

# Specification and estimation of exponential random graph models for social (and other) networks

Tom A.B. Snijders

University of Oxford

March 24, 2009

Modeling social networks

## Modeling social networks

This presentation treats the  $p_1$  and  $p_2$  models  
and the  $p^*$  = ERGM = Exponential Random Graph Model  
for single observations of social networks;  
the social network here is the *dependent variable*.

¿How to explain an observed network?

## Notation

Set of  $n$  actors, with a dichotomous = binary relation, represented as a *directed graph (digraph)*.

Tie variable from  $i$  to  $j$  indicated by  $Y_{ij}$  :

$$Y_{ij} = \begin{cases} 1 & \text{if there is a tie} \\ 0 & \text{if there is no tie.} \end{cases}$$

(Diagonal values  $Y_{ii}$  meaningless, formally defined as  $Y_{ii} = 0$ .)

$Y_{ij}$  is the indicator of the *arc* or *directed line* from  $i$  to  $j$ .

Matrix  $Y$  is *adjacency matrix* of digraph.

*Same approach is possible for non-directed graphs.*

Existence of ties can be explained on the basis of

1. Explanatory variables = covariates;  
can be function of individual actors (actor-based covariate)  
or of directed or undirected pairs of actors  
(dyad-based covariates).
2. Patterns of further ties in the network;  
this reflexivity / endogenous feedback is the big difficulty  
in modeling social networks.  
*Complicated dependence structure between ties.*

Some interesting kinds of dependence between ties:

- ★ *Reciprocity* :  
dependence between  $Y_{ij}$  and  $Y_{ji}$ .  
The pair  $(Y_{ij}, Y_{ji})$  is called a *dyad*.
- ★ The dependence within each *row*:  
outgoing relations of the same actor.
- ★ The dependence within each *column*:  
incoming relations of the same actor.

- ★ *Transitivity*:  
"a friend of my friend is also my friend",  
⇒ dependence between triples of actors.
- ★ *Popularity*:  
choices lead to more choices.
- ★ *Balance*:  
preference for others who make the same choices  
as the actor him/herself  
(similar to transitivity, not the same).

The difficulty with stochastic models for social networks is that they have to represent *dependence*, and cannot be built – like most stochastic models – on broad *independence* assumptions.

*Some models for social networks:*

1. Logistic regression
2.  $p_1$  model
3.  $p_2$  model
4. ERGM (exponential random graph model) =  $p^*$  model
5. MRQAP  
(multiple regression quadratic assignment procedure)

## 1. Logistic regression

$Y_{ij}$  is a binary (0–1) dependent variable, so the most straightforward model would be *logistic regression*:

$$\text{logit}(\mathbb{P}\{Y_{ij} = 1\}) = \gamma_0 + \gamma_1 W_{1ij} + \dots + \gamma_p W_{pij}$$

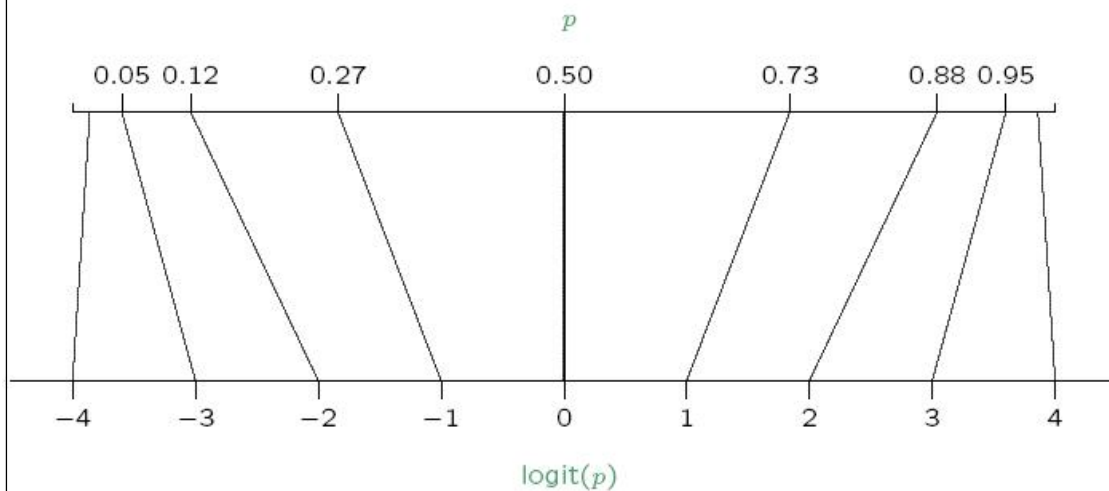
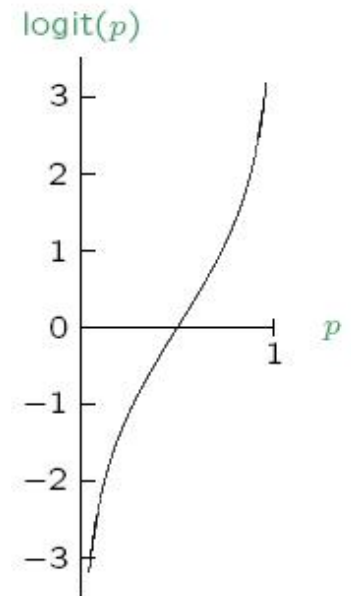
where  $W_1, \dots, W_p$  are explanatory variables, and where

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right).$$

The *logarithm* transforms a multiplicative to an additive scale and transforms the set of positive real numbers to the whole real line. A widely used transformation of probabilities is the *log odds*, defined by

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right),$$

where  $\ln(x)$  is the natural logarithm of  $x$ . The logit function, shown here, is defined for numbers between 0 and 1, and goes from minus infinity to plus infinity. E.g.,  $p = 0.269$  is transformed to  $\text{logit}(p) = -1$  and  $p = 0.982$  to  $\text{logit}(p) = 4$ . The logit of  $p = 0.5$  is exactly 0.



Correspondence between  $p$  and  $\text{logit}(p)$ .

The logistic regression model is a model where  $\text{logit}(p)$  is a linear function of the explanatory variables. In spite of the attractive properties of the logit function, it is by no means the only suitable function for transforming probabilities to arbitrary real values.

The general term for such a transformation function is the *link function*, as it links the probabilities (or more generally, the expected values of the dependent variable) to the explanatory variables.

The probit function (which is the inverse cumulative distribution function of the standard normal distribution) also is often used as a link function for dichotomous variables.

A generalized linear model for a dichotomous outcome with the probit link function is called a probit regression model.

For still other link functions see, e.g., the textbooks Long (1997) or McCullagh and Nelder (1989).

Statistical theory tells us the following about such a procedure, which neglects the dependence structure in the model for the data:

1. The parameter estimates are *consistent*, i.e., for large  $n$  they are reasonable;
2. however, they are not *efficient*, i.e., their precision is not optimal;
3. the standard errors are totally unreliable.

## Example: Friendship between Lazega's lawyers

Example based on Lazega's research on a New England law firm (Lazega, 2001; Lazega and Pattison, 1999).

*Friendship* relation between the 36 partners.

Actor covariates:

- ▶ seniority (rank number of entry in the firm)
- ▶ gender
- ▶ office (three different cities)
- ▶ years in the firm
- ▶ age

- ▶ practice (litigation or corporate law)
- ▶ law school attended  
(Ivy League, non-I.L. in the region, other).

For categorical covariates (office and law school), similarity represented by within-dyad identity:

$$I\{x_i = x_j\} = \begin{cases} 1 & \text{if } x_i = x_j, \\ 0 & \text{otherwise.} \end{cases}$$

Density = 0.21, average degree = 7.4.

In-degrees vary from 2 to 16, out-degrees from 0 to 21.

For the actor covariates  $W_i$ ,

we distinguish between the following effects on  $Y_{ij}$  :

1. the ego or out-degree effect, leading to correlation between the covariate  $W_i$  and the out-degrees  $Y_{i+}$
2. the alter or in-degree effect, leading to correlation between the covariate  $W_j$  and the in-degrees  $Y_{+j}$
3. similarity or dissimilarity effects, leading to correlation between the absolute difference  $|W_i - W_j|$  and the tie indicators  $Y_{ij}$  .

An inadequate but simple model can be fitted  
by applying logistic regression, assuming independent tie variables.

