The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

International Atomic Energy Agency

SMR.1656 - 22

School and Workshop on
Structure and Function of Complex Networks

16 - 28 May 2005

Actor-driven Models for Network Dynamics

Tom SNIJDERS
Department of Sociology
University of Groningen
Grote Rozenstraat 31
Groningen 9712 TG
THE NETHERLANDS

ACTOR-DRIVEN MODELS FOR NETWORK DYNAMICS

Tom A.B. Snijders

ICS

University of Groningen,

The Netherlands

May 2005

Networks in Social Science

In the empirical study of networks in social sciences,
it is important to use models
that link *network processes* to *statistical inference*.

Requirements:

⇒ good representation of empirical reality

⇒ assessment of uncertainty in conclusions from empirical data (real, not stylized facts)

⇒ procedures for estimating and testing parameters

⇒ procedures for assessing the fit of the model to the data

⇒ flexibility to adapt model if the fit is not satisfactory.

This presentation

1. Stochastic actor-oriented model
   - ⇒ basic model: objective function
   - ⇒ extensions: rate function, gratification function
2. Procedures for parameter estimation
3. Example: friendship dynamics in student group
4. Extension: networks and behavior.

Notation:

n nodes: *social actors*,
with a binary ("on/off") relation,
represented as a *directed graph (digraph)*.

Existence of tie from i to j indicated by X_{ij} :

$$X_{ij} = \begin{cases} 1 & \text{if there is a tie from } i \text{ to } j \\ 0 & \text{if there is no such tie} \end{cases}$$

indicating *arc* from i to j.
(Diagonal values X_{ii} meaningless.)

Matrix X is *adjacency matrix* of digraph.
X_{ij} is a *tie indicator* or *tie variable*.

Data: ≥ 2 repeated observations of a network / digraph.
Set of nodes (actors) is fixed, or changes exogenously.
Think of small node sets:
30-100 pupils, 10-100 colleagues, 30-500 firms.

Model: networks depend on a continuous time parameter:
matrix $X(t)$, element $X_{ij}(t)$.

Actor-driven models :

each actor "controls" his outgoing relations,
collected in the row vector $\big(X_{i1}(t), ..., X_{in}(t)\big)$.

At stochastic times (*rate function* λ),
the actors may change these outgoing relations.

The actors try to attain a rewarding position in the network.
The appreciation by actor *i* of his/her position in the network x
is expressed by the *objective function* $f_i(x)$.
The objective, or aim, of actor i is
to achieve a high value of the objective function $f_i(x)$.

The functions λ and f depend on K-dimensional
statistical parameter $\theta \in \Theta \subset \mathbb{R}^K$
(to be estimated from data).

Model for changes:

At random moments,
one random actor is designated to make a change in one relation:
on $\Rightarrow$ off, or off $\Rightarrow$ on.

This actor tries to improve his/her objective function
and looks only on its value immediately after this change
(*myopia*) .

This absence of strategy or farsightedness in the model
implies the *interpretation* of objective function as
"what the actors try to achieve in the short run".

Simple model specification:

* The actors all change their relationships at random moments, at the same rate ρ.

* Each actor tries to optimize an *objective function* with respect to the network configuration,

$$f_i(\beta, x)\ , \quad i = 1, ..., n, \quad x \in \mathcal{X}\ ,$$

which indicates the preference of actor i for the relational situation represented by x; objective function depends on *parameter* β.

Whenever actor i may make a change,
he changes only one relation, say x_{ij}.

The new network is denoted by $x(i \leadsto j)$.

Actions are propelled also by a *random component*,
expressing *unexplained change* ('residual term').

Actor i chooses the "best" j by maximizing

$$f_i\Big(\beta, x(i \leadsto j)\Big) \;+\; U_i(t, x, j)\,.$$

$$\Uparrow$$

random component

For a convenient choice of the distribution
of the random component,
(type 1 extreme value = Gumbel distribution)
given that i is allowed to make a change,
the probability that i changes his relation with j is

$$p_{ij}(\beta, x) = \frac{\exp\big(f(i,j)\big)}{\sum_{h=1, h\neq i}^{n} \exp\big(f(i,h)\big)} \quad (j \neq i).$$

where

$$f(i,j) = f_i\Big(\beta, x(i \rightsquigarrow j)\Big) \ .$$

This is the multinomial logit form of a *random utility* model.

The Gumbel distribution has variance $\pi^2/6 = 1.645$ and s.d. 1.28.

Intensity matrix

This specification implies that X follows a *continuous-time Markov chain* with intensity matrix

$$q_{ij}(x) = \lim_{dt \downarrow 0} \frac{\mathrm{P}\Big\{X(t+dt) = x(i \leadsto j) \mid X(t) = x\Big\}}{dt} \qquad (i \neq j)$$

given by

$$q_{ij}(x) = \rho \, p_{ij}(\beta, x) \, .$$

Model specification :

The objective functions f_i reflect
network effects (endogenous) and covariate effects (exogenous).

Covariates can be actor-dependent: v_i
or dyad-dependent: w_{ij} .

Convenient definition of objective function f_i
is a weighted sum

$$f_i(\beta, x) = \sum_{k=1}^{L} \beta_k \, s_{ik}(x) \, ,$$

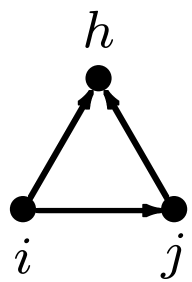
where weights β_k are statistical parameters
$s_{ik}(x)$ are statistics which can also depend on v_i and w_{ij} .

