

University of Groningen

StOCNET

Huisman, M.; van Duijn, M.A.J.

Published in:
Connections

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2003

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Huisman, M., & van Duijn, M. A. J. (2003). StOCNET: Software for the statistical analysis of social networks. *Connections*, 25(1), 7-26.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

StOCNET: Software for the statistical analysis of social networks¹

Mark Huisman and Marijtje A.J. van Duijn ²

ICS / Statistics & Measurement Theory, University of Groningen, the Netherlands

StOCNET³ is an open software system in a Windows environment for the advanced statistical analysis of social networks. It provides a platform to make a number of recently developed and therefore not (yet) standard statistical methods available to a wider audience. A flexible user interface utilizing an easily accessible data structure is developed such that new methods can readily be included in the future. As such, it will allow researchers to develop new statistical tools by combining their own programs with routines of the StOCNET system, providing a faster availability of newly developed methods.

In this paper we show the current state of the developments. The emphasis is on the implementation and operation of the programs that are included in StOCNET: BLOCKS (for stochastic blockmodeling), p2 (for analyzing binary network data with actor and/or dyadic covariates), SIENA (for analyzing repeated measurements of social networks), and ZO (for calculating probability distributions of statistics). Moreover, we present an overview of future contributions, which will be available in the near future, and of planned activities with respect to the functionality of the StOCNET software. StOCNET is a freeware PC program, and can be obtained from the StOCNET website at <http://stat.gamma.rug.nl/stocnet/>.

INTRODUCTION

At the Sunbelt XX conference in Vancouver, Zeggelink, Snijders, and Boer (2000) announced the development of an open software system called StOCNET for the advanced statistical analysis of social networks. Two years later, at the Sunbelt XXII conference in New Orleans, an update of the project was given by Huisman and Van Duijn (2002). Now, we are presenting the newest release: StOCNET Version 1.4 (Boer, Huisman, Snijders, and Zeggelink, 2002).

¹ This paper was presented at the 23rd Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences, Tilburg, The Netherlands, July 1-3, 2002.

² Correspondence to: Mark Huisman, ICS/ Statistics & Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. Tel: +31 50 3636345; Fax: +31 50 3636304; Email: m.huisman@ppsw.rug.nl, URL: www.ppsw.rug.nl/~huisman

³ The development of StOCNET is supported by the Dutch Research Council (NWO), grant 400-20-020.

Many different methods that exist for the analysis of social networks are covered by programs like GRADAP (Sprenger and Stokman, 1989), NetMiner (Cyram, 2002), Pajek (Batagelj and Mrvar, 2002), or UCINET (Borgatti, Everett, and Freeman, 1999). However, a number of recently developed statistical methods and the accompanying computer programs have not yet reached (potentially interested) empirical researchers in social networks. The main reason is that most of these methods are available in privately owned programs that are not quite fit for public distribution. StOCNET provides a Windows platform to make these statistical methods freely accessible.

In this paper the current state of the StOCNET software is presented. A general and flexible data structure and user interface were developed such that new methods can easily be included and new tools can be developed by combining existing programs with routines of the StOCNET system. At this moment, the latest release of StOCNET, Version 1.4, contains five statistical methods for social network analysis:

- BLOCKS, for stochastic blockmodeling of relational data (Nowicki and Snijders, 2001).
- p_2 , for the analysis of binary network data with actor and/or dyadic covariates and random effects (Van Duijn, 1995).
- SIENA, for the statistical analysis of longitudinal network data (Snijders, 2001).
- ULTRAS, for estimating latent transitive structures in social networks (Schweinberger and Snijders, forthcoming).
- ZO, for determining probability distributions of statistics of random graphs (Snijders, 1991; Molloy and Reed, 1995).

These modules are presented in this paper. Furthermore, modules that will be available in the near future are mentioned, as well as other functionalities of the StOCNET software to be implemented.

The paper starts with a brief introduction to the StOCNET project. In the following section, the current version of StOCNET (version 1.4) is presented. The supported options are described and an example data set consisting of friendship networks between university freshmen is used to illustrate the StOCNET software.

The statistical methods forming the core of the StOCNET software are introduced in Section 3: BLOCKS, p_2 , SIENA, ULTRAS, and ZO. The university freshmen data were analyzed with these procedures, and some results are presented. In Section 4 the Examine functionality, which gives the user the possibility to explore the network and covariate data is described. In Section 5, planned activities with respect to new procedures and functionalities are introduced, and the paper ends with directions for potential contributors and the presentation of the StOCNET website.

1. THE StOCNET PROJECT

Methods for the analysis of social networks are not covered in standard statistical packages. A number of software packages were specifically designed for social network analysis. The most widely known and widely used programs are GRADAP (Sprenger and Stokman, 1989), Pajek (Batagelj and Mrvar, 2002), and UCINET (Borgatti, Everett, and Freeman, 1999). Other (less well known and

more specialized) programs are FATCAT (Richards and Seary, 1993), MultiNet (Richards and Seary, 2000), NetMiner (Cyram, 2002), and STRUCTURE (Burt, 1991)⁴. These programs contain many state-of-the-art analysis methods, but they cover only specialized methods, or the coverage of established methods is not complete. Especially, and not surprisingly, recently developed statistical methods are underrepresented in these programs. Stimulated by general developments in statistical analysis that were facilitated by expanded computing force, such as Gibbs sampling, computer-intensive techniques have proved to be quite important contributions. These methods are usually implemented in computer programs constructed by their developers, which are not always easily available or easy to use, and usually miss sufficient documentation. Therefore, the methods have not yet (or hardly) reached the empirical social network researchers. For a comprehensive overview of software available for social network analysis, see Huisman and Van Duijn (2003).

The purpose of the StOCNET project is the development of an open software system for the advanced statistical analysis of social networks. The system serves three goals:

- the incorporation of important recently developed statistical methods for network analysis in user-friendly and easily available software,
- more efficiency in the implementation of such new methods by setting up a system with a common data structure and a common user interface, and
- faster availability of new methods.

The aim is not competition with but complementarity to other social network analysis software. Whereas other programs are usually closed systems offering a wide range of mostly non-statistical methods, StOCNET is an open system focusing on more advanced statistical methods. Partly because of their being based on simulation methods, these methods require more expertise of the user.

A second aspect of the StOCNET project is the definition of requirements and formulation of instructions for potential contributors to the system. The definition of requirements is aimed at user-interface, data definition, and output, thus developing StOCNET standards, with the purpose of reducing the programming task of the contributor. The requirements also include (precise) documentation of the method, which is facilitated because the set-up of the program, data definition, and technical specifications are already documented in StOCNET.

2. THE StOCNET SYSTEM

2.1 StOCNET sessions

An analysis within StOCNET takes place within a so-called session, and usually consists of five more or less sequential steps. The steps start with the data definition and result in specified output, after which all or some steps can be repeated. Within a session the user generally uses the same (selection of) data sets. Transformations, selections and the latest model specifications for these data are all saved by saving the session, and can easily be activated again when opening the same session a second time.

For most users, the sequential process of five steps will soon become a cyclic process, possibly skipping certain steps. The interactive features of StOCNET imply that any revised analysis can easily

⁴ Some of these programs are available as freeware: see the INSNA website at <http://www.sfu.ca/~insna/>.

be undertaken in the current session or in a new session. The different steps in a session are entered by clicking the corresponding buttons in the StOCNET-toolbar (see Figure 1).