Parameter	est.	s.e.
Constant term	-1.750	0.097
Same office	2.129	0.183
Seniority alter	0.003	0.008
Seniority ego	0.023	0.008
Seniority dissimilarity	-0.077	0.011
Practice (corp. law) alter	0.184	0.156
Practice (corp. law) ego	0.565	0.157
Same practice	0.396	0.155

Logistic regression estimates of covariate effects  
for Lazega's friendship data (partners).

A next step could be to include row ('*sender*') and column ('*receiver*') effects in the model:

$$\text{logit}(\mathbb{P}\{Y_{ij} = 1\}) = \alpha_i + \beta_j + \gamma_1 W_{1ij} + \dots + \gamma_p W_{pij}.$$

Then the variables  $W_1, \dots, W_p$  all must be dyadic covariates, because actor-dependent covariates, depending only on  $i$  or  $j$ , would be collinear with the sender and receiver effects.

The sender effects  $\alpha_i$  represent the *activity / outgoingness* of actor  $i$ , while the receiver effects  $\beta_j$  represent the *popularity / attractiveness* of actor  $j$ .

## 2. $p_1$ model

Holland and Leinhardt (1981) proposed the  $p_1$  model, which does not include covariate effects, but which does include *sender and receiver effects* as well as *reciprocity*.

The model includes the following parameters:

- ▶  $\mu$ , parameter for the density ( $\mu$  higher  $\Rightarrow$  more ties)
- ▶  $\alpha_i$ , activity parameter for actor  $i$   
( $\alpha_i$  higher  $\Rightarrow i$  has more outgoing ties)  
restriction  $\sum_i \alpha_i = 0$
- ▶  $\beta_i$ , popularity parameter for actor  $i$   
( $\beta_i$  higher  $\Rightarrow i$  has more incoming ties)  
restriction  $\sum_i \beta_i = 0$
- ▶  $\rho$ , reciprocity parameter  
( $\rho$  higher  $\Rightarrow Y_{ij}$  and  $Y_{ji}$  tend to be more alike)

This represents actor differences and reciprocity, not network structure or covariate effects.

The probability distribution for each dyad  $(Y_{ij}, Y_{ji})$  is defined by

$$P\{(Y_{ij}, Y_{ji}) = (a, b)\} = k_{ij} \exp(a(\mu + \alpha_i + \beta_j) + b(\mu + \alpha_j + \beta_i) + ab\rho) ;$$

this formula holds if  $a$  and  $b$  are 0 or 1 (possible outcomes of  $Y_{ij}$  and  $Y_{ji}$ );  
 $k_{ij}$  is a number (not depending on  $(a, b)$ ) ensuring that the four probabilities (for outcomes  $(0,0)$ ,  $(0,1)$ ,  $(1,0)$ ,  $(1,1)$ ) sum to 1.

Different dyads are assumed to be statistically independent.

*Note 1:*

This form of the probability distribution, consisting of an exponential function of a linear function of the parameters, is well-known in mathematical statistics:

a so-called *exponential family of distributions*.

*Exercise:*

Using the properties of the exponential function and the definition of statistical independence, prove that the relations  $Y_{ij}$  and  $Y_{ji}$  are independent if  $\rho = 0$ .

**Note 2:**

The  $p_1$  model can be regarded as a bivariate logistic regression model for the dyads, where the different dyads have different parameters depending on the actors  $i$  and  $j$  involved in the dyad.

To see the correspondence with the usual logistic regression model, write out the logit of the probability; recall that  $\text{logit}(p) = \ln(p/(1 - p))$ ; the result is

$$\begin{aligned} \text{logit}(P(Y_{ij} = 1 \mid Y_{ji} = b)) &= \ln \left( \frac{P\{Y_{ij} = 1, Y_{ji} = b\}}{P\{Y_{ij} = 0, Y_{ji} = b\}} \right) \\ &= \mu + \alpha_i + \beta_j + b\rho. \end{aligned}$$

This makes clear that

$\alpha_i$  is a main effect of the "sender" ( $i$ ) of the tie,

$\beta_j$  is a main effect of the "receiver" ( $j$ ) of the tie,

and  $\rho$  is the effect of the other relation ( $Y_{ji}$ ) within the dyad.

For a single dyad, this would be a rather foolish model:  
four outcomes and six parameters...;  
but we have  $n(n - 1)/2$  dyads to estimate the parameters from.

Note, however,  
that there are two parameters associated with every actor.  
The total number of parameters is  $2 + 2n$ ;  
since there are two equality constraints ( $\sum_i \alpha_i = \sum_i \beta_i = 0$ ),  
the number of independent parameters is  $2n$ .

The  $p_1$  model can be estimated using  
methods for estimating loglinear models.  
Estimation techniques and elaborations of the  $p_1$  model  
were developed by Wasserman and co-workers.  
(See Wasserman and Faust, 1994.)

In this model, *dyads* ( $Y_{ij}, Y_{ji}$ ) are independent.  
This is a severe limitation.

However, this model was an important first step  
for the statistical modeling of social networks.

Actor	Alpha	Beta	Actor	Alpha	Beta	Rho	Theta
1	-0.365	-0.515	19	-1.155	-0.544	3.574	-2.851
2	-1.202	1.022	20	1.094	-0.555		
3	-	0.291	21	-0.610	1.319		
4	1.182	0.496	22	0.537	-0.456		
5	-1.000	0.058	23	-2.739	0.965		
6	-	-0.575	24	1.707	0.024		
7	-0.916	-1.109	25	0.696	0.279		
8	-2.420	0.426	26	-0.800	2.049		
9	-1.125	1.856	27	0.152	1.264		
10	1.802	-1.773	28	0.640	-0.285		
11	-1.099	1.346	29	0.495	0.012		
12	2.774	-1.240	30	-0.001	-0.703		
13	0.178	0.923	31	2.116	-1.373		
14	0.428	-0.650	32	-1.000	0.058		
15	-0.375	-1.346	33	1.489	-2.439		
16	-0.198	0.589	34	-1.344	0.665		
17	1.843	0.561	35	-0.358	0.033		
18	0.311	-0.879	36	-0.735	0.207		

### 3. $p_2$ model

It is not very attractive that differences between actors are represented in the statistical model by *parameters*, two parameters for each actor.

E.g., this precludes the possibility of testing effects of actor-dependent explanatory variables, because the differences between actors are already represented completely by the parameters  $\alpha_i$  and  $\beta_i$ .

This is similar to the choice between representing group differences by fixed or by random effects in multilevel analysis.

## $p_2$ model

To overcome this problem, van Duijn (1995) proposed the  $p_2$  model, (also see van Duijn and Lazega, 1997; van Duijn, Snijders, and Zijlstra, 2004) which can be regarded as a *random effects* version, or also as a *multilevel* version, of the  $p_1$  model.

In this model, the sender and receiver effects  $\alpha_i$  and  $\beta_j$  are regarded as *latent (i.e., unobserved) random variables*; and these latent variables can be explained by explanatory actor-dependent variables. Dyad-dependent explanatory variables can also be included.

This does not represent network structure.

Since it is usual to denote random variables by capital instead of greek letters, we now write  $U_i$  and  $V_j$ , respectively, as the *unexplained* parts of the sender and receiver effects of actor  $i$

– unexplained, that is, given the explanatory variables  $W^{(1)}, \dots, W^{(p)}$ , where  $W^{(h)}$  has values  $W_{ij}^{(h)}$ , which could depend on  $i$  (the sender) only or on  $j$  (the receiver) only, but also on  $i$  and  $j$  simultaneously.

The vectors  $(U_1, \dots, U_n)$  and  $(V_1, \dots, V_n)$  are denoted  $U$  and  $V$ , respectively.

The probability distribution for each dyad  $(Y_{ij}, Y_{ji})$  is defined under the  $p_2$  model by

$$P\{(Y_{ij}, Y_{ji}) = (a, b) \mid U, V\} = k_{ij} \exp \left( a \left( \sum_h \gamma_h W_{ij}^{(h)} + \mu + U_i + V_j \right) + b \left( \sum_h \gamma_h W_{ji}^{(h)} + \mu + U_j + V_i \right) + ab\rho \right)$$

where  $a$  and  $b$  again are 0 or 1 (possible outcomes of  $Y_{ij}$  and  $Y_{ji}$ ).

The numbers  $\gamma_h$  are statistical parameters which are completely analogous to coefficients in logistic regression.

