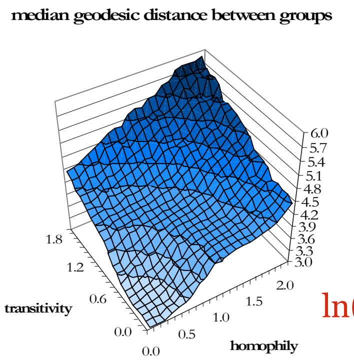
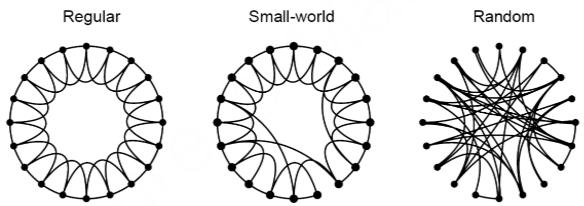
university of groningen / behavioural and social sciences / sociology

Workshop Social Network Analysis 2011

Exponential Random Graph models

Christian Steglich c.e.g.steglich@rug.nl





$$\ln\left(\frac{\Pr(x^c \rightarrow_i x^b)}{\Pr(x^c \rightarrow_i x^a)}\right) = \sum_{k=1}^{K} \beta_k \left(s_{ik}(x^b) - s_{ik}(x^a)\right)$$

Overview of topics

- Revisiting interdependence
- The exponential random graph approach
  - Markov dependence (Frank & Strauss, 1986)
  - Higher order dependence
- Local characterisation of ERGMs
- An application: Gossip about the boss (Ellwardt, 2008)
- A little bit on estimation / model identification

Revisiting interdependence

How do the different approaches studied so far …

(a) allow to test hypotheses on network data?

Typically, the "actual hypotheses" are not so much about structure at all – but about who is central, who links up with whom, etc.

(b) address the issue of interdependent data?

Interdependence in the above view is seen as a "nuisance", i.e., as something that needs to be taken into account, but not as something of focal interest.

Conditional uniform models

(e.g., tie independence, tie independence given the dyad census, tie independence given the degree distributions…)

a) Hypotheses are tested by working with a network distribution that enforces a selection of structural constraints.

b) Interdependence is taken into account as far as the structural constraints already imply it.

Here, structure is not treated as *endogenous* (part of the dependent variables, “to be explained”), but as *exogenous* (here even: enforced). It’s not always clear what to enforce. Approach is not very flexible.

Permutation-based modelling

a) Hypotheses are tested (like in conditional uniform tests) by working with a network distribution that enforces structural constraints – here the “total structure” even.
b) Interdependence is taken into account by completely fixing the network structure.

Also here, structure is treated as exogenous. This means it is difficult to study how dependencies due to structure and dependencies due to explanatory variables interact. Better would be an approach where both figure in similar positions in the model – which brings us to…

The p2 model and the social relations model

a) Hypothesis tests are done based on parameter estimation / model fitting. The distribution of networks is not fixed (as in previous approaches) but modelled by a parametric family of models.

b) Within-dyad dependence is modelled through correlation, between-dyad dependence through common sender and/or common receiver effects.

Here, structure is treated as endogenous – but in a limited sense. Triad level dependencies that are not due to common sender or common receiver effects cannot be expressed this way: transitivity, social balance, preferential attachment… ERGMs allow for this!

The exponential random graph approach

- Definition ERG model
- Markov dependence (Frank & Strauss, 1986)
- Higher order dependence
- Local characterisation of ERGMs
- An application: Gossip about the boss (Ellwardt, 2008)

Defining the Exponential Random Graph model

- history: Frank & Strauss "Markov random graph" model (1986), Frank (1991) and Wasserman & Pattison (1996) generalised to exponential family distribution:

$$\Pr(X = x) \propto \exp\left(\sum_{k=1}^{K} \beta_k s_k(x)\right)$$

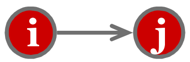
'proportional to'
proportionality constant unknown and practically uncalculable(!)

linear combination of network statistics
very flexible!

*Likelihood of macro structure **x** is explained by prevalence of micro structures **s(x)**, testable via parameters β.*

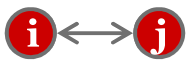
Some possible statistics *(directed case)*

- tie count statistic



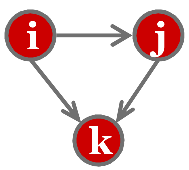
$$s_k(x) = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} x_{ij}$$

- reciprocity statistic
  (*for directed graphs*)



$$s_k(x) = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} x_{ij} x_{ji}$$

- transitive triplets

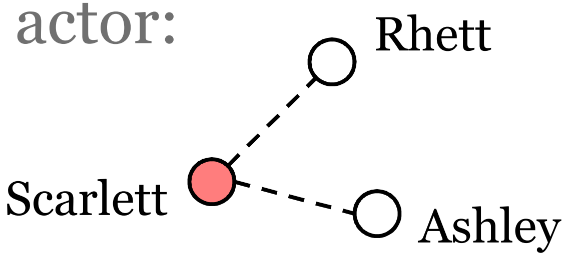


$$s_k(x) = \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} x_{ij} x_{jk} x_{ik}$$

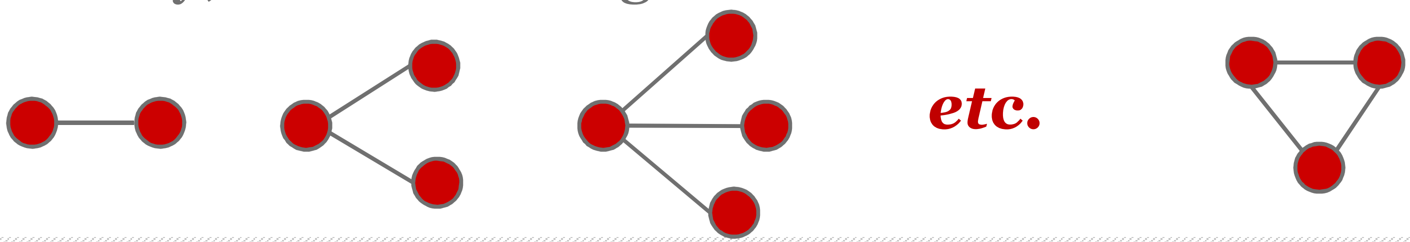
A positive parameter β attached to any of these statistics means: *presence* of the configuration is more likely than *absence*.

