# A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes

Miranda J. Lubbers [a,*], Tom A.B. Snijders [b,c]

[a] *Groningen Institute for Educational Research (GION), University of Groningen, The Netherlands*
[b] *Nuffield College, University of Oxford, UK*
[c] *ICS, Department of Sociology, University of Groningen, The Netherlands*

## Abstract

This paper describes an empirical comparison of four specifications of the exponential family of random graph models (ERGM), distinguished by model specification (dyadic independence, Markov, partial conditional dependence) and, for the Markov model, by estimation method (Maximum Pseudolikelihood, Maximum Likelihood). This was done by reanalyzing 102 student networks in 57 junior high school classes. At the level of all classes combined, earlier substantive conclusions were supported by all specifications. However, the different specifications led to different conclusions for individual classes. PL produced unreliable estimates (when ML is regarded as the standard) and had more convergence problems than ML. Furthermore, the estimates of covariate effects were affected considerably by controlling for network structure, although the precise specification of the structural part (Markov or partial conditional dependence) mattered less.
© 2007 Elsevier B.V. All rights reserved.

## 1. Introduction

The exponential family of random graph models (ERGMs) for social networks, also known as $p*$ models (Frank and Strauss, 1986; Frank, 1991; Wasserman and Pattison, 1996) were developed

---

* Corresponding author. Present address: The Autonomous University of Barcelona (UAB), Department of Social and Cultural Anthropology - Egoredes, Edifici B, 08193 Bellaterra (Barcelona), Spain.
*E-mail address:* MirandaJessica.Lubbers@UAB.es (M.J. Lubbers).

for the statistical analysis of complete networks. The present paper concerns models for single observations of digraphs, for which ERGMs specify the probability distribution for a digraph represented by its adjacency matrix $Y = (Y_{ij})$, where $1 \leq i, j \leq n$, $n$ is the number of actors in the network, and $Y_{ij}$ is the indicator variable of a tie from $i$ to $j$. The parameters of this distribution represent structural tendencies in the network (e.g., mutuality, transitivity), and if available, actor and dyadic attributes (e.g., gender, similarity in gender).

The $p1$ model (Holland and Leinhardt, 1981) assumes that dyads in the digraph (i.e., pairs $(Y_{ij}, Y_{ji})$) are mutually independent. The $p2$ model (Van Duijn et al., 2004) makes this assumption conditional on latent nodal variables. Such dyadic independence assumptions are usually not realistic for social networks. This led Frank and Strauss (1986) to propose Markov graphs, and Frank (1991) and Wasserman and Pattison (1996) to propose more general ERGMs. The Markov model assumes conditional dependencies only for tie variables between any two pairs $(i, j)$ and $(h, k)$ of actors that have an actor in common (e.g., $i = h$). Under the condition that all isomorphic configurations have the same probability under the model, Frank and Strauss (1986) proved that this implies that sufficient statistics for the model are counts of mutuals, triads of any kind, and stars of any order.

For the analysis of ERGMs, approximate pseudolikelihood estimation techniques (PL) were initially proposed by Frank and Strauss (1986), Straus and Ikeda (1990), and Wasserman and Pattison (1996). Pseudolikelihood estimation is quite convenient as it is algorithmically equivalent to logistic regression. However, the PL techniques were criticized because their statistical properties were unknown and the standard errors seemed to be underestimated (e.g., Snijders, 2002). Theoretically, it may be expected that the PL estimator overestimates structural effects in models that are characterized by a strong interrelatedness among the units, which is true for most social networks, and to perform adequately when the interrelatedness is weak (see Geyer and Thompson, 1992). In the case of independent units, PL and ML are the same.

More recently, Monte Carlo Markov Chain (MCMC) algorithms have been developed for ERGMs, which produce approximate Maximum Likelihood (ML) estimators. With the introduction of these algorithms, a new problem with ERGMs for social networks was revealed (Handcock, 2003; Snijders, 2002): In empirical networks, which are often characterized by high levels of transitivity, Markov models were often degenerate or near degenerate. This means in this case that for many empirical data sets, the ML estimates of parameters correspond to network distributions which will produce either low-density graphs or almost-complete graphs. As a consequence, Markov models are not good at representing real-life networks and the estimation procedure for these models frequently encounters convergence problems.

