Actor-driven alternatives

to exponential random graph models

Christian Steglich

ICS / University of Groningen

Sunbelt XXVI, Vancouver, April 24-30, 2006

Outline

1. Stochastic models for complete networks
   - ▷ *exponential random graph models*
   - ▷ *actor-driven models*
2. Why these models?
   - ▷ *reasons, objections, possible purposes*
3. An example: modelling network cohesion
   - ▷ *triadic closure and model degeneracy*
   - ▷ *fit issues*
4. Perspectives

1. Stochastic models for complete networks

- specify a probability distribution Pr
- over the state space $\Delta^{\frac{n(n-1)}{2}}$ of all networks $X = (x_{ij})$
- possible on a given actor set $\{1, ..., n\}$.

Δ *stands for the set of possible network states constrained to a single pair* $\langle i, j \rangle$ *of actors* $i, j \in \{1, ..., n\}$.

For undirected graphs: $x_{ij} = x_{ji} \in \Delta = \{0, 1\}$.

For directed graphs: $\langle x_{ij}, x_{ji} \rangle \in \Delta = \{\langle 0, 0\rangle, \langle 0, 1\rangle, \langle 1, 0\rangle, \langle 1, 1\rangle\}$.

Problem for stochastic inference: **state space is very big!**

Exponential Random Graph (ERG, aka p*) models

(Wasserman & Pattison 1996)

Probability of network X is given by

$$\Pr(X) \propto \exp\left(\sum_k \gamma_k g_k(X)\right)$$

with model parameters γ_k weighting network statistics $g_k(X)$.

<u>Local characterisation</u> can be given in terms of conditional log-odds for changes of a tie variable $x_{ij} \rightarrow (1 - x_{ij})$ (tie from node i to node $j \neq i$, '*micro step*').

Estimation can be based on construction of a Markov chain that has the ERG model as equilibrium distribution.

Actor-driven models for network evolution

(Snijders 1996)

Network change over time t is propelled by *actions* of the nodes:

- nodes i get *opportunities* to change their neighbourhood at rate $\lambda_i(X(t)) = \exp\left(\alpha_0 + \sum_k \alpha_k a_{ki}(X(t))\right)$
- given an opportunity, node i optimises its neighbourhood
  - based on an objective function $f_i(X(t)) = \sum_k \beta_k b_{ki}(X(t))$
  - over a choice set consisting of micro steps $x_{ij} \to (1 - x_{ij})$ (change tie to node $j \neq i$), plus the option of no change,

with parameters α_k, β_k weighting node-specific network statistics $a_{ki}(X)$, $b_{ki}(X)$.

Actor-driven models for a single network

The dynamic model describes a *continuous-time Markov process* and implies a stationary distribution over the state space $\Delta^{\frac{n(n-1)}{2}}$.

$\Rightarrow$ *Model the single observed network as drawn from this equilibrium distribution.*

rate function models differential network activity of nodes

objective function models behavioural rules of nodes

Estimation can be based on existing methods for network evolution (MoM, Snijders 2001; ML, Snijders 2006; Bayesian, Schweinberger 2006) and a limit construction.

Both models

- ▷ describe equilibrium distributions of a stochastic process
- ▷ in terms of *rules for tie change* (micro steps)

ERG models assume a global-structural perspective

- ▷ model the impact of a micro-step on global statistics
- ▷ the network-generating process is of secondary concern

Actor-driven models assume a constrained, local perspective

- ▷ model impact of a micro-step on neighbourhood statistics
- ▷ the network-generating process is central

2. Why these models?

Motive: understanding network evolution models

- ▷ problems with longitudinal analysis of certain data sets
- ▷ extrapolation into the future
- ▷ how would the absence of evolution look like?

Problematic: stationarity assumption

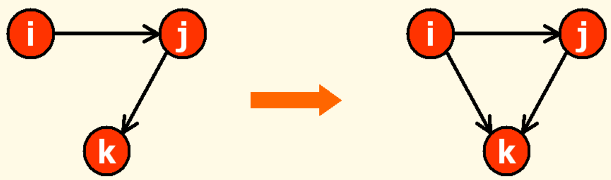
- ▷ rarely is a social system in equilibrium, but
- ▷ evolution of many empirical networks slows down over time

Potential: conditional realism

- ▷ if a research perspective of methodological individualism is warranted, the model seems a good choice

3. An example: modelling network cohesion

Empirical phenomenon (I): **transitive closure**



"friends of my friends are my friends"

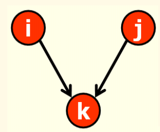
In many networks, there are significantly more transitive triplets than expected under random tie formation.

⇒ include main effect of transitive closure in any explanatory model for such data

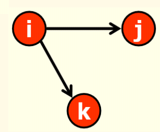
Main effect of transitive closure

Ties are more likely to be present when their removal would result in...

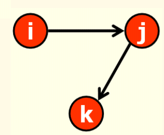
- an *in-2-star*,



- an *out-2-star*, or



- a *2-path*.



A positive main effect of transitive closure implies that the *presence* of such configurations is unlikely. But...

Empirical phenomenon (II): **this is not necessarily the case!**

in-2-stars, out-2-stars and 2-paths co-exist with transitive triplets.

