The SAGE Handbook of Social Network Analysis

Statistical Models for Ties and Actors

Contributors: Marijtje A.J. van Duijn & Mark Huisman
Edited by: John Scott & Peter J. Carrington
Book Title: The SAGE Handbook of Social Network Analysis
Chapter Title: "Statistical Models for Ties and Actors"
Pub. Date: 2014
Access Date: February 14, 2019
Publishing Company: SAGE Publications Ltd
City: London
Print ISBN: 9781847873958
Online ISBN: 9781446294413
DOI: http://dx.doi.org/10.4135/9781446294413.n31
Print pages: 459-483

Statistical Models for Ties and Actors

Marijtje A.J.van Duijn and MarkHuisman

Introduction

Can we predict friendship between researchers if we know how many email interactions they have? Do researchers on a conference prefer to get acquainted with colleagues from the same research area or do they interact instead with colleagues who have a high citation index? Could it be that email contact, homophily, and scientific status are all important to explain friendship or acquaintanceship, and if so, which of these effects is strongest? Can we distinguish different groups of researchers based on how well they know each other, and if so, is this grouping related to a research field or scientific status?

To answer these types of questions we need statistical models that can deal with the combination of network data and individual actor and/or dyadic attributes. They are presented in this chapter. The statistical models are categorized by the type of research question they can handle. Two streams of analysis are distinguished: the first are relationship-level models, modeling the ties between actors, and the other are actor-level models with an emphasis on differences between or grouping of actors.

The first three questions stated above can be answered with models that explain or predict the occurrence or value of ties in the network, using additional information on the relationships or actors if available. The last question is addressed using stochastic a posteriori block models that categorize or group actors based on their ties to each other, again using additional covariate information if available. The principle of stochastic blockmodels is to identify (groups of) stochastically equivalent actors, that is, actors who have the same probability distribution of ties to the other actors.[1]

The two different modeling approaches can be viewed as analyses of the same data (a social relational system according to Wasserman and Faust [1994: 89]) with a different emphasis. In the models for ties, the focal variable is formed by the relationship, expressed as a tie variable or pair of tie variables in a dyad whose outcome is observed and may be explained by attributes. In the models for actors, the focal variable is the group membership of the actors, expressed as a latent (unobserved) actor variable, whose value may be derived from the observed ties between actors and additional actor and dyadic characteristics.

The models presented here can be viewed as a sequel to the two earlier chapters by Hanneman and Riddle (this volume) that described the basic concepts of the analysis of egocentric networks and of complete network data, including some of their statistical properties used for testing simple hypotheses. Other introductions to statistical network models can be found in Wasserman and Faust (1994) and Scott (2000). The exponential random graph models treated by Robins (this volume) have the same goal as the models for ties in this chapter but model the complete network using statistics to represent more complex dependence structures than the dyadic dependence that is handled by most of the models presented in this chapter. The stochastic actor-oriented or actor-based models for longitudinal network data in the chapter by Snijders (this volume) are aimed at explaining changes in the observed networks and in the observed actor characteristics. O'Malley and Marsden (2008) present a broad overview of social network analysis with illustrations of exponential random graph models and individual outcome regression models. Goldenberg et al. (2009) review an even wider range of statistical models for social network analysis.

The methods and models are presented in a nontechnical manner, avoiding the use of formulas. Our aim is to convey the main objective of a model and relate it to the research question it can answer. Moreover, we try to compare and link models that, although their definition or assumptions may differ, answer similar questions. We use the EIES data (Freeman and Freeman, 1979) and various software programs to illustrate the methods and models presented in this chapter.[2]

Description of the Eies Data

The EIES data (Freeman and Freeman, 1979, taken from Wasserman and Faust, 1994, Appendix B, Tables 8–11) contain two observations of an acquaintanceship network. Acquaintanceship is a directed valued network, measured on a five-point scale ranging from (0) “don't know the other” to (4) “close personal friend.” Complete network data are available for 32 out of 50 researchers participating in a study carried out in 1978, which investigated the influence of electronic communication using the Electronic Information Exchange System (EIES). Participants were able to use the then novel technology to send each other email. The number of messages were recorded over the eight-month period that the experiment lasted and can thus be regarded as a directed valued network. The acquaintanceship network was determined at the beginning (time 1) and at the end of the study (time 2). In addition to the acquaintanceship and communication relations, actor attributes are available: The number of citations of the researchers, which we consider as a scientific status measure, and their primary disciplinary affiliation (research field). Four categories are distinguished: sociology, anthropology, mathematics/statistics, and a rest category “other” discipline. Because of the skewedness of the number of messages and the number of citations, the square roots of these (sometimes large) numbers were used in the analyses. Two more sociomatrices were constructed: the first by taking the difference in status, to express the status hierarchy between the researchers, and the second by taking the absolute difference in status, measuring their distance.[3] Moreover, a (symmetric adjacency) matrix was created to indicate the similarity or homophily of the researchers with respect to their research field (including “other”).

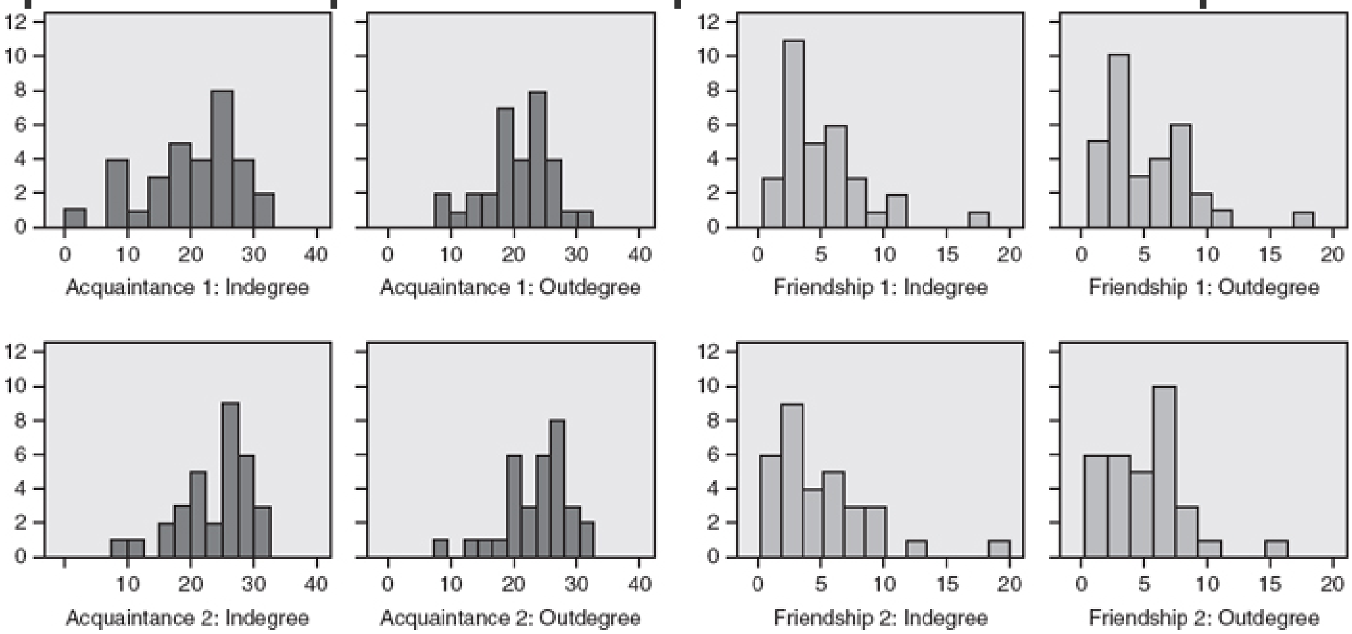
The data are summarized in Table 31.1. For most social network analyses the acquaintanceship network(s) had to be dichotomized, which we chose to indicate friendship: categories 3 and 4 (“friend” or “close personal friend”) versus “did not know,” “had heard of but not met,” and “had met” the other, categories 0–2. Measures are calculated on the complete network (n = 32), and on subnetworks defined by research field. The summary statistics were calculated with the programs StOCNET (Boer et al., 2006) and NetMiner 3 (Cyram, 2009).

Table 31.1 Description of the EIES network attribute data obtained with StOCNET and NetMiner

	Research field				
	Socio.	Anthro.	St/Ma.	Other	Total
n	17	6	3	6	32
Acquaintance 1					
Mean sum	28.88	12.50	3.67	4.33	42.44
SD column sum[1]	8.64	2.59	2.31	2.25	17.36
SD row sum[2]	6.98	3.15	1.53	3.08	13.29
Acquaintance 2					
Mean sum	31.76	13.17	4.67	5.00	52.09
SD column sum[1]	8.09	3.65	2.31	2.19	16.39
SD row sum[2]	7.92	2.32	0.58	3.52	14.33
Friendship 1					
Mean degree	3.53	2.50	0.67	0.17	4.78
SD indegree	2.15	1.38	0.47	0.37	3.47
SD outdegree	2.52	1.38	0.47	0.37	3.53
Density	0.22	0.50	0.33	0.03	0.15
Reciprocity[3]	0.67	0.53	1.00	0.00	0.56
Transitivity[4]	0.44	0.50	–	–	0.38
Friendship 2					
Mean degree	4.82	2.67	0.67	0.50	6.38
SD indegree	2.73	1.80	0.47	0.50	4.78
SD outdegree	3.00	0.94	0.47	0.50	3.82
Density	0.30	0.53	0.33	0.10	0.21
Reciprocity[3]	0.66	0.63	1.00	0.67	0.59
Transitivity[4]	0.44	0.50	–	0.00	0.40
Citations (square root)					
Mean	4.61	1.93	3.66	4.22	3.95
SD	2.27	0.72	1.85	4.68	2.75
Range	0.0–8.0	1.0–3.0	2.0–5.7	1.0–13.0	0.0–13.0
Messages sent (square root)					
Mean[5]	2.10	2.88	1.28	1.87	2.13
SD	2.47	2.06	0.90	1.81	2.15

For the acquaintance network the mean weighted (column and row) sum of the ties is 42.4 at the first time point, which increases to 52.1 at the second time point. The incoming tie sum has a larger standard deviation than the outgoing tie sum. Note that the acquaintance ties are valued (0–4), making the numbers somewhat hard to interpret. Therefore, the degree distributions of the dichotomized acquaintance network (0→0 and >0→1) are presented in Figure 31.1. Both indegree and outdegree distributions at the two time points are slightly left skewed. At time 1, the average number of ties is 20.3 (out of 31 possible relations; not reported in table), indicating that the mean value of these 20.3 relations is about 2.1 (“had met the other”). At time 2, the average number of ties has increased to 23.7, making the mean value slightly larger, 2.2. The acquaintance networks of the smaller groups of anthropologists and the statisticians/mathematicians have on average the strongest relations: at time 1 the mean values are 2.7 and 2.5, respectively, and at time 2 the mean values are 2.8 and 2.35 (numbers not presented in Table 31.1).

