Tensions and Treasures

Tom A.B. Snijders

University of Oxford
University of Groningen

July 1, 2010

1. tensions
2. treasures
3. tell me who your friends
4. the road goes ever

1. tensions
2. treasures
3. tell me who your friends
4. the road goes ever

1. tensions
2. treasures
3. tell me who your friends
4. the road goes ever

1. tensions
2. treasures
3. tell me who your friends
4. the road goes ever

Tensions

though redundant to calculate $x(t_2)$ from $x(t_1)$, are retained because they facilitate the computation of the likelihood. Define $x^{(k)} = x(T_k)$; the digraphs $x^{(k)}$ and $x^{(k-1)}$ differ in element (I_k, J_k) provided $I_k \neq J_k$, and in no other elements.

The probability function of the sample path, conditional on $x(t_1)$, is given by

$$
\begin{aligned}
&p_{\mathrm{sp}}\{V = \big((i_1, j_1), \ldots, (i_R, j_R)\big); \alpha, \beta\} = \\
(14) \qquad &\qquad P_{\alpha,\beta}\{T_R \le t_2 < T_{R+1} \mid x^{(0)}, (i_1, j_1), \ldots, (i_R, j_R)\} \\
&\qquad \times \prod_{r=1}^{R} \pi_{i_r}(\alpha, x^{(r-1)})\, p_{i_r, j_r}(\beta, x^{(r-1)}) \ ,
\end{aligned}
$$

where π_i is defined in (2), and p_{ij} in and just after (5). Denote the first component of (14) by

$$
\begin{aligned}
(15) \quad &\kappa\big(\alpha, x^{(0)}, (i_1, j_1), \ldots, (i_R, j_R)\big) \\
&\qquad = P_{\alpha,\beta}\{T_R \le t_2 < T_{R+1} \mid x^{(0)}, (i_1, j_1), \ldots, (i_R, j_R)\} \ .
\end{aligned}
$$

Conditioning on $x^{(0)}, (i_1, j_1), (i_2, j_2), \ldots,$ (and not on $x(t_2)$!), the differences $T_{r+1} - T_r$ are independently exponentially distributed with parameters $\lambda(\alpha, x^{(r)})$. Hence under this conditioning the distribution of $T_R - t_1$ is

Tensions? No formulae (almost).

Statistics is traditionally based on

Statistics is traditionally based on

- ⇒ sampling from a population
- ⇒
- ⇒

Statistics is traditionally based on

- ⇒ sampling from a population
- ⇒ generalizing from sample to population
- ⇒

Statistics is traditionally based on

- ⇒ sampling from a population
- ⇒ generalizing from sample to population
- ⇒ independence assumptions .

Social network analysis traditionally is concerned with

- ⇒ sampling from a population
- ⇒ generalizing from sample to population
- ⇒ independence assumptions .

Social network analysis traditionally is concerned with

- ⇒ study of one group
- ⇒ generalizing from sample to population
- ⇒ independence assumptions .

Social network analysis traditionally is concerned with

- ⇒ study of one group
- ⇒ interest in this particular group
- ⇒ independence assumptions .

Social network analysis traditionally is concerned with

- ⇒ study of one group
- ⇒ interest in this particular group
- ⇒ focus on dependencies .

Statistics is traditionally based on

Social network analysis traditionally is concerned with

⇒ sampling from a population
study of one group

⇒ generalizing from sample to population
interest in this particular group

⇒ independence assumptions .
focus on dependencies .

How can these tensions be overcome?

.

What is the specific nature of how statistical inference contributes to scientific progress?

What is the specific nature of how statistical inference contributes to scientific progress?

very impolitely ...

1. process data to reveal complicated dependencies

   hard computation

What is the specific nature of how statistical inference contributes to scientific progress?

very impolitely ...

1. process data to reveal complicated dependencies

   hard computation

2. gauge uncertainty in the results

What is the specific nature of how statistical inference contributes to scientific progress?

very impolitely ...

1. process data to reveal complicated dependencies
   hard computation
2. gauge uncertainty in the results
   p-values, standard errors, posterior distributions

What is the specific nature of how statistical inference contributes to scientific progress?

very impolitely ...

1. process data to reveal complicated dependencies
   hard computation
2. gauge uncertainty in the results
   p-values, standard errors, posterior distributions
3. combining these:
   assess quality of models / theories
   to enable moving forward in the empirical cycle.

First I shall talk
about the models for complicated dependencies
(models for networks as dependent variables;
with some historical remarks)

First I shall talk
about the models for complicated dependencies
(models for networks as dependent variables;
with some historical remarks)

and then about the probability assumptions
used for gauging uncertainty

First I shall talk
about the models for complicated dependencies
(models for networks as dependent variables;
with some historical remarks)

and then about the probability assumptions
used for gauging uncertainty

and then again about models for dependencies.

Models for network dependencies

Models for network dependencies

Network dependencies were noted long ago.

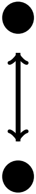
Models for network dependencies

Network dependencies were noted long ago.

1. *Reciprocity* was noted by Moreno (1934) and has been investigated a lot.

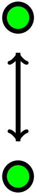
   An $n \times n$ adjacency matrix can be split into $\binom{n}{2}$ dyads;

   $\Rightarrow$ reciprocity is a manageable dependency.



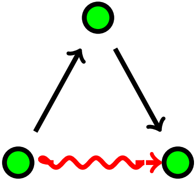
2. *Homophily* was discovered by Lazarsfeld and Merton (1954) and is related to exogenous variables (as opposed to endogenous structure);

   ⇒ homophily is a manageable dependency.



3

Transitivity was discussed already by Rapoport (1954).
Davis, Holland and Leinhardt (1970s) had an impressive research program demonstrating the ubiquity of transitivity in sociometric data.



Holland and Leinhardt (1976) systematized this to measures for *triadic dependencies*, evaluated under *null models*.

4. Differential node centrality was studied since late 1940s as reported in Freeman (1979); de Solla Price (1976) elaborated the “rich get richer" phenomenon for degrees which later was popularized (‘scalefree’) by Barabási and Albert (1999).

5. Differential node centrality was studied since late 1940s as reported in Freeman (1979); de Solla Price (1976) elaborated the “rich get richer" phenomenon for degrees which later was popularized (‘scalefree’) by Barabási and Albert (1999).

   Feld and Elmore (1982) noted that differential degrees may lead to many transitive triads.

6. Differential node centrality was studied since late 1940s as reported in Freeman (1979); de Solla Price (1976) elaborated the “rich get richer" phenomenon for degrees which later was popularized (‘scalefree’) by Barabási and Albert (1999).

   Feld and Elmore (1982) noted that differential degrees may lead to many transitive triads.

Two different structural effects: this requires dependencies represented in estimated model, not only in statistics tested in a null model.

Of course this is the usual flexibility
that we know of multivariate regression
and generalized linear models

using a linear predictor

$$\sum_k \beta_k X_k$$

Of course this is the usual flexibility
that we know of multivariate regression
and generalized linear models

using a linear predictor

$$\sum_k \beta_k X_k$$

This allows researchers
to model several theories simultaneously.

(Cf. control variables – simple version of this.)

Treasures

A great step was made by Frank and Strauss (1986)
who developed Markov graphs,
a model representing transitivity
in a kind of generalized linear model,
the probability of observing a particular network
depending on an expression of the kind

$$\sum_k \beta_k X_k$$

where the 'variables' X_k depend on the network itself :
so-called *endogenous* effects.

A great step was made by Frank and Strauss (1986)
who developed Markov graphs,
a model representing transitivity
in a kind of generalized linear model,
the probability of observing a particular network
depending on an expression of the kind

$$\sum_k \beta_k X_k$$

where the 'variables' X_k depend on the network itself :
so-called *endogenous* effects.

In the linear combination we can include also
exogenous effects: attributes of actors, etc.

Frank (1991) and Wasserman and Pattison (1996)
took the next great step
by generalizing this to “any" endogenous effects
— not only transitivity; ⇒ p^* .

Frank (1991) and Wasserman and Pattison (1996)
took the next great step
by generalizing this to “any" endogenous effects
— not only transitivity; ⇒p^* .

However, there were problems.

These seemed to be of a technical nature,
but they turned out to be problems of the model.

Frank (1991) and Wasserman and Pattison (1996)
took the next great step
by generalizing this to “any" endogenous effects
— not only transitivity; $\Rightarrow p^*$.

However, there were problems.

These seemed to be of a technical nature,
but they turned out to be problems of the model.

On to *probability models*,
then back to *dependencies*.

Statistical theory differentiates between

⇒ sampling-based inference:
probability model founded on collecting data
by a probability sample;

Statistical theory differentiates between

- ⇒ sampling-based inference:
  probability model founded on collecting data by a probability sample;
- ⇒ model-based inference:
  probability model is assumed by modeler;

Statistical theory differentiates between

- ⇒ sampling-based inference:
  probability model founded on collecting data
  by a probability sample;
- ⇒ model-based inference:
  probability model is assumed by modeler;
  'goodness of fit' of assumptions can be checked,
  but always incompletely.