Choose possible network effects for actor i, e.g.:
(others to whom actor i is tied are called here i's 'friends')

1. *out-degree effect*,
   $s_{i1}(x) = x_{i+} = \sum_j x_{ij}$
2. *reciprocity effect*, number of reciprocated relations
   $s_{i2}(x) = \sum_j x_{ij} \, x_{ji}$

Two effects related to network closure:

3. *transitivity effect*,
   number of transitive patterns in i's relations
   $(i \rightarrow j,\ j \rightarrow h,\ i \rightarrow h)$
   $s_{i3}(x) = \sum_{j,h} x_{ij}\, x_{jh}\, x_{ih}$



transitive triplet

4. *indirect relations effect*,
   number of actors j to whom i is indirectly related
   (through at least one intermediary: $x_{ih} = x_{hj} = 1$)
   but not directly ($x_{ij} = 0$),
   = number of geodesic distances equal to 2,
   $s_{i4}(x) = \#\{j \mid x_{ij} = 0,\ \max_h(x_{ih}\, x_{hj}) > 0\}$

Differences between these two network closure effects:

$\Rightarrow$ transitive triplets effect:
i more attracted to j if there are
more indirect ties $i \rightarrow h \rightarrow j$;

$\Rightarrow$ negative indirect connections effect:
i more attracted to j if there is
at least one such indirect connection .

Non-formalized sociological theories usually do not distinguish between these different closure effects.
It is possible to 'let the data speak for themselves'
and see what is the best formal representation of closure effects.

Good representation of such details may be necessary
for a reliable representation also of *other* effects.

Three kinds of objective function effect associated with actor covariate v_i :

5. *covariate-related popularity*,
   sum of covariate over all of i 's friends
   $s_{i5}(x) = \sum_j x_{ij} v_j$;
6. *covariate-related activity*,
   i's out-degree weighted by covariate
   $s_{i6}(x) = v_i x_{i+}$;
7. *covariate-related similarity*,
   sum of measure of covariate similarity between i and his friends, e.g.
   $s_{i7}(x) = \sum_j x_{ij} \Big(1 - |v_i - v_j|\Big)$
   if V has values between 0 and 1.

Objective function effect for dyadic covariate w_{ij} :

8. *covariate-related preference*,
   sum of covariate over all of i's friends,
   i.e., values of w_{ij} summed over all others to whom i is related,
   $s_{i8}(x) = \sum_j x_{ij} \, w_{ij}$.

   If this has a positive effect, then the value of a tie $i \rightarrow j$ becomes higher when w_{ij} becomes higher.

Statistical estimation

Suppose that at least 2 observations on $X(t)$ are available,
for observation moments $t_1,\ t_2$ (or more).

How to estimate $\theta = (\beta, \rho)$?

Condition on $X(t_1)$:
the first observation is accepted as given,
contains in itself no observation about θ.

No assumption of a stationary network distribution.

Method of moments :

choose a suitable statistic $Z = (Z_1, \ldots, Z_K)$,
i.e., K variables which can be calculated from the network;
the statistic Z must be *sensitive* to the parameter θ
in the sense that higher values of θ_k
lead to higher values of the expected value $\mathrm{E}_\theta(Z_k)$;

determine value of θ for which
observed and expected values of suitable statistic are equal,

$$\mathrm{E}_{\widehat{\theta}}\{Z\} = z \ .$$

Issues:

* What is a suitable (K-dimensional) statistic Z ?
  Based on observed amount of change
  and components of objective function.

* Solve this equation in θ by stochastic approximation.
  Iteration step:

$$\hat{\theta}_{N+1} = \hat{\theta}_N - a_N \, D^{-1}(z_N - z) \ , \tag{1}$$

  where D is a suitable matrix,
  z_N is a simulation of Z with parameter $\hat{\theta}_N$,
  and a_N is a sequence $a_N \rightarrow 0$.

Covariance matrix

The method of moments yields estimator with covariance matrix

$$\text{cov}(\hat{\theta}) \approx D_\theta^{-1}\,\Sigma_\theta\,D_\theta'^{-1}$$

where

$$\begin{aligned}\Sigma_\theta &= \text{cov}\{Z \,|\, X(t_1) = x(t_1)\}\\ D_\theta &= \frac{\partial}{\partial\theta}\text{E}\{Z \,|\, X(t_1) = x(t_1)\}\,.\end{aligned}$$

(Note: Z is function of $X(t_1)$ and $X(t_2)$).

Summary of estimation algorithm

3 phases:

1. brief phase for preliminary estimation of $\partial \mathrm{E}_{\hat{\theta}}\{Z\}/\partial\theta$ for defining D;
2. estimation phase with Robbins-Monro updates, where a_N remains constant in *subphases* and decreases between subphases;
3. final phase where θ remains constant at its estimated value; this phase is for checking that

