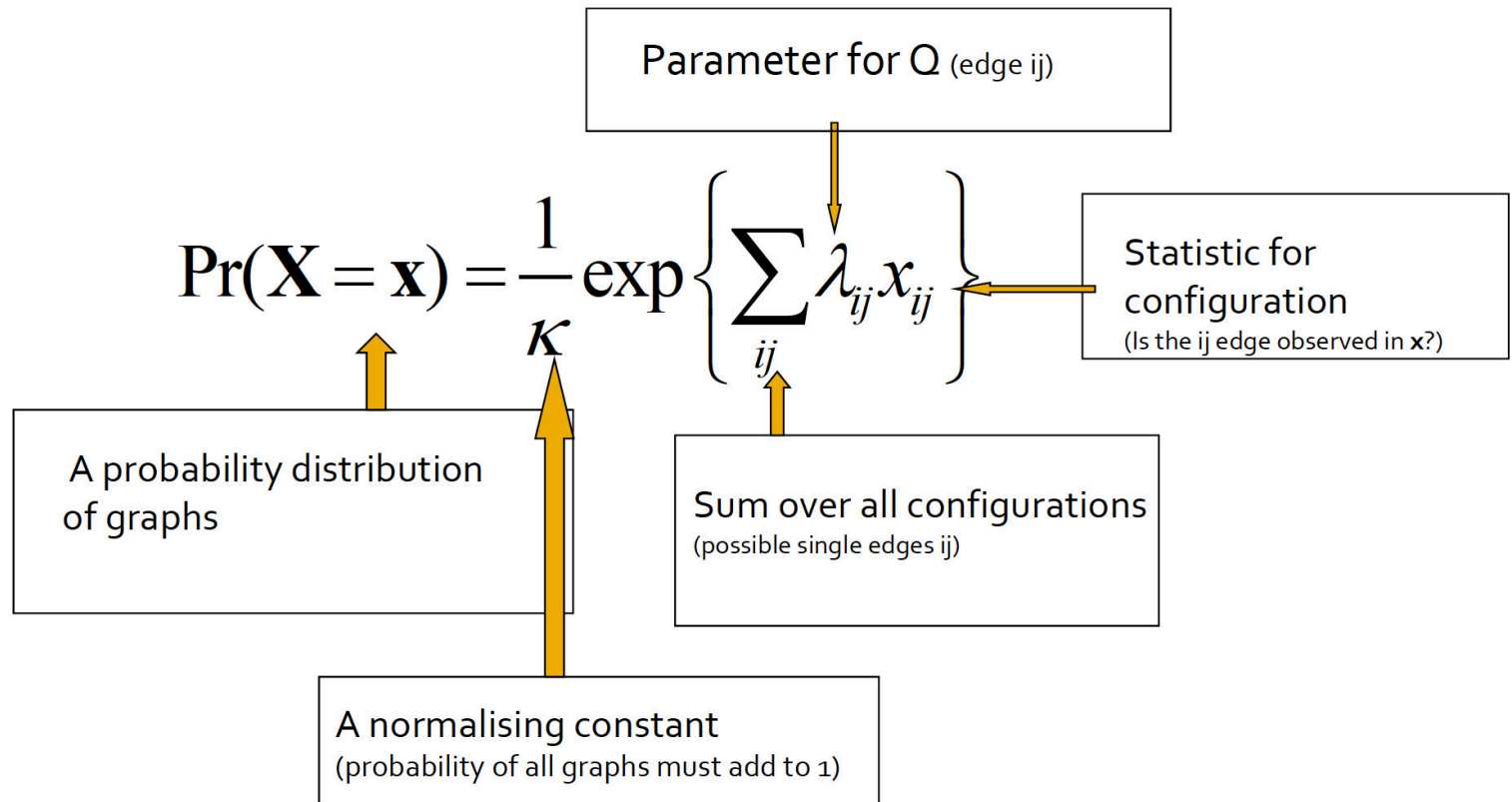
# Bernoulli distribution

- A **graph distribution** is the set of all possible graphs (in this case on 38 nodes) with a probability assigned to each graph.
- **Bernoulli distribution** of graphs: each edge in the graph occurs independently with a fixed probability:
  - Like tossing a coin many times
  - Leading to a probability for each graph
  - For 38 nodes, if we make this probability 0.06259, then across the entire distribution, the average density will be 0.06259 and the average number of edges will be 44.