Step 1. Data definition: Specification and description of the network(s) and the actor attributes in separate data files. In this step, the network data and the attribute data are specified. Both may be contained in multiple data files. The network must be presented as an adjacency matrix (saved) in ASCII format. This means that each network is presented by n lines with n integer numbers separated by blanks, and each line is ended by a hard return. Once a file has been selected, the network in that file is added to the set of available networks for that session. The same holds for attribute files (actor and/or dyadic attributes). For k actor covariates, the data file must contain n lines, with on each line k numbers separated by blanks which are read as real numbers. This data definition was chosen because most current statistical methods require networks with not more than moderately many nodes (up to a few hundred⁵). The StOCNET user-interface for this step is presented in Figure 1.

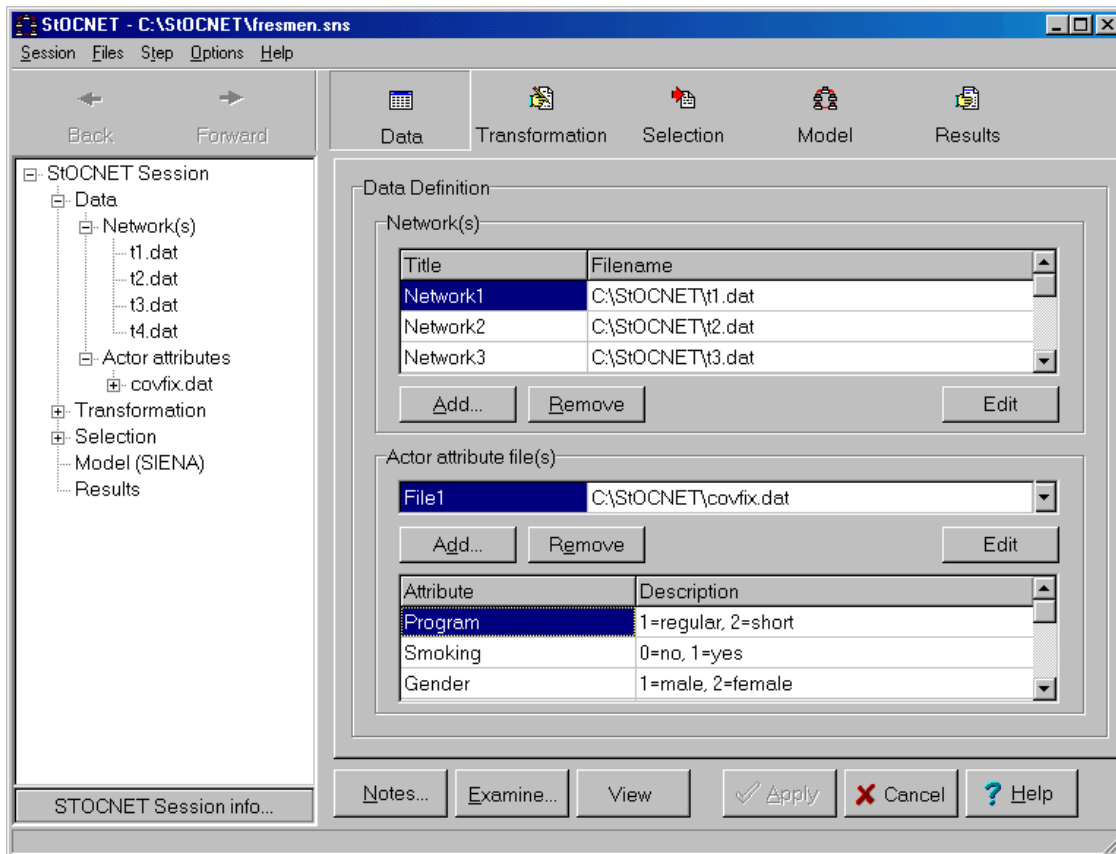


Figure 1. Step1 in a StOCNET session – Data definition.

Step 2. Transformation: Recoding network values and/or actor variables and specification of missing values. A full recode functionality is implemented to transform all available data (network and attributes). Also missing value codes can be defined (multiple codes for network values). How the missing values are treated depends on the procedure chosen for the analysis. The only thing

⁵ In programs that are designed to handle large networks like Pajek or MultiNet, data are usually defined by node and link (arcs/edges) lists.

StOCNET does in this respect is to pass on the code to the module. All transformations can be inspected by clicking the View-button to examine the data. The transformed data is saved in a temporary file with the same name as the old file preceded by '~'. These files can also be saved under different names.

Step 3. Selection: Selection of actors based on several possible criteria. There are four different ways to select a set of actors:

1. Select all n actors that are available in the specified network files.
2. Define a subset of actors by specifying a range of actors, for instance, the first ten or twenty actors in the adjacency matrix. The program will automatically select the corresponding rows and columns of the network data.
3. Selection based on the values of the attributes. For instance, using the attribute gender, the female actors can be selected. For this purpose, first the attribute file containing the desired covariate has to be specified then the covariate and its value to use as selection criterion.
4. The last selection method is the most complex one. It involves an examination of the network data and computation of some network statistics. Subsequently a subset of actors is selected that fulfills a certain network requirement. The current version of StOCNET, only the selection of actors based on their in- or out-degree is supported.

Selections can be examined and saved the same way as transformations.

Step 4. Model specifications and analysis: Choice of the program for data analysis, specification of the model parameters in the model specific user interface, and running the method. The models that are available in the current version are described in Section 3.

Step 5. View results: Inspection of the output and results from the analyses. StOCNET allows for a structured view through the output by selecting certain output items. The items are indicated in the output file by the symbol @1 for chapters, @2 for sections, @3 for sub-sections, and so on. These items are shown in the session-tree in the left part of the StOCNET window, and the user can select an item by double clicking on it.

In every step of a StOCNET session, the session-tree (or history-tree) contains global information on the history of the present session. The session-tree is shown in the left part of the window. The operation of this tree is similar to standard options in Windows Explorer, with the difference that here an overview is given of actions taken together with details of these actions. The details can be viewed by clicking the corresponding '+'. Double clicking the step name results in a move towards the corresponding step in this session. The session-tree is shown in Figure 1.

2.2 Menu bar

The menu bar of the StOCNET program consists of four items. The first one, labeled *Session*, and the last one, labeled *Help* are standard Windows functionalities for opening/saving files (sessions) and on-line help, respectively. The Session-menu also contains Export functions to other network programs. In order to exploit the large amount of methods and statistics that are available in other programs (e.g., UCINET, Pajek, or NetMiner), and especially the graphical procedures in some of

these programs,⁶ export functions are included in the new version of StOCNET. Thus input files are created that can easily be used in MultiNet (Richards and Seary, 2002), NetMiner (Cyram, 2002), Pajek (Batagelj and Mrvar, (2002), and STRUCTURE (Burt, 1991).

With the second item, labeled *Step*, the consecutive steps in a StOCNET session are entered. These steps are also available via separate buttons in the StOCNET-toolbar (see Figure 1). The third option, labeled *Options*, contains options to set the toolbar on or off and to set working directories.

2.3 Example: friendships between university freshmen

Van Duijn, Huisman, Stokman, Wasseur, and Zeggelink (2003) studied friendship relations in a network of freshmen students. The data were collected at five time points during the years 1996 and 1997, and are repeated measures of the friendship network of 38 Dutch university freshmen in sociology. The dichotomized relation studied is defined as ‘at least a friendly relationship’.

