

# Goodness of Fit for Multilevel Dynamic Network Models

Tom A.B. Snijders

University of Groningen  
University of Oxford

December, 2023



# Multilevel Longitudinal Network Analysis

Suppose we have a sample indexed by  $j = 1, \dots, N$  from a population of networks on disjoint node sets, where the networks are independent ‘replications’ of each other in the following sense:

# Multilevel Longitudinal Network Analysis

Suppose we have a sample indexed by  $j = 1, \dots, N$  from a population of networks on disjoint node sets, where the networks are independent ‘replications’ of each other in the following sense:

they all are regarded as realizations of processes obeying the same model, but having different parameters  $\theta_1, \dots, \theta_j, \dots, \theta_N$ .

Each disjoint network is called a **group**.

We assume we have network panel data for each group, and wish to analyze these by fitting Stochastic Actor-Oriented Models (‘SAOMs’, **RSiena**).

## Random Coefficient Models

This presentation considers *random coefficient models* for such multilevel longitudinal network models.

The groupwise parameters  $\theta_1, \dots, \theta_j, \dots, \theta_N$  are assumed to be a sample from a multivariate normal distribution.

A Bayesian estimation procedure was developed by Johan H. Koskinen and Tom A. B. Snijders (2023), Multilevel Longitudinal Analysis of Social Networks. *Journal of the Royal Statistical Society, Series A*, 186, 376–400.

It is implemented in function `sienaBayes` in package `multiSiena`.

The presentation is about assessing **goodness of fit** for such models.

## What is the Bayesian approach?

*Bayesian statistics*, distinct from the more usual *frequentist statistics*, postulates a probability distribution not only for the data, but also for the parameters.

The summary is that the data  $X$  have a conditional distribution given the parameters  $\theta$ , which is the statistical model (here: SAOM); the parameters  $\theta$  have a **prior distribution**, reflecting what the researcher knows about them in advance; the statistical analysis leads to the **posterior distribution**, which is the conditional distribution of the parameters given the data.

For many models (including SAOM), a sample from the posterior distribution can be obtained through an MCMC procedure.

## Goodness of fit

A good model for network dynamics should represent the important features of the network in a good way.

E.g.: distribution of indegrees and outdegrees, triad census, perhaps also distribution of geodesic distances.

This was brought forward as a goodness-of-fit criterion for ERGMs by Hunter, Goodreau, and Handcock (*JASA*, 2008), and elaborated for SAOMs by Lospinoso & Snijders (*Methodological Innovations*, 2019).

For non-multilevel SAOMs, implemented in the function **sienaGOF**.

This is done by simulations of ‘auxiliary statistics’.  
The fit is good if the average values  
of the auxiliary statistics over many simulation runs  
do not deviate too much from the values observed in the data.

A Monte Carlo test based on the Mahalanobis distance  
is used to calculate  $p$ -values.

This is a case where you wish the  $p$ -values to be *large* enough!

A **plot** method can be used to diagnose poor fit.

# Multilevel Goodness of Fit

Two things are required to adapt this procedure to multilevel estimation of SAOMs using **sienaBayes**:

- ⇒ Combine across many different networks (with the same structure and specification) ;
- ⇒ adapt to the Bayesian instead of frequentist estimation .



# Multilevel Goodness of Fit

Two things are required to adapt this procedure to multilevel estimation of SAOMs using **sienaBayes**:

- ⇒ Combine across many different networks (with the same structure and specification) – **similar to many periods**;
- ⇒ adapt to the Bayesian instead of frequentist estimation .

# Multilevel Goodness of Fit

Two things are required to adapt this procedure to multilevel estimation of SAOMs using **sienaBayes**:

- ⇒ Combine across many different networks (with the same structure and specification) – **similar to many periods;**
- ⇒ adapt to the Bayesian instead of frequentist estimation – **using the method of**  
posterior predictive checks *PPC*  
**from Bayesian statistics.**

# Basic Idea of Posterior Predictive Checks (PPC)

From the data  $X^{\text{obs}}$  and the prior distribution we obtain the posterior distribution for parameter  $\theta$ .