Markov dependence (Frank & Strauss, 1986)

› Assume two tie variables are dependent when they share an actor:

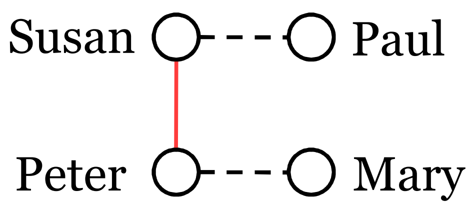


› Frank & Strauss showed that then tie probabilities can be expressed by an ERG model that includes just density, star and triangle effects.



Higher order dependence

- Pattison & Robins (2002) proposed more general dependence structures than Markov dependence, such as realisation-dependent models.



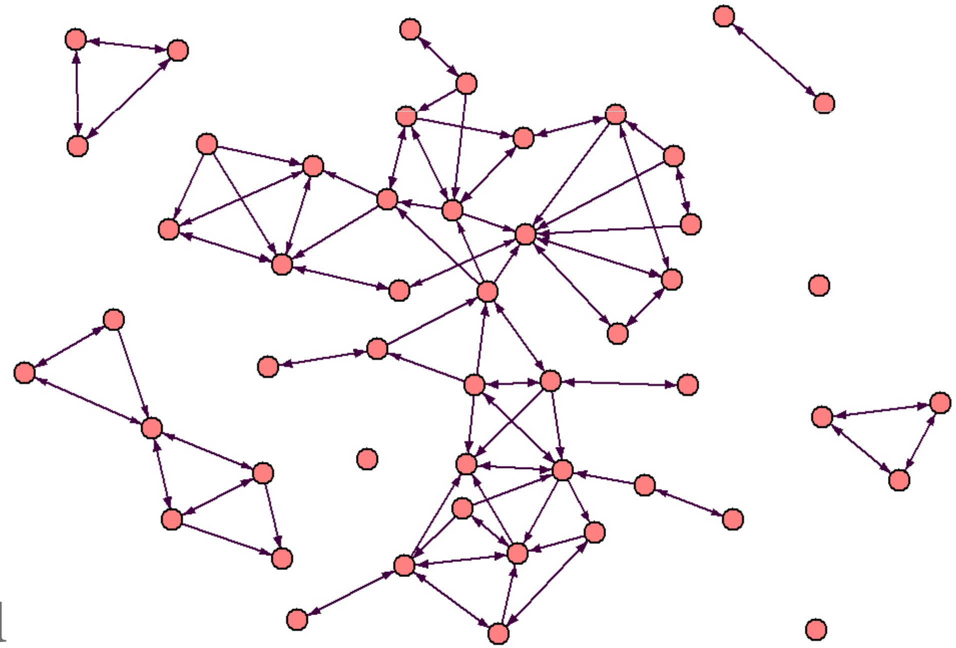
Two non-overlapping tie variables depend on each other if there is at least one tie present that connects one actor in the one tie variable to one actor in the other tie variable.

- Snijders, Pattison, Robins & Handcock (2006)

Let's try fitting a Markov specification 50 Scottish girls' nominated friendship

data from Michell & Amos (1997)

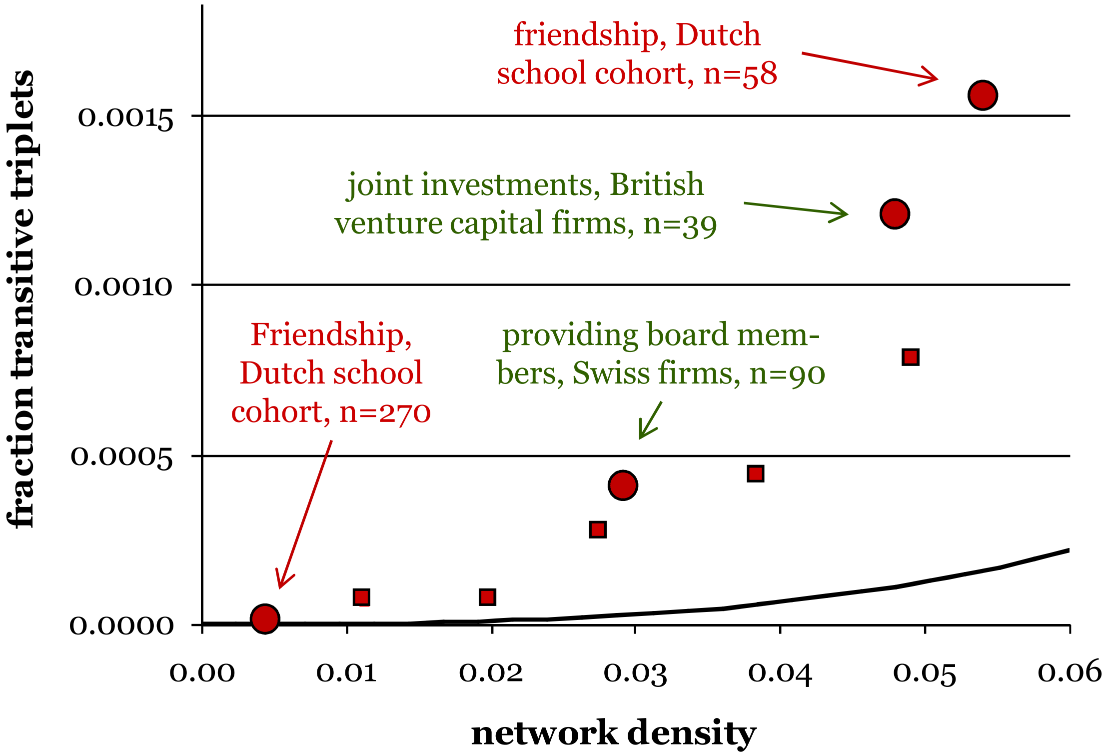
- 50 girls
- Scotland 1995
- 1^{st} year secondary
- 13 years old