$$\mathrm{E}_{\hat{\theta}}\{Z\} \approx z \ ,$$

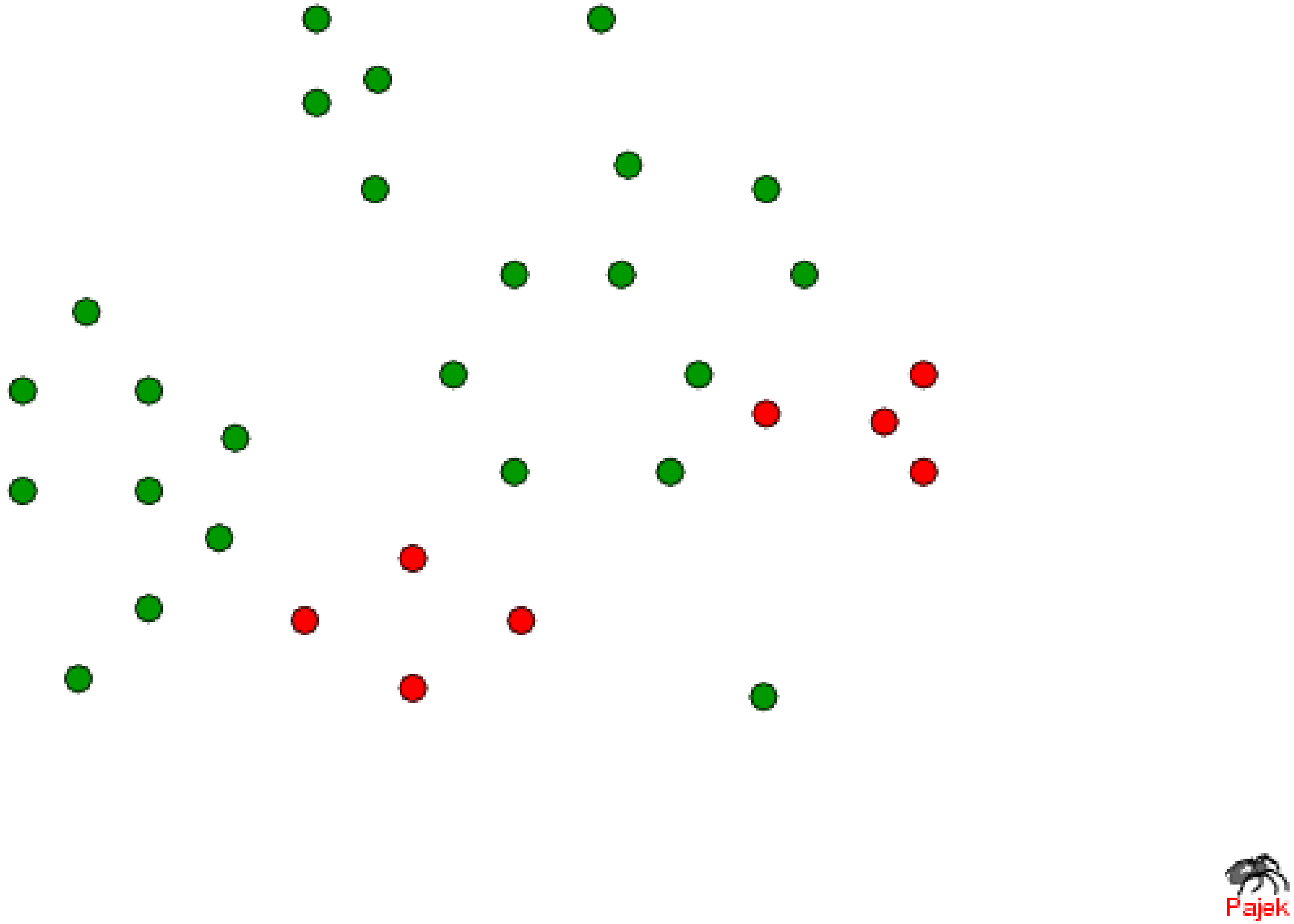
and for estimating D_θ and Σ_θ to calculate standard errors.

Example:
Studies Gerhard van de Bunt

1. Study of 32 freshman university students,
   7 waves in 1 year.
   See van de Bunt, van Duijn, & Snijders,
   Computational & Mathematical Organization Theory,
   5 (1999), 167 – 192.