Friendship data were collected at five time points: at the start of the academic year, and next at 3, 6, 13, and 35 weeks after the start of the year, respectively. We refer to these observation times as t_1 to t_5 . Although most students did not know each other at the beginning of the year, some relationships already present at t_1 are taken into account. At the observation times, the relations of 38, 25, 26, 18, and 18 students, respectively, were completely observed⁷.

At t_1 actor attribute data were collected. Three groups of attribute variables were distinguished: 1) opportunity variables (study program, smoking behavior), 2) visible attributes (gender), and 3) invisible attributes (doing sports, watching sports, playing music, religious involvement, club membership, study orientation, and social orientation); see Van Duijn et al. (2003) for a detailed description of all variables. Figure 1 shows the definition of the freshmen data in StOCNET.

3. IMPLEMENTED MODULES

3.1 BLOCKS

The module BLOCKS (version 1.53) is designed for stochastic blockmodeling of relational data according to the methods described in Nowicki and Snijders (2001). For more detailed information on the program BLOCKS or stochastic blockmodeling, the reader is referred to Nowicki and Snijders (2001) and to the user's manual of BLOCKS (Snijders and Nowicki, 2001).

Posterior blockmodeling consists of finding equivalent groups of actors (with respect to relational patterns) based on the observed relations between the actors. When the observed data are assumed to be generated by some stochastic mechanism, this approach to blockmodeling is called stochastic blockmodeling. The method implemented in BLOCKS tries to find out – a posteriori – how many different (latent) classes of actors can be distinguished and which actors belong together in the same class by estimating the posterior probability distribution of the configuration of the class structure given the network data. The parameter estimates are obtained with Gibbs sampling.

⁶ It was decided not to develop a graphical module for the StOCNET system because of the large amount of programming time it would require, and the availability of other programs with good graphic procedures.

⁷ Between t_3 and t_4 one student left the network due to a change of discipline. The program SIENA can handle this kind of composition change (Huisman and Snijders, 2002).

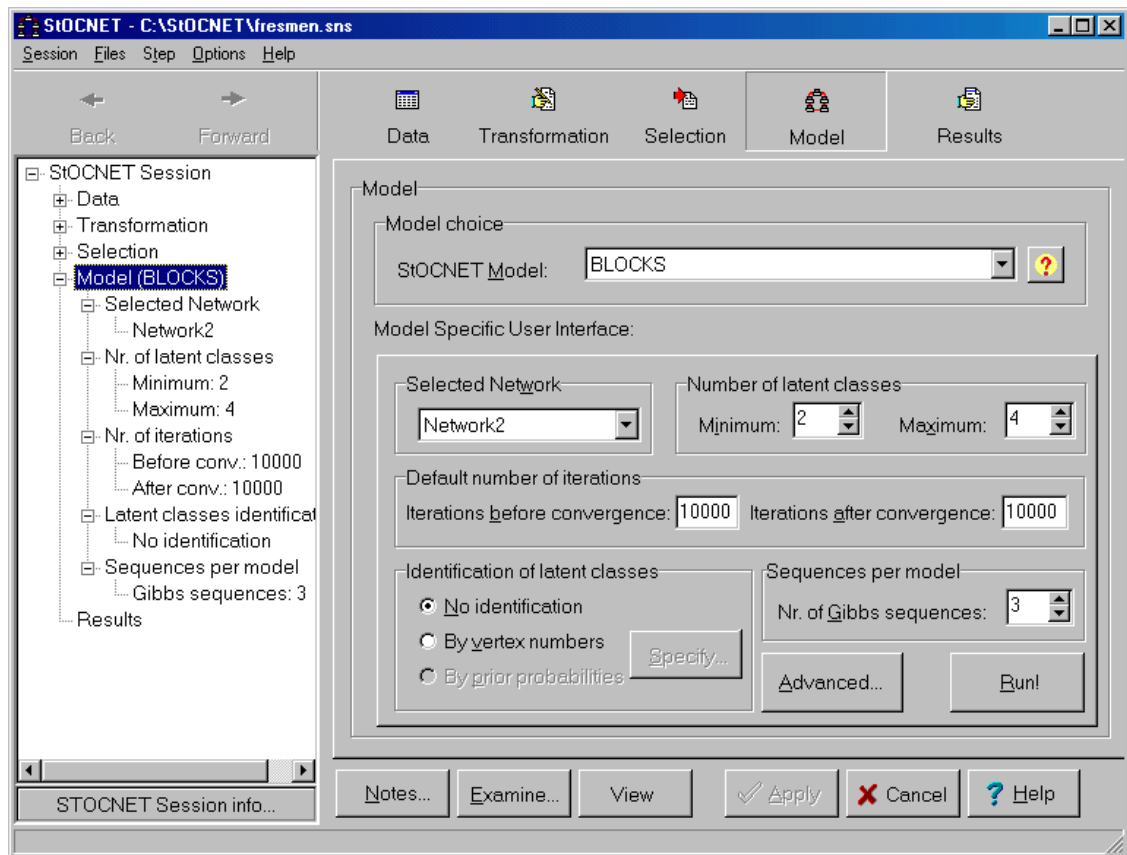


Figure 2. BLOCKS – Model-specific user interface.

In Figure 2 the model-specific user interface for BLOCKS is shown. The options to be specified by the user are the network data set, the number of latent classes, the number of iterations, and the number of Gibbs sequences in the estimation procedure. If prior information is available on the latent classes, the classes can be identified. Identification can be accomplished in two ways: 1) by specifying for each class one vertex number that has a high probability of belonging to that class, or 2) by specifying prior probabilities for a vertex to belong to a certain class.

University freshmen

The data observed on t_2 are analyzed with the program BLOCKS to find distinct groups of students. Solutions with 2, 3, and 4 blocks of students were studied. In Table 1 the values of two fit indices, I_y and H_x (Nowicki and Snijders, 2001), are presented; I_y indicates the extra information contained in observing the relations, if grouping is known a priori; H_x indicates the clarity of the block structure. The indices have values between 0 and 1, with 0 indicating a good fit.

Table 1: Goodness-of-fit of three blockmodels.

| Blocks | I_y | H_x |
|--------|-------|-------|
| 2 | 0.633 | 0.673 |
| 3 | 0.572 | 0.576 |
| 4 | 0.558 | 0.500 |

Inspection of the fit indices shows that there is little improvement in model fit if more than 2 blocks are distinguished. Other results of the analysis (not reported here), show that the 2 blocks solution is the best, and distinguishing more groups does not result in stable results. In the data at t_2 two groups of students can be distinguished. Inspection of the attributes of the students shows that these two groups correspond exactly with one of the opportunity variables, program. The groups consist of those students in one of the two study programs: the regular four-year program, or the short two-year program for students with previous training.

3.2 p_2

The module p_2 (version 2.00) is designed for the analysis of binary social network data with actor and/or dyadic covariates. The program carries out the estimation of a random effects model with the dyadic ties as the dependent variable according to the Iterative Generalized Least Squares algorithm for nonlinear multilevel models as described by Van Duijn (1995). For more detailed information the reader is referred to Van Duijn (1995) and Lazega and Van Duijn (1998).

The p_2 model is a type of logistic regression model for the ties in a network, to which a reciprocity effect is added as well as random sender and receiver effects (representing differential activity and popularity, respectively). In that respect it can be regarded as an extension of the well-known p_1 model (Holland and Leinhardt, 1981), where the actor parameters are replaced by random effects and actor and dyadic attributes can be included. Fitting the p_1 model in StOCNET is discussed in Section 5.