If  $\theta$  is drawn according to this posterior distribution, we can generate a hypothetical new data set  $X$ .

This is called the *posterior predictive distribution*.

The fit is assessed by checking how well the observed data  $X^{\text{obs}}$  corresponds with the posterior predictive distribution.

This involves double stochasticity: sample  $\theta$ , then sample  $X$ .

This involves double stochasticity: sample  $\theta$ , then sample  $X$ .

Different from regular/frequentist Goodness of Fit,  
where only one parameter vector  $\theta^{\text{estimated}}$  is considered.

Interpretation of posterior predictive distribution:

hypothetical distribution of data sets that might have been observed,  
given the current knowledge about the parameter vector.

## Multilevel version of PPC

Suppose the MCMC procedure yields (after thinning) a sample from the posterior distribution of the parameters for all groups:

$\theta^{(g)}(s)$  is the  $s$ 'th sampled parameter vector for group  $g$ .

The whole set of simulated parameters can be represented as

$$\begin{pmatrix} \theta^{(1)}(1) & \theta^{(2)}(1) & \dots & \theta^{(g)}(1) & \dots & \theta^{(G)}(1) \\ \theta^{(1)}(2) & \theta^{(2)}(2) & \dots & \theta^{(g)}(2) & \dots & \theta^{(G)}(2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta^{(1)}(s) & \theta^{(2)}(s) & \dots & \theta^{(g)}(s) & \dots & \theta^{(G)}(s) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta^{(1)}(S) & \theta^{(2)}(S) & \dots & \theta^{(g)}(S) & \dots & \theta^{(G)}(S) \end{pmatrix}.$$

Rows are samples  $s = 1, 2, \dots, S$ , columns are groups  $g = 1, 2, \dots, G$ .

(This really is a 3-way array because the  $\theta$ 's are vectors.)

For each parameter  $\theta^{(g)}(s)$  we sample, for group  $g$ ,  
a network data set  $\mathbf{x}^{(g)}(s)$ .

This leads to a sequence of simulated multilevel data sets

$$\begin{pmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \dots \\ \mathbf{x}(s) \\ \dots \\ \mathbf{x}(S) \end{pmatrix} = \begin{pmatrix} \mathbf{x}^{(1)}(1) & \mathbf{x}^{(2)}(1) & \dots & \mathbf{x}^{(g)}(1) & \dots & \mathbf{x}^{(G)}(1) \\ \mathbf{x}^{(1)}(2) & \mathbf{x}^{(2)}(2) & \dots & \mathbf{x}^{(g)}(2) & \dots & \mathbf{x}^{(G)}(2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{x}^{(1)}(s) & \mathbf{x}^{(2)}(s) & \dots & \mathbf{x}^{(g)}(s) & \dots & \mathbf{x}^{(G)}(s) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{x}^{(1)}(S) & \mathbf{x}^{(2)}(S) & \dots & \mathbf{x}^{(g)}(S) & \dots & \mathbf{x}^{(G)}(S) \end{pmatrix},$$

Each row is a hypothetical multilevel data set.

Goodness of fit analyses can be done per group (column) and then combined,  
or for the generated entire multilevel samples  $\mathbf{x}(1), \dots, \mathbf{x}(s), \dots, \mathbf{x}(S)$ .

## How do we proceed in practice?

- 1 Estimate the model using **sienaBayes** and assess convergence.
- 2 This yields a sample  $\theta^{(g)}(s)$  ( $s = 1, \dots, S$ ;  $g = 1, \dots, G$ ) of simulated parameters from the posterior.
- 3 For each group  $g$ , use **siena07** with **simOnly=TRUE** to simulate data sets  $\mathbf{x}^{(g)}(s)$  ( $s = 1, \dots, S$ ;  $g = 1, \dots, G$ ).  
This produces **predictive posterior** multilevel data sets  $\mathbf{x}(s)$ ,  $s = 1, \dots, S$ .
- 4 Use function **sienaGOF** (adapted for this purpose) to **check** the fit either for all groups  $g$  separately, or across all groups.



