University of Oxford

Simulation Studies of Power and Robustness in Models for Network Dynamics

by

Jin Wang

St Catherine's College

A dissertation submitted in partial fulfilment of the degree of Master of Science in Applied Statistics.

Department of Statistics, 1 South Parks Road,
Oxford OX1 3TG

September 2007

This is my own work (except where otherwise indicated)

Candidate: Jin Wang

Signed:.................................

Date:...................................

Abstract

Statistical methods for network dynamics and for co-evolving network and actor behaviours have been proposed recently. Like many other statistical models, they are sensitive to the model assumptions. In this project, some issues about these models' robustness are studied by simulation. We use a true model with known parameters to generate data, and then use a postulated model with different model specifications to estimate parameters based on the generated data. As a baseline situation, the case where the postulated model is the same as the true model will also be studied. Some robustness issues are chosen to design the misspecification, and analysis is based on these issues. Generally the estimates in network dynamics models are good, but in the model for co-evolving network and actor behaviours estimates are not as good as in network dynamics models. The results show that most of the misspecifications affect estimates' precision and significance especially for parameters which are strongly related to the misspecification effect.

Acknowledgements

I would like to take this opportunity to express my thankfulness to all the people that have been helpful in the completion of this project. Special thanks to my supervisor, Prof. Tom Snijders, for his guidance, support and patience. Also I greatly appreciate the love, support and encouragement of my parents through out my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A social network is a set of actors (or points, or nodes, or agents) that may have relationships (or edges, or ties) with one another. For instance, the actors can be individuals and the ties are friendships between them; or the actors can be companies and the ties are alliances between them. Usually, the data structure can be represented by a directed graph.

Many types of models are used to study social networks. Some of them deal with static networks, where neither the set of nodes nor the set of ties varies over time. However in this thesis, we concentrate on network dynamic models for modeling the evolution of social networks. In other words, we are concerned with models dealing with longitudinal network data.

Statistical methods for network dynamics have been developed by Snijders (2001, 2005). Methods for the simultaneous dynamics of networks and individual attributes ('behaviour') of the individuals ('nodes', 'actors') in the networks were proposed by Snijders, Steglich and Schweinberger (2007). These methods using continuous Markov chain models are called 'actor-oriented' models. They assume the network dynamics is, to some extent, driven by the social actors. For example, the actors can control their own outgoing ties and behaviours.

The same as other statistical models, these methods are sensitive to model assumptions. This is currently one of the major practical questions about the application of these methods. For example, in the model for network dynamics, many effects will affect the network dynamics, such as the reciprocation of ties ("since you are my friend, I will become your friend"), transitive closure ("friends of friends become friends") and variables representing actor behaviours (whether the individual is smoking) etc. The model should be specified before used to estimate or simulate, which means we have to decide which effects should be included in this model. In practice, misspecification happens. There are so many effects and some of them may have similar meanings in the model. The choice can hardly be correct all the time. This kind of problems also exists in the models for the simultaneous dynamics of network and individual behaviours, where both the evolution of network and individual behaviours are analyzed. The number of choices of effects then is even larger.

The aim of this project is to study the degree of sensitivity to model assumptions, or expressed alternatively, the robustness to misspecifications. This will be done by simulations. The data are simulated for the dynamics of networks or of networks and attributes. These simulated data sets then are analyzed under the model specifications which are different from that used for the data generating, and which therefore can be regarded as the incorrect specifications. As simulation studies to investigate properties of statistical procedures, such as power, coverage rates of confidence intervals, and the good estimation of standard errors, some analyses will also be done under the correct model specifications. But most attention will be paid to robustness issues.

1.1 Outline

In chapter 2, an introduction to models of network dynamics and of the simultaneous dynamics of networks and individual attributes ('behaviour') of the individuals ('actors') in the networks is given. A practical example is also given in this chapter. In chapter 3, we explain the methodology of this robustness study and the issues in this study. The details of the simulations, such as the detailed design of simulations and the analysis of simulation results are given in chapter 4. A conclusion that based on the discussion on this robustness study will be given in chapter 5.

Chapter 2

Models for Network Dynamics and Models for Co-evolving Networks and Behavior Dynamics

Stochastic models are used to model the social network dynamics under the continuous Markov chain assumptions. Based on this, models for co-evolving social networks and individual behaviors are also proposed, which are not only concerned with the network dynamics but also with the changes in the actor behaviours in this network. The detailed explanations about these models are given in this chapter.

This chapter is based on Snijders (2006), Snijders (2005), Burk et al. (2007), Snijders, Steglich & Schweinberger (2007) and the *SIENA* manual version 3.11.

2.1 Some basic ideas of social network data

Before we start these stochastic models for network, it is necessary to illustrate some notations and configurations of the network data.

In a social network, there are a certain number of actors and some ties which indicate the relationship between them. Suppose the set of actors is N= {1,….., n}, and the ties can be defined as x_{ij} which indicate the relationship from i to j. Typically, $x_{ij} = 1$ indicates a tie from i to j, and $x_{ij} = 0$ indicates no such tie. A social network can either be represented by a directed graph where the nodes represent the actors, or an adjacency matrix with a structurally zero diagonal. For each actor, there will be some characteristics of the actors; these variables are called individual covariates and can be indicated by v_i. The attributes of individuals can either be constant or changeable over time. In the models for co-evolving networks and actor behaviours, the evolution of changeable attributes is considered to be influenced by the network dynamics.

There usually is a strong dependence between the tie variables x_{ij}, e.g. the social processes of reciprocity ("since you are my friend, I will become your friend") will lead to a dependence between x_{ij} and x_{ji}, and transitivity of choices ("friends of my friends are my friends") implies that when $x_{ij}=1$ and $x_{jh}=1$, there will be a tendency that $x_{ih}=1$.

The variable x_{ij} will be written as X_{ij} when it is regarded as a stochastic variable in the stochastic model.

Some important network configurations should be known, such as reciprocating arcs, transitive triplet and two-path etc, which are shown in Figure 2-1. These configurations play important roles in the following models because of their usefulness to represent the network structure.

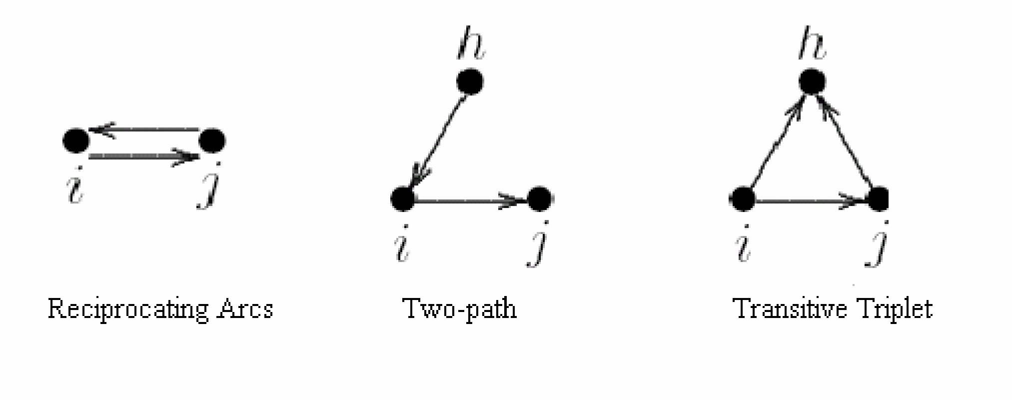


Figure 2-1 Some network configurations

As we concentrate on the network dynamics, the data should be longitudinal network data, which is typically shown as panel data. For $M \geq 2$ time points an observation $X(t)$ is available of the network on the same set N of actors.

2.2 Models for network dynamics

2.2.1 Assumptions

The stochastic models are built under the assumption that the network evolution is a continuous-time Markov process, where changes are the result of a series of small changes. Observations are made at discrete time points. The assumption of small changes also implies that there is no more than one tie variable that can change its value at a single moment.

2.2.2 Stochastic models

In actor-oriented models, the actors are assumed to control their outgoing ties. The network changes only by one tie at a single moment. Therefore the model includes two parts. The first one is determining the time when the change happens. The second part is to determine what kind of change the actor makes. The probability distributions in these two steps will depend on the current network structures and actor's observed attributes. The first part is determined by the so-called rate function and the second one is determined by the so-called objective function. In the second part, the actor is assumed to determine the change by optimize the objective function.

Some more general specifications like gratification function (see Snijders, 2001) may be included in the second part, but it will not be discussed here.

The rate function

The rate function $\lambda_i(x)$ for actor i is the rate at which there occur changes in this actor's outgoing ties. It can be a positive constant to be estimated from the data. It is also possible that the actor covariate and network structures effects are included in the specification of the rate function.

In this study, the rate function will be independent of the actors, and it will be assumed to be constant between observation moments.

The objective function

The objective function $f_i(x)$ of actor *i* is interpreted as the value attached by this actor to the network configuration *x*. It can be represented as a weighted sum of meaningful aspects of the networks as follows

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

The function $s_{ik}(x)$ can represent the structure of the network and covariates (actors' attributes) as follows. Only a few regular specifications are given below. The mathematical expressions and more specifications can be found in Snijders (2005) and the Manual for SIENA version 3.11.

1. Density effect (out-degree effect) , defined by the out-degree
2. Reciprocity effect, defined by the number of reciprocated ties, which is shown in Figure 2-1. A high value of the parameter means actors have a preference for reciprocal ties.
3. Transitivity effect, defined by the number of transitive triplets, which is also shown in Figure 2-1. A high value of the parameter means actors have a preference of network closure.
4. The distances two effect expresses network closure inversely: When this effect has a negative parameter, actors will have a preference for having network closure.
5. The covariate-similarity effect. A positive parameter implies that actors prefer ties to others with similar values on this variable
6. The effect on the actor's activity (covariate-ego); a positive parameter will imply the tendency that actors with higher values on this covariate increase their out-degrees more rapidly.
7. The effect on the actor's popularity to other actors (covariate-alter); a positive

parameter will imply the tendency that the in-degrees of actors with higher values on this covariate increase more rapidly.

8. The interaction between the value of the covariate of ego and of the other actor (covariate ego × covariate alter); a positive effect here means that actors with a higher value of the covariate have a greater tendency to have ties to others actors who also have higher values of this covariate; this is similar to a positive similarity effect. In models where the ego and alter effects are also included and the covariate is binary, it even is equivalent to the similarity effect (although expressed differently)

The changing probabilities and intensity matrix

The model assumes that when actor *i* changes one of his tie variables with some other actor *j*, the value of

$$f_i (x(i \rightarrow j)) + U (j) \quad (2)$$

is maximized. Where $f_i (x(i \rightarrow j))$ can be interpreted as "the directed graph obtained from **x** when *i* changes the tie variable to *j*" , and $U (j)$ are assumed to be random variables distributed symmetrically about 0 and independently generated for each new change.

The $U (j)$ is always assumed to be Gumbel distribution with mean 0 and scale parameter 1 (Maddala, 1983). Under these assumptions, the probability that *i* choose to change x_{ij} for another *j*, is given as

$$p_{ij}(x) = \frac{\exp(f_i(x(i \to j)))}{\sum_{h=1,h\neq i}^{g} \exp(f_i(x(i \to h))}\qquad (j \neq i) \quad (3)$$

So the intensity matrix (Sidney I. Resnick ``*Adventures in Stochastic Processes*'', 1992) of this Markov chain can be written as

$$q_{ij}(x) = \lambda_i(x) p_{ij}(x) \quad (4)$$

2.3 Models for co-evolving networks and actor behaviours

The models for co-evolving networks and actor behaviours are quite similar to the models for network dynamics. They are also actor-oriented models. The difference here is that the actors can control not only their outgoing ties but also their behaviours like smoking, drinking and so on. This model assumes that the network and the behaviour develop in a mutual dependence. Social selection and influence processes occur. Considering an example where the relation (network) is friendship and the behaviour is smoking, the selection process is that people who smoke may like to choose friends who also smoke, and the influence process is that people who don't smoke may start to smoke because their friends smoke. Here $Z_i(t)$ is given as the behaviour variable of actor i; and $X_{ij}(t)$ is given as the tie variable.

2.3.1 Assumptions

This is also an actor-oriented model for longitudinal social network data, so it also has the continuous-time Markov chain assumption, which means the entire outcome (X, Z), where X is the adjacency matrix and Z is the vector ($Z_1, \ldots, Z_n$) jointly is a Markov process with parameter t. And at any single moment, no more than one of all the variables X_{ij} and Z_i can be changed.

Each actor gains, at random moments, the opportunity to change one outgoing tie and the opportunity to change his or her behaviour.

Another important assumption is that the actors react to the change of each other in network ties and behaviour, but do not negotiate or make joint changes based on a prior agreement. Changes of X_{ij} show the selection part and changes of Z_i give the influence part.

2.3.2 Stochastic Models

The models can also be regarded as two parts which are the same as the network dynamics models. The first part is to determine the time to change ties or behaviours which is determined by the rate function. The second part is to choose which tie or behaviour to change to maximize the objective function.

So there are two functions included in this model, the rate function and objective function.

These functions are in the same form as that in the network dynamic models. But there are two for one kind of function, which are the tie variable X_{ij} and the behaviour variable Z_i.

The rate function

The moments when any given actor i has the opportunity to make a decision to change its outgoing or behaviour are determined randomly and follow Poisson processes. For each actor i, there is one rate function $\lambda_i^{[X]}$ for the network and one $\lambda_i^{[Z]}$ for each behavioural dimension. Rate functions can depend on the time period, and are also allowed to depend on the network characteristics or behaviour variables.

Also in our study, the rate functions will be independent of the actors, and are assumed to be constant between observation moments.

The objective function

There are also two objective functions written as $f_i^{[X]}$ and $f_i^{[Z]}$, which are for the evolution of network dynamics and behaviour dynamics separately. Both of them are modeled as weighted sums.

$$f_i^{[X]}(\beta,x)=\sum_{k=1}^{l}\beta^{[X]}{}_k s_{ik}{}^{[X]}(x) \qquad (5)$$

$$f_i^{[Z_h]}(\beta,x)=\sum_{k=1}^{l}\beta^{[Z_h]}{}_k s_{ik}{}^{[Z_h]}(x) \qquad (6)$$

The specification of $s_{ik}{}^{[X]}(x)$ is almost the same as that in models for network dynamics which discussed above. There are some examples of $s_{ik}{}^{[Z_h]}(x)$ below, the mathematical expressions and more specifications can be found in Snijders, Steglich & Schweinberger (2007) and Manual for SIENA version 3.11

1. The tendency effect, expressing the tendency that toward high value on *Z*
2. The effect of the behavior Z on itself, which is relevant only if the number of behavioral categories is 3 or more. This can be interpreted as giving a quadratic preference function for the behavior. With a negative coefficient, this represents that the most desired behavior can lie somewhere between the minimum and maximum values of the behavioral variable.
3. The total similarity effect, expressing the preference of actors to being similar to their alters, where the total influence of the alters is proportional to the number of alters.
4. The average similarity effect, expressing the preference of actors to being similar with respect to *Z* to their alters, where the total influence of the alters is the same regardless of the number of alters.
5. The average alter effect, expressing that actors whose alter have a higher average value of the behavior Z, also have themselves a stronger tendency toward high values on the behavior.

The changing probabilities

The same with models for network dynamics there is a random residual ε included when determining which tie or which behaviour should be changed. It is assumed to be independent and follows a standard Gumbel distribution.

So for the network decisions, the probability that actor i will change the tie x_{ij} to another *j* is

$$p(x(i \to j) \mid x(t), z(t)) = \frac{\exp(f_i^{[X]}(x(i \to j)(t), z(t)))}{\sum_k \exp(f_i^{[X]}(x(i \to j)(t), z(t)))} \qquad (7)$$

And the in the behaviour decisions, the probability of actor *i* changing behaviour *h* is

$$p(z(i \to h) \mid x(t), z(t)) = \frac{\exp(f_i^{[Z]}(x(t), z(i \to h)(t)))}{\sum \exp(f_i^{[Z]}(x(t), z(i \to h)(t)))} \qquad (8)$$

2.4 Estimations and software

These stochastic models are too complicated for explicit calculation of expected values. Instead the Robbins Monro approximation method and Monte Carlo simulation method are used to obtain approximate expected values of relevant statistics and parameters Snijders (1994, 1996, 2001). Usually, the method of moment is used here. Since the parameter estimates are approximately normally distributed, the corresponding t-statistics (the t-ratio, estimate divided by standard error) are used to determine approximate significance. Tests based on the t-ratio will be referred as the t-test here.

All the estimations and simulations can be done by SIENA (Simulation Investigation for Empirical Network Analysis) in practice. In this thesis, we use SIENA version 3.1 to estimate and simulate.

2.5 An example

To illustrate the models clearly, a simple example is given. The data is part of the West-Pearson Glasgow data set (Pearson and West 2003). The West-Pearson Glasgow data were recorded for a cohort of pupils in the West of Scotland. The panel data were recorded over a three year period starting in 1995, when the pupils were aged 13, and ending in 1997. A total of 160 pupils took part in the study, 129 of whom were present at all three measurement points. The friendship networks were formed by allowing the pupils to name up to twelve best friends.

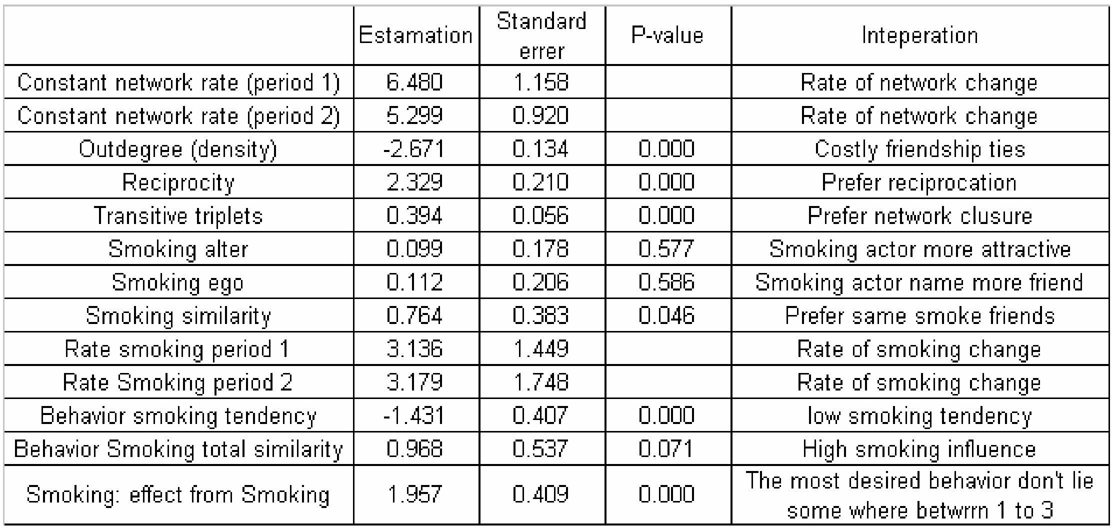
Pupils were also asked about substance use and adolescent behaviour associated with, for instance, lifestyle, sporting behaviour and tobacco, alcohol and cannabis consumption.

The analysis reported here is based on the subset of 50 girls from the whole dataset. For the behaviour attributes only smoking is counted. A model for co-evolving network and actor behaviours can be built based on the chosen dataset.

As usual, $x_{ij} = 1$ indicates that there are friendship between girl i and j, and $x_{ij} = 0$ indicates that there is no friendship between them. Smoking was coded as: 1(non), 2(occasional) and 3(Regular i.e. more than once per week)

We choose some regular effects in this model. The estimates of parameters and correspondence standard errors obtained from SIENA are shown in Table 2-1

Table 2-1 SIENA estimation results from the example model



	Estamation	Standard errer	P-value	Inteperation
Constant network rate (period 1)	6.480	1.158		Rate of network change
Constant network rate (period 2)	5.299	0.920		Rate of network change
Outdegree (density)	-2.671	0.134	0.000	Costly friendship ties
Reciprocity	2.329	0.210	0.000	Prefer reciprocation
Transitive triplets	0.394	0.056	0.000	Prefer network clusure
Smoking alter	0.099	0.178	0.577	Smoking actor more attractive
Smoking ego	0.112	0.206	0.586	Smoking actor name more friend
Smoking similarity	0.764	0.383	0.046	Prefer same smoke friends
Rate smoking period 1	3.136	1.449		Rate of smoking change
Rate Smoking period 2	3.179	1.748		Rate of smoking change
Behavior smoking tendency	-1.431	0.407	0.000	low smoking tendency
Behavior Smoking total similarity	0.968	0.537	0.071	High smoking influence
Smoking: effect from Smoking	1.957	0.409	0.000	The most desired behavior don't lie some where betwrrn 1 to 3

The p-values given refer to two-sided t-tests based on the t-ratio defined as parameter estimate divided by standard error, testing whether the corresponding parameter deviated significantly from zero. It doesn’t make sense for rate parameters (the fact that any change has occurred indicates that the rate cannot be zero). So p-value is only given for the parameters in the objective function. Usually, when p-value is greater than 0.05, the estimation is considered as non-significant. Interpretations are also given in Table 2-1.

Chapter 3

Methodology and Issues of the power and robustness study by simulation

The models introduced in chapter 2 show that there are many kinds of specifications that can be chosen to build the model, but which of them should be chosen is a problem. We cannot guarantee our choice is always right in practice. What if the choice is not suited for the real data? How sensitive is the model to these specifications? The aim of this project is to find out the degree of sensitivity of the estimates and the tests obtained for these models to the model specification. This is studied by means of simulations.

3.1 Methodology

In this study, firstly, network dynamics data or network and behaviour dynamics data are generated based on a known model, where the values of all parameters are known. This model is called model *A*, or the true model. These data are analyzed under different model specifications. The models used to estimate and test parameters, in most cases having different specifications from model *A,* are called model *B* or the postulated model. After a reasonable number of repeats (in this study 500 times usually), we get many estimates under the assumption of model *B*. By analyzing those results, we can find how different these estimates are from the value in model *A*. In this way, the model's degree of sensitivity (or robustness) to misspecification is obtained.

The case that model *A* =model *B* will also receive some attentions as a baseline situation, to investigate the properties of statistical procedures such as power, coverage rates of confidence intervals, and the good estimation of standard errors.

3.2 Concerned Issues of the study

The issues concerning in this study is different when model A=model B from when model $A \neq$ model B.

3.2.1 When model A= model B

When model A=model B, we analyze the simulated data using the exactly right model specification. The results of the estimation should be, ideally, the same as we proposed in model A. Using these results, we can check the estimation of these models.

The tests proposed for this model by Snijders (2001) assume that the estimators of the parameters are normally distributed. So the 500 estimates should be normally distributed. The standard error of a parameter's estimations should be approximately the same value as the average standard deviation of this parameter's estimator. The proportion of data sets for which the hypothesis that the parameter equals its true value was rejected by the two-sided t-test should be no more than 5%, which implies there are no more than 5% incorrect rejections. We can also test whether the estimator is biased in this kind of model by using a t-test.

3.2.2 When model $A \neq$ model B

There are many possibilities of the choice of model B. It is impossible to study all of them in this project, so we chose some of them. Because of the difference of model specifications in the models for network dynamics and for co-evolving networks and actor behaviours, the issues are different.

In models for network dynamics

In models for network dynamics, attentions are mainly paid to two kinds of effects.

These are the effects that can represent the transitive closure and the effects that can represent the covariate (behaviour) similarity.

There are different effects that can represent the transitive closure of the network, like the transitive triplets effect, the balance effect, the direct and indirect ties effect and the distances two effect (SIENA manual 3.11). Normally, including one or two of these four effects is enough for the model. Which effect should express the transitive closure for the model needs to be decided. Also the covariate similarity can be represented by different specifications, like the covariate similarity effect and covariate ego × covariate alter effect. The covariate similarity effect can also lead to a tendency toward transitivity, and therefore the transitive closure effects and the covariate similarity effects could also be collinear to some extent.

.