Figure 31.1 Degree distributions of the dichotomized acquaintanceship and friendship network at both time points

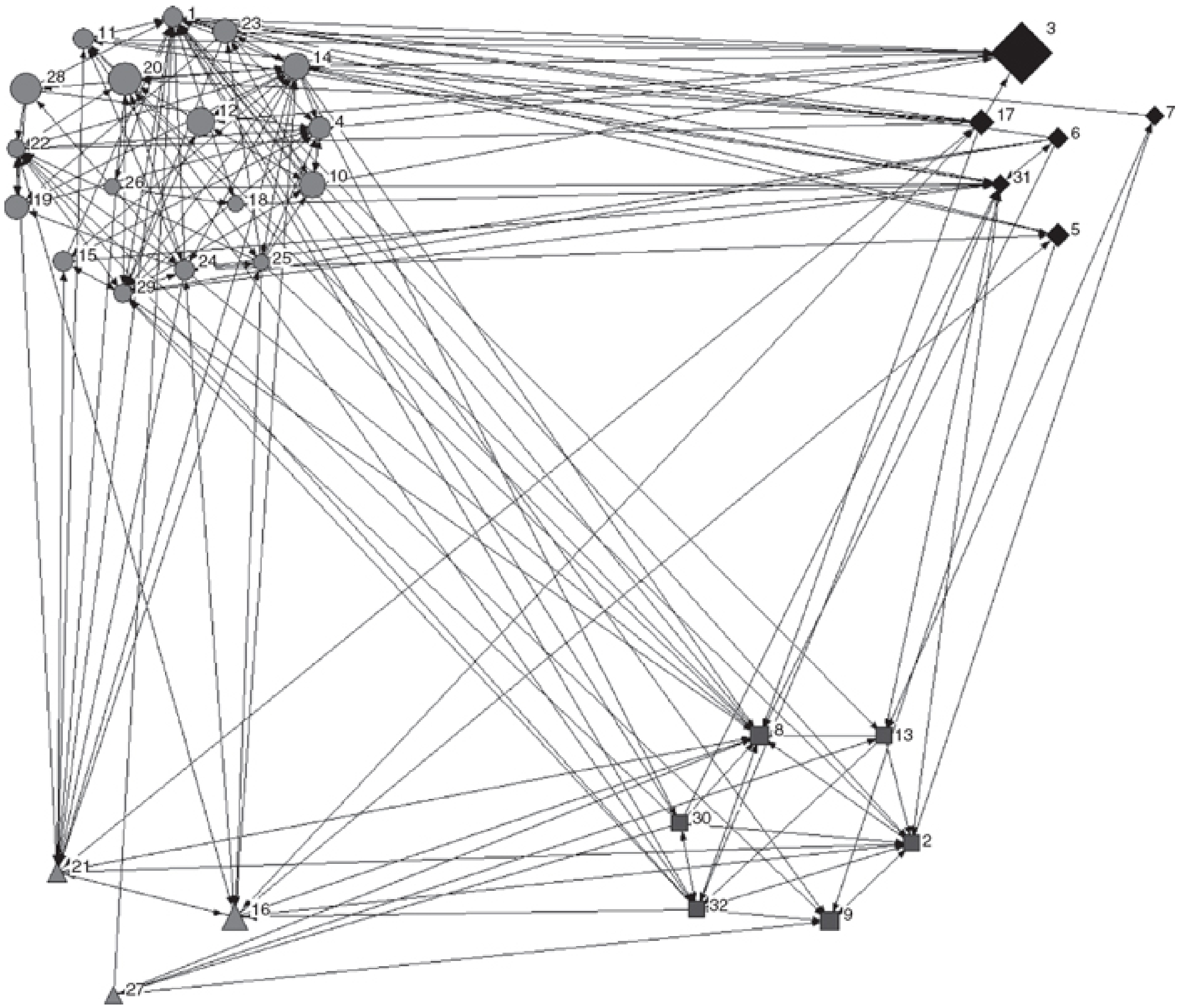


The friendship network has mean degrees of 4.8 and 6.4 at times 1 and 2, respectively, with densities 0.15 and 0.21. The increase in the number of friends is largest in the sociologist network, whereas in the other fields the mean degrees stay more or less the same. Thus, friendships between actors of different fields were established. The subnetworks of the anthropologists and statisticians/mathematicians have the highest densities (at both time points). All (sub)networks have become slightly denser at the second time point, with more mutual relations, except the (small) network of the statisticians/mathematicians, which did not change. The statisticians/mathematicians have the highest reciprocity score, 1, because there is one mutual dyad and two null dyads in this group of three. At the second time point, all (sub)networks have reciprocity scores of 0.59 or higher, showing that friendship nominations are often reciprocated. Transitivity only increased slightly in the total network and remained unchanged for the networks of sociologists and of anthropologists. It is not defined for the small group of statisticians/mathematicians at both time points and at time 1 for the group of “other” researchers.

For citations, messages sent, and messages received, the mean score, standard deviation, and range are reported in Table 31.1. The fields of sociologists and “other” contain actors with the highest number of citations and the highest variation in number of citations. The anthropologists have the smallest number of citations, but on average the highest number of messages sent and received. Interestingly, the statisticians/mathematicians received more messages than they sent, where for the other fields the numbers are almost the same.

The friendship network at time 2 is depicted in Figure 31.2. The graph was made using the layout option group by categorical attribute in NetDraw (Borgatti, 2002) to show the ties within and between the different fields. In the upper left corner the largest group of 17 sociologists (represented by circles) is positioned. Their status (depicted by the node size) varies from small to medium, approximately. The researcher with the largest number of citations belongs to the group of six researchers with an “other” research field (reflected in the large standard deviation of the number of citations in Table 31.1). This group is much less dense (with only one mutual and one asymmetric dyad) than the groups with specific disciplines. This is visually most clear in the comparison with the group of the same size in the lower right corner, consisting of six anthropologists with relatively low status. The small group in the lower left corner are the three statisticians/mathematicians of whom one is connected mainly with the group of sociologists, the second mainly to the anthropologists, and the third, with the highest number of citations, has about the same number of ties to all three other groups.

Figure 31.2 EIES friendship network at time 2. The actors are distinguished by research field (node shape and color) and by number of citations (node size). The graph was made with NetDraw



Modeling Ties

An interesting research question for the EIES data concerns the impact of the communication during the experiment as measured by the number of messages sent on the acquaintanceship at time 2, taking into account the research area and status of the researchers, and possibly also their acquaintanceship at time 1. Such a question implies modeling one complete network (acquaintanceship at time 2) using covariate network information (communication, hierarchy, distance, acquaintanceship at time 1) as well as actor attributes (status, research field).

Several types of statistical models for (complete) network data are available. They differ in measurement level of the tie variable (dichotomous or continuous) on the one hand and the statistical modeling tradition on the other hand.

The section starts with QAP (Quadratic Assignment Procedure). This method is based on regression models and permutation tests for continuous tie variables (see Dekker et al., 2007 for an overview of its development). Next, the Social Relations Model (SRM) is explained. The SRM was first proposed by Kenny and La

Voie (1984) and is rooted in an ANOVA-tradition. It has been used in many applications in social psychology and other fields (see Kenny et al., 2006 for a more detailed account). An important extension is the random effects or multilevel model version of the SRM proposed by Snijders and Kenny (1999, see also Snijders and Bosker, 1999, Chapter 11.3.3). QAP and SRM assume continuous tie variables (or at least a relation with a sufficient number of values).

Based on loglinear modeling, Holland and Leinhardt (1981) proposed the $p1$ model for the analysis of binary complete network data. The $p2$ model (Van Duijn et al., 2004) combines this tradition with a random effects approach borrowed from the SRM, including actor and dyadic attributes. A different extension of the $p1$ model is the so-called p^* model or exponential random graph model (Wasserman and Pattison, 1996; Snijders et al., 2006; Robins et al., 2007). This model is not treated in this chapter but is the topic of the chapter by Robins (this volume). The final model shared under the models for ties is the bilinear model developed by Hoff (2005). The bilinear model resembles the SRM and $p2$ in its way of incorporating dyadic and actor covariates, with random and correlated sender and receiver effects. It is more than dyadic, however, because it incorporates additional parameters to capture third-order dependence (such as transitivity and balance) between the actors. The extra parameters define a (latent) space in which the actors are positioned relative to each other. The bilinear model is related to the latent cluster model (Handcock et al., 2007) presented later in this chapter.

QAP

When investigating the association between two (or more) network matrices one has to take into account the dependence inherent in the data due to the fact that actors send and receive multiple ties. Thus, the observed outcomes of the tie variable are not independent. One could use OLS correlation and regression models to estimate the association, implicitly assuming independence over and within dyads. This approach results in incorrect tests of the correlation and regression coefficients (due to underestimation of the standard errors). The quadratic assignment procedure (QAP), applied to social networks by Hubert (1987) and Krackhardt (1987), tests the null hypothesis of no correlation between two matrices, Y and X, by repeatedly permuting the order of the rows (and columns) of one of the matrices, Y, while keeping X intact. The resulting sample of product-moment correlations of the permuted Y and X after vectorization provides the distribution of the correlation coefficient under the null hypothesis to which the observed correlation can be compared. When the association between two matrices is investigated while controlling for a third matrix Z, a more complex procedure is needed because of the dependence between X and Z, comparable to multicollinearity in multiple regression (MR). Several MR-QAP procedures have been proposed, of which the residual permutation methods are found to perform best (Dekker et al., 2007). In these approaches the residuals of either the regression of Y on Z or X on Z obtained in a first step in the analysis are permuted and included in the regression equation in the second step to compute the association between Y and X controlling for Z. The latter approach is the Double-Semi-Partialing approach used in the application.

Application to Acquaintance Network

The QAP and MR-QAP methods are implemented in UCINET (Borgatti et al., 2002), and SNA (Butts, 2008). As a first step, correlations are computed between the five available social networks with continuous tie variables. These are presented in Table 31.2.

Table 31.2 QAP correlations for the EIES complete network data obtained with UCINET based on 5000 permutations (p-values in parentheses; QAP correlations of square-root transformed variables below the diagonal)

		1	2	3	4	5
1	Network time 1	–	0.800 (<0.0001)	0.214 (0.001)	−0.039 (0.153)	0.001 (0.490)
2	Network time 2	–	–	0.355 (<0.0001)	−0.043 (0.114)	−0.076 (0.215)
3	Communication	0.250 (<0.0001)	0.431 (<0.0001)	0.888 (<0.0001)	−0.020 (0.170)	−0.157 (0.001)
4	Hierarchy	−0.070 (0.031)	−0.060 (0.039)	−0.045 (0.077)	0.928 (<0.0001)	0.000 (1.000)
5	Distance	−0.126 (0.067)	−0.180 (0.021)	−0.189 (0.009)	0.000 (1.000)	0.905 (<0.0001)

The QAP correlations between the EIES networks are highest for the strength of acquaintanceship at time 1 and time 2 (0.80) and with the number of messages sent between the researchers stronger at time 2 than at time 1, even more so when the square-root transformed variable is used. The correlations of the absolute difference between actors in number of citations (distance) and the difference measure (hierarchy) with acquaintanceship at both time points are all negative, weak and only significant for the square root transformed variables, except for distance and acquaintance time 1.

The expected negative effect of distance reflects that researchers far apart in status are less well acquainted; the negative effect of hierarchy indicates the asymmetric phenomenon that researchers tend to rate the intensity of acquaintanceship higher with others who are "higher in rank" in terms of the number of citations than vice versa, which may be called a status effect or aspiration effect. There is some evidence of a weak relation between communication and distance, as researchers at a smaller distance tend to communicate more ($r = -0.16$ and -0.19 for the transformed variables). The correlations between the explanatory variables and their square-root transformed version are quite high, as expected.

In Table 31.3, the results of an MR-QAP analysis are given with the acquaintanceship at time 2 as outcome variable. We start with hierarchy and distance as explanatory variables to decide which of the two effects of citations is stronger: hierarchy or distance. Next, the effect of the messages sent is added to the model (Model 2 in Table 31.3). The results are computed using MR-QAP with the residual permutation method proposed by Dekker et al. (2007), called "Double Dekker Semi-Partialling" in UCINET (based on the 2000 permutations and the *t*-statistic). No *p*-value for the intercept is reported. The QAP and MR-QAP procedures are also available in the R package sna and take slightly longer to produce (almost) identical results.

Table 31.3 MR-QAP analyses results for the EIES acquaintanceship data at time 2 obtained with UCINET based on 2000 permutations (p-values in parentheses)

	Model 1	Model 2	Model 3
Intercept	1.866	1.458	0.576
	–	–	–
Communication		0.175	0.102
		(0.001)	(0.001)
Hierarchy	−0.017	−0.012	0.000
	(0.041)	(0.103)	(0.454)
Distance	−0.060	−0.047	−0.019
	(0.028)	(0.080)	(0.079)
Network time 1			0.735
			(0.001)
R^2 (adjusted)	0.035	0.196	0.698
	(0.012)	(<0.0001)	(<0.0001)

Model 1 in Table 31.3 shows that distance has a stronger effect on acquaintanceship than hierarchy. When communication (i.e., the square root of the number of messages sent) is added to the model, the effects of distance and hierarchy are reduced and not significant anymore. The effect of communication is strong, as is also shown by a large increase in (adjusted) R^2, the proportion of variance explained in the analysis. Adding the degree of acquaintanceship at time 1 as explanatory variable to the model increases the amount of explained variance even more. The parameter value of communication is smaller while still significant. The effect of hierarchy is gone, apparently completely absorbed in or represented by the acquaintanceship at the start of the experiment. Thus, evidence is found for a positive impact of communication on acquaintanceship.

Social Relations Model (SRM)

The Social Relations Model (SRM; Kenny and La Voie, 1984) assigns the variance observed in dyadic relations Y_{ij} to parts attributable to senders of the relations, to their receivers, and to their interaction. This definition makes it clear that the model was originally formulated as an ANOVA model (Kenny and La Voie, 1984). It was later reformulated as a random effects or multilevel model (Snijders and Kenny, 1999) incorporating explanatory dyadic covariates and using actor characteristics to further model (random) sender and receiver effects. Note that the dyadic and actor covariates can be categorical or continuous. The SRM can be regarded as a cross-nested multilevel (regression) model where dyads are nested within actors, who for each of the directed relations act as sender or receiver (Snijders and Bosker, 1999, Chapter 11).[4]

The SRM requires continuous dyadic outcomes in order to make the assumption of a normal error distribution of the residuals at tie-level and at actor-level. The dependence between the two roles of each actor is represented by the covariance between the *random* sender and receiver effects. The random effects and their (co)variances can be viewed as measures of (unexplained) actor sociability. In multilevel terminology, the regression parameters for the actor and dyadic covariates are *fixed* effects. The SRM does not require complete network data and is easily extended to the situation of multiple (independent) networks by adding an extra level (see, for example, Gerlsma et al., 1997). The SRM for observations of independent dyads (networks with only two actors) is known as the Actor Partner Interdependence Model (APIM) developed by Cook and Kenny (2005); see also the textbook on dyadic data analysis by Kenny et al. (2006).