## Essential difference with earlier case of **sienaGOF**

An essential difference with the use of **sienaGOF** for the known case of modeling single longitudinal networks is that for a single network data set, the MoM estimation using **siena07** and the checks using **sienaGOF** both are based on forward simulation;

for a multilevel network data set, the estimation using **sienaBayes** is based on bridge simulation of the sequence of ministeps connecting  $Y(t_1)$  to  $Y(t_2)$ , while the checks using **sienaGOF** are based on forward simulation.

## Example: data Andrea Knecht

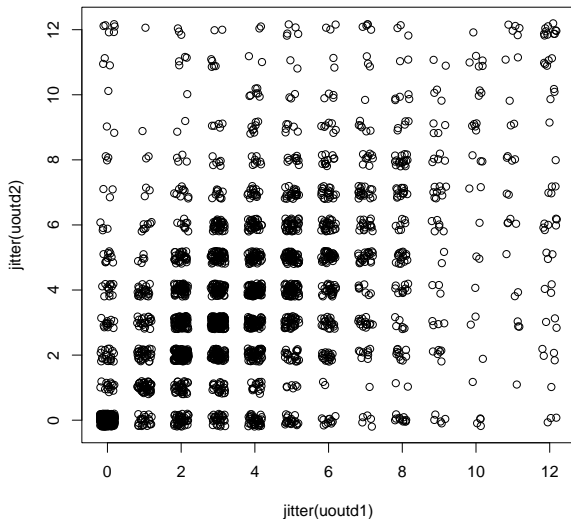
The data set used is about friendship networks in 44 school classes from the study by Andrea Knecht (PhD thesis Utrecht, 2008);

see Knecht, Snijders, Baerveldt, Steglich, & Raub, 'Friendship and Delinquency: Selection and Influence Processes in Early Adolescence', *Social Development*, 2010.

This is a selection from the data set also used in Koskinen & Snijders (2023).

We consider a model for a longitudinal study with 2 waves ('X' and 'Y'), dependent variable friendship.

The students were asked to mention at most 12 friends.



Plot of jittered outdegrees for waves 1 and 2, all 104 groups.

The over-representation of outdegrees equal to 12 leads to problems for GoF.

Therefore it was decided to use only the 44 classes without any outdegrees equal to 12.

I started with the regular model specification.

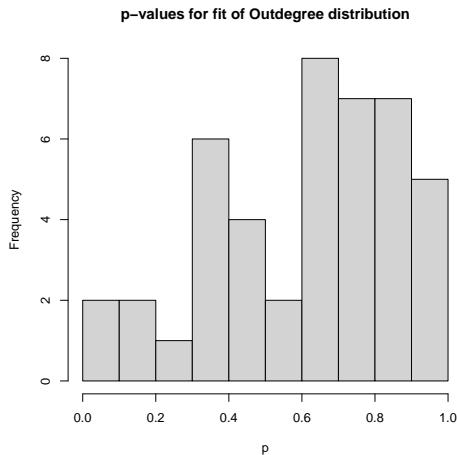
Non-reported GoF exercises led to also including the out-Isolate effect, the outdegree-truncated-to-3 effect, and the  $g_{wespBB}$  effect (which is cyclical).

This led to the model with estimates reported at the following slide.

| Effect                                                       | par.    | (psd)   | betw. sd |
|--------------------------------------------------------------|---------|---------|----------|
| outdegree (density)                                          | -2.640  | (0.285) | 0.651    |
| reciprocity                                                  | 3.135   | (0.181) | 0.656    |
| GWESP I $\rightarrow$ K $\rightarrow$ J                      | 2.057   | (0.093) | 0.289    |
| GWESP I $\leftarrow$ K $\leftarrow$ J                        | -0.124  | (0.048) | .        |
| indegree - popularity ( $\surd$ )                            | -0.510  | (0.065) | 0.181    |
| outdegree - popularity ( $\surd$ )                           | -0.333  | (0.045) | .        |
| outdegree - activity ( $\surd$ )                             | 0.414   | (0.068) | 0.199    |
| reciprocal degree ( $\surd$ ) - activity                     | -0.372  | (0.068) | 0.243    |
| outdegree truncated at 3                                     | 0.712   | (0.122) | 0.423    |
| outdegree higher than 12                                     | -30.000 | (fixed) |          |
| out-isolate                                                  | 1.117   | (0.280) | 0.665    |
| sex alter                                                    | 0.025   | (0.043) | 0.213    |
| sex ego                                                      | -0.017  | (0.041) | 0.193    |
| same sex                                                     | 0.411   | (0.073) | 0.274    |
| reciprocity $\times$ GWESP I $\rightarrow$ K $\rightarrow$ J | -0.929  | (0.110) | 0.317    |