For the random variables  $U_i$  and  $V_i$  it is assumed that they are independent for different  $i$  and normally distributed; however,  $U_i$  and  $V_i$ , which variables refer to the same actor, can be correlated.

The statistical parameters of this model are

1. the parameters  $\mu$  and  $\rho$ ,
2. the regression coefficients  $\gamma_h$ ,
3. and the variances  $\sigma_U^2$  and  $\sigma_V^2$ ,
4. and the correlation  $\rho_{UV}$  of the actor effects.

Note that  $\rho_{UV}$  has nothing to do with  $\rho$ , although the same letter is used.

*Remark:*

The complete specification of the  $p_2$  model also includes explanatory variables associated with the reciprocity effect  $\rho$ , i.e., interactions between explanatory variables and reciprocity. This is omitted here.

Parameter estimation methods for the  $p_2$  model have been developed, using 'MCMC' (Markov chain Monte Carlo) estimation methods.

The  $p_1$  and  $p_2$  models represent structural network effects only to a very limited extent: e.g., no transitivity!

$\rho_1$  and  $\rho_2$  models

Random effects:

	parameter estimate	standard error
sender variance:	1.1405	0.2356
receiver variance:	0.5615	0.1375
sender receiver covariance:	-0.3464	0.1375

Fixed effects:

Overall effects:

	parameter estimate	standard error
Density:	-3.5695	0.7524
Reciprocity:	1.9784	0.2687

$\rho_1$  and  $\rho_2$  models

Overall covariate effects:

Overall effects of covariates

including diff and absdiff manipulations.

Covariate	Wald test		
	statistic	df	P
Seniority	30.7112	3	0.0000
practice	6.1933	3	0.1026
Same Office	73.8774	1	0.0000

Specific covariate effects:

Sender covariates:

	parameter	standard
Covariate	estimate	error
Seniority	0.0342	0.0196
practice	0.6004	0.4003

Receiver covariates:

	parameter	standard
Covariate	estimate	error
Seniority	-0.0091	0.0152
practice	-0.0062	0.3067

Density covariates:

	parameter	standard
Covariate	estimate	error
abs_diff_Seniority	-0.0523	0.0099
abs_diff_practice	-0.2573	0.1327
Same Office	1.5320	0.1782

## 4. Exponential random graph models

Frank (1991) and Wasserman and Pattison (1996) proposed the  $p^*$  model for social networks, generalizing the Markov graph distribution of Frank and Strauss (1986), also called the *Exponential Random Graph Model*, *ERGM*.

The probability distribution for the *ERGM* can be defined by

$$P\{Y = y\} = \frac{\exp(\sum_k \theta_k s_k(y))}{\kappa(\theta)}$$

where the  $s_k(y)$  are statistics of the digraph,  
 $\theta$  is a vector of statistical parameters,  
 and  $\kappa(\theta)$  is a normalization factor,  
 ensuring that the probabilities sum to 1.

Note that this is again an exponential family of distributions.

This formula is extremely general,  
because  $s(y)$  could be anything.

In this general form, the formula could be used  
to represent any probability distribution for a directed graph.

In practice,  $s(y)$  will contain effects of covariate effects  
as well as structural graph effects representing  
the dependence of the existence of ties within the network.

The mostly used statistics  $s(y)$  depend on  
the *density* or (equivalently) the total number of ties,

$$Y_{++} = \sum_{i,j} Y_{ij} ,$$

the *number of mutual dyads*,

$$R = \sum_{i < j} Y_{ij} Y_{ji} ,$$



mutual = reciprocated dyad

various statistics expressing *transitivity*,  
and explanatory variables.

The structures representing transitivity are based on *triad* or *triplet* counts, which indicate the number of times that a given triad occurs as a subgraph in the entire graph.

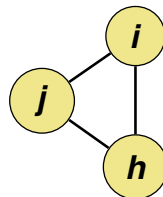
*Triads* are directed graphs with three nodes; *triplets* are ‘incomplete triads’, where the presence of only a subset of the tie variables is counted; e.g., the 3-cycle defined by  $y_{12} = y_{23} = y_{31} = 1$ ,  $y_{21} = y_{32} = y_{13} = 0$ .

A triad count is the number of times that a given triad occurs among the  $\binom{n}{3}$  triads generated by all subsets of three nodes within the total digraph  $y$  (where the order of the three nodes is not taken into account).

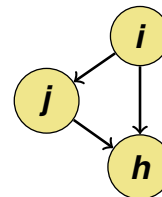
A *triplet* is an incomplete triad: it defines the values of some, but not all, of the six tie variables  $y_{12}$ ,  $y_{21}$ ,  $y_{13}$ ,  $y_{31}$ ,  $y_{23}$ ,  $y_{32}$ .

The triplet count is defined in analogous fashion.

Examples are the transitive triplet defined by  $y_{12} = y_{23} = y_{13} = 1$  and the intransitive triplet defined by  $y_{12} = y_{23} = 1$ ,  $y_{13} = 0$ .



transitive triad (nondirected)



transitive triplet (directed)

The definition of transitivity is:

*if  $i$  is tied to  $j$  and  $j$  is tied to  $h$ , then  $i$  is tied to  $h$ .*

Thus, the configuration  $Y_{ij} = Y_{jh} = 1$ , called a *two-path*, is a condition which must be fulfilled in order for transitivity to be a non-vacuous property.

If one wishes to test for transitivity, then it is relevant to control for this condition.

This can be done by including among the  $s_k(y)$  the number of two-stars (graphs) or number of two-paths (directed graphs).

The number of two-paths can be expressed as

$$\sum_{\substack{i,j,h=1 \\ i \neq h}}^g Y_{ij} Y_{jh} = \sum_{j=1}^n \sum_{\substack{i,h=1 \\ i \neq h}}^g Y_{ij} Y_{jh}$$

$$= \sum_{j=1}^n \left( Y_{+j} Y_{j+} - \sum_{i=1}^n Y_{ij} Y_{ji} \right) = \left( \sum_{j=1}^n Y_{+j} Y_{j+} \right) - 2R,$$

where  $R$  is the number of mutual dyads.

This shows that the number of two-paths is a function of the degrees and the number of mutual dyads.

Thus, when testing for transitivity using the number of transitive triplets as a test statistic, it is meaningful either to control for the number of two-paths, or for the degrees and the number of mutual dyads.

If the elements of the vector  $s(y)$  are the following:

1. total number of ties  $Y_{++}$
2. number of reciprocated ties  $R$
3. out-degrees  $Y_{i+}$
4. in-degrees  $Y_{+i}$ ,

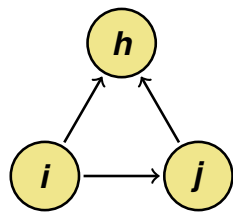
then the *ERGM* is just the  $p_1$  model.

(This requires some equality restrictions between the parameters, because there are linear equalities among the statistics.)

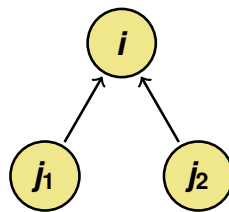
A more fruitful set of statistics is the following, which is a particular case of a *Markov graph* :

1. total number of ties  $Y_{++}$
2. number of reciprocated ties  $R$
3. number of out-2-stars  $\sum_{i,j,h} Y_{ij} Y_{ih}$
4. number of in-2-stars  $\sum_{i,j,h} Y_{ji} Y_{hi}$
5. number of 2-paths  $\sum_{i,j,h} Y_{ij} Y_{jh}$
6. number of transitive triplets  $\sum_{i,j,h} Y_{ij} Y_{jh} Y_{ih}$ .

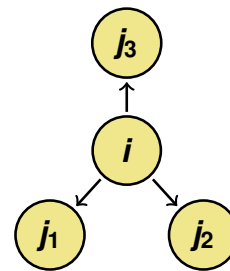
An explanation follows:



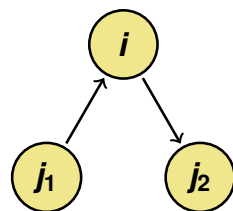
transitive  
triplet



in-two-star



out-three-star



two-path = mixed two-star

The number of out-2-stars is

$$\begin{aligned} \frac{1}{2} \sum_{\substack{i,j,h=1 \\ i \neq h}}^g Y_{ij} Y_{ih} &= \sum_{i=1}^n \sum_{\substack{j,h=1 \\ j \neq h}}^g Y_{ij} Y_{ih} = \sum_{j=1}^n \left( Y_{+j}^2 - \sum_{i=1}^n Y_{ij} \right) \\ &= n \text{var}\{Y_{+j}\} + \frac{1}{n} Y_{++}^2 - Y_{++}. \end{aligned}$$

This implies that if the total number of ties, or equivalently the average degree ( $= Y_{++}/n$ ) is fixed, then the number of out-2-stars is a function of the out-degree variance.