The first question is that to what extent tests for transitive closure are sensitive for misspecifications of the model for transitive closure (like the different choice of the four effects for transitive closure); and the second question is that to what extent tests for covariate effects are sensitive for misspecifications of the model for transitive closure. Vice versa, a third question is to what extent tests for transitive closure are sensitive for misspecifications of the covariate effects, like the covariate similarity.

In models for co-evolving networks and actor behaviours

In models for co-evolving network dynamics, we are mainly concerned with the effects which represent the ‘social influence’ and ‘social selection’. For example there are three different specifications of social influence which are the behaviour total similarity effect, the average similarity effect, and the average alter effect. There are also different specifications of social selection, like the covariate similarity and covariate ego × covariate alter.

The questions are the robustness of the test for social influence for the precise specification of the influence terms in the model and for the specification of the selection terms in the model, and the robustness of the test for social selection for the precise specification of the selection terms in the model and for the specification of the influence terms in the model.

.

Chapter 4

Simulation

To solve the issues proposed in last chapter, the simulations are done by using the methodology which was explained in chapter 3. The detailed simulation design is shown in this chapter. After some statistical analysis of the simulation results, interpretations of the simulation results are given.

4.1 Simulation design

The simulation design is based on the data used in the example model given in section 2.5, which is an excerpt of 50 girls from the West-Pearson Glasgow dataset (detailed explanation of this dataset is given in section 2.5). Model *A* is built to be able to generate data that are similar to this dataset, which includes 50 actors, 238 ties and behaviour data. Some small changes have been made to adapt this study's aim.

4.1.1 Network dynamics models

In models for network dynamics, we did not use smoking as a dependent behaviour as what was done in the example model. So there are no effects about the behaviour evolution. The other specifications are the same as that in the example model. We want to include the plausible effects in model A, and effects included in model A should be able to satisfy our study's requirements. The specifications included in model *A* for the network dynamics models are shown in Table 4-1, and the estimates of these parameters under model *A* using the original dataset are given in Table 4-1.

Table 4-1 Estimation results of network dynamics model *A* specifications using original data

	Estimate	Standard error	P-value
Constant network rate (period 1)	6.550	1.112	
Constant network rate (period 2)	5.388	0.934	
Outdegree (density)	-2.682	0.123	0.000
Reciprocity	2.388	0.209	0.000
Transitive triplets	0.413	0.053	0.000
Smoking alter	0.103	0.112	0.358
Smoking ego	0.104	0.126	0.405
Smoking similarity	0.519	0.227	0.022

Another important thing for model *A* is to decide on the values of parameters of these effects. We use an approximation of the estimates based on the original dataset (50 girls dataset) under the model *A* specification here.

Some changes are made at this step because of the study's purpose. Firstly, in this case, there are two main concerned effects which are the transitive triplets effect and the smoking similarity effect. We want to know the difference of simulation results when those two parameters are given different values. So beside the value given by the approximation of estimation, these two effects may also be 0. As each of them has two possible values, there are four conditions because of the combination. Under this condition, we actually have four kinds of model *A*. We use the symbol *00* to represent the case where both parameters are 0, *01* to represent that the parameter for the transitive triplets effect is 0 and smoking similarity is not 0. Similarly *10* means that transitive triplets is not 0 and smoking similarity is 0, and *11* means both of them are not 0.

The value of outdegree (density) parameter together with the value of other parameters determines the average number of ties in the network dynamics. If we use the approximation of the estimated values based on the 50 girls data and just set one or both of the mainly concerned parameters to 0, the number of ties becomes very small (sometimes smaller than 100, and we have 50 actors), whereas the number of ties of the 50 girls data is 238. With a small number (less than 100) of ties, the estimation can hardly make sense and also the convergence of the estimation algorithm can be very poor. So to make sure the number of ties is around 238, we change the value of outdegree parameter. Different values are given under each of the different conditions *00, 01, 10*, and *11*. These values were obtained by some trial and error, aiming at an average number of ties not too far from the observed 238.

The final parameter values for model *A* are given in Table 4-2. The average number of ties in the simulations is also shown in Table 4-2

Table 4-2 Model *A* (true model) parameter values used for simulation in Network dynamics models

	Parameter Values			
	00 case	*01* case	*10* case	*11* case
Constant network rate (period 1)	7.0	7.0	7.0	7.0
Constant network rate (period 2)	5.0	5.0	5.0	5.0
Outdegree (density)	-1.9	-2.0	-2.4	-2.3
Reciprocity	2.0	2.0	2.0	2.0
Transitive triplets	0.0	0.0	0.3	0.3
Smoking alter	0.1	0.1	0.1	0.1
Smoking ego	0.1	0.1	0.1	0.1
Smoking similarity	0.0	0.4	0.0	0.4
Average number of ties of dataset simulated	285.13	252.90	195.02	248.77

The model *B* design is based on the issues discussed in chapter 3. Details are shown in Table 4-3 :

Table 4-3 Model *B* (postulated model) specifications in Network dynamics model

The model symbol	The model *B* specifications
a	The same as the plausible model *A*
b	The Number of actors at distance 2 is used instead of the Transitive triplets, further as *a*.
c	The Transitive triplets is left out, further as *a*.
d	The ego × alter interaction is used instead of the Smoking similarity, further as *a*.
e	The Smoking similarity is left out, further as *a*.

4.1.2 Co-evolving network and actor behaviours models

The specification of Model *A* for co-evolving network and behaviour is the same as that of the example model explained in section 2.5. Based on the estimation results of the example model all these effects are plausible and they also satisfy our study's requirement. The effects and their estimation results using the original data are already shown in Table 2-1.

In this model the values of parameters are decided based on the same principles as in the models for network dynamics.

There are also two mainly concerned effects which are smoking similarity (in the network dynamics part of the model) and the behavior smoking total similarity. The symbol *00* means both of them are 0, and *01* means smoking similarity is 0 but behavior smoking total similarity is not 0. Further *10* represents smoking similarity is not 0 but behavior smoking total similarity is 0, and *11* means both of them are not 0.

Like in the preceding subsection the value for density parameter has been changed to make sure the number of ties is around 238. The values of the model A parameters and the average number of ties of these four combinations are given in Table 4-4.

Table 4-4 Model *A* (true model) parameter values used for simulation in Co-evolving network and behaviour dynamics model

	Parameter Values			
	00 case	*01* case	*10* case	*11* case
Constant network rate (period 1)	7	7	7	7
Constant network rate (period 2)	5	5	5	5
Outdegree (density)	-2.3	-2.3	-2.3	-2.3
Reciprocity	2	2	2	2
Transitive triplets	0.3	0.3	0.3	0.3
Smoking alter	0.1	0.1	0.1	0.1
Smoking ego	0.1	0.1	0.1	0.1
Smoking similarity	0	0	0.4	0.4
Rate smoking period 1	3	3	3	3
Rate Smoking period 2	3	3	3	3
Behavior smoking tendency	-1.4	-1.4	-1.4	-1.4
Behavior Smoking total similarity	0	1	0	1
Smoking: effect from Smoking	2	2	2	2
Average number of ties of dataset simulated	263.45	252.29	259.31	256.32

Also the design of model *B* is based on the concerned issues mentioned in chapter 3. The details are given in Table 4-5.

Table 4-5 Model *B* (postulated model) specifications in Co-evolving network and behaviour dynamics model

The model symbol	The model *B* (postulated model) specifications
a	The same as the plausible model *A*
b	The Behaviour smoking average similarity effect is used instead of the Smoking total similarity further as *a*
c	The Smoking alter and Smoking ego are left out, further as *a*
d	The ego x alter interaction is used instead of the Smoking similarity, and Behaviour smoking average alter is used instead of Behaviour smoking total similarity, further as *a*
e	The Effect smoking on smoking is left out, further as *a*
f	The Smoking similarity is left out, further as *a*
g	The Behaviour smoking total similarity is left out, further as *a*

4.2 Simulation Results and integration of the results

Before we start to analyze the simulation results, some notation for the models needs to be clarified. In section 4.3, models are represented by notations like *Na-00*. The *N* means it is for Network dynamics. If it is *Ca-00*, it represents a Co-evolving network and actor behaviours model. *a* in this notation indicates the model *B* (postulated model) specification type, which can refer to Table 4-3 and Table 4-5. *00, 01, 10* or *10* indicate the parameters values used to generate data in model *A (*true model*),* which can refer to Table 4-2and Table 4-4.

For every type of model, we generated at least 500 network dynamics or network and behaviour dynamics datasets. Subsequently we estimated the parameters for every dataset under the model *B* specifications (postulated model). This yields at least 500 estimated values for each parameter of every type of model. For a small number of these estimations, the algorithm has not converged properly. As their number is very small (around 10 for each type of model) compared with the total number of simulations (500 or more), they are discarded. For each combination of models, some statistical properties of these estimations are given. Details are listed below

1. True value: The parameter’s value in model *A* (true model)
2. Mean: The mean of the estimates for this parameter
3. Proportion of significant estimates (PSE): The proportion of estimates of this parameter deviating significantly from 0, according to the two-sided t-test (t-ratio test) which uses the nominal significance level of 5%. This is an estimate of the power of the test that the parameter is equal to 0.
4. Estimated type-I error rate (ETER): Proportion of estimates for which the parameter deviated significantly from its true value according to a two-sided t-test (t-ratio test) which uses the nominal significance level of 5%
5. T-test P-value (TP): P-value of the two-sided t-test for unbiasedness of this parameter's estimator, where the H_0 is that the expected value of the estimator is

equal to the true value. The significance level of 5% is used here.

6. MSD: The mean of this parameter's estimates' standard error
7. SD: The standard deviation of this parameter's estimator
8. Difference ratio of MSD and SD (DMS): $\frac{\left|\text{MSD - SD}\right|}{\text{SD}}$
9. Estimation KS-test p-value (EKSP): P-value of Kolmogorov-Smirnov test for this parameter's estimator, which is to check the normality of the estimator.
10. Standard test statistics KS-test p-value (SKSP): P-value of Kolmogorov-Smirnov test for the standard test statistics of this parameter's estimates, which is to check the normality of the standard test statistics. The standard test statistic is $\frac{\text{Estimate - True value}}{\text{This estimate's standard error}}$
11. Sqrt KS-test p-value (SQKSP): P-value of Kolmogorov-Smirnov test for this parameter's estimates' square roots, which is to check the normality of the estimator's square root.

4.2.1 Results analysis when model *A* = model *B*

When model *A* = model *B*, the model used to estimate the parameters is well specified. We hope the mean is very close to the true value and the estimator is normally distributed. ETER (Estimated type-I error rate) and DMS (difference ratio of MSD and SD) should not be greater than 0.05. PSE (Proportion of significant estimates), which is an estimate of the power of the test, should be bigger than 0.95, when the parameter is indeed nonzero. TP (T-test P-value) should not be less than 0.05.

Network dynamics

Appendix Table 1 shows all the statistical results of network dynamics models *Na*. Some conclusions can be obtained based on it.

Most of the TP are much greater than 0.05. But the TP of Transitive triplets in all these four models and Reciprocity in model *Na-10* and *Na-11* are less than 0.05, which means these estimations are biased. But the bias is not big. The bias is only around 0.02. So generally the estimator is very close to the true value.

PSE in these models are satisfying except Smoking alter and smoking ego. The power of significance tests of Outdegree, Reciprocity and Transitive triplets are good. Their PSE are all larger than 0.95 when their true values are not zero, and are less than 0.05 when their true values are zero. But powers of Smoking alter and smoking ego's significance tests are not so good. In all these four models, their PSE are around 0.20.

All the parameter's ETER are not large. Although not all of them are ideally less than 0.05, the one greater than 0.05 is only around 0.06. With 500 replications, the standard error of these proportions is $(0.05*0.95/500)^{0.5} = 0.01$, so the deviations from 0.05 can be regarded as chance fluctuations. The estimates are quite precise.

Most of the DMS are less than 0.05. Only few parameters have a 0.06 or 0.07 ratio. So the estimation of standard error is also good.

Most of the estimations can be regarded as normal distributed except the Transitive triplets and Constant network rate (period 2) in model *Na-10*., because most of the EKSP are greater than 0.05 except the Transitive triplets and Constant network rate (period 2) in model *Na-10*. We can also find the normality of rate parameters and Transitive triplets are not very good. Some of them should skew slightly as evidenced by their poor EKSP although these are not less than 0.05. Most of the SKSP are larger than EKSP. This means, especially for those whose EKSP is less than 0.05 like Constant network rate (period 2), outliers are compensated by large values of the standard errors. However this is different for period 1 rate parameter in model *Na-10* and *Na-1I*. Their SKSP are less than EKSP.

For the rate parameters' estimators, they may follow a distribution similar in shape to a Poisson distribution times some scale factor because of the model definition, so a square root transformation is tried. The normality of the square roots is better than the raw data. All the square roots' ks-test p-value are greater than 0.05. This can be identified by comparing the EKSP and KS-test p-values of these square roots.

All the normality conclusions can also be obtained by looking at these data's qq-plots, which are shown in Appendix Figure 1.

As the t-tests are based on the assumption that the estimators follow a normal distribution, the results of tests may be not so reliable when the parameters' estimators and standard test statistics are not normally distributed.

Generally, estimations using this network dynamics model are satisfying, either their significance or precision, and most of the estimators are normally distributed, except some rate parameters and the transitive triplets parameter.

Co-evolving network and actor behaviours models

Statistical test results of *Ca* models are given in Appendix Table 2.

The TP in co-evolving network and behaviour dynamics models (*Ca* models) are not as good as that in network dynamic models. Almost the two thirds of them are less than 0.05, and some of them are close to 0. These estimators are biased. However, for most of the parameters the bias is not large, which is around 0.1, except for Behavior smoking tendency (BST) and Behavior Smoking total similarity (BSTS). True value of BST is -1.40, but the mean of BST is around 0.49 in the case *00* and *10*, around 0.8 in *10* case and around 1.0 in the case *11*. The bias of BST is also quite large. When BSTS' true value is 1.0, the mean of estimates is around 0.75.

The PSE of network structure effects parameters (network rate parameter, outdegree, reciprocity, and transitive triplets) are all greater than 0.95 and so do the Smoking effect from smoking. Powers of their significance test are very good. When Smoking similarity and Behaviour smoking total similarity's true value are 0, their PSE is close to 0, which gives the evidence that the parameter should be 0. PSE of Smoking rate parameters are around 0.88 and they only get about 0.75 in the *11* case. PSE of the smoking alter and ego are only around 0.13. When Smoking similarity's true value is 0.4, its power is only about 0.30. The power of Behaviour smoking total similarity's significance test is the worst. When its true value is 1, its PSE is less than 0.02.

ETER are around 0.05 and not greater than 0.1 except the Rate smoking period 1, BST (Behavior smoking tendency) and BSTS (Behavior Smoking total similarity). Corresponding to the large bias of BST reported above, the ETER of BST is also quite large which is almost 0.80 in the case *00* and *10*, 0.43 in the case *01* and 0.25 in the case *11*. When the true value of BSTS is 1, the ETER is about 0.2, which is also quite large. The ETER of Rate smoking period 1 is around 0.14 in the case *01* and *11* (when true value of BSTS is 1.0), and is around 0.11 in the case *00* and *10* (when true value of BSTS is 0).

All parameters' DMS are less than 0.05 except BST (Behaviour smoking tendency) and SES (Smoking: effect from Smoking). For BST, it should be caused by the large bias. The DMS of BST are all around 0.3. And the DMS of SES is all around 0.15.

The normality of estimations in this model is not quite satisfying. Almost half of its parameters EKSP is less than 0.05, like smoking rate parameters, BSTS (Behavior Smoking total similarity), SES (Smoking: effect from Smoking), and network rate parameters in some case. Also the SKSP are always greater than EKSP except few odd estimates. Most of the SKSP are greater than 0.05 except the rate parameters for both network and behaviour and BSTS. So most of the parameters' standard test statistics can be regarded as normally distributed except the rate parameters and BSTS. Details can be found in Appendix Table 2.

For the rate parameters' estimators, the same as models for network dynamics they may follow a distribution similar in shape to a Poisson distribution times some scale factor because of the model definition, so a square root transformation is tried. The normality of the square roots is better than the row data, which can be identified by comparing the EKSP and KS-test p-value of these square roots. But for some of the behaviour rate parameters, the square roots still seriously skews.

All the normality conclusions can also be obtained by looking at these data's qq-plot, which are shown in Appendix Figure 2. From the qq-plots we can find that some distributions of the raw estimates are heavy-tailed, whereas for the standardized test statistics they are light tailed like for BSTS, BESS (Behaviour effect smoking on smoking). This confirms that outlying estimates are counterbalanced by large standard errors and even by standard error which are too large.

As the same to models of network dynamics, the t-tests are done with the assumption that all the estimators are normally distributed. When the estimators and standardized test statistics are not normally distributed, the results of tests are not reliable.

Generally, some estimates in models for co-evolving network and behaviour dynamics are good, but not all. It is not as good as the models for network dynamics. The precision of some effects' parameter's estimates like behaviour smoking tendency is quite biased. Some power of parameters significance tests has problems, like Smoking similarity and Behaviour smoking total similarity. Almost half of the effects' parameters' estimators' can not be regarded as normally distributed. However most of the estimators' standard test statistics follow a normal distribution except the rate parameters and behaviour smoking total similarity.

4.2.2 Results analysis when the model $A \neq$ model B

This part is concerned with the estimator's significance and precision when the data are estimated under incorrect model specifications.

In Network dynamics models

In the models for network dynamics, we are mainly concerned by the effects with respect to transitive closure and smoking behaviour. Most of the analysis is about their changes under misspecifications. Other parameters estimates are also reported in the tables, but less attention is given.

Model *Nb*

In model *Nb*, the place of Transitive triplets effect is taken by Number of actors at distance 2. The statistical test results are shown in Appendix table 3.

We know that both Transitive triplets and Number of actors at distance 2 can represent the network closure. When there is a positive value for Transitive triplets parameter, the value for Number of actors at distance 2 is expected to be negative.

Firstly, in the *00* case, where both Transitive triplets (TT) and smoking similarity (SS) are 0, the mean of Number of actors at distance 2 (NAD) is also 0. The TP is large enough to support the conclusion that in this case the estimator of NAD is very close to 0. The PSE of NAD also supports this conclusion. No significant difference from the model *Na-00* is found while checking other estimates.

In the case *01*, SS's true value is 0.4. The mean of NAD is -0.022. Others TP are all greater than 0.05. The PSE of NAD is 0.058, and it shows most of the estimates are still can be regarded as 0. Other test values are almost the same as that in *Na-01.* So generally, no significant change has taken place.

In the case *10*, when TT's true value is 0.3, the mean of NAD is -0.164. The power of NAD's significance test is not so good. Its PSE is only 0.30. All the estimates in this model are biased based on the fact that all the TP are 0. This didn't happen in model *Na-10.* Smoking alter's ETER is 0.16 which means the estimator is not precise enough. The outdegree and reciprocity's ETER are also too large.

In the case *11*, the mean of NAD is -0.226, even smaller than that in the case *10*. The power of significance test of NAD is better than that in the case *10*; the PSE is 0.53 here. The power of SS's significance test has been improved, where its PSE is 0.65 in this case compared with 0.59 in model *Na-11*. The Smoking alter's ETER is still large, so the imprecision still exists, so does the outdegree and reciprocity's estimates. All the TP here are close to 0, which is not in model *Na-11*. Some of the estimators, like SS, are biased because of the misspecification.

In summary, when we change the TT into NAD and TT's true value is not 0, models estimates are affected. The NAD will have on average a negative value, but a small power of significance test. The knowledge of the true model increases the power of significance test. Some of the parameters' estimators become biased. The Smoking alter and outdegree's preciseness has been affect. But there is no significant change when TT's true value is 0 and estimates of NAD are on average 0.

Model *Nc*

In model *Nc*, the Transitive triplets (TT) effect is deleted from the model specification in model *B.* The test results are shown in Appendix Table 4.

Firstly, in the case *00* and *01* where TT's true value is 0, not many changes of these estimations can be found. Only the fact that TP of outdegree is less than 0.05 is not found in model *Na-00* and *Na-01.* Actually in such a case, the 0 value of TT implies that model *B* is correct. We have a well specified model here.

In the case *10* and *11*, where TT's true value is 0.3, there is something changed. Firstly, all the TP are close to 0. The Smoking similarity, smoking ego and smoking alter were not biased in model *Na-10* and *Na-11.* It means that they become biased because of the misspecification. All the ETER are greater than 0.05. So Smoking similarity, smoking ego and smoking alter become more imprecise because of the misspecification. From the value of ETER, it looks like Smoking similarity's precision doesn't significantly change. However, the significance of these estimates is still as good as that in *Na* models.

In summary, when we leave out Transitive triplets and its true value is not 0, the estimations' preciseness will be seriously affected. But the power of significance test for other parameters won't change significantly. There is no significant affection when TT's true value is 0.

Model *Nd*

In model *Nd*, the Smoking ego × alter interaction effect is used instead of the Smoking similarity. As mentioned before, a positive ego × alter interaction effect means just like a positive similarity effect. In our model, where the ego and alter effects are included, it should be equivalent to the similarity effect (although expressed differently). The tests results are shown in Appendix table 5.

In the case *00* and *10*, where the SS (Smoking similarity)'s true value is 0. The means of Smoking ego × alter interaction (SEA)'s parameter's estimates are very close to 0 too. The power of their significance test is small, where the PSE are around 0.05. The SEA should also be 0, when SS's true value is 0. Other parameters' estimates are not different from that in model *Na.*

In the case *01* and *11*, where the SS's true value is 0.4, the mean of SEA is around 0.20, and the PSE is around 0.50. Obviously, SEA has a positive parameter and the significance is acceptable. TP of other parameters' estimates come to be very small, which means the misspecification makes them biased. Especially for smoking alter and smoking ego, their power of significance test decreases dramatically, which are almost less than 0.05 (they used to be around 0.20 in model *Na*). Smoking alter and Smoking ego's ETER are also changed significantly, which increase to almost 0.2 (they used to be around 0.05 in model *Na*).

So when we change SS into SEA and SS's true value is not 0, this misspecification will seriously affect Smoking alter and Smoking ego's power of significance tests and precision. At the same time, other parameters become biased because of this misspecification. However, when the true value of SS is 0, other parameters' estimates are not affected.

Model *Ne*

In model *Ne*, we leave out the Smoking similarity (SS) effect. The statistical tests results are shown in Appendix Table 6.

In the case *00* when SS and TT (Transitive triplets)'s true value are both 0, most of the parameters' estimates are not affected.

In the case *10*, when SS's true value is still 0, but TT's true value is 0.3, the Smoking ego and Reciprocity's TP become very small. They are biased because of the misspecification. At the same time, TT's ETER is 0.36, which is too large compared with 0.036 in model *Na-10.*

In the case *01* and *11*, when the SS's true value is 0.4, almost all parameters' estimates' are close to 0. Similar with the model *Nd* case, the estimates of Smoking alter and Smoking ego's parameters are affected dramatically. Their PSE and ETER both change to be poor values. Smoking alter and Smoking ego's power of significance tests are around 0.05 and ETER are around 0.16, sometimes almost 0.2.

When Smoking similarity is left out by misspecification and both SS and TT's true value is 0, nothing is affected. But when TT's true value is not 0 although SS's true value is 0, TT's precision is seriously affected. When SS's true value is not 0, all other parameters' estimators become biased. Smoking alter and Smoking ego's precision and significance are affected seriously.

In Co- evolving Network and actor behaviours models

In models for co-evolving network and actor behaviours, how the effects which represent the 'social influence' and 'social selection' are affected by the misspecification is the main issue. Most of the analysis is about them. The test results of other parameters' estimates are also reported in the tables, but little analysis of them is done.

Model *Cb*

In model *Cb*, the Behaviour Smoking average similarity (BSAS) effect is used instead of the Behaviours Smoking total similarity (BSTS). Both of them express the preference of actors to being similar to their alters. In BSTS the total influence of the alters is proportional to the number of alters and in BSAS the total influence of the alters is the same regardless of the number of alters. The tests results are shown in Appendix Table 7.