Low density,
reciprocity, and
triangulation can be observed

Triangulation is ubiquitous in network data:

In many data sets, significantly more triangles are observed than would be expected under random tie formation: evidence for <u>*groups*</u>.



Red markers stand for empirically observed data sets, **black line** for expectations under randomness.

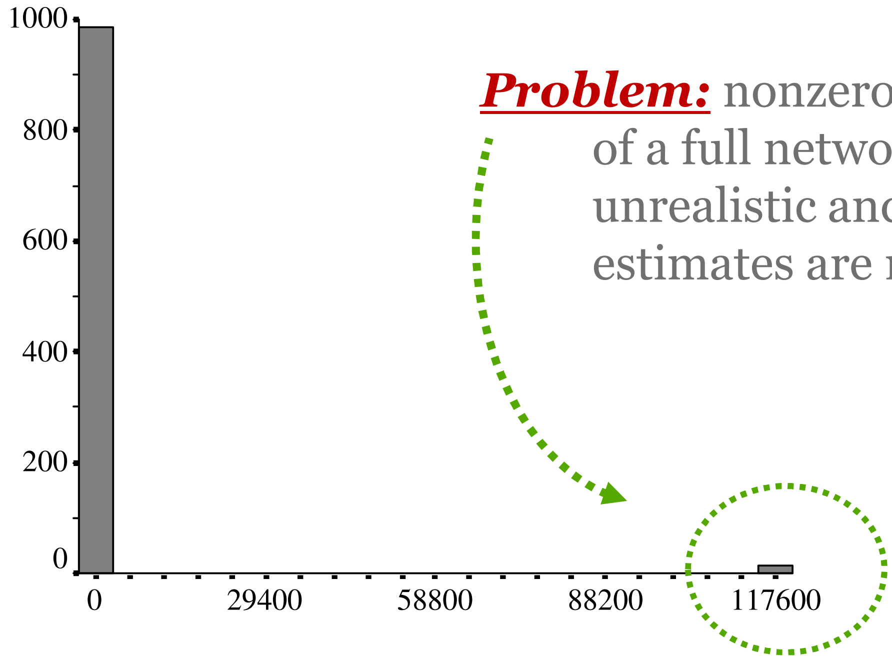
Method of moments estimates for a very simple model with a linear triangulation effect

Just three parameters (*the rate parameter was fixed at value* λ=625):

Parameter	***estimate***	***st.error***		t_{conv}
outdegree	-4.55	(0.13)	***	0.090
reciprocity	5.99	(0.23)	***	0.072
transitive triplets	0.28	(0.03)	***	0.054

Convergence statistics might suggest reasonable model fit…

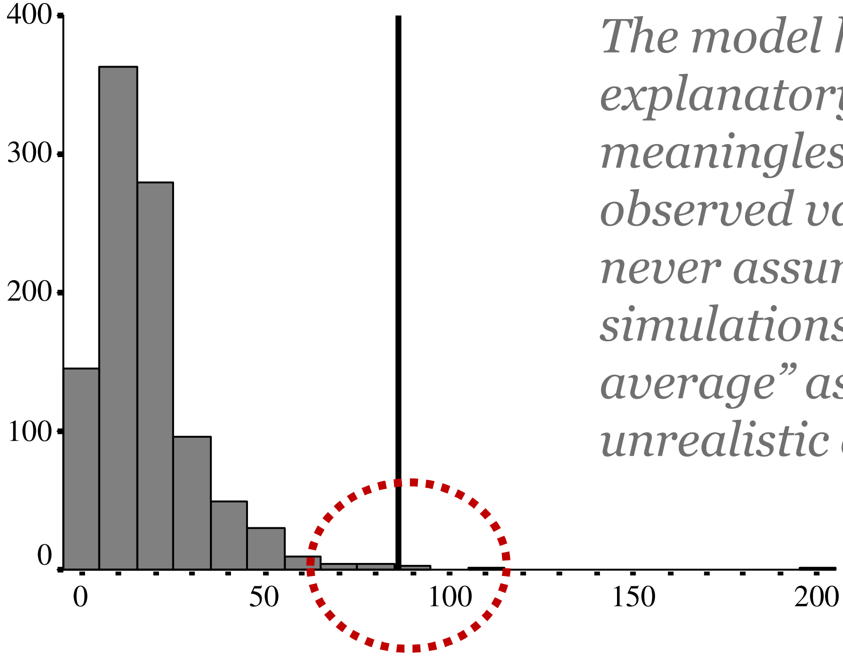
*…**but** see here the distribution of the number of transitive triplets over simulated networks from such a model…*



Problem: nonzero probability of a full network! This is unrealistic and implies that estimates are meaningless…