A similar relation holds for the in-2-stars.

This implies that the numbers of out- and in-2-stars represent the variability among the out- and in-degrees.

Further, the number of two-paths represents the covariance between out- and in-degrees.

These degree variances and covariance are to some extent functionally similar to the variance and covariance of  $U_i$  and  $V_i$  in the  $\rho_2$  model.

The preceding specification of the *ERGM* represents popularity and activity differences between actors as well as the two primary structural effects: reciprocity and transitivity.

## Markov graphs

In probability models for graphs, usually the set of nodes is fixed and the set of edges (or arcs) is random.

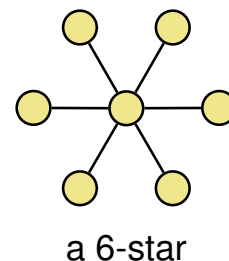
Frank and Strauss (1986) defined that a probabilistic graph is a *Markov graph* if for each set of 4 *distinct* actors  $i, j, h, k$ , the tie indicators  $Y_{ij}$  and  $Y_{hk}$  are *independent, conditionally* on all the other ties.

This generalizes the concept of Markov dependence for time series, where random variables are ordered by time, to graphs where the random edge indicators are ordered by pairs of nodes.

Frank and Strauss (1986) proved that a probability distribution for graphs, under the assumption that the distribution does not depend on the labeling of nodes, is Markov if and only if it can be expressed as

$$P\{Y = y\} = \frac{\exp(\theta L(y) + \sum_{k=2}^{n-1} \sigma_k S_k(y) + \tau T(y))}{\kappa(\theta, \sigma, \tau)}$$

where  $L$  is the edge count,  
 $T$  is the triangle count,  
 $S_k$  is the  $k$ -star count, and  
 $\kappa(\theta, \sigma, \tau)$  is a normalization constant to let the probabilities sum to 1.



It is in practice not necessary to use all  $k$ -star parameters, but only parameters for lower-order stars, like 2-stars and 3-stars.

Varying the parameters leads to quite different distributions. E.g., when using  $k$ -stars up to order 3, we have:

- ▶ higher  $\theta$  gives more edges  $\Rightarrow$  higher density;
- ▶ higher  $\sigma_2$  gives more 2-stars  $\Rightarrow$  more degree dispersion;
- ▶ higher  $\sigma_3$  gives more 3-stars  $\Rightarrow$  more degree skewness;
- ▶ higher  $\tau$  gives more triangles  $\Rightarrow$  more transitivity.

But note that having more triangles and more  $k$ -stars also implies a higher density!

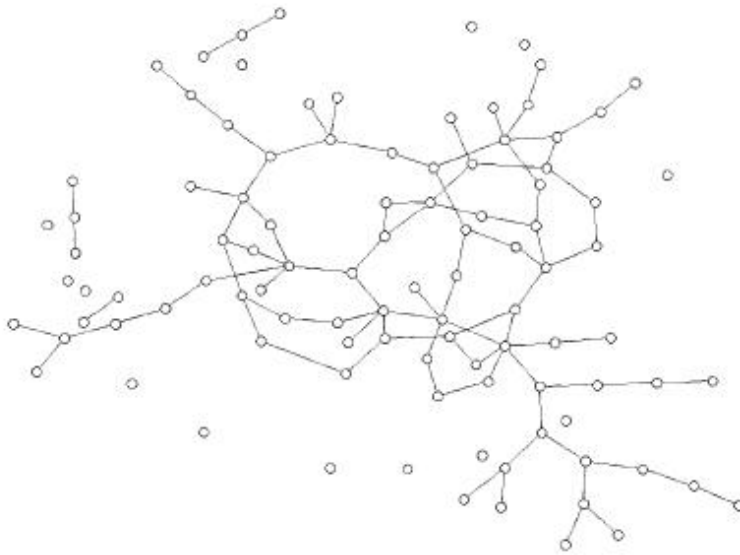
## Small and other worlds

Robins, Woolcock and Pattison (2005) studied these distributions in detail and investigated their potential to generate *small world networks* (Watts, 1999) defined as networks with many nodes, limited average degrees, low geodesic distances and high transitivity.

(Note that high transitivity in itself will lead to long geodesics.)

They varied in the first place the parameters  $\tau, \sigma_k$  and then adjusted  $\theta$  to give a reasonable average degree.

All graphs have 100 nodes.



Bernoulli  
graph:  
random

FIG. 9.—A Bernoulli graph



$(\theta, \sigma_2, \sigma_3, \tau) =$   
 $(-4, 0.1, -0.05, 1.0)$   
small-world graph:  
high transitivity,  
short geodesics

FIG. 5.—A small world graph

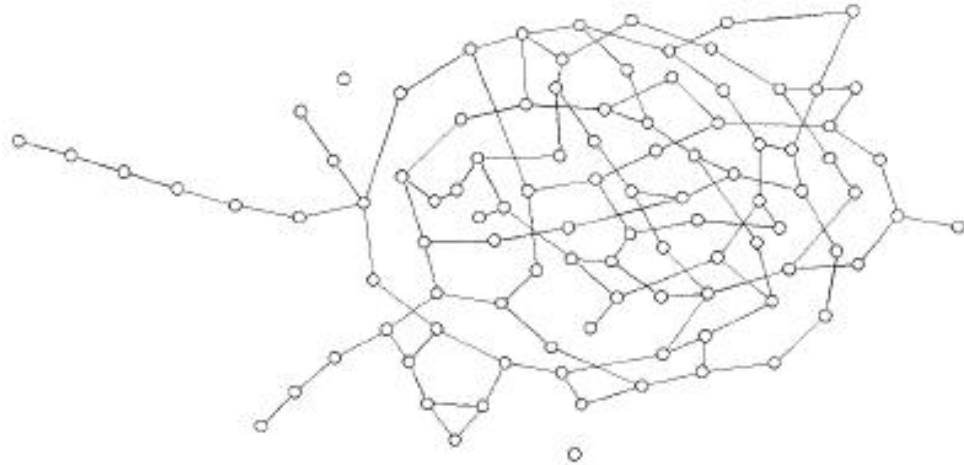


FIG. 7.—A graph with long median paths

$$(\theta, \sigma_2, \sigma_3, \tau) = (-1.2, 0.05, -1.0, 1.0)$$

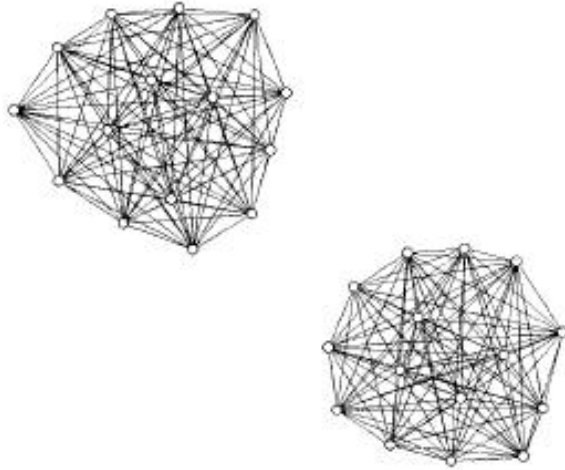
long paths; few high-order stars



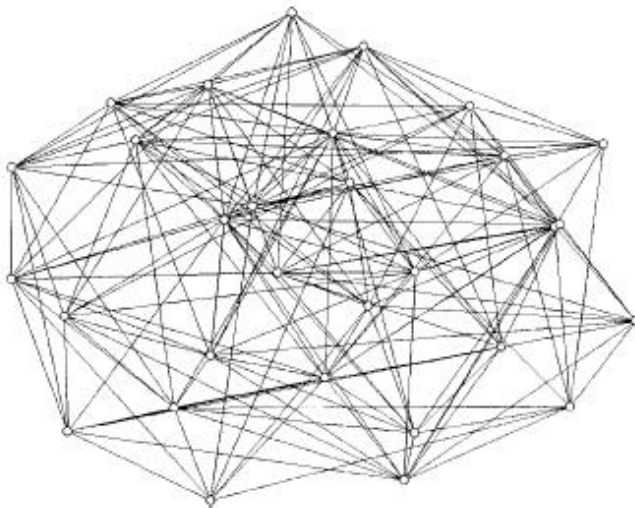
FIG. 8.—A long path graph with low clustering

$$(\theta, \sigma_2, \sigma_3, \tau) = (-2.0, 0.05, -2.0, 1.0)$$

long paths  
low transitivity



$(\theta, \sigma_2, \sigma_3, \tau) =$   
 $(-3.2, 1.0, -0.3, 3.0)$   
caveman world



$(\theta, \sigma_2, \sigma_3, \tau) =$   
 $(-0.533, 0.167,$   
 $-0.05, 0.5)$   
heated  
caveman world

FIG. 11.—Effects of parameter scaling for two temperatures

Thus we see that by varying the parameters,  
many different graphs can be obtained.