# Some notations

- Regard each network tie  $x_{ij}$  as a random variable which is the unit of the analysis
- Probability of network  $x$  is given by a sum of network statistics ( $X$ 's)
  - weighted,
  - inside an exponential (binary distribution)
  - normalized,
  - expresses the counts of network configurations in network  $x$

# What we estimate



Possible edges are independent of one another. Configurations in this model relate to single possible edges ( $x_{ij}$ ) (and nothing beyond that in a Bernoulli graph).

# What we estimate

- If there was one parameter  $\lambda_{ij}$  for every possible edge – simply too many.
- Homogeneity assumption:  $\lambda_{ij} = \theta$  for all  $i, j$ . Assumes that the edge effect is the same across the entire network.

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{K} \exp\left(\sum \lambda_{ij} x_{ij}\right) = \frac{1}{K} \exp\left(\theta \sum x_{ij}\right) = \frac{1}{K} \exp(\theta L)$$

- where  $L$  is the number of edges in the observed network  $\theta$  is an edge or density parameter

# What are we trying to do?

- We do not know the theoretical model
- We want to find out the parameters (= weights for local structural configurations)
- We have many  $[n(n-1)]$  observations of ties and the local network structure around them
- These ties are assumed to be the outcomes of the same theoretical model (our ERGM)
- We can use them to estimate the model parameters
- We want to know how the global network structure might have been built up out of small local substructures
- The parameter estimates permit us to make inferences about this
- Positive parameter estimates indicate more configurations observed in the network than expected by chance
- Negative parameter estimates indicate fewer configurations than expected by chance

# Exercise 1

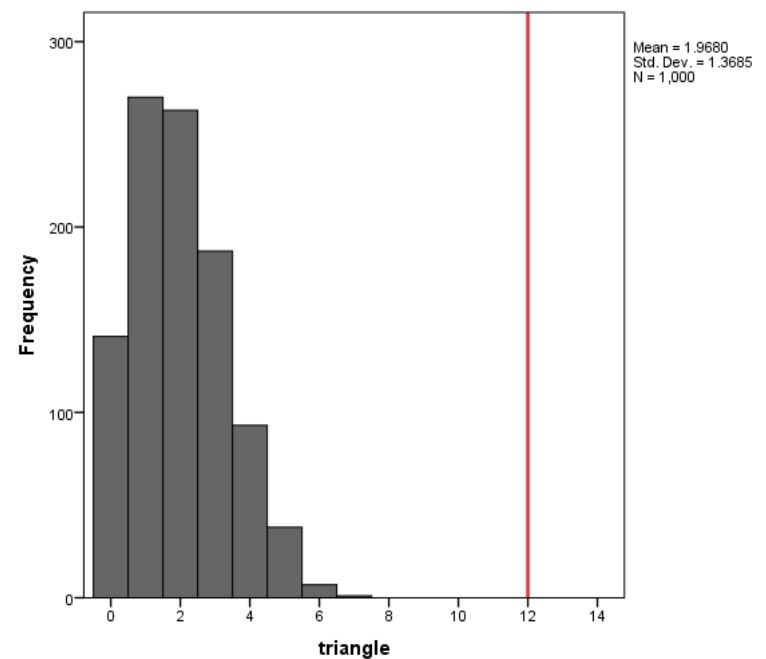
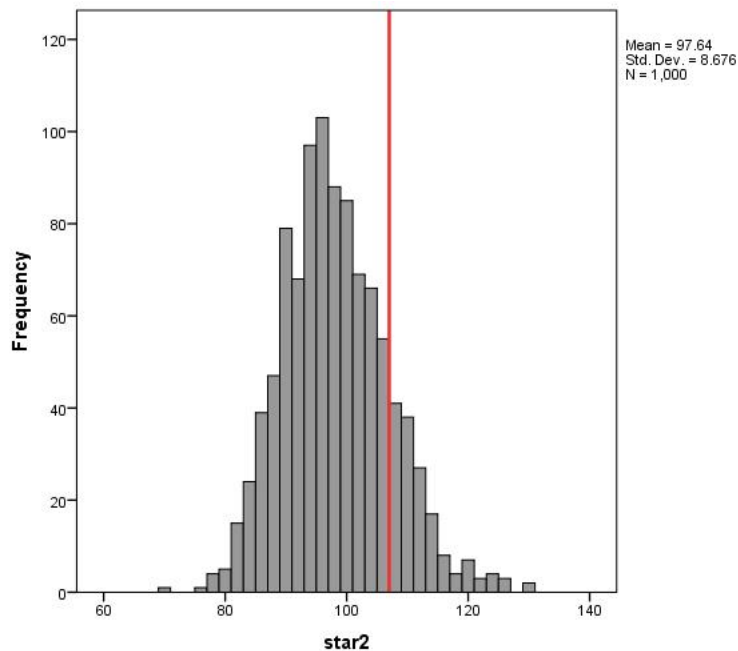