Snijders et al. (2006) diagnosed these problems – or at least partially – as deficiencies in the specification of the ERGM. They proposed new specifications for ERGMs that deal with the problem of degeneracy of distributions for many networks. Three types of statistics were proposed: (1) geometrically weighted degrees, to represent the distribution of degrees, (2) alternating transitive $k$-triangles, to represent transitivity, and (3) alternating independent two-paths, to represent the preconditions for transitivity. Further explanations can be found in Snijders et al. (2006) and Robins et al. (2007). The new parameters do not satisfy the classical assumption of Markov dependence, which is often too stringent, but they satisfy the assumption of partial conditional dependence (Pattison and Robins, 2002). That is, conditional dependencies are assumed for tie variables (1) between any two pairs of actors that have an actor in common (the classical Markov dependence), and in addition (2) between any two disjoint pairs of actors, if ties are observed between the two actors within each pair (so that ties also between the pairs would yield a four-

cycle). For some empirical data sets it was shown that the new specifications improve possibilities for parameter estimation.

As the new specifications were proposed only recently, there is little experience with the new specifications in empirical data. Robins et al. (2007) described their experience with the model based on twenty well-known data sets that are available in UCINET (Borgatti et al., 1999). They found that Markov models were degenerate for almost half of the non-directed data sets and three quarters of the directed data sets, whereas the models incorporating the new specifications (the partial conditional dependence model) were not degenerate for any of the data sets. Furthermore, PL standard errors were about 15% underestimated. They recommended that PL estimation should be used primarily as an exploratory method, to indicate the strength and direction of effects, but concluded that ML estimation is more accurate and should be used when available.

The aim of the present paper was to gain experience with the new specifications by analyzing 102 empirical networks of high school students in 57 school classes (boys and girls were analyzed separately; groups of fewer than 10 students were excluded). The data concern directed relations of liking among students in their school classes, further details are given below. Boys and girls were analyzed separately because they hardly nominated each other. This led to small networks, with sizes ranging from 10 to 20. It should further be noted that students could nominate up to three classmates (this was a sacrifice that had to be made for the large-scale data collection).

The data were analyzed earlier with PL estimation of a traditional Markov specification of the ERGM (Lubbers, 2003), and were here reanalyzed with ML estimation to make the comparison. Decisions about the model specification were taken in such a way as to enable a direct comparison with the results of Lubbers.

Two sets of comparisons were considered: (1) between Pseudolikelihood and Maximum Likelihood estimation of the Markov specification of the model used by Lubbers (2003); (2) between several model specifications, each estimated with Maximum Likelihood: (a) a dyadic independence model, (b) a traditional specification of the model as used by Lubbers, which satisfies the Markov dependence assumption, and (c) a new specification of the model proposed by Snijders et al. (2006), which satisfies the assumption of partial conditional dependence. The second set of comparisons was included to give insight into the consequences of how the structural part of the model is specified for the estimates for effects of covariates. The following sections describe each of these model specifications in more detail. For each approach, we carried out a meta-analysis to summarize the findings of the 102 networks, and explored whether the different approaches led to different substantive conclusions. The meta-analysis was performed analogously to the procedure used by Lubbers (2003) and is described in Section 4. Section 5 describes the data. Section 6 discusses the results of each of the four models at meta-level, and it also compares the coefficients of the different models and their *t*-values per class. The article finishes with a conclusion and discussion.

## 2. The four approaches

Four models were fitted to a total of 102 directed networks of high school students, described below in Section 5. The models included dependence on three dyadic covariates: same primary school attended, absolute difference in prior school performance, and different ethnic status. Further effects of these (and other) covariates would be substantively interesting as well, but these are not considered here because the focus is on the comparison of the methods.