Application to Acquaintance Network

The SRM is estimated with MLwiN (Rasbash et al., 2005), using a macro written by Tom Snijders, available from his Web site (http://stat.gamma.rug.nl/). Note that the model can be estimated with any software that allows specification of random effects model with a complex variance structure (see David Kenny's Web site, http://davidakenny.net/srm/srm.htm, for more specific information about software for estimating the SRM). The SRM is applied to the EIES acquaintanceship data in four models presented in Table 31.4. The first one is a so-called null or empty model. This model serves as a baseline model, to obtain information on the overall mean strength of acquaintanceship, while taking into account individual differences between the researchers acting as both senders and receivers of ties through the sender and receiver variances. Moreover, an estimate of within-dyad reciprocity is obtained, in the form of the covariance (or correlation) between the residuals of the directed ties. An estimate of the tie variance is also available, which can be viewed as a measure of the within-actor variability of the acquaintanceship intensity. These parameters are present in all subsequent models. Next, an SRM is estimated with exactly the same effects as the first MR-QAP model, hierarchy, and distance, dyadic covariates based on the actor covariate status (the square root of the number of their citations). As a third model, separate sender and receiver effects of status are used instead of the dyadic hierarchy effect to demonstrate the properties of the SRM. The dyadic communication effect and homophily effect of research field are added as well. In the fourth and final model, the first measurement of the acquaintanceship network is added, to judge the strength of the previously present effects and for comparison with the MR-QAP results. The models are compared by deviance tests. Although the deviance serves to assess improvement of the model after adding (or deleting) covariates, it does not provide a direct measure of the variance explained by the model. Because these models are *nested*, the usual procedures of model comparison apply.

Table 31.4 SRM analyses results for the EIES acquaintanceship data at time 2 obtained with MLwiN (standard errors in parentheses)

	Model 0	Model 1	Model 2	Model 3
Intercept	1.680	2.109	1.059	0.661
	(0.17)	(0.19)	(0.22)	(0.012)
Communication			0.120	0.0655
			(0.014)	(0.00091)
Hierarchy		−0.0150		
		(0.0098)		
Distance		−0.142	−0.123	−0.0365
		(0.022)	(0.020)	(0.012)
Network time 1				0.690
				(0.021)
Same field			0.331	−0.0362
			(0.084)	(0.050)
Sender status			0.0655	0.00654
			(0.026)	(0.015)
Receiver status			0.0929	0.00830
			(0.026)	(0.014)
Sender variance	0.198	0.234	0.128	0.0366
	(0.057)	(0.065)	(0.038)	(0.012)
Receiver variance	0.264	0.266	0.124	0.0301
	(0.073)	(0.073)	(0.037)	(0.010)
Sender/Receiver	0.193	0.215	0.0842	0.0206
covariance	(0.059)	(0.063)	(0.031)	(0.0087)
Residual tie variance	0.869	0.809	0.717	0.328
	(0.047)	(0.043)	(0.037)	(0.015)
Reciprocity	0.531	0.471	0.370	0.0534
within-dyad covariance	(0.047)	(0.043)	(0.037)	(0.015)
Deviance	2563.7	2521.6	2437.2	1769.1

Model 0 in Table 31.4 shows that the overall average strength of acquaintanceship is 1.68, a little left from the middle of the 0–4 scale. The actor receiver variance is somewhat larger than the sender variance, corresponding to what is found in Table 31.1. The covariance of the random sender and receiver effects is equal to 0.19, implying a high correlation of 0.84 (0.19/√(0.20 . 0.26)) between them. This means that an actor with a positive sender effect (i.e., who has a higher than average mean strength of outgoing acquaintanceships) tends to also have a higher than average mean strength of received acquaintanceship ratings (a positive receiver effect). Note that this model describes the overall structure in the network, without using any of the covariate information. The residual tie variance is much larger than the actor variances. The overall reciprocity in the network is high, with a correlation of 0.61(0.53/0.87) between the two directed ties.

Like in the MR-QAP analysis, the effect of distance is stronger than that of hierarchy, as is evident from the small and nonsignificant effect of hierarchy in Model 1. The model is a clear improvement over the empty model with a difference in deviance of 42.1 (2563.7 −2521 .6), which is significant at the .001 level with two degrees of freedom (due to adding two parameters). The reduced total variance leads to a lower residual tie variance and slightly higher sender and receiver variances (a phenomenon well known in multilevel analysis, cf. Snijders and Bosker, 1999: 100). The correlation between sender and receiver effects and the unexplained reciprocity are approximately the same as in Model 0 (0.86 and 0.58, respectively).

In Model 2 both actor variances and residual tie variances are reduced after adding sender and receiver covariates of status, communication, and field homophily. The parameters pertaining to these effects are all significant, as is assessed by a *t*-test (or approximate z-test) dividing the parameter estimate by its standard error. A value larger than 2 is a rough indication of significance at the 5 percent level. After adding acquaintance at time 1 in Model 3, the covariate effects are greatly reduced, and all parameter estimates except communication and distance are no longer significant. The variances are reduced by more than half, with an unexplained reciprocity of only 0.16(0.05/0.33). The correlation between random sender and receiver effects remains strong at 0.62 (0.02/√(0.03.0.04)).

An interpretation of these changes is that much of the acquaintance network at time 2 can be explained by its state at time 1. Especially the effects of field similarity and (sender and receiver) status are captured by the network at time 1, which makes sense, and to a lesser extent the effect of distance. The amount of communication, however, adds to the explanation of the acquaintanceship at time 2. This seems very much in line with the purpose of the study and could be interpreted as a success of the experiment with computer communication.

The p1 and p2 model

In a separate development, statistical models for dichotomous relations (in a directed graph) were proposed that distinguish the four possible outcomes of a dyad: one null (0,0), one mutual (1,1), and two asymmetric (0,1) and (1,0) dyadic states. Holland and Leinhardt (1981) proposed the first so-called p_1 model. It can be considered as the loglinear pendant of the Kenny and La Voie (1984) ANOVA SRM, distinguishing the sender and receiver roles of all *n* actors. The model contains 2*n* actor parameters in addition to a density and reciprocity parameter representing the overall propensity to engage in any relationship (no matter the direction) and a mutual relationship, respectively.

Similar to the Snijders and Kenny (1999) SRM, the p_1 model was extended with random sender and receiver effects to the so-called p_2 model (Lazega and Van Duijn, 1997). The model provides the possibility to include (actor) sender and receiver effects as well as dyadic covariates for density and reciprocity effects. In the p_2 model, the four outcomes of a dyad for binary relations are modeled explicitly (comparable to polytomous logistic regression) with the null dyad (0,0) as the reference category. It has correlated random sender and receiver effects (like in the SRM). Its density parameter represents the log-odds of a directed tie (vs. no tie), regardless of the outcome of the other tie in the dyad. The reciprocity parameter represents the interaction effect of the increase in log-odds of a mutual dyad in comparison to the sum of the two log-odds of the asymmetric dyads.

Unlike the $p1$ model, which can be estimated easily using straightforward methods for loglinear or generalized linear models, the IGLS estimation first proposed by Van Duijn et al. (2004) for the $p2$ model was improved by using Markov Chain Monte Carlo methods (Zijlstra et al., 2009). They used a Bayesian model formulation in the tradition of Wong (1987), which was extended to categorical covariates by Wang and Wong (1987) and Gill and Swartz (2004). The $p2$ model can also be estimated for multiple (independent) networks (Zijlstra et al., 2006) and multiple relations (Zijlstra, 2008).

Application to Friendship Network

We turn our attention to the binary friendship network, to demonstrate the $p1$ and $p2$ models. The $p2$ model is estimated using the P2 module in StOCNET. The results are presented in Table 31.5. The $p1$ estimates were obtained with UCINET and only reported in the text. After estimating the $p1$ model, four $p2$ models are fitted to the EIES friendship data, starting with the null model. The next model is comparable to SRM Model 2 in Table 31.4, estimating the effect of communication, distance, and status. The third model shows the specific feature of the $p2$ model to include covariates for the reciprocity parameter. In the fourth model this model is extended with the (density) effect of friendship at the start of the experiment.

Table 31.5 p1 and p2 analyses results for the EIES friendship data at time 2 obtained with StOCNET (standard errors in parentheses)

	p_1	p_2			
	Model	Model 0	Model 1	Model 2	Model 3
Density	−3.34	−3.01	−4.67	−4.12	−5.22
		(0.27)	(0.50)	(0.53)	(0.76)
Communication			0.426	0.424	0.484
			(0.061)	(0.067)	(0.078)
Distance			−0.224	−0.365	−0.340
			(0.055)	(0.084)	(0.121)
Same field			0.717	0.736	0.677
			(0.20)	(0.20)	(0.34)
Network time 1					6.59
					(0.62)
Reciprocity	4.33	3.82	3.30	2.43	0.839
		(0.45)	(0.48)	(0.60)	(0.71)
Distance				0.364	0.519
				(0.16)	(0.21)
Sender status			0.171	0.155	0.108
			(0.12)	(0.13)	(0.14)
Receiver status			0.0962	0.0851	0.0251
			(0.077)	(0.087)	(0.10)
Sender variance	1.74[1]	1.06	3.40	3.49	2.97
		(0.40)	(1.20)	(1.29)	(1.34)
Receiver variance	2.23[2]	1.46	0.730	0.740	0.731
		(0.52)	(0.35)	(0.34)	(0.48)
Sender/ Receiver covariance	−1.26[3]	−0.829	−1.15	−1.18	−0.901
		(0.40)	(0.51)	(0.55)	(0.55)
Deviance (approx.)		663.7	531.2	526.8	258.5

[1] Based on the variance of the *n* estimated sender parameters.
[2] Based on the variance of the *n* estimated receiver parameters.
[3] Based on the covariance of the *n* pairs of sender and receiver parameters.

The analysis of the EIES data with the p_1 model results in a negative estimate of the density (–3.34). This implies that the probability of a tie is much smaller than 0.50 (corresponding to the observed overall density of 0.21). The reciprocity parameter is positive (4.33), slightly larger in absolute value than the density parameter. These values reflect that the occurrence of a mutual dyad is somewhat more likely than one of the asymmetric dyadic outcomes (corresponding to the observed number of 60 mutual dyads, which is more than half the number of asymmetric dyads, 82). The 32 individual estimates for sender and receiver effects are summarized by their variances and covariances 1.74, 2.23, and −1.26, respectively (the latter implying a correlation of −0.64).

The p_2 null model (Model 0 in Table 31.5) shows, as expected, a negative density parameter estimate and positive reciprocity parameter. The receiver variance is, again, higher than the sender variance and the covariance (and correlation) negative. Different than for the SRM, this correlation cannot be linked to the observed correlation between in- and outdegree, because of the inclusion of the reciprocity parameter in the model. Finally, a deviance value is reported, that due to the nonlinearity of the model is only approximate (cf. random effects logistic regression). No formal tests for model comparison can be based on the deviance (see Zijlstra et al., 2005).

Model 1 in Table 31.5 is analogous to SRM Model 2 in Table 31.4 and has the same substantive interpretation with positive effects of communication and field similarity and a negative effect of distance on friendship. The actor status effects, however, are too small to be significant. (Direct comparison of the regression parameters to the previous model is difficult, because they tend to increase in size with more variables in the model. This is a well-known phenomenon in logistic regression with random effects, see, for example, Snijders and Bosker [1999].) Note that the sender variance has increased as well. This is best interpreted in comparison to the receiver variance. Apparently the network effects included in the model explain more of the differences between the friendship ties received than those sent. A similar pattern, although not quite as strong, was also observed in the SRM analysis.

In Model 2, the typical p_2 feature of reciprocity covariates is demonstrated, by including distance in the model, not only for density but also for reciprocity. An interesting positive effect of distance for reciprocity is found, which can be interpreted in combination with the negative effect of distance for density. The negative effect of distance on density implies that the overall probability of a friendship tie is smaller for researchers with a larger status difference. The reciprocity distance parameter is an interaction effect, implying that the negative effect of distance on density is reduced in mutual dyads by the positive effect of density on reciprocity, thus making mutual and asymmetric ties approximately equally likely for dyads with the same distance between actors.