This suggests that the Markov graphs will provide  
a good statistical model for modeling observed social networks.

For some time, so-called *pseudo-likelihood methods* were used  
for parameter estimation;  
however, this procedure has poor statistical properties  
and is unreliable.

A better procedure was proposed by Wasserman & Crouch (1999),  
Snijders (2002), Handcock (2003).

This procedure uses *Markov chain Monte Carlo* (MCMC)  
methods to approximate the maximum likelihood estimator.

MLE  $\hat{\theta}$  is the solution of

$$E_{\hat{\theta}}\{s(Y)\} = s(y).$$

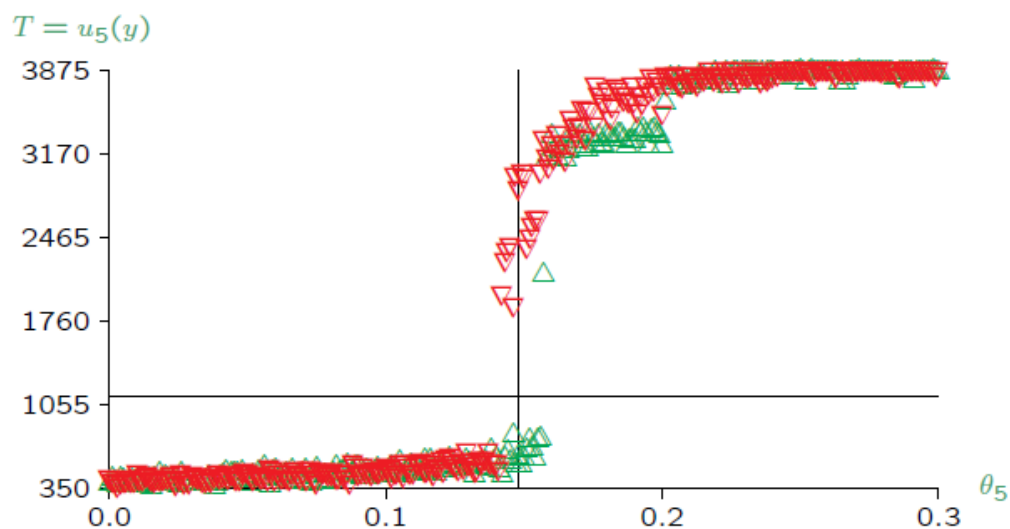
Procedure:

1. Simulate  $Y$  using Gibbs or Metropolis-Hastings algorithm
2. Use stochastic approximation to approximate  $\hat{\theta}$ .

MH steps conditional on total number of ties defined as follows:  
select random  $(i, j) \neq (h, k)$  (with  $i < j$ ,  $h < k$  for graphs);  
proposal distribution is to swap  $Y_{ij}$  and  $Y_{hk}$   
(this has an effect only if  $Y_{ij} \neq Y_{hk}$ ).

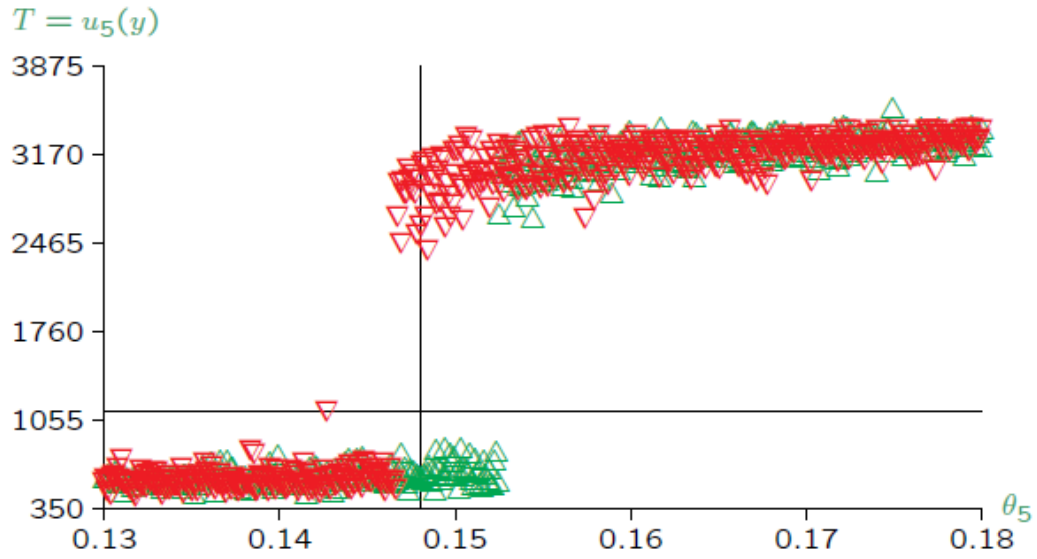
Running the stochastic approximation algorithm to solve the likelihood (= moment) equations may lead to a 'solution' which is not satisfactory because checks indicate that the observed values are not reproduced accurately.

The parameter for the number of transitive triplets is obtained for the Lazega data as  $\hat{\theta}_5 = 0.148$ . Varying this parameter while keeping fixed the others, and simulating digraphs accordingly, leads to the following pictures (note the hysteresis).



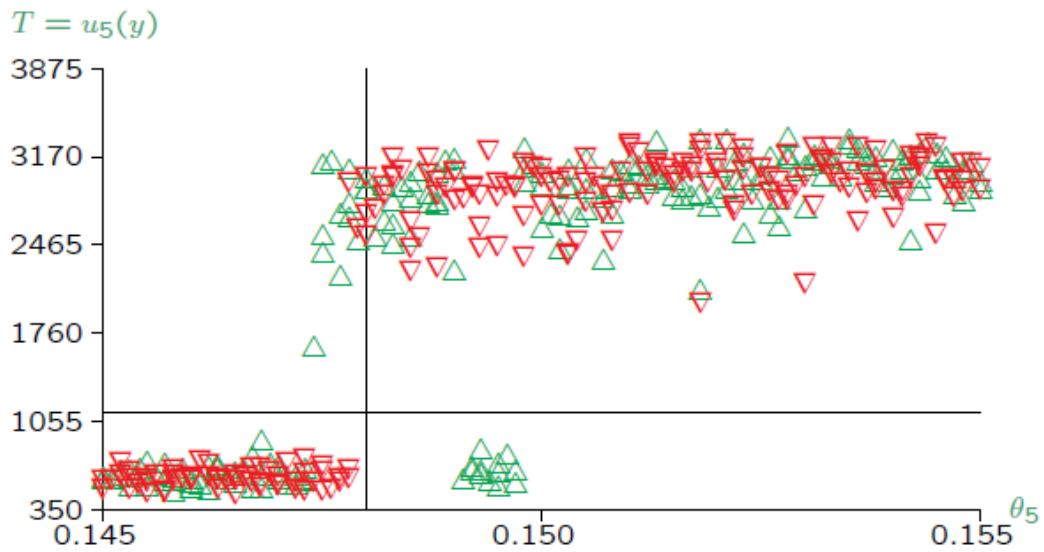
Simulated values of  $T$  with  $\theta_5$  going up ( $\triangle$ ) or down ( $\nabla$ ), sequentially simulated each with 200,000 MH steps.

Exponential Random Graph Models



Simulated values of  $T$  with  $\theta_5$  going up ( $\triangle$ ) or down ( $\nabla$ ), sequentially simulated each with 400,000 MH steps.