# Exercise 1

- Simulating graphs with 38 nodes and fixed density of 0.06259 (UIL = 44)
  - Simulate the UIL=44 distribution.
  - Take a sample of graphs from the distribution
  - Calculate the number of 2-stars and triangles for each sampled graph.
  - See whether the numbers of 2-stars and triangles are consistent with the observed data.

# Exercise 1

- stars2 are okay-ish.
- Observed network has 44 edges, 107 2-stars and 12 triangles
- Not a good model to describe triangles: triangulation in this data very unlikely to arise from random graphs conditional



# What we want instead:

- A statistical model for a network which can be parameterised so that important effects in the data are not extreme when the model is simulated.
- “Important effects” here means the presence of small subgraphs, called network configurations.
- ERGMs can provide such models for a range of configurations relevant to social network theory.

# Dependencies

- Once we move beyond simple random graph models, we introduce dependencies among network tie variables.
- These express (earlier discussed) various types of network self organisation.
- And we assume that the network is built up of these configurations.

# Generations of dependence assumptions

1. Bernoulli graphs: Network variables are independent of each other
2. Dyadic dependence: for directed graphs – dependence within dyads
3. Markov dependence: Network variables are (conditionally) independent unless they share at least one node.
4. Social circuit dependence: Network variables are (conditionally) dependent if they create 4-cycles.

Keep in mind dependence also arising from actor attributes.

## Exercise 2

# Exercise 2

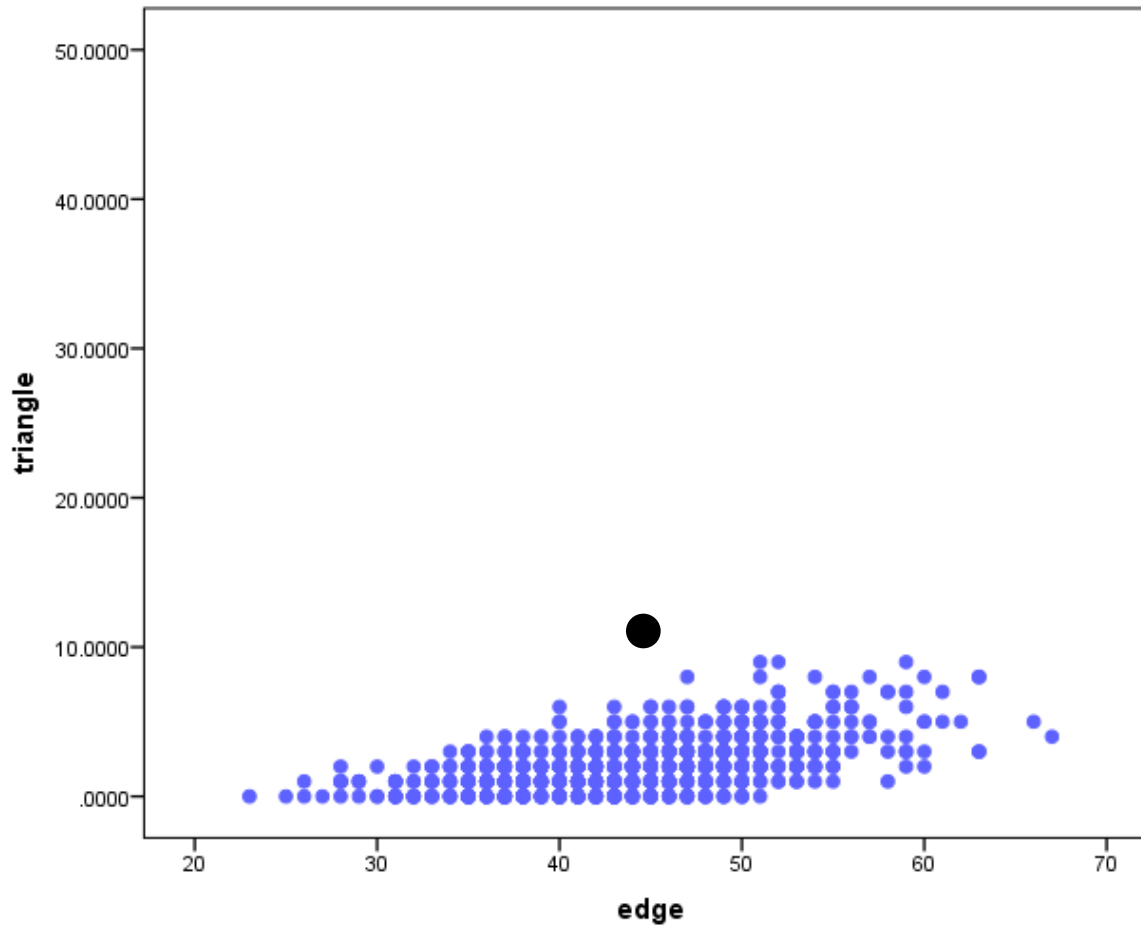
- Use the undirected communication network
- Try to estimate an edge-only model and save the results
- Calculate the probability of an existing edge, what is this number?
- Try to estimate an edge/2-star/3-star/triangle Markov model
- How did the edge parameter change? Why?
- (You may find the convergence ratios very bad)