When the true value of BSTS is 0, in the *00* and *10* case, the mean of BSAS is around -0.25. The PSE of BSAS is only 0.001. So the parameter of BSAS should still be regarded as 0. Other parameters' estimates don't seem to be affected.

In the case *01* and *11*, where the true value of BSTS is 1, the mean of BSAS is around 1.75. The PSE of BSAS is around 0.017. As the PSE of BSTS is also not large (around 0.015) in model *Ca* model, 0.017 is reasonable for BSAS's power of significance test. Other parameters are not affected significant either in this case.

In summary, changing BSTS into BSAS by mistake won't affect other parameters' estimates significantly. When BSTS's true value is 0, BSAS tends to be 0 too; and when BSTS's true value is not 0, BSAS doesn't tend to be 0 either.

Model *Cc*

In model *Cc*, the Smoking alter effect and Smoking ego effect are left out, which should be related to the Smoking similarity (SS). The tests results are given in Appendix table 8.

Firstly for BSTS (Behaviour smoking total similarity), no big change can be found compared with values in Model *Ca*. All the tests result are similar with that in Model *Ca*.

Then for SS, when the true value of SS is 0, the PSE of estimates increased because of this misspecification. When the true value of SS is 0.4, the PSE of SS decrease and ETER of SS increase. This means both the power of SS's significance test and the precision of SS become worse.

So when the smoking alter and smoking ego are left out by mistake, the BSTS won't be affected. However, the SS will be affected dramatically. When SS's true value is 0, the power of significance test of SS become greater; and when SS's true value is not 0, its significance and precision become worse.

Model *Cd*

In model *Cd*, the Smoking ego x alter interaction (SEA) is used instead of the Smoking similarity (SS), and Smoking average alter (SAA) is used instead of

Behaviour Smoking total similarity (BSTS). The Smoking average alter is another way to represent the ‘social influence’. The ego x alter interaction is also another effects to represent ‘social selection’. The test results are shown in Appendix Table 9.

As SEA can represent SS, to some extent, when the true value of SS is 0, the mean and PSE of SEA are both close to 0, which means the parameter of SEA is very close to 0 under this condition. When the true value of SS is 0.4, the mean of SEA are around 0.2 and the PSE are around 0.3 which is similar to the PSE of SS in model *Ca*. There should be a positive value for SEA’s parameter.

When the true value of BSTS is 0, the mean of SAA are around -0.11 and the PSE of SAA are close to 0. This means the parameter of SAA should still be regarded as 0. When the true value of BSTS is 1, the mean of SAA are also around 1 and the PSE of BSAA are around 0.15 which are similar to the PSE of BSTS in model *Ca*.

The Smoking alter and Smoking ego’s PSE and ETER are affected when SEA’s true value is non-zero, like the situation in network dynamics model *Nd*.

Generally, when these two exchanges happen together, the new effects’ estimates are quite similar to the previous effects. The parameter’s estimate of both SEA and SAA are almost 0 when SS and BSTS’ true value are 0, and they have a positive value when SS and BSTS’s true value are positive. SEA and SAA’s powers of significance test are close to SS and BSTS’s in model *Ca* when SS and BSTS’s true value are non-zero. The estimators of Smoking alter and Smoking ego will be affected because of the misspecification of SS.

Model *Ce*

In model *Ce*, the Smoking: effect from smoking (SES) effect is left out, which means the quadratic preference function for the behavior has been deleted. The tests results are shown in Appendix Table 10

In this case, the estimates of SS (Smoking similarity) are almost the same to that in model *Ca*. No obvious affection can be found.

For BSTS (Behaviour smoking total similarity), when its true value is 1, the PSE of its estimates are smaller compared with that in model *Ca*. The PSE are only less than 0.02 which are around 0.15 in model *Ca*. When BSTS and SS's true value are both 0, the PSE of BSTS is greater than that in model *Ca* (0.015 compared with 0.001), which means the estimator of BSTS become less close to 0 under this condition. But the PSE are still less than 0.05 when the true value is 0, which means we can still regard BSTS as 0. Other tests results are almost the same with that in model *Ca*.

In summary, this misspecification doesn't affect the estimates of SS, but the powers of significance tests of BSTS's parameter are affected. When BSTS's true value is not 0, the significance become poorer, but when BSTS's true value is 0, the power of significance become larger.

Model *Cf*

In model *Cf*, the SS (Smoking similarity) effects which represent the 'social selection' in model *A* is deleted. The tests results are given in Appendix Table 11.

When the true value of SS is 0, no obvious change can be found compared with model *Ca*.

When the true value of SS is 0.4, the Smoking alter and Smoking ego's estimates'

PSE become smaller, which is also found in network dynamics model *Ne* where the similar misspecification happens. However, from the tests results we cannot find some evidence that this misspecification have affected the estimates of BSTS which represents the ‘social influence’ in this model.

This misspecification doesn’t affect the estimates dramatically except for Smoking alter and Smoking ego’s parameter.

Model *Cg*

In mode *Cg*, we leave out the BSTS (Behaviour smoking total similarity) effect, which represent the ‘social influence’ in model *A*. The test results can be seen in Appendix Table 12.

Under this misspecification, we can’t find significant changes of SS’s estimates compared that in model *Ca*. All the statistical tests results of SS are similar to that in model *Ca*, no matter whether the true value of BSTS is 0 or not. Other parameters estimates are not affected either. This result should be caused by BSTS’s poor power of significance tests.

Chapter 5

Conclusions

The statistical methods for network dynamics and methods for co-evolving network and individual behaviours have been developed by Snijders (2001, 2005) and Snijders, Steglich, Schweinberger (2007). However, the power and robustness of these models, in another word how sensitive are these models to misspecifications, have not earlier been studied. In this paper we use simulations to study these models' power and robustness.

A true model called model *A*, which has known parameters and plausible specifications, is given. Data are generated based on this true mode, and a postulated model called model *B*, which has different specifications from model *A*, is used to estimate the parameters using the generated datasets. These estimates are analyzed to study these models' power and robustness. As a baseline situation, the case where model *A*=model *B* are also analyzed to study the properties of these statistical models.

When model *A*=model *B*, we find that the estimates of models for network dynamics are good. Their significance and precision are all satisfying. Most of the estimators and their standard test statistics can be regarded as normally distributed, but the rate parameters and transitive parameters' raw data are slightly skewed. A square root transformation for the rate parameters was proposed based on its definition in the model. The square roots' normality is better.

When model *A*=model *B*, the estimates of models for co-evolving network and actor behaviours are not as good as the model for network dynamics. The power of some estimators' significance tests and precision of some estimates are good, but few estimators are quite biased. Some estimates' significance have problems. Almost half of the estimators can not be regarded as normally distributed, although most of their standard test statistics can be regarded as normally distributed.

When model $A \neq$ model B, the issues are different in models for network dynamics and in models for co-evolving network dynamics and actor behaviours.

In models for network dynamics, we are mainly concerned by the sensitivity of tests about transitive closure effects and covariate effects to misspecifications. In this case, in model B, transitive closure effects and attribute similarity effects are changed or deleted. The results of estimates under model B specifications show that these misspecifications always affect the precision and significance of estimates especially for the parameters which are strongly related to the changed effect's parameter. Some parameters' estimators become biased and imprecise because of the misspecification, and some of them become insignificant.

In models for co-evolving network dynamics and actor behaviours, effects which could represent 'social influence' and 'social selection' are mainly studied. These effects are changed or left out in model B. In the analysis of simulation and estimation results, only these effects' parameters are studied carefully. Generally, the significance and precision of estimates are affected by these misspecifications, but when some very similar effects exchanged with each other, or the effect with poor power of significance test in model A deleted in model B, the estimates are not affected significantly.

For further study, the normality problems of estimators in these models found in this project may be interesting. There are still many other issues about the models' robustness worth to study e.g. how these estimates will be affected when changing the actor-oriented network dynamics model into tie-oriented network dynamics model or when some observations errors are added in data used to estimate.

Bibliography

Hanneman, Robert A. and Mark Riddle. 2005. *Introduction to social network methods.* Riverside, CA: University of California, Riverside

Snijders, T.A.B., 2006. Statistical Methods for Network Dynamics. In: S.R. Luchini et al. (eds.), *Proceedings of the XLIII Scientific Meeting*, Italian Statistical Society, pp. 281–296. Padova: CLEUP.

Van Duijn, M.A.J., E.P.H. Zeggelink, M. Huisman, F.N. Stokman, and F.W.Wasseur. 2003. Evolution of Sociology Freshmen into a Friendship Network. *Journal of Mathematical Sociology* 27, 153–191.

Snijders, T.A.B. 2005. Models for Longitudinal Network Data. Chapter 11 in P. Carrington, J. Scott, and S. Wasserman (Eds.), *Models and methods in social network analysis*. New York: Cambridge University Press.

Burk, William J., Christian E.G. Steglich, and Tom A.B. Snijders. 2007. Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors. *International Journal of Behavioral Development*, in press.

Pearson, M., Ch. Steglich & T.A.B.Snijders. Homophily and assimilation among sport-active adolescent substance users. *Connections* 27(1) (2006), 47-63.

Snijders, Tom A.B., Steglich, Christian E.G., and Schweinberger, Michael. 2007. Modeling the coevolution of networks and behavior. In: *Longitudinal models in the behavioral and related sciences*, edited by Kees van Montfort, Han Oud and Albert Satorra, pp. 41–71. Mahwah, NJ: Lawrence Erlbaum.

The *SIENA* webpage:

http://stat.gamma.rug.nl/siena.html

The *StOCNET webpage:*

http://stat.gamma.rug.nl/stocnet/

Pearson, M., and P.West. 2003. Drifting Smoke Rings: Social Network Analysis and Markov Processes in a Longitudinal Study of Friendship Groups and Risk-Taking. *Connections*, 25(2), 59–76.

Snijders, T.A.B. 2001. *The statistical evaluation of social network dynamics*. Pp. 361-395 in Sociological Methodology – 2001, edited by M.E. Sobel and M.P. Becker. Boston and London: Basil Blackwell.

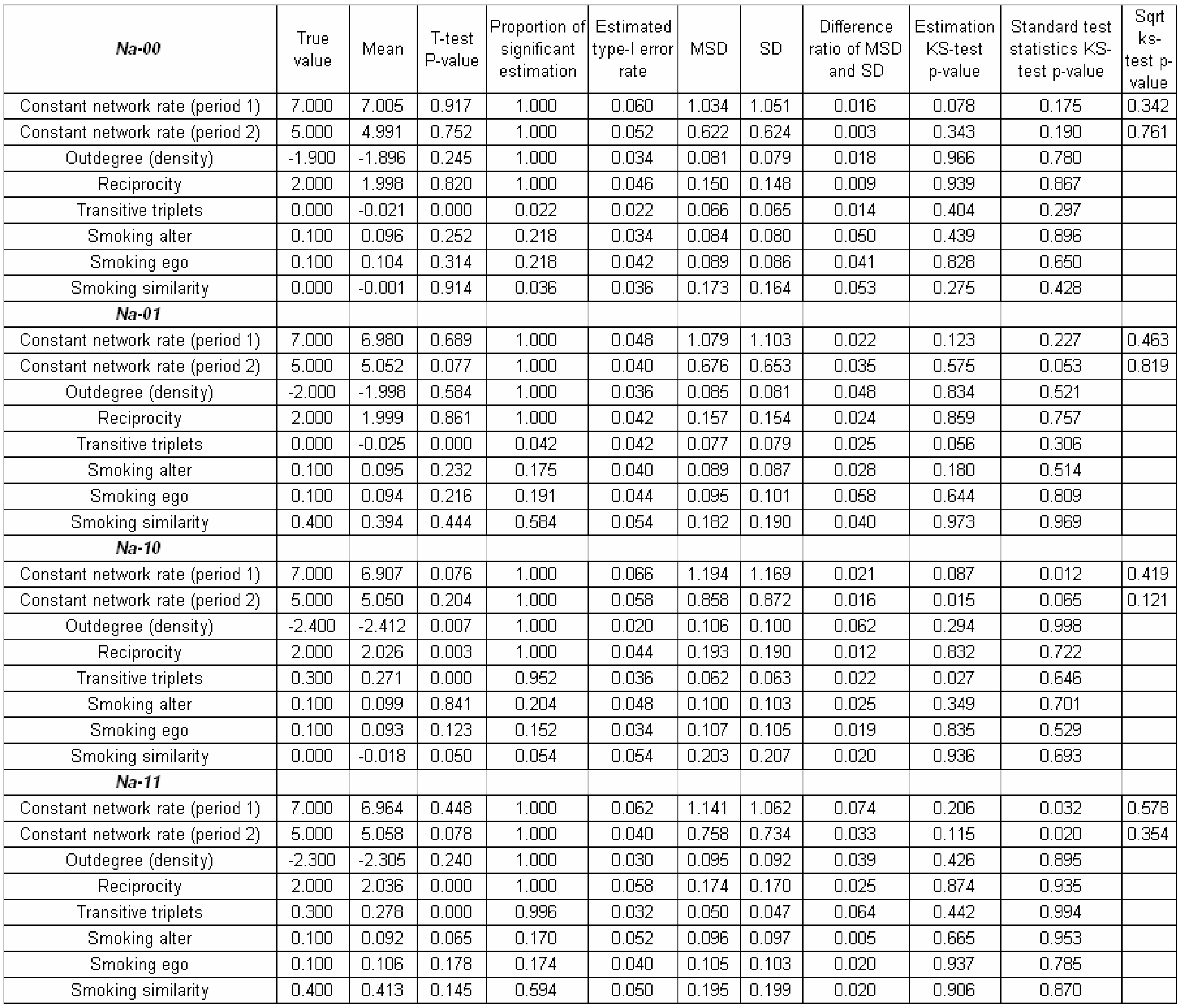
Sidney I. Resnick, 1992, *Adventures in Stochastic Processes*, Birkhauser Verlag AG

Snijders, Tom A.B., Christian E.G. Steglich, Michael Schweinberger, and Mark Huisman. 2007. *Manual for SIENA version 3.11*. Groningen: University of Groningen, ICS. Oxford: University of Oxford, Department of Statistics. http://stat.gamma.rug.nl/stocnet

Snijders, TAB. 1996. Stochastic Actor-Oriented Models for Network Change. *Journal of Mathematical Sociology* 21, 149-72.

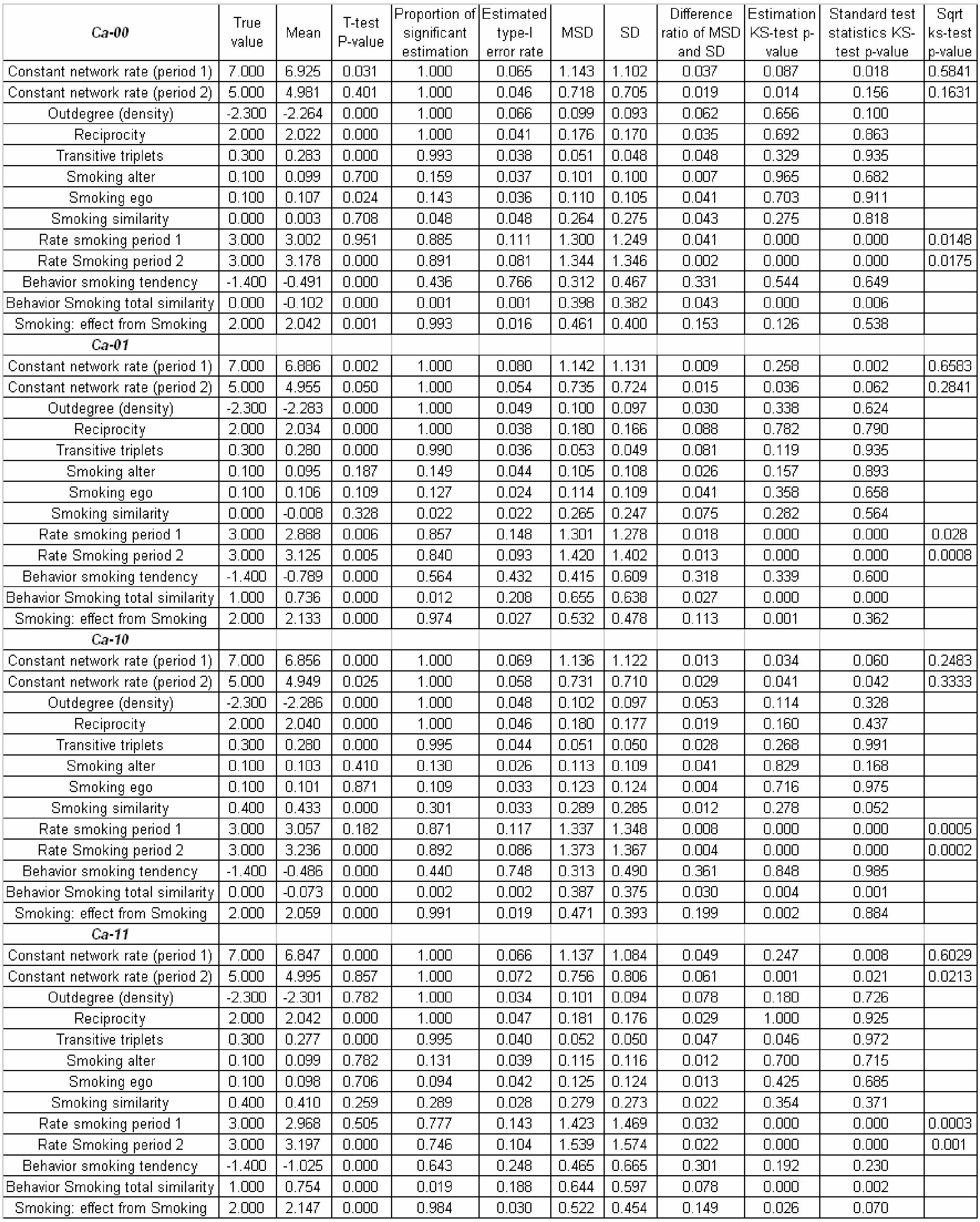
Appendix

Appendix Table 1 Statistical analysis of model *Na* simulations results



Na-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate	MSD	SD	Difference ratio of MSD and SD	Estimation KS-test p-value	Standard test statistics KS-test p-value	Sqrt ks-test p-value
Constant network rate (period 1)	7.000	7.005	0.917	1.000	0.060	1.034	1.051	0.016	0.078	0.175	0.342
Constant network rate (period 2)	5.000	4.991	0.752	1.000	0.052	0.622	0.624	0.003	0.343	0.190	0.761
Outdegree (density)	-1.900	-1.896	0.245	1.000	0.034	0.081	0.079	0.018	0.966	0.780	
Reciprocity	2.000	1.998	0.820	1.000	0.046	0.150	0.148	0.009	0.939	0.867	
Transitive triplets	0.000	-0.021	0.000	0.022	0.022	0.066	0.065	0.014	0.404	0.297	
Smoking alter	0.100	0.096	0.252	0.218	0.034	0.084	0.080	0.050	0.439	0.896	
Smoking ego	0.100	0.104	0.314	0.218	0.042	0.089	0.086	0.041	0.828	0.650	
Smoking similarity	0.000	-0.001	0.914	0.036	0.036	0.173	0.164	0.053	0.275	0.428	
Na-01											
Constant network rate (period 1)	7.000	6.980	0.689	1.000	0.048	1.079	1.103	0.022	0.123	0.227	0.463
Constant network rate (period 2)	5.000	5.052	0.077	1.000	0.040	0.676	0.653	0.035	0.575	0.053	0.819
Outdegree (density)	-2.000	-1.998	0.584	1.000	0.036	0.085	0.081	0.048	0.834	0.521	
Reciprocity	2.000	1.999	0.861	1.000	0.042	0.157	0.154	0.024	0.859	0.757	
Transitive triplets	0.000	-0.025	0.000	0.042	0.042	0.077	0.079	0.025	0.056	0.306	
Smoking alter	0.100	0.095	0.232	0.175	0.040	0.089	0.087	0.028	0.180	0.514	
Smoking ego	0.100	0.094	0.216	0.191	0.044	0.095	0.101	0.058	0.644	0.809	
Smoking similarity	0.400	0.394	0.444	0.584	0.054	0.182	0.190	0.040	0.973	0.969	
Na-10											
Constant network rate (period 1)	7.000	6.907	0.076	1.000	0.066	1.194	1.169	0.021	0.087	0.012	0.419
Constant network rate (period 2)	5.000	5.050	0.204	1.000	0.058	0.858	0.872	0.016	0.015	0.065	0.121
Outdegree (density)	-2.400	-2.412	0.007	1.000	0.020	0.106	0.100	0.062	0.294	0.998	
Reciprocity	2.000	2.026	0.003	1.000	0.044	0.193	0.190	0.012	0.832	0.722	
Transitive triplets	0.300	0.271	0.000	0.952	0.036	0.062	0.063	0.022	0.027	0.646	
Smoking alter	0.100	0.099	0.841	0.204	0.048	0.100	0.103	0.025	0.349	0.701	
Smoking ego	0.100	0.093	0.123	0.152	0.034	0.107	0.105	0.019	0.835	0.529	
Smoking similarity	0.000	-0.018	0.050	0.054	0.054	0.203	0.207	0.020	0.936	0.693	
Na-11											
Constant network rate (period 1)	7.000	6.964	0.448	1.000	0.062	1.141	1.062	0.074	0.206	0.032	0.578
Constant network rate (period 2)	5.000	5.058	0.078	1.000	0.040	0.758	0.734	0.033	0.115	0.020	0.354
Outdegree (density)	-2.300	-2.305	0.240	1.000	0.030	0.095	0.092	0.039	0.426	0.895	
Reciprocity	2.000	2.036	0.000	1.000	0.058	0.174	0.170	0.025	0.874	0.935	
Transitive triplets	0.300	0.278	0.000	0.996	0.032	0.050	0.047	0.064	0.442	0.994	
Smoking alter	0.100	0.092	0.065	0.170	0.052	0.096	0.097	0.005	0.665	0.953	
Smoking ego	0.100	0.106	0.178	0.174	0.040	0.105	0.103	0.020	0.937	0.785	
Smoking similarity	0.400	0.413	0.145	0.594	0.050	0.195	0.199	0.020	0.906	0.870	

