Social Influence and Actor Heterogeneity

Tom A.B. Snijders

University of Oxford
University of Groningen

Sunbelt 2012

My voice and throat are not well.

But ... you can read!

That even goes quicker than how I could talk about it.

My pantomime skills are limited.
Let me try to guide you silently through the slides.

Coevolution of Networks and Behavior

The stochastic actor-oriented model was elaborated to study the co-evolution of networks and behavior (Steglich, Snijders & Pearson, *Soc. Meth., 2010*).

This is a methodology with the purpose of estimating and testing social influence in a dynamic setting, while controlling for homophilous and other behavior-dependent selection of network partners.

Note: social influence is understood here as influence of network ties & position on individual behavior and performance.

What are the threats to such inferences?

To what extent can such results be causally interpreted?
(and alternative explanations excluded!)

This talk has two parts:

This talk has two parts:

1. Some brief thoughts about causality.

This talk has two parts:

1. Some brief thoughts about causality.
2. Proof of concept of ‘fixed effect estimator’, which provides a bit of protection against alternative explanations.

1. Causality

To position the discussion,
recall the distinction between *experimental*
and *observational* studies:

1. Causality

To position the discussion,
recall the distinction between *experimental*
and *observational* studies:

in experimental studies,
the main ‘independent’ variables
are under control of the researcher;

in observational studies,
the main ‘independent’ variables are observed,
without control, in the setting of the data collection.

Methods and models for causal inference tend to focus on experimental studies as the ideal design:
'*no [evidence for] causation without experimentation*'.

The counterfactual model
developed by Paul Holland and Donald Rubin
('what would have happened
if the treatment had been different?')
is the most well-known and most fruitful approach here.

Methods and models for causal inference tend to focus on experimental studies as the ideal design:
‘*no [evidence for] causation without experimentation*’.

The counterfactual model
developed by Paul Holland and Donald Rubin
(‘what would have happened
if the treatment had been different?’)
is the most well-known and most fruitful approach here.

Studies about causal inference in observational studies
tend to focus on methods that attempt to
exclude alternative explanations,
using experimental studies as the ideal reference point.

However,
for questions about social influence in networks,
a quasi-experimental approach risks missing the point:

However,
for questions about social influence in networks,
a quasi-experimental approach risks missing the point:

the ‘choices’ by social actors of their network positions
and their behaviors
are entwined in an inseparable process
of how the actors cope with their social environment
and try to make the best of it, or at least to get by.

However,
for questions about social influence in networks,
a quasi-experimental approach risks missing the point:

the ‘choices’ by social actors of their network positions
and their behaviors
are entwined in an inseparable process
of how the actors cope with their social environment
and try to make the best of it, or at least to get by.

Shalizi & Thomas (*SMR* 2011):
‘disentangling’ influence and selection
cannot be done without depending on model assumptions.

Provisional conclusion:

[1] economists wish to see proof of this

Provisional conclusion:

⇒ The paradigm that idealizes experiments can be helpful for making the point that social influence exists [1], but is of limited importance for finding out the finer structure of how social influence operates.

⇒ It is important to check model assumptions; but some assumptions may be non-testable within the framework of the current data and model.

[1] economists wish to see proof of this

Provisional conclusion:

⇒ The paradigm that idealizes experiments can be helpful for making the point that social influence exists [1], but is of limited importance for finding out the finer structure of how social influence operates.

⇒ It is important to check model assumptions; but some assumptions may be non-testable within the framework of the current data and model.

But then –
how to make inferential progress about causation?

[1] economists wish to see proof of this

Sir Ronald Fisher: Make your theories elaborate.

Sir Ronald Fisher: Make your theories elaborate.

Sir David Cox (*JRSS-A* 1992): important is
“an explicit notion
of an underlying process or understanding
at an observational level that is deeper than
that involved in the data under immediate analysis”.