Finally, the friendship network at time 1 is added to the model.[5] Model 3 shows that this covariate has by far the largest influence, and like in the SRM analysis, the effects of field similarity and distance are reduced (although not for reciprocity), whereas the effect of communication remains relatively strong.

The Bilinear Model

Hoff's (2005) bilinear model builds on the earlier presented dyadic models, by incorporating dyadic and actor covariate effects. The model goes beyond dyadic dependence, because it includes additional parameters to capture specific forms of third-order dependence, defined as balance or clusterability. This can be understood as looking for further structure in the residuals of the model. The residuals are defined as a function of latent actor characteristics, forming a distance or space. Thus, a more complex dependence structure is modeled than in the SRM or $p2$ model, although not as complete or complex as in the Exponential Random Graph Models. The bilinear model is set up as a Bayesian Generalized Linear Model, accommodating normal, Poisson and binomial tie distributions, and is estimated using MCMC methods. The dimensionality of the latent space is to be decided on by the researcher, with the help of (Bayesian) fit statistics. If the dimensions of the latent space are chosen equal to zero, the bilinear model with a normal distribution for the ties amounts to the Social Relations Model. For dichotomous ties there is no equivalence with the $p2$ model because the bilinear model does not include a reciprocity parameter in the fixed (mean) part of the model. Bayesian model summary and selection tools are available to decide whether the model fits the data satisfactorily.

Application to Acquaintance Network

The bilinear model is not part of the statnet suite in R (Handcock et al., 2003), but for its estimation R source code is available at Peter Hoff's Web site (http://www.stat.washington.edu/hoff/Code/GBME/). The bilinear model presented here is used for illustration and is not intended to be presented as the best fitting model to the EIES data. A model is fitted with parameters equal to those of Models 2 of the SRM model and $p1$ model, that is, including the effect of communication, distance, and field similarity, as well as sender and receiver effects of status but without the acquaintanceship network at time 1 (as this explains so much of the acquaintanceship at time 2). Further, the latent space is chosen to be two-dimensional to facilitate graphical representation obtained after a Procrustes rotation.[6]

Comparing the model results of the bilinear model, presented in Table 31.6 to those of the Social Relations Model (Model 2 in Table 31.4 and also obtained with the bilinear model setting the number of latent space dimensions equal to 0), it becomes clear that the effects of the covariates are considerably reduced, notably the effects of field homophily, distance and sender status, and, to a lesser extent, of receiver status. The sender and receiver variances are practically the same, but the residual tie variance has gone down. The variances of the two latent dimensions are in size between the sender and receiver variances and the residual tie variance, and (compared to the SRM) picking up some of the residual variance.

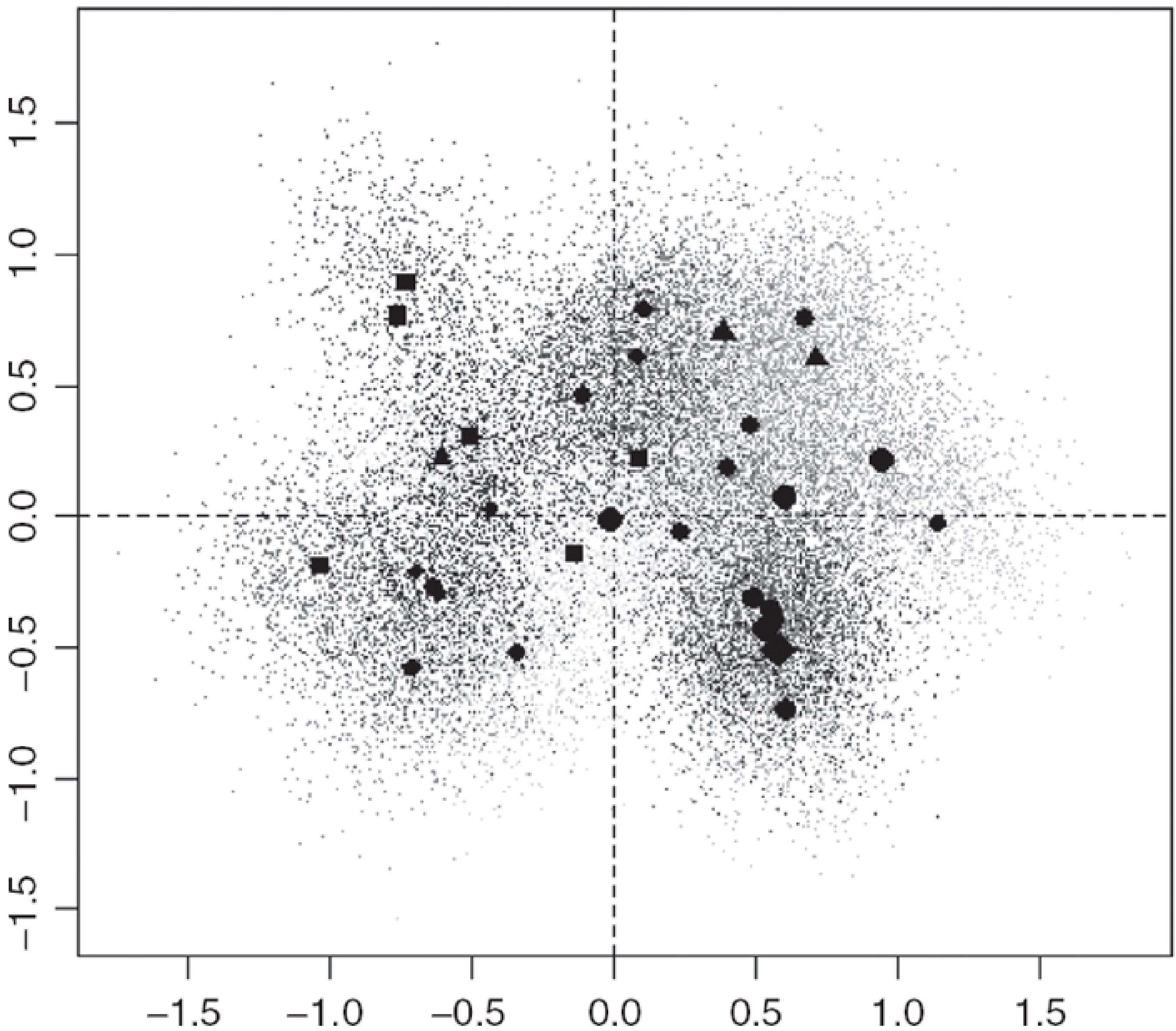
Table 31.6 Bilinear model analysis results for the EIES acquaintanceship data at time 2 obtained with StOCNET (standard errors in parentheses)

Density	1.14	(0.25)
Communication	0.107	(0.014)
Distance	–0.0496	(0.026)
Same field	0.175	(0.089)
Sender status	0.0363	(0.028)
Receiver status	0.0650	(0.030)
Sender variance	0.117	(0.039)
Receiver variance	0.136	(0.045)
Sender/Receiver covariance	0.0880	(0.036)
Residual tie variance	0.544	(0.029)
Within-dyad covariance	0.363	(0.041)

Latent dimension 1 variance	0.317	(0.10)
Latent dimension 1 variance	0.312	(0.094)

Our interpretation of the results is that, since the latent space serves to capture third order dependence, some of these structural effects were hidden in field similarity and distance. This seems plausible for effects like transitivity and balance on acquaintanceship, which may well work through colleagues in the same field or with a similar status. Interesting is that the (truly) dyadic effect of communication is only slightly reduced. Figure 31.3 depicts the position of the actors in the latent space. The symbols represent the locations of the actors, obtained as posterior mode from the MCMC runs. The tiny dots are the realizations obtained in all runs of the MCMC sampler, and they give an indication of the spread of nodal locations and of the overlap between nodes. Actors who are located close together have many ties to each other and/or a similar pattern of ties to others (Hoff, 2005). Some grouping seems to be present; the grouping of actors will be further investigated in the next section, after the summary of the results obtained with the tie models.

Figure 31.3 Bilinear model results for the EIES friendship network at time 2



Summary and Discussion of the Tie Models

Four different models with increasing complexity were discussed and illustrated in this section. With the MR-QAP model, we were able to identify a sizable influence of communication on the EIES acquaintanceship network at time 2, also (be it less strong) after controlling for the relations at time 1.

All other models are more geared toward social network data in the sense that they specifically model dyadic dependence. In addition to the dyadic effect of communication, the analyses with these models showed a symmetric dyadic effect of status (distance), also after controlling for the first measurement of the network (not computed for the bilinear model). If the distance between actors is large, both tie values reported are lower, on average. The individual status effects were stronger in the SRM and bilinear model (with only a significant receiver effect) than in the *p*2 model. Because the *p*2 model models “friendship” (defined as dichotomized acquaintanceship), this difference might be interpreted substantively as that acquaintanceship is more sensitive to individual status than friendship. Such an interpretation is tentative because an alternative explanation could be lack of power due to the dichotomous tie variable (as indicated by the larger standard errors for all parameters in the $p2$ model).

The interpretation of the random effects included in the models may seem difficult at first but is informative. The most important information in the SRM and bilinear model is captured by the reciprocity effect, reporting the association between the two ties in the dyad. The sender and receiver variances can be interpreted relative to each other, so for the EIES data it was found that researchers differ more from each other in reporting ties than in receiving ties. The positive covariance implies that researchers who report more than average tie values also tend to receive such values, which could be interpreted as a general sociability effect. Reciprocity is treated as a fixed effect in the $p2$ model and also showed the expected positive effect. The results of the bilinear model with a two-dimensional latent space showed that the effects obtained with the other models were probably somewhat inflated due to the presence of third-order effects.

Investigating Groups of Actors

In this section we take a different perspective on the social relational system. Instead of focusing on the relation(s), trying to explain the observed social network using dyadic and/or actor characteristics, we now turn to the actors. For the EIES data this amounts to how we can categorize the researchers best, based on the individual covariates available or on the acquaintance relations and the communication network. Before we proceed with presenting some stochastic a posteriori block models, we mention other methods and models with the aim to compare actors across known groups (for instance, according to research field) or to explain or predict an actor outcome variable using the network ties and possibly actor and dyadic covariates (for instance, scientific status using the acquaintanceship or friendship network and field).

Comparing Groups of Actors

The use of well-known methods like *t*-tests and ANOVA for the comparison of actors and their network characteristics suffers from the incorrect assumption of independence of observations (see, e.g., Hanneman and Riddle, 2005: Chapter 18). It is easy to summarize relational data to the actor level (although the choice for summary statistics may be overwhelming), but the resulting “actor” data are dependent by definition. As for correlations, a solution is to use permutation-based methods (cf. QAP), which are implemented in UCINET for *t*-tests, ANOVA and regression with actor outcomes. In the case of the EIES researchers it would be natural to compare the researchers by their disciplinary affiliation.

To investigate the influence of relations on actor characteristics, contagion models were proposed by, for example, Doreian (1980), Burt (1987), Friedkin (1998) and Leenders (2002). In these models the outcome variable is an actor characteristic, where the dependence between actors is represented by including (some form of) the social network in the regression equation. A contagion model is thus a kind of spatial regression model (cf. Ord, 1975; Anselin, 1988), modeling autoregression or (network) autocorrelation or both. Contagion models can be estimated using software for spatial regression or sna (Butts, 2008). See O'Malley and Marsden (2008) for a good overview and application of these models. We do not illustrate these models here for the reason that continuous time models are available in which contagion can be distinguished from selection effects (cf. Steglich et al., 2010).

Identifying Groups of Actors

The aim of a stochastic blockmodel analysis is the identification of groups (or positions) of actors. As in latent class analysis, one of the questions in such an analysis is to determine how many groups are needed to distinguish the set of actors sufficiently, with the immediate next question whether the group assignment is consistent with the researcher's expectation or can be understood given other quantitative (or qualitative) information. In some stochastic blockmodels additional information on the positions of the actors is obtained, to be used for a graphical representation. Note that the bilinear model presented in the previous section also had this feature. Early examples of stochastic blockmodels can be found in Anderson and Wasserman (1992), who group the sender and receiver parameters of the p_1 model.

In the remainder of this section we present four stochastic blockmodels, which will be indicated by the name of the software in which they are implemented. The first stochastic blockmodel, BLOCKS, was proposed by Nowicki and Snijders (2001) for (valued) social network data. This model follows an earlier model for (binary) network data to assign actors to (latent) groups (Snijders and Nowicki, 1997). Tallberg (2005) proposed an extension of the model by further modeling the group probabilities using covariate information (not illustrated here because no software is readily available). The next model presented is the stochastic block model proposed by Frank (1995), KliqueFinder. In addition to a group assignment, a graphical display of the positions of the actors is obtained. A somewhat different approach is the third model, proposed by Schweinberger and Snijders (2003) and based on the ultrametric distances between actors involving triads of actors and resulting in hierarchically clustered groups. As a final model the latent cluster models of Handcock et al. (2007), latentnet, are presented, in which dyadic and actor covariate information is incorporated. These models are related to the models presented by Hoff et al. (2002), Shortreed et al. (2006), and Krivitsky et al. (2009).