## 2.1. A Markov specification of the ERGM, with PL estimation

As explained in Wasserman and Pattison (1996), a convenient and illuminating way to specify the ERGM is by specifying the conditional probabilities of ties, given the rest of the digraph. For any tie variable $Y_{ij}$, denote by $Y_{(ij)}^c$ the set of all elements of the adjacency matrix except for $Y_{ij}$. The ERGM implies a linear model for the log odds that a tie is present, conditional on the rest of the digraph:

$$\text{logit}\,(Pij) = \log\left(\frac{\Pr(Y_{ij} = 1|Y_{(ij)}^c)}{\Pr(Y_{ij} = 0|Y_{(ij)}^c)}\right) = \log\left(\frac{\Pr(Y_{ij} = y_{(ij)}^+)}{\Pr(Y_{ij} = y_{(ij)}^-)}\right) = \theta'(z(y_{(ij)}^+) - z(y_{(ij)}^-)) \tag{1}$$

where $y_{(ij)}^+$ denotes the adjacency matrix $Y_{(ij)}^c$ complemented with the tie $Y_{ij} = 1$, and $y_{(ij)}^-$ denotes the adjacency matrix $Y_{(ij)}^c$ complemented with the tie $Y_{ij} = 0$, $\theta$ denotes the vector of model parameters, and $z(y)$ the vector of network statistics, which are functions of the network $y$. The vector of change statistics, $(z(y_{(ij)}^+) - z(y_{(ij)}^-))$, thus expresses the vector of changes in network statistics when variable $Y_{ij}$ changes from 0 to 1, while the rest of the network is unaltered.

Based on the sufficient statistics for a directed Markov graph, seven structural effects were estimated for all networks: density (denoted $\phi$), mutuality ($\rho$), 2-out-star ($\sigma_O$), 2-in-star ($\sigma_I$), 2-mixed-star ($\sigma_M$), transitivity ($\tau_T$), and cyclicity ($\tau_C$). Stars of order three or more were not included because of the sampling scheme of our data. Besides the structural effects, the effects of the three dyadic covariates were added: same primary school attended (samesch$_{ij}$, parameter $\phi_{\text{school}}$), absolute difference in prior school performance (difPEFT$_{ij}$, $\phi_{\text{PEFT}}$), and different ethnic status (difethnic$_{ij}$, $\phi_{\text{ethnic}}$).

The corresponding network statistics for the 10 effects were defined as follows:

Density

$$z_0(y) = L = y_{i+} = \sum\nolimits_{i,j} y_{ij} \tag{2}$$

Mutuality

$$z_1(y) = M = \sum\nolimits_{i<j} y_{ij} y_{ji} \tag{3}$$

2-in-star

$$z_3(y) = S_I = \sum\nolimits_{i,j,k} y_{ji} y_{ki} \tag{4}$$

2-out-star

$$z_2(y) = S_O = \sum\nolimits_{i,j,k} y_{ij} y_{ik} \tag{5}$$

2-mixed star (two-paths)

$$z_4(y) = S_M = \sum\nolimits_{i,j,k} y_{ji} y_{ik} \tag{6}$$

Transitivity

$$z_5(y) = T_T = \sum\nolimits_{i,j,k} y_{ij} y_{jk} y_{ik} \tag{7}$$

Cyclicity

$$z_6(y) = T_C = \sum_{i,j,k} y_{ij} y_{jk} y_{ki} \tag{8}$$

Same primary school attended

$$z_7(y) = D_S = \sum_{i,j} y_{ij} \text{samesch}_{ij} \tag{9}$$

Difference in prior performance

$$z_8(y) = D_P = \sum_{i,j} y_{ij} \text{difPEFT}_{ij} \tag{10}$$

Difference in ethnic status

$$z_9(y) = D_E = \sum_{i,j} y_{ij} \text{difethnic}_{ij} \tag{11}$$

where $i$, $j$, and $k$ denote distinct students in a class. The variables $\text{samesch}_{ij}$, $\text{difPEFT}_{ij}$, and $\text{difethnic}_{ij}$ are explained in Section 5.2. The network statistics are subgraph counts. The network change statistics for these configurations are the changes between the value calculated for the network where the tie $y_{ij}$ is present and the value for the network in which the tie is absent.

