New Specifications for Exponential Random Graph Models: Some empirical results

Garry Robins*
Tom Snijders**
Peng Wang*

Sunbelt XXV
February 2005

*University of Melbourne, Australia
**University of Groningen, the Netherlands

- *What counts as a "good" model for a social network.*
- *Exponential random graph (p*) models*
    - *Markov models and the new specifications*
- *Fitting models to small networks – examples from UCINET*
- *Goodness of fit*
- *Fitting models to Italian transport industry network*

What counts as a “good” statistical model for an observed social network?

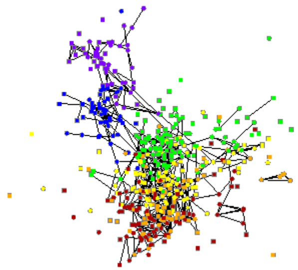
1. Models must be estimable from data.

2. Model parameters should imply model statistics consistent with those of the observed graph.

3. A good model will imply graphs with other features that are consistent with the observed graph.
   - *path lengths (geodesic distribution)*
   - *clustering (triangle formation)*
   - *degree distribution*
   - *denser regions (cohesive subsets of nodes)*

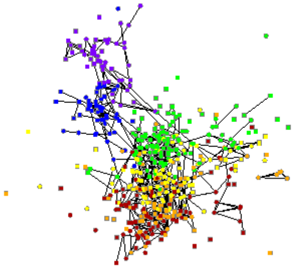
4. An excellent model will also successfully predict the presence or absence of network ties.

DATA



MODEL

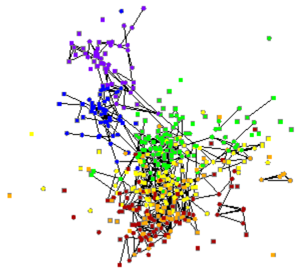
DATA



MODEL

MODEL COEFFICIENTS

DATA

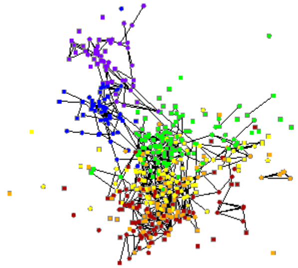


MODEL

MODEL COEFFICIENTS

SIMULATED DATA
(draws from the prob. dist.)

DATA



MODEL

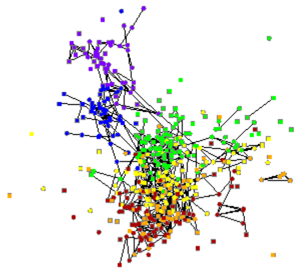
MODEL COEFFICIENTS

SIMULATED DATA
(draws from the prob. dist.)

GRAPH STATISTICS OF DATA

GRAPH STATISTICS OF SIMULATED DATA

DATA



MODEL

MODEL
COEFFICIENTS

SIMULATED DATA
(draws from the prob. dist.)

GRAPH STATISTICS
OF DATA

GRAPH STATISTICS
OF SIMULATED DATA

GOODNESS OF FIT
OF MODEL
TO DATA

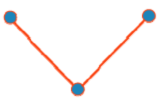
A Markov random graph model:

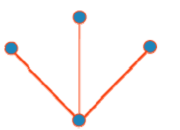
Nondirected networks

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = (1/\kappa) \exp\{\theta L + \sigma_2 S_2 + \sigma_3 S_3 + \tau T\}$$

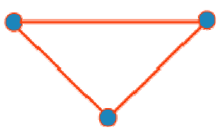
- *Edge parameter (θ)*
  - L … number of edges

- *Star parameters (σ_k)*
  - Propensities for individuals to have connections with multiple network partners



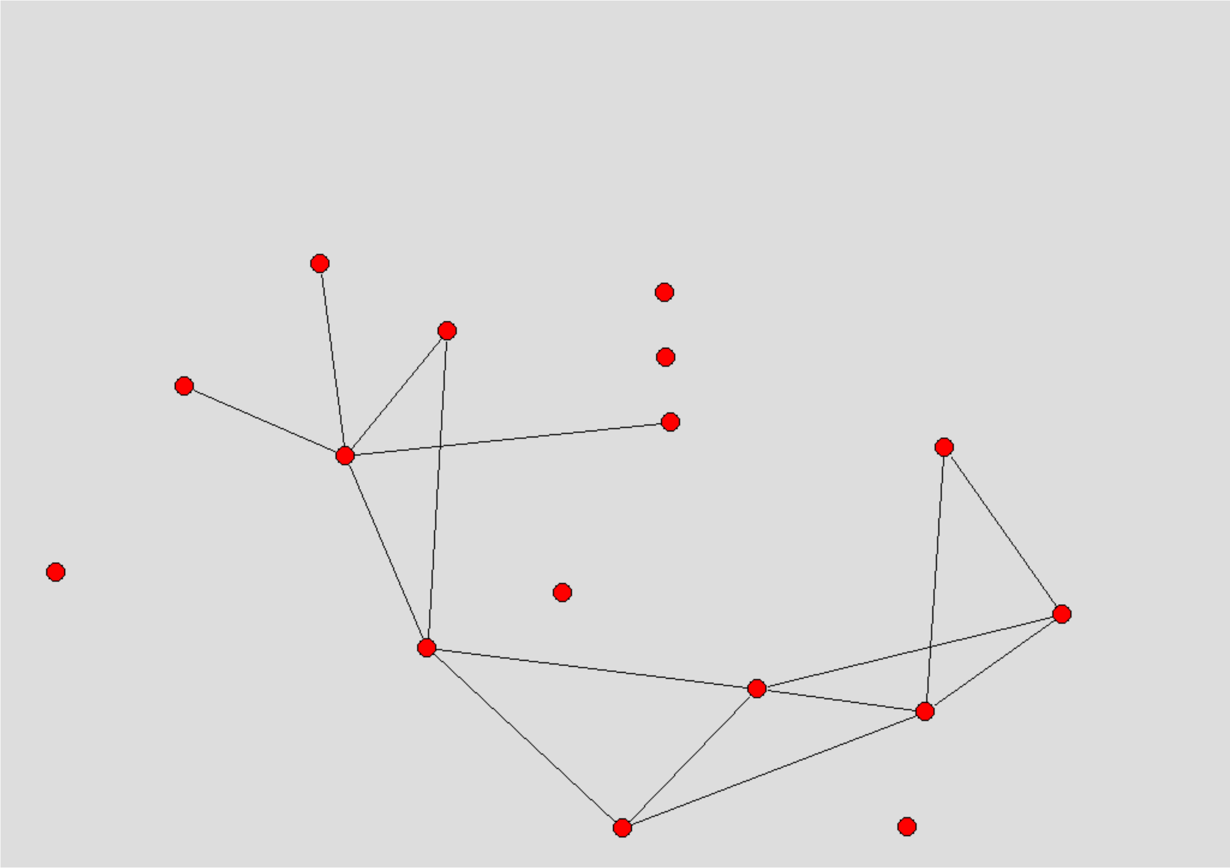


- *Triangle parameter (τ)*
  - represents clustering



If θ is the only nonzero parameter, this is a Bernoulli random graph model.