Problem for both model families

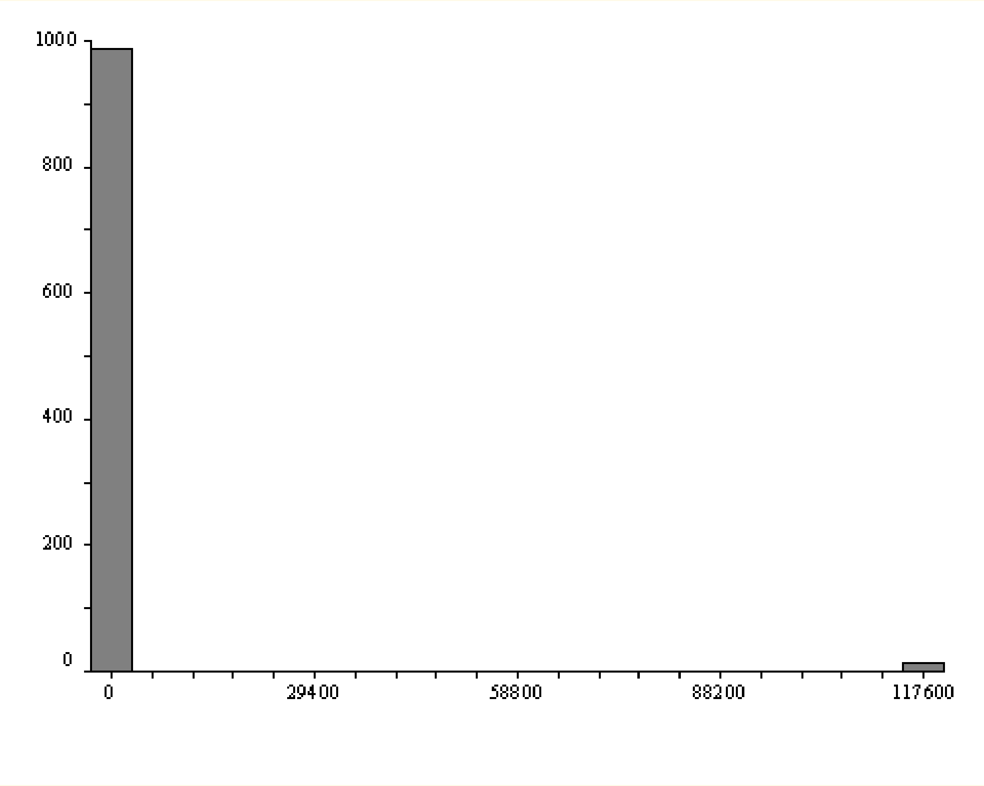
Naïve inclusion of transitive closure effects leads to what is known by the names of

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

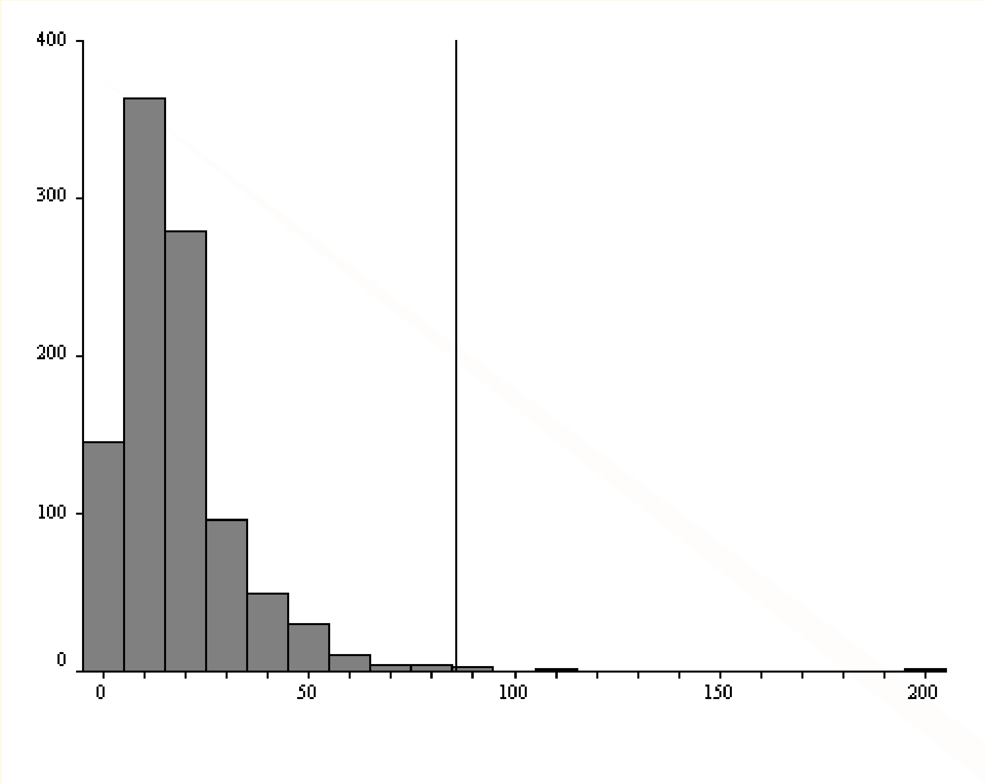
because such effects are inherently self-accelerating.