Appendix Table 2 Statistical analysis of model *Ca* simulations results

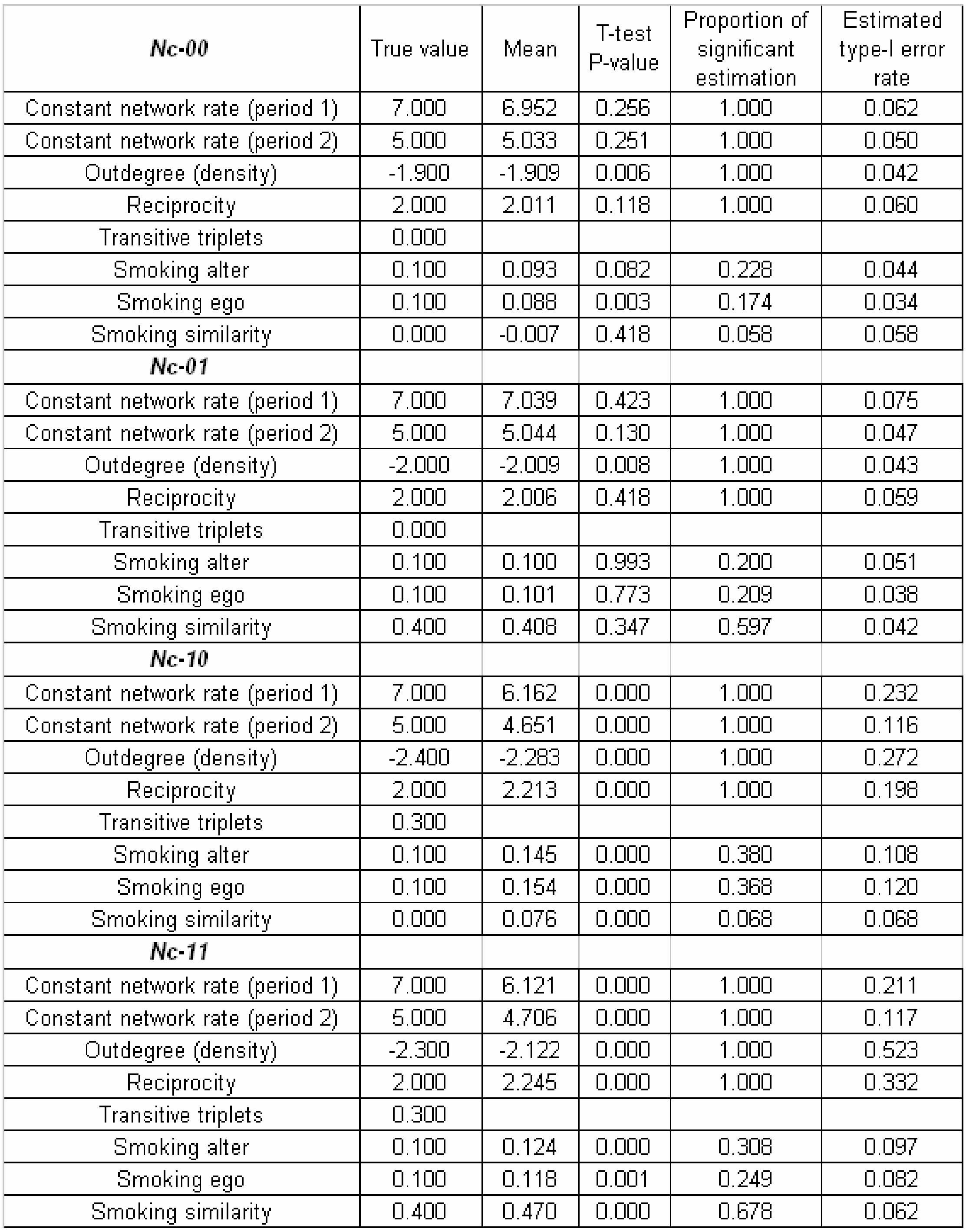


Ca-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate	MSD	SD	Difference ratio of MSD and SD	Estimation KS-test p-value	Standard test statistics KS-test p-value	Sqrt ks-test p-value
Constant network rate (period 1)	7.000	6.925	0.031	1.000	0.065	1.143	1.102	0.037	0.087	0.018	0.5841
Constant network rate (period 2)	5.000	4.981	0.401	1.000	0.046	0.718	0.705	0.019	0.014	0.156	0.1631
Outdegree (density)	-2.300	-2.264	0.000	1.000	0.066	0.099	0.093	0.062	0.656	0.100	
Reciprocity	2.000	2.022	0.000	1.000	0.041	0.176	0.170	0.035	0.692	0.863	
Transitive triplets	0.300	0.283	0.000	0.993	0.038	0.051	0.048	0.048	0.329	0.935	
Smoking alter	0.100	0.099	0.700	0.159	0.037	0.101	0.100	0.007	0.965	0.682	
Smoking ego	0.100	0.107	0.024	0.143	0.036	0.110	0.105	0.041	0.703	0.911	
Smoking similarity	0.000	0.003	0.708	0.048	0.048	0.264	0.275	0.043	0.275	0.818	
Rate smoking period 1	3.000	3.002	0.951	0.885	0.111	1.300	1.249	0.041	0.000	0.000	0.0148
Rate Smoking period 2	3.000	3.178	0.000	0.891	0.081	1.344	1.346	0.002	0.000	0.000	0.0175
Behavior smoking tendency	-1.400	-0.491	0.000	0.436	0.766	0.312	0.467	0.331	0.544	0.649	
Behavior Smoking total similarity	0.000	-0.102	0.000	0.001	0.001	0.398	0.382	0.043	0.000	0.006	
Smoking: effect from Smoking	2.000	2.042	0.001	0.993	0.016	0.461	0.400	0.153	0.126	0.538	
Ca-01											
Constant network rate (period 1)	7.000	6.886	0.002	1.000	0.080	1.142	1.131	0.009	0.258	0.002	0.6583
Constant network rate (period 2)	5.000	4.955	0.050	1.000	0.054	0.735	0.724	0.015	0.036	0.062	0.2841
Outdegree (density)	-2.300	-2.283	0.000	1.000	0.049	0.100	0.097	0.030	0.338	0.624	
Reciprocity	2.000	2.034	0.000	1.000	0.038	0.180	0.166	0.088	0.782	0.790	
Transitive triplets	0.300	0.280	0.000	0.990	0.036	0.053	0.049	0.081	0.119	0.935	
Smoking alter	0.100	0.095	0.187	0.149	0.044	0.105	0.108	0.026	0.157	0.893	
Smoking ego	0.100	0.106	0.109	0.127	0.024	0.114	0.109	0.041	0.358	0.658	
Smoking similarity	0.000	-0.008	0.328	0.022	0.022	0.265	0.247	0.075	0.282	0.564	
Rate smoking period 1	3.000	2.888	0.006	0.857	0.148	1.301	1.278	0.018	0.000	0.000	0.028
Rate Smoking period 2	3.000	3.125	0.005	0.840	0.093	1.420	1.402	0.013	0.000	0.000	0.0008
Behavior smoking tendency	-1.400	-0.789	0.000	0.564	0.432	0.415	0.609	0.318	0.339	0.600	
Behavior Smoking total similarity	1.000	0.736	0.000	0.012	0.208	0.655	0.638	0.027	0.000	0.000	
Smoking: effect from Smoking	2.000	2.133	0.000	0.974	0.027	0.532	0.478	0.113	0.001	0.362	
Ca-10											
Constant network rate (period 1)	7.000	6.856	0.000	1.000	0.069	1.136	1.122	0.013	0.034	0.060	0.2483
Constant network rate (period 2)	5.000	4.949	0.025	1.000	0.058	0.731	0.710	0.029	0.041	0.042	0.3333
Outdegree (density)	-2.300	-2.286	0.000	1.000	0.048	0.102	0.097	0.053	0.114	0.328	
Reciprocity	2.000	2.040	0.000	1.000	0.046	0.180	0.177	0.019	0.160	0.437	
Transitive triplets	0.300	0.280	0.000	0.995	0.044	0.051	0.050	0.028	0.268	0.991	
Smoking alter	0.100	0.103	0.410	0.130	0.026	0.113	0.109	0.041	0.829	0.168	
Smoking ego	0.100	0.101	0.871	0.109	0.033	0.123	0.124	0.004	0.716	0.975	
Smoking similarity	0.400	0.433	0.000	0.301	0.033	0.289	0.285	0.012	0.278	0.052	
Rate smoking period 1	3.000	3.057	0.182	0.871	0.117	1.337	1.348	0.008	0.000	0.000	0.0005
Rate Smoking period 2	3.000	3.236	0.000	0.892	0.086	1.373	1.367	0.004	0.000	0.000	0.0002
Behavior smoking tendency	-1.400	-0.486	0.000	0.440	0.748	0.313	0.490	0.361	0.848	0.985	
Behavior Smoking total similarity	0.000	-0.073	0.000	0.002	0.002	0.387	0.375	0.030	0.004	0.001	
Smoking: effect from Smoking	2.000	2.059	0.000	0.991	0.019	0.471	0.393	0.199	0.002	0.884	
Ca-11											
Constant network rate (period 1)	7.000	6.847	0.000	1.000	0.066	1.137	1.084	0.049	0.247	0.008	0.6029
Constant network rate (period 2)	5.000	4.995	0.857	1.000	0.072	0.756	0.806	0.061	0.001	0.021	0.0213
Outdegree (density)	-2.300	-2.301	0.782	1.000	0.034	0.101	0.094	0.078	0.180	0.726	
Reciprocity	2.000	2.042	0.000	1.000	0.047	0.181	0.176	0.029	1.000	0.925	
Transitive triplets	0.300	0.277	0.000	0.995	0.040	0.052	0.050	0.047	0.046	0.972	
Smoking alter	0.100	0.099	0.782	0.131	0.039	0.115	0.116	0.012	0.700	0.715	
Smoking ego	0.100	0.098	0.706	0.094	0.042	0.125	0.124	0.013	0.425	0.685	
Smoking similarity	0.400	0.410	0.259	0.289	0.028	0.279	0.273	0.022	0.354	0.371	
Rate smoking period 1	3.000	2.968	0.505	0.777	0.143	1.423	1.469	0.032	0.000	0.000	0.0003
Rate Smoking period 2	3.000	3.197	0.000	0.746	0.104	1.539	1.574	0.022	0.000	0.000	0.001
Behavior smoking tendency	-1.400	-1.025	0.000	0.643	0.248	0.465	0.665	0.301	0.192	0.230	
Behavior Smoking total similarity	1.000	0.754	0.000	0.019	0.188	0.644	0.597	0.078	0.000	0.002	
Smoking: effect from Smoking	2.000	2.147	0.000	0.984	0.030	0.522	0.454	0.149	0.026	0.070	

Appendix Table 3 Statistical analysis of model *Nb* simulations results

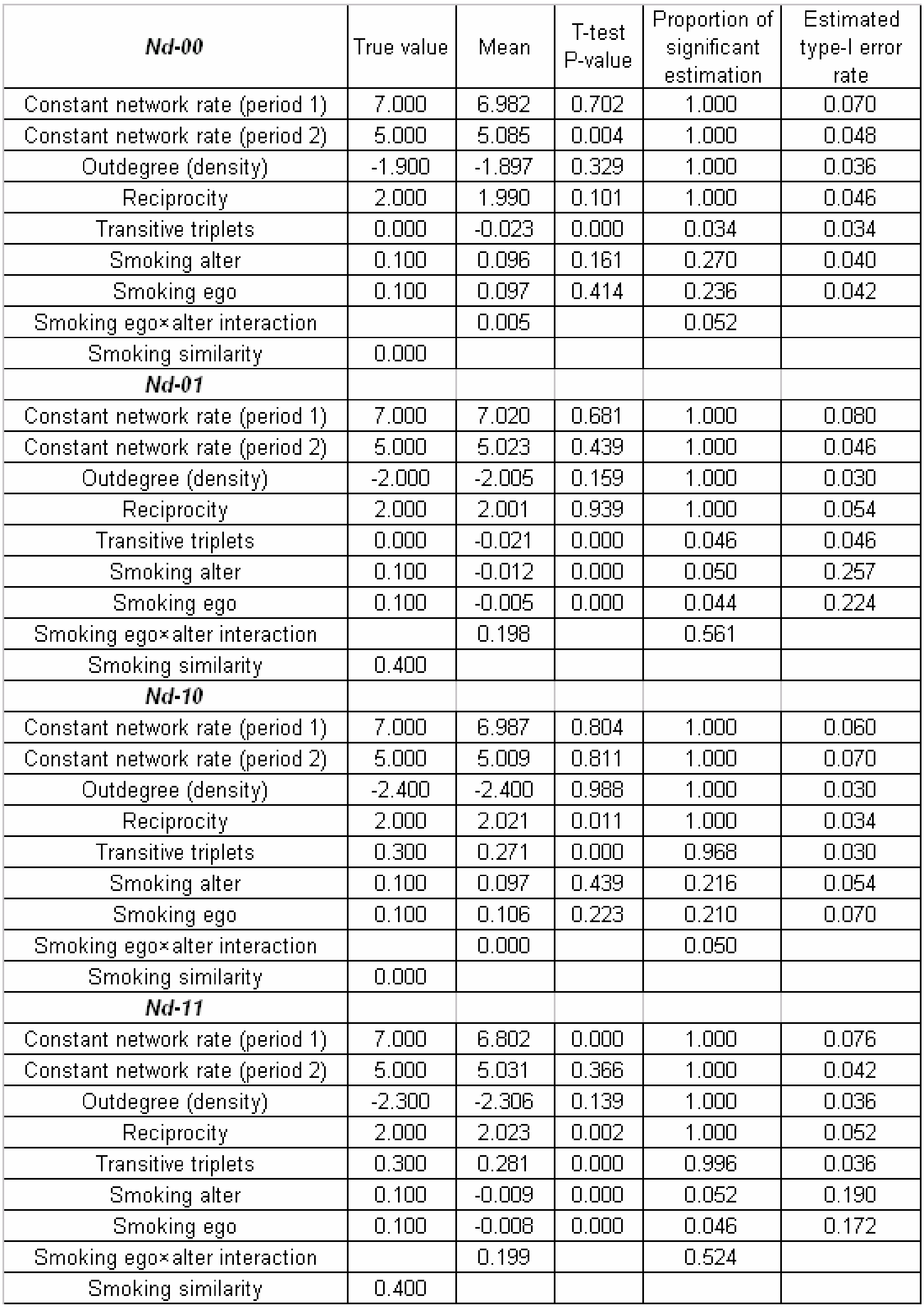
Nb-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	7.038	0.394	1.000	0.054
Constant network rate (period 2)	5.000	5.045	0.134	1.000	0.050
Outdegree (density)	-1.900	-1.911	0.027	1.000	0.022
Reciprocity	2.000	2.009	0.161	1.000	0.048
Transitive triplets	0.000				
Number of actors at distance 2		0.000		0.016	
Smoking alter	0.100	0.097	0.434	0.248	0.040
Smoking ego	0.100	0.094	0.099	0.156	0.034
Smoking similarity	0.000	0.002	0.807	0.032	0.032
Nb-01					
Constant network rate (period 1)	7.000	7.072	0.127	1.000	0.060
Constant network rate (period 2)	5.000	5.102	0.001	1.000	0.044
Outdegree (density)	-2.000	-1.994	0.278	1.000	0.048
Reciprocity	2.000	2.005	0.486	1.000	0.048
Transitive triplets	0.000				
Number of actors at distance 2		-0.022		0.058	
Smoking alter	0.100	0.093	0.065	0.204	0.034
Smoking ego	0.100	0.094	0.197	0.162	0.050
Smoking similarity	0.400	0.394	0.463	0.585	0.062
Nb-10					
Constant network rate (period 1)	7.000	6.415	0.000	1.000	0.150
Constant network rate (period 2)	5.000	4.755	0.000	1.000	0.090
Outdegree (density)	-2.400	-2.166	0.000	1.000	0.440
Reciprocity	2.000	2.172	0.000	1.000	0.150
Transitive triplets	0.300				
Number of actors at distance 2		-0.164		0.300	
Smoking alter	0.100	0.153	0.000	0.408	0.164
Smoking ego	0.100	0.133	0.000	0.268	0.078
Smoking similarity	0.000	0.048	0.000	0.044	0.044
Nb-11					
Constant network rate (period 1)	7.000	6.492	0.000	1.000	0.146
Constant network rate (period 2)	5.000	4.807	0.000	1.000	0.062
Outdegree (density)	-2.300	-1.963	0.000	1.000	0.722
Reciprocity	2.000	2.194	0.000	1.000	0.188
Transitive triplets	0.300				
Number of actors at distance 2		-0.226		0.534	
Smoking alter	0.100	0.115	0.003	0.302	0.120
Smoking ego	0.100	0.110	0.037	0.212	0.054
Smoking similarity	0.400	0.444	0.000	0.648	0.062

Appendix Table 4 Statistical analysis of model *Nc* simulations results



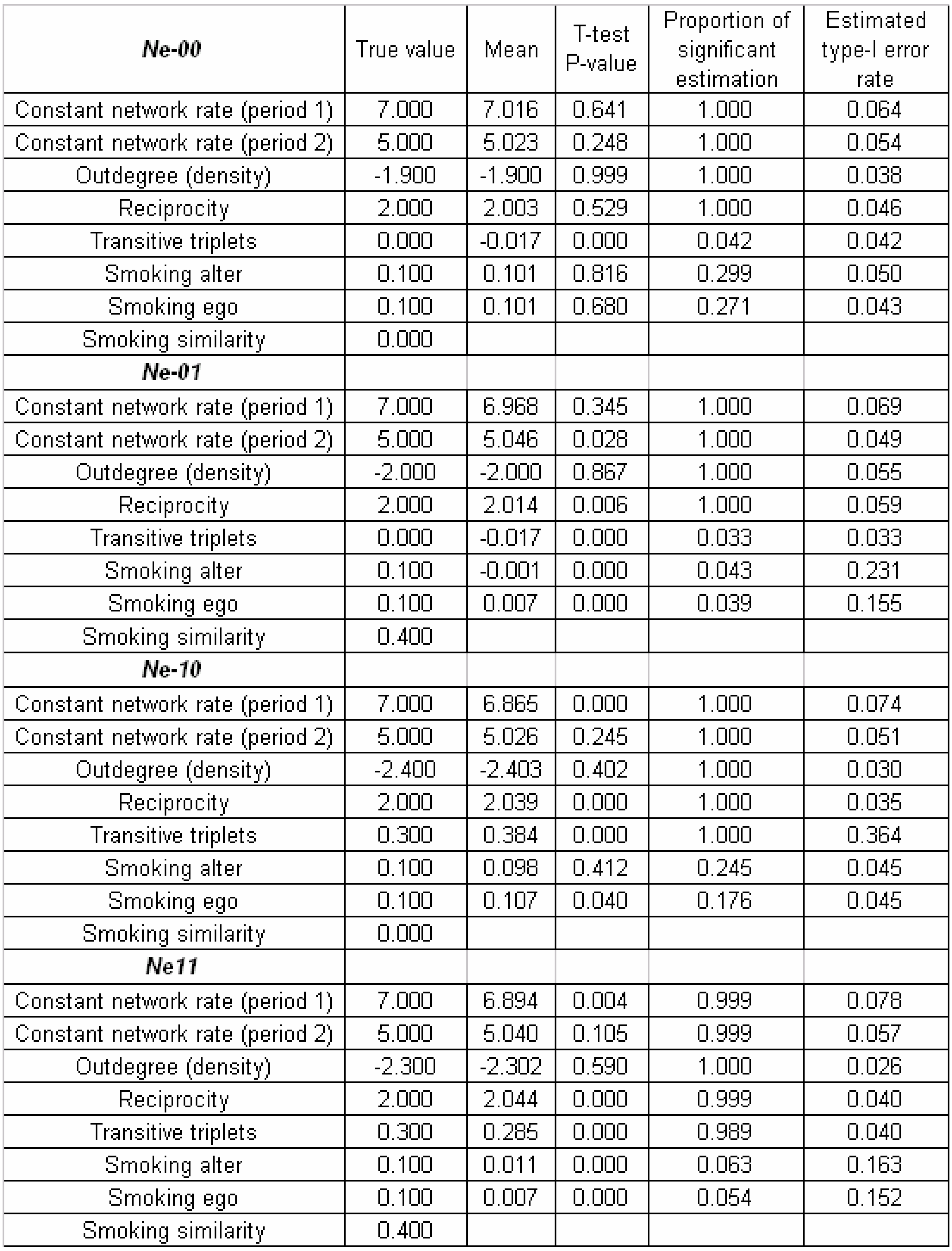
Nc-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.952	0.256	1.000	0.062
Constant network rate (period 2)	5.000	5.033	0.251	1.000	0.050
Outdegree (density)	-1.900	-1.909	0.006	1.000	0.042
Reciprocity	2.000	2.011	0.118	1.000	0.060
Transitive triplets	0.000				
Smoking alter	0.100	0.093	0.082	0.228	0.044
Smoking ego	0.100	0.088	0.003	0.174	0.034
Smoking similarity	0.000	-0.007	0.418	0.058	0.058
Nc-01					
Constant network rate (period 1)	7.000	7.039	0.423	1.000	0.075
Constant network rate (period 2)	5.000	5.044	0.130	1.000	0.047
Outdegree (density)	-2.000	-2.009	0.008	1.000	0.043
Reciprocity	2.000	2.006	0.418	1.000	0.059
Transitive triplets	0.000				
Smoking alter	0.100	0.100	0.993	0.200	0.051
Smoking ego	0.100	0.101	0.773	0.209	0.038
Smoking similarity	0.400	0.408	0.347	0.597	0.042
Nc-10					
Constant network rate (period 1)	7.000	6.162	0.000	1.000	0.232
Constant network rate (period 2)	5.000	4.651	0.000	1.000	0.116
Outdegree (density)	-2.400	-2.283	0.000	1.000	0.272
Reciprocity	2.000	2.213	0.000	1.000	0.198
Transitive triplets	0.300				
Smoking alter	0.100	0.145	0.000	0.380	0.108
Smoking ego	0.100	0.154	0.000	0.368	0.120
Smoking similarity	0.000	0.076	0.000	0.068	0.068
Nc-11					
Constant network rate (period 1)	7.000	6.121	0.000	1.000	0.211
Constant network rate (period 2)	5.000	4.706	0.000	1.000	0.117
Outdegree (density)	-2.300	-2.122	0.000	1.000	0.523
Reciprocity	2.000	2.245	0.000	1.000	0.332
Transitive triplets	0.300				
Smoking alter	0.100	0.124	0.000	0.308	0.097
Smoking ego	0.100	0.118	0.001	0.249	0.082
Smoking similarity	0.400	0.470	0.000	0.678	0.062

Appendix Table 5 Statistical analysis of model *Nd* simulations results



Nd-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.982	0.702	1.000	0.070
Constant network rate (period 2)	5.000	5.085	0.004	1.000	0.048
Outdegree (density)	-1.900	-1.897	0.329	1.000	0.036
Reciprocity	2.000	1.990	0.101	1.000	0.046
Transitive triplets	0.000	-0.023	0.000	0.034	0.034
Smoking alter	0.100	0.096	0.161	0.270	0.040
Smoking ego	0.100	0.097	0.414	0.236	0.042
Smoking ego×alter interaction		0.005		0.052	
Smoking similarity	0.000				
Nd-01					
Constant network rate (period 1)	7.000	7.020	0.681	1.000	0.080
Constant network rate (period 2)	5.000	5.023	0.439	1.000	0.046
Outdegree (density)	-2.000	-2.005	0.159	1.000	0.030
Reciprocity	2.000	2.001	0.939	1.000	0.054
Transitive triplets	0.000	-0.021	0.000	0.046	0.046
Smoking alter	0.100	-0.012	0.000	0.050	0.257
Smoking ego	0.100	-0.005	0.000	0.044	0.224
Smoking ego×alter interaction		0.198		0.561	
Smoking similarity	0.400				
Nd-10					
Constant network rate (period 1)	7.000	6.987	0.804	1.000	0.060
Constant network rate (period 2)	5.000	5.009	0.811	1.000	0.070
Outdegree (density)	-2.400	-2.400	0.988	1.000	0.030
Reciprocity	2.000	2.021	0.011	1.000	0.034
Transitive triplets	0.300	0.271	0.000	0.968	0.030
Smoking alter	0.100	0.097	0.439	0.216	0.054
Smoking ego	0.100	0.106	0.223	0.210	0.070
Smoking ego×alter interaction		0.000		0.050	
Smoking similarity	0.000				
Nd-11					
Constant network rate (period 1)	7.000	6.802	0.000	1.000	0.076
Constant network rate (period 2)	5.000	5.031	0.366	1.000	0.042
Outdegree (density)	-2.300	-2.306	0.139	1.000	0.036
Reciprocity	2.000	2.023	0.002	1.000	0.052
Transitive triplets	0.300	0.281	0.000	0.996	0.036
Smoking alter	0.100	-0.009	0.000	0.052	0.190
Smoking ego	0.100	-0.008	0.000	0.046	0.172
Smoking ego×alter interaction		0.199		0.524	
Smoking similarity	0.400				