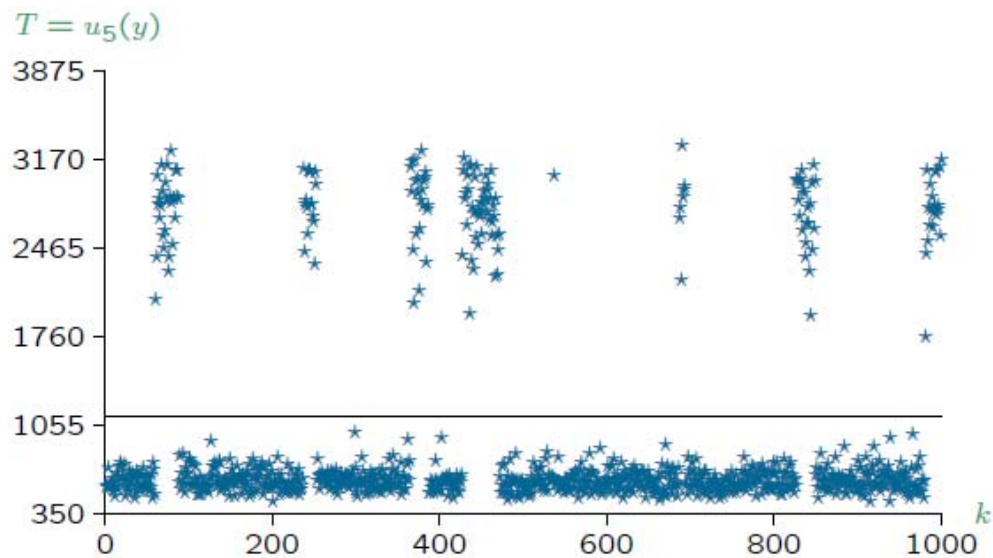
Exponential Random Graph Models



Simulated values of  $T$  with  $\theta_5$  going up ( $\triangle$ ) or down ( $\nabla$ ), sequentially simulated each with 2,000,000 MH steps.

The last picture suggests that for  $\theta_5$  close to 0.148, the distribution of  $T$  is mainly concentrated in two disjoint intervals:  $[400, 800]$  and  $[2000, 3200]$ .

This is investigated by repeated simulation of this model for  $\theta_5 = 0.148$ .



1000 simulated values of  $T$  for  $\theta_5 = 0.148$   
sequentially simulated each with 2,000,000 MH steps.

*Conclusion :*

Representing transitivity in Markov graph models by the number of transitive triplets does not represent well the extent of transitivity as observed in many social networks:

in almost all graphs generated by these models  
the transitivity is either quite weak, or much too strong.

In this example, if the likelihood equations is successfully solved, the observed value  $t$  is well reproduced by  $E_{\theta} T$  not because the distribution of  $T$  is concentrated on values close to  $t = 1125$ , but because the distribution has appropriate probabilities for the separate intervals  $[400, 800]$  and  $[2000, 3200]$  (while probability of values between these intervals is very low).

Handcock (2003) and Snijders (2002) argued that the problems of the Markov model are that, depending on the specification, this ERGM can have a *bimodal* or *degenerate nature*:

e.g., for many parameter values,  
this distribution can be concentrated with a high probability on the full (all  $y_{ij} = 1$ ) or empty (all  $y_{ij} = 0$ ) graph.

Therefore, the specification of the structural statistics, which express the dependencies of ties within the network, must be chosen so that this degeneracy does not occur.

This is discussed now – the ‘new’ (2006) specifications.

## More general specifications

Markov graph models are not flexible enough to represent the degree of transitivity observed in social networks.

It is usually necessary for a good representation of empirical data to generalize the Markov model and include in the exponent also higher-order subgraph counts.

This means that the Markov dependence assumption of Frank and Strauss is too strong, and less strict conditional independence assumptions must be made.

Snijders, Pattison, Robins, and Handcock (2006) proposed a new specification for the ERGM model, defined by other choices of the statistics  $s_k(y)$ .

The new statistics are:

1. alternating in/out- $k$ -star combinations
2. alternating independent 2-paths combinations
3. alternating  $k$ -transitive triplets combination.

The alternating in/out- $k$ -star combinations represent the distribution of the in/out-degrees;  
 the alternating independent 2-paths represent the covariance between in- and out-degrees,  
 and the *conditions* for transitivity;  
 the alternating  $k$ -transitive triplets represent transitivity.

These are explained as follows.

## Alternating in/out- $k$ -star combinations

The number of out-2-stars is a function of out-degree variance.

Therefore, if we wish to represent graphs with a high dispersion of the out-degrees, then we wish to have many out-2-stars; but not too many, since that could bring us too close to a complete graph or another type of degenerate graph where some nodes have out-degree 0 and all others have the same out-degree  $K$  for some high value  $K$ .

This can be represented as follows:  
we wish many 2-stars, but not too many 3-stars.

More generally, an out-degree equal to  $k + h$  contributes  $\binom{k+h}{k}$   $k$ -stars, which is very large if  $h$  is large.

Therefore, if we wish not too have degenerate graphs, a high number of  $k$ -stars should be balanced by a not-too-high number of  $(k + 1)$ -stars.

If we denote the number of out- $k$ -stars by  $S_k$ , this leads to the idea of combining the  $k$ -star counts with *alternating signs*. This is implemented by the following definition of *alternating out- $k$ -star combinations* :

$$\begin{aligned}
u(y) &= S_2 - \frac{S_3}{\lambda} + \frac{S_4}{\lambda^2} - \dots + (-1)^{n-2} \frac{S_{n-1}}{\lambda^{n-3}} \\
&= \sum_{k=2}^{n-1} (-1)^k \frac{S_k}{\lambda^{k-2}} \\
&= \lambda^2 \sum_{i=1}^n \left\{ \left(1 - \frac{1}{\lambda}\right)^{y_{i+}} + \frac{y_{i+}}{\lambda} - 1 \right\}.
\end{aligned}$$

The number  $\lambda \geq 1$  is a constant, chosen by the researcher, which ensures that the contribution of higher-order star counts is downweighted.

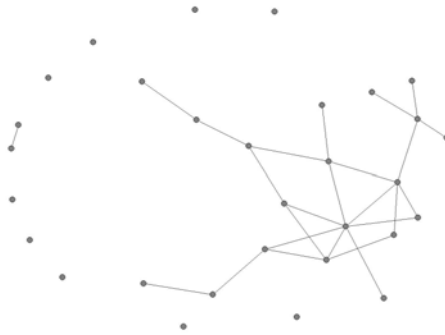
The formula above shows that this linear combination of  $k$ -star counts is a function of the out-degrees  $y_{i+}$ ; the function

$$\sum_{i=1}^n \left\{ \left(1 - \frac{1}{\lambda}\right)^{y_{i+}} + \frac{y_{i+}}{\lambda} - 1 \right\}$$

is an increasing function of the degrees  $y_{i+}$ , and high values of the degrees  $y_{i+}$  are downweighted more strongly when the parameter  $\lambda$  is closer to 1.

Therefore, including this statistic in the vector  $s(y)$  of the ERGM can help to give a good fit to the distribution of the out-degrees. This can be in addition to, or in replacement of, the number of out-2-stars.

The following graph is the outcome of simulating  
 with an edge parameter of  $-4.3$   
 and an alternating  $k$ -star parameter ( $\lambda = 2$ ) of  $0.65$ .  
 This is a low density graph with 25 edges and a density of  $0.06$ .



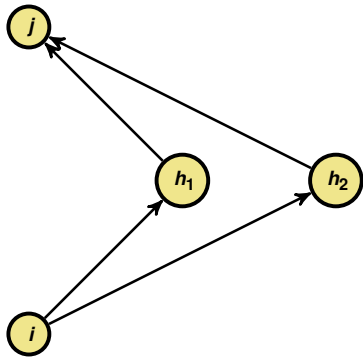
## Alternating independent 2-paths

We saw earlier that two-paths are a precondition for a meaningful (non-vacuous) concept of transitivity; and the number of two-paths is also related to the covariance between in- and out-degrees.

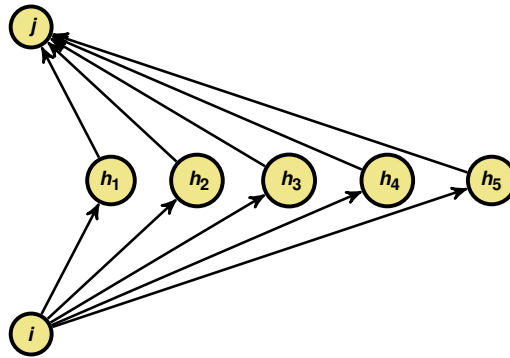
Similar to the alternating  $k$ -star combinations, having many but not too many two-paths between two nodes  $i$  and  $j$  can be obtained by combining two-path counts of higher order with alternating signs.