Blocks

The stochastic a posteriori model proposed by Nowicki and Snijders (2001) defines groups of stochastic equivalent actors, having the same probability of dyadic outcomes with actors in their own group and the same probability of dyadic outcomes with actors in other groups. It is based on the assumption of dyad independence in the observed social network, conditional on the latent (unknown) group membership of the actors (their so-called colors). Together with a color distribution defining a priori probabilities for each color (similar to latent class analysis), the joint distribution of dyadic outcome and color is obtained. Parameter estimates of the a posteriori probabilities for each color (per actor) are obtained using MCMC estimation, which is implemented in BLOCKS. Note that this model only uses the observed (binary) network and no covariate information.

Each actor is assigned to the color with the highest posterior probability for that actor. To determine the number of groups, two fit statistics pertaining to the extra information (I_y) and clarity (H_x) of the obtained actor positions are available. For both measures, values close to zero are preferred. The statistics are compared across models with a (predetermined) different number of groups to obtain the best solution.

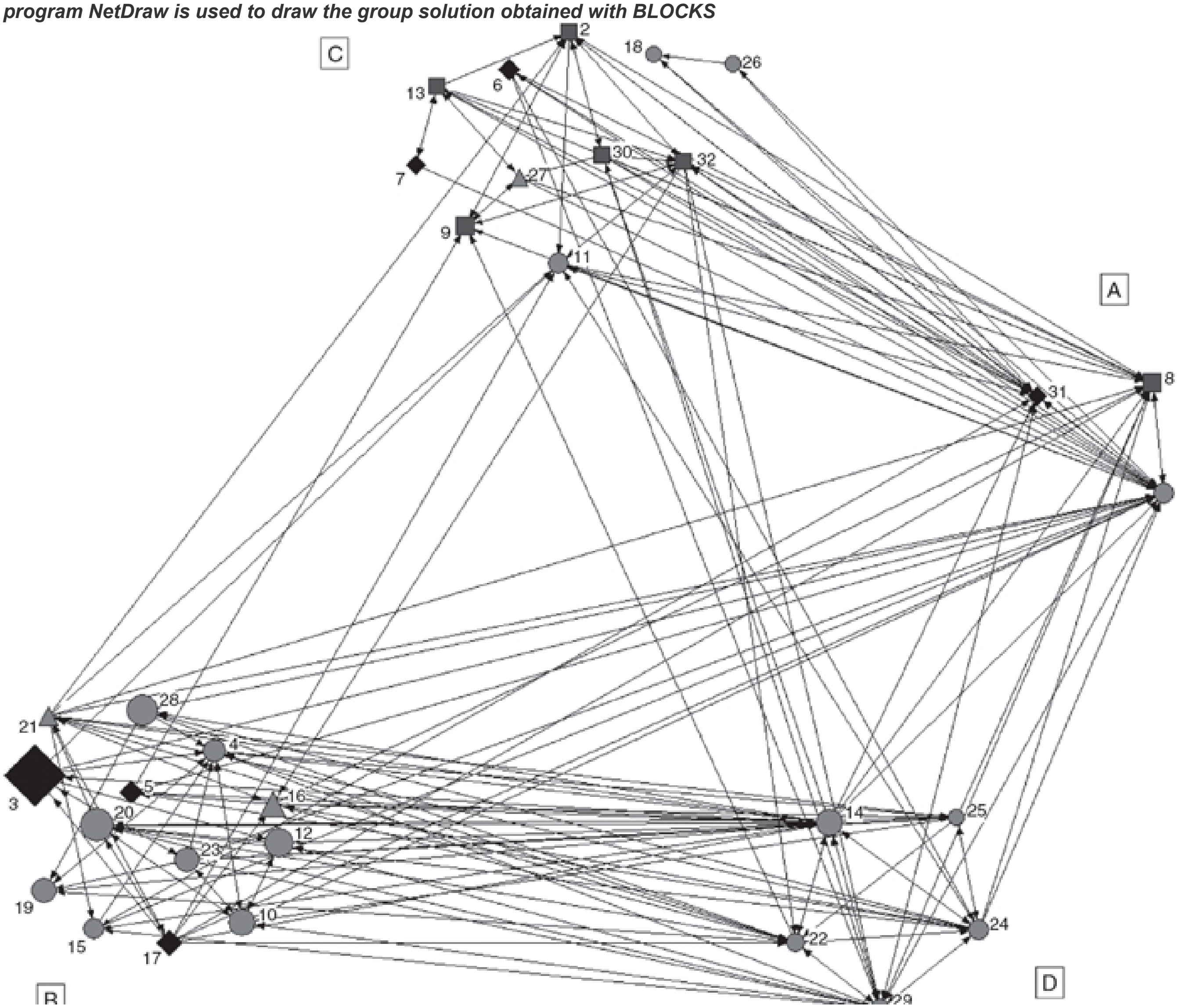
Application to Friendship Data

BLOCKS was run on the EIES friendship data in an exploratory approach to obtain solutions for models with two to six classes. It turned out that none of the group assignments could be classified as very informative or clear. We focused on the solutions for three to five groups (colors). BLOCKS has the option to discard stepwise the “worst” fitting actors in order to obtain a network with a reduced number of actors who are more easily assigned to groups such that the probability of within-group ties is higher than between-group ties. Using the stepwise approach, only four groups remained in the five-group solution. We therefore chose four as the desired number of groups. For reasons of comparability with the other stochastic block models we did not want to leave out actors of the network (although it can be an attractive feature to identify “ungroupable” actors), and therefore we used the option to reassign actors to groups. The algorithm gives the three or four best-fitting actors, that is, actors who are unambiguously assigned to different groups. The first three actors all had different research fields and we chose this as a starting configuration for the four-group solution. The obtained group assignment is depicted in Figure 31.4 and Table 31.7. For this solution, the values $H_X = 0.27$ and $I_y = 0.70$ were obtained, indicating that the solution is not particularly informative, which is at least partly due to our choice to assign all actors to a group.

Table 31.7 BLOCKS four-group solution for the EIES friendship data at time 2 (excerpt from output obtained with StOCNET; slightly adapted to distinguish groups (Grp: A–D), actors (Id), and research fields (Fld: Sociology, Anthropology, Statistics/ Mathematics, and Other)

```
                 Group And Actor Id
                |AAA|BBBBBBBBBBBBB|CCCCCCCCCCC|DDDDD|
                |   |             |           |     |
                |  3|   1111112222|    1112233|12222|
  Grp  Id  Fld|181|3450256790138|26791386702|42459|
--------------+---+-------------+-----------+-----+
   A    1  S   |A22|3.33.3.3..333|23322222332|3.3.2|
   A    8  A   |2A.|......33..3..|33..23..332|3.333|
   A   31  O   |2.A|.....2.......|22..2223.32|3...2|
--------------+---+-------------+-----------+-----+
   B    3  O   |4..|B4..4..3...2.|....4......|2....|
   B    4  S   |...|3B.2.3..2..33|...........|33.33|
   B    5  O   |4..|..B...3......|...4.......|3.3..|
   B   10  S   |4..|.2.B2....2.2.|...........|...32|
   B   12  S   |...|3..2B....4...|...........|3...3|
   B   15  S   |4.2|.4...B....3..|...........|2...2|
   B   16  SM  |.4.|..4...B3..2..|..........3|2233.|
   B   17  O   |44.|4.....4B.42..|....4......|24...|
   B   19  S   |...|.2......B..33|...........|32...|
   B   20  S   |...|...23..3.B4..|...........|22222|
   B   21  SM  |44.|.....422.3B..|4...4......|3232.|
   B   23  S   |4..|24.2....4..B.|...........|2..34|
   B   28  S   |4..|.4......4...B|...........|343..|
--------------+---+-------------+-----------+-----+
   C    2  A   |242|..........3..|C..243...22|.....|
   C    6  O   |442|.............|.C.........|..4.4|
   C    7  O   |4..|.............|..C..2.....|.....|
   C    9  A   |2..|..3..........|2..C....2.2|.3...|
   C   11  S   |222|3......3..3..|3...C.....2|..3..|
   C   13  A   |242|.............|4.2..C..2.4|.....|
```

Figure 31.4 EIES friendship network at time 2. The actors are assigned to four different groups (distinguished by location) and are distinguished by research eld (node shape and color; same as in [Figure 31.2](#)) and by number of citations (node size). The program NetDraw is used to draw the group solution obtained with BLOCKS



[Table 31.7](#) and [Figure 31.4](#) show that the groups indicate no association with the research field as there are researchers from different disciplines (depicted by the shapes of their nodes in the graph) in all groups except for group D, positioned in the lower righthand corner. There are many ties among the groups, especially between the groups B and D on the lower side of the picture. The picture, however, seems to show a hierarchy in citation status among the groups: the largest group, B (bottom left), contains the researchers with high citation status, followed by the smaller group, D positioned bottom right, consisting of five sociologists. The second largest group, C (top left), and the small group, A (top right), consist of the researchers with different fields with the lowest status, where the researchers in group C have a lower status than those in group A. Some more precise information can be derived from [Table 31.7](#). It represents the adjacency matrix, as available in the output of BLOCKS, slightly adapted by replacing the original 1s (indicating a null relation) by dots and adding the information on the actors' fields. The 2s indicate mutual relations, the 3s and 4s asymmetric relations, from column to row and vice versa, respectively. Group A (top right in [Figure 31.4](#)) has mutual relations within the group and mainly sends relations to the other groups (in view of the many 3s in the rows of group A), with also a good deal of mutual relations with group C (top left in Figure 31.4), the group with a slightly lower status. Group B (bottom left in Figure 31.4) is the largest, with the highest status, and exhibits all forms of dyadic relations within the group (with many null dyads), and only sends some ties to the first group and even fewer to the third group but receives a lot of ties (some reciprocal) to the fourth group. Group C is large and like group B has relatively few ties within the group. It has hardly any ties with groups B and D and more ties (some reciprocal) with group A. Finally, group D is quite dense and its actors send ties to especially groups A and B (some reciprocal) but have hardly any ties with researchers in group C. The many ties with group B are well visible in Figure 31.4. In terms of hierarchy defined as receiving many choices by other groups, the order of the groups is A, C, B, D.

KliqueFinder

The procedure implemented in KliqueFinder (Frank, 1995, 1996, 2009) is based on a $p1$-like model[7] extended with a categorical actor covariate (group), in the form of a group similarity (homophily) effect on the density. The underlying idea is that of cohesion: Actors in the same group should have a higher probability to interact with each other than with actors from other groups. Just like in BLOCKS no other information is necessary than one observed social network. Unlike BLOCKS, KliqueFinder makes no distinction between the direction of ties.

The model was developed before MCMC estimation became more common and is estimated with an algorithm called iterative partitioning. The actors are preassigned to a group, starting with a clique of three actors, and this assignment is changed iteratively until no further improvement of the objective function is possible. The objective function is usually defined to maximize the probabilities of in-group ties, but some other, related, definitions are also available. Preassignment of actors to different groups is optional. A graphical representation of the groups and the actors within them is obtained using MDS on distances between actors and groups defined by the amount of interaction between them.

Application to Friendship Data

The EIES friendship data were analyzed with KliqueFinder. This resulted in five groups, depicted in Figure 31.5. It is clear that the five groups do not correspond to the four research fields distinguished in the EIES data. A possible interpretation is a hierarchy of actors according to status but with a different result than the BLOCKS solution with four groups. The upper group (B in Table 31.8) consists of researchers with a relatively low number of citations and is far away from the middle two groups (A and E) containing most actors. Of these middle groups, only three or four (in the lower-middle group E) have a higher status. The two overlapping lower groups (C and D) consist of fewer actors. These actors have the highest status and are all sociologists except for the most-cited researcher from an “other” discipline but are further apart within the groups. Figure 31.6, obtained with UCINET, confirms this view of the KliqueFinder solution and gives some more information on the middle groups (bottom left and top right), which have a high density and also many relations with each other. Finally, the adjacency matrix representation in Table 31.8 (slightly adapted from the KliqueFinder output to indicate the original research fields) shows the high density within the five groups, where the 1s indicate either asymmetric or mutual relations and the dots no relation.