Appendix Table 6 Statistical analysis of model *Ne* simulations results

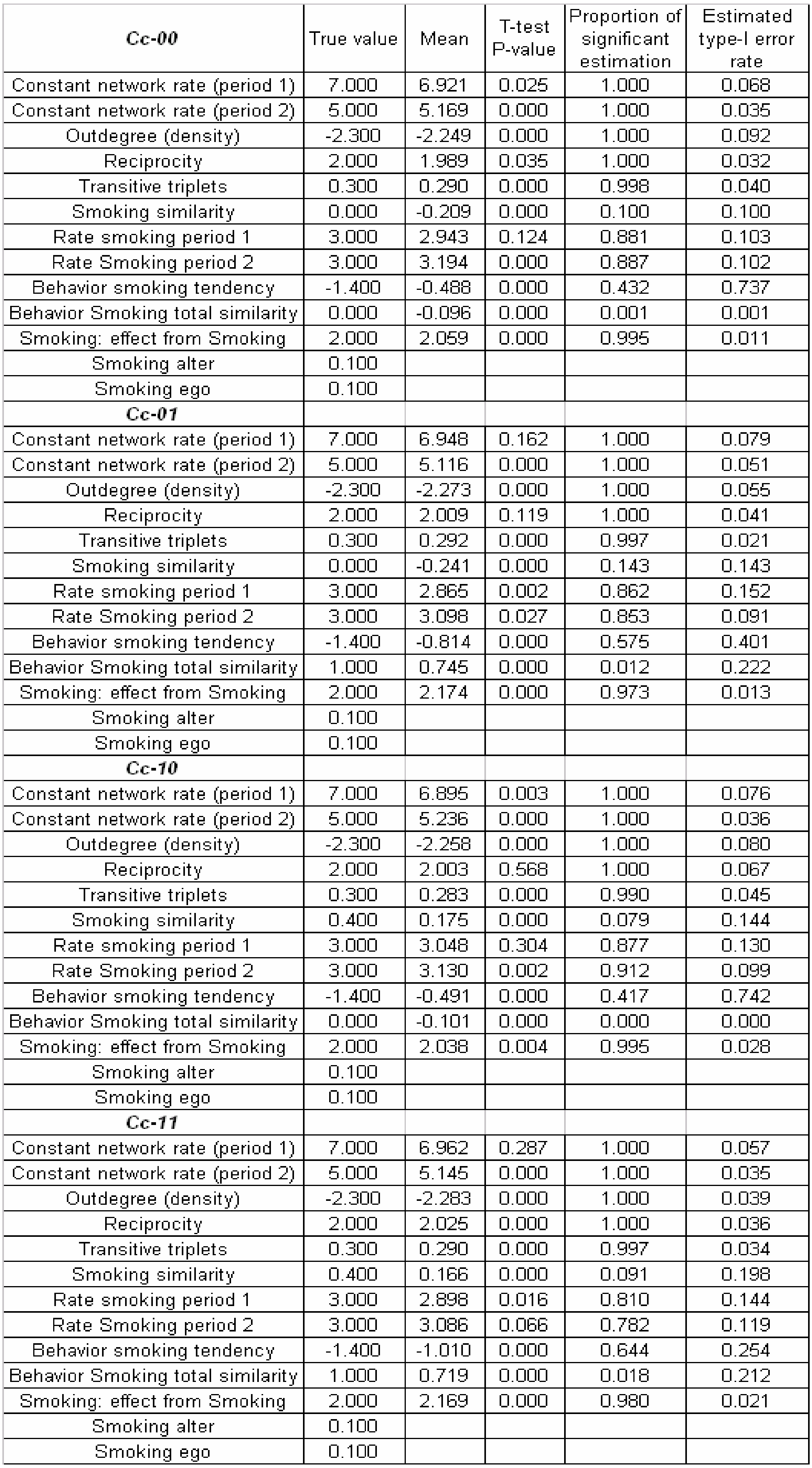


Ne-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	7.016	0.641	1.000	0.064
Constant network rate (period 2)	5.000	5.023	0.248	1.000	0.054
Outdegree (density)	-1.900	-1.900	0.999	1.000	0.038
Reciprocity	2.000	2.003	0.529	1.000	0.046
Transitive triplets	0.000	-0.017	0.000	0.042	0.042
Smoking alter	0.100	0.101	0.816	0.299	0.050
Smoking ego	0.100	0.101	0.680	0.271	0.043
Smoking similarity	0.000				
Ne-01					
Constant network rate (period 1)	7.000	6.968	0.345	1.000	0.069
Constant network rate (period 2)	5.000	5.046	0.028	1.000	0.049
Outdegree (density)	-2.000	-2.000	0.867	1.000	0.055
Reciprocity	2.000	2.014	0.006	1.000	0.059
Transitive triplets	0.000	-0.017	0.000	0.033	0.033
Smoking alter	0.100	-0.001	0.000	0.043	0.231
Smoking ego	0.100	0.007	0.000	0.039	0.155
Smoking similarity	0.400				
Ne-10					
Constant network rate (period 1)	7.000	6.865	0.000	1.000	0.074
Constant network rate (period 2)	5.000	5.026	0.245	1.000	0.051
Outdegree (density)	-2.400	-2.403	0.402	1.000	0.030
Reciprocity	2.000	2.039	0.000	1.000	0.035
Transitive triplets	0.300	0.384	0.000	1.000	0.364
Smoking alter	0.100	0.098	0.412	0.245	0.045
Smoking ego	0.100	0.107	0.040	0.176	0.045
Smoking similarity	0.000				
Ne11					
Constant network rate (period 1)	7.000	6.894	0.004	0.999	0.078
Constant network rate (period 2)	5.000	5.040	0.105	0.999	0.057
Outdegree (density)	-2.300	-2.302	0.590	1.000	0.026
Reciprocity	2.000	2.044	0.000	0.999	0.040
Transitive triplets	0.300	0.285	0.000	0.989	0.040
Smoking alter	0.100	0.011	0.000	0.063	0.163
Smoking ego	0.100	0.007	0.000	0.054	0.152
Smoking similarity	0.400				

Appendix Table 7 Statistical analysis of model *Cb* simulations results

Cb-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.850	0.000	1.000	0.080
Constant network rate (period 2)	5.000	4.998	0.947	1.000	0.056
Outdegree (density)	-2.300	-2.270	0.000	1.000	0.054
Reciprocity	2.000	2.036	0.000	1.000	0.041
Transitive triplets	0.300	0.281	0.000	0.994	0.050
Smoking alter	0.100	0.102	0.461	0.153	0.033
Smoking ego	0.100	0.104	0.281	0.135	0.031
Smoking similarity	0.000	-0.012	0.153	0.025	0.025
Rate smoking period 1	3.000	3.033	0.462	0.862	0.107
Rate Smoking period 2	3.000	3.240	0.000	0.893	0.084
Behavior smoking tendency	-1.400	-0.494	0.000	0.444	0.727
Behavior Smoking average similarity		-0.279		0.001	
Smoking: effect from Smoking	2.000	2.084	0.000	0.990	0.013
Behavior Smoking total similarity	0.000				
Cb-01					
Constant network rate (period 1)	7.000	6.827	0.000	1.000	0.081
Constant network rate (period 2)	5.000	4.983	0.467	1.000	0.064
Outdegree (density)	-2.300	-2.285	0.000	1.000	0.042
Reciprocity	2.000	2.021	0.000	1.000	0.045
Transitive triplets	0.300	0.280	0.000	0.984	0.038
Smoking alter	0.100	0.097	0.338	0.130	0.018
Smoking ego	0.100	0.102	0.586	0.119	0.034
Smoking similarity	0.000	-0.011	0.167	0.024	0.024
Rate smoking period 1	3.000	2.855	0.000	0.838	0.136
Rate Smoking period 2	3.000	3.114	0.006	0.855	0.081
Behavior smoking tendency	-1.400	-0.864	0.000	0.598	0.400
Behavior Smoking average similarity		1.750		0.017	
Smoking: effect from Smoking	2.000	2.159	0.000	0.987	0.026
Behavior Smoking total similarity	1.000				
Cb-10					
Constant network rate (period 1)	7.000	6.879	0.000	1.000	0.069
Constant network rate (period 2)	5.000	4.986	0.564	1.000	0.062
Outdegree (density)	-2.300	-2.289	0.001	1.000	0.050
Reciprocity	2.000	2.044	0.000	1.000	0.039
Transitive triplets	0.300	0.278	0.000	0.993	0.046
Smoking alter	0.100	0.106	0.081	0.132	0.035
Smoking ego	0.100	0.099	0.823	0.103	0.033
Smoking similarity	0.400	0.417	0.053	0.291	0.035
Rate smoking period 1	3.000	3.056	0.190	0.868	0.101
Rate Smoking period 2	3.000	3.169	0.000	0.895	0.089
Behavior smoking tendency	-1.400	-0.480	0.000	0.425	0.751
Behavior Smoking average similarity		-0.236		0.001	
Smoking: effect from Smoking	2.000	2.052	0.000	0.993	0.025
Behavior Smoking total similarity	0.000				
Cb-11					
Constant network rate (period 1)	7.000	6.858	0.000	1.000	0.068
Constant network rate (period 2)	5.000	4.948	0.033	1.000	0.074
Outdegree (density)	-2.300	-2.305	0.081	1.000	0.033
Reciprocity	2.000	2.041	0.000	1.000	0.044
Transitive triplets	0.300	0.279	0.000	0.991	0.043
Smoking alter	0.100	0.100	0.927	0.138	0.037
Smoking ego	0.100	0.097	0.447	0.088	0.035
Smoking similarity	0.400	0.404	0.611	0.288	0.029
Rate smoking period 1	3.000	3.024	0.614	0.792	0.130
Rate Smoking period 2	3.000	3.170	0.001	0.770	0.094
Behavior smoking tendency	-1.400	-1.074	0.000	0.666	0.231
Behavior Smoking average similarity		1.755		0.017	
Smoking: effect from Smoking	2.000	2.200	0.000	0.993	0.018
Behavior Smoking total similarity	1.000				

Appendix Table 8 Statistical analysis of model *Cc* simulations results



Cc-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.921	0.025	1.000	0.068
Constant network rate (period 2)	5.000	5.169	0.000	1.000	0.035
Outdegree (density)	-2.300	-2.249	0.000	1.000	0.092
Reciprocity	2.000	1.989	0.035	1.000	0.032
Transitive triplets	0.300	0.290	0.000	0.998	0.040
Smoking similarity	0.000	-0.209	0.000	0.100	0.100
Rate smoking period 1	3.000	2.943	0.124	0.881	0.103
Rate Smoking period 2	3.000	3.194	0.000	0.887	0.102
Behavior smoking tendency	-1.400	-0.488	0.000	0.432	0.737
Behavior Smoking total similarity	0.000	-0.096	0.000	0.001	0.001
Smoking: effect from Smoking	2.000	2.059	0.000	0.995	0.011
Smoking alter	0.100				
Smoking ego	0.100				
Cc-01					
Constant network rate (period 1)	7.000	6.948	0.162	1.000	0.079
Constant network rate (period 2)	5.000	5.116	0.000	1.000	0.051
Outdegree (density)	-2.300	-2.273	0.000	1.000	0.055
Reciprocity	2.000	2.009	0.119	1.000	0.041
Transitive triplets	0.300	0.292	0.000	0.997	0.021
Smoking similarity	0.000	-0.241	0.000	0.143	0.143
Rate smoking period 1	3.000	2.865	0.002	0.862	0.152
Rate Smoking period 2	3.000	3.098	0.027	0.853	0.091
Behavior smoking tendency	-1.400	-0.814	0.000	0.575	0.401
Behavior Smoking total similarity	1.000	0.745	0.000	0.012	0.222
Smoking: effect from Smoking	2.000	2.174	0.000	0.973	0.013
Smoking alter	0.100				
Smoking ego	0.100				
Cc-10					
Constant network rate (period 1)	7.000	6.895	0.003	1.000	0.076
Constant network rate (period 2)	5.000	5.236	0.000	1.000	0.036
Outdegree (density)	-2.300	-2.258	0.000	1.000	0.080
Reciprocity	2.000	2.003	0.568	1.000	0.067
Transitive triplets	0.300	0.283	0.000	0.990	0.045
Smoking similarity	0.400	0.175	0.000	0.079	0.144
Rate smoking period 1	3.000	3.048	0.304	0.877	0.130
Rate Smoking period 2	3.000	3.130	0.002	0.912	0.099
Behavior smoking tendency	-1.400	-0.491	0.000	0.417	0.742
Behavior Smoking total similarity	0.000	-0.101	0.000	0.000	0.000
Smoking: effect from Smoking	2.000	2.038	0.004	0.995	0.028
Smoking alter	0.100				
Smoking ego	0.100				
Cc-11					
Constant network rate (period 1)	7.000	6.962	0.287	1.000	0.057
Constant network rate (period 2)	5.000	5.145	0.000	1.000	0.035
Outdegree (density)	-2.300	-2.283	0.000	1.000	0.039
Reciprocity	2.000	2.025	0.000	1.000	0.036
Transitive triplets	0.300	0.290	0.000	0.997	0.034
Smoking similarity	0.400	0.166	0.000	0.091	0.198
Rate smoking period 1	3.000	2.898	0.016	0.810	0.144
Rate Smoking period 2	3.000	3.086	0.066	0.782	0.119
Behavior smoking tendency	-1.400	-1.010	0.000	0.644	0.254
Behavior Smoking total similarity	1.000	0.719	0.000	0.018	0.212
Smoking: effect from Smoking	2.000	2.169	0.000	0.980	0.021
Smoking alter	0.100				
Smoking ego	0.100				

Appendix Table 9 Statistical analysis of model *Cd* simulations results

Cd-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.941	0.250	1.000	0.075
Constant network rate (period 2)	5.000	4.968	0.309	1.000	0.065
Outdegree (density)	-2.300	-2.267	0.000	1.000	0.063
Reciprocity	2.000	2.031	0.000	1.000	0.028
Transitive triplets	0.300	0.282	0.000	0.996	0.046
Smoking alter	0.100	0.106	0.151	0.179	0.032
Smoking ego	0.100	0.103	0.483	0.133	0.034
Ego x alter interaction		-0.004		0.030	
Rate smoking period 1	3.000	3.116	0.084	0.847	0.115
Rate Smoking period 2	3.000	3.240	0.000	0.885	0.093
Behavior smoking tendency	-1.400	-0.466	0.000	0.433	0.728
Smoking average alter		-0.111		0.000	
Smoking: effect from Smoking	2.000	2.105	0.000	0.990	0.014
Smoking similarity	0.000				
Behavior Smoking total similarity	0.000				
Cd-01					
Constant network rate (period 1)	7.000	6.895	0.003	1.000	0.067
Constant network rate (period 2)	5.000	5.000	0.993	1.000	0.050
Outdegree (density)	-2.300	-2.284	0.000	1.000	0.046
Reciprocity	2.000	2.038	0.000	1.000	0.035
Transitive triplets	0.300	0.273	0.000	0.982	0.048
Smoking alter	0.100	0.104	0.207	0.167	0.032
Smoking ego	0.100	0.103	0.393	0.135	0.043
Ego x alter interaction		0.001		0.028	
Rate smoking period 1	3.000	2.815	0.000	0.853	0.157
Rate Smoking period 2	3.000	3.094	0.038	0.839	0.092
Behavior smoking tendency	-1.400	-0.987	0.000	0.685	0.366
Smoking average alter		1.038		0.013	
Smoking: effect from Smoking	2.000	2.022	0.372	0.979	0.041
Smoking similarity	0.000				
Behavior Smoking total similarity	1.000				
Cd-10					
Constant network rate (period 1)	7.000	6.779	0.000	1.000	0.094
Constant network rate (period 2)	5.000	4.996	0.881	1.000	0.058
Outdegree (density)	-2.300	-2.303	0.297	1.000	0.027
Reciprocity	2.000	2.038	0.000	1.000	0.037
Transitive triplets	0.300	0.279	0.000	0.994	0.029
Smoking alter	0.100	0.046	0.000	0.060	0.059
Smoking ego	0.100	0.058	0.000	0.068	0.039
Ego x alter interaction		0.211		0.294	
Rate smoking period 1	3.000	3.131	0.003	0.849	0.080
Rate Smoking period 2	3.000	3.195	0.000	0.892	0.081
Behavior smoking tendency	-1.400	-0.470	0.000	0.420	0.742
Smoking average alter		-0.133		0.001	
Smoking: effect from Smoking	2.000	2.113	0.000	0.986	0.016
Smoking similarity	0.400				
Behavior Smoking total similarity	0.000				
Cd-11					
Constant network rate (period 1)	7.000	6.777	0.000	1.000	0.098
Constant network rate (period 2)	5.000	4.992	0.728	1.000	0.053
Outdegree (density)	-2.300	-2.307	0.016	1.000	0.023
Reciprocity	2.000	2.026	0.000	1.000	0.036
Transitive triplets	0.300	0.277	0.000	0.992	0.050
Smoking alter	0.100	0.019	0.000	0.043	0.098
Smoking ego	0.100	0.020	0.000	0.036	0.078
Ego x alter interaction		0.211		0.328	
Rate smoking period 1	3.000	2.912	0.042	0.762	0.140
Rate Smoking period 2	3.000	3.123	0.014	0.770	0.122
Behavior smoking tendency	-1.400	-1.249	0.000	0.771	0.225
Smoking average alter		0.989		0.015	
Smoking: effect from Smoking	2.000	2.089	0.000	0.970	0.020
Smoking similarity	0.400				
Behavior Smoking total similarity	1.000				

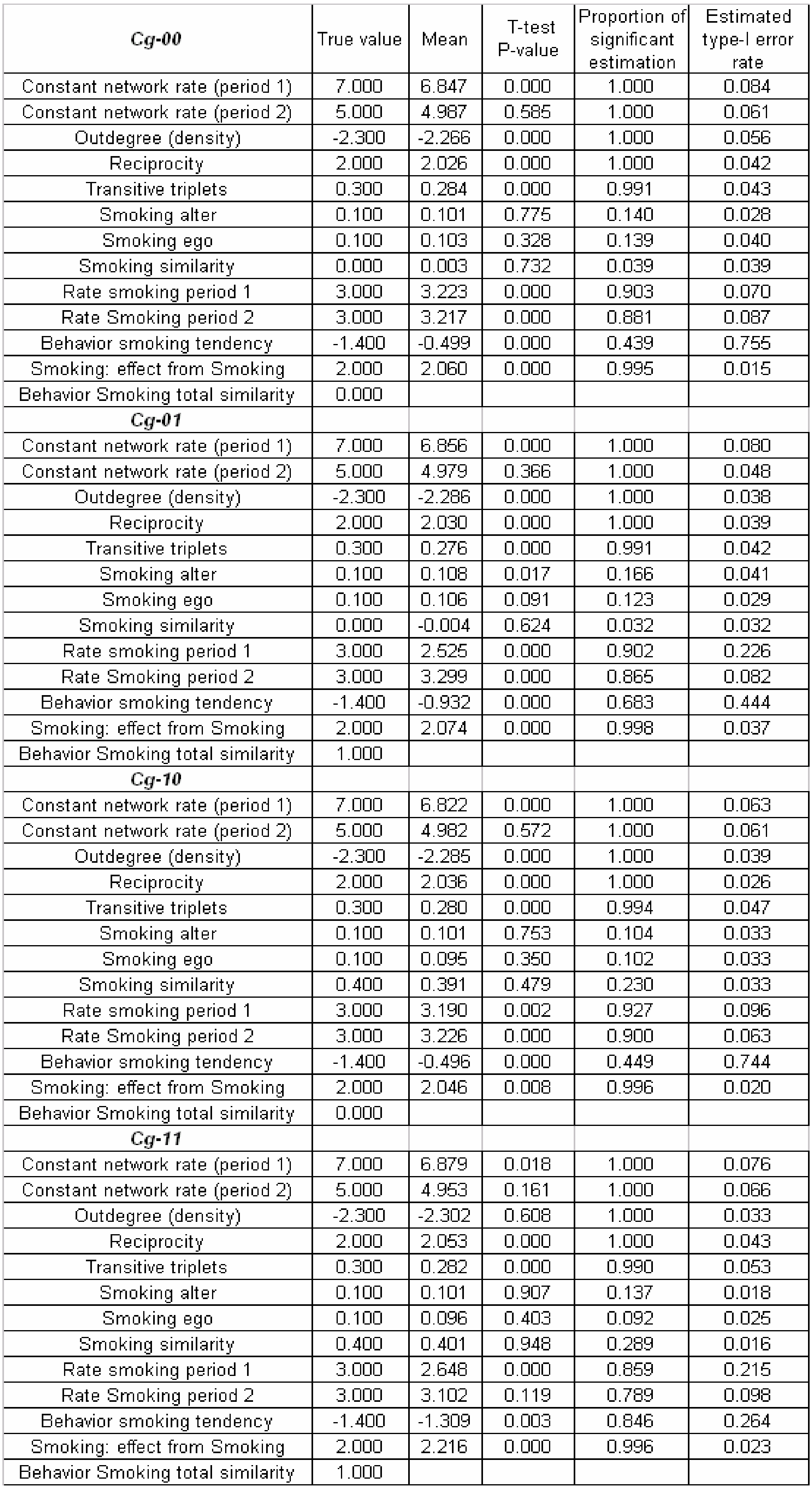
Appendix Table 10 Statistical analysis of model *Ce* simulations results

Ce-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.960	0.439	1.000	0.066
Constant network rate (period 2)	5.000	4.967	0.323	1.000	0.070
Outdegree (density)	-2.300	-2.272	0.000	1.000	0.048
Reciprocity	2.000	2.032	0.000	1.000	0.048
Transitive triplets	0.300	0.283	0.000	0.998	0.055
Smoking alter	0.100	0.101	0.829	0.138	0.053
Smoking ego	0.100	0.113	0.015	0.144	0.044
Smoking similarity	0.000	-0.025	0.052	0.035	0.035
Rate smoking period 1	3.000	0.983	0.000	1.000	0.991
Rate Smoking period 2	3.000	1.159	0.000	1.000	0.937
Behavior smoking tendency	-1.400	0.887	0.000	0.289	0.939
Behavior Smoking total similarity	0.000	-1.152	0.000	0.015	0.015
Smoking: effect from Smoking	2.000				
Ce-01					
Constant network rate (period 1)	7.000	6.862	0.011	1.000	0.060
Constant network rate (period 2)	5.000	5.021	0.550	1.000	0.048
Outdegree (density)	-2.300	-2.279	0.000	1.000	0.055
Reciprocity	2.000	2.017	0.037	1.000	0.034
Transitive triplets	0.300	0.280	0.000	0.988	0.036
Smoking alter	0.100	0.111	0.048	0.180	0.026
Smoking ego	0.100	0.104	0.426	0.111	0.046
Smoking similarity	0.000	-0.011	0.441	0.036	0.036
Rate smoking period 1	3.000	0.864	0.000	1.000	1.000
Rate Smoking period 2	3.000	1.065	0.000	1.000	0.976
Behavior smoking tendency	-1.400	0.833	0.000	0.108	0.863
Behavior Smoking total similarity	1.000	0.772	0.000	0.000	0.144
Smoking: effect from Smoking	2.000				
Ce-10					
Constant network rate (period 1)	7.000	6.775	0.000	1.000	0.073
Constant network rate (period 2)	5.000	4.992	0.824	1.000	0.073
Outdegree (density)	-2.300	-2.287	0.004	1.000	0.038
Reciprocity	2.000	2.045	0.000	1.000	0.035
Transitive triplets	0.300	0.280	0.000	0.986	0.052
Smoking alter	0.100	0.110	0.073	0.146	0.028
Smoking ego	0.100	0.111	0.056	0.106	0.031
Smoking similarity	0.400	0.436	0.017	0.276	0.042
Rate smoking period 1	3.000	0.965	0.000	1.000	0.993
Rate Smoking period 2	3.000	1.088	0.000	1.000	0.974
Behavior smoking tendency	-1.400	0.884	0.000	0.231	0.906
Behavior Smoking total similarity	0.000	-0.886	0.000	0.000	0.000
Smoking: effect from Smoking	2.000				
Ce-11					
Constant network rate (period 1)	7.000	6.808	0.000	1.000	0.060
Constant network rate (period 2)	5.000	4.944	0.131	1.000	0.074
Outdegree (density)	-2.300	-2.316	0.001	1.000	0.044
Reciprocity	2.000	2.036	0.000	1.000	0.053
Transitive triplets	0.300	0.282	0.000	0.995	0.030
Smoking alter	0.100	0.102	0.792	0.120	0.037
Smoking ego	0.100	0.115	0.021	0.118	0.028
Smoking similarity	0.400	0.464	0.000	0.340	0.032
Rate smoking period 1	3.000	0.857	0.000	1.000	0.993
Rate Smoking period 2	3.000	0.976	0.000	1.000	0.993
Behavior smoking tendency	-1.400	0.642	0.000	0.037	0.926
Behavior Smoking total similarity	1.000	0.824	0.000	0.002	0.111
Smoking: effect from Smoking	2.000				