Goldthorpe (*ESR* 2001) discusses that causality can be approached in different ways:

1. robust dependence;
   (across situations; not disappearing
   when controlling for alternative explanations)

Goldthorpe (*ESR* 2001) discusses that causality can be approached in different ways:

1. robust dependence;
   (across situations; not disappearing
   when controlling for alternative explanations)
2. consequence of manipulation;
   (as in experimental research;
   ‘counterfactual’ approach)

Goldthorpe (*ESR* 2001) discusses that causality can be approached in different ways:

1. robust dependence;
   (across situations; not disappearing
   when controlling for alternative explanations)
2. consequence of manipulation;
   (as in experimental research;
   'counterfactual' approach)
3. generative process:
   mechanism more fundamental
   than the observed association.

Conclusion:

Possibilities to moving forward in understanding social influence in networks will have to be based on studying mechanisms at deeper levels:

Conclusion:

Possibilities to moving forward in understanding social influence in networks will have to be based on studying mechanisms at <u>deeper levels</u>:

⇒ more data
(ranging from intervening variables and multivariate networks to focus groups discussing subjective experiences of social influence)

Conclusion:

Possibilities to moving forward in understanding social influence in networks will have to be based on studying mechanisms at <u>deeper levels</u>:

⇒ more data
(ranging from intervening variables and multivariate networks to focus groups discussing subjective experiences of social influence)

⇒ more theory
(why, when, how will actors influence & be influenced?)

Conclusion:

Possibilities to moving forward in understanding social influence in networks will have to be based on studying mechanisms at deeper levels:

⇒ more data
(ranging from intervening variables and multivariate networks to focus groups discussing subjective experiences of social influence)

⇒ more theory
(why, when, how will actors influence & be influenced?)

⇒ fancy ‘causal statistical analysis’ will not help a lot.

2. Unobserved Heterogeneity: Fixed Effect Estimator

But now, let us think nevertheless about
how statistical methods might be able to help.

An important type of deviation from assumptions
in many longitudinal models is
unobserved heterogeneity:

here, this amounts to
unknown differences between the social actors.

Suppose a ‘Siena’ analysis leads to significant evidence for social influence in the sense of
network ties leading to similarity w.r.t. behavior Z.

Suppose a ‘Siena’ analysis leads to significant evidence for social influence in the sense of network ties leading to similarity w.r.t. behavior Z.

The strict interpretation of this is that actors who are tied tend to become or remain similar in their behavior Z, more so than non-tied actors.

This *might* be actual social influence.

Suppose a ‘Siena’ analysis leads to significant evidence for social influence in the sense of network ties leading to similarity w.r.t. behavior Z.

The strict interpretation of this is that actors who are tied tend to become or remain similar in their behavior Z, more so than non-tied actors.

This *might* be actual social influence.

An alternative possibility is that the ties were *first* formed based on an unobserved variable V that *later* leads to development of Z.

For example: friendship formation could be based on earlier homophily w.r.t. $V =$ sensation seeking, that later leads to $Z =$ antisocial behavior.

In analysis of non-network panel data,
there is available the so-called

fixed effects estimator

which permits to control for arbitrary
time-fixed differences between individuals.

In analysis of non-network panel data,
there is available the so-called

fixed effects estimator

which permits to control for arbitrary
time-fixed differences between individuals.

Can something similar be developed
for actor-based models for networks & behavior?

Proof of concept

Further plan of presentation:

1. Small simulation study about sensitivity of ‘regular’ estimator to non-observed heterogeneity.
2. Fixed effects estimators for actor-oriented models.

Proof of concept

Further plan of presentation:

1. Small simulation study about sensitivity of 'regular' estimator to non-observed heterogeneity.
2. Fixed effects estimators for actor-oriented models.

The heterogeneity considered here refers to unobserved differences between actors that do not change over time.

Proof of concept

Further plan of presentation:

1. Small simulation study about sensitivity of ‘regular’ estimator to non-observed heterogeneity.
2. Fixed effects estimators for actor-oriented models.

The heterogeneity considered here
refers to unobserved differences between actors
that do not change over time.