Data sets tend to be modelled as a mixture of two uninformative networks: the full network (or close-to-full networks) and the empty network (or close-to-empty networks).

This is undesirable – worthless modelling!



Distribution of transitive triplets under an actor-driven model with naïve closure modelling.



Same graph with emphasis on the 'empty end'.

The horizontal line indicates the observed number of transitive triplets.

Modelling transitive closure

...is all about balancing closure tendencies.

Presence of the 'precursor configurations' may *not under all circumstances* translate into higher probabilities for presence of the transitive triad.

Task, thus, is to *differentiate these circumstances*.

Here, different strategies have been followed...

ERG approaches

In the ERG approach, it is natural to think in terms of prevalence (counts) of local configurations as predictors of global structure.

- ▷ include prevalence of precursor configurations in the model (Snijders, 2002), or
- ▷ include prevalence of consequential configurations in the model: *(alternating) k-triplets* (Snijders, Pattison, Robins & Handcock, 2004).

Closure is being curbed by effects that act as “local brakes”.

Additionally, the whole model family can be nonlinearly adjusted as to not so easily translate opportunities for closure into high closure probabilities (Hunter & Handcock, 2005).

Actor-driven approach

Due to actor's local, neighbourhood-focused perspective, localisation here is built into the model already.

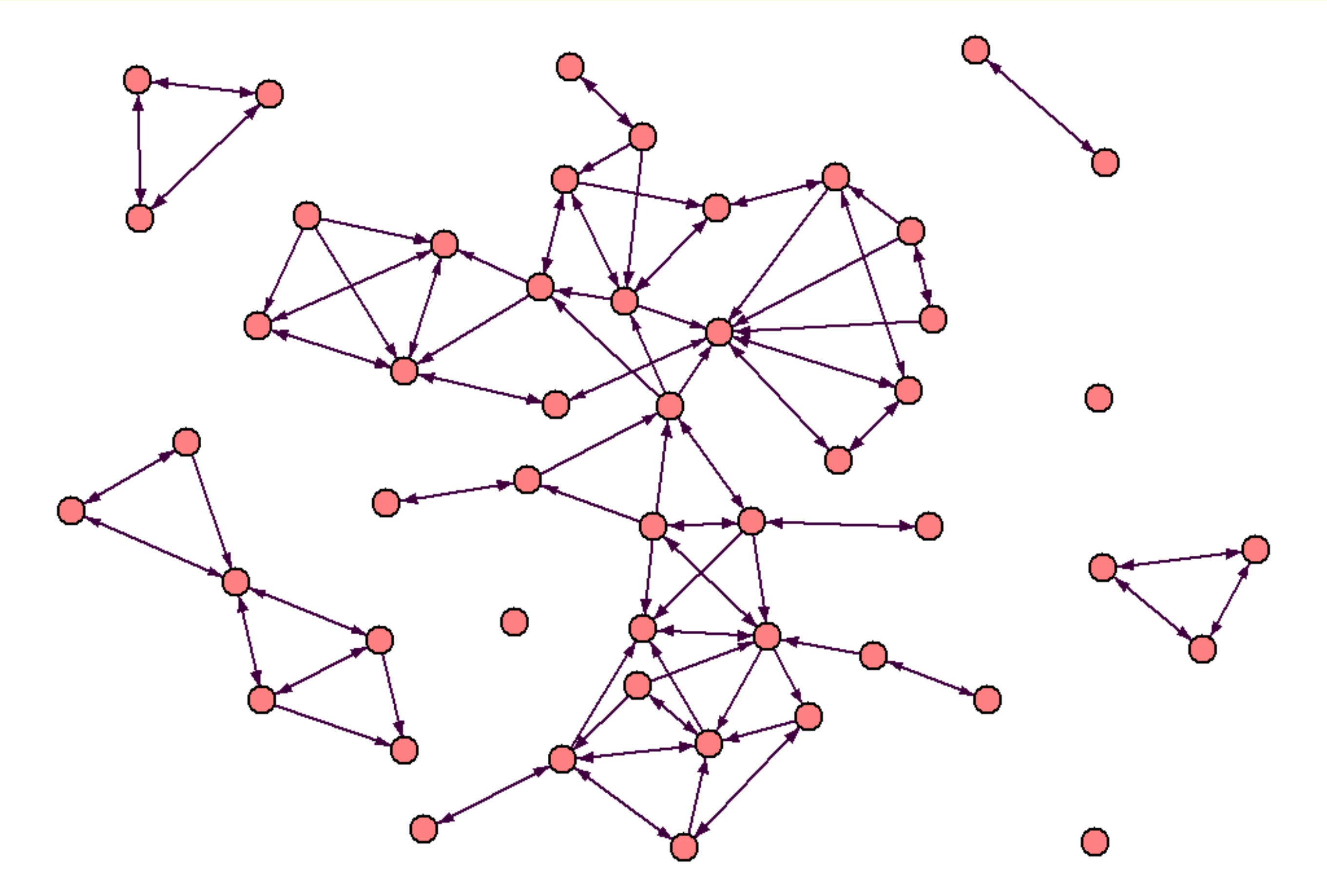
1. freeze network activity of transitively embedded actors,
2. decreasing marginal returns on transitive embeddedness,
3. extra incentives to destroy accumulated transitive embedding.