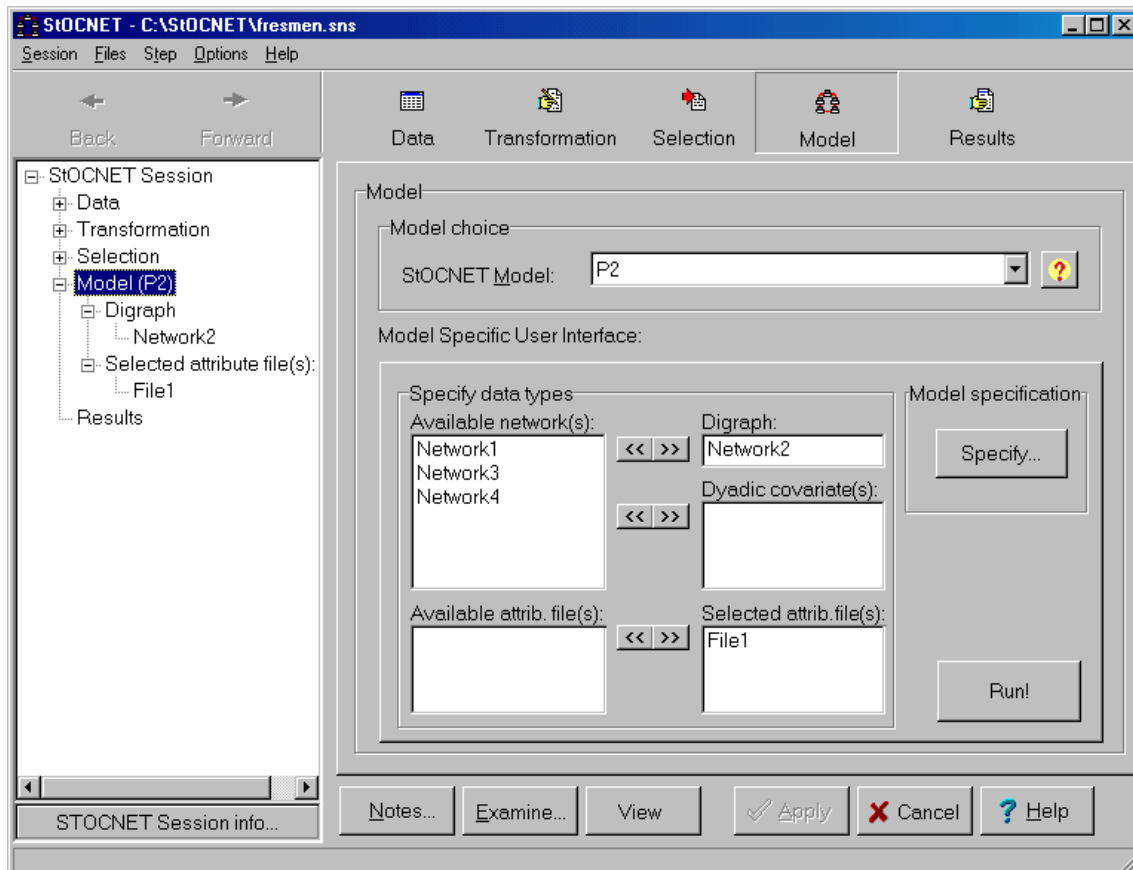


Figure 3. p_2 – Model-specific user interface.

In Figure 3 the model-specific user interface for p_2 is shown. The options that have to be specified by the user are the network and actor attribute data, and the model parameters, that is, the effects to be included in the model. There are four different kinds of effects, density, reciprocity, sender, and receiver effects, which are all based on the covariates.

University freshmen

The data observed at the second time point t_2 are analyzed with the p_2 model. Only the actor attributes program, smoking, and gender are included. The results are presented in Table 2.

Table 2: Parameters estimates for the university freshmen data at observation time t_2 , obtained with the program p_2 (only significant effects).

| | Parameter | Est. | (s.e.) |
|-----------------|---------------------|-------|--------|
| Sender | Variance s^2_A | 0.68 | -0.19 |
| Receiver | Variance s^2_B | 1.21 | -0.27 |
| Sender-receiver | Covariance s_{AB} | -0.09 | -0.18 |
| Density | μ | -4.21 | -0.38 |
| | Similarity program | 1.9 | -0.26 |
| | Similarity gender | 1.16 | -0.22 |
| Reciprocity | ρ | 1.12 | -0.3 |

In the p_2 analysis no significant sender and receiver effects are found. Similarity with respect to program and gender both have a significant and positive effect on density (i.e., there are more relations between students in the same program, between male students, and between female students). There is a general reciprocity effect, but this is not differentiated according to similarity variables. The receiver variance is much larger than the sender variance, due to the missing reports of 13 of the 38 students, whereas those missing students did receive choices of the others.

3.3 SIENA

The module SIENA (Simulation Investigation for Empirical Network Analysis; version 1.97) is a program that carries out the statistical estimation of models for the evolution of social networks according to the dynamic actor-oriented model of Snijders (2001). For more detailed information on the program or on stochastic actor-oriented models, the reader is referred to Snijders (2001) and to the user's manual of SIENA (Snijders and Huisman, 2002).

Stochastic actor-oriented models are used to model longitudinal network data. The dependent variable is the evolving relation network, represented by repeated measurements of a directed graph. The network evolution is modeled as the consequence of actors initiating new relations or withdrawing existing relations such that a more rewarding configuration for the actor in the network emerges. This goal is modeled in a so-called objective function the actors try to maximize. The models are continuous-time Markov chains that are implemented as simulation models.

In Figure 4 the model-specific user interface for SIENA is shown. The model specifications are presented in four groups: specification of network types (digraphs and dyadic covariates), actor attributes (varying and non-varying covariates, and times of composition change), estimation options (model specification and advanced estimation options), and simulation options (to simulate the network evolution without parameter estimation).