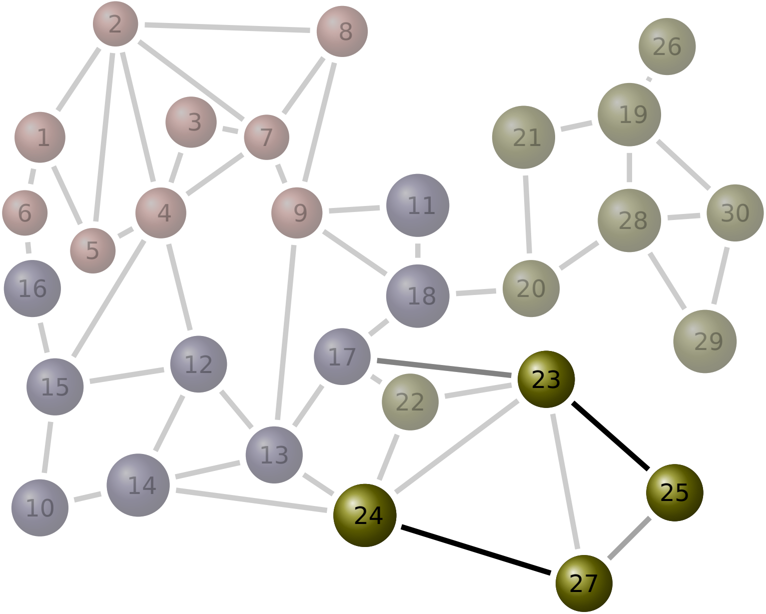
Statistical theory differentiates between

- ⇒ sampling-based inference:
  probability model founded on collecting data
  by a probability sample;
- ⇒ model-based inference:
  probability model is assumed by modeler;
  'goodness of fit' of assumptions can be checked,
  but always incompletely.

Assumptions can be about independence,
conditional independence, distributional shape, etc.

The model of Frank and Strauss, *Markov graph*,
was based on a conditional independence assumption
which they called Markov dependence:

non-adjacent ties are independent,
conditionally on the rest of the graph.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Frank and Strauss proved that this type of conditional independence corresponds to measuring transitivity by counting closed triangles or transitive triplets.

In research by Mark Handcock, Pip Pattison, Garry Robins, and Tom Snijders (2001-2006) (⇒ paper in *Sociological Methodology* 2006) it was found that this model does not fit well to empirical social science network data.

In research by Mark Handcock, Pip Pattison,
Garry Robins, and Tom Snijders (2001-2006)
($\Rightarrow$ paper in *Sociological Methodology* 2006)
it was found that this model does not fit well
to empirical social science network data.

Substantive conclusion about the social world:

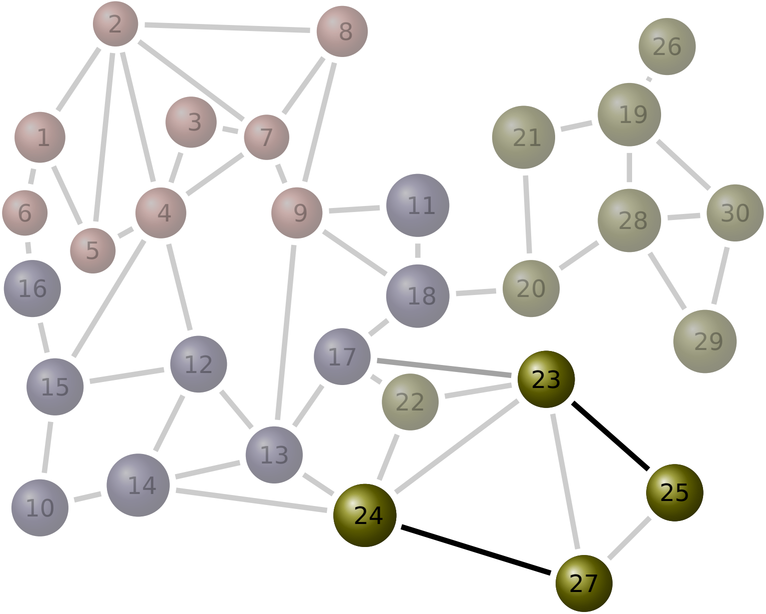
Independence between existence of non-adjacent ties
(i.e., two ties between four distinct nodes)
is not generally plausible;
an extra requirement is necessary:

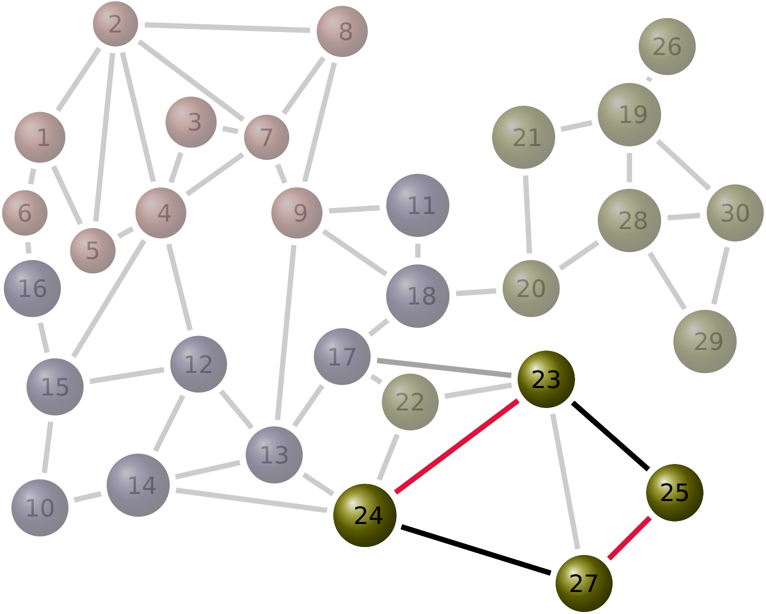
In research by Mark Handcock, Pip Pattison,
Garry Robins, and Tom Snijders (2001-2006)
(⇒ paper in *Sociological Methodology* 2006)
it was found that this model does not fit well
to empirical social science network data.

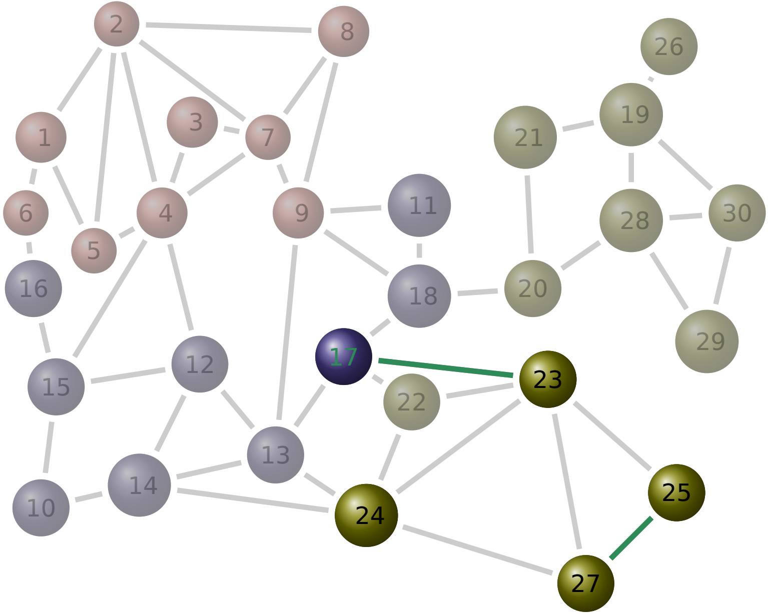
Substantive conclusion about the social world:

Independence between existence of non-adjacent ties
(i.e., two ties between four distinct nodes)
is not generally plausible;
an extra requirement is necessary:

no social circuits.


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30


2
8
26
1
3
7
19
21
6
4
9
11
28
30
5
16
18
20
17
23
29
12
15
22
13
14
25
10
24
27

These developments were part of the rise of the
Exponential Random Graph Model (p^*)
which can

These developments were part of the rise of the
Exponential Random Graph Model (p^*)
which can

⇒ represent many different kinds of dependencies;

These developments were part of the rise of the
Exponential Random Graph Model (p^*)
which can

- ⇒ represent many different kinds of dependencies;
- ⇒ include these dependencies simultaneously,
  also effects of measured (‘exogenous’) variables;

These developments were part of the rise of the
Exponential Random Graph Model (p^*)
which can

- ⇒ represent many different kinds of dependencies;
- ⇒ include these dependencies simultaneously,
  also effects of measured (‘exogenous’) variables;
- ⇒ be estimated using Statnet and pnet,
  yielding also goodness of fit assessments.

How does this help in gauging uncertainty?

All assertions about uncertainty
– standard errors, hypothesis tests,
posterior distributions –
are based on the validity of the probability model
(allowing for some grains of salt).

How does this help in gauging uncertainty?

All assertions about uncertainty
– standard errors, hypothesis tests,
posterior distributions –
are based on the validity of the probability model
(allowing for some grains of salt).

The basic purpose of such assertions is
grappling with the question
(thanking Ivo Molenaar, 1988)

.

What would happen if you did it again?

The probability distributions stand for variability
that would come out when doing the research again:

The probability distributions stand for variability that would come out when doing the research again:

⇒ at a different moment

The probability distributions stand for variability that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different respondents

The probability distributions stand for variability that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different groups or organizations

The probability distributions stand for variability that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different groups or organizations
- ⇒ with different measurement instruments ?

The probability distributions stand for variability that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different groups or organizations
- ⇒ with different measurement instruments ?
- ⇒ by a different researcher?

The probability distributions stand for variability
that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different groups or organizations
- ⇒ with different measurement instruments ?
- ⇒ by a different researcher?