Figure 31.5 EIES friendship network at time 2. The actors are assigned to five different groups (distinguished by location) and are distinguished by research eld (node shape and color; same as in Figure 31.2) and by number of citations (node size). The

program NetDraw is used to draw the graph obtained with KliqueFinder

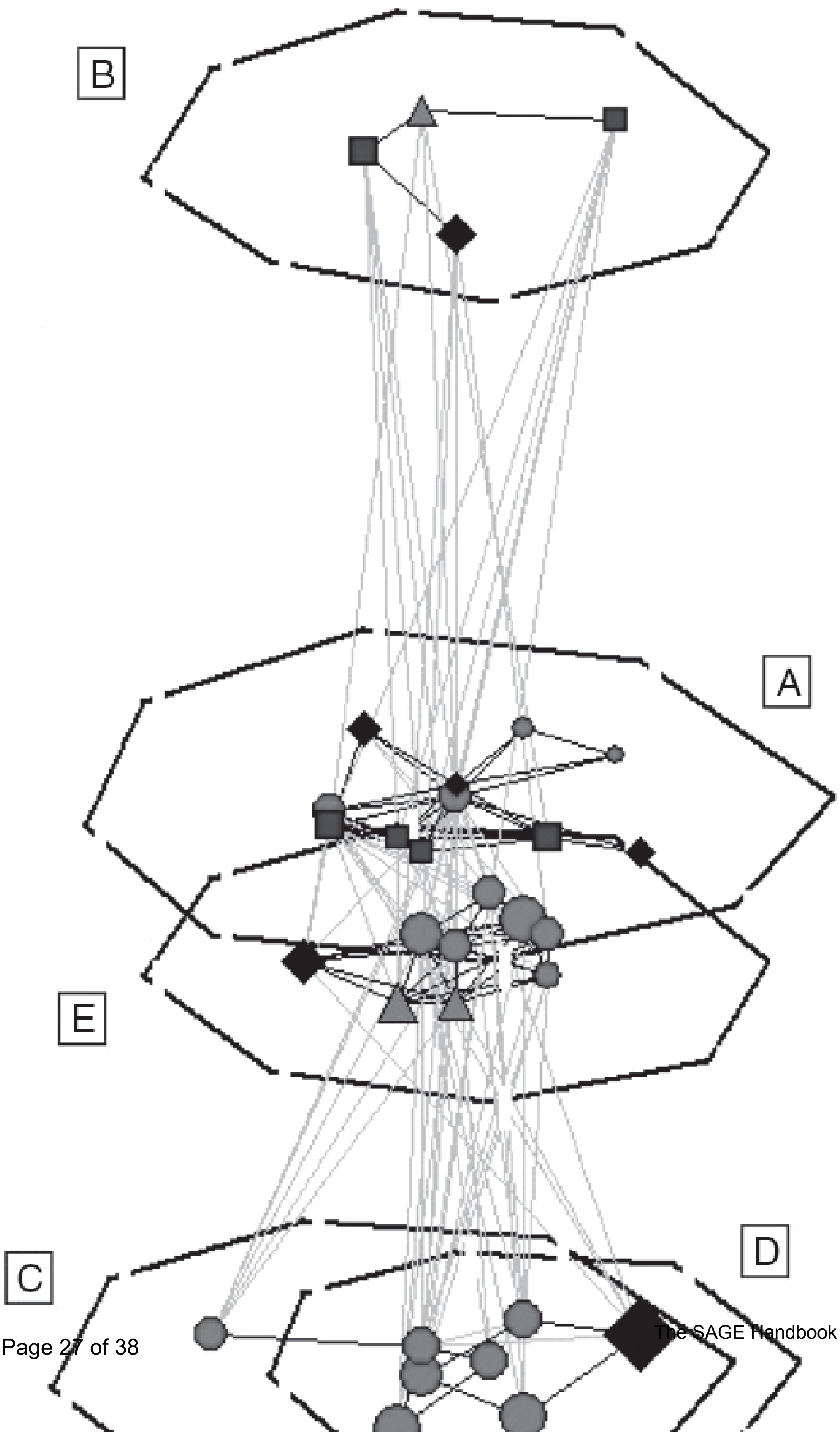

B
A
E
C
D

Figure 31.6 EIES friendship network at time 2. The actors are assigned to five different groups (distinguished by location) and are distinguished by research eld (node shape and color; same as in [Figure 31.2](#)) and by number of citations (node size). The program NetDraw is used to draw the group solution obtained with KliqueFinder

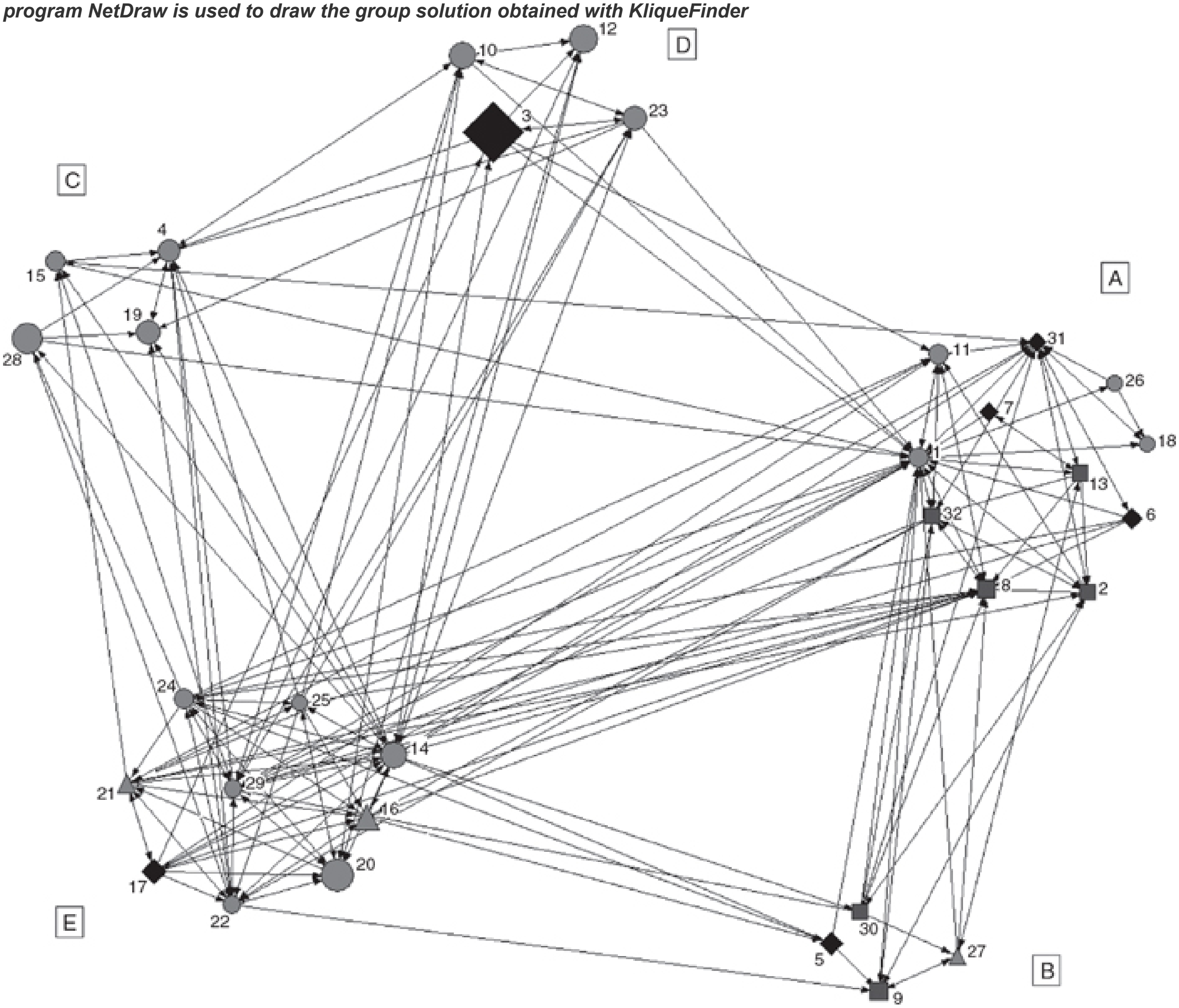


Table 31.8 KliqueFinder five-group (Grp: A–E) solution for the EIES friendship data at time 2. Actors (Id) and Research fields (Fld: Sociology, Anthropology, Statistics/Mathematics, and Other) are distinguished

```
 N             Group And Actor Id
 32           |AAAAAAAAAAA|BBBB|CCCC|DDDD|EEEEEEEEE|
              |           |    |    |    |         |
              |  1313  2 1| 2 3| 112|112 |122221221|
  Grp  Fld Id|81118227663|9750|4958|2033|612094457|
------------+-----------+----+----+----+---------+
  1A   A    8|A11..1.....|....|....|....|.........|
  1A   S    1|1A11111.1.1|1...|....|....|....1....|
  1A   S   11|11A1.1.....|....|....|....|.........|
  1A   O   31|.11A111..11|....|..1.|....|....1....|
  1A   S   18|.1.1A......|....|....|....|.........|
  1A   A   32|1111.A1....|1...|....|....|1.1.1....|
  1A   A    2|1111.1A....|1..1|....|....|.........|
  1A   O    7|.1.....A..1|....|....|....|.........|
  1A   S   26|.1.11...A..|....|....|....|.........|
  1A   O    6|11.1.....A.|....|....|....|....1.1..|
  1A   A   13|11.1.111..A|.1..|....|....|.........|
------------+-----------+----+----+----+---------+
  2B   A    9|.1...11....|B1..|....|....|.........|
  2B   SM  27|11........1|1B..|....|....|.........|
  2B   O    5|.1.........|1.B.|....|....|.........|
  2B   A   30|11.1.11....|.1.B|....|....|....1....|
------------+-----------+----+----+----+---------+
  3C   S    4|...........|....|C1..|.1..|.........|
  3C   S   19|...........|....|1C..|....|..1......|
  3C   S   15|.1.1.......|....|1.C.|....|....11...|
  3C   S   28|.1.........|....|11.C|....|..1......|
------------+-----------+----+----+----+---------+
  4D   S   12|...........|....|....|D1..|...1.....|
```