par = posterior mean; psd = posterior standard deviation;  
betw. sd = posterior between-groups stand. deviation.

# Goodness of Fit for Outdegree Distribution



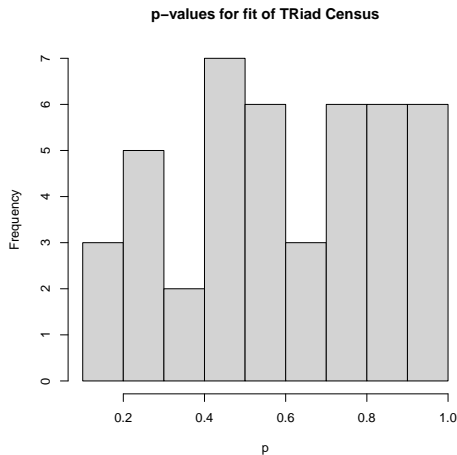
A histogram for the  $p$ -values per group for the fit of the outdegree distribution.

If the fit is good, they should be uniformly distributed, or higher.

The smallest value was 0.029.

Fit per group is excellent.

# Goodness of Fit for Triad Census



A histogram for the  $p$ -values per group for the fit of the triad census.

The smallest value was 0.14.

Fit per group is excellent.

## Aggregated fit across all groups

For the aggregated fit across all groups, frequencies (of counts of degrees equal to  $k$  and of triads) can be added.



## New since EUSN-2023

For estimation by MoM, if estimation converges, the average '*deviations from targets*', averaged over all groups, will be close to 0.

*But for estimation by ML or Bayesian methods this is not the case.*

Worse: I found it very hard to get an acceptable GoF (by the method of posterior predictive checks) for the deviations from targets.

Note: the difference between ML (or, generally, efficient) estimators and other estimators can be used as a test for misspecification.

This is the principle of the *Hausman test*.

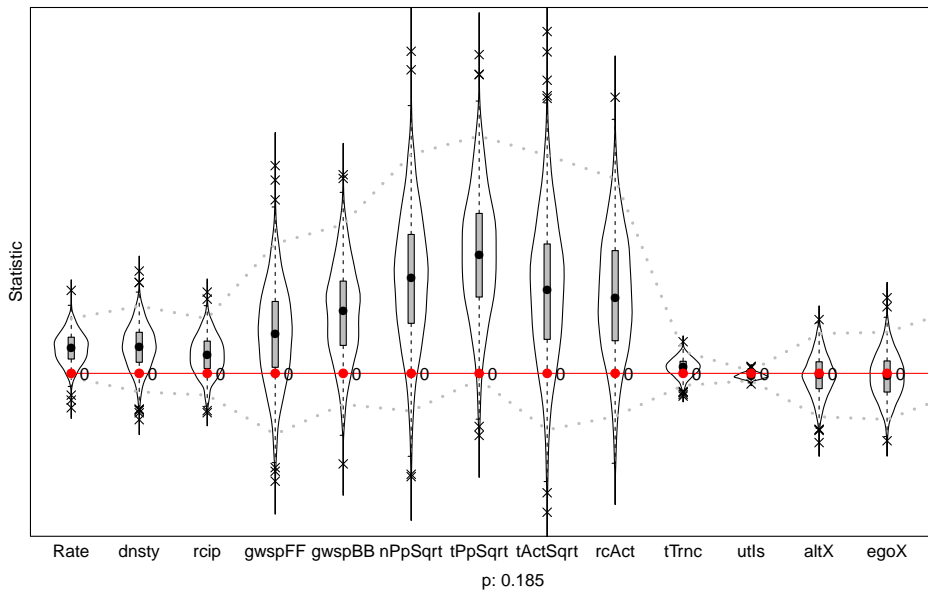
This data set (Andrea Knecht's data, waves  $W$  and  $X$ , only classes with maximum degree less than 12 for both waves) is (up to now) the only one for which I got an acceptable GoF for the deviations from targets.