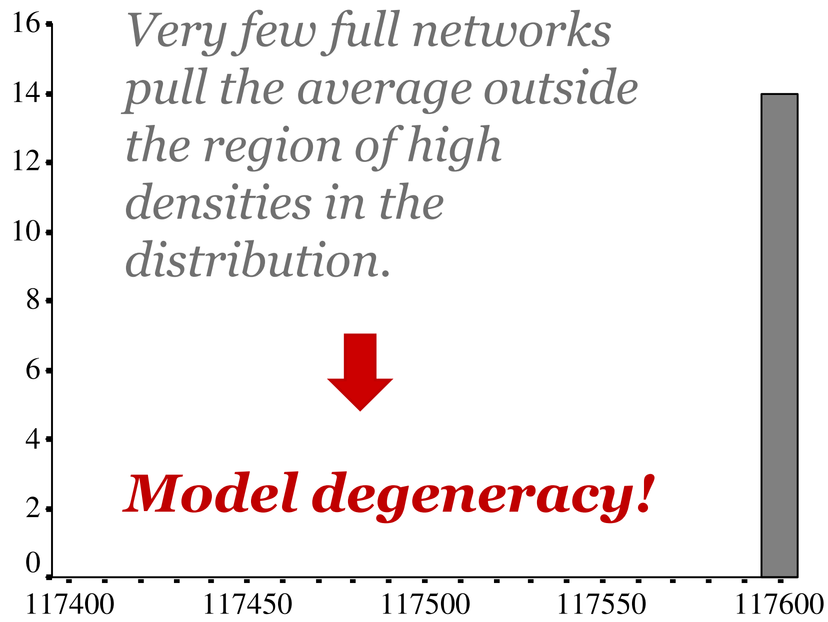
…zooming in on the lower end of the scale ("realistic networks"):

vertical line = observed value



The model has no explanatory power. It is meaningless because the observed value is almost never assumed by simulations – only "on average" as mixture of two unrealistic extremes!

...zooming in on the upper end of the scale (unrealistic networks):



Very few full networks pull the average outside the region of high densities in the distribution.

Model degeneracy!

What is the problem?

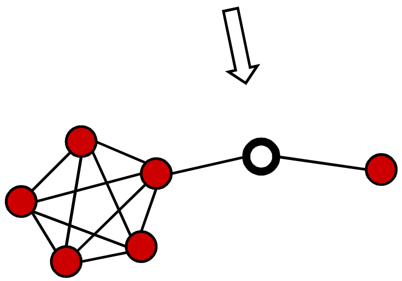
Naive inclusion of a linear transitive closure effect leads to what is known by the names of

- ***explosions*** (Snijders, 2002), ***avalanches*** (Handcock, 2003), or
- ***phase transitions*** (Bianconi, Marsili & Vega-Redondo, 2005)

in the model-based simulations because the mechanism is *inherently self-accelerating*.

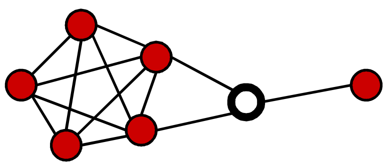
A dense network region can act as ‘critical mass’ to pull in connected nodes…

Illustration of triangulation avalanche



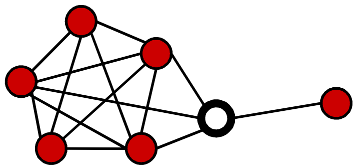
The marked actor is peripherally attached to a small clique of size ***n*** .

- The ***n*-1** other clique members pull with the weight of *one* transitive triad each.
- The bigger the clique, this higher the chances such a tie will be established.



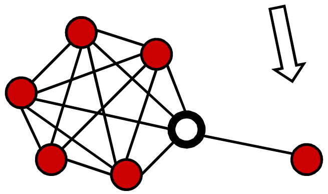
Eventually, the actor establishes another tie to the clique.

- Now the ***n*-2** other clique members pull with the weight of *two* transitive triads each.
- The attractiveness to establish more connections to the clique increases.



Eventually, the actor establishes a third tie to the clique.

- Now the ***n*-3** other clique members pull with the weight of <u>*three*</u> transitive triads each.
- etc.

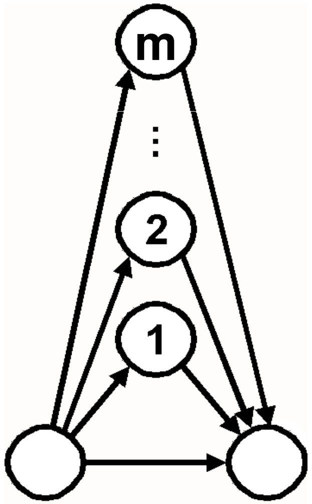


Eventually, the actor has become absorbed by the clique.

- Next connected actor experiences same forces, but due to increased size of the clique, the pull will be stronger.
- Clique acts as 'black hole' sucking connected nodes in – and stochastically, the whole network is connected…

What is the solution?

operationalise triangulation differently …



'new specifications' as introduced by (Snijders et al. 2006)

$$s_k(x) = \sum_{m=1}^{n-2} (-1)^{m+1} \frac{\tau_m}{2^{m-1}}$$

"alternating m-triangles statistic"

τ_m = number of m-triangles (configurations on the left)

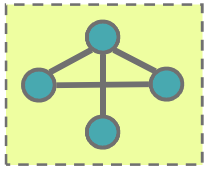
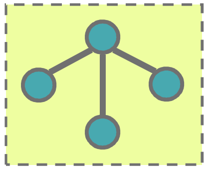
More on this later…