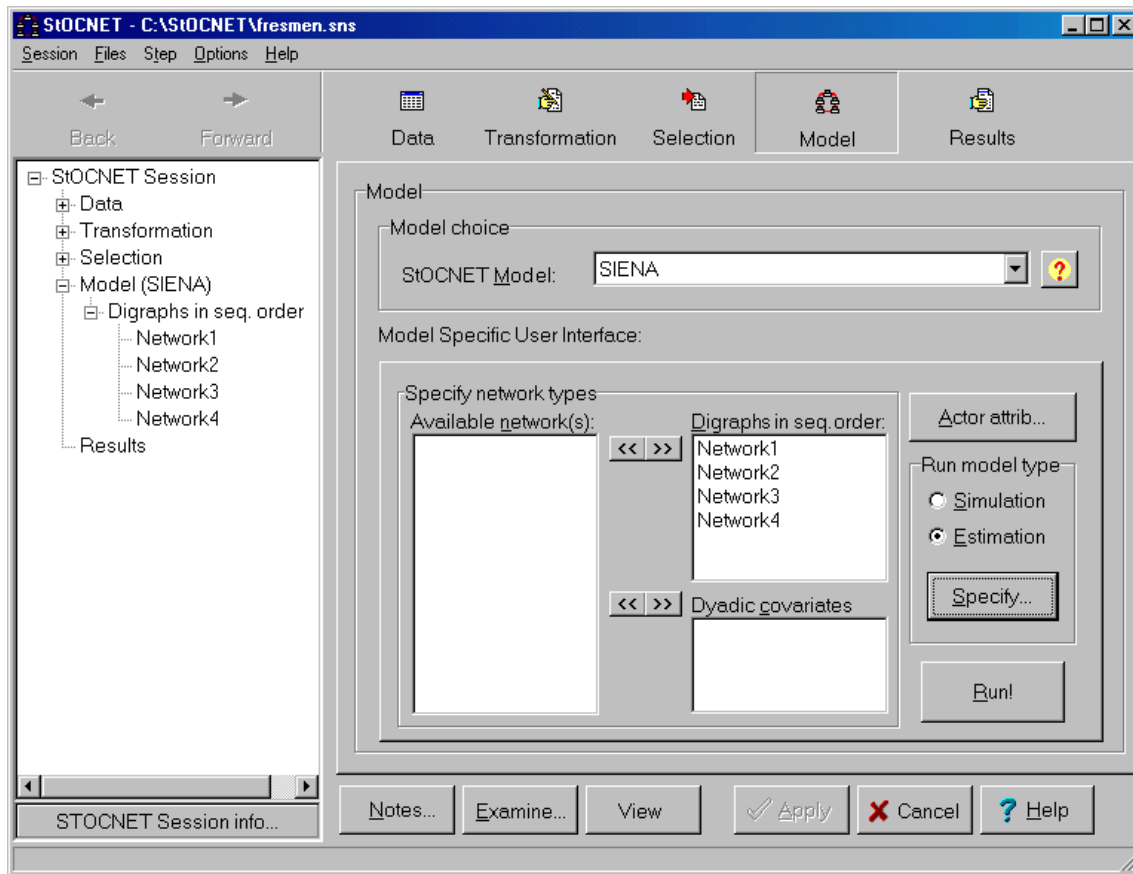


Figure 4. SIENA - Model-specific user interface

University freshmen

Data at four observation times (t_1 to t_4) are used to model the network evolution with the program SIENA⁸. The effects included in the actors' objective functions are reciprocity, balance, and popularity (network effects), and program, smoking behavior, and gender (actor attributes). For the three attributes dissimilarity effects are calculated, that is, the preference for dissimilar others. Four

⁸ For a more detailed discussion of the results see Van Duijn et al., 2003

different models are estimated: for the evolution from t_1 to t_2 , t_2 to t_3 , t_3 to t_4 , and the complete period t_1 to t_4 . The results (parameter estimates and standard errors) are presented in Table 3.

The rate parameters in table 3 shows that in each of the three periods the actors made about 5, 3, and 5 relationship changes on average, respectively. In each period the actors have a preference for establishing reciprocated relations (large, positive reciprocity effects) and balanced network configurations (closed networks). Only in the last period the actors show a preference for initiating relations with popular alters (i.e., alters with large in-degrees). The proximity variables program and smoking, and the visible attribute gender, are only important in the early stages of the evolutionary process. In the first period, students initiate new relations with others in the same study program as do smokers with other smokers. Also, women tend to choose other women and men other men (negative dissimilarities).

From the fourth model it follows that the attributes program and gender are important effects to model network changes for the complete time period, although they are only significant in the early stages of the evolutionary process. The popularity effect is also an important effect for the complete time period, whereas it is only significant in model 3.

Table 3: Parameters for models estimated using observations t_1 to t_4 of the freshmen data. Only the parameters for which the t -value (estimate divided by standard error) is larger than 1.5 are depicted.

| Effect | Model 1: $t_1 - t_2$ | | Model 2: $t_2 - t_3$ | | Model 3: $t_3 - t_4$ | | Model 4: $t_1 - t_4$ | |
|-------------------------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
| | Est. | (s.e.) | Est. | (s.e.) | Est. | (s.e.) | Est. | (s.e.) |
| Rate (period 1) | 5.26 | | | | | | 7.01 | |
| Rate (period 2) | | | 2.96 | | | | 3.06 | |
| Rate (period 3) | | | | | 5.30 | | 5.28 | |
| Density | - | | -1.61 | (0.48) | -2.17 | (0.84) | -1.52 | (0.56) |
| Reciprocity | 2.61 | (1.04) | 3.51 | (1.63) | 3.07 | (0.98) | 2.92 | (0.58) |
| Balance | 9.25 | (3.95) | 6.62 | (3.04) | 9.18 | (2.74) | 7.25 | (2.10) |
| Popularity | - | | - | | 6.55 | (2.05) | 7.08 | (1.23) |
| Program – dissimilarity | -1.10 | (0.19) | - | | - | | -0.79 | (0.13) |
| Smoking – dissimilarity | -0.37 | (0.22) | - | | - | | - | |
| Gender – dissimilarity | -0.60 | (0.22) | - | | - | | -0.40 | (0.14) |

3.4 p^* SIENA

Another application of the SIENA module is the estimation of the exponential random graph model, also called the p^* model (Frank and Strauss, 1986; Wasserman and Pattison, 1996), using Markov chain Monte Carlo (MCMC) methods described in Snijders (2002a). For more detailed information about the estimating the p^* model with SIENA the reader is referred to these articles and to the SIENA user's manual (Snijders and Huisman, 2002).

An exponential random graph model, that is, p^* model, is estimated if only one observation moment is specified in the SIENA program. The module carries out MCMC estimation for this model. If the algorithm works properly, the computed estimates are approximations of the maximum likelihood estimates. However, it is discussed in Snijders (2002a) that there are problems for estimating parameters of the likelihood distribution, and for many data sets it is impossible to achieve satisfactory estimates—perhaps it is next to impossible with any method. To use p^* SIENA it is advisable to first read Snijders (2002a). SIENA also gives maximum pseudolikelihood estimates (cf. Frank and Strauss, 1986; Wasserman and Pattison, 1996).

University freshmen

The data observed at t_2 , the first occurrence of the network, are analyzed with the program SIENA. The results are presented in Table 4. The estimated network effects are number of ties (density), reciprocity and transitive triplets. Although the models show the same general picture (also found in the longitudinal analysis), Table 4 shows some differences between the two estimation procedures. The MCMC method implemented in SIENA is recommended, because the properties of the pseudolikelihood estimators are unknown. It should be noted, however, that the convergence of the SIENA p^* model was doubtful. When the transitivity effect was deleted from the model, good convergence was obtained.

Table 4: Parameter estimates for the university freshmen data at observation time t_2 , obtained with the program SIENA (MCMC and pseudolikelihood).

| Effect | MCMC | | pseudolikelihood | |
|-------------------------|-------|--------|------------------|--------|
| | Est. | (s.e.) | Est. | (s.e.) |
| Number of ties | -3.61 | (0.96) | -5.09 | (0.39) |
| Reciprocity | 3.99 | (0.37) | 1.97 | (0.29) |
| Transitive triplets | 0.17 | (0.04) | 0.57 | (0.05) |
| Program – dissimilarity | -0.75 | (0.25) | -1.10 | (0.26) |
| Smoking – dissimilarity | -0.04 | (0.03) | -0.14 | (0.04) |
| Gender – dissimilarity | -0.41 | (0.16) | -0.80 | (0.22) |

3.5 ULTRAS

The module ULTRAS (version 1.1) aims at estimating latent transitive structures in social networks. The program estimates latent structures expressed by ultrametrics as presented by Schweinberger and Snijders (2003) and identifies and describes groups according to this latent structure. For more detailed information the reader is referred to Schweinberger and Snijders (2003) and to the user's manual of ULTRAS (Schweinberger, 2002).