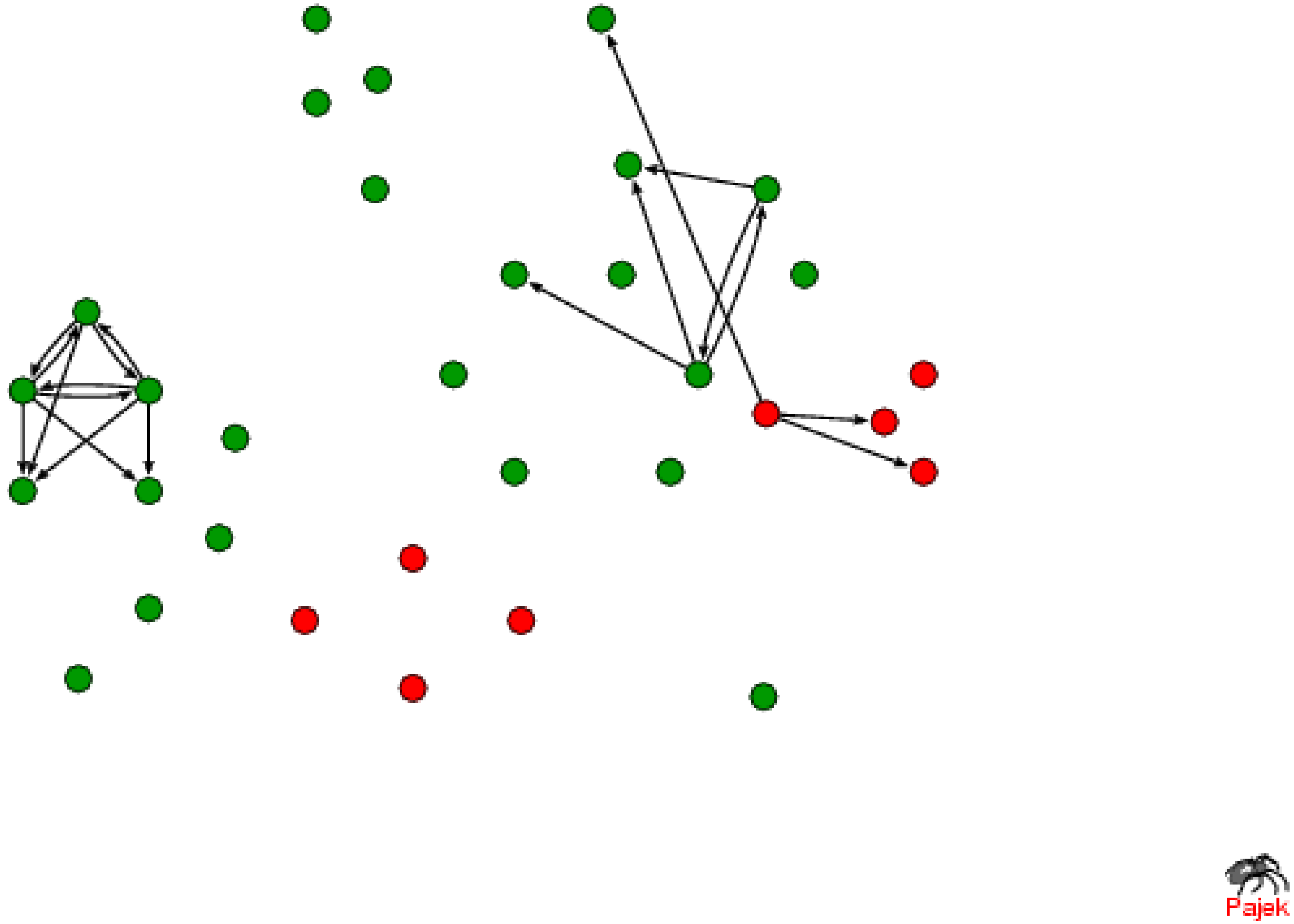
2. Study of hospital employees,
   2 departments (49 and 30 actors), 4 waves.

This presentation concentrates on the first data set,
which can be pictured by the following graphs
(arrow stands for 'best friends').



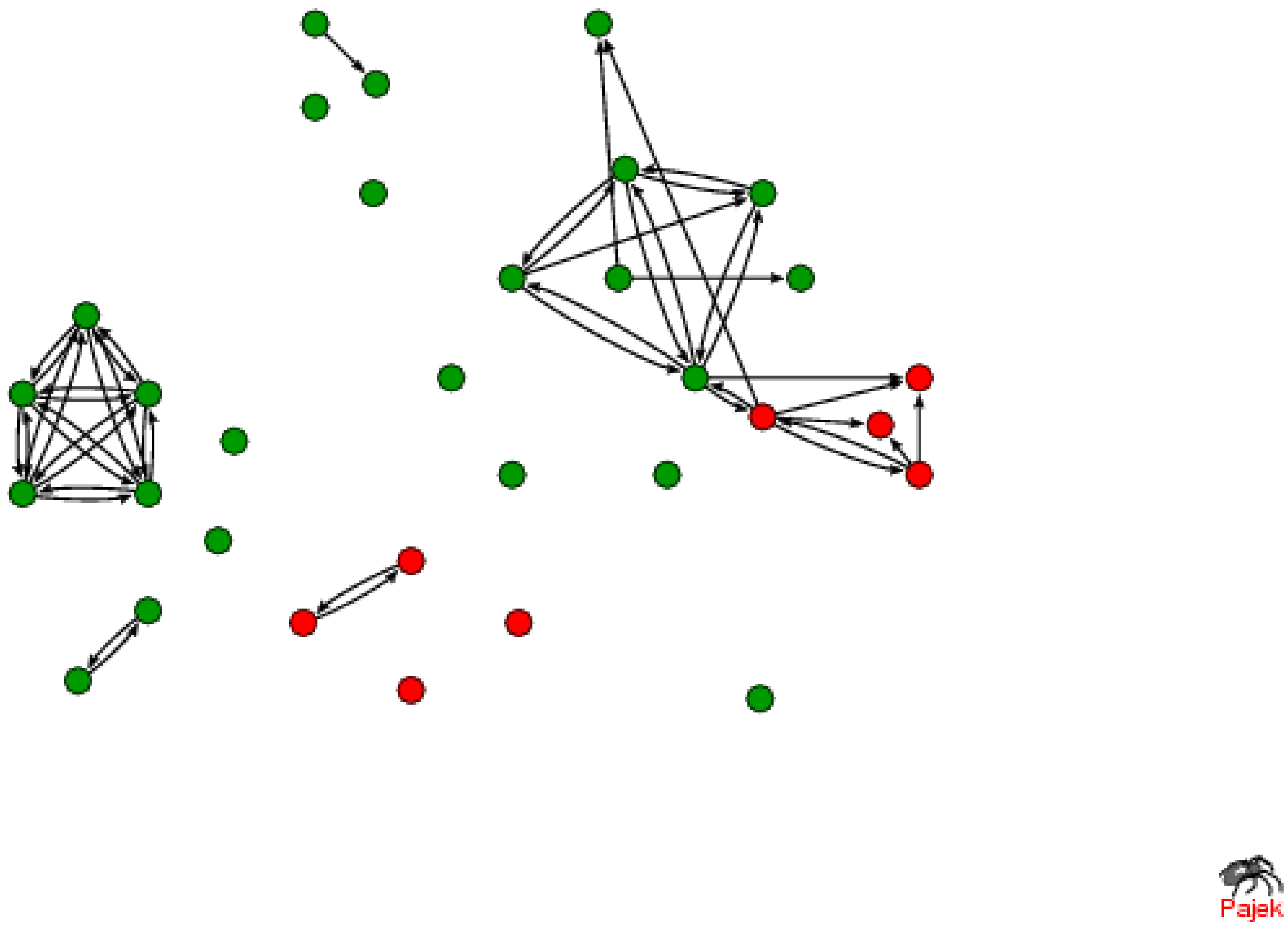
Friendship network time 1.