Good conditions for a proof-of-concept study:

1. Many waves ($\Rightarrow$ good estimation of actor effects);
2. Few actors ($\Rightarrow$ limited computing time).

General setup 'simulated reality'

1. Waves $0, 1, \ldots, M = 10$;
   period $t_0 - t_1$ is used to set the stage,
   the analysed waves are $t_1, \ldots, t_M$.
2. Actors $1, \ldots, n = 30$.
3. Dependent variables:
   Network X, Behavior Z with categories $1, 2, 3, 4, 5$
4. Time-constant covariate $V \sim \mathcal{N}(0, 1)$ (unobserved)
5. $X(t_0)$ is random, parameters between t_0 and t_1 include
   homophily of network X w.r.t. V,
   so that $X(t_1)$ has network autocorrelation w.r.t. V.
6. Also later on homophily on V in dynamics of X.
7. V has positive effect on dynamics of Z after t_1.

E.g., X = Friendsh.; Z = Delinq.; V = Sensation seeking.

Network model:

The initial network $X(t_1)$ is generated with a strong V-similarity parameter.

Network dynamics from t_1 to t_{10} is determined by:

1. Rate parameters $\rho_m^X = 2$ (all periods m)
2. Outdegree effect $\beta_{\mathrm{d}}^X = -1.8$
3. Reciprocity effect $\beta_{\mathrm{rec}}^X = 2$
4. Transitive triplets effect $\beta_{\mathrm{tt}}^X = 0.3$
5. 3-cycles effect $\beta_{\mathrm{tc}}^X = -0.3$
6. Z-similarity effect $\beta_Z^X = 0.5$
7. V-similarity effect $\beta_V^X = 1$ in data, not in analysis.

Behavior model:

1. Rate parameters $\rho_m^Z = 1$ (all periods m)
2. Linear tendency effect $\beta_1^Z = 0$
3. Quadratic tendency effect $\beta_2^Z = 0$
4. Average alter effect β_{avalt}^Z, social influence
   (effect of average of my friends behavior on my behavior)
5. V-effect β_V^Z in data, not in analysis
   unobserved heterogeneity.

The parameters varied in the simulations are those representing unobserved heterogeneity:
they are used in 'simulated reality'
but ignored in the data analysis:

1. β^X_{V0} , the homophily parameter on the unobserved variable V before the start of observations.
2. β^Z_V , the effect of the unobserved variable V on Z.

The parameter investigated is

3. the estimated $\hat{\beta}^Z_{\text{avalt}}$, the social influence effect,

which is 0 in 'simulated reality',
but may be estimated as positive because V is unobserved.

What do we expect?

What do we expect?

1. For $\beta_V^Z = 0$ (no unobserved heterogeneity),
   the model is well specified,
   and the test statistic for social influence

$$\frac{\hat{\beta}^Z_{\text{avalt}}}{s.d.\left(\hat{\beta}^Z_{\text{avalt}}\right)}$$

   has approximately a standard normal distribution;

What do we expect?

1. For $\beta^Z_V = 0$ (no unobserved heterogeneity), the model is well specified, and the test statistic for social influence

$$\frac{\hat{\beta}^Z_{\text{avalt}}}{s.d.\left(\hat{\beta}^Z_{\text{avalt}}\right)}$$

has approximately a standard normal distribution;

2. For $\beta^X_{V0} > 0$, i.e., initial homophily on a variable that later leads to higher Z, the test for β^Z_{avalt} is positively biased: rejection rate higher than $\alpha = 0.05$. This is because the model is misspecified.

This is tested in a very small simulation study.

Is the regular estimator/test sensitive for unobserved heterogeneity?

Is the regular estimator/test sensitive for unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^{Z}_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^{X}_{V0}	β^{Z}_{V}	rejection rate
2	2	.54
1	2	.37
2	1	.38

Is the regular estimator/test sensitive for unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	2	.54
1	2	.37
2	1	.38

Conclusion:

Is the regular estimator/test sensitive for unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	2	.54
1	2	.37
2	1	.38

Conclusion: Yes.
(It should have been 0.05.)

Is the regular estimator/test adequate if there is no unobserved heterogeneity?

Is the regular estimator/test adequate if there is no unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of no unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	0	.006
1	0	.003
0	0	.007

Is the regular estimator/test adequate if there is no unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^{Z}_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of no unobserved heterogeneity.