See next page

shortNames abbreviated...

The auxiliary statistics are here the deviations from targets, divided by the number of actors (since targets are sums across actors).

## PPC GOF average deviations from targets

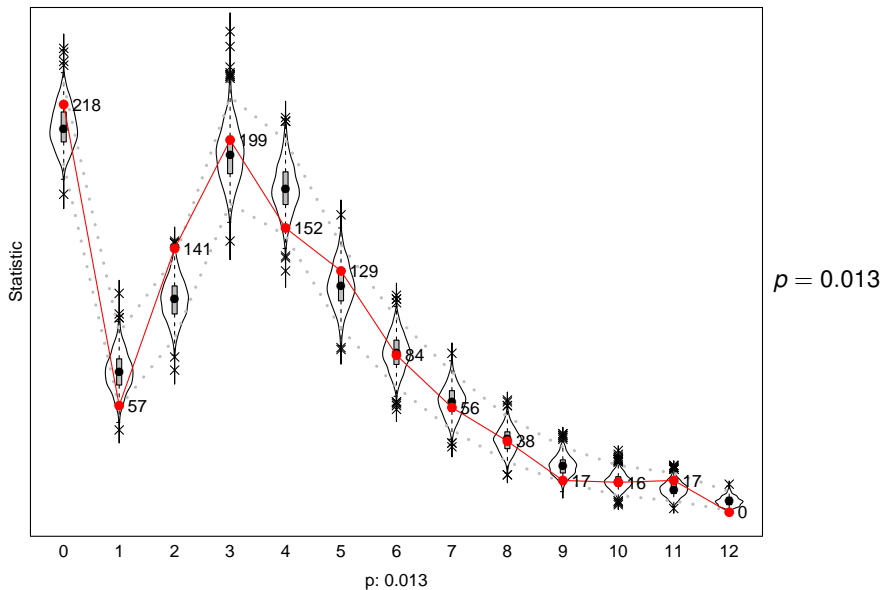


The fit with  $p = 0.185$  is acceptable.

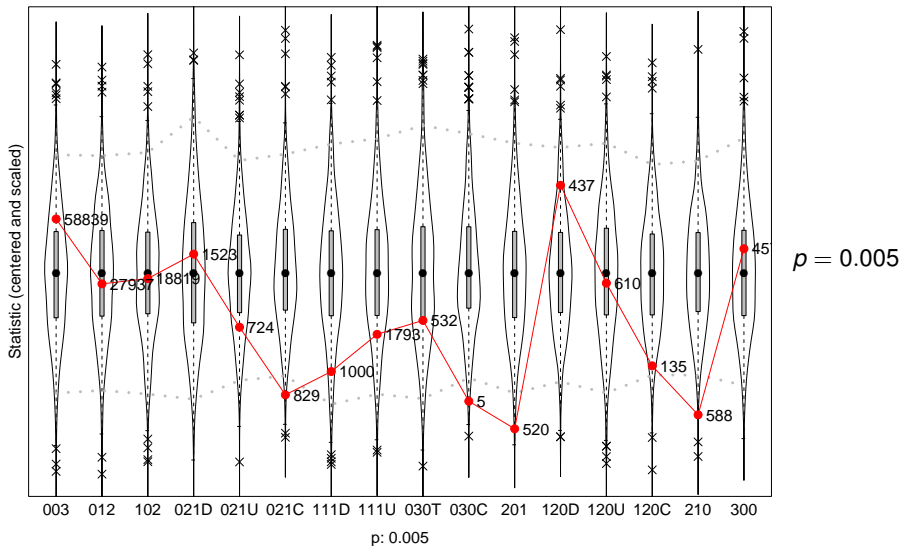
This means that it is relevant  
also to consider other auxiliary statistics.