Appendix Table 11 Statistical analysis of model *Cf* simulations results

Cf-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.931	0.158	1.000	0.067
Constant network rate (period 2)	5.000	5.004	0.903	1.000	0.069
Outdegree (density)	-2.300	-2.264	0.000	1.000	0.071
Reciprocity	2.000	2.031	0.000	1.000	0.038
Transitive triplets	0.300	0.283	0.000	0.992	0.043
Smoking alter	0.100	0.102	0.688	0.158	0.038
Smoking ego	0.100	0.109	0.039	0.158	0.026
Rate smoking period 1	3.000	3.030	0.610	0.897	0.123
Rate Smoking period 2	3.000	3.161	0.005	0.895	0.073
Behavior smoking tendency	-1.400	-0.498	0.000	0.411	0.757
Behavior Smoking total similarity	0.000	-0.116	0.000	0.000	0.000
Smoking: effect from Smoking	2.000	2.065	0.000	0.988	0.018
Smoking similarity	0.000				
Cf-01					
Constant network rate (period 1)	7.000	6.893	0.037	1.000	0.079
Constant network rate (period 2)	5.000	4.975	0.467	1.000	0.052
Outdegree (density)	-2.300	-2.278	0.000	1.000	0.050
Reciprocity	2.000	2.045	0.000	1.000	0.039
Transitive triplets	0.300	0.275	0.000	0.983	0.037
Smoking alter	0.100	0.105	0.259	0.174	0.027
Smoking ego	0.100	0.107	0.150	0.157	0.025
Rate smoking period 1	3.000	2.919	0.179	0.837	0.122
Rate Smoking period 2	3.000	3.123	0.058	0.853	0.101
Behavior smoking tendency	-1.400	-0.791	0.000	0.541	0.438
Behavior Smoking total similarity	1.000	0.746	0.000	0.014	0.229
Smoking: effect from Smoking	2.000	2.123	0.000	0.969	0.017
Smoking similarity	0.000				
Cf-10					
Constant network rate (period 1)	7.000	6.872	0.001	0.999	0.084
Constant network rate (period 2)	5.000	5.041	0.074	1.000	0.057
Outdegree (density)	-2.300	-2.277	0.000	1.000	0.055
Reciprocity	2.000	2.052	0.000	1.000	0.043
Transitive triplets	0.300	0.279	0.000	0.992	0.039
Smoking alter	0.100	0.076	0.000	0.120	0.052
Smoking ego	0.100	0.082	0.000	0.101	0.040
Rate smoking period 1	3.000	3.068	0.102	0.886	0.092
Rate Smoking period 2	3.000	3.235	0.000	0.886	0.081
Behavior smoking tendency	-1.400	-0.504	0.000	0.445	0.731
Behavior Smoking total similarity	0.000	-0.088	0.000	0.000	0.000
Smoking: effect from Smoking	2.000	2.069	0.000	0.994	0.021
Smoking similarity	0.400				
Cf-11					
Constant network rate (period 1)	7.000	6.826	0.000	1.000	0.063
Constant network rate (period 2)	5.000	4.998	0.945	1.000	0.051
Outdegree (density)	-2.300	-2.289	0.008	1.000	0.039
Reciprocity	2.000	2.054	0.000	1.000	0.045
Transitive triplets	0.300	0.286	0.000	0.994	0.047
Smoking alter	0.100	0.055	0.000	0.092	0.047
Smoking ego	0.100	0.059	0.000	0.061	0.049
Rate smoking period 1	3.000	2.947	0.368	0.800	0.126
Rate Smoking period 2	3.000	3.205	0.002	0.758	0.079
Behavior smoking tendency	-1.400	-1.097	0.000	0.692	0.242
Behavior Smoking total similarity	1.000	0.749	0.000	0.014	0.189
Smoking: effect from Smoking	2.000	2.234	0.000	0.988	0.018
Smoking similarity	0.400				

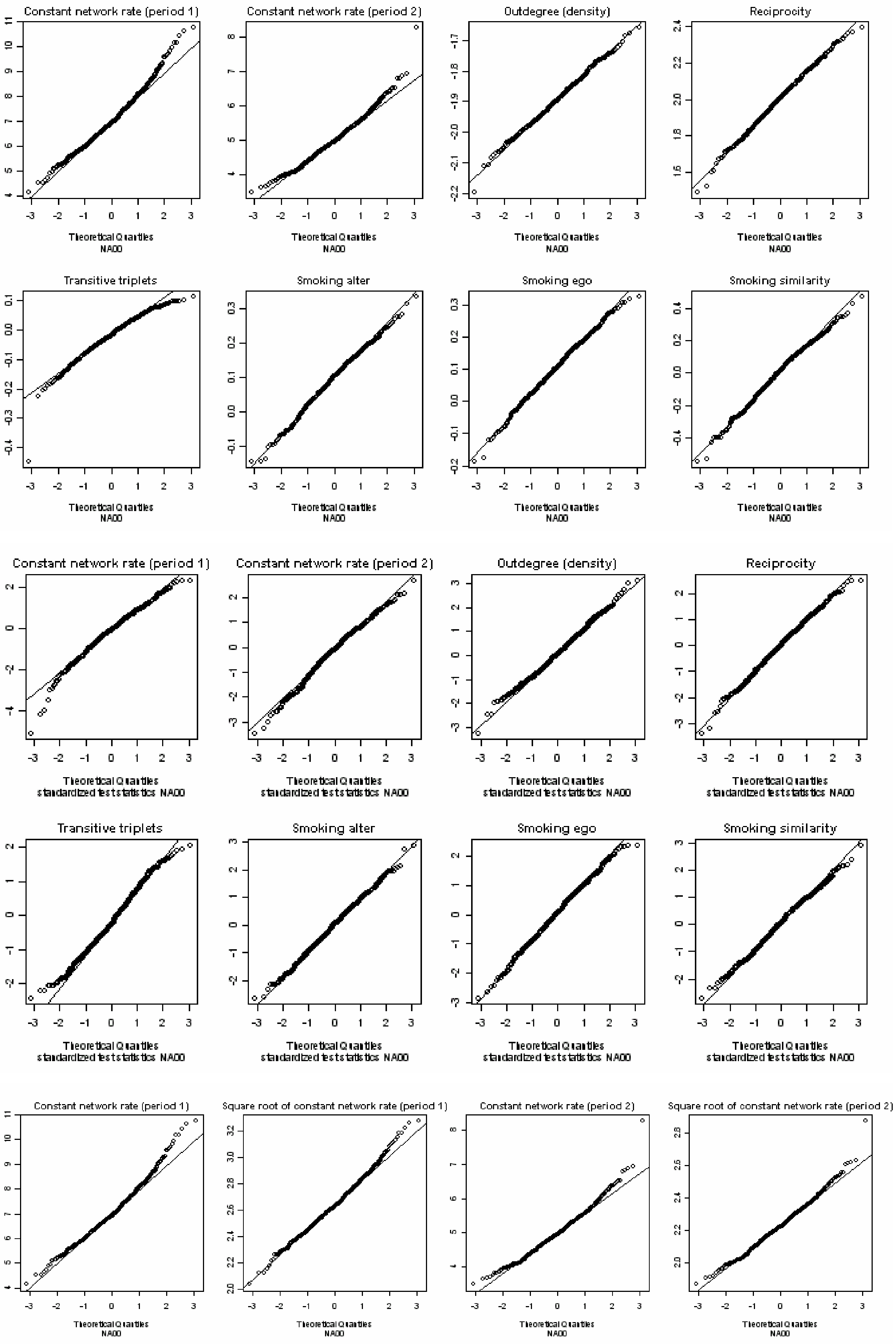
Appendix Table 12 Statistical analysis of model *Cg* simulations results

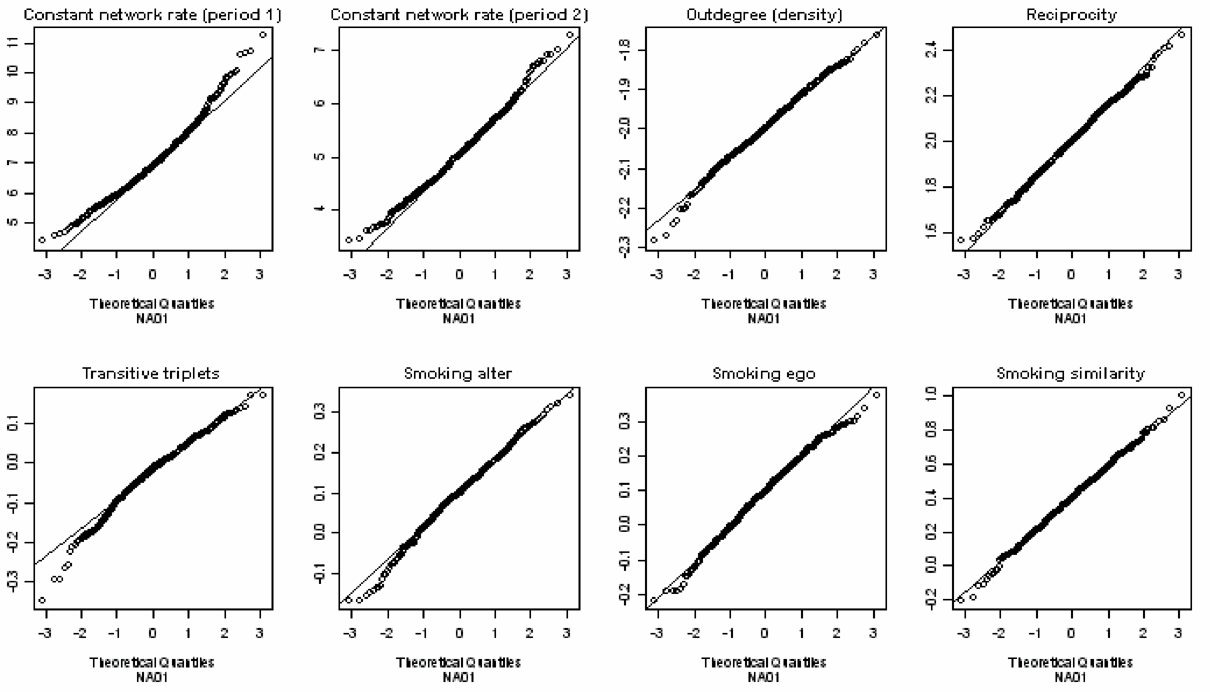


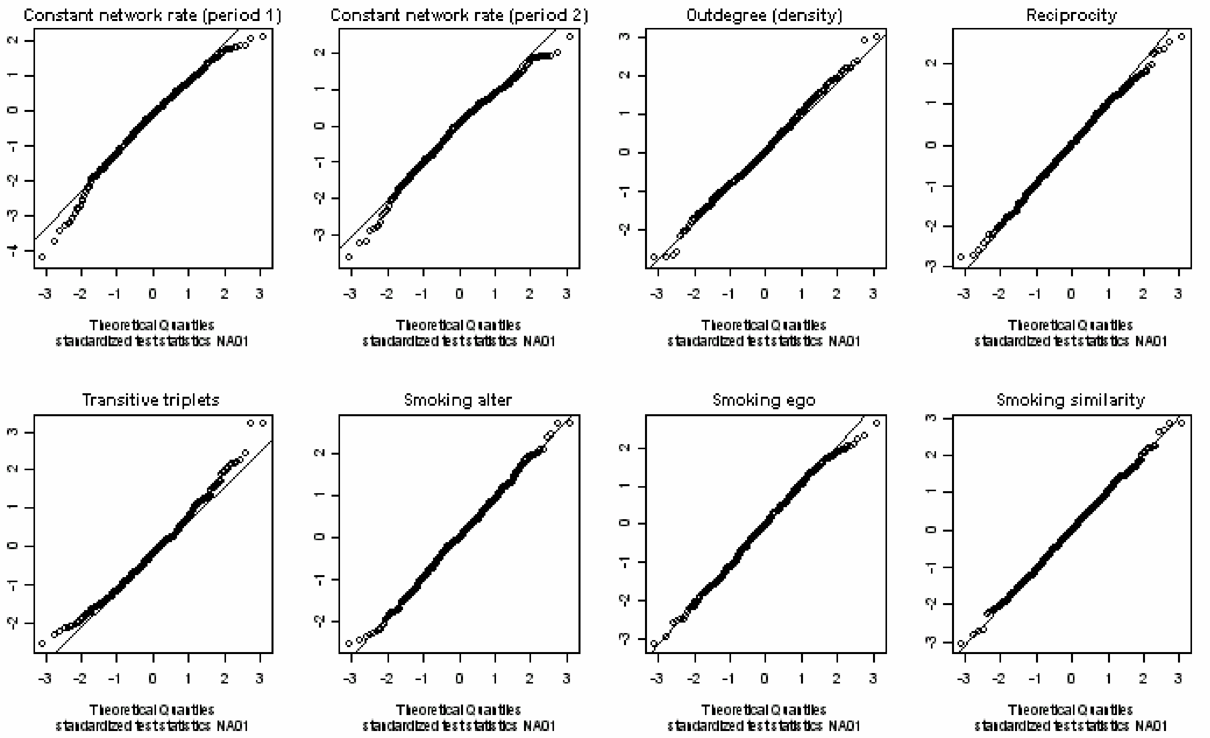
Cg-00	True value	Mean	T-test P-value	Proportion of significant estimation	Estimated type-I error rate
Constant network rate (period 1)	7.000	6.847	0.000	1.000	0.084
Constant network rate (period 2)	5.000	4.987	0.585	1.000	0.061
Outdegree (density)	-2.300	-2.266	0.000	1.000	0.056
Reciprocity	2.000	2.026	0.000	1.000	0.042
Transitive triplets	0.300	0.284	0.000	0.991	0.043
Smoking alter	0.100	0.101	0.775	0.140	0.028
Smoking ego	0.100	0.103	0.328	0.139	0.040
Smoking similarity	0.000	0.003	0.732	0.039	0.039
Rate smoking period 1	3.000	3.223	0.000	0.903	0.070
Rate Smoking period 2	3.000	3.217	0.000	0.881	0.087
Behavior smoking tendency	-1.400	-0.499	0.000	0.439	0.755
Smoking: effect from Smoking	2.000	2.060	0.000	0.995	0.015
Behavior Smoking total similarity	0.000				
Cg-01					
Constant network rate (period 1)	7.000	6.856	0.000	1.000	0.080
Constant network rate (period 2)	5.000	4.979	0.366	1.000	0.048
Outdegree (density)	-2.300	-2.286	0.000	1.000	0.038
Reciprocity	2.000	2.030	0.000	1.000	0.039
Transitive triplets	0.300	0.276	0.000	0.991	0.042
Smoking alter	0.100	0.108	0.017	0.166	0.041
Smoking ego	0.100	0.106	0.091	0.123	0.029
Smoking similarity	0.000	-0.004	0.624	0.032	0.032
Rate smoking period 1	3.000	2.525	0.000	0.902	0.226
Rate Smoking period 2	3.000	3.299	0.000	0.865	0.082
Behavior smoking tendency	-1.400	-0.932	0.000	0.683	0.444
Smoking: effect from Smoking	2.000	2.074	0.000	0.998	0.037
Behavior Smoking total similarity	1.000				
Cg-10					
Constant network rate (period 1)	7.000	6.822	0.000	1.000	0.063
Constant network rate (period 2)	5.000	4.982	0.572	1.000	0.061
Outdegree (density)	-2.300	-2.285	0.000	1.000	0.039
Reciprocity	2.000	2.036	0.000	1.000	0.026
Transitive triplets	0.300	0.280	0.000	0.994	0.047
Smoking alter	0.100	0.101	0.753	0.104	0.033
Smoking ego	0.100	0.095	0.350	0.102	0.033
Smoking similarity	0.400	0.391	0.479	0.230	0.033
Rate smoking period 1	3.000	3.190	0.002	0.927	0.096
Rate Smoking period 2	3.000	3.226	0.000	0.900	0.063
Behavior smoking tendency	-1.400	-0.496	0.000	0.449	0.744
Smoking: effect from Smoking	2.000	2.046	0.008	0.996	0.020
Behavior Smoking total similarity	0.000				
Cg-11					
Constant network rate (period 1)	7.000	6.879	0.018	1.000	0.076
Constant network rate (period 2)	5.000	4.953	0.161	1.000	0.066
Outdegree (density)	-2.300	-2.302	0.608	1.000	0.033
Reciprocity	2.000	2.053	0.000	1.000	0.043
Transitive triplets	0.300	0.282	0.000	0.990	0.053
Smoking alter	0.100	0.101	0.907	0.137	0.018
Smoking ego	0.100	0.096	0.403	0.092	0.025
Smoking similarity	0.400	0.401	0.948	0.289	0.016
Rate smoking period 1	3.000	2.648	0.000	0.859	0.215
Rate Smoking period 2	3.000	3.102	0.119	0.789	0.098
Behavior smoking tendency	-1.400	-1.309	0.003	0.846	0.264
Smoking: effect from Smoking	2.000	2.216	0.000	0.996	0.023
Behavior Smoking total similarity	1.000				

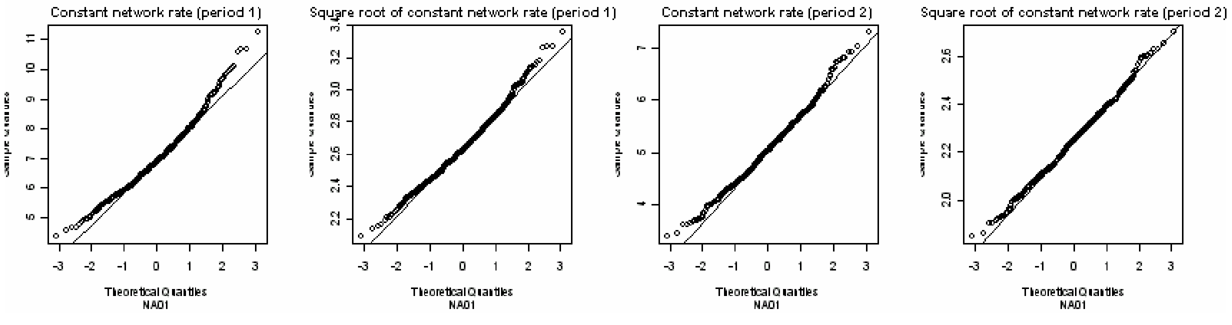
Appendix Figure 1 QQ-plot of *Na* models estimation and standard test statistics

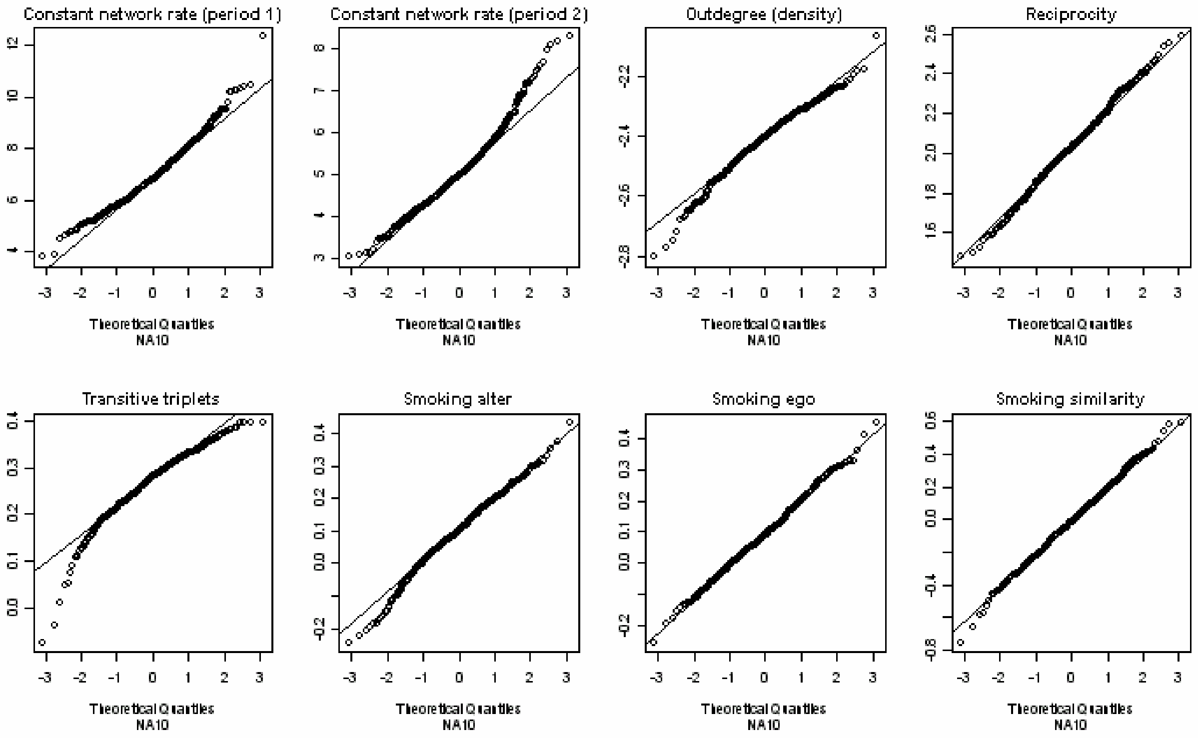
NA in subtitle of these figures in Appendix Figure 1 means *Na*.