Average degree 0.0; missing fraction 0.0.



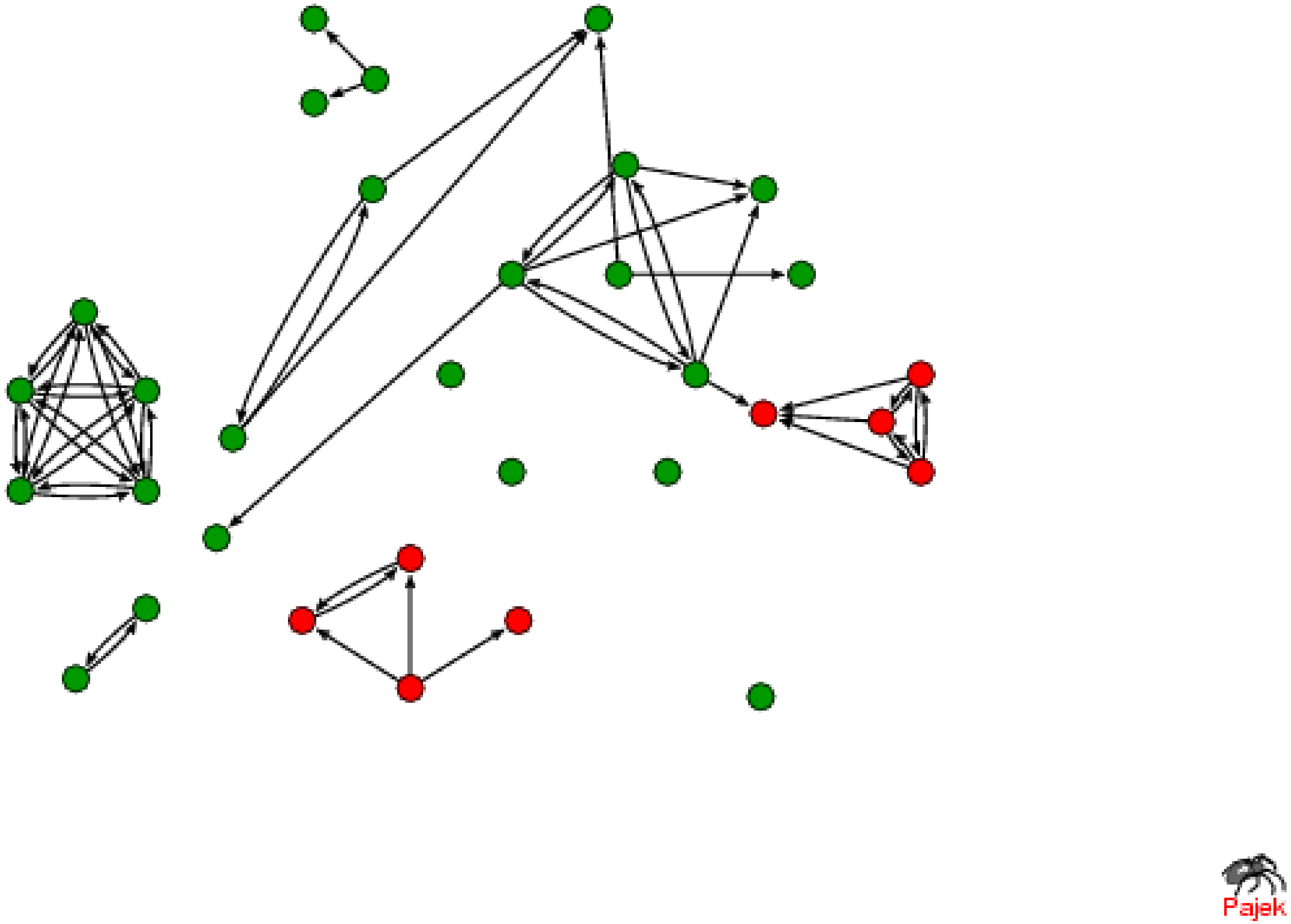
Friendship network time 2.

Average degree 0.7; missing fraction 0.06.