β^{X}_{V0}	β^{Z}_{V}	rejection rate
2	0	.006
1	0	.003
0	0	.007

Conclusion:

Is the regular estimator/test adequate if there is no unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^{Z}_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of no unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	0	.006
1	0	.003
0	0	.007

Conclusion: Yes but conservative.
(It should have been 0.05.)

Potential Solution: Fixed effects estimator

The fixed effects estimator for behavior
is the regular estimator for a model that has
actor-specific effects (dummy variables) for all actors
in the model for behavior.

These actor-specific effects absorb
all time-constant differences between the actors,
so that conclusions about social influence are made
only based on within-actor over-time comparisons,
excluding any information of between-actor comparisons.

This must lead to considerable loss of power,
like always is the case for fixed effects estimators.

The model specification for the estimation model for behavior includes:

1. Rate parameters ρ_m^Z (all periods m)
2. Linear tendency effect β_1^Z
3. Quadratic tendency effect β_2^Z
4. Average alter effect β_{avalt}^Z (social influence)
5. Effects $\beta_{\text{act}(i)}^Z$ for $i = 1, \ldots, n-1$ of dummy variables for actors (control for unobserved heterogeneity).

In linear models, fixed effects estimators can be implemented more easily; here we must work with a model with a very large number of parameters.

For this large number of waves and moderate number of actors, it runs with few problems.

Does the fixed effects estimator give protection against unobserved heterogeneity?

Does the fixed effects estimator give protection against unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	2	.09

Does the fixed effects estimator give protection against unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^{Z}_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^{X}_{V0}	β^{Z}_{V}	rejection rate
2	2	.09

Conclusion:

Does the fixed effects estimator give protection against unobserved heterogeneity?

Rejection rates for the true null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
in case of unobserved heterogeneity.

β^X_{V0}	β^Z_V	rejection rate
2	2	.09

Conclusion: Pretty good but not totally.
(It should have been 0.05.)

Note that the estimation model still is misspecified
because it ignores the continuing homophily w.r.t. V
(part of the network model, not the behavior model).

Does the protection afforded by the fixed effects estimator come at a loss of power?

Does the protection afforded by the fixed effects estimator come at a loss of power?

Rejection rates for the false null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
for the regular estimator and the fixed effects estimator.

β^X_{V0}	β^Z_V	β^Z_{avalt}	rej. r. regular	rej. r. FEE
1	0	1	0.88	0.24

Does the protection afforded by the fixed effects estimator come at a loss of power?

Rejection rates for the false null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
for the regular estimator and the fixed effects estimator.

β^X_{V0}	β^Z_V	β^Z_{avalt}	rej. r. regular	rej. r. FEE
1	0	1	0.88	0.24

Conclusion:

Does the protection afforded by the fixed effects estimator come at a loss of power?

Rejection rates for the false null hypothesis $\beta^Z_{\text{avalt}} = 0$
with $\alpha = 0.05$ (one-sided),
for the regular estimator and the fixed effects estimator.

β^X_{V0}	β^Z_V	β^Z_{avalt}	rej. r. regular	rej. r. FEE
1	0	1	0.88	0.24

Conclusion: Yes.
(0.24 much less than 0.88.)

Provisional conclusions for fixed effects estimator

1. The principle is feasible.

Provisional conclusions for fixed effects estimator

1. The principle is feasible.

2. Like all fixed effect estimators, it gives protection only against specific kinds of unobserved heterogeneity: differences between actors that are constant over time.

Provisional conclusions for fixed effects estimator

1. The principle is feasible.

2. Like all fixed effect estimators, it gives protection only against specific kinds of unobserved heterogeneity: differences between actors that are constant over time.

3. The loss of power seems considerable; and that in the situation where high power for discovering social influence requires quite a lot of data.

Provisional conclusions for fixed effects estimator

1. The principle is feasible.

2. Like all fixed effect estimators, it gives protection only against specific kinds of unobserved heterogeneity: differences between actors that are constant over time.

3. The loss of power seems considerable; and that in the situation where high power for discovering social influence requires quite a lot of data.

More is to follow; but the perspective is not very bright, underlining the necessity of 'understanding at a deeper level' and the limitations of the attempts of statistically fixing the issue.

Yes/No questions are preferred.