Example:
Florentine families business network
(Padgett & Ansell, 1993)



Markov model estimates:
Florentine families business network

Model containing edges, 2-stars, 3-stars, triangles

Monte Carlo Max. Likelihood estimates *(pnet)*

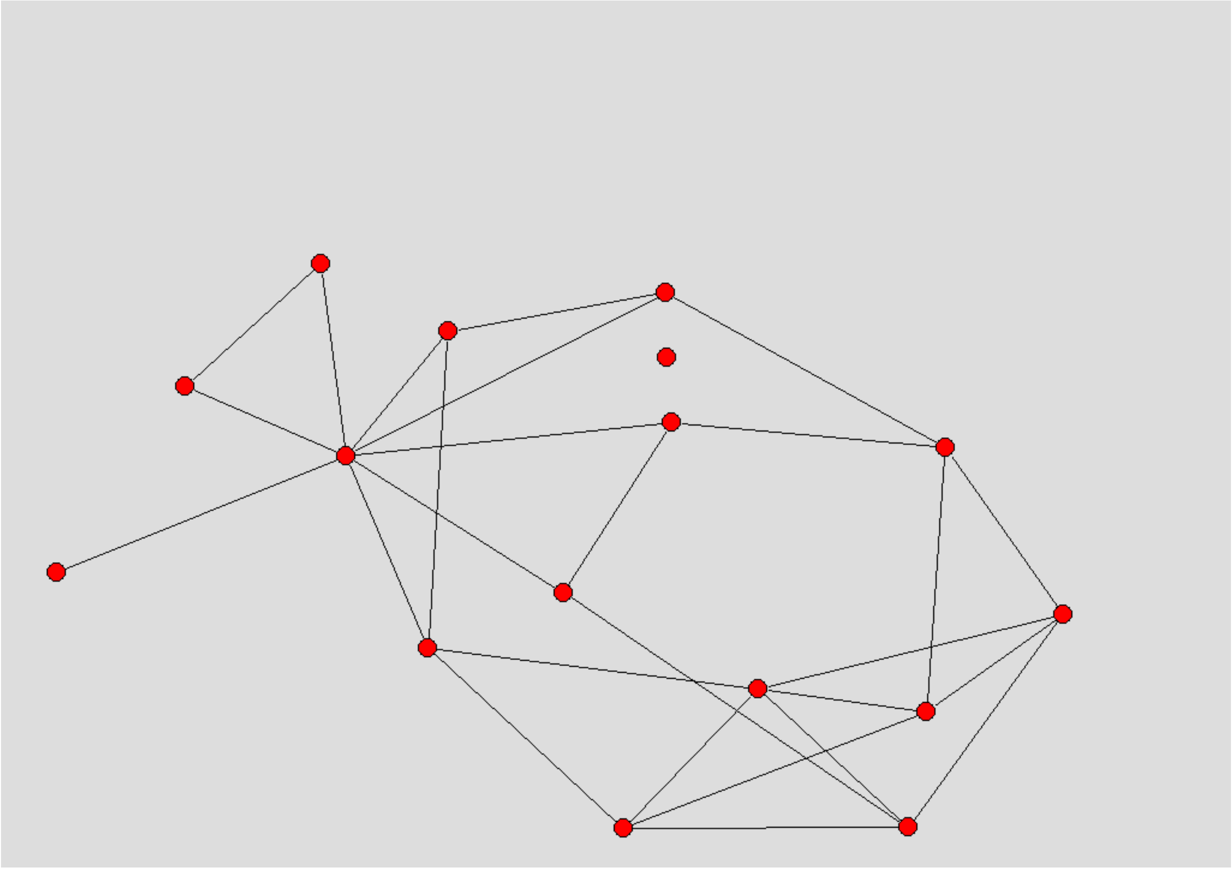
Edge = - 4.27 (1.13)*

2-star = 1.09 (0.65)

3-star = -0.67 (0.41)

Triangle= 1.32 (0.65)*

But the same model is degenerate for the combined business and marriage networks: there are no coherent parameter estimates



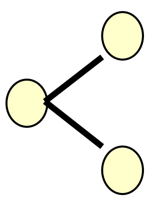
New specifications for exponential random graph models

Snijders, Pattison, Robins & Handcock (2005)

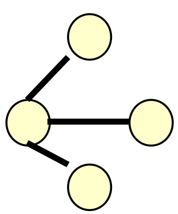
1. Models for degree sequences:

Alternating *k*-star models:

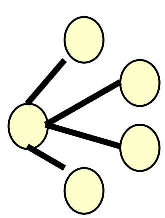
S_2



S_3



S_4



$$z(\mathbf{x}) = S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \ldots + (-1)^{n-2} \frac{S_{n-1}}{\lambda^{n-3}}$$

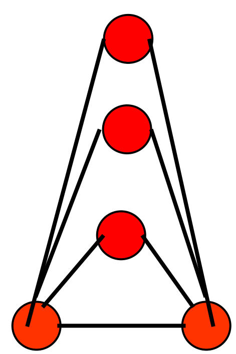
New specifications for exponential random graph models

Snijders, Pattison, Robins & Handcock (2005)

2. Models for cohesive subsets of nodes:

Alternating *k*-triangles:

3-triangle (T_1)
(T_3)



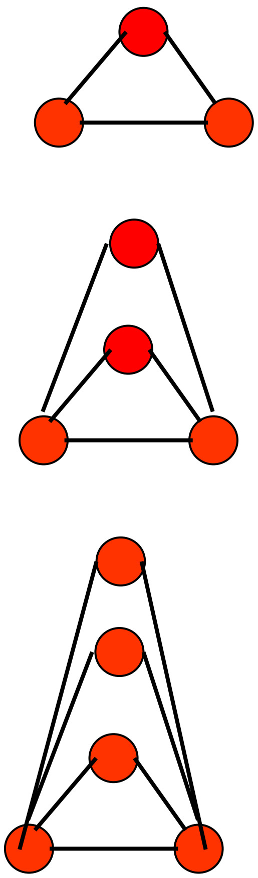
New specifications for exponential random graph models