Some example calculations for ERGMs

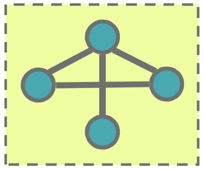
- Consider an ERG model for an *undirected* network with parameters for these three statistics:
  - (1) number of edges $s_{edges}(x) = \sum_{i<j} x_{ij}$
  - (2) number of 2-stars $s_{2-stars}(x) = \sum_{i;j<k} x_{ij}x_{ik}$
  - (3) number of triangles $s_{triangles}(x) = \sum_{i<j<k} x_{ij}x_{jk}x_{ik}$
- Then the 3-parameter ERG distribution function is this one:

$$\Pr(X = x) \propto \exp(\beta_{edges} \times s_{edges}(x) + \beta_{2-stars} \times s_{2-stars}(x) + \beta_{triangles} \times s_{triangles}(x))$$

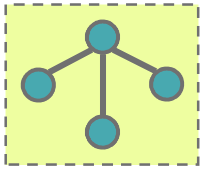
› ...and consider the following two 4-node-networks & their statistics:

		
	x^a	x^b
s_{edges}	4	3
$s_{2\text{-}stars}$	5	3
$s_{triangles}$	1	0

Probabilities are (only) given up to a proportionality factor:



x^a

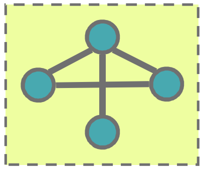


x^b

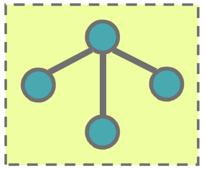
$$\Pr(x^a) \propto \exp(4\times\beta_{edges} + 5\times\beta_{2\text{-}stars} + 1\times\beta_{triangles})$$

$$\Pr(x^b) \propto \exp(3\times\beta_{edges} + 3\times\beta_{2\text{-}stars})$$

...but the ratio of probabilities can be calculated!



x^a



x^b

$$\frac{\Pr(x^a)}{\Pr(x^b)} = \frac{\exp(4\beta_{edges} + 5\beta_{2-stars} + \beta_{triangles})}{\exp(3\beta_{edges} + 3\beta_{2-stars})}$$

$$= \exp\Big((4-3)\beta_{edges} + (5-3)\beta_{2-stars} + \beta_{triangles}\Big)$$

$$= \exp(\beta_{edges} + 2\beta_{2-stars} + \beta_{triangles})$$

Local characterisation of ERGMs

› ERG distribution is considered as a collection of local conditional tie probabilities

$$\frac{\Pr(x^a)}{\Pr(x^b)} = \exp\left(\sum_{k=1}^{K} \beta_k \underbrace{\left(s_k(x^a) - s_k(x^b)\right)}\right)$$

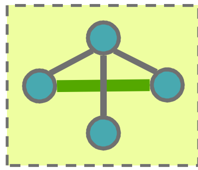
compared are a network $\boldsymbol{x^a}$ and a "neighbouring network" $\boldsymbol{x^b}$

differences in model statistics between $\boldsymbol{x^a}$ and $\boldsymbol{x^b}$

What this equation refers to is the (conditional) probability of the one tie by which they differ!

model parameters

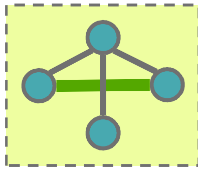
- …so how do these 'conditional odds' for the middle tie to exist (vs. not to exist) look like in applications?
  - Suppose, in a (larger) trade network, estimation gave:
    - redundant ties are avoided: $\beta_{triangles}$= -0.4
    - positive degree variance: $\beta_{2\text{-}stars}$= 0.1
    - low density: β_{edges}= -1.5



- Then the equation becomes:

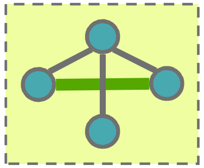
$$\frac{\Pr(\text{middle tie})}{\Pr(\text{no middle tie})} = \exp(\beta_{edges} + 2\beta_{2-stars} + \beta_{triangles})$$
$$= \exp(-1.5 + 2 \times 0.1 - 0.4)$$
$$= \exp(-1.7)$$
$$\approx 0.183$$
$$\approx 1/5.5$$

i.e., the middle tie is about
5.5 times as likely NOT to exist
as to exist [given the rest of the network].



› Same model in a different application:

- Suppose, in a (larger) <u>friendship</u> network, estimates were:
  - closure: $\beta_{triangles} = 0.4$
  - pos. degree variance: $\beta_{2\text{-}stars} = 0.1$
  - low density: $\beta_{edges} = -1.5$

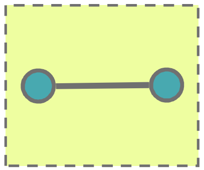


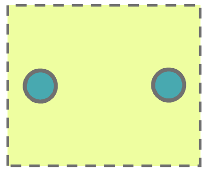
- Then the equation becomes:

$$\frac{\Pr(\text{middle tie})}{\Pr(\text{no middle tie})} = \exp(-0.9) \approx 0.407 \approx 1/2.5$$

- Still two½ times as likely NOT to exist – but always keep in mind that *all* ties are random…

…random ties are benchmark for evaluating such conditional odds!
Hence, consider these “networks” as baseline comparison:





$$\frac{\Pr(\text{tie})}{\Pr(\text{no tie})} = \exp(\beta_{edges}) = \exp(-1.5) \approx 0.223 \approx 1/4.5$$

- In the *friendship context* (odds=2.5), the tie in question is *relatively likely* to be present,
- In the *trade context* (odds=5.5), it is *relatively unlikely* to be present, given the network’s density.