Transitive structures can be used to identify close-knit subsets of actors in large social networks, which may contain even hundreds of actors. Latent transitive structures can be expressed by ultrametrics, measuring distances between actors based on triads. The module ULTRAS estimates ultrametrics, given one observed network, under the assumption that the probability of a tie depends

on the ultrametric distance between the two actors. The observed network may be generated by one, two, or even more ultrametrics. The program can handle dichotomous, discrete, as well as continuous network data. Maximum Likelihood Estimation is done using a non-greedy optimization algorithm, and Bayesian inference with uniform priors is implemented using hybrid MCMC methods (Schweinberger and Snijders, 2003).

In Figure 5 the model-specific user interface for ULTRAS is shown. To specify the model, the user has to determine the maximum value the ultrametric distance between any two actors may take on, and to indicate how the network ties are distributed (Bernoulli, Poisson, or Gaussian). To estimate the ultrametrics, either a Maximum Likelihood or Bayesian procedure can be used. Clicking the Specify-button gives the possibility to specify certain basic ingredients to the algorithm and output options.

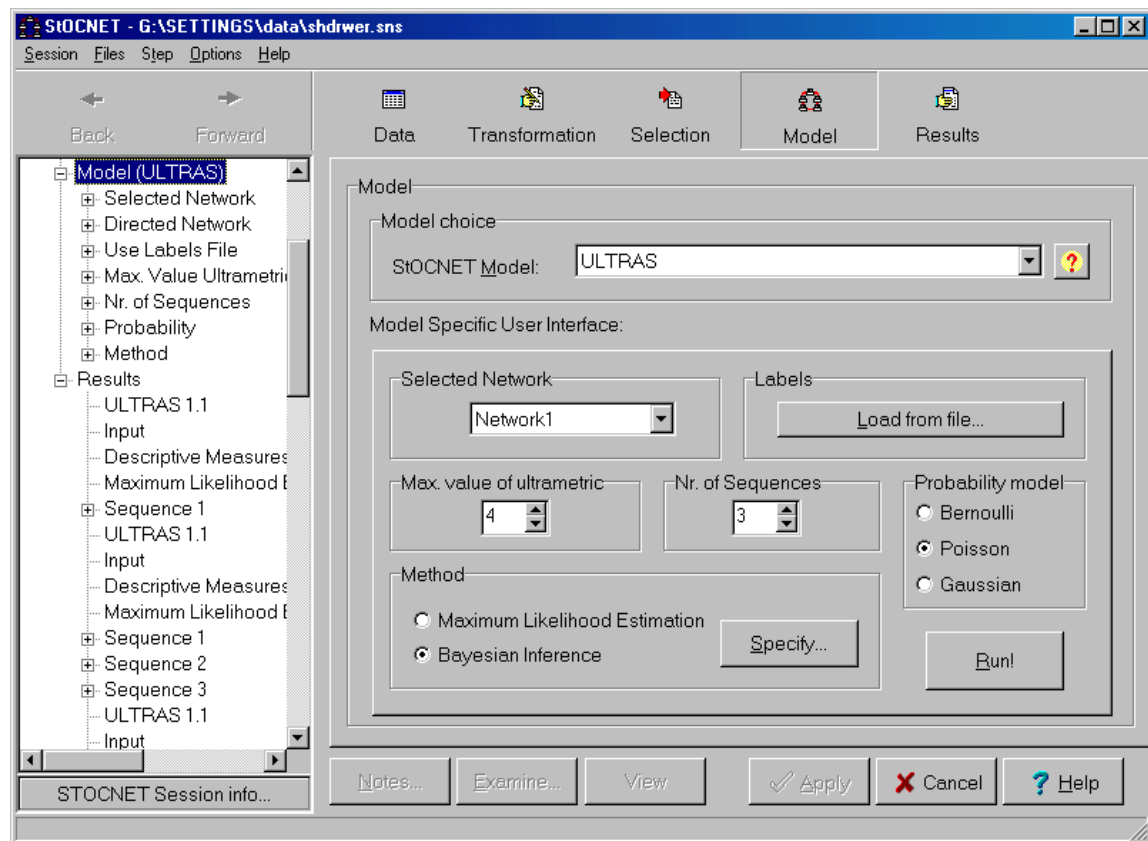


Figure 5. ULTRAS – Model-specific user interface.

University freshmen

The data observed at the second time point t_2 are analyzed with ULTRAS. Since the outcome spaces of the tie variables are discrete, the Poisson probability model is used. Maximum likelihood estimation is executed, assuming that one ultrametric has generated the observed network. The maximum log likelihood value is found to be -538.29 . The main results are presented in Table 5 on the next page.

Table 5: ML estimates and posterior means of theta for the university freshmen data at observation time t_2 , obtained with the program ULTRAS.

| Parameter | MLE | | Bayesian approach | |
|-----------|------|--------|-------------------|------------------|
| | Est. | (s.e.) | Posterior mean | Posterior (s.e.) |
| theta(1) | 2.86 | (0.19) | 2.34 | (0.40) |
| theta(2) | 1.17 | (0.07) | 1.78 | (0.15) |
| theta(3) | 0.65 | (0.10) | 0.84 | (0.09) |

The estimates of $\theta(h)$ have to be interpreted as the estimated tie value, given that the ultrametric distance between the actors takes the value h . The ultrametric which maximizes the log likelihood function (not shown) reveals that, given distance $h = 2$, there is one large group and one small group. The ML-ultrametric can also be represented as an ultrametric tree, similar to a dendrogram in cluster analysis.

Executing Bayesian inference, the average log likelihood value over the ultrametries sampled from the posterior distribution is found to be -572.60 . The posterior standard deviations (and other output not presented here) indicate that the model is uncertain about the latent structures in the network. This could be due to the small network size and the many missing values. The Bayesian procedure shows that actors 18 and 32 may decrease the global model fit. When these actors are excluded from the analysis, the average log likelihood changes to -526.41 , improving the model fit.

3.6 ZO

ZO (Zero-One; version 2.3) is a program that calculates the probability distribution of any statistic of random graphs with given degrees and random digraphs with given in- and out-degrees. In addition it is possible to request a given number of mutual dyads and/or a connected graph. ZO carries out simulation and/or enumeration analysis of the graphs according to the algorithms of Snijders (1991) and Molloy and Reed (1995). For more detailed information on the program or probability distributions of statistics of random (di)graphs, the reader is referred to Snijders (1991) and Molloy and Reed (1995), and the user's manual of ZO (Snijders, 2002b).

ZO determines the distribution of statistics for general (rectangular) random 0-1 matrices with given row and column sums, including matrices with structural zeros (i.e., the restriction that a given set of entries is equal to 0). The distribution of the random 0-1 matrices is uniform, that is, each matrix satisfying the restrictions has the same probability of occurrence. The results include p-values that can be used for testing reciprocity or transitivity while controlling for the in- and out-degrees. For very small matrices (up to 8-12 rows and columns) this can be done by enumerating all matrices satisfying these constraints; the simulation method is more generally applicable.