Snijders, Pattison, Robins & Handcock (2005)

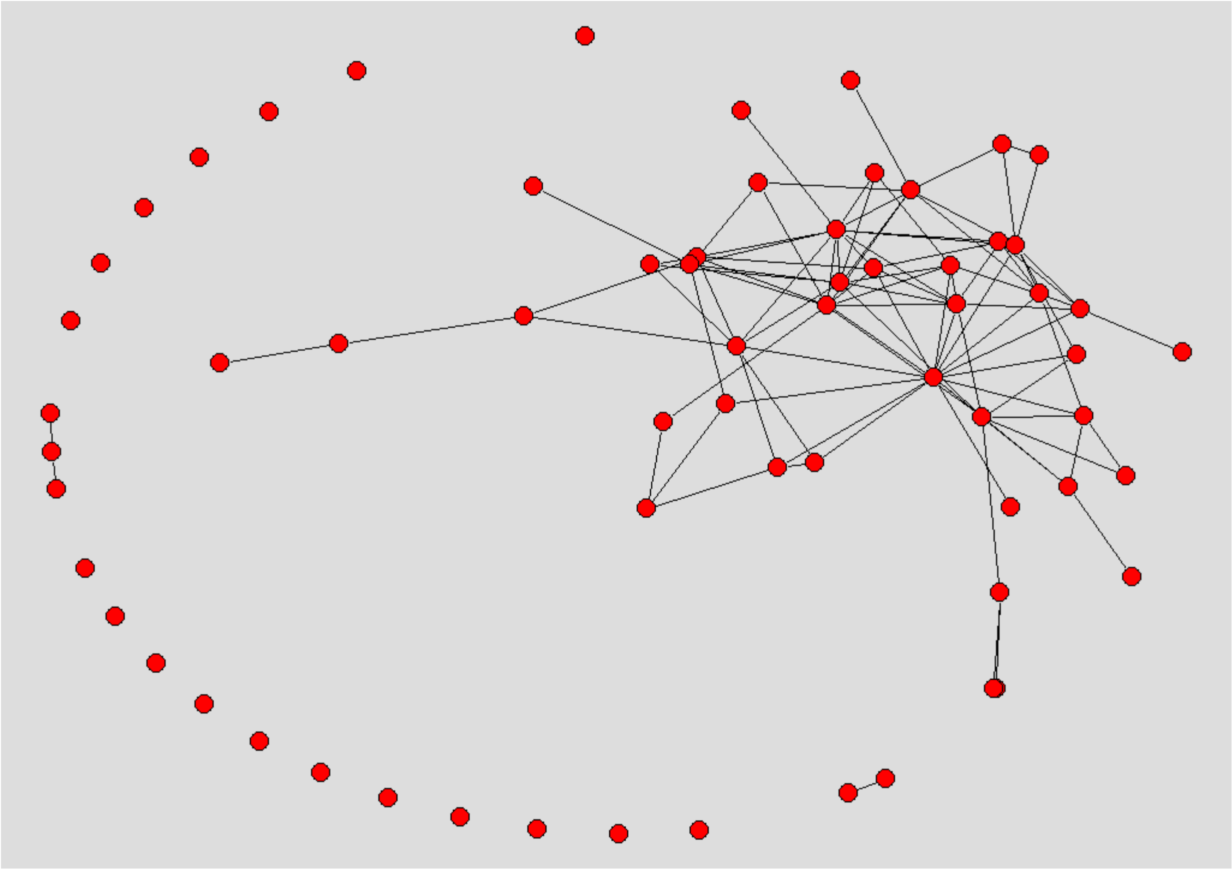
2. Models for cohesive subsets of nodes:

Alternating *k*-triangles:

$$u(\mathbf{x}) = T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} - \ldots + (-1)^{n-2}\frac{T_{n-2}}{\lambda^{n-3}}$$



etc…



Higher order model

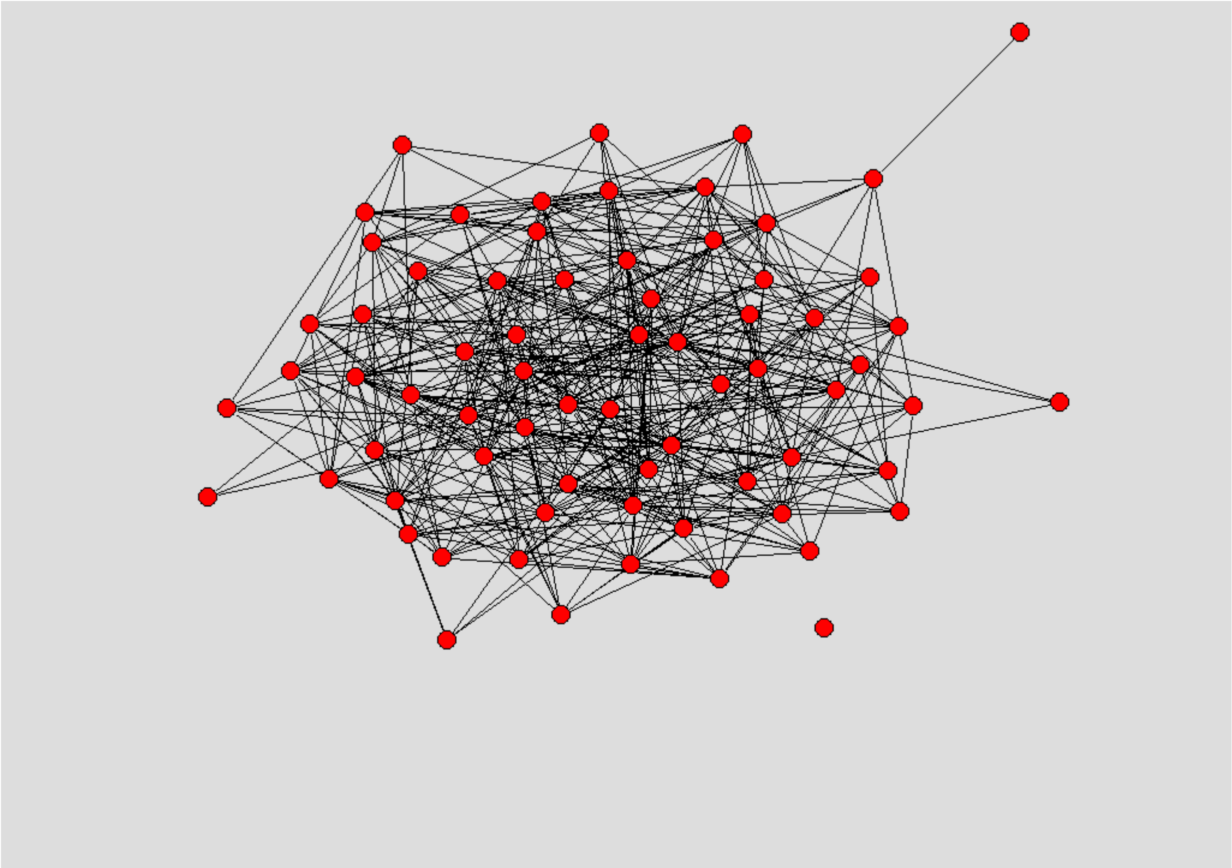
Positive *k*-triangle parameter:

Parameters:

Edge = -4.5

Alt. k-triangle=1.3

65 nodes



Higher order model

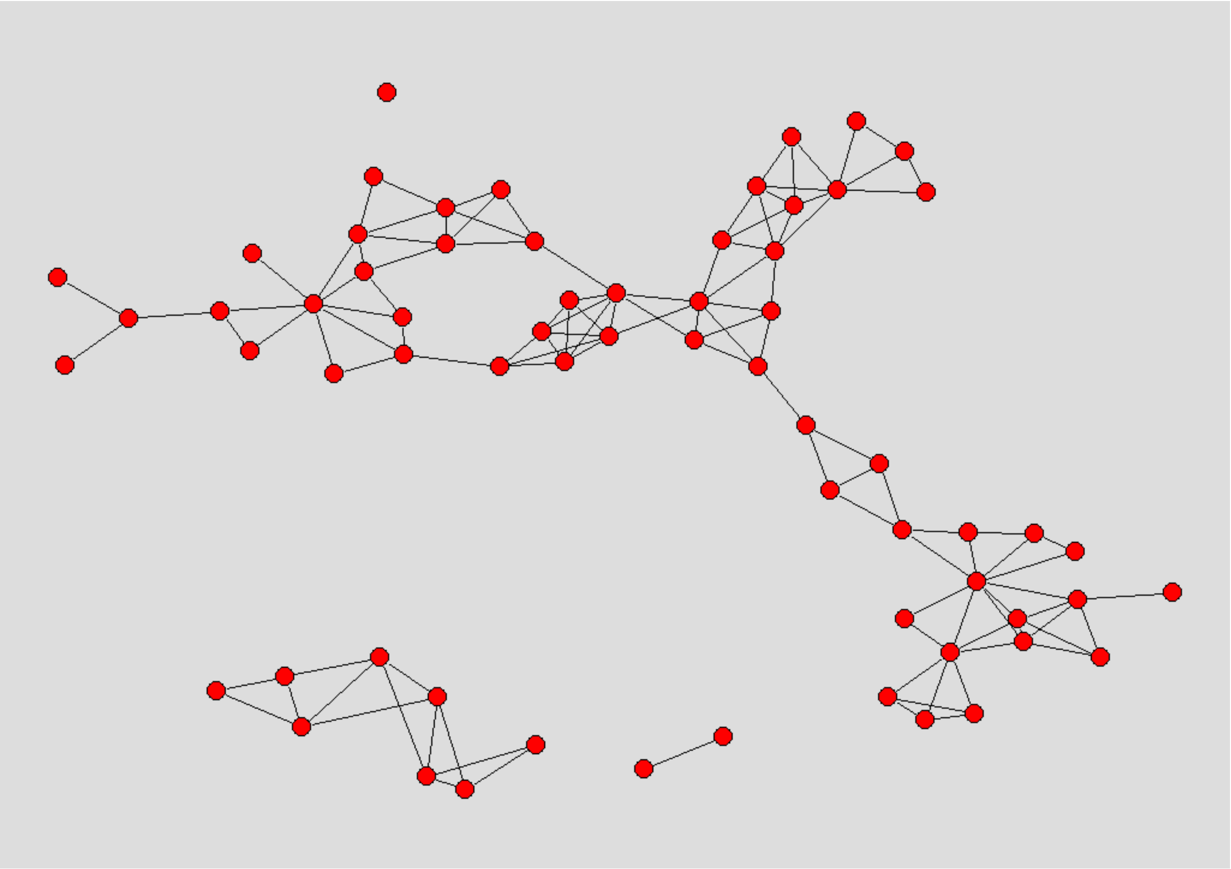
Positive *k*-triangle parameter:

Parameters:

Edge = -4.5

Alt. k-triangle=1.3

70 nodes



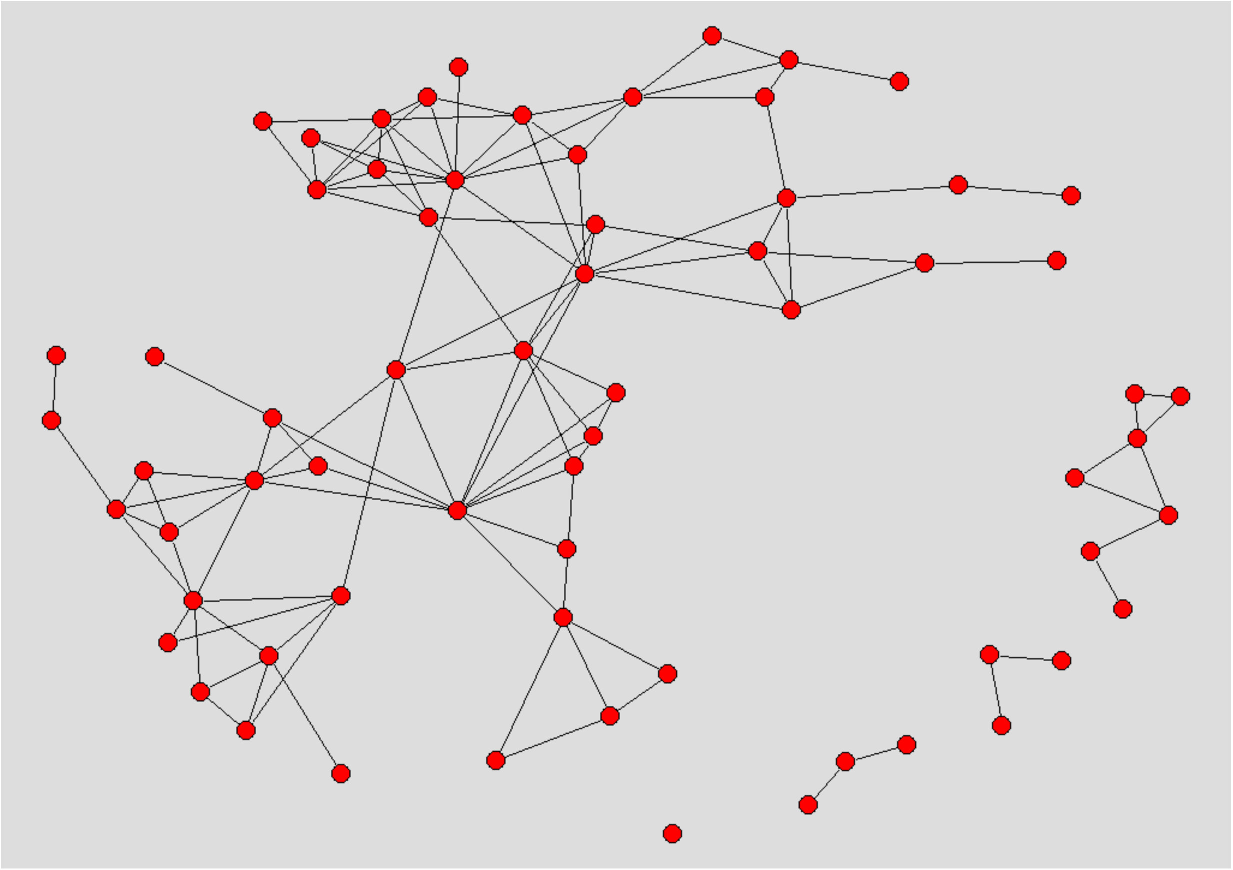
Higher order model

Positive *k*-triangle parameter & negative *k*-star parameter:

Parameters:

Edge = -0.5, alt. k-star = -1.5, alt. k-triangle = 2.0

65 nodes



Higher order model

Positive *k*-triangle parameter & negative *k*-star parameter:

Parameters:

Edge = -0.5, alt. k-star = -1.5, alt. k-triangle = 2.0

70 nodes

New specifications for exponential random graph models

Snijders, Pattison, Robins & Handcock (2005)

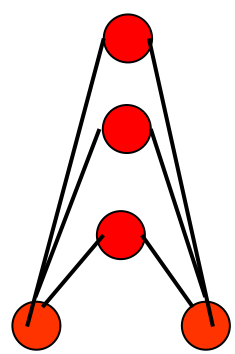
3. Models for multiple connectivity:

Alternating independent 2-paths:

3-indept.2-paths (U_3)

2-indept.2-paths (U_2)

2-path (U_1)



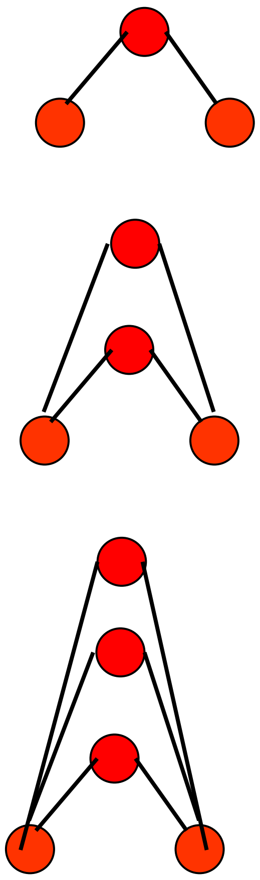
New specifications for exponential random graph models

Snijders, Pattison, Robins & Handcock (2005)

3. Models for cohesive subsets of nodes:

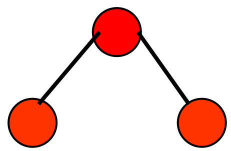
Alternating independent 2-paths:

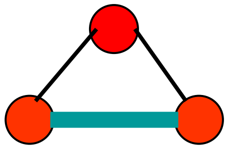
$$z(\mathbf{x}) = U_1 - \frac{2U_2}{\lambda} + \sum_{k=3}^{n-2} \left(\frac{-1}{\lambda} \right)^{k-1} U_k$$

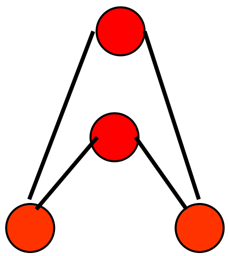


etc…

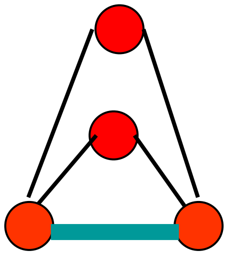
Why might we want Alternating independent 2-paths?

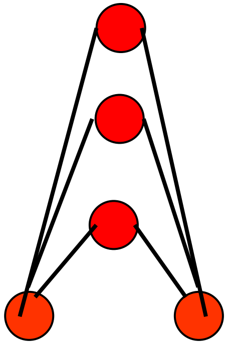




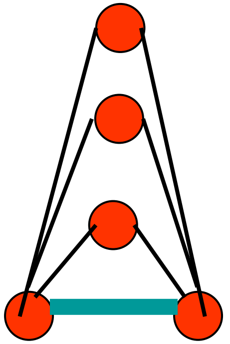


Indpt. 2-paths are lower order to *k*-triangles.





Helps assess whether a clustering (*k*-triangle) effect relates to the formation of the base of the k-triangle



Classes of exponential random graph models

(Single binary networks without node attributes)

	Non-directed networks	Directed networks
Dyadic independence		
Markov random graphs		
Higher order dependence		

Classes of exponential random graph models

(Single binary networks without node attributes)

	Non-directed networks	Directed networks
Dyadic independence	Edges *(Bernoulli graphs; simple random graphs)*	Edges; Mutuality *(Also p1 models)*
Markov random graphs		
Higher order dependence		

Classes of exponential random graph models

(Single binary networks without node attributes)

	Non-directed networks	Directed networks
Dyadic independence	Edges *(Bernoulli graphs; simple random graphs)*	Edges; Mutuality *(Also p1 models)*
Markov random graphs	Edges, stars, triangles	Edges, Mutuality, in-,out-,mixed-stars,transitive and cyclic triads
Higher order dependence		

Classes of exponential random graph models

(Single binary networks without node attributes)