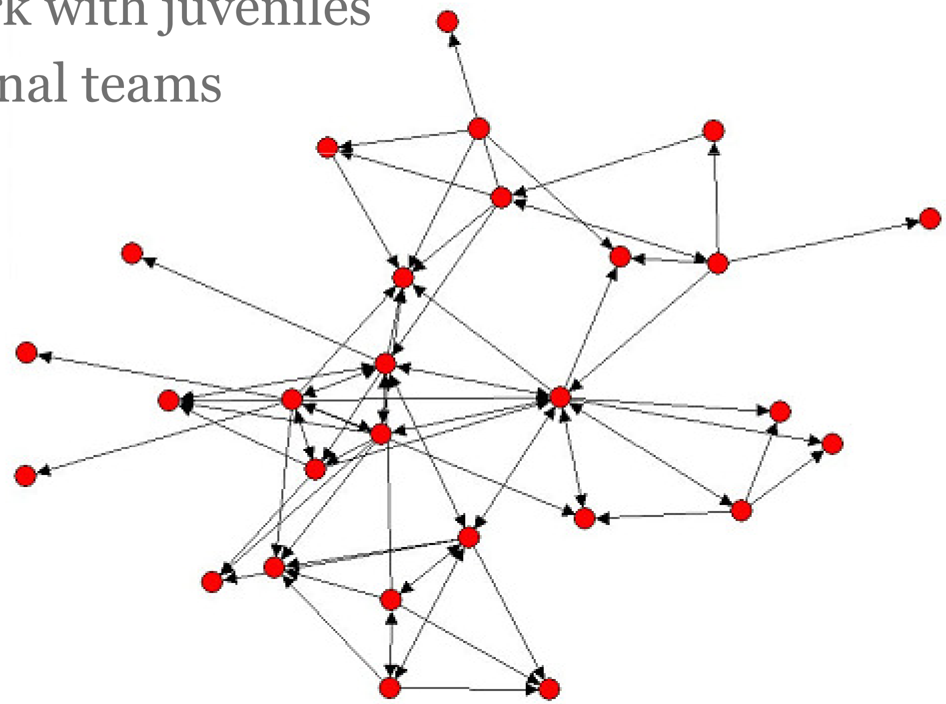
Some ERGM features

- high flexibility due to the many possibilities of choosing statistics & controlling effects for each other +
- can, e.g., express small world networks (see Robins et al. 2005) but also other network types
- network probability model: not tied to any particular algorithm+
- problems: estimation, model specification, interpretation –

An example: gossip about the boss

› **data from Lea Ellwardt (2008)**
  - organisation for social work with juveniles
  - one department with internal teams
  - 28 employees

› *gossip about the boss in this department was largely <u>negative</u>*



Hypotheses:

1. closeness / proximity is associated with gossip
    - friendship / liking predicts gossip
    - communication frequency predicts gossip
    - team structure predicts gossip
    - gossip occurs in local clusters
2. information asymmetry is associated with gossip
    - dissimilarity in / lack of contacts with the boss predicts g'sp.
3. "negative attitude" is associated with gossip
    - distrust in management predicts gossip
    - disliking the boss predicts gossip
4. attitude similarity is associated with gossip

Results from an exponential random graph analysis (1)

Parameter	*Estimate*	*SE*	
NETWORK EFFECTS			
1. Reciprocity	0.9356	0.6254	
2. Transitive triplets	-0.1851	0.2873	… (1)
3. 3-cycles	-0.1362	0.3619	… (1)
4. Alternating out-k-stars	**0.8823***	**0.3179**	← – – – **Unpredicted: heterogeneity of outdegrees**
5. Alternating in-k-stars	-0.2607	0.3701	
6. Alternating k-triangles	**0.9762***	**0.4834**	✓ (1)
7. Team membership	**0.7866***	**0.2835**	✓ (1)
8. Communication ego-alter (symmetric)	**0.9111***	**0.3254**	✓ (1)
9. Liking of alter by ego (out-degree)	**1.8782***	**0.3305**	✓ (1)

Results from an exponential random graph analysis (2)

Parameter	*Estimate*	*SE*	
SENDER EFFECTS			
10. Trust in management	-0.1261	0.1337	… (3)
11. Liking of manager	**-0.5071***	**0.2289**	✓ (3)
12. Communication with manager	**-0.3239***	**0.0915**	✓ (2)
RECEIVER EFFECTS			
13. Trust in management	0.2187	0.1996	… (3)
14. Liking of manager	0.2244	0.3588	… (3)
15. Communication with manager	-0.0445	0.1411	… (2)

Results from an exponential random graph analysis (3)

Parameter	*Estimate*	*SE*	
SIMILARITY EFFECTS			
16. Trust in management (similar)	**-1.3708***	**0.7921**	× (4)
17. Liking of manager (same)	0.3919	0.2878	… (4)
18. Communication with manager (similar)	-0.4359	0.6343	… (4)

Against prediction: similar attitude implies LESS gossip

MAYBE ALSO information asymmetry?

* p<0.05 (one-sided test)

Conclusion: the analysis supports the hypotheses (1), (2) and (3), but refutes hypothesis (4)