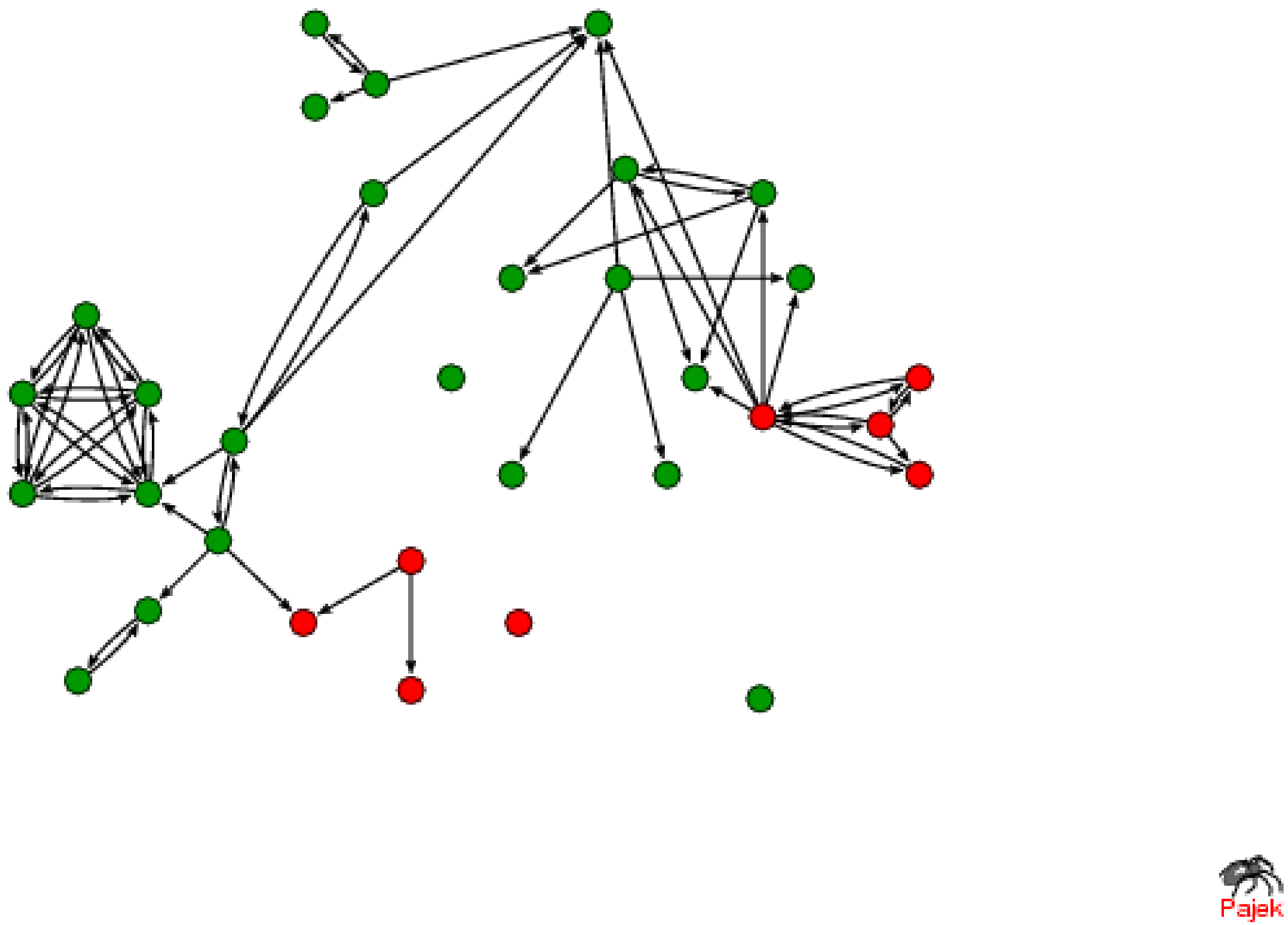
Friendship network time 3.

Average degree 1.7; missing fraction 0.09.



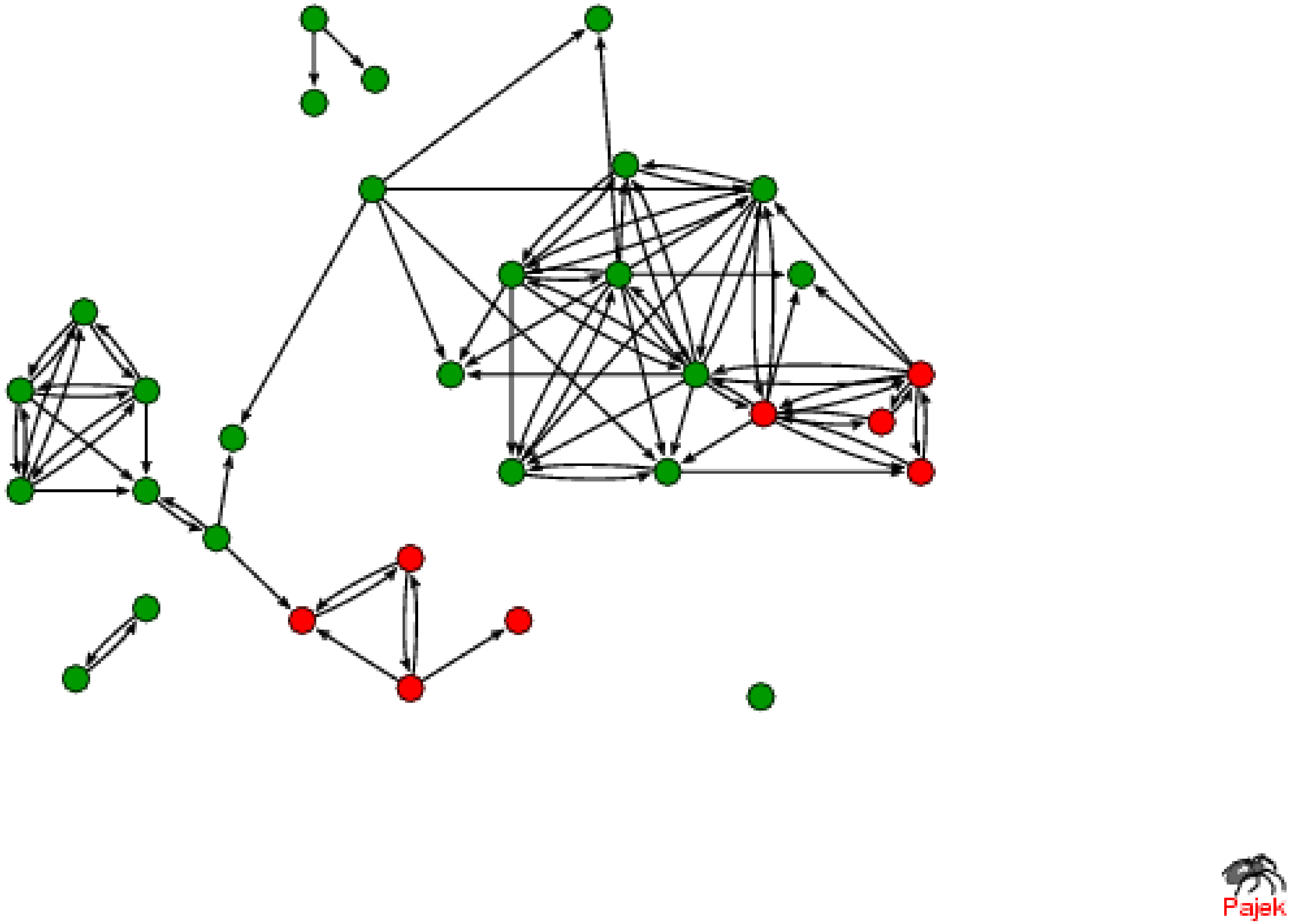
Friendship network time 4.