Here only strategy *2.* is discussed.

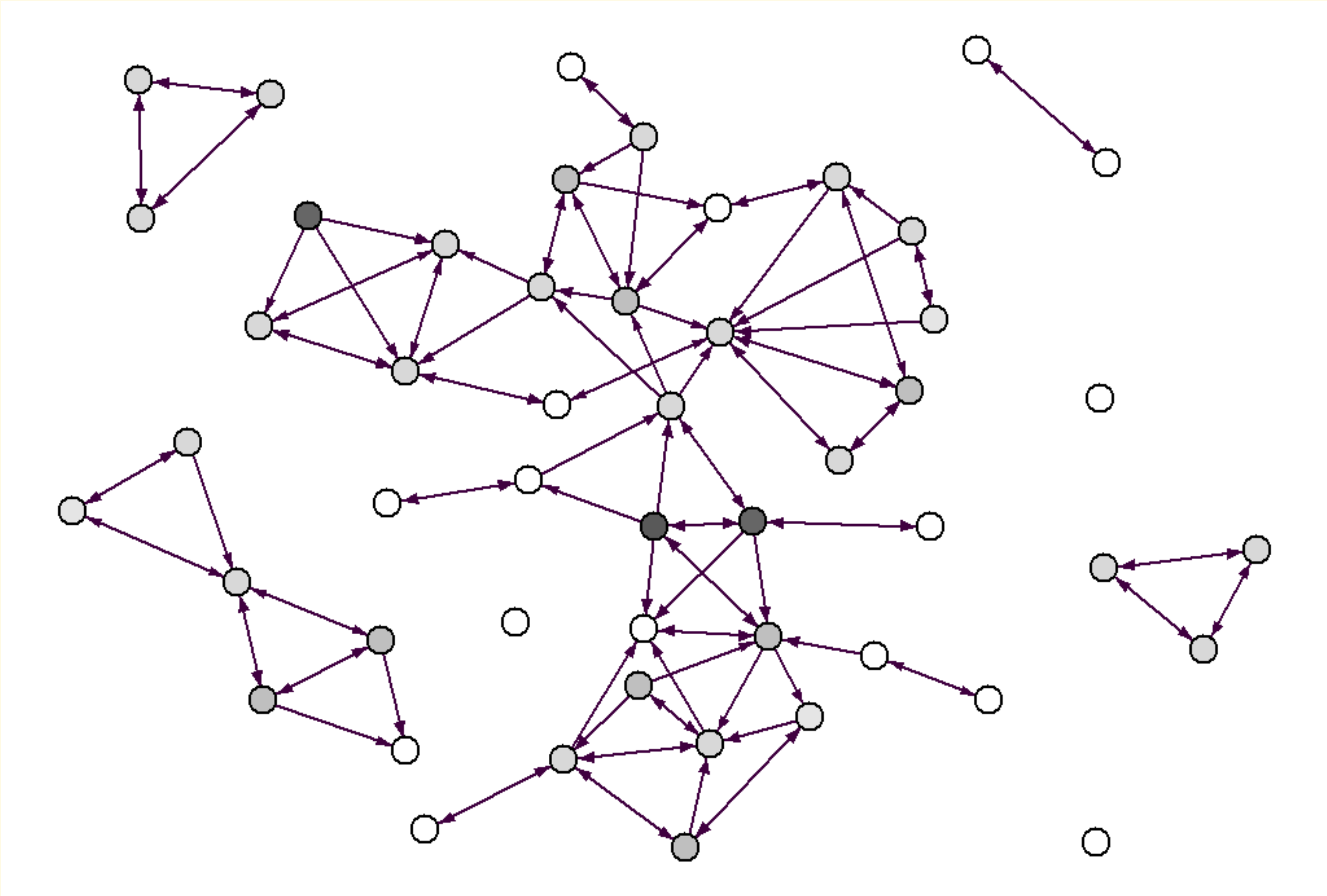
Sample data: standard SIENA **data set s50**

The 50 most active girls in the *Teenage Friends and Lifestyle Study* (Michell & Amos, 1997).

- ▷ 13 year old girls from one school cohort
- ▷ in the West of Scotland,
- ▷ network ties indicate friendship (dichotomous),
- ▷ here only first observation (1995).



First observation of the s50 network (Michell & Amos, 1997).