But look at the violins  
for the number of changes and the average degree:  
these are close to unacceptable.

## Goodness of Fit of OutdegreeDistribution



## Goodness of Fit of TriadCensus



The fit for the Outdegree Distribution and the Triad Census is low – on the boundary of what is acceptable ( $0 < p \leq 0.05$ ).

## Varying group sizes

One issue that plagues multilevel network modeling is the difficulty in comparing parameter values of the ERGM and SAOM across group sizes (number of actors in the networks).

The reason is, briefly, that when the number of actors is larger while further the network evolution is similar, there are more ties that should not be created.

For ERGMs this was shown in research by Krivitsky, Kolaczyk, and Butts.

At the EUSN-2022 I presented about this for the SAOM.

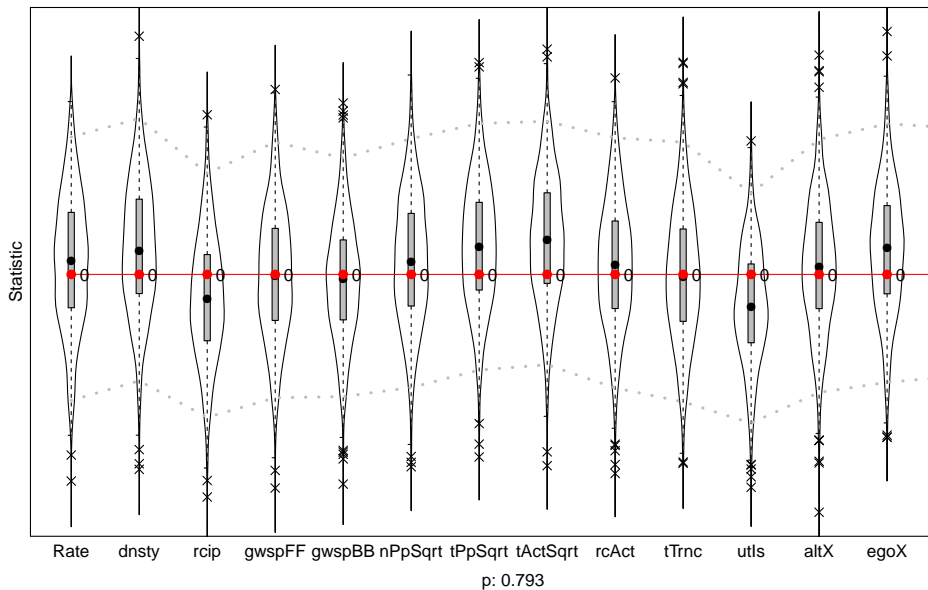
The dependence of the parameter values on group size  $n$  is an unknown function of  $\log(n)$ , in the first place for the density parameter.



The global GoF approach for multilevel networks offers a way to check whether, for a given multilevel data set analyzed by **sienaBayes**, this dependence on group size is a reason for concern.

This can be done by considering the Posterior Predictive Distribution of the **correlation** across groups between  $\log(n)$  and the **deviations between simulated and observed mean degrees and the other estimation statistics for MOM**.

## PPC GOF correlation between av. dev. from targets and log(n)



The correlations of  $\log(n)$  with the deviations for mean degree, and also the other estimation statistics averaged across actors, are nicely distributed about 0.

No reason for concern about varying group sizes.

*My experience up to now, with classroom data sets, has been that it is very hard to get an acceptable PPC fit for the average deviations from targets, but that this is relatively easy for the correlations of av. deviations from targets with  $\log(\text{group size})$ .*

# Discussion

- ⇒ Goodness of Fit extended to multilevel SAOM, following the framework of Posterior Predictive Checks.
- ⇒ Checks per group, and then combined, are more lenient than aggregate checks.
- ⇒ In frequentist Method of Moments estimation, the statistics used for estimation will show perfect fit; up to now, a good fit for these statistics seems very hard to attain for Bayesian estimation.
- ⇒ Checks for correlations of deviation statistics with  $\log(n)$  provide a direct way of checking dependence on group sizes.