Random terms will stand for sources of variability
left out of your model

The probability distributions stand for variability that would come out when doing the research again:

- ⇒ at a different moment
- ⇒ with different groups or organizations
- ⇒ with different measurement instruments ?
- ⇒ by a different researcher?

Random terms will stand for sources of variability left out of your model –
but also model approximations, etc.

As researchers, in most cases we wish to generalize to wider behavioral and social regularities

(cf. *generalizability theory* as coined by Cronbach)

As researchers, in most cases we wish to generalize to wider behavioral and social regularities

(cf. *generalizability theory* as coined by Cronbach)

and statistically gauged variability provides a *lower bound* to the uncertainty in our results.

As researchers, in most cases we wish to generalize to wider behavioral and social regularities

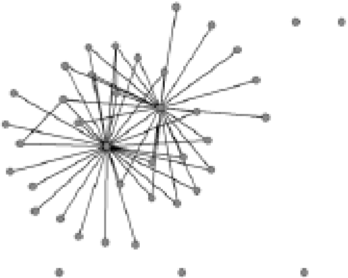
(cf. *generalizability theory* as coined by Cronbach)

and statistically gauged variability provides
a *lower bound* to the uncertainty in our results.

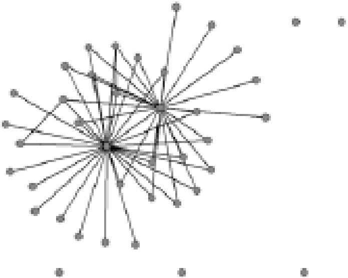
Statistical models for networks are based usually on, instead of independence: *exchangeability* of actors which sometimes
is more reasonable than some other times.

Garry Robins sampled this network from a particular specification of an ERGM.

A priori the nodes are exchangeable;



Garry Robins sampled this network from a particular specification of an ERGM.



A priori the nodes are exchangeable;

estimates of variability for samples from this ERG distribution would themselves be highly variable...

The difficulty of statistical modeling of
single observations (‘cross-sectional data’)
of networks is that
‘everything depends on everything else’
and this dependence must be limited in some way.

The difficulty of statistical modeling of
single observations (‘cross-sectional data’)
of networks is that
‘everything depends on everything else’
and this dependence must be limited in some way.

Network structure is contingent on
attractions and events from the past.

The difficulty of statistical modeling of
single observations (‘cross-sectional data’)
of networks is that
‘everything depends on everything else’
and this dependence must be limited in some way.

Network structure is contingent on
attractions and events from the past.

In modeling longitudinal data we can exploit
the arrow of time :
the present depends on the past, not on the future.

Holland & Leinhardt

Holland & Leinhardt (1977) framework for network dynamics:

Holland & Leinhardt (1977) framework for network dynamics:

1. continuous-time Markov models for panel data
   this allows expressing feedback:
   network builds upon itself;

Holland & Leinhardt (1977) framework
for network dynamics:

1. continuous-time Markov models for panel data
   this allows expressing feedback:
   network builds upon itself;
2. decompose change in smallest constituents,
   i.e., single tie changes.

Holland & Leinhardt (1977) framework
for network dynamics:

1. continuous-time Markov models for panel data
   this allows expressing feedback:
   network builds upon itself;
2. decompose change in smallest constituents,
   i.e., single tie changes.

Elaborated by Stanley Wasserman (1977-1981)
in models representing (separately)
two types of dependence:
reciprocity and popularity.

Holland & Leinhardt (1977) framework extended
to include triadic and other dependencies
by taking an actor-oriented perspective
(Tom Snijders, 1996-2001):

Holland & Leinhardt (1977) framework extended
to include triadic and other dependencies
by taking an actor-oriented perspective
(Tom Snijders, 1996-2001):

in a directed network,
tie changes are modeled as resulting from
actions by nodes = actors to change their outgoing ties;

Holland & Leinhardt (1977) framework extended
to include triadic and other dependencies
by taking an actor-oriented perspective
(Tom Snijders, 1996-2001):

in a directed network,
tie changes are modeled as resulting from
actions by nodes = actors to change their outgoing ties;

probabilities of tie changes depend on
(static!) functions (‘objective’, ‘evaluation’)
representing ‘where the actors tend to go’.

Holland & Leinhardt (1977) framework extended
to include triadic and other dependencies
by taking an actor-oriented perspective
(Tom Snijders, 1996-2001):

in a directed network,
tie changes are modeled as resulting from
actions by nodes = actors to change their outgoing ties;

probabilities of tie changes depend on
(static!) functions ('objective', 'evaluation')
representing 'where the actors tend to go'.

Again, linear predictor to combine theories,
dependencies, measured variables, controls ...

Stochastic actor-oriented models :

Stochastic actor-oriented models :

$\Rightarrow$ process model for network dynamics;

Stochastic actor-oriented models :

- ⇒ process model for network dynamics;
- ⇒ in the classical framework for statistical inference;

Stochastic actor-oriented models :

- ⇒ process model for network dynamics;
- ⇒ in the classical framework for statistical inference;
- ⇒ elaborated for network panel data;

Stochastic actor-oriented models :

- ⇒ process model for network dynamics;
- ⇒ in the classical framework for statistical inference;
- ⇒ elaborated for network panel data;
- ⇒ for dynamic network analysis,
  focusing on agency;

Stochastic actor-oriented models :

⇒ process model for network dynamics;

⇒ in the classical framework for statistical inference;

⇒ elaborated for network panel data;

⇒ for dynamic network analysis,
focusing on agency;

⇒ estimation by Siena, now RSiena .

Some things we have learned

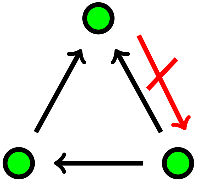
Some things we have learned

- It works!

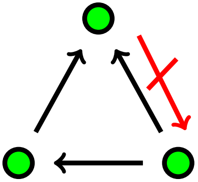
Some things we have learned

- It works!
- In affective relations like friendship,
  reciprocation often has a strength of about 2
  on a logistic scale:
  for making new ties or maintaining existing ties,
  the tie being reciprocated
  increase chances by factor 5-10 ($e^2 = 7.389$).

- Affective relations often have a locally hierarchical nature: transitive triplets yes, 3-cycles no. (See already Davis 1970.)



- Affective relations often have a locally hierarchical nature: transitive triplets yes, 3-cycles no. (See already Davis 1970.)
- Hierarchical structures (in-degrees $\sim$ high status) can be distinguished from locally transitive (clustered) structures.



- For effects of psychological traits on friendship dynamics:
  Big Five:
  openness, extraversion, agreeableness, neuroticism and conscientiousness.

- For effects of psychological traits on friendship dynamics:
  tendencies to homophily on
  openness, extraversion, agreeableness;
  not on neuroticism and conscientiousness.
  (Selfhout/van Zalk et al., *J. Personality*, 2010)

- For effects of psychological traits on friendship dynamics:
  tendencies to homophily on
  openness, extraversion, agreeableness;
  not on neuroticism and conscientiousness.
  (Selfhout/van Zalk et al., *J. Personality*, 2010)

  And, unsurprising perhaps:
  extraverts make more friends;
  agreeable persons get more friends.

Tell me who your friends are

An example of additional detail, multiple theories

An example of additional detail, multiple theories

Homophily well known
(Lazarsfeld & Merton 1954;
McPherson, Smith-Lovin & Cook 2001):

ties more likely between similar actors.

An example of additional detail, multiple theories

Homophily well known
(Lazarsfeld & Merton 1954;
McPherson, Smith-Lovin & Cook 2001):

ties more likely between similar actors.

⇒ I am similar to my friends ;

An example of additional detail, multiple theories

Homophily well known
(Lazarsfeld & Merton 1954;
McPherson, Smith-Lovin & Cook 2001):

ties more likely between similar actors.

$\Rightarrow$ I am similar to my friends ;
$\Rightarrow\Rightarrow$I am similar to friends of my friends

An example of additional detail, multiple theories

Homophily well known
(Lazarsfeld & Merton 1954;
McPherson, Smith-Lovin & Cook 2001):

ties more likely between similar actors.

⇒ I am similar to my friends ;
⇒⇒I am similar to friends of my friends

'homophily at distance 2'.

Various theoretical arguments for distance-2 homophily, e.g.:

Various theoretical arguments for
distance-2 homophily, e.g.:

1. *social identity* : “tell me who your friends are ..."

Various theoretical arguments for distance-2 homophily, e.g.:

1. *social identity* : “tell me who your friends are ..."
2. *uncertainty reduction* :
   “if this person gets along with others like me ..."

Various theoretical arguments for distance-2 homophily, e.g.:

1. *social identity* : “tell me who your friends are ..."
2. *uncertainty reduction* :
   “if this person gets along with others like me ..."
3. *signal unreliability* : if ego’s observation of alter’s attribute is unreliable,
   and ego assumes that homophily operates,
   then dist.-2 similarity suggests direct similarity;

Various theoretical arguments for
distance-2 homophily, e.g.:

1. *social identity* : “tell me who your friends are ..."
2. *uncertainty reduction* :
   “if this person gets along with others like me ..."
3. *signal unreliability* : if ego’s observation of alter’s attribute is unreliable,
   and ego assumes that homophily operates,
   then dist.-2 similarity suggests direct similarity;
4. negative *diversity, social capital* :
   alters bridging to different third actors.