	Non-directed networks	Directed networks
Dyadic independence	Edges *(Bernoulli graphs; simple random graphs)*	Edges; Mutuality *(Also p1 models)*
Markov random graphs	Edges, stars, triangles	Edges, Mutuality, in-,out-,mixed-stars,transitive and cyclic triads
Higher order dependence	*The above plus:* Alternating k-stars, k-triangles, alt. 2-paths	*The above plus:* Alternating k-stars, k-triangles, alt. 2-paths

Fitting Models:
20 network data sets from UCINET5

(Borgatti, Everett & Freeman, 1999)

Non directed networks:

Kapferer mine: kapfmm, kapfmu (16 nodes)
Kapferer tailor shop: kapfts1, kapfts2 (39 nodes)
Padgett Florentine families: padgb, padgm (16 nodes)
Read Highland tribes: gamapos (16 nodes)
Zachary karate club: Zache (34 nodes)
Bank wiring room: rdpos, rdgam (14 nodes)
Taro exchange: Taro (22 nodes)
Thurman office: Thurm (15 nodes)

Fitting Models: 20 network data sets from UCINET5

(Borgatti, Everett & Freeman, 1999)

Directed networks:

Kapferer tailor shop: kapfti1, kapfti2 (39 nodes)
Wolf primates: wolfk (20 nodes)
Krackhardt hi-tech managers: friend, advice (21 nodes)
Bank wiring room: rdhlp (14 nodes)
Knoke bureaucracies: knokm, knoki (10 nodes)

Non directed networks

Data set	Markov
Kapfmm	OK
Kapfmu	OK
Kapfts1	Does not converge
Kapfts2	Does not converge
Padgm	OK
Padgb	OK
Gamapos	Does not converge
Zache	Does not converge
Rdpos	OK
Rdgam	OK
Taro	Does not converge
Thurm	OK
TOTAL	7/12

Directed networks

Data set	Markov
Kapfti1	Does not converge
Kapfti2	Does not converge
Wolfk	Does not converge
Krackhardt friend	Does not converge
Krackhardt advice	Does not converge
Rdhlp	OK
Knokm	Does not converge
Knoki	OK
TOTAL (directed)	2/8
TOTAL (nondir.)	7/12
TOTAL	9/20

Non directed networks

Data set	Markov	Higher order
Kapfmm	OK	
Kapfmu	OK	
Kapfts1	Does not converge	
Kapfts2	Does not converge	
Padgm	OK	
Padgb	OK	
Gamapos	Does not converge	
Zache	Does not converge	
Rdpos	OK	
Rdgam	OK	
Taro	Does not converge	
Thurm	OK	
TOTAL	7/12	

Non directed networks

Data set	Markov	Higher order
Kapfmm	OK	OK
Kapfmu	OK	OK
Kapfts1	Does not converge	OK
Kapfts2	Does not converge	OK
Padgm	OK	OK
Padgb	OK	OK
Gamapos	Does not converge	OK
Zache	Does not converge	OK
Rdpos	OK	OK
Rdgam	OK	OK
Taro	Does not converge	OK
Thurm	OK	OK
TOTAL	7/12	12/12

Directed networks

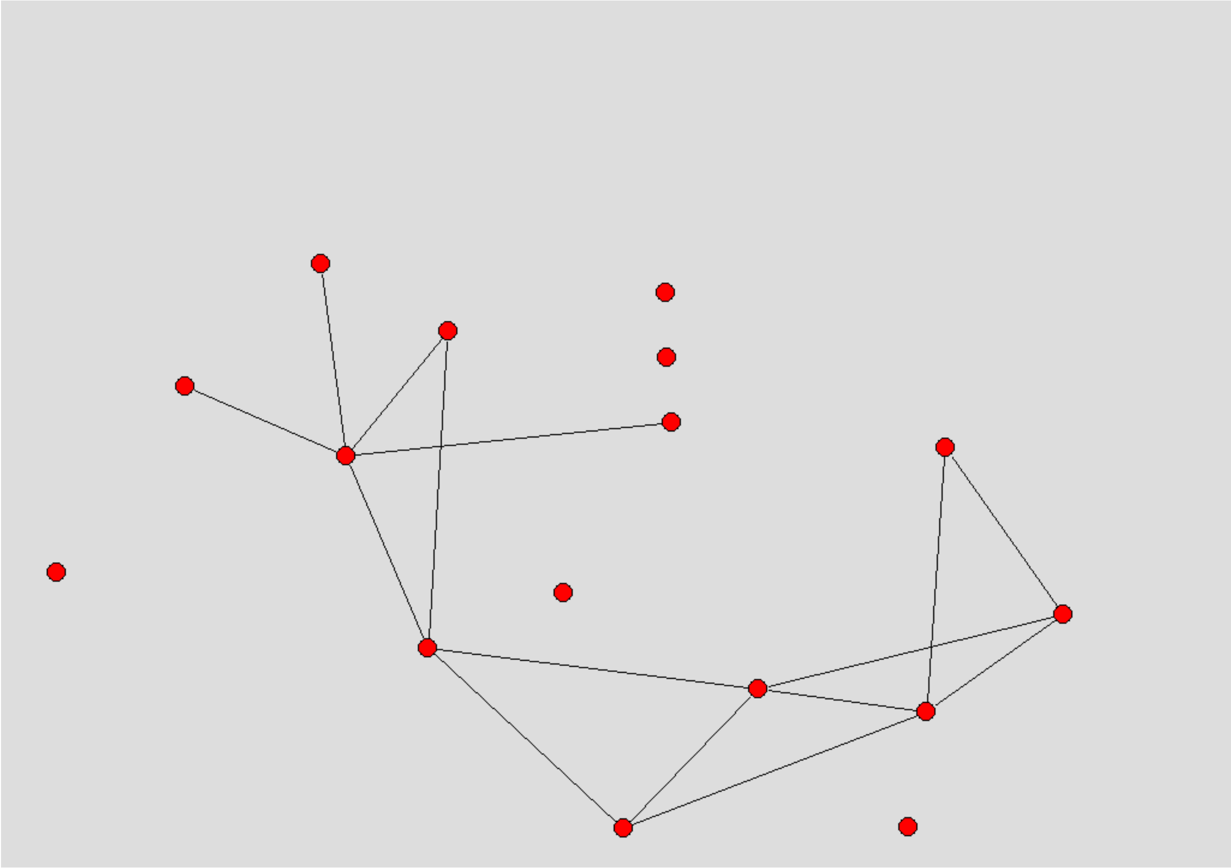
Data set	Markov	Higher order
Kapfti1	Does not converge	
Kapfti2	Does not converge	
Wolfk	Does not converge	
Krackhardt friend	Does not converge	
Krackhardt advice	Does not converge	
Rdhlp	OK	
Knokm	Does not converge	
Knoki	OK	
TOTAL (directed)	2/8	
TOTAL (nondir.)	7/12	
TOTAL	9/20	