# Markov model

- Frank and Strauss drew on the work of Besag (1974) in spatial statistics
  - In particular, the Hammersley-Clifford theorem that sets out constraints on model form implied by dependence assumptions
  - Dependence graph (for more information see Lusher et al, 2013)
- They proposed a network dependence assumption (Markov dependence):
  - Two tie variables are conditionally independent unless they share a node
- Edges are conditionally dependent if and only if they share a node (Frank & Strauss, 1986)
- Frank and Strauss showed that configurations in this model comprised edges, stars and triangles

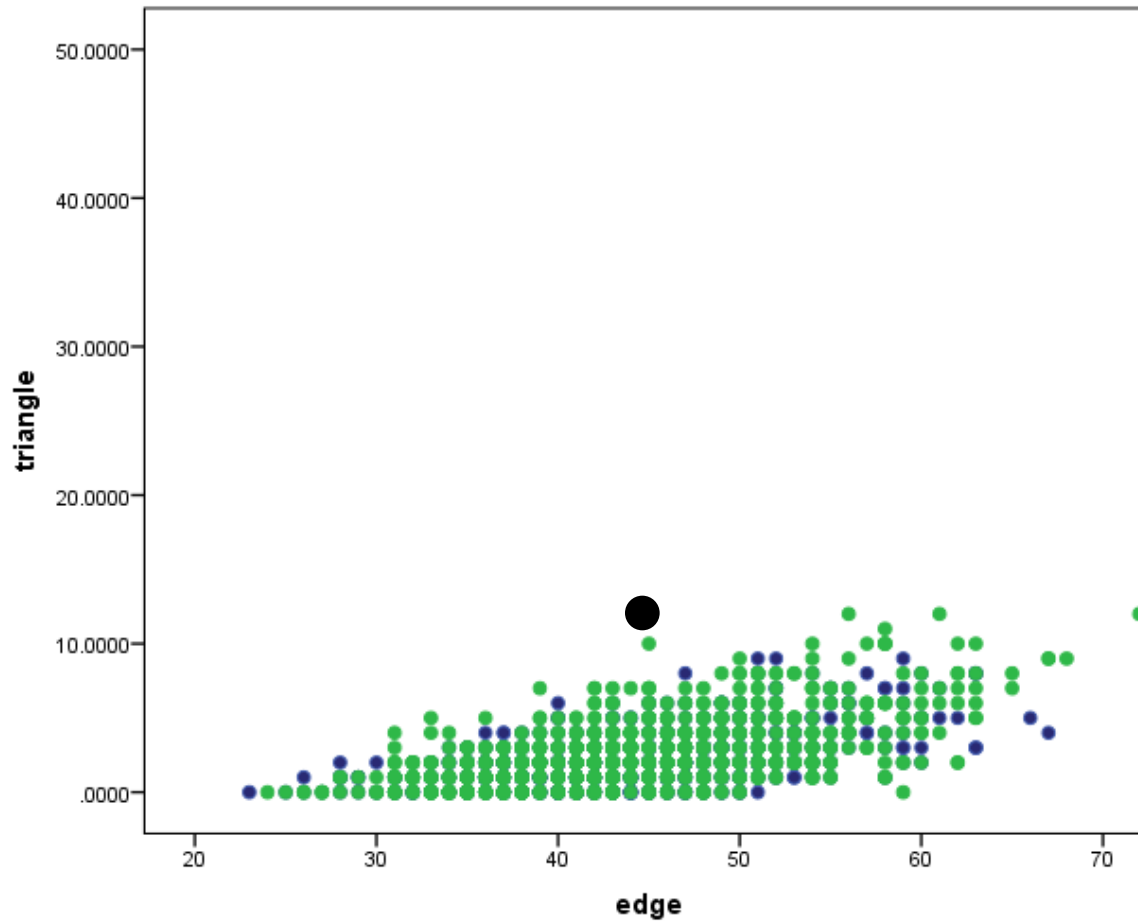


# Simulated results from Bernoulli graph



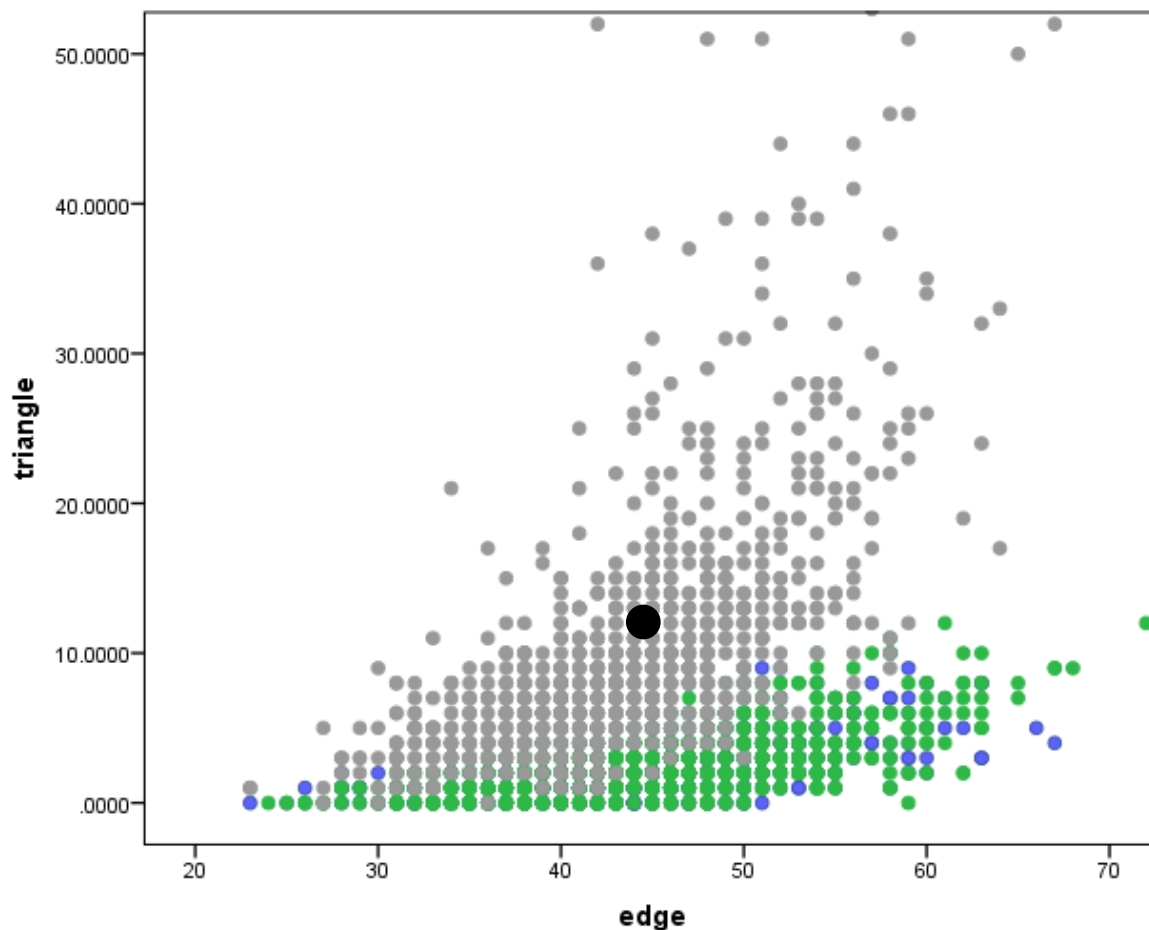
Statistics from a simulated sample of Bernoulli graph distribution (blue) around the observed statistic (N of triangles) ●