¿ is there a tendency to homophily at distance 2,
while controlling for (regular) homophily ?

¿ is there a tendency to homophily at distance 2,
while controlling for (regular) homophily ?

↬ We also have to control for transitivity.

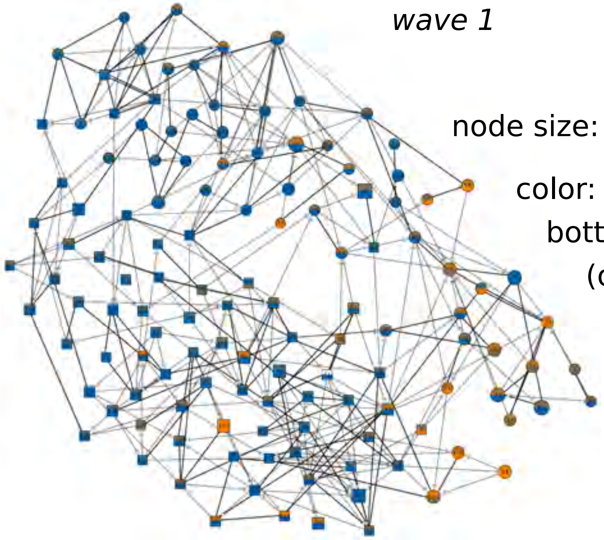
.

Example :

Study of smoking initiation and friendship
(following up on P. West, M. Pearson & others;
cf. Steglich, Snijders & Pearson, *Sociol. Methodology*, 2010).

One school year group from a Scottish secondary school starting at age 12-13 years, monitored over 3 years;
129 (out of 160) pupils present at all 3 observations;
three waves, at appr. 1 year intervals.

Smoking: values 1–3; drinking: values 1–5;
covariates:
gender, smoking of parents and siblings (binary),
money available (range 0–40 pounds/week).

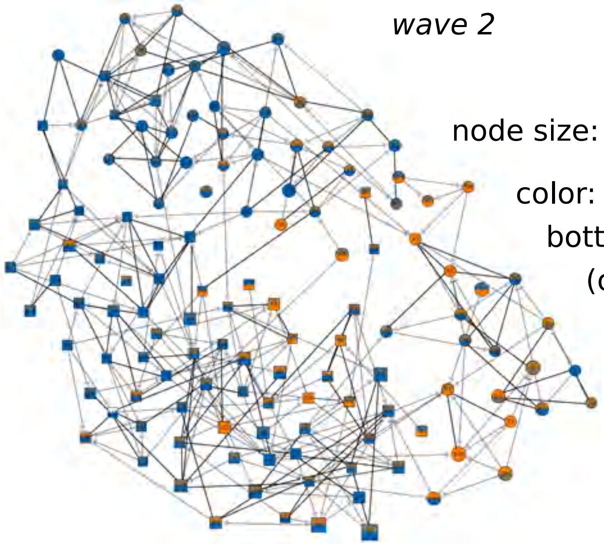


wave 1

girls: circles
boys: squares

node size: pocket money

color: top = drinking
bottom = smoking
(orange = high)

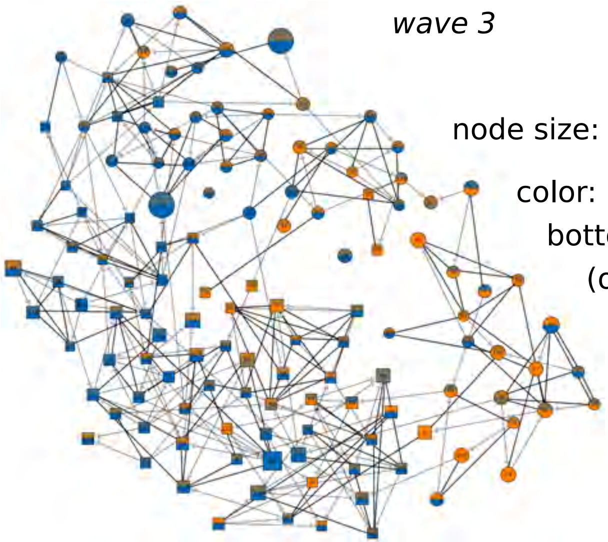


wave 2

girls: circles
boys: squares

node size: pocket money

color: top = drinking
bottom = smoking
(orange = high)



wave 3

girls: circles
boys: squares

node size: pocket money

color: top = drinking
bottom = smoking
(orange = high)

Some descriptives

Average degree varies between 3.5 and 3.7.

Table of tie changes

	$0 \Rightarrow 0$	$0 \Rightarrow 1$	$1 \Rightarrow 0$	$1 \Rightarrow 1$	Jaccard
Per. 1	15,800	235	265	212	$\frac{212}{235+265+212} = .30$
Per. 2	15,838	227	210	237	$\frac{237}{227+210+237} = .35$

Wave 1 $\Rightarrow$ wave 2: 235 ties created, 265 ties dissolved.
Wave 2 $\Rightarrow$ wave 3: 227 ties created, 210 ties dissolved.

Structural effects

		estimate	(s.e.)
1 .	outdegree (density)	−0.76*	(0.38)
2 .	reciprocity	1.94***	(0.12)
3 .	transitive triplets	0.45***	(0.05)
4 .	3-cycles	−0.21**	(0.09)
5 .	transitive ties	0.68***	(0.10)
6 .	indegree − popularity (sqrt)	*n.s.*	
7 .	outdegree − popularity (sqrt)	−0.86***	(0.12)
8 .	outdegree − activity (sqrt)	−0.65***	(0.10)

Attribute effects

		estimate	(s.e.)
9 .	sex alter	.	
10.	sex ego	.	
11.	sex similarity	.	
12.	sex similarity at distance 2	.	
13.	drinking similarity	.	
14.	drinking sim. at distance 2	.	
15.	smoking similarity	.	
16.	smoking sim. at distance 2	.	
17.	money alter	.	
18.	money similarity	.	
19.	money sim. at distance 2	.	

.

Attribute effects

		estimate	(s.e.)
9 .	sex alter	−0.08	(0.10)
10.	sex ego	0.13	(0.12)
11.	sex similarity	0.60***	(0.12)
12.	sex similarity at distance 2	0.64	(0.40)
13.	drinking similarity	0.27°	(0.15)
14.	drinking sim. at distance 2	3.64*	(1.66)
15.	smoking similarity	0.29*	(0.12)
16.	smoking sim. at distance 2	*n.s.*	.
17.	money alter	0.017**	(0.006)
18.	money similarity	1.25***	(0.31)
19.	money sim. at distance 2	*n.s.*	.

Conclusion :

Evidence for distance-2 homophily
for drinking behavior,
weakly for sex,
not for smoking or pocket money.

Other example :

Large-scale study of adolescent development
initiated by Håkan Stattin and Margaret Kerr (Ørebro).

All 12-18 year olds in a small town in Sweden,
yearly waves.

Here: 339 individuals,
cohort of all 13 year olds in given year, 3 waves.

Collaboration with Håkan Stattin, Margaret Kerr,
Bill Burk, Paulina Preciado Lopez.

Some descriptives

Relation: friends & important persons.

Average degree increases from 4.0 to 4.7.

Table of tie changes

	$0 \Rightarrow 0$	$0 \Rightarrow 1$	$1 \Rightarrow 0$	$1 \Rightarrow 1$	Jaccard
Per. 1	105,901	750	542	719	$\dfrac{719}{750+542+719} = .36$
Per. 2	108,450	630	588	888	$\dfrac{888}{630+588+888} = .42$

Wave 1 $\Rightarrow$ wave 2: 750 ties created, 542 ties dissolved.
Wave 2 $\Rightarrow$ wave 3: 630 ties created, 588 ties dissolved.

Structural effects

		estimate	(s.e.)
1 .	outdegree (density)	.	
2 .	reciprocity	.	
3 .	transitive triplets	.	
4 .	3-cycles	.	
5 .	indegree − popularity (sqrt)	.	
6 .	indegree − pop. (sqrt) × time	.	
7 .	outdegree − popularity (sqrt)	.	
8 .	outdegree − activity (sqrt)	.	

.

Structural effects

		estimate	(s.e.)
1 .	outdegree (density)	−2.81 * * *	(0.20)
2 .	reciprocity	2.34 * * *	(0.08)
3 .	transitive triplets	0.61 * * *	(0.02)
4 .	3-cycles	−0.54 * * *	(0.05)
5 .	indegree − popularity (sqrt)	0.11 *	(0.05)
6 .	indegree − pop. (sqrt) × time	−0.09 * *	(0.03)
7 .	outdegree − popularity (sqrt)	−0.92 * * *	(0.06)
8 .	outdegree − activity (sqrt)	−0.29 * * *	(0.05)

.

Settings effects

		estimate	(s.e.)
9 .	log distance	.	
10.	same school	.	
11.	same class	.	
12.	same class × time	.	
13.	same ethnicity	.	

.

Settings effects

		estimate	(s.e.)
9 .	log distance	-0.08^{***}	(0.02)
10.	same school	0.96^{***}	(0.06)
11.	same class	0.56^{***}	(0.05)
12.	same class × time	0.26^{**}	(0.08)
13.	same ethnicity	0.26^{**}	(0.08)

.