Directed networks

Data set	Markov	Higher order
Kapfti1	Does not converge	OK
Kapfti2	Does not converge	OK
Wolfk	Does not converge	OK
Krackhardt friend	Does not converge	OK
Krackhardt advice	Does not converge	OK
Rdhlp	OK	OK
Knokm	Does not converge	OK
Knoki	OK	OK
TOTAL (directed)	2/8	
TOTAL (nondir.)	7/12	
TOTAL	9/20	20/20

Goodness of fit:
Florentine families business network

(Padgett & Ansell, 1993)



Goodness of fit: Florentine families business network

For **Bernoulli graph** model: *t*-statistics

Model statistics
Edges: t = 0.15

Other graph statistics
2-stars: t = 0.95
3-stars: t = 1.04
triangles: t = 3.16
k-stars: t = 0.83
k-triangles: t = 3.06

Degree distribution
std dev degree dist: t = 2.29
skew deg dist: t = -0.05

Clustering
global clustering: t = 2.92
local clustering: t = 3.52

Goodness of fit:
Florentine families business network

For **Bernoulli graph** model: *t*-statistics

Model statistics
Edges: t = 0.15

Other graph statistics
2-stars: t = 0.95
3-stars: t = 1.04
triangles: t = 3.16
k-stars: t = 0.83
k-triangles: t = 3.06

Degree distribution
std dev degree dist: t = 2.29
skew deg dist: t = -0.05

Clustering
global clustering: t = 2.92
local clustering: t = 3.52

Markov model estimates:
Florentine families business network

Model containing edges, 2-stars, 3-stars, triangles

Maximum Likelihood (pnet) estimates

Edge = - 4.27 (1.13)*

2-star = 1.09 (0.65)

3-star = -0.67 (0.41)

Triangle= 1.32 (0.65)*

Goodness of fit:
Florentine families business network

For Markov random graph model: *t*-statistics

Model statistics
Edges: t = 0.03
2-stars: t = 0.03
3-stars: t = 0.03
triangles: t = 0.00

Other graph statistics
k-stars: t = 0.02
k-triangles: t = 0.23

Degree distribution
std dev degree dist: t = 0.54
skew deg dist: t = 0.08

Clustering
global clustering: t = 0.24
local clustering: t = 0.93

Markov model estimates: Florentine families business network

Model containing edges, k-stars, k-triangles

Pnet estimates

Edge = - 2.75 (0.85)

k-star = -0.06 (0.49)

K-triangle= 0.86 (0.47)

Goodness of fit:
Florentine families business network

For kstar/ktriangle model: *t*-statistics

Markov statistics
Edges: t = -0.05
2-stars: t = -0.13
3-stars: t = -0.29
triangles: t = -0.13

Higher order statistics
k-stars: t = -0.03
k-triangles: t = -0.01

Degree distribution
std dev degree dist: t = 0.39
skew deg dist: t = -0.28

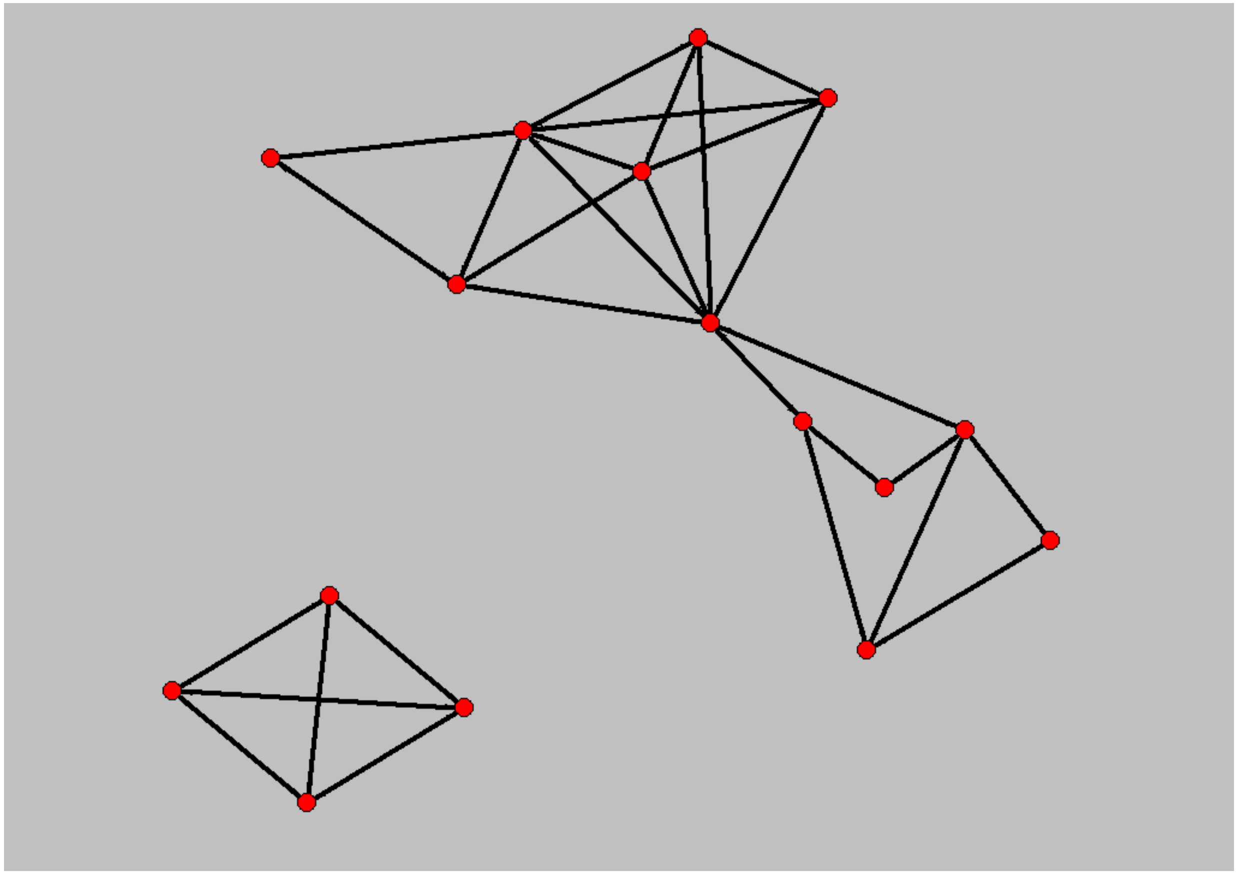
Clustering
global clustering: t = 0.32
local clustering: t = 0.66