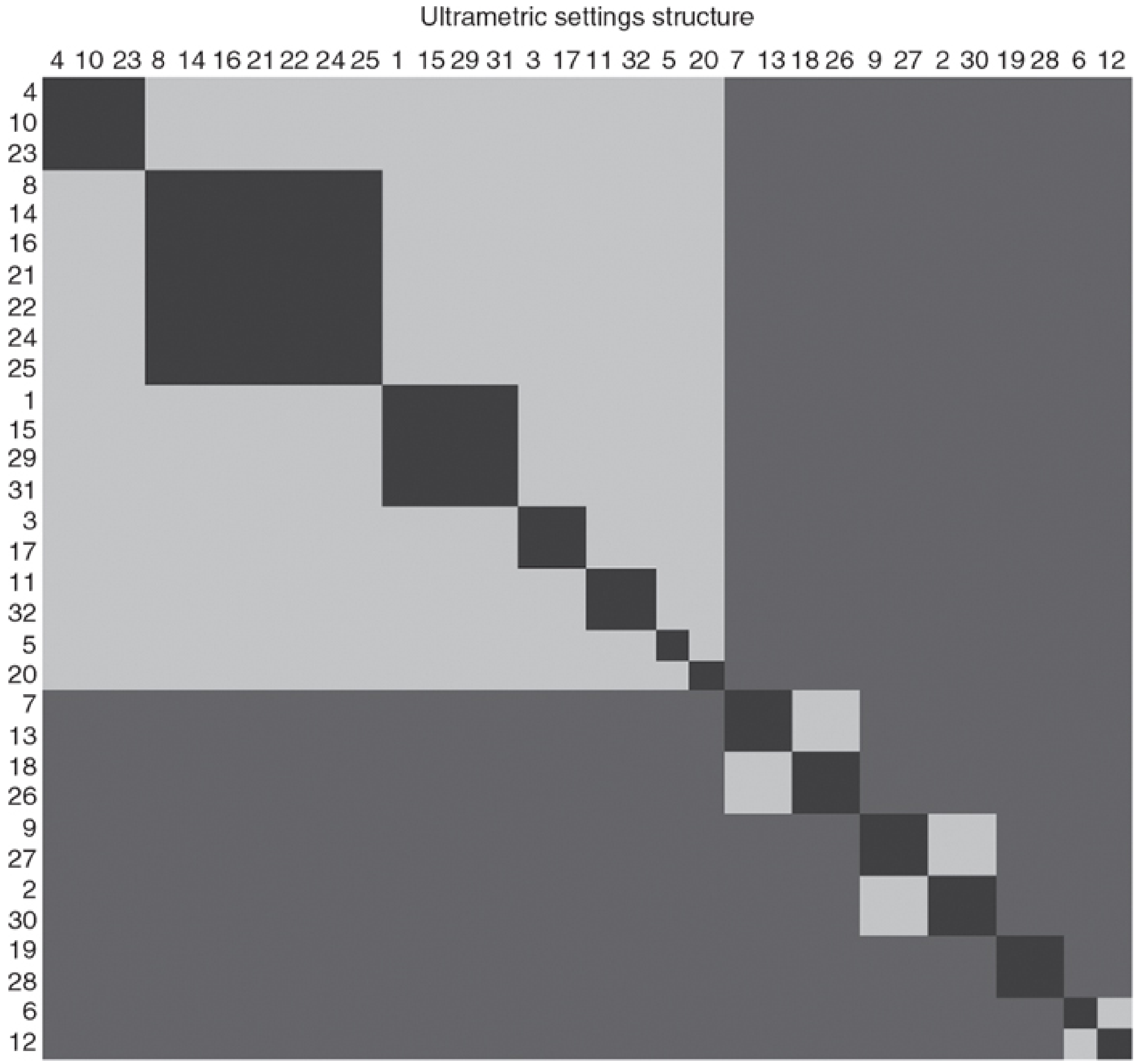
ULTRAS

The model proposed by Schweinberger and Snijders (2003) to find groups (or settings as they call it) is based on the assumption that there is an unknown distance between the actors in the network that governs the probability of ties between them, leading to a nested grouping of the actors. The distance (or proximity) is measured by an ultrametric (latent) space defined for pairs of actors, relative to their distance to third actors, and implies a transitive structure. The larger the distance between actors, the lower their probability of a tie. The nested grouping leads to regions with higher density like a geographic map with contour lines (see Schweinberger and Snijders, 2003: Figure 2). The ULTRAS model requires a symmetric matrix and can deal with dichotomous, count, or continuous tie variables, using, respectively, a Bernoulli, Poisson, or normal distribution for the tie probabilities, intensities, or strengths. It is estimated using Maximum Likelihood or Bayesian methods.

Application to the Symmetrized Friendship Data

The "ultrametric" model is implemented in the STOCNET module ULTRAS. The results for the EIES friendship data are based on a standard setup with three runs of 10,000 iterations of the MCMC sampler. The first step is to decide how many levels are needed in the grouping of the actors. Running the module for 2 to 9 levels, it was found that four levels sufficed in the sense that solutions with five or more levels had all very small probabilities of ties between actors at a distance of five or more. We did not run extensive checks for convergence or model fit. The solution is presented in Figure 31.7, not as a map but in the form of an adjacency matrix, consisting of several "blocks" or "settings," where the last simply comprises the whole network, with a very low overall probability (approximately 0.001) of a tie between actors not belonging to a further nesting (here, none of the actors). The lower right block consisting of actors 7 through 12 (as ordered by ULTRAS, see Figure 31.7) has a probability of 0.05 of a tie, whereas within the upper left block (actors 4 through 20) this probability has increased to 0.20. This division roughly corresponds to the two largest groups in the BLOCKS solution. Within these two settings a large number of groups are nested, many of which contain only pairs of actors. The highest density is found in the darker blocks, with a tie probability of 0.70. The largest block of actors this close together is the second block in the adjacency matrix (consisting of actors 8 through 25), which shows some resemblance with the fifth group of the KliqueFinder solution (see Table 31.8).

Figure 31.7 Block structure obtained with ULTRAS for the symmetrized EIES friendship network at time 2



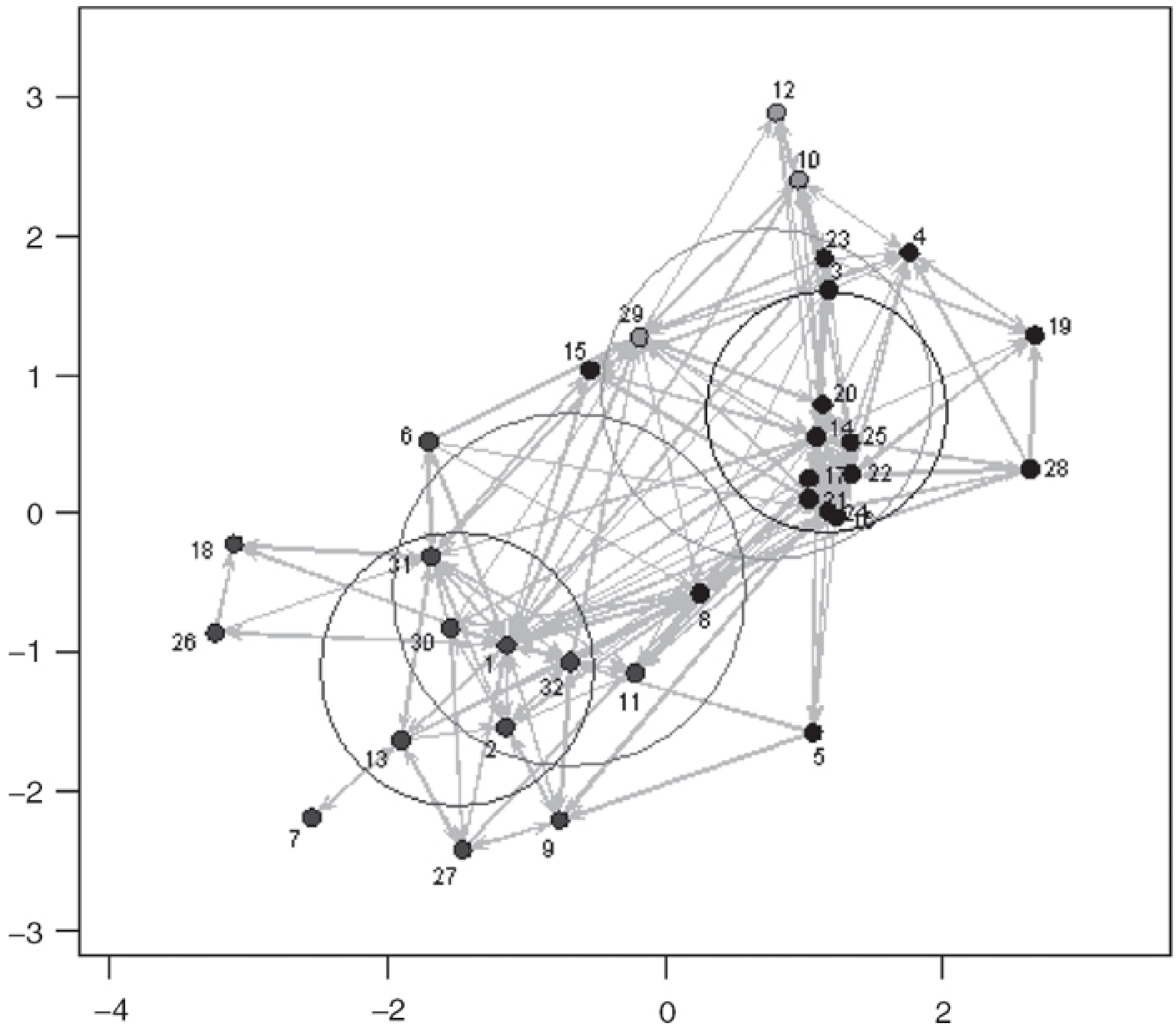
Latentnet

The latent position cluster model proposed by Handcock et al. (2007) is based on a different concept of latent space, which was previously discussed for Hoff's (2005) bilinear model. The latent space model proposed by Hoff et al. (2002) does not have the feature of assigning group positions to actors. It was extended to the latent position cluster model by assuming unknown group (latent cluster) membership of the actors on the dimension(s) in the latent space. Actor and dyadic covariates can be included in the model as well, which makes the model more elaborate than the previously presented stochastic a posteriori blockmodels. The model is estimated with advanced Bayesian (MCMC) techniques and has several options for the definition of distance on which graphical representation is based (see Shortreed et al., 2006). Various statistics are available to make the sometimes difficult choice on the optimal number of latent clusters (as in the BLOCKS model), and the number of dimensions of the latent space. The model was extended with random sender and receiver effects (comparable to the bilinear model) by Krivitsky et al. (2009).

Application to Friendship Data

The EIES friendship data are analyzed using the R package latentnet (see Krivitsky and Handcock, 2008) with the same covariates as in the bilinear model. Again, we did not search for a best-fitting model but compared the four and five latent cluster solutions, including the dyadic covariates for communication, distance, and field homophily. The four latent cluster solution was to be preferred according to the Bayesian Information Criterion. The regression parameters (not shown in a table) revealed that the effect of communication was strongest, but interestingly, there was also a significant effect of field similarity, whereas the effect of distance was small and insignificant. (Recall that in the bilinear model with random sender and receiver effects both field similarity and distance did not reach significance.) The graphical representation of the solution in Figure 31.8 shows large overlap in the four clusters and a lot of variation around the cores of the clusters. A general positive association between the positions on the two dimensions is visible. The resemblance with the BLOCKS solution is larger than with the KliqueFinder solution.

Figure 31.8 Latent position and latent cluster graph obtained with latentnet for the EIES friendship network at time 2



Summary

Four different stochastic a posteriori block models were presented. Although they have the same goal, finding a (good) classification of the actors in a network, their properties are very different. The first three models do not use covariate information and aim to find groups consisting of stochastic equivalent actors, which in the case of KliqueFinder is defined simply as having an as high as possible probability of within-group ties (and therefore lower between-group ties). This is also one of the assumptions in ULTRAS, whose definition of distance leads to a classification of actors in nested groups. In BLOCKS the definition of stochastic equivalence is more general and extended to all possible dyadic outcomes (four in the case of binary data). The concept of stochastic equivalence is not as clearly defined in the latent position cluster model, because group membership is defined on the latent space dimensions (possibly while taking into account covariate effects on the probability of a tie). The interpretation of these dimensions is not always easy or immediately clear. ULTRAS is conceptually related to latentnet but employs a different definition of latent space and distance.

Given the different definitions of stochastic equivalence it is not surprising that rather different solutions are found. KliqueFinder gave a five-group solution with high within-group density (see Figures 31.5 and 31.6) which seemed to be somewhat related to the scientific status of the actors and not so much to research field, although scientific status is related to field as the sociologists have the highest number of citations (cf. Table 31.1). For the four-group solution found by BLOCKS, a weak relation with scientific status was suggested as well. The nested group structure obtained with ULTRAS could be roughly related to both earlier solutions but contained many more smaller groups. The analysis with latentnet indicated that a solution with four groups was to be preferred over a five-group solution, although from the overlap between clusters a solution with fewer clusters might have been better. The assignment of actors to groups (not precisely shown here) resembled the group assignment by BLOCKS and is therefore also weakly related to scientific status of the actors. Interestingly, no significant effect of scientific status (through dyadic distance) was found. A possible interpretation is that distance does not directly influence the (dyadic) probability of a tie but is present in the latent dimensions that capture third-order dependence.

Concluding Remarks

Although we did not find an answer to all of the questions posed in the introduction of this chapter, the models and methods presented gave a lot of insight into the structure of the EIES data and the association between the networks and covariates. Some tentative answers are that communication is an important explanatory variable for the observed acquaintanceship at time 2 for the EIES researchers. This could be regarded as a proof of the success of the experiment that led to the data. Status effects are less clear, and the influence of working in the same area on the relationship seems to depend to some extent on the model applied. Whether and how the actors could be assigned to different groups has not become very clear, as the methods presented gave different solutions. If anything, the grouping seems related to scientific status, not to research area.

The difference in (ease of) interpretation underlines the differences between the two classes of stochastic models we discussed. Modeling the ties using dyadic and actor covariate information is relatively straightforward because the focus is on the explanatory power of the covariates and less on unexplained differences between actors. In the models for finding groups of actors, there seems to be a need to interpret the groups a posteriori using additional (covariate) information as well.

The latent space models with covariates form in this sense an interesting hybrid class of models with the possibility of using covariate information and finding groups, although this does not solve the problem of interpreting the latent space dimension(s), capturing a form of triadic dependence. The random effects in the models for ties can be regarded as latent variables as well, forming a two-dimensional latent space of (unexplained) sender and receiver effects related to the actors. Such an interpretation is easier and more straightforward than the interpretation of the latent space defined by the bilinear model and latent space models. Even if no further interpretation can be obtained, it is expected that modeling network dependence explicitly improves the estimates of the covariate parameters.

All models and methods presented can be estimated with (more or less readily) available software. The development of the R packages for social network analysis have made it almost easy to apply the complex latent position cluster models. The interpretation of social network data, however, remains not very easy. It requires the expertise of applied researchers, who, just like the EIES researchers, may want to have colleagues with statistical expertise in their network.