Attribute effects

		estimate	(s.e.)
14.	sex alter	.	
15.	sex ego	.	
16.	sex similarity	.	
17.	sex similarity at distance 2	.	
18.	delinquency alter	.	
19.	delinquency ego	.	
20.	delinquency similarity	.	
21.	delinquency sim. distance 2	.	

.

Attribute effects

		estimate	(s.e.)
14.	sex alter	0.07	(0.06)
15.	sex ego	0.08	(0.09)
16.	sex similarity	0.30**	(0.09)
17.	sex similarity at distance 2	1.08***	(0.32)
18.	delinquency alter	0.15***	(0.02)
19.	delinquency ego	0.04	(0.02)
20.	delinquency similarity	0.23*	(0.10)
21.	delinquency sim. distance 2	3.41***	(0.88)

Conclusion :

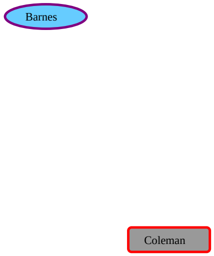
1. Evidence for distance-2 homophily
for sex and delinquent behavior.

Conclusion :

1. Evidence for distance-2 homophily
   for sex and delinquent behavior.
2. More generally:
   statistical modeling can bridge between
   network analysis and mainstream social science.

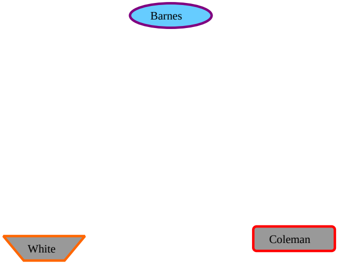
bridging intermezzo

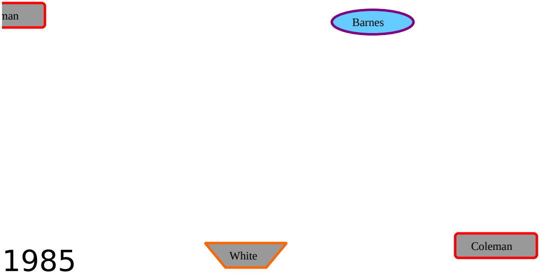
1982

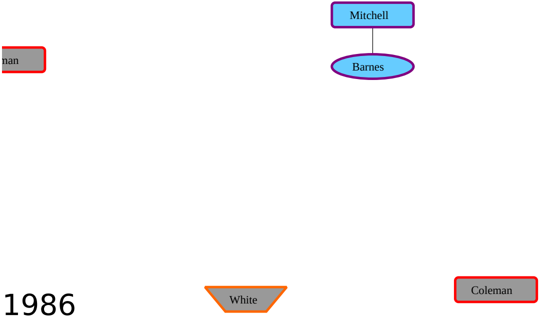


1983

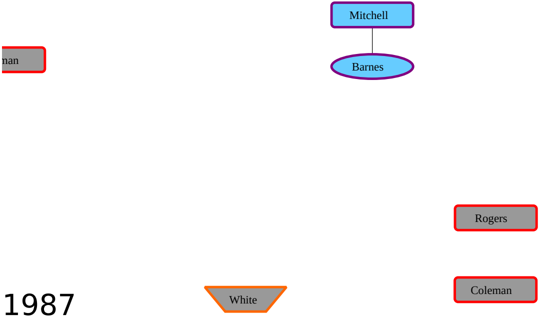
1984



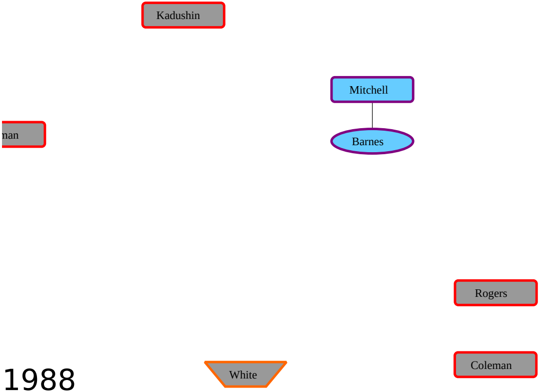

nan
Barnes
1985
White
Coleman



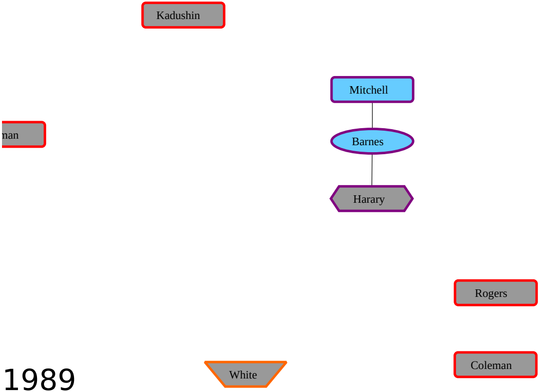
1986



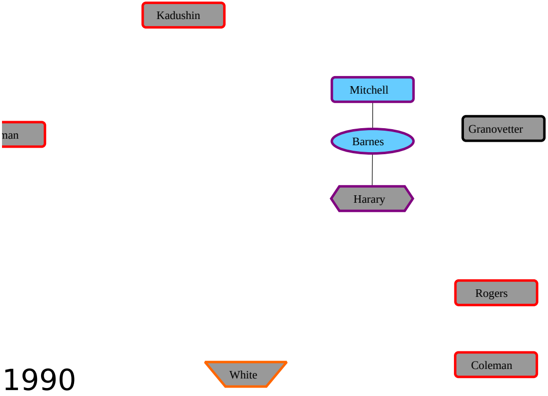
1987

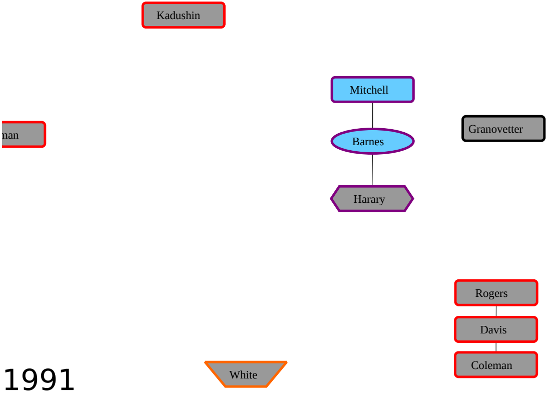


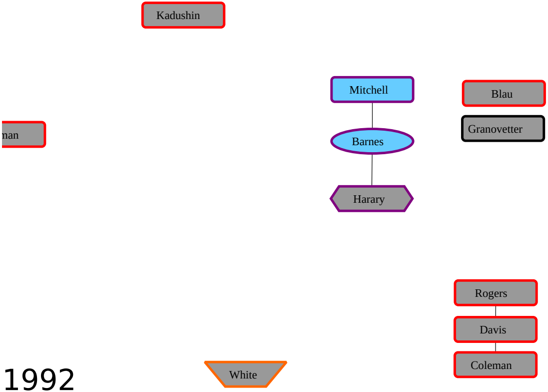
1988



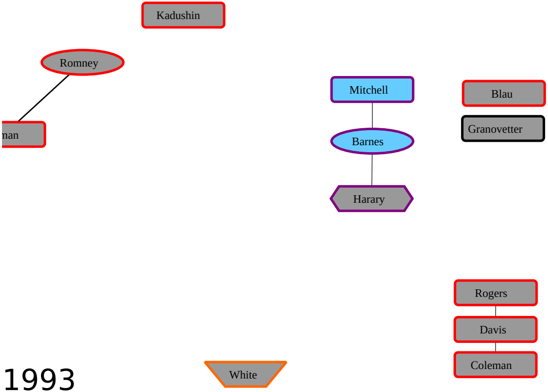
1989


Kadushin
Mitchell
nan
Granovetter
Barnes
Harary
Rogers
Coleman
1990
White

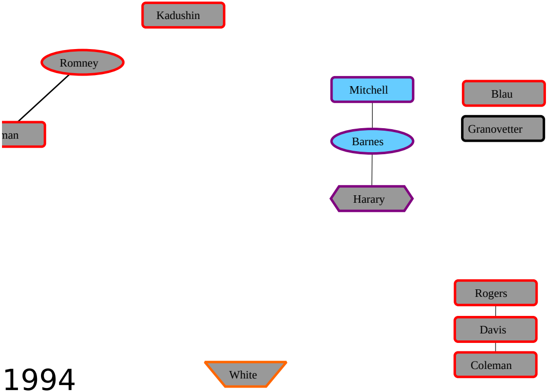

Kadushin
Mitchell
nan
Barnes
Granovetter
Harary
Rogers
Davis
1991
White
Coleman


Kadushin
Mitchell
Blau
Granovetter
nan
Barnes
Harary
Rogers
Davis
Coleman
White

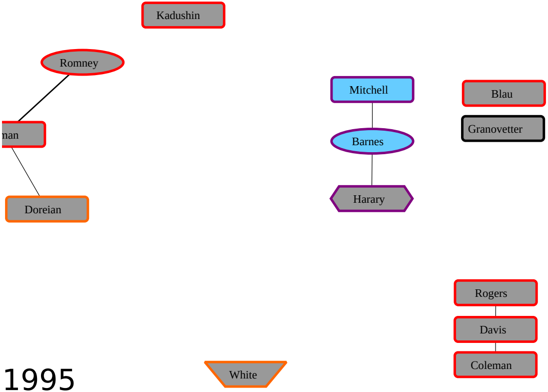
1992


Kadushin
Romney
man
Mitchell
Barnes
Harary
Blau
Granovetter
Rogers
Davis
Coleman
White