In Figure 6 the model-specific user interface for ZO is presented. The user has to specify the following: 1. network data (adjacency matrix or separate file containing only row and column sums), 2. type of matrix to be analyzed, 3. the run-model type (simulation or enumeration), and 4. options

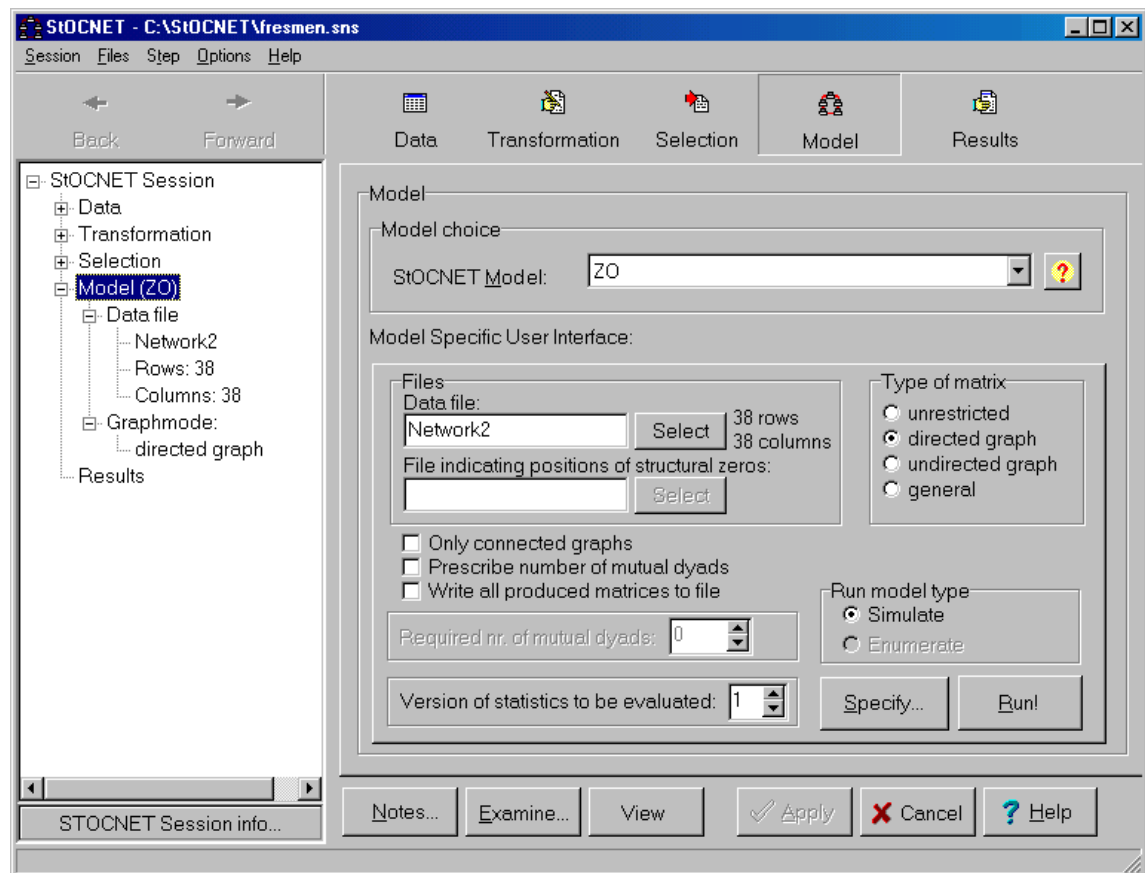


Figure 6. ZO — Model-specific user interface.

that depend on previous choices (e.g., specifying that the graphs that are generated by ZO should be connected – only for graph-type matrices, or specifying the number of mutual dyads – only for digraphs).

University freshmen

The data observed at the second time point t_2 are analyzed with the ZO model. Part of the ZO results are presented in Figure 7 on the next page.

A test for reciprocity is carried out on the freshmen data by testing the number of mutual dyads M in the $U | X_{i+}, X_{+i}$ distribution (see Wasserman and Faust, Ch. 13). The observed number of mutual dyads in the freshmen data is 136. With 10,000 simulation runs, a probability distribution of the mutual dyads was calculated (presented in Figure 5). The observed number of 136 is so high under the null distribution, that such high values did not even occur once in 10,000 simulations (the maximum number was 124). It can be concluded that there is a very significant tendency toward reciprocity, given the in- and out-degrees.

In the same way a tendency towards transitivity is tested; null distributions will be the $U | X_{i+}, X_{+i}, M$ distribution, with $M = 136$ (because the observed value of M is very unlikely giving the $U | X_{i+}, X_{+i}$ distribution, the number failed attempts will be large and therefore the number of simulations was reduced to 1000 to make it less time-consuming). The observed transitivity is 0.87, which is

larger than the maximum value found in the simulations (concentrating round 0.85, variance 0.08), indicating a strong significant tendency towards transitivity, controlling for in- and out-degrees and the number of mutual dyads.

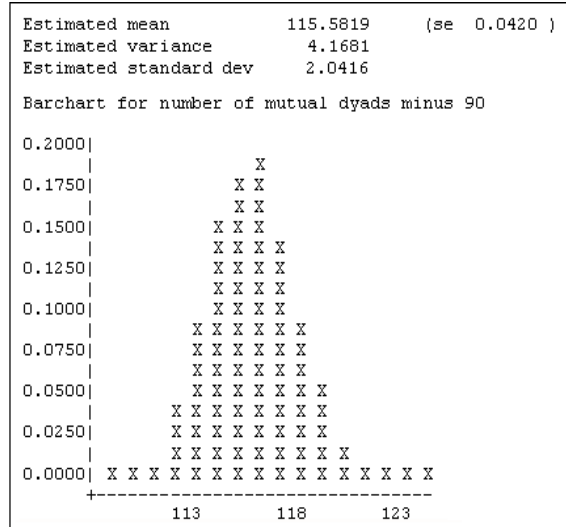


Figure 7. Part of the ZO output.

4. DESCRIPTIVE STATISTICS: EXAMINE

In four of the five steps of a StOCNET-session, the user has the opportunity to examine the data that are available up to and including the functionalities of the current step. After specification of the relevant options in each step-specific interface, clicking the *Examine-button* results in some descriptive analyses of the network and attribute data. The Examine-button is active in the steps Data definition, Transformation, Selection, and Model Specification, but not in the last step, Results.

In every step, clicking the Examine-button gives descriptives of all available network and attribute data (i.e., all network files and all attributes in all attribute files). The only exception is the fourth step (specification of a statistical model), where only those data files are analyzed that are specified in that step. Each time the specifications are changed in some session-step, for example, when the data are recoded (step 2) or when a different set of actors is selected (step 3), descriptives of the 'new' data sets are presented.

The following network descriptives are available in the current version of StOCNET (some statistics are only available for symmetric and/or dichotomous networks; see the StOCNET user's manual for more details):

- Step 1. Relation count and dyad count.
- Step 2. Missing values, density, degrees, degree variances, index of heterogeneity, dyad census, degree of reciprocity, triad count, (normalized) degree of transitivity, triad census, degree of segmentation, number of isolates and number of weak components.
- Step 3. Same descriptives as in step 2.

Step 4. For the program p_2 , clicking the Examine-button results in fitting the p_1 model (Holland and Leinhardt, 1981). For the program SIENA, clicking the Examine-button results in tables containing change statistics (change in arcs, dyads, and triplets). For the other three programs (BLOCKS, Ultras, and ZO) there are no descriptive analyses available.

It should be noted that apart from the descriptive statistics that are generated by the examine-function, the output of some of the available models also contains some descriptive statistics, which are too important to miss (according to the contributors of the programs).

5. PLANNED ACTIVITIES

Some new StOCNET options and functionalities, as well as some new modules will become available in future versions of the program. The planned activities are presented in this section.

- Modules PALNET and PACNET: Pattison and Wasserman (1995) and Pattison, Wasserman, Robins, and Kanfer (2000) present methods for the construction and fitting of structural models for local (i.e., ego-centered) and complete social networks, respectively.