Acknowledgments

The authors would like to thank Christian Steglich, Tom Snijders, and Peter Carrington for helpful comments, and Ken Frank for advice on the use of KliqueFinder.

Notes

1 The principles of blockmodeling are laid out by Doreian et al. (this volume); see also De Nooy et al. (2005) and Wasserman and Faust (1994, Chapter 10).

2 Although the availability of software is discussed to some extent, a more complete overview of software for network analysis can be found in Huisman and van Duijn's chapter (this volume). Full-color versions of the figures are available on http://www.gmw.rug.nl/~huisman/sna/figures.html.

3 This is just a simple and straightforward way to express status differences among the researchers and is here mainly used for illustration. We have not investigated other definitions and cannot claim that this would be the best or most meaningful way to operationalize status. The square-root transformation is preferred over the log-transformation, another commonly used transformation of right-skewed data, because it preserves the 0 and 1, both frequently occurring as citation numbers.

4 A straightforward multilevel model is obtained for egocentric data, where (ties to) alters are nested in egos, under the assumption of nonoverlapping ego networks, leading to a simple distinction of within-ego and between-ego variance (see also Van Duijn et al., 1999).

5 Note that it would have been possible to use the acquaintanceship network instead of the binary friendship network, because the network is used as a covariate.

6 The results are based on a sample of 1,000 draws from a sample of 50,000 from the posterior distribution of the parameters, taking every 50th draw, after a burn-in of 50,000.

7 Note that the p_1 model and its early extensions with categorical actor variables are also considered stochastic blockmodels (with groups defined a priori) because in these models the actors belonging to the same group are stochastically equivalent.

References

Anderson, C.J. and Wasserman, S. (1992) 'Building stochastic blockmodels', Social Networks, 14(1–2): 137–61. http://dx.doi.org/10.1016/0378-8733%2892%2990017-2
Anselin, L. (1988) Spatial Econometrics: Methods and Models. Dordrecht, Netherlands: Kluwer Academic. http://dx.doi.org/10.1007/978-94-015-7799-1
Boer, P., Huisman, M., Snijders, T.A.B., Steglich, C.E.G., Wichers, L.H.Y. and Zeggelink, E.P.H. (2006) StOCNET: An Open Software System for the Advanced Statistical Analysis of Social Networks. Version 1.8. Groningen: ICS, University of Groningen/SciencePlus.
Borgatti, S.P. (2002) NetDraw: Graph Visualization Software. Harvard: Analytic Technologies.
Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) UCINET 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
Burt, R.S. (1987) 'Social contagion and innovation: Cohesion versus structural equivalence', American Journal of Sociology, 92(6): 1287–335. http://dx.doi.org/10.1086/228667
Butts, C.T. (2008) 'Social network analysis with sna', Journal of Statistical Software, 24(6): 30.
Cook, W.L. and Kenny, D.A. (2005) 'The actor-partner interdependence model: A model of bidirectional

effects in developmental studies', International Journal of Behavioral Development, 29(2): 101–9. http://dx.doi.org/10.1080/01650250444000405
Cyram (2009) Cyram NetMiner 3. Seoul: Cyram Co., Ltd.
Dekker, D., Krackhardt, D. and Snijders, T.A.B. (2007) 'Sensitivity of MRQAP tests to collinearity and autocorrelation conditions', Psychometrika, 72(4): 563–81. http://dx.doi.org/10.1007/s11336-007-9016-1
De Nooy, W., Mrvar, A. and Batagelj, V. (2005) Exploratory Social Network Analysis with Pajek. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511806452
Doreian, P. (1980) 'Linear models with spatially distributed data.'Sociological Methods and Research, 9(1): 29–60. http://dx.doi.org/10.1177/004912418000900102
Frank, K.A. (1995) 'Identifying cohesive subgroups', Social Networks, 17(1): 27–56. http://dx.doi.org/10.1016/0378-8733%2894%2900247-8
Frank, K.A. (1996) 'Mapping interactions within and between cohesive subgroups', Social Networks, 18(2): 93–119. http://dx.doi.org/10.1016/0378-8733%2895%2900257-X
Frank, K. (2009) KliqueFinder for Windows. Version 0.11. East Lansing: Michigan State University.
Freeman, S.C. and FreemanL.C. (1979) 'The networks network: A study of the impact of a new communications medium on sociometric structure', Social Science Research Reports 46. Irvine: University of California.
Friedkin, N.E. (1998) A Structural Theory of Social Influence. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527524
Gerlsma, C., Snijders, T.A.B., van Duijn, M.A.J. and Emmelkamp, P.M.G. (1997) 'Parenting and psychopathology: Differences in family members perceptions of parental rearing styles', Personality and Individual Differences, 23(2): 271–82. http://dx.doi.org/10.1016/S0191-8869%2897%2900033-0
Gill, P.S. and Swartz, T.B. (2004) 'Bayesian analysis of directed graphs data with applications to social networks', Journal of the Royal Statistical Society. Series C (Applied Statistics), 53(2): 249–60. http://dx.doi.org/10.1046/j.1467-9876.2003.05215.x
Goldenberg, A., Zheng, A.X., Fienberg, S.E. and Airoldi, E.M. (2009) 'A survey of statistical network models', Foundations and Trends in Machine Learning, 2(2): 129–233. http://dx.doi.org/10.1561/2200000005
Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P.N. and Morris, M. (2003) Statnet: Software Tools for the Statistical Modeling of Network Data. Version 2.1-1.31.
Handcock, M.A., Raftery, A.E. and Tantrum, J.M. (2007) 'Model-based clustering for social networks. With discussion', Journal of the Royal Statistical Society A, 170(2): 301–54. http://dx.doi.org/10.1111/j.1467-985X.2007.00471.x
Hanneman, R.A. and Riddle, M. (2005) Introduction to Social Network Methods. Riverside: University of California. http://faculty.ucr.edu/~hanneman/.
Hoff, P.D. (2005) 'Bilinear mixed-effects models for dyadic data', Journal of the American Statistical Association, 100(469): 286–95. http://dx.doi.org/10.1198/016214504000001015
Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002) 'Latent space approaches to social network analysis', Journal of the American Statistical Association, 97(460): 1090–98. http://dx.doi.org/10.1198/016214502388618906
Holland, P.W. and Leinhardt, S. (1981) 'An exponential family of probability distributions for directed graphs', Journal of the American Statistical Association, 76(373): 33–50. http://dx.doi.org/10.1080/01621459.1981.10477598
Hubert, L.J. (1987) Assignment Methods in Combinatorial Data Analysis. New York: Marcel Dekker.
Kenny, D.A., Kashy, D.A. and Cook, W.L. (2006) Dyadic Data Analysis. New York: Guilford.
Kenny, D.A. and La Voie, L. (1984) 'The social relations model', in L.Berkowitz (ed.), Advances in Experimental Social Psychology, vol. 18. New York: Academic Press. pp. 141–82.
Krackhardt, D. (1987) 'QAP partialling as a test of spuriousness', Social Networks, 9(2): 171–86. http://dx.doi.org/10.1016/0378-8733%2887%2990012-8
Krivitsky, P.N. and Handcock, M.S. (2008) 'Fitting latent cluster models for networks with latentnet', Journal of Statistical Software, 24(5): 1–23.
Krivitsky, P.N., Handcock, M.S., Raftery, A.E. and Hoff, P.D. (2009) 'Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models', Social Networks31(3): 204–13. http://dx.doi.org/10.1016/j.socnet.2009.04.001
Lazega, E. and van Duijn, M.A.J. (1997) 'Formal structure and exchanges of advice in a law firm: A

logistic regression model for dyadic network data', Social Networks, 19(4): 375–97. http://dx.doi.org/10.1016/S0378-8733%2897%2900006-3
Leenders, R. Th. A. J. (2002) 'Modeling social influence through network autocorrelation: Constructing the weight matrix', Social Networks, 24(1): 21–47. http://dx.doi.org/10.1016/S0378-8733%2801%2900049-1
Nowicki, K. and Snijders, T.A.B. (2001) 'Estimation and prediction for stochastic blockstructures', Journal of the American Statistical Association, 96(455): 1077–87. http://dx.doi.org/10.1198/016214501753208735
O'Malley, A.J. and Marsden, P.V. (2008) 'The analysis of social networks', Health Services and Outcomes Research Methodology, 8(4): 222–69. http://dx.doi.org/10.1007/s10742-008-0041-z
Ord, K. (1975) 'Estimation methods for models of spatial interaction', Journal of the American Statistical Association, 70(349): 120–26. http://dx.doi.org/10.1080/01621459.1975.10480272
Rasbash, J., Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (2005) MLwiN Version 2.02. Bristol: Centre for Multilevel Modelling, University of Bristol.
Robins, G.L., Snijders, T.A.B., Wang, P., Handcock, M.S. and Pattison, P. (2007) 'Recent developments in exponential random graph (p_ p*) models for social networks', Social Networks, 29(2): 192–215. http://dx.doi.org/10.1016/j.socnet.2006.08.003
Schweinberger, M. and Snijders, T.A.B. (2003) 'Settings in social networks: A measurement model', Sociological Methodology, 33(1): 307–41. http://dx.doi.org/10.1126/science.1070120
Scott, J. (2000) Social Network Analysis: A Handbook. London: Sage.
Shortreed, S., Handcock, M.S. and Hoff, P.D. (2006) 'Positional estimation within a latent space model for networks', Methodology, 2(1): 24–33.
Snijders, T.A.B. and Bosker, R.J. (1999) Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London: Sage.
Snijders, T.A.B. and Kenny, D.A. (1999) 'The social relations model for family data: A multilevel approach', Personal Relationships, 6(4): 471–86. http://dx.doi.org/10.1111/j.1475-6811.1999.tb00204.x
Snijders, T.A.B. and Nowicki, K. (1997) 'Estimation and prediction for stochastic blockmodels for graphs with latent block structure', Journal of Classification, 14(1): 75–100. http://dx.doi.org/10.1007/s003579900004
Snijders, T.A.B., Pattison, P.E., Robins, G.L. and Handcock, M.S. (2006) 'New specifications for exponential random graph models', Sociological Methodology, 36(1): 99–153. http://dx.doi.org/10.1086/210318
Steglich, C., Snijders, T.A.B. and Pearson, M. (2010) 'Dynamic networks and behavior: Separating selection from influence', Sociological Methodology, 40(1): 329–393. http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x
Tallberg, C. (2005) 'A Bayesian approach to modeling stochastic blockstructures with covariates', Journal of Mathematical Sociology, 29(1): 1–23. http://dx.doi.org/10.1080/00222500590889703
Van Duijn, M.A.J., Snijders, T.A.B. and Zijlstra, B.J.H. (2004) 'p2: A random effects model with covariates for indirected graphs', Statistica Neerlandica, 58(2): 234–54. http://dx.doi.org/10.1046/j.0039-0402.2003.00258.x
Van Duijn, M.A.J., van Busschbach, J.T. and Snijders, T.A.B. (1999) 'Multilevel analysis of personal networks as dependent variables', Social Networks, 21(2): 187–209. http://dx.doi.org/10.1016/S0378-8733%2899%2900009-X
Wang, Y.J. and Wong, G.Y. (1987) 'Stochastic blockmodels for directed graphs', Journal of the American Statistical Association, 82(397): 8–19. http://dx.doi.org/10.1080/01621459.1987.10478385
Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815478
Wasserman, S. and Pattison, P. (1996) 'Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p*', Psychometrika, 61(3): 401–25. http://dx.doi.org/10.1007/BF02294547
Wong, G.Y. (1987) 'Bayesian models for directed graphs', Journal of the American Statistical Association, 82(397): 140–48. http://dx.doi.org/10.1080/01621459.1987.10478406
Zijlstra, B.J.H. (2008) 'Random effects models for directed graphs with covariates'. Unpublished Ph.D. thesis, University of Groningen, The Netherlands.
Zijlstra, B.J.H., van Duijn, M.A.J. and Snijders, T.A.B. (2005) 'Model selection in random effects models for directed graphs using approximated Bayes factor', Statistica Neerlandica, 59(1): 107–18. http://dx.doi.org/10.1111/j.1467-9574.2005.00283.x
Zijlstra, B.J.H., van Duijn, Marijtje A.J., and Snijders, Tom A.B. (2006) 'The multilevelp2 model: A random effects model for the analysis of multiple social networks', Methodology2(1): 42–47.

Zijlstra, B.J.H., van Duijn, M.A.J. and Snijders, T.A.B. (2009) 'MCMC estimation of the p_2 network regression model with crossed random effects', British Journal of Mathematical and Statistical Psychology, 62(1): 143–66.

- covariates
- actors
- dyads
- citations
- random effects
- homophily
- reciprocity

http://dx.doi.org/10.4135/9781446294413.n31