The graph-theoretical definition is that two-paths are independent if they are non-intersecting.

This statistic is chosen in a way that is combined with the following statistic (alternating  $k$ -transitive triplets) and must be understood together with the latter statistic.



independent 2-two-paths



independent 5-two-paths

Define  $L_{2ij}$  as the number of two-paths connecting  $i$  and  $j$ ,

$$L_{2ij} = \sum_{h \neq i, j} y_{ih} y_{hj}.$$

Then the number of  $k$ -independent 2-paths can be defined by the formula

$$\begin{aligned} U_k &= \#\left\{((i, j), \{h_1, h_2, \dots, h_k\}) \mid \right. \\ &\quad \left. i \neq j, y_{ih_\ell} = y_{h_\ell j} = 1 \text{ for } \ell = 1, \dots, k\right\} \\ &= \sum_{i, j} \binom{L_{2ij}}{k}. \end{aligned}$$

The corresponding statistic  
with alternating and geometrically decreasing weights is

$$\begin{aligned} u(y) &= U_1 - \frac{1}{\lambda} U_2 + \dots + \left(\frac{-1}{\lambda}\right)^{n-3} U_{n-2} \\ &= \lambda \sum_{i < j} \left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}. \end{aligned}$$

The term

$$\left\{ 1 - \left(1 - \frac{1}{\lambda}\right)^{L_{2ij}} \right\}$$

increases from 0 for  $L_{2ij} = 0$  to almost 1 for  $L_{2ij} = n - 2$ .

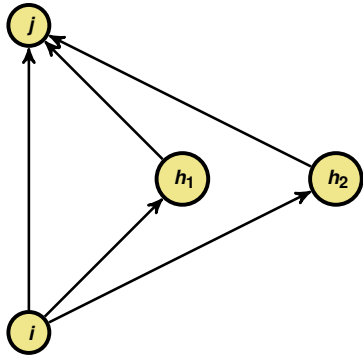
For  $\lambda = 1$  the statistic reduces to

$$u(y) = \sum_{i,j} I\{L_{2ij} \geq 1\},$$

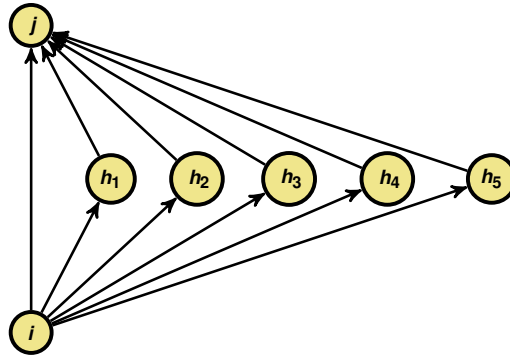
the number of ordered pairs  $(i, j)$   
indirectly connected by at least one two-path.

## Alternating $k$ -transitive triplets

$k$ -transitive triplets are defined as multiple transitive relations on the same "base tie"  $i \rightarrow j$ :



2-transitive triplet



5-transitive triplet

The formula for the number of  $k$ -transitive triplets is

$$\begin{aligned}
 T_k &= \#\left\{((i, j), \{h_1, h_2, \dots, h_k\}) \mid \right. \\
 &\quad \left. y_{ij} = 1 \text{ and } y_{ih_\ell} = y_{h_\ell j} = 1 \text{ for } \ell = 1, \dots, k\right\} \\
 &= \sum_{i < j} y_{ij} \binom{L_{2ij}}{k}.
 \end{aligned}$$

The linear combination of  $k$ -transitive triplets with alternating signs is

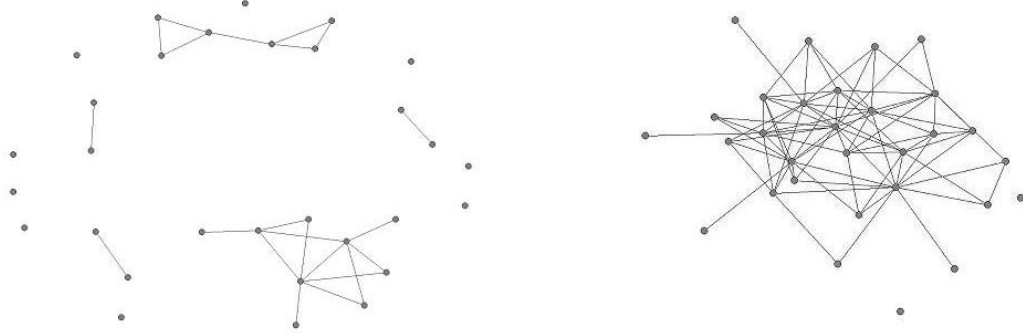
$$\begin{aligned} u(y) &= T_1 - \frac{T_2}{\lambda} + \frac{T_3}{\lambda^2} - \dots + (-1)^{n-3} \frac{T_{n-2}}{\lambda^{n-3}} \\ &= \lambda \sum_{i < j} y_{ij} \left\{ 1 - \left( 1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\}. \end{aligned}$$

This counts the number of transitive triplets but downweights the large numbers of transitive triplets formed by the single tie  $i \rightarrow j$  when there are many actors linked to  $i$  as well as  $j$  — i.e., when  $L_{2ij}$  is large.

For  $\lambda = 1$  the statistic is equal to

$$u(y) = \sum_{i < j} y_{ij} I\{L_{2ij} \geq 1\},$$

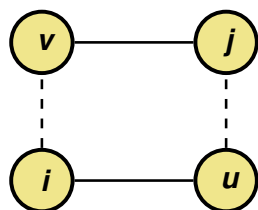
the number of ordered pairs  $(i, j)$  that are directly linked ( $y_{ij} = 1$ ) but also indirectly linked ( $y_{ih} = y_{hj} = 1$  for at least one other node  $h$ ).



$k$ -transitive triplets graphs, low – high densities

Edge parameter =  $-3.7$  for both;  
 alternating  $k$ -transitive triplets parameter =  $1.0$  and  $1.2$ ,  
 respectively ( $\lambda = 2$ ).

These statistics satisfy a weaker dependence assumption than Markov graphs:  
 partial conditional independence (Pattison & Robins, 2002).  
 This assumption states that two possible edges in the graph are dependent, conditionally on the rest of the graph, when they are incident with a common node (this is Markov dependence),  
 or when the presence of these arcs would create a 4-cycle.  
 (For digraphs, directionality is not considered in this respect.)



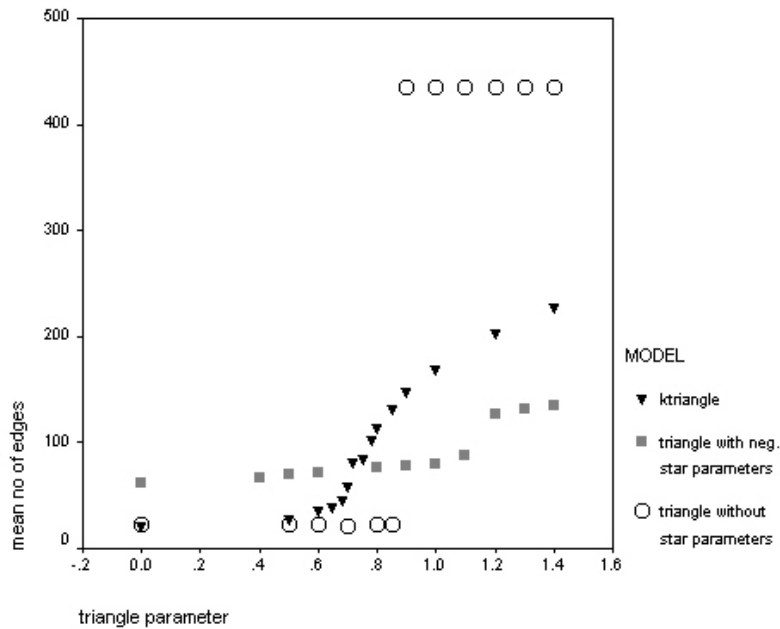
Creation of a 4-cycle  
 by edges  $i - v$  and  $u - j$ .