# Simulated results from Bernoulli graph



Statistics from a simulated sample of Bernoulli graph distribution (blue), plus 2-stars (green) around the observed statistic (N of triangles) ●

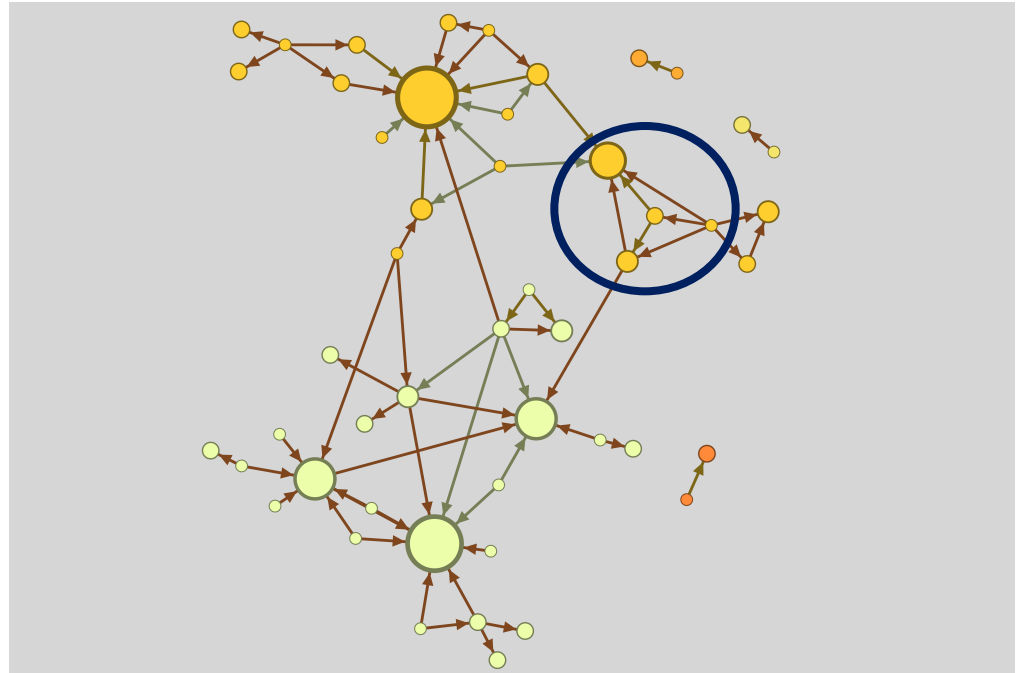
# Simulated results from Bernoulli graph



Statistics from a simulated sample of Bernoulli graph distribution (blue), plus 2-stars (green), plus 3-stars and triangle (grey: Markov model) around the observed statistic (N of triangles) ●

# Markov model

- Markov random graph distributions provide statistical models for social networks based on plausible assumptions and importantly can represent clustering through the triangle parameter 😊🧐



- The “leakage” shows a common problem with Markov models – they are not always stable; and may be degenerate 😱😓
- They can’t capture tie formation processes in the denser regions