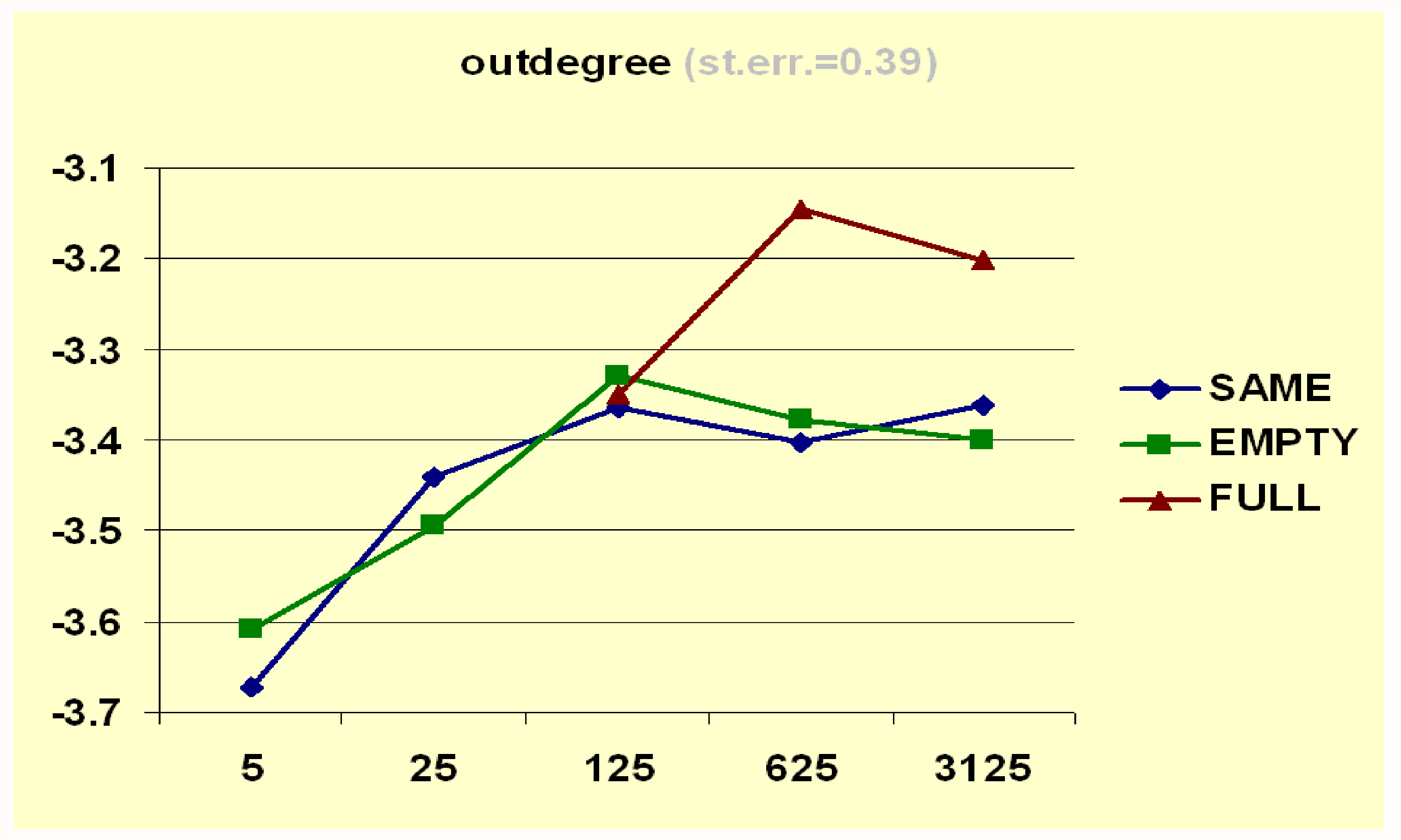
The s50 network coloured by actorwise transitive triplet count.

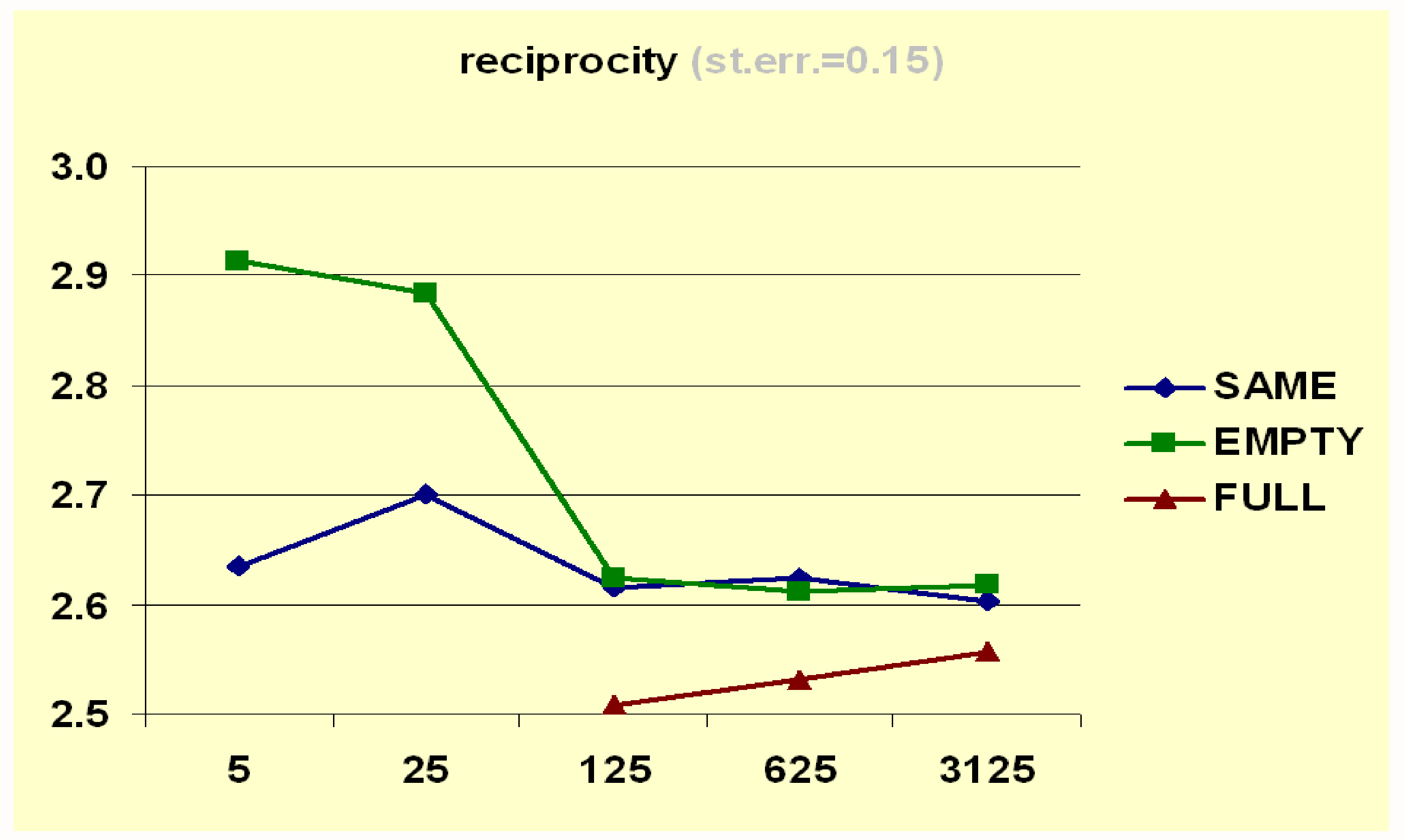
Estimating an actor-driven model with decreasing marginal returns on transitive embeddedness