## 2.2. A Markov specification of the ERGM, estimated with ML

We estimated the same model as in Section 2.1, but we conditioned on the number of ties, as unconditional estimation led to many convergence problems. Snijders et al. (2006) proposed to condition on the number of ties as a standard approach. Therefore, the density parameter was removed. Furthermore, we used Monte Carlo Maximum Likelihood estimation (further referred to as ML). This estimation procedure was implemented in the software SIENA (Snijders et al., 2005), which is part of the StOCNET environment (Boer et al., 2003).

## 2.3. A dyadic independence specification of the ERGM, estimated with ML

This model is defined by the effects of mutuality (Eq. (3)), same primary school attended (Eq. (9)), difference in prior performance (Eq. (10)), and difference in ethnic status (Eq. (11)). Under this model, the dyads are independent. This is an unrealistic model and almost always an inadequate representation of network structure, and the aim of the ERGM is precisely to improve on it. For the estimation, we used Monte Carlo Maximum Likelihood as implemented in the software SIENA, again conditioning on the number of ties. For this model, this conditioning is superfluous for obtaining convergence, but it makes very little difference in the results.

## 2.4. A partial conditional dependence specification of the ERGM, estimated with ML

The Markov specification is repeated (Eqs. (2)–(11)), but instead of 2-in-star, 2-out-star, 2-mixed-star, and transitive triplets (Eqs. (4)–(7)), the following statistics were included, in accordance with Snijders et al. (2006). Snijders et al. argued that these statistics are analogous to the four mentioned Markov model statistics, in that they represent, respectively, variability of in-degrees, variability of out-degrees, association of ingoing and outgoing ties for the same vertices,

and transitive closure, and that they satisfy the partial conditional independence condition, rather than the Markov condition.

Alternating in-*k*-star combinations

$$z_{10}(y) = S_{\mathrm{I}}^{\mathrm{a}} = \sum_{i=1}^{n} \mathrm{e}^{-\alpha y_{+i}} \tag{12}$$

Alternating out-*k*-star combinations

$$z_{11}(y) = S_{\mathrm{O}}^{\mathrm{a}} = \sum_{i=1}^{n} \mathrm{e}^{-\alpha y_{i+}} \tag{13}$$

Alternating independent two-paths

$$z_{12}(y) = S_{\mathrm{M}}^{\mathrm{a}} = \lambda \sum_{i,j} \left\{ 1 - \left( 1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\} \tag{14}$$

Alternating transitive *k*-triangles

$$z_{13}(y) = T_{\mathrm{T}}^{\mathrm{a}} = \lambda \sum_{i,j} y_{ij} \left\{ 1 - \left( 1 - \frac{1}{\lambda} \right)^{L_{2ij}} \right\} \tag{15}$$

where $\alpha$ is the degree weighting parameter ($\alpha > 0$), $y_{+i}$ represents the in-degree of node $i$, $y_{i+}$ represents the out-degree of node $i$, $\lambda = \mathrm{e}^{\alpha}(\mathrm{e}^{\alpha} - 1)$, and $L_{2ij} = \sum_{i,j,k} y_{ik} y_{kj}$. We chose $\lambda = 2$, corresponding to $\alpha = \ln(2)$, for which Robins et al. (2007) found good results. Again, we conditioned on the number of ties, so there is no density parameter, and we used ML estimation.

## 3. Estimation procedure

The PL estimations were carried out using the logistic regression option in SPSS (see Lubbers, 2003, for further details). The ML estimations were carried out using the software SIENA (Snijders et al., 2006). The algorithm was described in Snijders (2002), and its use in Robins et al. (2007). Convergence of the algorithm is assessed by *t*-statistics for convergence (Snijders, 2002, Eq. (34)). The algorithm was deemed to have converged when these were less than 0.1 in absolute value for all parameter estimates. For all classroom data sets, the estimation algorithm was used repeatedly until convergence was attained (except for the few cases where this seemed unattainable), always using the provisional estimates obtained earlier as the initial values for the new operation of the algorithm[1]. It seems to be quite usual for the Snijders (2002) algorithm to require several successive