Average degree 2.1; missing fraction 0.16.



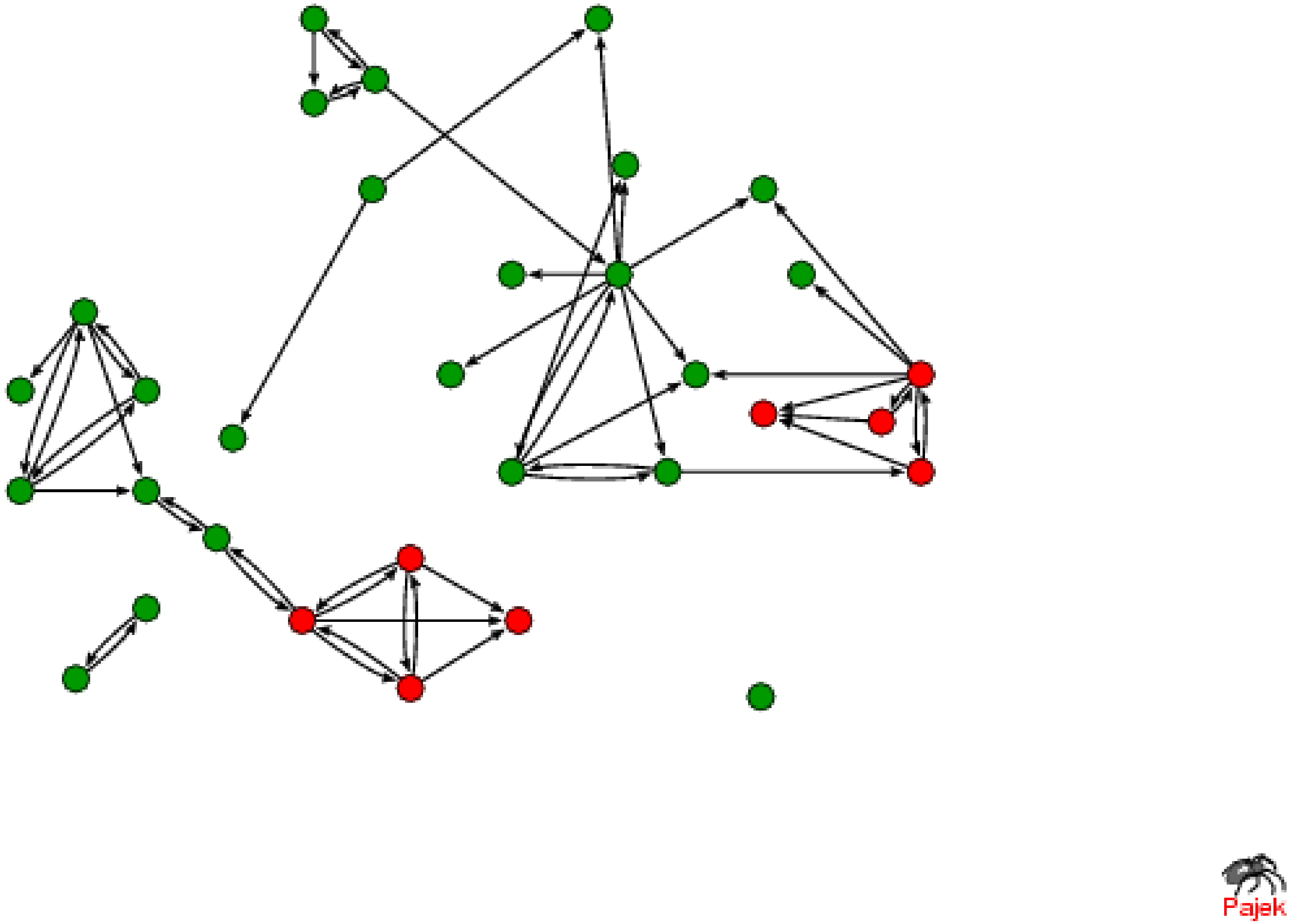
Friendship network time 5.

Average degree 2.5; missing fraction 0.19.



Friendship network time 6.

Average degree 2.9; missing fraction 0.04.



Friendship network time 7.