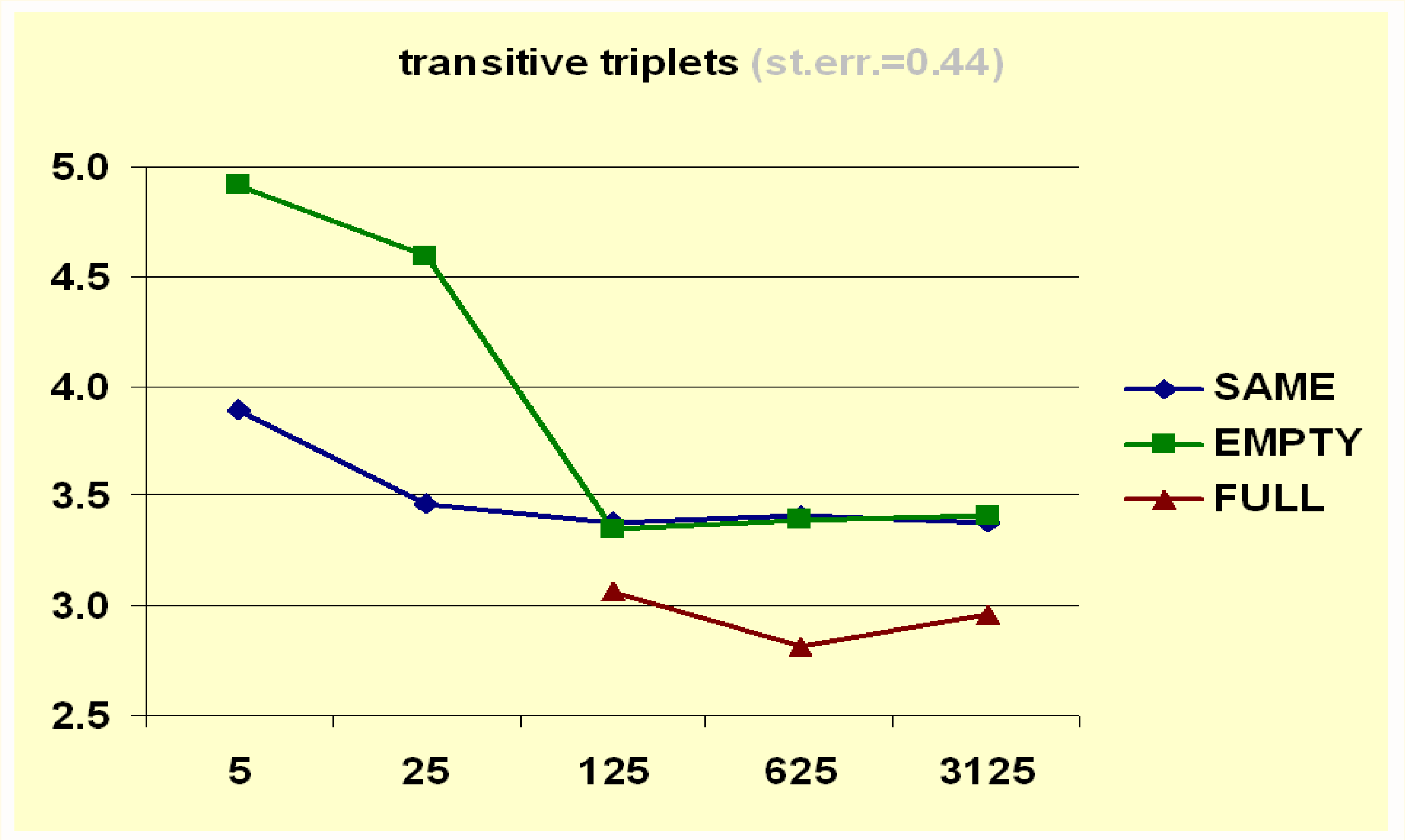
Standard model estimated with (fixed, high) rate parameter.