Estimation of ERGMs

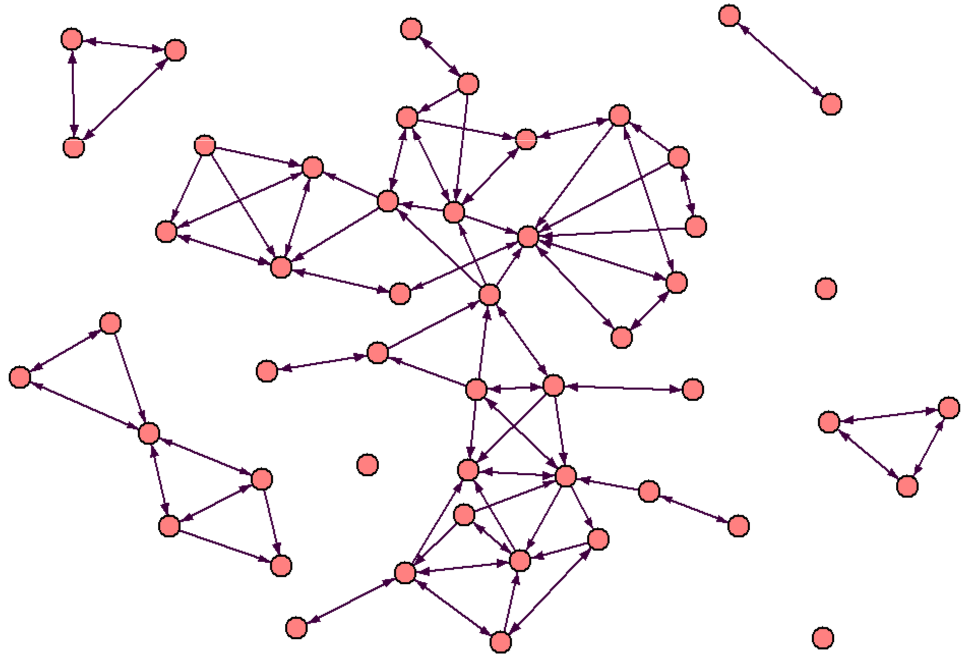
As for actor-based evolution models, estimation is based on simulation:

- Markov chains are constructed that deliver the ERGM's probability distribution over the network space as equilibrium distribution;
- using these, maximum likelihood estimates can be derived.

This unfortunately does not always work as it should, and notably triangulation effects cause problems…

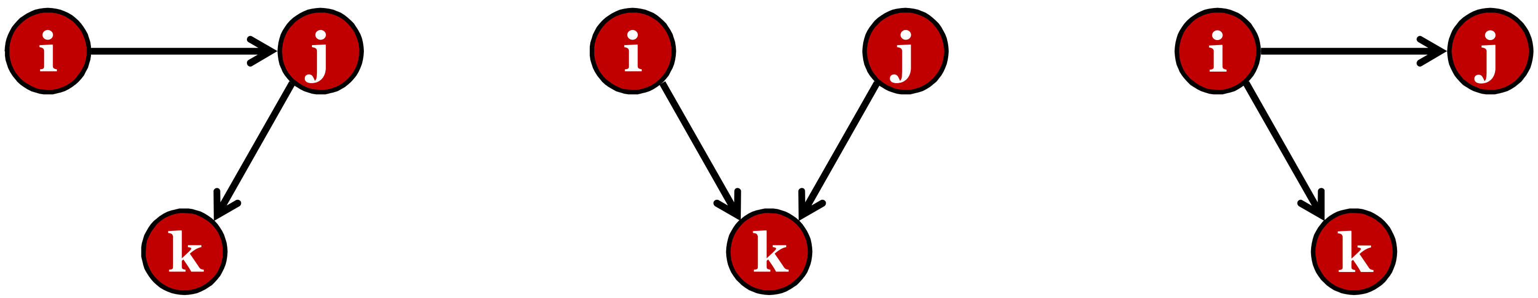
Back to the 50 Scottish girls' friendship network

How to model the triangulation tendencies in this data set when specifying exponential random graph models?



Naïve try: *estimate a relatively simple model with triangulation / transitive closure effects*

include *<u>precursor configurations</u>* in attempt to arrive at a meaningful model (Snijders 2002)



2-path in-2-star out-2-star

This is a quite typical '<u>p* estimation</u>' result: *a main effect of the transitive triplet count SEEMS TO BE identified…*

essentially, this is logistic regression (dubious, biased!)

```
Maximum Pseudolikelihood estimates

Estimates and standard errors

 1. degree (density)              -3.2617  (   0.4065)
 2. reciprocity                    4.0064  (   0.3278)
 3. transitive triplets            1.0976  (   0.1117)
 4. out-2-stars                   -0.3563  (   0.1453)
 5. in-2-stars                     0.1235  (   0.0970)
 6. 2-paths                       -0.3111  (   0.0934)
```

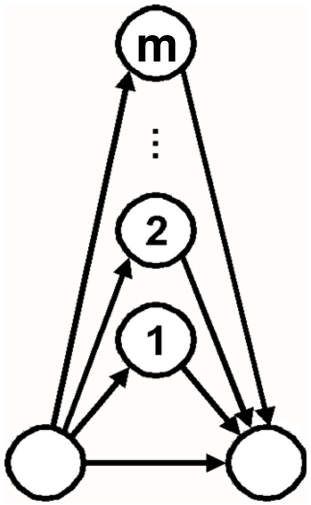
Trying to estimate the same model by 'proper ML estimation' fails: models with main effect of transitive triplet count diverge…

```
Information for convergence diagnosis: Averages, standard
   deviations, and t-ratios for deviations from targets.

  1.    -10.436    6.973   -1.497
  2.     -5.584    3.727   -1.498
  3.    -49.242   18.012   -2.734
  4.    -23.362   13.734   -1.701
  5.    -27.000   19.552   -1.381
  6.    -52.408   28.092   -1.866

One or more of the t-statistics are rather large.
Convergence of the algorithm is doubtful.
```

Second try: estimate another model with 'new specifications' (Snijders et al. 2006)



$$s_k(x) = \sum_{m=1}^{n-2} (-1)^{m+1} \frac{\tau_m}{2^{m-1}}$$

"alternating m-triangles statistic"

τ_m = number of m-triangles
(configurations on the left)

In the R-module 'ergm', a corresponding effect would be geometrically weighted edgewise shared partners, a.k.a. 'gwesp'.