Average degree 2.3; missing fraction 0.22.

Model with two network closure effects

	Model 1	
Effect	par.	(s.e.)
Rate $t_1 - t_2$	3.87	(0.57)
Rate $t_2 - t_3$	3.12	(0.48)
Density	−1.45	(0.26)
Reciprocity	1.90	(0.33)
Transitive triplets	0.22	(0.12)
Indirect relations	−0.32	(0.07)

Conclusion from t-values (estimate / standard error):
negative indirect relations effect
much stronger than transitive triplets effect.

Model with only indirect relations effect

	Model 2	
Effect	par.	(s.e.)
Rate (period 1)	3.78	
Rate (period 2)	3.14	
Out-degree	-1.05	(0.19)
Reciprocity	2.44	(0.40)
Indirect relations	-0.557	(0.083)

Conclusion: high value for reciprocity;
negative value for non-reciprocated ties;
negative value of indirect relations $\sim$ preference for network closure.

Model with indirect relations and gender effects

(Gender: F $= -0.25$, M $= 0.75$.)

	Model 3	
Effect	par.	(s.e.)
Rate (period 1)	3.91	
Rate (period 2)	3.07	
Out-degree	-1.13	(0.22)
Reciprocity	2.52	(0.37)
Indirect relations	-0.502	(0.084)
Gender activity	-0.60	(0.28)
Gender popularity	0.64	(0.24)
Gender dissimilarity	-0.42	(0.24)

Conclusion:
Women more active;
men more popular;
no significant sex
dissimilarity effect,

Extended model specification

1. Gratification function / endowment effect $g_i(\gamma, x, j)$

This represents the "gratification" experienced by the actor when he *makes* a particular *change* in his relations, rather than when he *has* a particular configuration of relations.

Is used to represent models where certain effects work differently for *creation* of ties ($0 \longrightarrow 1$) than for *termination* of ties ($1 \longrightarrow 0$).

Again, linear combination of theoretically proposed effects.

Extended model specification

2. Non-constant rate function $\lambda_i(\alpha, x)$.

This means that some actors change their relations
more quickly than others,
depending on covariates or network position.

Dependence on network position and covariates:

$$\lambda_i(\alpha, x) = \exp(\sum_h \alpha_h v_{hi}) \ ,$$

where v_{hi} can refer to a covariate
(constant, or exogenously changing)
or an indicator of network position such as degree
(endogenously changing).

Current work:

Extend this type of modeling to the joint dynamics
of networks and behavior tendencies, to get
separate estimates of social influence and social selection processes.

Model sketch :

⇒ Separate changes for networks and for behavior

⇒ Rate functions for networks and for behavior

⇒ Objective & gratification functions for both.

The term

$$\beta_k \sum_j x_{ij} \, \mathrm{sim}(i,j) \ ,$$

where x_{ij} is the network, $\mathrm{sim}(i,j)$ a measure for behavior similarity, represents influence when it is a component of the behavior objective function,
and selection when it is a component of the network objective function.

Results for data of one school class (130 pupils) (Scotland),
3 waves of data collection (13-15 years).

First the model for friendship dynamics:

Effect	par.	(s.e.)
Rate $t_1 - t_2$	12.44	(1.51)
Rate $t_2 - t_3$	9.29	(1.06)
Out-degree	−2.21	(0.32)
Reciprocity	2.05	(0.16)
Transitive triplets	0.15	(0.04)
Indirect relations	−0.82	(0.03)
Classmate	0.01	(0.05)
Sex (F) popularity	−0.21	(0.10)
Sex (F) activity	0.20	(0.09)
Sex similarity	0.83	(0.21)
Smoking popularity	−0.10	(0.04)
Smoking activity	0.17	(0.09)
Smoking similarity	0.47	(0.34)
Alcohol popularity	0.06	(0.04)
Alcohol activity	−0.07	(0.10)
Alcohol similarity	0.72	(0.31)

Effect	smoking par.	smoking (s.e.)	alcohol par.	alcohol (s.e.)
Rate $t_1 - t_2$	1.24	(1.60)	1.57	(0.40)
Rate $t_2 - t_3$	1.03	(0.85)	2.43	(0.71)
Tendency	−2.07	(0.76)	0.26	(0.26)
Similarity w. friends	0.55	(0.45)	0.87	(0.21)
Sex (F)	0.18	(0.12)	1.03	(0.85)
Parent smoking	0.31	(0.36)	0.05	(0.15)
Sibling smoking	−0.50	(0.42)	0.11	(0.20)
Other behavior (alc / sm)	0.64	(0.25)	0.02	(0.17)

Conclusion :
social influence *and* social selection on alcohol;
alcohol promotes smoking, not vice versa.

Issue :
robustness to model specification.

The procedure is implemented in the program

S imulation

I nvestigation for

E mpirical

N etwork

A nalysis

(current version 2.0) which can be downloaded from
http://stat.gamma.rug.nl/snijders/siena.html
(programmed by Tom Snijders, Mark Huisman, Christian Steglich, Michael Schweinberger).

A Windows shell is contained in the StOCNET package
(current version 1.6)
developed by Peter Boer
(contributions by Evelien Zeggelink, Mark Huisman, Christian Steglich)
http://stat.gamma.rug.nl/stocnet/

Further work on this line of modeling

* Models for non-directed ties, where two actors are involved in deciding to create and break ties.
* Other richer data structures (multivariate, valued ties, etc.)
* Maximum likelihood estimation.
* Procedures for assessing model fit.
* Unobserved heterogeneity of actors.
* Various applications.