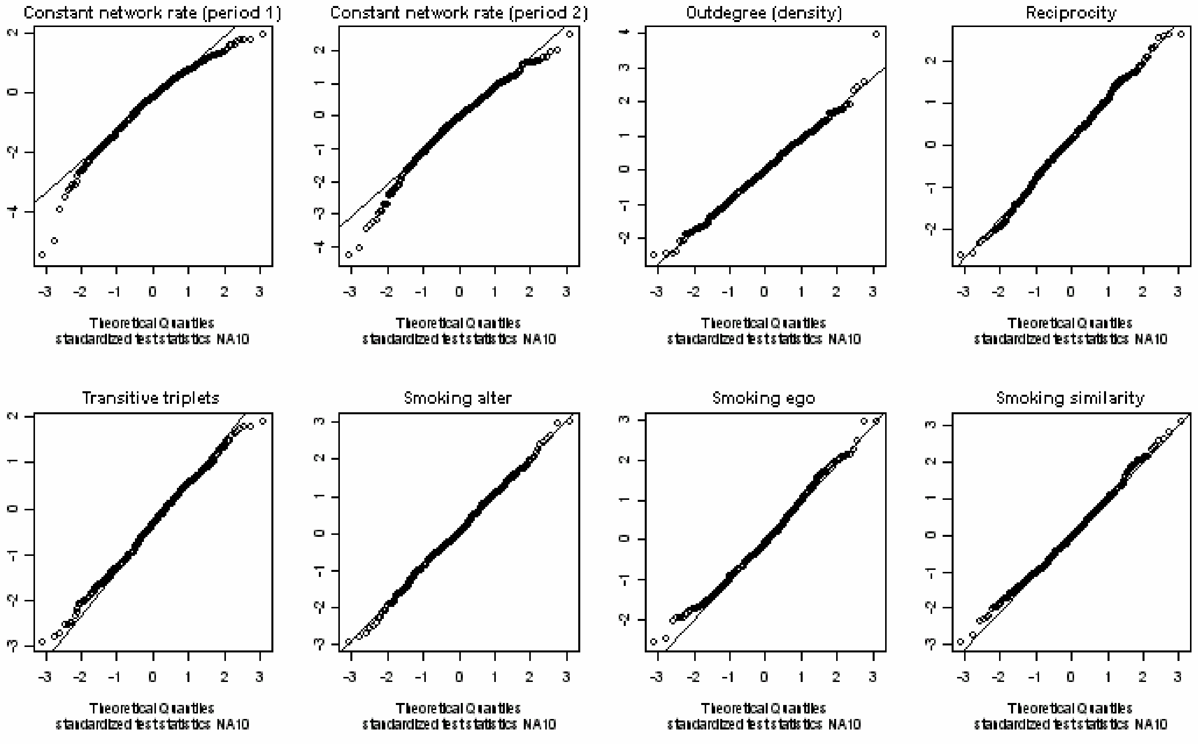


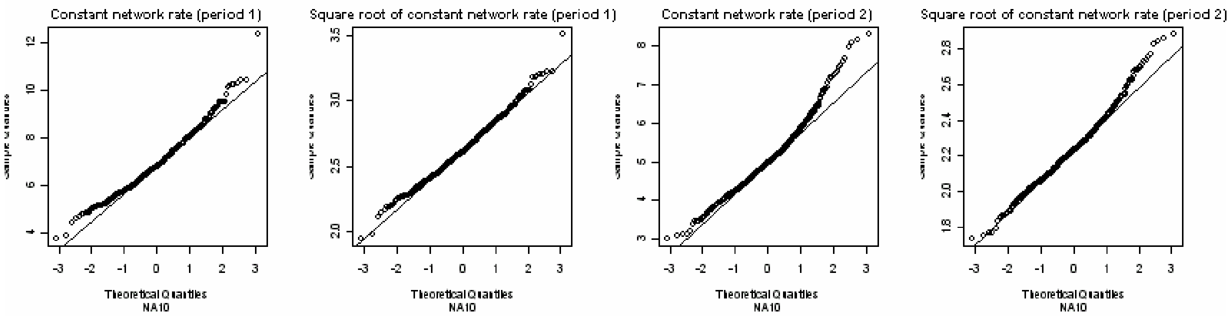

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Theoretical Quantiles
NA01

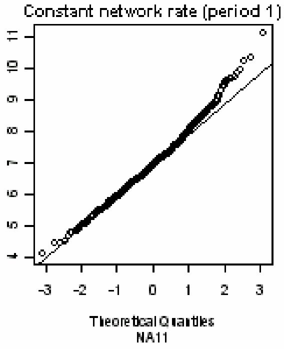

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Theoretical Quantiles
standardized test statistics NA01

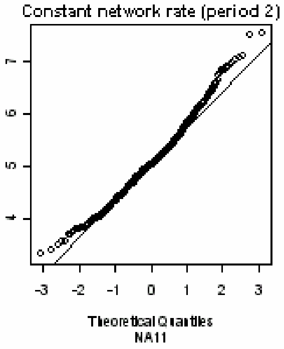

Constant network rate (period 1)
Square root of constant network rate (period 1)
Constant network rate (period 2)
Square root of constant network rate (period 2)
Sample Quantiles
Theoretical Quantiles
NA01

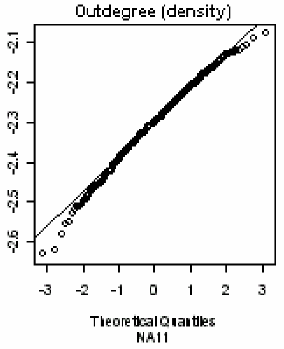

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Theoretical Quantiles
NA10

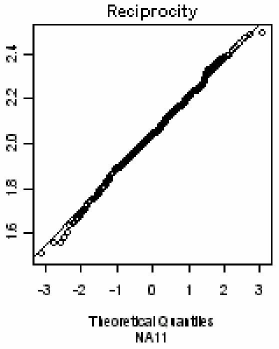

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Theoretical Quantiles
standardized test statistics NA10

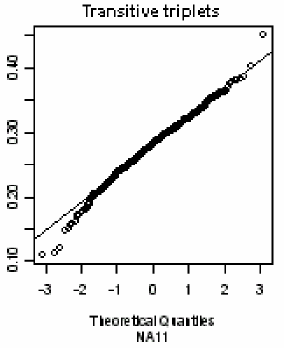

Constant network rate (period 1)
Square root of constant network rate (period 1)
Constant network rate (period 2)
Square root of constant network rate (period 2)
Sample Quantiles
Theoretical Quantiles
NA10

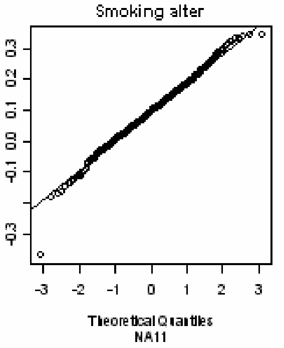

Constant network rate (period 1)
4 5 6 7 8 9 10 11
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

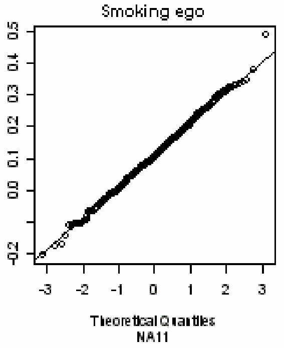

Constant network rate (period 2)
4 5 6 7
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

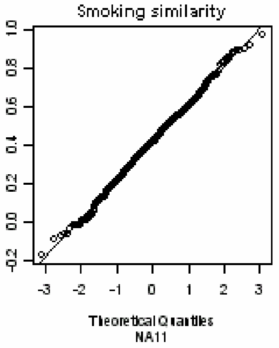

Outdegree (density)
-2.6 -2.5 -2.4 -2.3 -2.2 -2.1
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

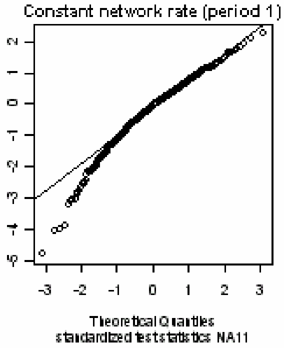

Reciprocity
1.6 1.8 2.0 2.2 2.4
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

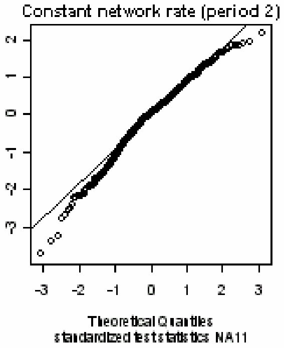

Transitive triplets
0.10 0.20 0.30 0.40
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

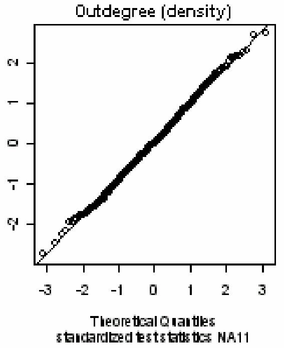

Smoking alter
-0.3 -0.1 0.0 0.1 0.2 0.3
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

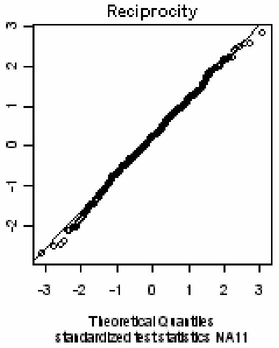

Smoking ego
-0.2 0.0 0.1 0.2 0.3 0.4 0.5
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

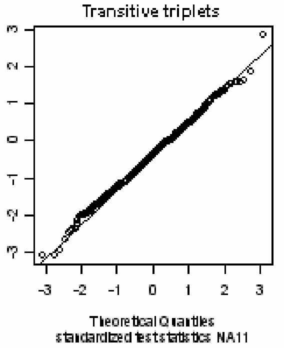

Smoking similarity
-0.2 0.0 0.2 0.4 0.6 0.8 1.0
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

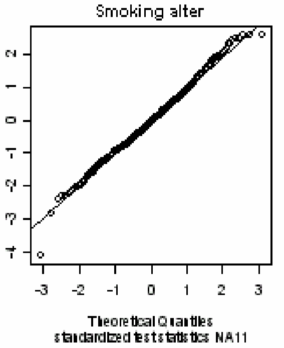

Constant network rate (period 1)
-5 -4 -3 -2 -1 0 1 2
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

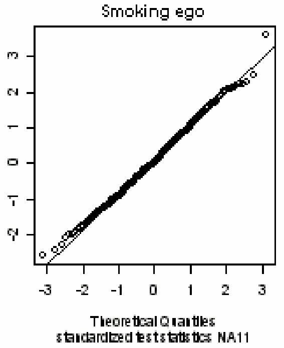

Constant network rate (period 2)
-3 -2 -1 0 1 2
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

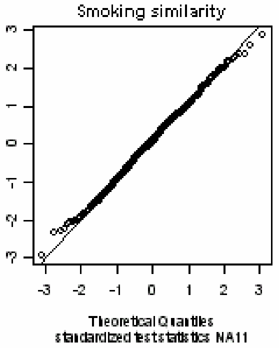

Outdegree (density)
-2 -1 0 1 2
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

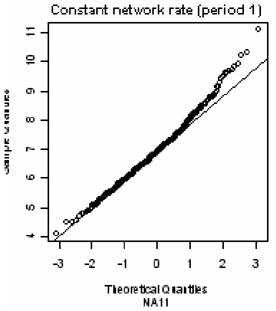

Reciprocity
-2 -1 0 1 2 3
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

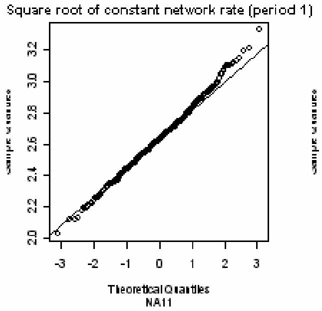

Transitive triplets
-3 -2 -1 0 1 2 3
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

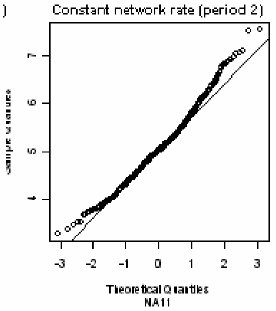

Smoking alter
-4 -3 -2 -1 0 1 2
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11

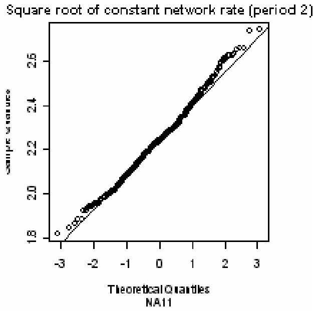

Smoking ego
-2 -1 0 1 2 3
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11


Smoking similarity
-3 -2 -1 0 1 2 3
-3 -2 -1 0 1 2 3
Theoretical Quantiles
standardized test statistics NA11


Constant network rate (period 1)
Sample Quantiles
4 5 6 7 8 9 10 11
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

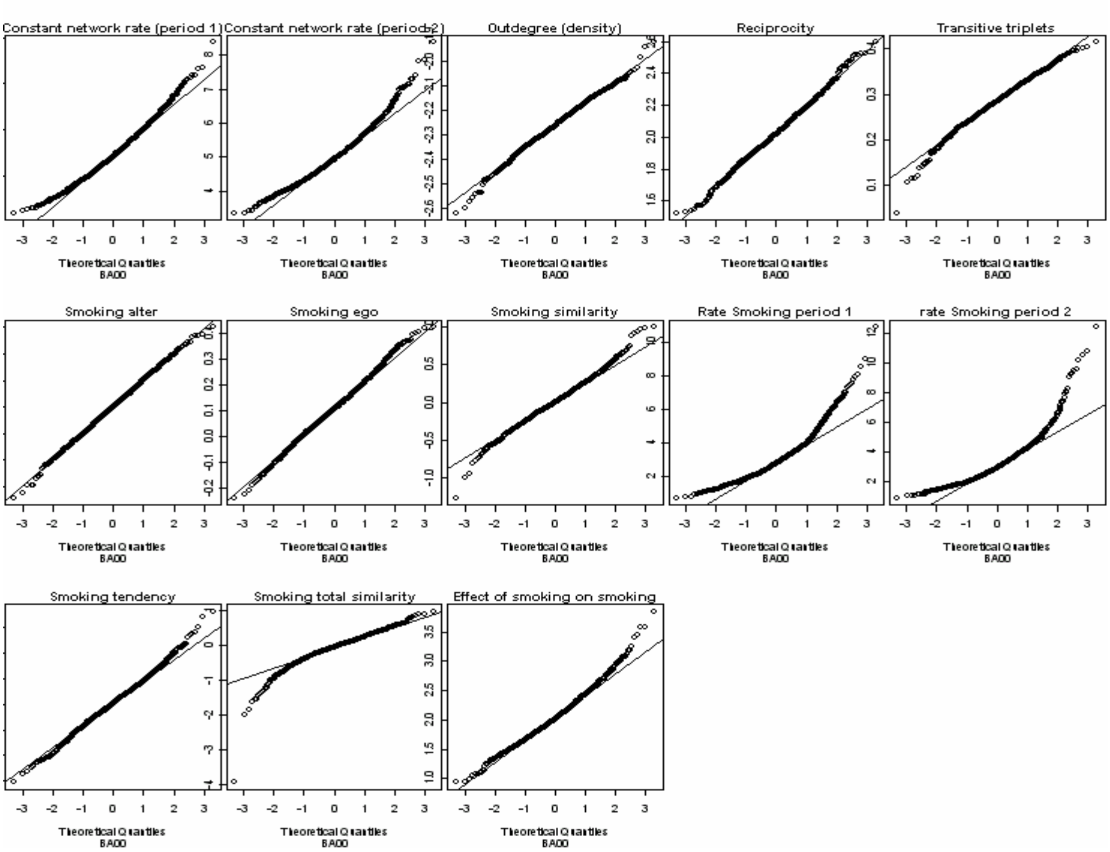

Square root of constant network rate (period 1)
Sample Quantiles
2.0 2.2 2.4 2.6 2.8 3.0 3.2
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

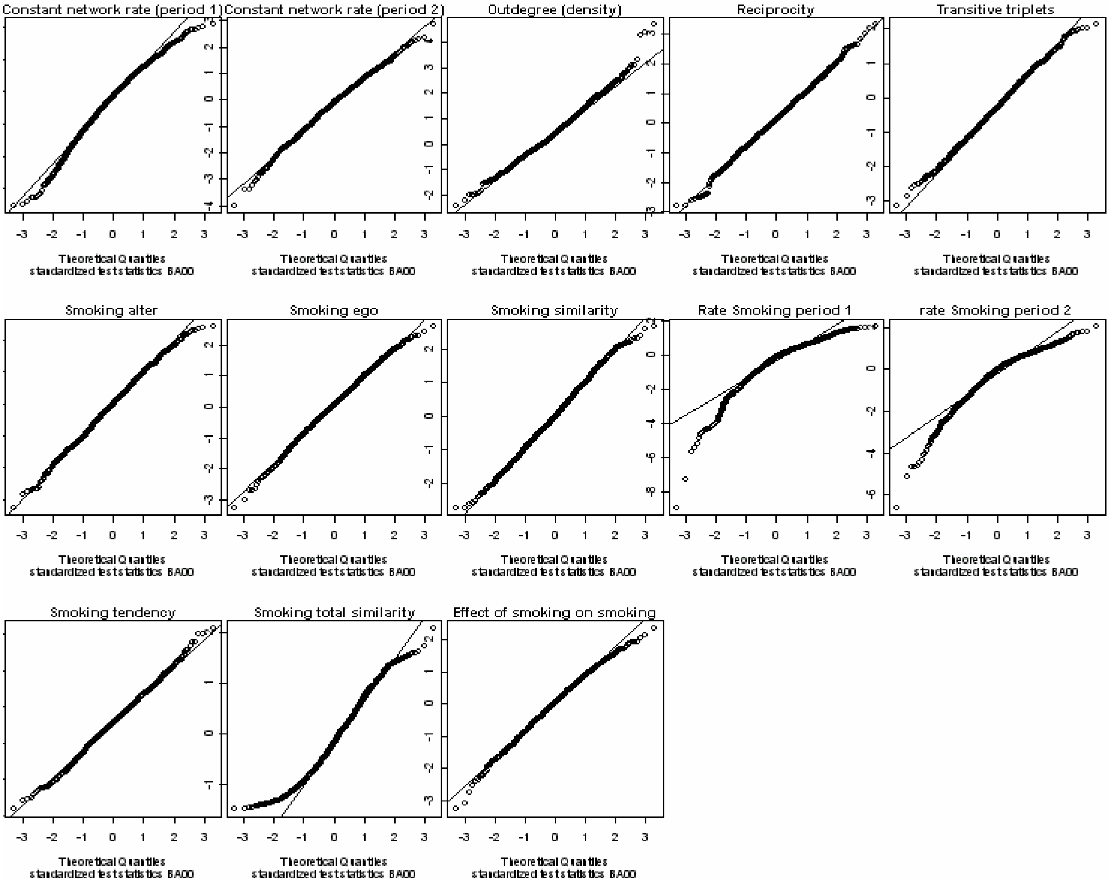

Constant network rate (period 2)
Sample Quantiles
4 5 6 7
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

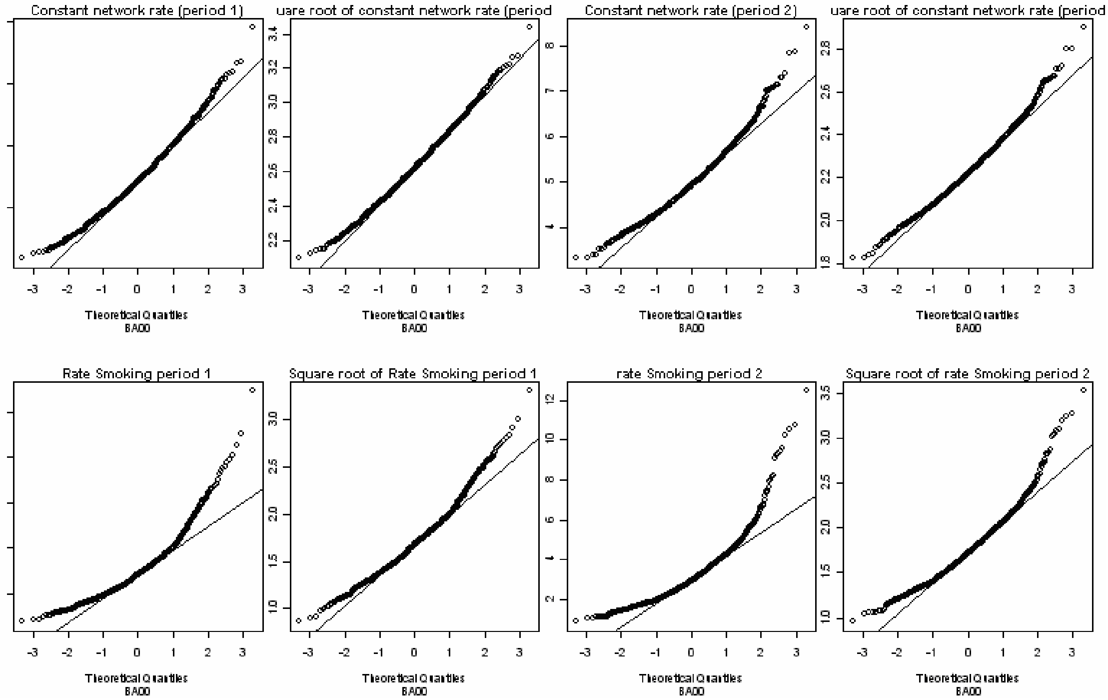

Square root of constant network rate (period 2)
Sample Quantiles
1.8 2.0 2.2 2.4 2.6
-3 -2 -1 0 1 2 3
Theoretical Quantiles
NA11

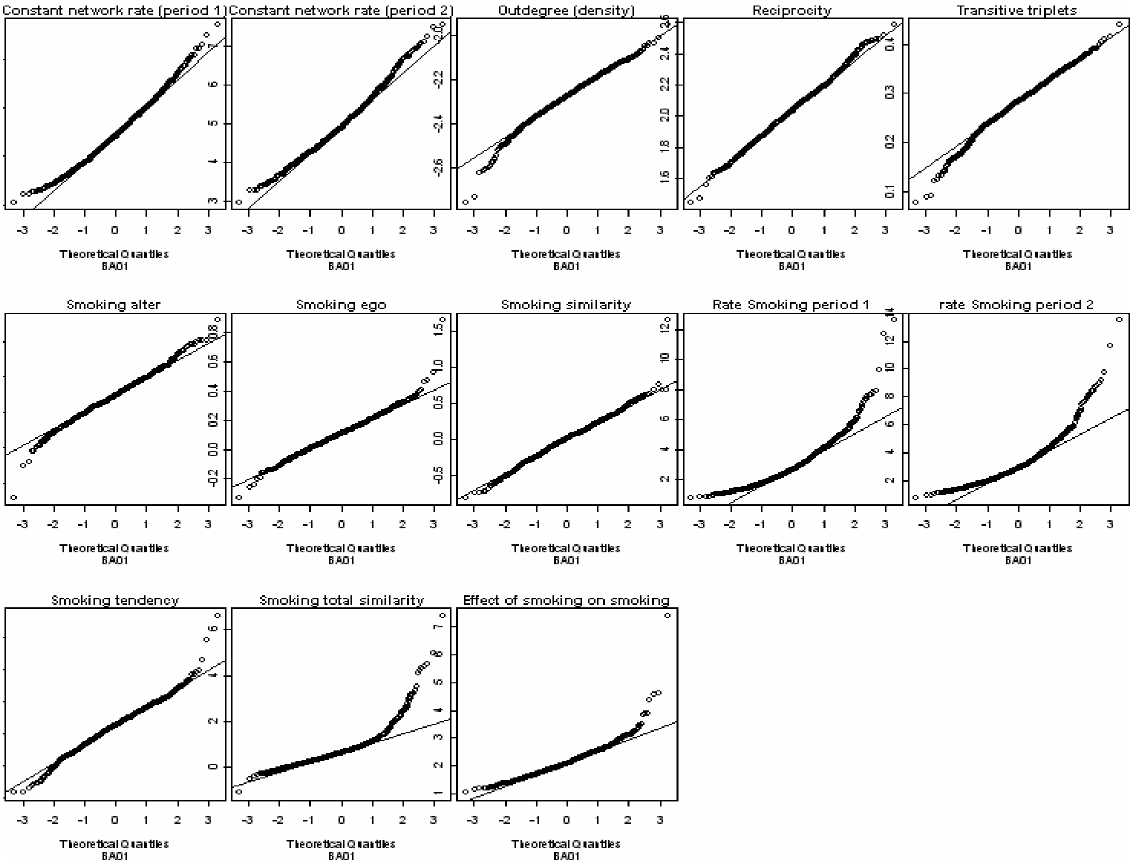
Appendix Figure 2 QQ-plot of Ca models estimations and standard test statistics