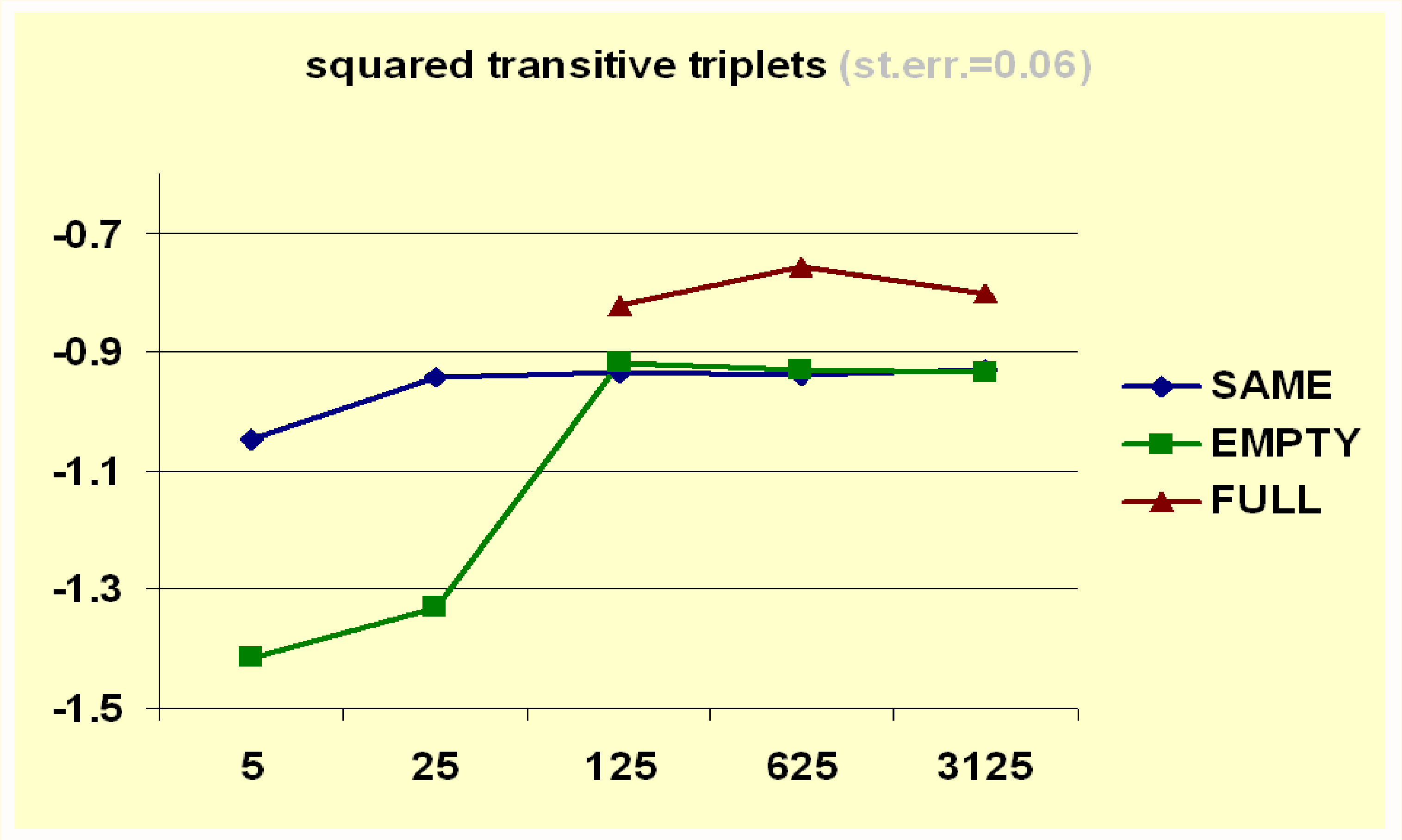
- outdegree: -3.36 (st.err. 0.39)
- reciprocity: 2.60 (st.err. 0.15)
- transitive triplets: 3.37 (st.err. 0.44)
- transitive triplets squared: -0.93 (st.err. 0.06)

The following pages show sample trajectories of estimates for increasing, fixed rate parameter. All simulation paths start from one of three initial networks: the full network, the empty network, and the s50 network itself.


outdegree (st.err.=0.39)
-3.1
-3.2
-3.3
-3.4
-3.5
-3.6
-3.7
5
25
125
625
3125
SAME
EMPTY
FULL


reciprocity (st.err.=0.15)
3.0
2.9
2.8
2.7
2.6
2.5
5
25
125
625
3125
SAME
EMPTY
FULL