The advantage of the new specifications is that they allow to model empirically observed networks that have a tendency toward transitivity in the range usually observed in practice.

E.g., transitivity coefficients between 0.1 and 0.6.

This is impossible for Markov graph specifications except for quite small graphs.

The picture on the following page illustrates that in Markov models, the generated graphs sometimes depend almost in a discontinuous way on the parameters, whereas in the new specifications the dependence on parameters is nicely continuous.



Numbers of edges for some generated  $k$ -triangle and Markov graphs, as a function of the transitivity parameter.

The proposed way of specifying the ERGM now is:  
 normally, condition on the total number of edges  $S_1(y)$ ;  
 choose a value of  $\lambda$  — e.g., 1, 2, or 3;  
 and include the following statistics:

1. If (for some reason) one does not condition on the total number of edges, include it among the fitted statistics;
2. the alternating out- $k$ -star combinations and in- $k$ -star combinations, to reflect the heterogeneity of the degrees;
3. (this can be supplemented, or replaced, by the number of out-two-stars, in-two-stars, and two-paths, *provided* that this leads to a good model fitting procedure;)

4. the alternating  $k$ -transitive triplet combinations and the alternating combinations of  $k$ -independent two-paths, to reflect transitivity;
5. (this can be supplemented with the triad count  $T(y) = T_1(y)$ , if a satisfactory estimate can be obtained for the corresponding parameter, and if this yields a better fit as shown from the  $t$ -statistic for this parameter).

The choice of a suitable value of  $\lambda$  depends on the data set.

Actor and dyadic covariate effects can also be added.

Three different effects are possible for each actor covariate  $v_j$ .

These are represented by the following statistics

(potential elements of  $s(y)$ )

1. *covariate-related sender effect* ('ego'),  
sum of sender-covariate over all ties,  $\sum_{i,j} y_{ij} v_i$ ;
2. *covariate-related receiver effect* ('alter'),  
sum of receiver-covariate over all ties  $\sum_{i,j} y_{ij} v_j$ ;
3. *covariate-related similarity*,  
1 minus sum of standardized absolute covariate differences over all ties  $\sum_{i,j} y_{ij} \left(1 - \frac{|v_i - v_j|}{r_v}\right)$ ,  
where  $r_v$  is the range of variable  $v$ ,  $r_v = \max_j v_j - \min_j v_j$ ;

## 4. as an alternative to the similarity:

*interaction covariate-ego*  $\times$  *covariate-alter*,

sum of sender-receiver covariate cross-products

over all ties,  $\sum_{i,j} y_{ij} v_i v_j$ ;

choice between similarity and product interaction

can be based on theory or best fit to data.

Other interactions are also possible;

e.g., consider only reciprocated ties in these effects.

Graphs according to the ERGM can be simulated, and the parameters of the ERGM can be estimated

by a Markov Chain Monte Carlo ('MCMC') procedure described in Snijders (2002)

and implemented in the *SIENA*, *pnet* and *statnet* programs.

The stability of the estimation process is much improved if the total number of ties is kept fixed ('conditioned upon').

The following table presents MCMC parameter estimates for the friendship relation between Lazega's lawyers, with  $\lambda = 2$ .

Parameter	Model 1		Model 2	
	est.	s.e.	est.	s.e.
Mutual dyads	1.653	0.257	2.405	0.291
Out-two-stars	–	–	0.131	0.015
In-two-stars	–	–	0.149	0.024
two-paths	–	–	–0.090	0.019
Alternating out- $k$ -stars	0.235	0.295	–0.333	0.341
Alternating in- $k$ -stars	–1.143	0.551	–2.096	0.927
Alternating $k$ -trans. triplets	0.661	0.142	0.704	0.146
Alternating indep. two-paths	–0.140	0.031	–0.065	0.034
Same office	0.582	0.125	0.859	0.169
Seniority alter	–0.002	0.007	0.001	0.007
Seniority ego	0.013	0.007	0.010	0.006
Seniority dissimilarity	–0.039	0.008	–0.041	0.008
Practice (corp. law) alter	–0.052	0.160	0.067	0.114
Practice (corp. law) ego	0.328	0.137	0.200	0.097
Same practice	0.279	0.125	0.292	0.129

Other effects, such as the number of transitive triplets and the numbers of three-stars and four-stars are also represented adequately, each with a difference between observed and estimated expected value of less than 1.5 standard deviation.

## 5. Multiple Regression Quadratic Assignment Procedure

The MRQAP approach was developed by Hubert (1987) and Krackhardt (1987, 1988).

The basic idea is to apply regular regression coefficients and OLS linear regression analysis to dyadic data collected in square matrices; compute  $p$ -values by a *permutational approach*: the null distribution is obtained by permuting  $X$  values and  $Y$  values with respect to each other, permuting rows and columns ('actors') simultaneously so that the network structure is respected.

It was shown by Dekker, Krackhardt and Snijders (2007) how to do this correctly when controlling for other variables (permute residuals; use pivotal statistics).

This does not model network structure, but controls for it. The MRQAP approach is useful if one is not interested in network structure per se, but wishes to study linear relations between independent and dyadic dependent variables in a network setting.

## Comparison between the diverse network models:

- ▶ Logistic regression and  $p_1$  are inadequate.
- ▶ MRQAP is for linear models – numerical data; the other models are for (di)graphs – binary data.
- ▶  $p_2$  and MRQAP do not model network structure (although  $p_2$  does model reciprocity), they do try to control for it:
  - ▶  $p_2$  controls for actor differences and reciprocity; this may not always be sufficient.
  - ▶ MRQAP controls based on the assumption of permutation-invariant residuals; here also more research is needed to study its adequacy.

- ▶  $p_2$  gives nicely interpretable regression coefficients and variance-covariance parameters.
- ▶ ERGM is the only model that represents details of network structure.

- ▶ MRQAP always works;  
 $p_2$  works well;  
ERGM requires some experience in model specification and estimation.
- ▶ The MCMC procedures for  $p_2$  and ERGM are time-consuming.
- ▶ Many data structures have been elaborated for the  $p_2$  model:  
multiplex networks, multilevel networks, missing data.

## Software

MCMC estimation of the ERGM is available in the programs

1. `statnet` & `ergm` (a set of R packages)
2. `pnet` (stand alone).
3. `SIENA` (stand alone / `StocNet`)  
(but `SIENA` is not kept up to date with newest developments).

## References

- ▶ Dekker, D., Krackhardt, D., and Snijders, Tom A.B. (2007). Sensitivity of MRQAP Tests (...). *Psychometrika*, 72, 563–581.
- ▶ Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica*, 45, 283 – 293.
- ▶ Frank, O., and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81, 832 – 842.
- ▶ M.S. Handcock, D.R. Hunter, C.T. Butts, S.M. Goodreau, M. Morris (2008). `statnet`: Software Tools for (...) of Network Data. *Journal of Statistical Software* 24.1 (Feb. 2008).
- ▶ Harrigan, N. (2007). PNET for dummies.  
<http://www.sna.unimelb.edu.au/pnet/pnet.html#download>
- ▶ Holland, P.W., and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, 76, 33 – 65.

- ▶ Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social Networks*, 9, 171–186.
- ▶ Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10, 359–381.
- ▶ Lazega, E., and van Duijn, M.A.J. (1997). Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data. *Social Networks* 19, 375 – 397.
- ▶ Robins, G., Pattison, P., Kalish, Y., and Lusher, D. 2007. "A introduction to exponential random graph ( $p^*$ ) models for social networks." *Social Networks* 29, 173–191.
- ▶ Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P. 2007. Recent developments in Exponential Random Graph ( $p^*$ ) Models for Social Networks. *Social Networks*, 29, 192–215.

- ▶ Robins, G.L., Woolcock, J., and Pattison, P. 2005. Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, 110, 894–936.
- ▶ Snijders, T.A.B. 2002. MCMC Estimation of Exponential Random Graph Models. *Journal of Social Structure*, Vol. 3 (2002), No. 2.
- ▶ Snijders, T.A.B., Pattison, P.E., Robins, G. and M.S. Handcock. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 99–153.
- ▶ van Duijn, M.A.J., Snijders, T.A.B., & Zijlstra, B.H. (2004).  $p_2$ : a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58, 234–254.
- ▶ Wasserman, S., and K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- ▶ Wasserman, S., and P. Pattison (1996). Logit models and logistic regression for social networks: I. An introduction to Markov graphs and  $p^*$ . *Psychometrika* 61, 401 – 425.