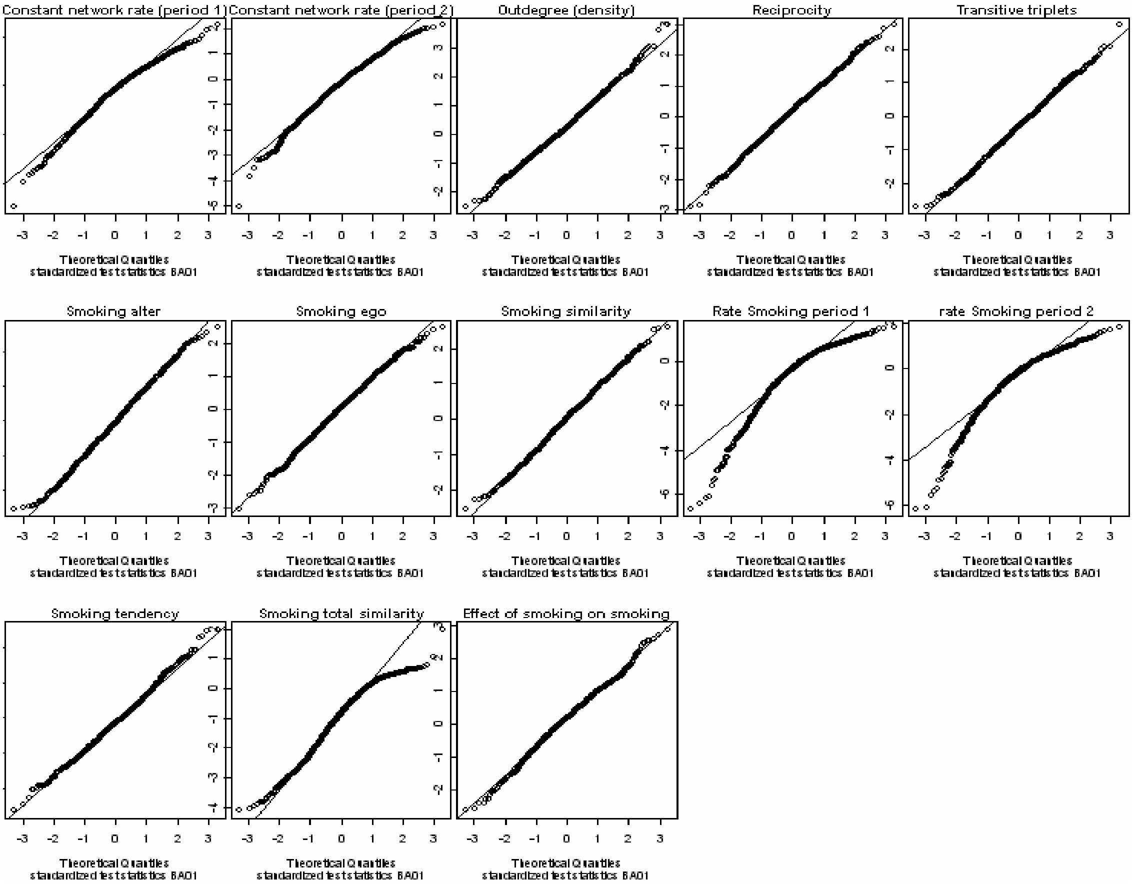
BA in subtitle of these figures in Appendix Fugure 2 means *Ca*

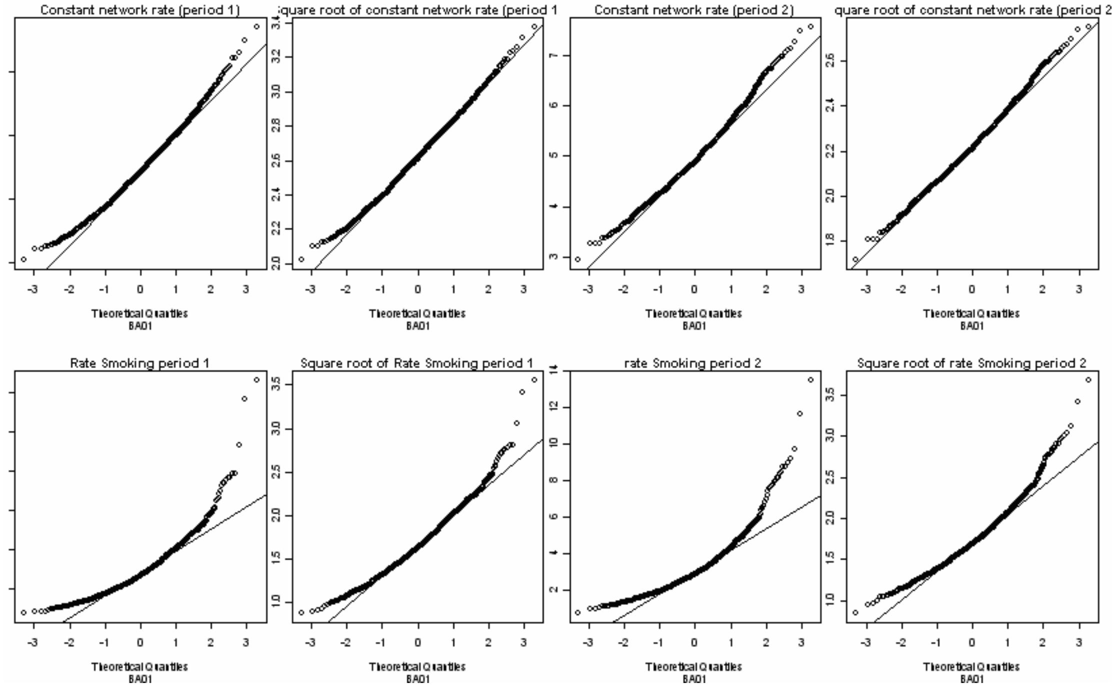


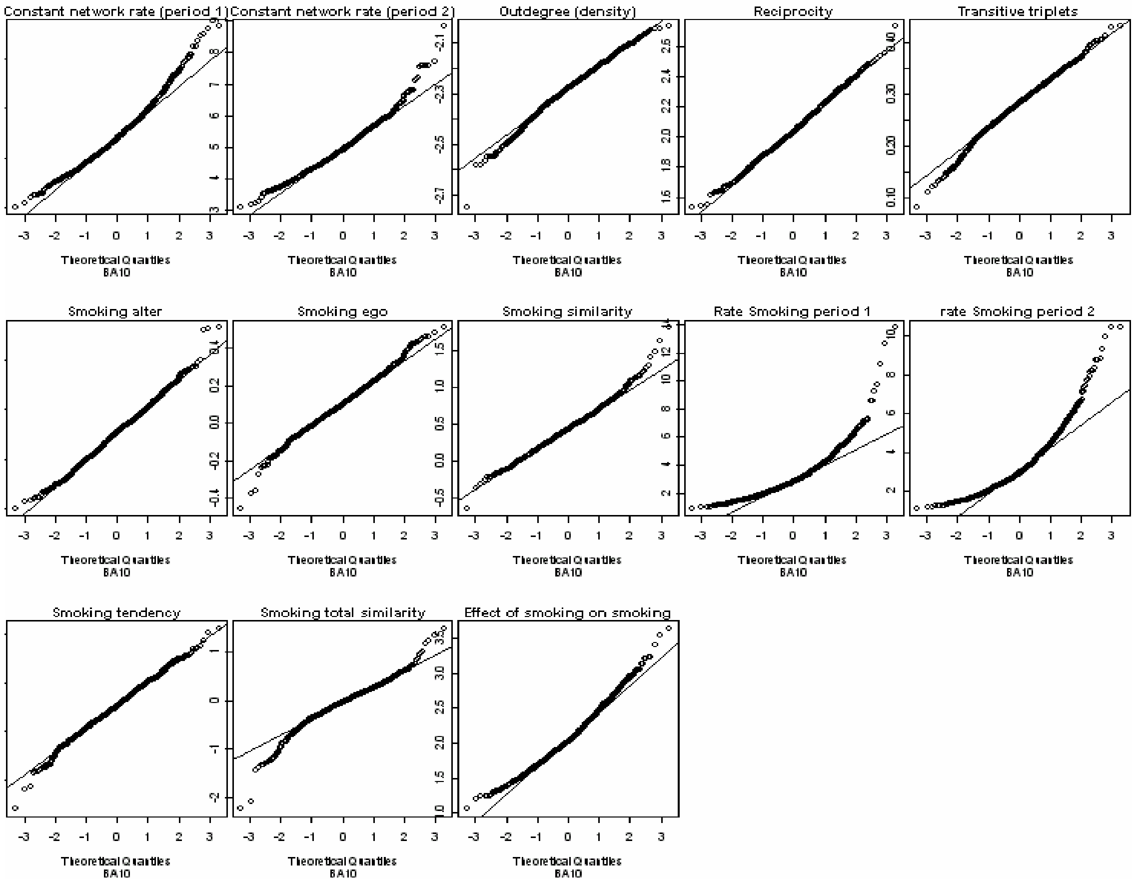


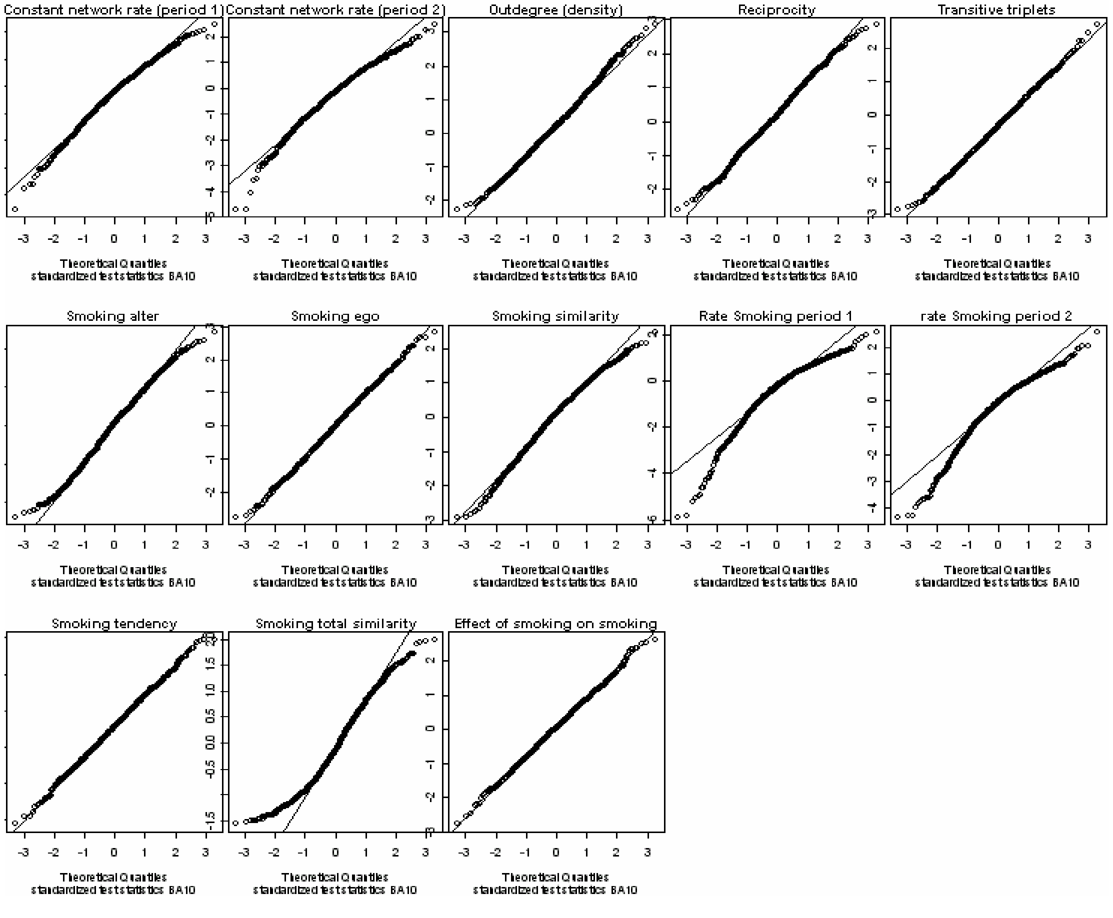

Constant network rate (period 1)
uare root of constant network rate (period
Constant network rate (period 2)
uare root of constant network rate (period
Rate Smoking period 1
Square root of Rate Smoking period 1
rate Smoking period 2
Square root of rate Smoking period 2
Theoretical Quantiles
BA00

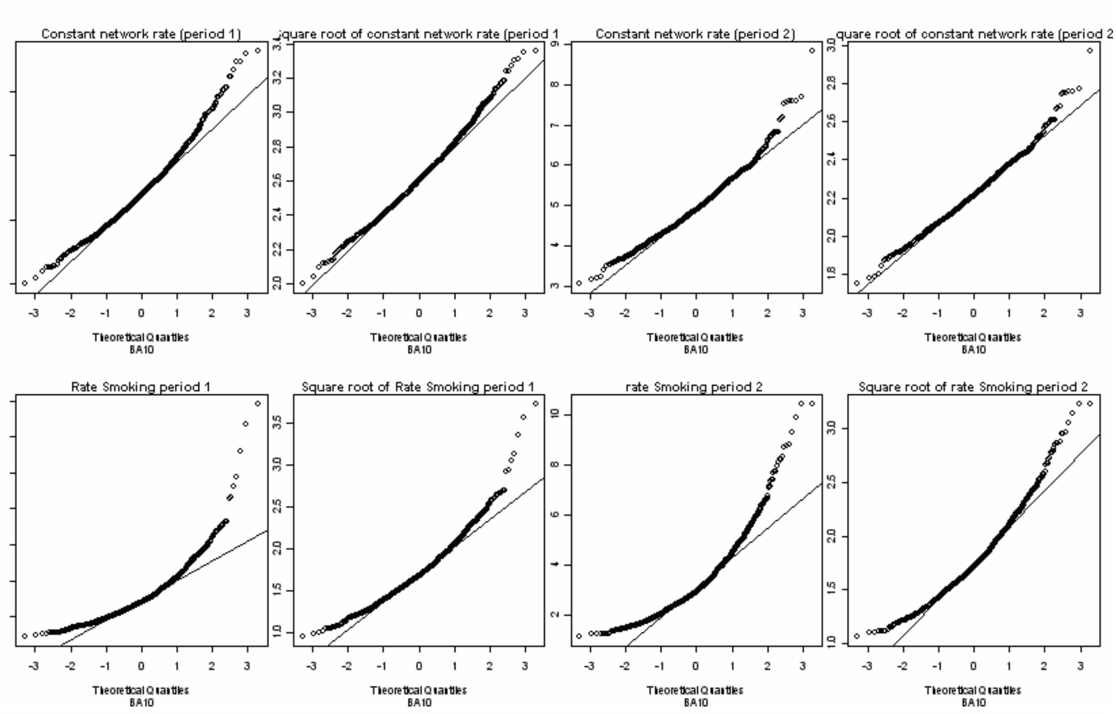

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
BA01

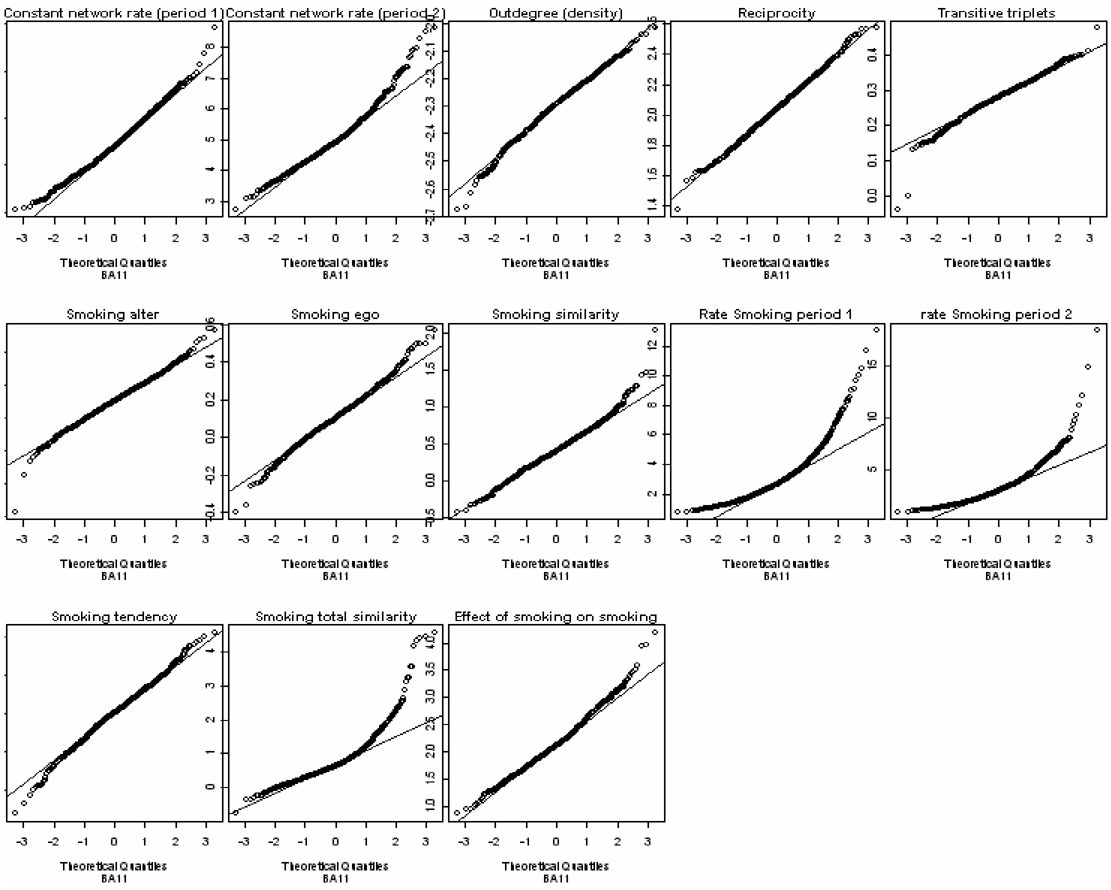

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
standardized test statistics BA01

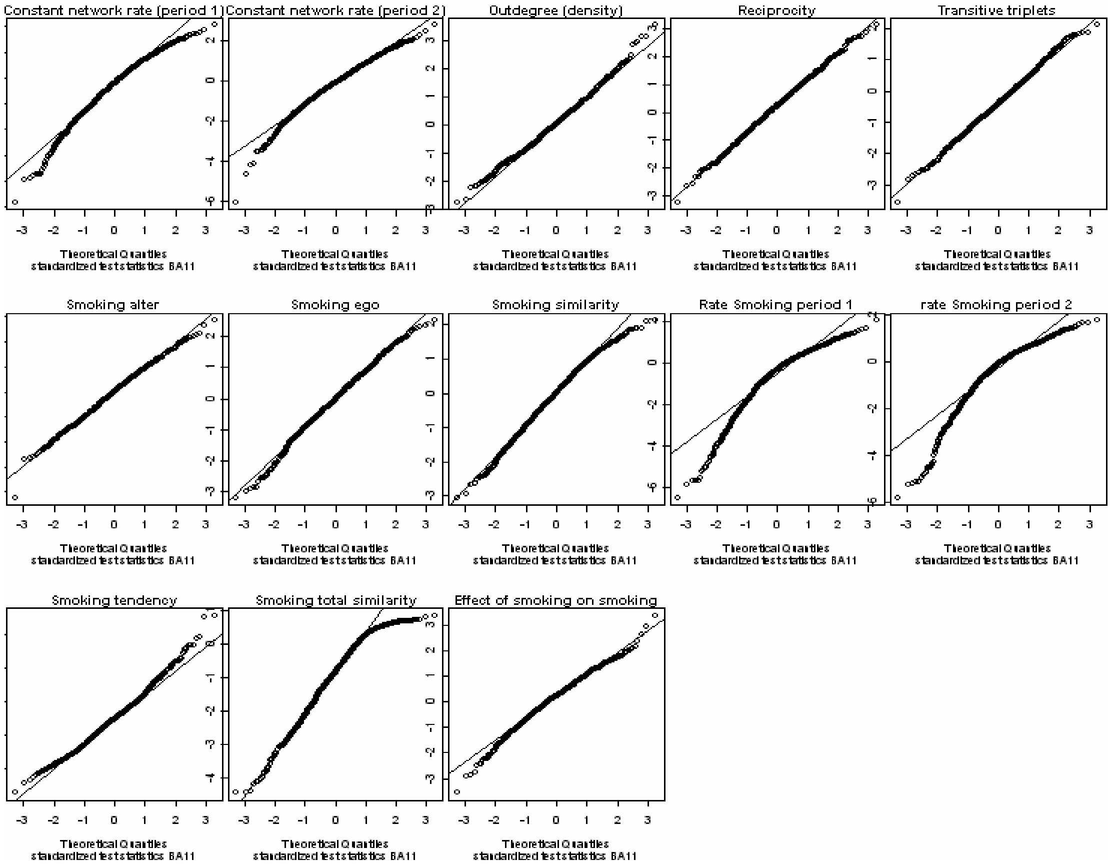

Constant network rate (period 1)
quare root of constant network rate (period 1
Constant network rate (period 2)
quare root of constant network rate (period 2
Rate Smoking period 1
Square root of Rate Smoking period 1
rate Smoking period 2
Square root of rate Smoking period 2
Theoretical Quantiles
BA01

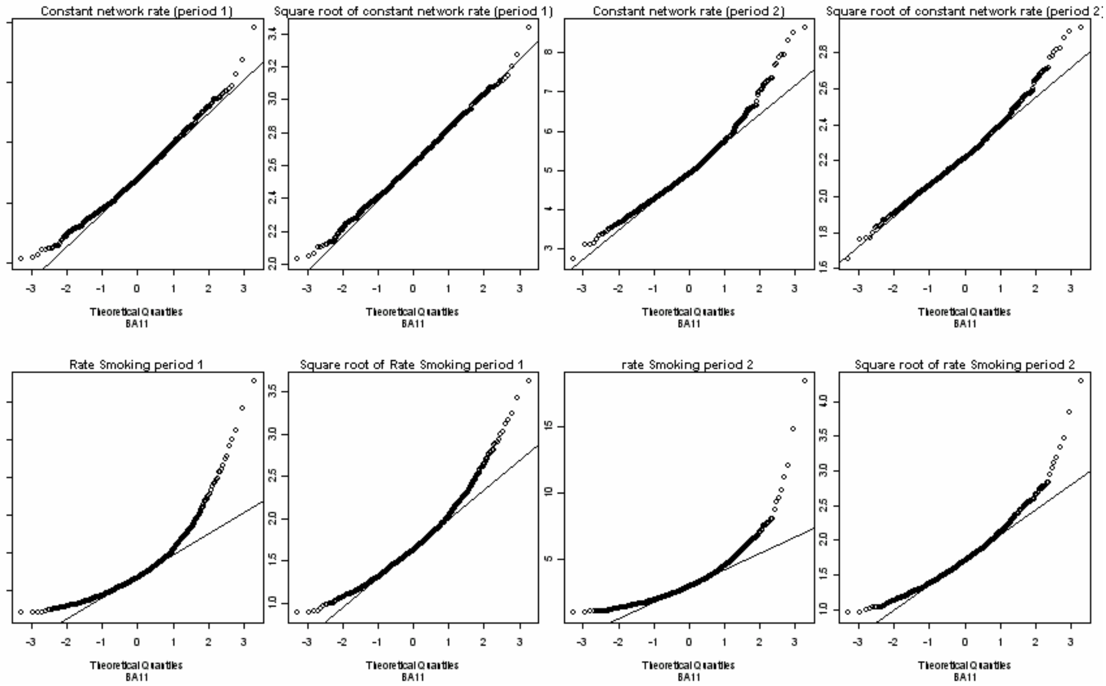

Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
BA10


Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
standardized test statistics BA10


Constant network rate (period 1)
Square root of constant network rate (period 1
Constant network rate (period 2)
Square root of constant network rate (period 2
Rate Smoking period 1
Square root of Rate Smoking period 1
rate Smoking period 2
Square root of rate Smoking period 2
Theoretical Quantiles
BA10


Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
BA11


Constant network rate (period 1)
Constant network rate (period 2)
Outdegree (density)
Reciprocity
Transitive triplets
Smoking alter
Smoking ego
Smoking similarity
Rate Smoking period 1
rate Smoking period 2
Smoking tendency
Smoking total similarity
Effect of smoking on smoking
Theoretical Quantiles
standardized test statistics BA11


Constant network rate (period 1)
Square root of constant network rate (period 1)
Constant network rate (period 2)
Square root of constant network rate (period 2)
Rate Smoking period 1
Square root of Rate Smoking period 1
rate Smoking period 2
Square root of rate Smoking period 2
Theoretical Quantiles
BA11