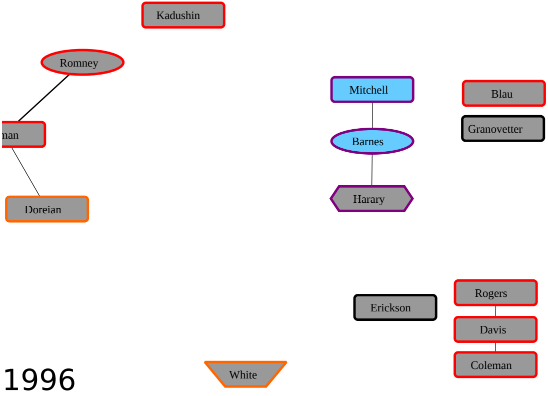
1993

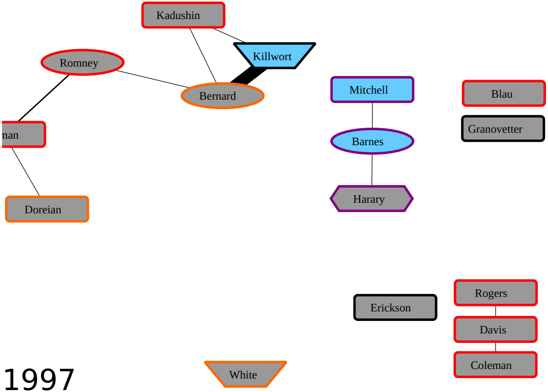

Kadushin
Romney
nan
Mitchell
Barnes
Harary
Blau
Granovetter
Rogers
Davis
Coleman
White

1994

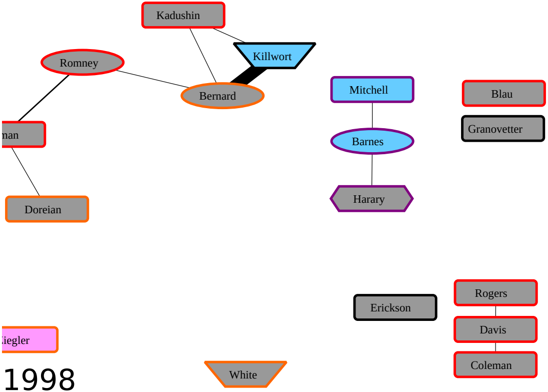

Kadushin
Romney
nan
Doreian
Mitchell
Barnes
Harary
Blau
Granovetter
Rogers
Davis
Coleman
White

1995

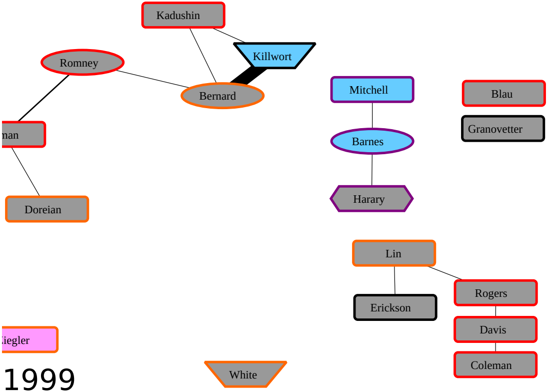

Kadushin
Romney
nan
Doreian
Mitchell
Barnes
Harary
Blau
Granovetter
Erickson
Rogers
Davis
Coleman
1996
White


Kadushin
Killwort
Romney
Bernard
Mitchell
Blau
Granovetter
nan
Barnes
Harary
Doreian
Erickson
Rogers
Davis
Coleman
White

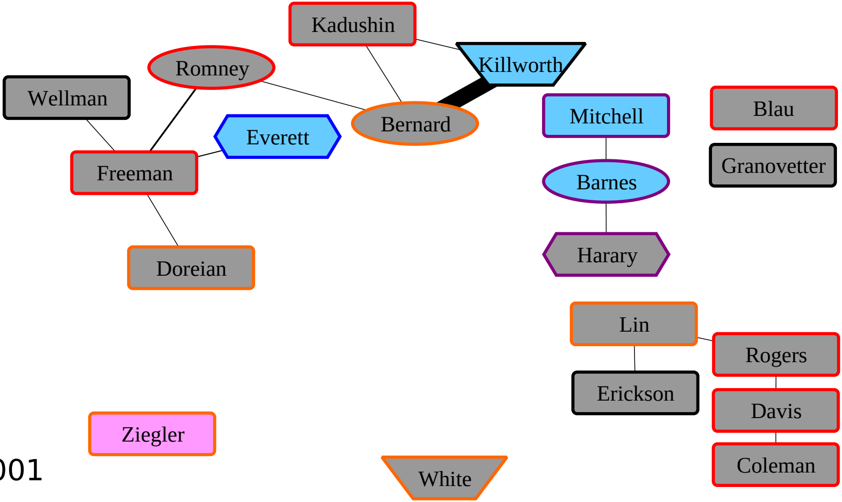
1997


Kadushin
Killwort
Romney
Bernard
Mitchell
Blau
Granovetter
Barnes
nan
Doreian
Harary
Erickson
Rogers
Davis
Coleman
iegler
White

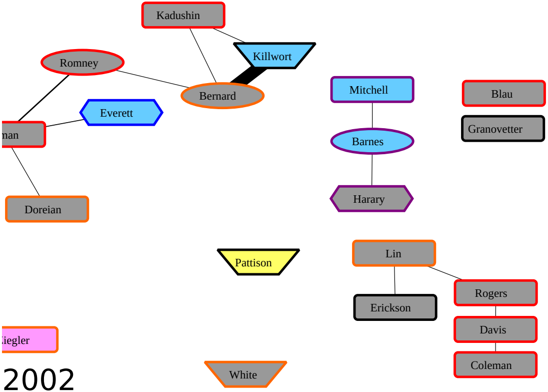
1998


Kadushin
Killwort
Romney
Bernard
Mitchell
Blau
Granovetter
nan
Barnes
Doreian
Harary
Lin
Rogers
Erickson
Davis
iegler
Coleman
White

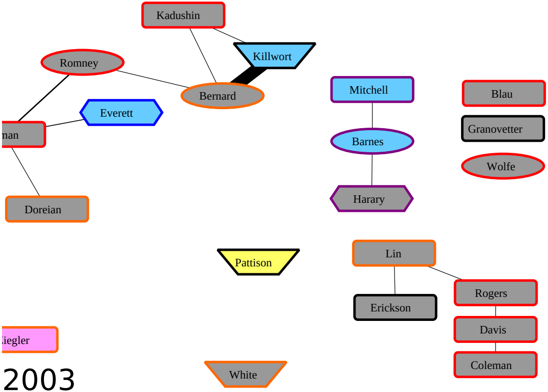
1999


Kadushin
Killworth
Romney
Wellman
Bernard
Mitchell
Blau
Everett
Freeman
Barnes
Granovetter
Harary
Doreian
Lin
Rogers
Erickson
Davis
Ziegler
Coleman
White

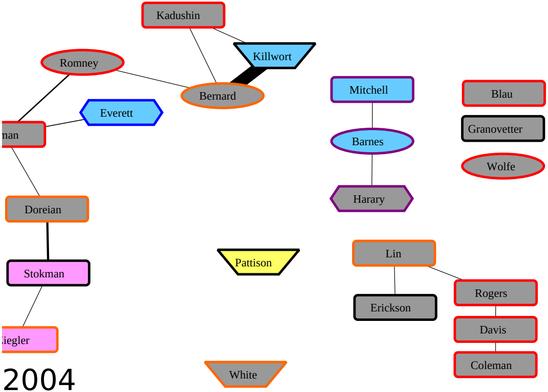
2001


Kadushin
Killwort
Romney
Bernard
Mitchell
Blau
Everett
nan
Granovetter
Barnes
Harary
Doreian
Pattison
Lin
Erickson
Rogers
Davis
iegler
Coleman
White

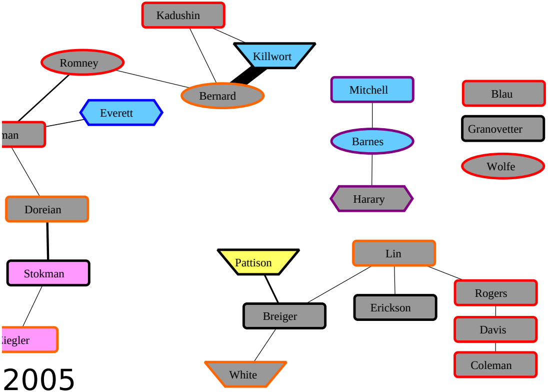
2002


Kadushin
Killwort
Romney
Bernard
Everett
nan
Doreian
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Pattison
Lin
Erickson
Rogers
Davis
Coleman
iegler
White

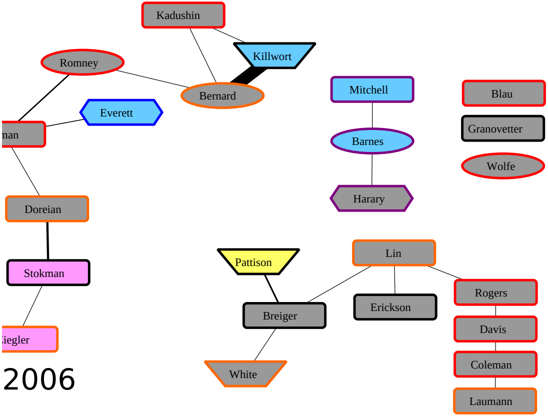
2003


Kadushin
Killwort
Romney
Bernard
Everett
nan
Doreian
Stokman
iegler
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Pattison
White
Lin
Erickson
Rogers
Davis
Coleman