Now, 'proper ML estimation' is successful as well: *models with new specifications for transitivity etc. generally converge…*

```
Information for convergence diagnosis: Averages, standard
   deviations, and t-ratios for deviations from targets.

  1.      -1.076   12.405   -0.087
  2.      -0.804    6.917   -0.116
  3.      -1.509   17.614   -0.086
  4.      -1.595   17.990   -0.089
  5.      -3.444   24.797   -0.139
  6.      -2.949   45.755   -0.064

Good convergence is indicated by the t-ratios being close to
   zero.
```

...the underlying reason is that under new specifications, the triplet count distribution is modelled as a unimodal one.

```
Estimates and standard errors

 1. degree (density)                              -2.3372  (0.5101)
 2. reciprocity                                    3.7823  (0.4436)
 3. alternating out-k-stars, par. 2               -1.0196  (0.3945)
 4. alternating in-k-stars, par. 2                -0.2799  (0.3421)
 5. alternating k-triangles, par. 2                1.0973  (0.1433)
 6. alternating independent twopaths, par. 2      -0.1429  (0.1200)
```

A bit problematic, however, is the not-so-straightforward interpretation of the new effects... *[but stay tuned, this is work in progress...]*

Goodness of fit criteria for ERGMs (1)

› The 'convergence t-statistics' indicate goodness of fit on the dimensions of the model's effects:

$$t = \frac{E_{\text{simulations}}(\text{simulated count} - \text{observed count})}{\text{st.dev.}_{\text{simulations}}(\text{simulated count} - \text{observed count})}$$

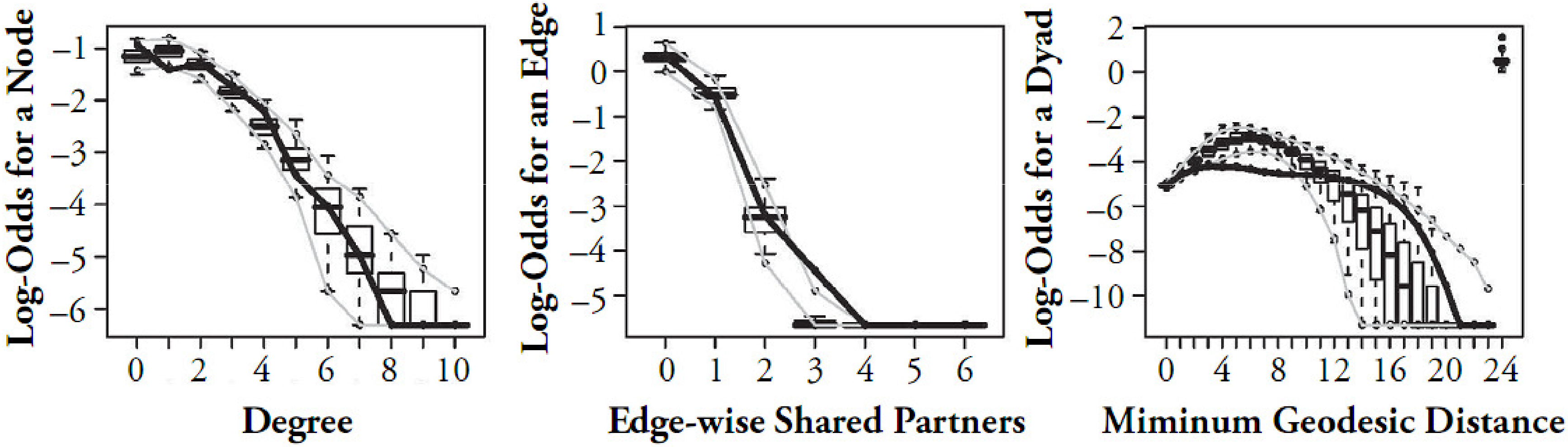
If small, simulated statistics (i.e., those that the identified model would predict) are on average identical to observed statistics.

› The same type of statistics can be evaluated also for *other dimensions of the network*, such as degree distributions, geodesic distributions, etc.

This is how SIENA & PNet evaluate goodness of fit.

Goodness of fit criteria for ERGMs (2)

But why focus only on the t-statistic?



› In the ergm-package, the *whole distribution* of simulated statistics can be checked against observed values.

Allows more detailed inspection of the model's predicted distributions.

Revisiting interdependence, again

- ERGMs instantiate interdependence of tie variables x_{ij} by including ‘beyond the tie’ local network configurations (“motifs”) as predictors
- “Markov dependence” (Frank & Strauss):
  - tie variables dependent when they share an actor
  - exhaustive model: edges, stars, triangles
- Higher order dependence (Pattison, Robins, Snijders, Handcock, Hunter)
  - 2006 models: tie variables dependent when there exists a dyad-connecting third tie variable
  - No easy ‘exhaustive model’

To keep in mind about ERGMs:

- Visualise ERG models as probability distributions on a (huge) space of all possible network,
- one observed network is modelled as drawn from that distribution.
- Model parameters β are
  - attached to network statistics ***s***,
  - these statistics in general correspond to subgraph counts (local patterns, 'motifs'),
  - the parameters describe the relative prevalence of the corresponding subgraph in the total graph.
- Interpretation of parameters is similar to actor-based evolution models; needs to take into account other parameters.

References for ERGMs

Frank, Ove. 1991. Statistical Analysis of Change in Networks. *Statistica Neerlandica*, 45: 283-293.

Frank, Ove, and David Strauss. 1986. Markov Graphs. *Journal of the American Statistical Association*, 81: 832-842.

Handcock, Mark S. (2003). Assessing degeneracy in statistical models of social networks. Working paper no. 39, Center for Statistics and the Social Sciences. Seattle: University of Washington.

Robins, Garry, Pip Pattison, Yuval Kalish, and Dean Lusher (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks,* 29: 173-191.

Snijders, Tom A.B. (2002) Markov Chain Monte Carlo Estimation of Exponential Random Graph Models Journal of Social Structure. Volume 3, number 2.

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock (2006). New specifications for exponential random graph models. *Sociological Methodology*.