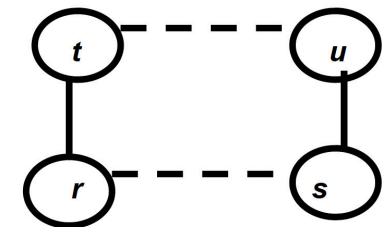
## Exercise 3

# Exercise 3

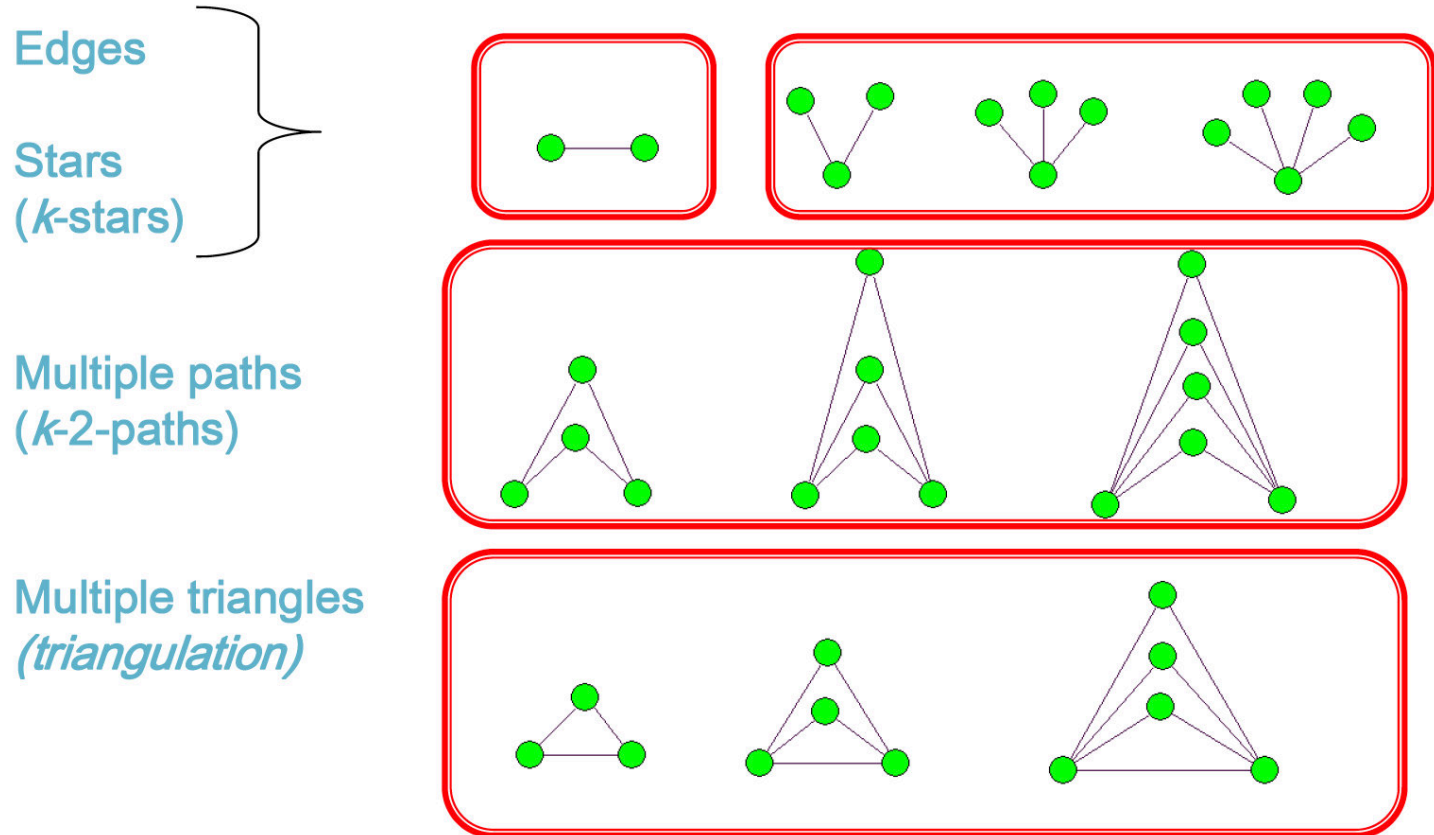
- Use the fishermen's network (85 nodes)
- Try to estimate an edge/2-star/3-star/triangle Markov model
- You may find the convergence ratios very bad