The methods proposed by Pattison and Wasserman (1995) give researchers a tool to adequately summarize social networks with different types of ties. They define partial algebraic structures from the collection of network paths that have a focal individual as their source, and present a method for deriving algebraic representations from local network data. The statistical criteria used in the method are based on permutation distributions. The methods are implemented in the program PALNET with which partial algebras can be constructed from network paths with a fixed, maximum length.

The purpose of the program PACNET is to construct a partial algebra for a complete network, using statistical criteria based on conditional uniform random graph distributions. Both PALNET and PACNET are written in C and still only available from the authors. However, the programs are in the process of being implemented in the StOCNET system and thus will become available to a wider audience.

In order to provide a new platform to make statistical programs available to a wider audience, the StOCNET system was set up in such a way that new modules be implemented with as little effort as possible. New contributions can be implemented as executables or as DLLs, and their source codes are allowed to be written in a large variety of programming languages (e.g., Delphi, C, C++).

The platform with its common data structure and user interface is provided by the StOCNET system, and the programs containing the statistical methods are treated as black boxes. All procedures will have similar interfaces and the contributors therefore only need to provide information with respect to data input, data representation, data output, parameter restrictions and so forth. Moreover, the procedures should have some general properties:

- proper documentation,
- the status of the calculations sent to the screen,

- user break possibility,
- proper error handling and error messages through error or log files,
- correct memory handling and allocation, and
- an ASCII input file for model definition and parameter input.

News about StOCNET can be found at the StOCNET website at <http://stat.gamma.rug.nl/stocnet/>. Here, new versions of the program and the corresponding documentation will be available for downloading. Also a brief history of the project is given, and of its goals and team members. Instructions for new contributors (including a questionnaire to be filled-in in order to facilitate inclusion of potential contributions) can be found as well, together with a list of potential contributors, whose programs are candidates for implementation in StOCNET.

REFERENCES

- Batagelj, V. and A. Mrvar. 2002. *Pajek. Version 0.84*. Ljubljana: University of Ljubljana. <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
- Boer, P., M. Huisman, T.A.B. Snijders, and E.P.H. Zeggelink. 2002. *StOCNET: An open software system for the advanced statistical analysis of social networks. Version 1.3*. Groningen: ProGAMMA/ICS. <http://stat.gamma.rug.nl/stocnet/>.
- Borgatti, S.P., M.G. Everett, and L.C. Freeman. 1999. *UCINET V for Windows: Software for social network analysis*. Natick: Analytic Technologies. http://www.analytictech.com/ucinet_5_description.htm.
- Burt, R.S. 1991. *STRUCTURE. A general purpose network analysis program providing sociometric indices, cliques, structural and role equivalence, density tables, contagion, autonomy, power and equilibria in multiple network systems. Version 4.2*. New York: Columbia University.
- Cyram. 2002. *Cyram NetMiner. Version 1.2.0*. Seoul: Cyram Co., Ltd. <http://www.netminer.com>.
- Frank, O. and D. Strauss. 1986. Markov graphs. *Journal of the American Statistical Association*, 81: 832-842.
- Holland, P.W. and S. Leinhardt. 1981. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76: 33-50.
- Huisman, M. and M.A.J. Van Duijn. 2002. *StOCNET in development: An update*. Paper presented at the International Sunbelt Social Network Conference XXII, New Orleans, USA. February 13-17, 2002. <http://stat.gamma.rug.nl/stocnet/>.
- Huisman, M. and M.A.J. Van Duijn. 2003. Software for social network analysis. To appear in P.J. Carrington, J. Scott, and S. Wasserman (Eds.) *Models and methods in social network analysis*. New York: Cambridge University Press. Preprints are available from the authors
- Huisman, M. and T.A.B. Snijders. 2003. *Statistical analysis of longitudinal network data with changing composition*. [submitted]

- Lazega, E. and M.A.J. Van Duijn. 1997. Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data. *Social Networks*, 19: 375-397.
- Molloy, M. & B. Reed. 1995. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6, 161-179.
- Nowicki, K. and T.A.B. Snijders. 2001. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96: 1077-1087.
- Pattison, P. and S. Wasserman. 1995. Constructing algebraic models for local social networks using statistical methods. *Journal of Mathematical Psychology*, 39: 57-72.
- Pattison, P., S. Wasserman, G. Robbins, and A. Kanfer. 2000. Statistical evaluation of algebraic constraints for social networks. *Journal of Mathematical Psychology*, 44: 536-568.
- Richards, W.D. and A.J. Seary. 1993. *FATCAT. Version 4.2*. Burnaby: Simon Fraser University. <http://www.sfu.ca/~richards/>.
- Richards, W.D. and A.J. Seary. 2002. *MultiNet. Version 4.0 for Windows*. Burnaby: Simon Fraser University. <http://www.sfu.ca/~richards/>.
- Schweinberger, M. 2002. *Manual for ULTRAS. Version 1.1*. Groningen: ICS/Dept. of Statistics & Measurement Theory, University of Groningen.
- Schweinberger, M. and T.A.B. Snijders. 2003. Settings in social networks: Represented by latent transitive structures. [submitted]
- Sprenger, C.J.A. and F.N. Stokman. 1989. *GRADAP 2. GRaph Definition and Analysis Package*. Groningen: ProGAMMA.
- Snijders, T.A.B. 1991. Enumeration and simulation models for 0-1 matrices with given marginals. *Psychometrika*, 56: 397-417.
- Snijders, T.A.B. 2001. The statistical evaluation of social network dynamics. In M.E. Sobel & M.P. Becker (Eds.) *Sociological Methodology*. (pp. 361-395) London: Basil Blackwell.
- Snijders, T.A.B. 2002a. Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3 (2). <http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html>
- Snijders, T.A.B. 2002b. *Manual for ZO version 2.3*. Groningen: ICS/Dept. of Statistics & Measurement Theory, University of Groningen. <http://stat.gamma.rug.nl/stocnet/>.
- Snijders, T.A.B. and M. Huisman. 2002. *Manual for SIENA. Version 1.95*. Groningen: ICS/Statistics & Measurement Theory, University of Groningen. <http://stat.gamma.rug.nl/stocnet/>.
- Snijders, T.A.B. and K. Nowicki. 2001. *Manual for BLOCKS. Version 1.53*. Groningen: ICS/Statistics & Measurement Theory, University of Groningen. <http://stat.gamma.rug.nl/stocnet/>.
- Snijders, T.A.B. and M.A.J. van Duijn. 2002. Conditional maximum likelihood estimation under various specifications of exponential random graph models. In Jan Hagberg (Ed.), *Contributions to social network analysis, information theory, and other topics in statistics; A festschrift in honour of Ove Frank* (pp. 117-134). University of Stockholm, Department of Statistics.

- Van Duijn, M.A.J. 1995. Estimation of a random effects model for directed graphs. In: SSS'95. *Symposium Statistische Software, nr. 7. Toeval zit overal: programmatuur voor random-coëfficiënt modellen*, T.A.B. Snijders, ed., 113-131. Groningen: ProGAMMA.
- Van Duijn, M.A.J., E.P.H. Zeggelink, M. Huisman, F.N. Stokman, and F. Wasseur. 2003. Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology*. [in press]
- Wasserman, S. and K. Faust. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Wasserman, S. and P. Pattison. 1996. Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika*, 61: 401-425.
- Zeggelink, E.P.H., T.A.B. Snijders, and P. Boer. 2000. *StOCNET in development*. Paper presented at the International Sunbelt Social Network Conference XX, Vancouver, Canada. April 13-16, 2000.