---

[1] To obtain good convergence, the parameters of the algorithm must be chosen in a suitable way. The main parameters that can be determined by the user were specified as follows. The number of subphases was held at the standard value of 4. The run length for simulating one randomly drawn ERGM is determined by SIENA as $fn(n-1)/(2d(1-d))$ where $f$ is the so-called multiplication factor, $n$ is the number of actors in the network, and $d$ is the density. If $f$ is too low, the algorithm will not perform well because networks that are successively generated are mutually too highly correlated; if $f$ is too high, the computing time will be unnecessarily long. The multiplication factor was determined depending on the autocorrelations between the statistics computed from successively drawn ERGMs, given in the SIENA output. The multiplication factor was initially chosen as $f = 5$, but if at least one of the autocorrelations was higher than .1, then $f$ was increased to 10, 20, or 50, until all autocorrelations were less than 0.1. Finally, the step size for parameter updates as determined by $a_1$ in Eq. (35) in Snijders (2002), also called the initial gain parameter, must be chosen. If $a_1$ is too small, it will take long before the algorithm arrives near the Maximum Likelihood estimate. If $a_1$ is too large, the algorithm will be unstable and move about the Maximum Likelihood estimate without really getting there. This step size was initially set

Table 1
Number and percentage of networks for which a good convergence was achieved

| Estimation method and model specification | Boys' networks ($N_{tot} = 48$) | | Girls' networks ($N_{tot} = 54$) | |
|---|---|---|---|---|
| | N | % | N | % |
| PL Markov | 35 | 73 | 34 | 63 |
| ML Dyadic independence | 48 | 100 | 54 | 100 |
| ML Markov | 47 | 98 | 50 | 93 |
| ML Partial conditional dependence | 45 | 94 | 50 | 93 |

operations of the algorithm to obtain good convergence for estimating ERGM parameters, note that using the threshold of 0.1 to decide about convergence is a rather strict condition.

Proceeding in this way, convergence was achieved for most data sets in 2–4 repeated operations of the algorithm. In a minority of cases, more repetitions (up to about 8) were required. In a few cases, convergence was not achieved (see Table 1). The fact that all networks had a maximum out-degree of 3, and most had most out-degrees equal to 3, appeared to cause problems in the estimation. In the boundary case where all out-degrees are equal, the estimate for the 2-out-stars parameter in the Markov model mathematically is minus infinity, which for numerical purposes in these models is indistinguishable from $-20$. The Metropolis–Hastings algorithm for drawing from the ERGM distribution (Snijders, 2002) mixes less well for such parameter values, which leads to the need for the high value of $f = 50$ for the multiplication factor. We also used a version of the algorithm that uses the extra constraint that out-degrees are not larger than 3. This led to better convergence in a few cases but to non-convergence in a few others. For the data sets where both methods converged, conditioning on the maximal degree made little difference to the estimates. The only difference concerned the larger estimates for outdegree in the unconstrained model. Here we report the results without this extra condition, to allow a more straightforward comparison with the model as estimated with PL.

## 4. Meta-analysis

The ERGM analyses yielded a set of parameter estimates and standard errors for each of the $N$ classes, separately for boys and girls. To summarize these findings over classes, we follow the meta-analysis procedure described earlier (Lubbers, 2003). The coefficients of the classes were split into an average coefficient and a class-dependent deviation. Let $\theta$ denote one of the parameters. Then, the meta-level regression equation can be written as:

$$\hat{\theta}_m = \mu_\theta + U_m + E_m \tag{16}$$

where $\hat{\theta}_m$ denotes the estimated parameter value for class $m$, $\mu_\theta$ the average coefficient in the population, $U_m$ the true deviation of class $m$, which has a mean of 0 and a variance $\sigma_\theta^2$, and $E_m$ denotes the estimation error associated with the true parameter value $\theta_m$. The variance of $E_m$ is assumed to be known, and equal to the squared standard error obtained when estimating $\theta_m$. To differentiate between true variance and error variance, and thus to obtain more precise estimators for $\mu_\theta$ and $\sigma_\theta^2$, we need to account for the considerable differences in standard errors between

---

at .01. If the algorithm seemed close to a solution without really getting there, as evidenced by all convergence $t$-statistics being less than .4 but some of them remaining larger than .1, the step size was decreased to .001 or .0001.

classes (see Snijders and Baerveldt, 2003). The program MLwiN (Goldstein et al., 1998) was used to estimate model (16) by iterated weighed least squares. Similarly to the procedure of modeling a meta-analysis with known level-1-variation (e.g., Bryk and Raudenbush, 1992; Goldstein, 1995), this was done by creating a pseudo-level in such a way that each unit at level 2 has only one unit at level 1. The error variance was constrained to be equal to the squared standard error, and the true variance was modeled at level 2. As a result of this procedure, classes with large standard errors have less influence on the average effect size than classes with small errors.