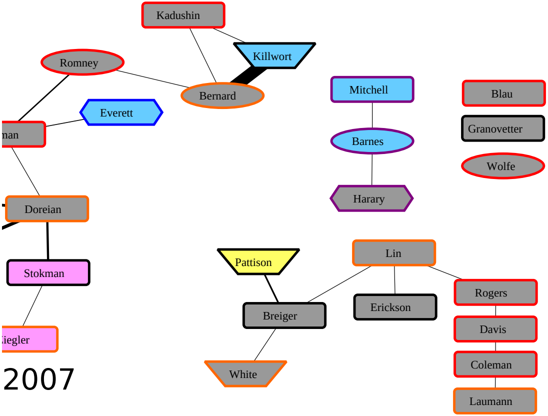
2004

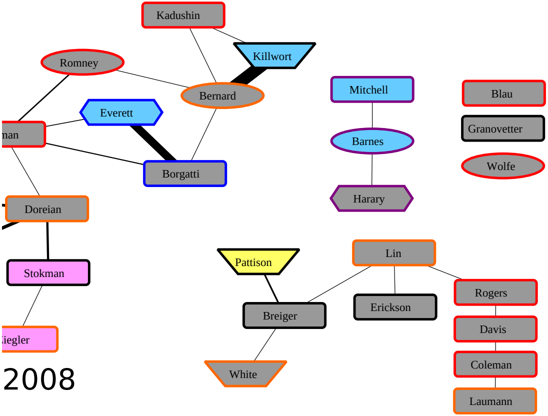

Kadushin
Killwort
Romney
Bernard
Everett
nan
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Doreian
Stokman
iegler
Pattison
Lin
Breiger
Erickson
Rogers
Davis
Coleman
White

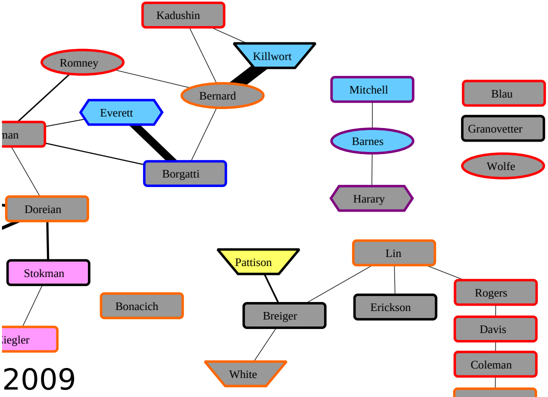
2005

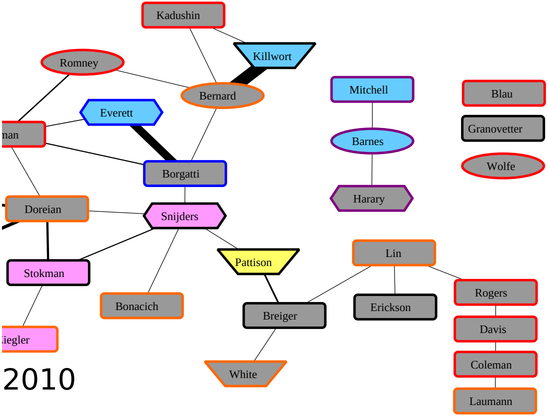

Kadushin
Killwort
Romney
Bernard
Mitchell
Blau
Everett
Granovetter
nan
Barnes
Wolfe
Harary
Doreian
Lin
Pattison
Stokman
Rogers
Erickson
Breiger
Davis
iegler
Coleman
White
Laumann

2006


Kadushin
Killwort
Romney
Bernard
Everett
man
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Doreian
Stokman
iegler
Pattison
Lin
Breiger
Erickson
Rogers
Davis
Coleman
Laumann
White
2007


Kadushin
Killwort
Romney
Bernard
Everett
nan
Borgatti
Doreian
Stokman
iegler
2008
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Pattison
Breiger
White
Lin
Erickson
Rogers
Davis
Coleman
Laumann


Kadushin
Killwort
Romney
Bernard
Everett
nan
Borgatti
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Doreian
Stokman
iegler
Bonacich
Pattison
Breiger
White
Lin
Erickson
Rogers
Davis
Coleman
Laumann
2009


Kadushin
Killwort
Romney
Bernard
Everett
Mitchell
Barnes
Harary
Blau
Granovetter
Wolfe
Borgatti
Doreian
Snijders
Stokman
Bonacich
Pattison
Breiger
White
Lin
Erickson
Rogers
Davis
Coleman
Laumann
2010

back to tensions,
treasures,
tell me who

The actor-oriented framework lends itself nicely for modeling *co-evolution*

network $\Longleftrightarrow$ individual outcomes

The actor-oriented framework lends itself nicely for modeling *co-evolution*

network $\Longleftrightarrow$ individual outcomes

Networks are dependent
as well as independent variables.

Social capital theory
here clearly argues the connection:

The actor-oriented framework lends itself nicely for modeling *co-evolution*

network $\Longleftrightarrow$ individual outcomes

Networks are dependent
as well as independent variables.

Social capital theory
here clearly argues the connection:

$\Rightarrow$ Our social network has an effect on us;

The actor-oriented framework lends itself nicely for modeling *co-evolution*

network $\Longleftrightarrow$ individual outcomes

Networks are dependent
as well as independent variables.

Social capital theory
here clearly argues the connection:

- $\Rightarrow$ Our social network has an effect on us;
- $\Rightarrow$ therefore we try to 'improve' our network.

Investigating network effects on individual outcomes now become fashionable as the study of

peer effects .

Investigating network effects on individual outcomes now become fashionable as the study of

peer effects .

Much non-network literature has an atomistic prejudice, regarding endogenous network choice as a nuisance, which best is being worked out of the model;
and using any ‘peer’ delineation that is available.

Investigating network effects on individual outcomes now become fashionable as the study of

peer effects .

Much non-network literature has an atomistic prejudice, regarding endogenous network choice as a nuisance, which best is being worked out of the model;
and using any ‘peer’ delineation that is available.

However –
the choice of our social neighborhood
is the essence of social science.

The question is not

The question is not

¿how much is added to R^2 by including network variables ?

The question is not

¿how much is added to R^2 by including network variables ?

but

¿what choice do actors have
in selecting their social surroundings,
and how does this affect their chances and outcomes ?

Christian Steglich, Tom Snijders, Mike Pearson,

Dynamic Networks and Behavior:

Separating Selection from Influence

(*Sociological Methodology, 2010*)

http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x

discusses methods for studying peer effects

Christian Steglich, Tom Snijders, Mike Pearson,

Dynamic Networks and Behavior:

Separating Selection from Influence

(*Sociological Methodology, 2010*)

http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x

discusses methods for studying peer effects

- from an actor perspective

Christian Steglich, Tom Snijders, Mike Pearson,
Dynamic Networks and Behavior:
Separating Selection from Influence
(*Sociological Methodology, 2010*)
http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x

discusses methods for studying peer effects

- from an actor perspective
- based on explicitly modeling the feedback

$$\text{network} \Longleftrightarrow \text{individual outcomes}$$

Christian Steglich, Tom Snijders, Mike Pearson,
Dynamic Networks and Behavior:
Separating Selection from Influence
(*Sociological Methodology, 2010*)
http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x

discusses methods for studying peer effects

- from an actor perspective
- based on explicitly modeling the feedback
  $$\text{network} \Longleftrightarrow \text{individual outcomes}$$
- incorporating the network as an interesting phenomenon rather than a nuisance.

'Selection and influence'

Co-evolution of networks and behavior:

model the mutually dependent dynamics of
networks ($\sim$ selection)
and behavior ($\sim$ influence)

'Selection and influence'

Co-evolution of networks and behavior:

model the mutually dependent dynamics of
networks ($\sim$ selection)
and behavior ($\sim$ influence)

as a way to studying peer effects
while simultaneously focusing on peer selection.

.

The advantage:

The advantage: more detail .

The advantage: more detail .

The difficulty:

The advantage: more detail .

The difficulty: dependence on the model.

Example, also from the Ørebro network studies:

See:

Burk et al., *Int. J. Behav. Development*, 2007;

Burk et al. *Revue Française de Sociologie*, 2008.

445 students initially aged 9–14 years (ave. 10.6):
all grade-4 pupils with complete data & their nominees;
5 waves, intervals 1 year.