# Social circuit model

- Snijders, T.A.B., Pattison, P., Robins, G.L., & Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36, 99-153.
- Social circuit dependence (Network ties self-organize within 4-cycles): two possible network ties are conditionally dependent if they would form a 4-cycle (Pattison & Robins, 2002; Snijders et al, 2006)
- Larger network configurations emerge: parameters for degree sequences, denser regions of triangulation, multiple connectivity.
- This dependence assumption that captures emergence may be necessary to model real social networks (Robins, Snijders, Wang, Handcock & Pattison, 2007)



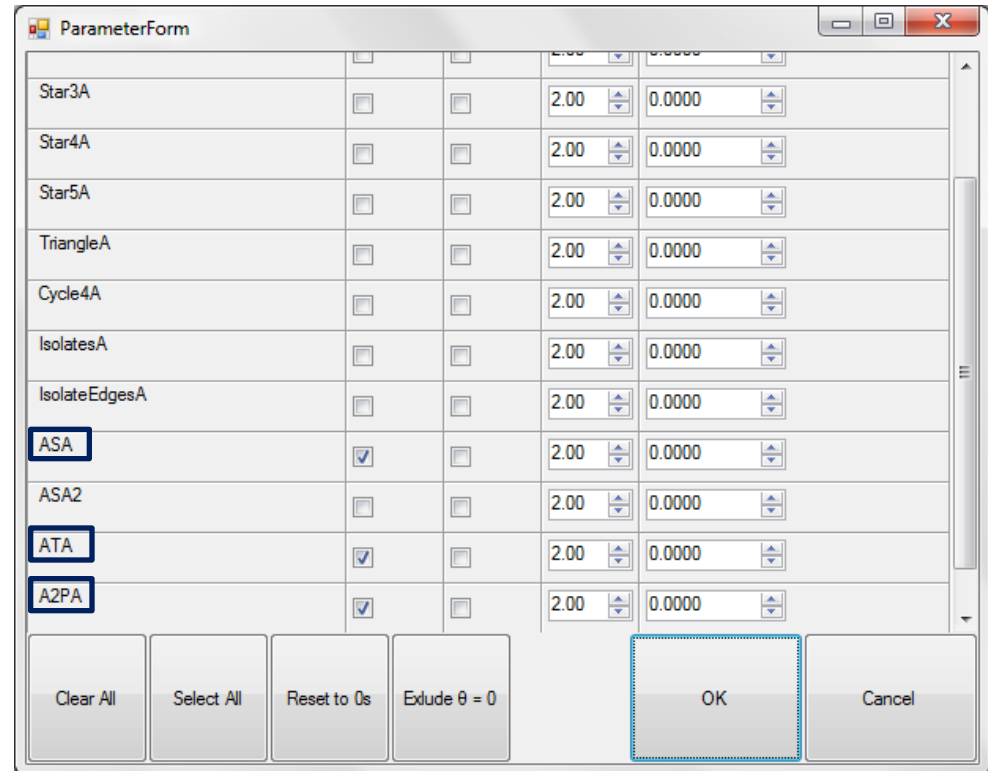
# Network configurations in SC models





# Possible parameters for SCMs in MPnet

- Edge
- Alternating star (AS)
- Alternating triangles (AT)
- Alternating 2-path (A2P)



Parameter	Checkbox 1	Checkbox 2	Value 1	Value 2
Star3A	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
Star4A	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
Star5A	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
TriangleA	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
Cycle4A	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
IsolatesA	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
IsolateEdgesA	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
ASA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
ASA2	<input type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
ATA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000
A2PA	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2.00	0.0000

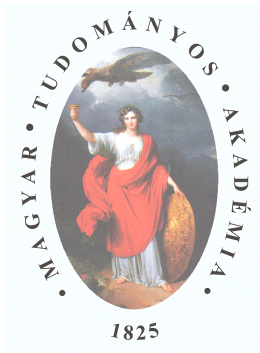
Buttons: Clear All, Select All, Reset to 0s, Exclude  $\theta = 0$ , OK, Cancel

## Exercise 4

# Exercise 4

- Use the fishermen's undirected network (85 nodes)
- Try to fit a Markov model
- Fit a SC model
- Use edges, AS, AT, A2P

## Questions ? !



European  
Research  
Council