Goodness of fit:
Read Highland tribes network

<u>The effect of alternating two paths</u>

<u>gamapos</u>

Read Highland tribes alliances



Goodness of fit:
Read Highland tribes alliances

Comparing to a sample of **Bernoulli graphs**: *t*-statistics

Model statistics

Edges: t = 0.052

Other graph statistics

2-stars: t = -0.17

3-stars: t = -0.24

triangles: t = 2.28

k-stars: t = -0.08

k-triangles: t = 1.59

Indpt.2-paths: t=-0.68

Degree distribution

std dev degree dist: t = -0.63

skew deg dist: t = 1.66

Clustering

global clustering: t = 4.92

local clustering: t = 4.92

Geodesic distribution:

Observed graph has extremely long 3rd quartile (>99% of sample) and 4th quartile geodesics (>100% of sample)

K-star/k-triangle model estimates: Read Highland tribes alliances

<u>pnet estimates</u>

Edge = 7.893 (4.544)

K-star = -3.472 (1.251)

K-triangle= 1.473 (0.316)

K-star/k-triangle/Alt 2-paths estimates: Read Highland tribes alliances

pnet estimates

Edge = 5.600 (3.824)

K-star = -1.801 (1.129)

K-triangle = 0.927 (0.353)

Alt.2p = -0.561 (0.278)

Goodness of fit:
Read Highland tribes alliances

Compared to kstar/ktriangle model: *t*-statistics

Markov statistics
Edges: t = 0.21
2-stars: t = 0.30
3-stars: t = -0.39
triangles: t = 0.99

Higher order statistics
k-stars: t = 0.22
k-triangles: t = 0.23
Alt-2p: t = -0.30

Degree distribution
std dev degree dist: t = 0.71
skew deg dist: t = 0.91

Clustering
global clustering: t = 0.98
local clustering: t = 0.46

Geodesic distribution:
Observed geodesic quartiles are not extreme compared to sample graphs.

Goodness of fit:
Read Highland tribes alliances

Compared to kstar/ktriangle/alt2p model: *t*-statistics

Markov statistics

Edges: t = 0.05

2-stars: t = 0.16

3-stars: t = -0.39

triangles: t = 0.23

Higher order statistics

k-stars: t = 0.05

k-triangles: t = 0.03

Alt-2p: t = 0.04

Degree distribution

std dev degree dist: t = 1.05

skew deg dist: t = 2.03

Clustering

global clustering: t = 0.30

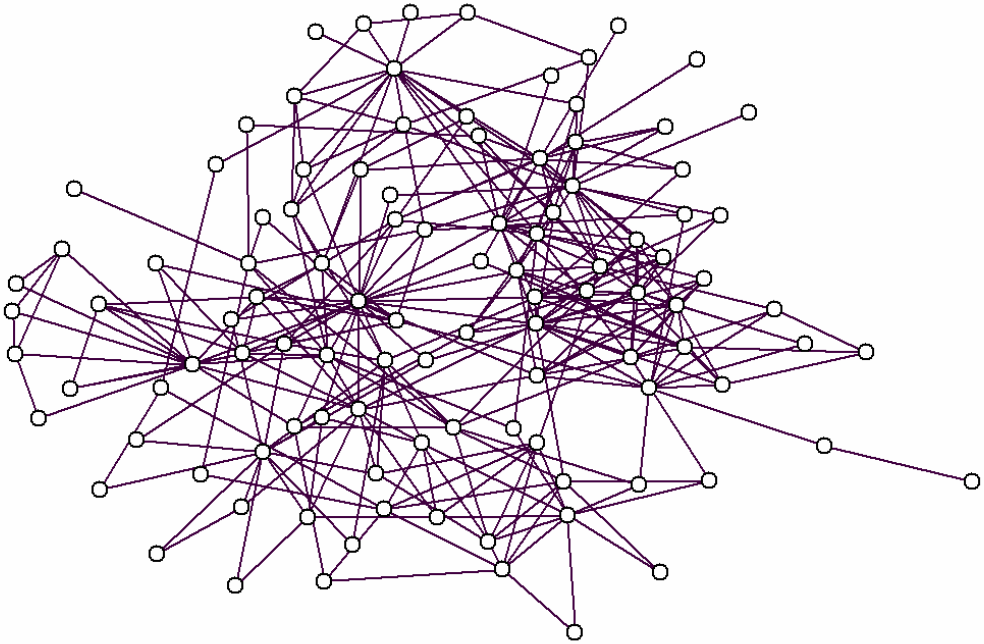
local clustering: t = 0.53

Geodesic distribution:

Observed geodesic quartiles are not extreme compared to sample graphs.

Italian transport industry network (106 nodes)

(Lomi & Pattison, 2004)



Italian transport industry network

Parameter	Estimate	Standard error	Convergence statistic
Edge	– 2.591	0.687	– 0.01
Alt. kstars	– 0.477	0.217	– 0.01
Alt. ktriangles	0.960	0.081	– 0.01
Alt. 2-paths	0.014	0.018	– 0.02

Italian transport industry network

Parameter	Estimate	Standard error	Convergence statistic
Edge	**– 2.591**	**0.687**	– 0.01
Alt. kstars	**– 0.477**	**0.217**	– 0.01
Alt. ktriangles	**0.960**	**0.081**	– 0.01
Alt. 2-paths	0.014	0.018	– 0.02

Goodness of fit: *t* statistics

Graph counts	
Edges	– 0.01
2-stars	0.12
3-stars	0.14
Triangles	0.36
Alt kstars	– 0.01
Alt ktriangles	– 0.01
Alt 2-paths	– 0.01
Degree distribution	
Standard deviation	0.63
skew	0.35
Clustering	
Global	**1.59**
Local	**1.77**

Conclusions

"…the science of fitting statistical models to complex network data is still in its infancy."

(Hunter & Handcock, 2004).

"The most promising class of statistical models for expressing structural properties of social networks is the class of Exponential Random Graph models …The new specifications increase the range and applicability of these models as a tool for the statistical analysis of social networks."

(Snijders, Pattison, Robins & Handcock, 2005)

Draft paper:

Robins, Snijders & Wang (2005). *Recent developments in exponential random graph (p*) models for social networks.*

at http://www.psych.unimelb.edu.au/staff/robins.html