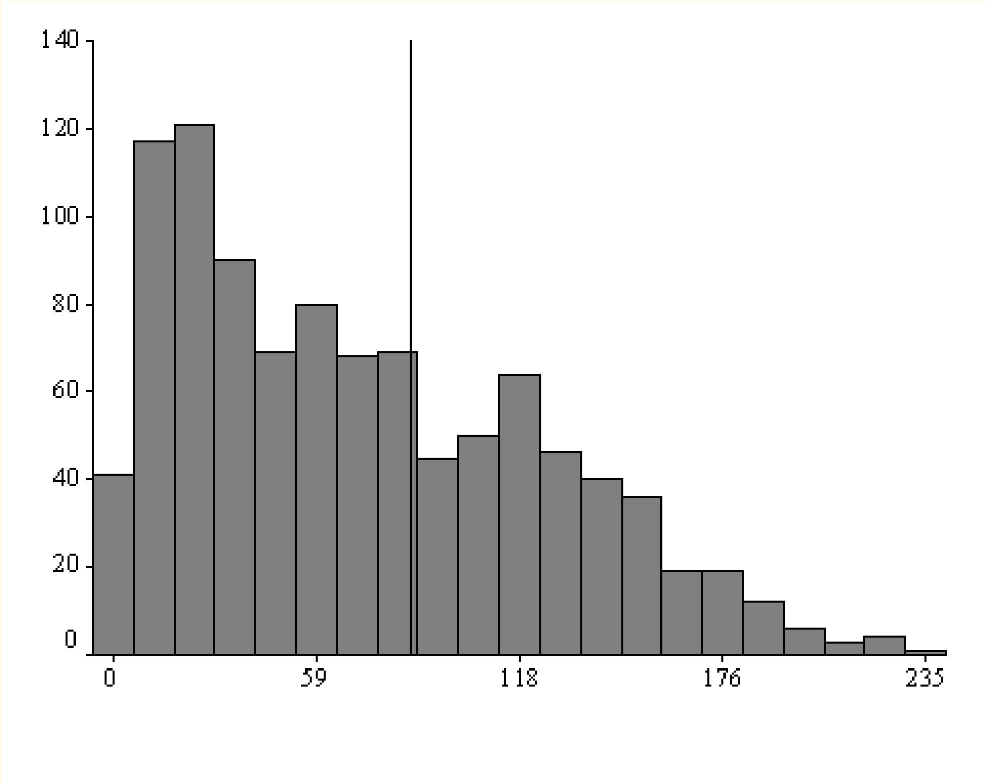

transitive triplets (st.err.=0.44)
5.0
4.5
4.0
3.5
3.0
2.5
5
25
125
625
3125
SAME
EMPTY
FULL


squared transitive triplets (st.err.=0.06)
-0.7
-0.9
-1.1
-1.3
-1.5
5
25
125
625
3125
SAME
EMPTY
FULL

Messages on model convergence

- ▷ convergence of model parameters can be diagnosed
- ▷ high rates are crucial (not yet satisfactory at rate=3125)
- ▷ keeping initial network the same is a bad strategy for actually estimating a model
- $\rightsquigarrow$ randomize initial networks per simulation run, ideally draw from limit distribution

What about model fit?



Distribution of transitive triplets under the squared triplets model. Still very skew, but at least no weight on the 'full end'.

FIT STATISTIC	OBS	SIMAV	STDEV	T-VAL
Number of ties	113	116.65	(64.34)	-0.0568
Number of reciprocated ties	78	75.57	(49.38)	0.0490
Number of transitive triplets	86	71.50	(51.99)	0.2787
Sum of actorwise squares of tr.trips	276	350.00	(320.75)	-0.2307
Number of directed distances equal to 2	123	280.41	(243.26)	-0.6471
Number of 3-cycles	21	23.35	(17.66)	-0.1333
Number of isolates (indegree up to 0)	4	15.36	(7.52)	-1.5092

Some more fit indicators, based on 1000 simulations.

Messages on model fit

- ▷ transitivity is not yet adequately modelled
- ▷ isolated nodes are overpredicted
- $\rightsquigarrow$ inclusion of distance-2 parameter as remedy?

4. Perspectives

Some topics for the future (in order of importance):

- investigation of fit (diverse distributions)
- get rid of longitudinal artefacts
- facilitate estimation
- comparison with ERG models on data sets

References

Bianconi, Ginestra, Matteo Marsili, and Fernando Vega-Redondo (2005). On the non-trivial dynamics of complex networks. *Physica A* **346**, 116 – 122.

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.

Hunter, David R., and Mark S. Handcock (2005). Inference in curved exponential family models for networks. To appear in *Journal of Graphical and Computational Statistics*.

Michell, Lynn, and Amanda Amos (1997). Girls, pecking order and smoking. *Social Science and Medicine* **44**, 1861 – 1869.

Snijders, Tom A.B. (1996). Stochastic Actor-Oriented Models for Network Change. *Journal of Mathematical Sociology* **21**, 149 – 172.

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

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock (2004). New specifications for exponential random graph models. To appear in *Sociological Methodology.*

Wasserman, Stanley, and Philippa E. Pattison (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika* **61**, 401 – 425.