Example, also from the Ørebro network studies:

See:

Burk et al., *Int. J. Behav. Development*, 2007;

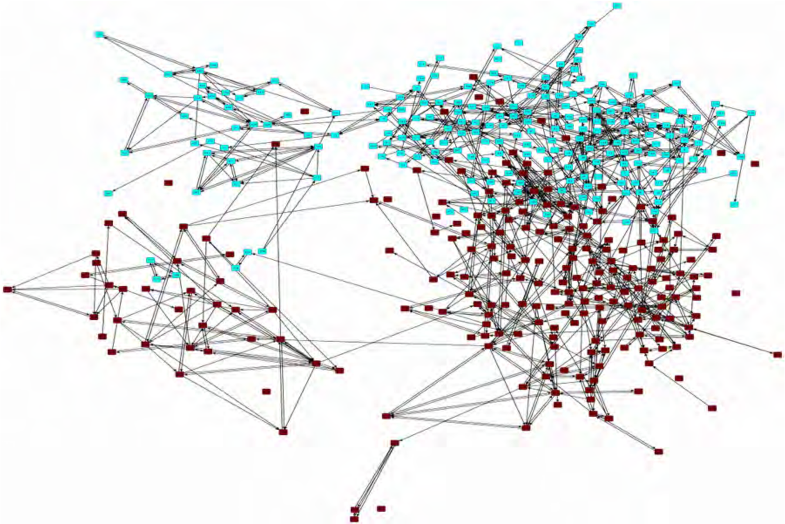
Burk et al. *Revue Française de Sociologie*, 2008.

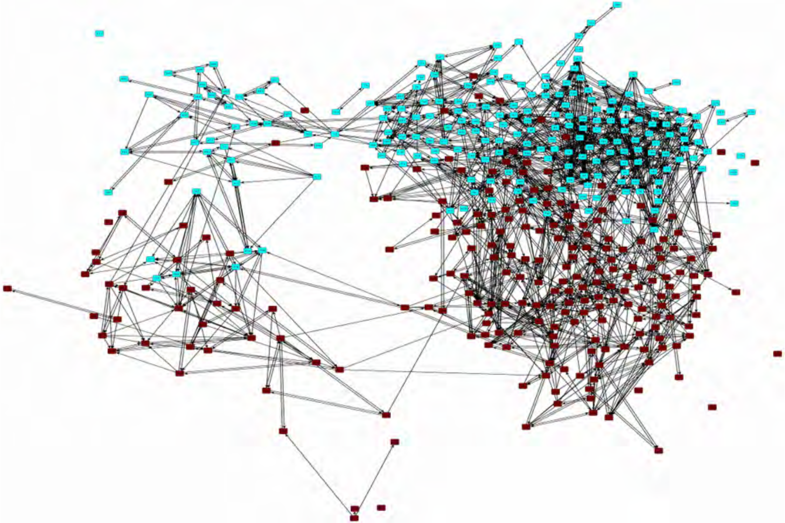
445 students initially aged 9–14 years (ave. 10.6):
all grade-4 pupils with complete data & their nominees;
5 waves, intervals 1 year.

Measures:
peer contacts (talk, hang out, do things);
delinquency; school involvement.

Descriptives

Wave	1	2	3	4	5
Mean degree	2.4	2.6	3.0	3.8	4.1
Reciprocity	0.42	0.46	0.45	0.49	0.49
Transitivity	0.31	0.38	0.37	0.45	0.46
Delinquency (1–3)	1.05	1.07	1.12	1.19	1.28
School inv. (1–5)	4.66	3.95	3.79	3.62	3.50





Friendship model

Outdegree	–2.73***	(0.23)
Reciprocity	2.59***	(0.07)
Transitive triplets	0.08***	(0.02)
Schoolmates	1.01***	(0.13)
Classmates	0.26	(0.45)
Age alter	0.24***	(0.03)
Age ego	–0.08*	(0.04)
Age similarity	1.94***	(0.18)
Sex (F) alter	–0.11	(0.06)
Sex ego	0.21*	(0.08)
Sex similarity	0.80***	(0.07)

Friendship model (continued)

Delinquency alter	0.15∗∗∗	(0.04)
Delinquency ego	–0.04	(0.13)
Delinquency similarity	1.55∗∗∗	(0.44)
School inv. alter	0.04	(0.02)
School inv. ego	0.00	(0.02)
School inv. similarity	0.35∗	(0.14)

Delinquency dynamics

Similarity to friends	1.444*	(0.643)
Age	0.152*	(0.075)
Sex (F)	–0.212*	(0.091)
School involvement	–0.008***	(0.001)

Delinquency dynamics

Similarity to friends	1.444*	(0.643)
Age	0.152*	(0.075)
Sex (F)	–0.212*	(0.091)
School involvement	–0.008***	(0.001)

School involvement dynamics

Similarity to friends	0.300	(0.189)
Age	0.007	(0.034)
Sex (F)	0.062	(0.045)
Delinquency	–0.064*	(0.030)

Conclusions

Friendship is influenced
by similarity on delinquent behavior
and (less strongly) on school involvement;

Conclusions

Friendship is influenced
by similarity on delinquent behavior
and (less strongly) on school involvement;
more delinquent pupils are slightly more popular;

Conclusions

Friendship is influenced
by similarity on delinquent behavior
and (less strongly) on school involvement;
more delinquent pupils are slightly more popular;
evidence for peer influence w.r.t. delinquency,
not w.r.t. school involvement;

Conclusions

Friendship is influenced
by similarity on delinquent behavior
and (less strongly) on school involvement;
more delinquent pupils are slightly more popular;
evidence for peer influence w.r.t. delinquency,
not w.r.t. school involvement;

mutually negative influence (but not strong)
between school involvement and delinquency.

Network delineation issues

Network delineation issues

1. What is
   – for a given relation and a given behavior –
   the most relevant set of ‘ties’ that influences us?

Network delineation issues

1. What is
   – for a given relation and a given behavior –
   the most relevant set of ‘ties’ that influences us?
   E.g.:
   direct ties; reciprocated ties; Simmelian ties;
   structurally equivalent others; larger setting …

Network delineation issues

1. What is
   – for a given relation and a given behavior –
   the most relevant set of ‘ties’ that influences us?
   E.g.:
   direct ties; reciprocated ties; Simmelian ties;
   structurally equivalent others; larger setting ...
2. How are the results of these statistical models
   affected by network delineation?

Other statistical models ...

This talk has attempted to provide some elucidation and perspective for *some* statistical network models

but there are many more —

network sampling: work by Frank, Thompson, Handcock, Gile, Heckathorn, Salganik, etc.;

latent variable models;

see book by Eric Kolaczyk;

overview article Goldenberg, Zheng, Fienberg, Airoldi *Foundations and Trends in Machine Learning* (2010);

special issue *Annals of Applied Statistics* (soon).

The road goes ever on and on

We are only beginning to scratch the surface.

We are only beginning to scratch the surface.

research problems lead to new methods

We are only beginning to scratch the surface.

new methods lead to new problems

We are only beginning to scratch the surface.

new problems lead to new methods

We are only beginning to scratch the surface.

new methods lead to new problems

We are only beginning to scratch the surface.

new problems lead to new methods

We are only beginning to scratch the surface.

new methods lead to new problems

Modeling and statistical issues:

Modeling and statistical issues:

1. Multilevel network studies

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence
3. goodness of fit

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence
3. goodness of fit
4. robustness for lack of fit

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence
3. goodness of fit
4. robustness for lack of fit
5. constructing models suitable for larger networks: social settings

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence
3. goodness of fit
4. robustness for lack of fit
5. constructing models suitable for larger networks: social settings
6. emergent properties

$\Rightarrow$ we are working with $N = \sharp$(networks) !

Modeling and statistical issues:

1. Multilevel network studies
2. delineating the network relevant for influence
3. goodness of fit
4. robustness for lack of fit
5. constructing models suitable for larger networks: social settings
6. emergent properties

   $\Rightarrow$ we are working with $N = \sharp(\text{networks})$!
7. mathematical proofs

but my main concerns are ....

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?

.

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   in view of what is interesting –

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – new methods lead to new insights

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – new insights lead to new research questions

.

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – new questions lead to new methods

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – in view of substantive interest

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – in view of substantive interest
   – and in view of goodness of fit
   & robustness to deviations from assumptions.

but my main concerns are ....

1. Risks of mechanistic use of statistical methods.
   Researchers should not jump to
   statistical network modeling while ignoring
   traditional SNA approaches;
   a combination of both is required.
2. How complicated does the model need to be ?
   – in view of substantive interest
   – and in view of goodness of fit
   & robustness to deviations from assumptions.
3. construction, documentation,
   dissemination of software.

... collaborators ....

Christian Steglich, Johan Koskinen,
Marijtje van Duijn, Mark Huisman, Michael Schweinberger,
Pip Pattison, Garry Robins, Mark Handcock,
with runners up Josh Lospinoso, Paulina Preciado Lopez,
Viviana Amati,
for methods development & & ;
John Light, Ruth Ripley, Krists Boitmanis,
for RSiena development & & ,

NWO, NIH for funding of software development;

.

... more collaborators ....

Gerhard van de Bunt, Evelien Zeggelink, Roger Leenders, Frans Stokman, Rafael Wittek, Siegwart Lindenberg, Andrea Knecht, Michael Pearson, John Light, John Scholz, Ainhoa de Federico de la Rúa, Chris Baerveldt, Emmanuel Lazega, Lise Mounier, Alessandro Lomi, Francesca Pallotti, Birgit Pauksztat, Mark Pickup, Paola Tubaro, Vanina Torlo, Bill Burk, Liesbeth Mercken, Isidro Maya Jariego, Håkan Stattin, Margaret Kerr, Maarten van Zalk, Filip Agneessens, Maurits de Klepper, Ulrik Brandes, Jürgen Lerner, Miranda Lubbers, José Luis Molina, Jo Halliday, Laurence Moore, ...

and others, for substantive discussions, theory, applications,