The average effect size of a variable, $\mu_\theta$, indicates to what extent the effect occurs in the set of all classes. To test whether the average effect size is zero, a *t*-ratio of the average parameter estimate $\hat{\mu}_\theta^{\text{WLS}}$ and the associated standard error S.E. $(\hat{\mu}_\theta^{\text{WLS}})$ is used. This statistic has approximately a standard normal distribution. The *t*-ratio is:

$$t_{\mu\theta} = \frac{\hat{\mu}_\theta^{\text{WLS}}}{\text{S.E.}\,(\hat{\mu}_\theta^{\text{WLS}})} \tag{17}$$

and the numerator and denominator are produced by MLwiN when estimating (16).

## 5. Data

### 5.1. Subjects

The data in this paper were described in more detail in our earlier paper (Lubbers, 2003). The data were collected as part of a large-scale study in The Netherlands, the 'Longitudinal Cohort Studies in Secondary Education—Cohort 1999' (e.g., Kuyper et al., 2003). This study follows students from the first year in junior high until they leave full-time secondary education. We selected 57 school classes on the basis of three criteria: (1) class size of at least 20 students; (2) all students completed the sociometric questionnaire; (3) no more than 2 students within the class had missing data on the covariate prior performance level. Missing data on the other covariates were far less frequent; remaining missing data were imputed on the basis of other available variables referring to the same concepts. All data used here were collected when the students were in Grade 1 (equivalent to Grade 7 in the US), half a year after they entered junior high. In total, there were 1466 students within these classes (53% females; 11% ethnic minorities; mean age 13.0 years).

Although the classes were mixed-gender, the networks appeared to be almost completely separated by gender (only 3% of all ties were cross-gender; Lubbers, 2003), hence we decided to analyze boys' and girls' networks separately. This approach was also employed in the three new models. As we decided to exclude networks of fewer than 10 boys or girls, the number of networks is $N = 48$ for boys (containing 620 boys) and $N = 54$ for girls (containing 748 girls).

### 5.2. Measures

#### 5.2.1. Social network

Relations of liking were measured using the sociometric item 'Whom of your classmates do you like most?' (free recall). The maximum of nominations in response to this item was limited to three. Eighty percent of the students in this selection nominated the maximum of three classmates, 13% nominated two classmates, 4% nominated one classmate, and 3% did not nominate any classmate in response to this question. Boys nominated fewer classmates than girls ($M_{\text{boys}} = 2.63$;

$M_{girls} = 2.77$; unequal-variances $t = -3.671$, $df = 1218$, $p < .001$). Ties were observed in less than 1% of all cross-gender dyads, and in 20% of all same-gender dyads.

### 5.2.2. Dyadic covariates

- *Same primary school attended.* The students were asked to write down the name and address of their primary school. On the basis of this information, a binary variable was constructed at dyad level, samesch$_{ij}$, which was coded 1 if students $i$ and $j$ had attended the same primary school and else 0. Students in 14% of all dyads attended the same primary school.
- *Absolute difference in entry level of school performance (prior performance).* The students' entry level was measured with their scores on the Primary Education Final Test (PEFT), developed by the National Institute for Educational Measurement (CITO). This standardized national test assesses the performance level of children in the final year of primary education in arithmetic, Dutch language, and information processing. The test scores were obtained via the student records of the junior high schools. They range theoretically between 501 and 550 (In the selection of cases for this paper, $M = 539.2$; S.D. $= 7.4$). On the basis of these test scores, a variable difPEFT$_{ij}$ was constructed that indicates the absolute difference between student $i$ and student $j$. The average difference in test scores between two students in a dyad was 4.5 with a standard deviation of 3.8.
- *Different ethnic status.* The ethnic status of students was determined on the basis of the country of birth of the students and their parents. When at least either both parents or one parent and the child were born outside the Netherlands, the student was characterized as non-Dutch. As the categories distinguishing between ethnicities all contained relatively few cases when studied per class, they were grouped together, so that ethnic status only reflects a distinction between ethnic minority and ethnic majority. The percentage of ethnic minorities in classes varied from 0 to 43% with an average of 11%. At dyad level, the binary variable difethnic$_{ij}$ was constructed, which was coded 0 if the students $i$ and $j$ were either both majority or both minority students and else 1. Eighteen percent of all dyads were mixed (majority–minority) dyads.

## 6. Results

### 6.1. A summary of the Markov specification of the ERGM, with PL estimation

In the original study (Lubbers, 2003), we had to omit 13 networks for boys (remaining $N = 35$) and 20 networks for girls (remaining $N = 34$; see also Table 1), due to high levels of collinearity or estimation problems. High collinearity usually concerned at least the change statistics for density and 2-out-stars, which is due to the facts that the maximum number of nominations was three and that many students chose precisely three classmates.

For both boys and girls, the strongest effects were observed for reciprocity, 2-out-stars, transitivity, and same primary school attended (see the PL columns in Table 2). Thus, these tendencies best describe the network structure. The negative average effect size of 2-out-stars merely reflects the design of the sociometric questionnaire, which has a fixed maximum on the number of nominations. The strong positive effect of reciprocity suggests that there is a high tendency toward reciprocity: The odds for a tie is on average between $e^{1.839} = 6$ and $e^{2.084} = 8$ times greater when the potential choice is reciprocated, given the other effects. The strong positive effect size of transitivity in the presence of the weaker negative average effect sizes of cyclicity and 2-mixed-stars suggest that two-paths as part of transitive triads are far more likely than intransitive two-paths

Table 2
Meta-analysis comparison of estimation methods—Markov model

| Variables | Boys' networks ($N=35$) | | | | | | Girls' networks ($N=34$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PL Markov | | | ML Markov | | | PL Markov | | | ML Markov | | |
| | $\hat{\mu}_{WLS}$ | (S.E.) | $\hat{\sigma}^2$ | $\hat{\mu}_{WLS}$ | (S.E.) | $\hat{\sigma}^2$ | $\hat{\mu}_{WLS}$ | (S.E.) | $\hat{\sigma}^2$ | $\hat{\mu}_{WLS}$ | (S.E.) | $\hat{\sigma}^2$ |
| Network effects | | | | | | | | | | | | |
| Density | .434 | (.279) | .683 | – | – | – | **2.329** | (.496) | 5.098 | – | – | – |
| Mutuality | **2.084** | (.202) | .770 | **2.015** | (.170) | .000 | **1.863** | (.186) | .535 | **2.048** | (.167) | .002 |
| 2-in-star | **−.089** | (.042) | .000 | .019 | (.042) | .001 | −.029 | (.045) | .012 | .033 | (.034) | .000 |
| 2-out-star | **−1.279** | (.131) | .360 | **−.957** | (.109) | .025 | **−1.871** | (.159) | .531 | **−1.299** | (.141) | .137 |
| 2-mixed-star | **−.276** | (.057) | .034 | **−.237** | (.053) | .000 | **−.385** | (.067) | .069 | **−.311** | (.052) | .000 |
| Transitivity | **.803** | (.062) | .045 | **.719** | (.059) | .018 | **.708** | (.055) | .026 | **.714** | (.052) | .006 |
| Cyclicity | −.368 | (.189) | .600 | **−.396** | (.161) | .000 | −.007 | (.168) | .427 | **−.332** | (.163) | .076 |
| Dyadic attributes | | | | | | | | | | | | |
| Same school | **.748** | (.146) | .033 | **.750** | (.066) | .000 | **.818** | (.133) | .000 | **.671** | (.053) | .000 |
| Dif. perform. | .016 | (.018) | .000 | −.031 | (.132) | .000 | −.042 | (.026) | .006 | **−.259** | (.131) | .034 |
| Dif. ethnic.[a] | .045 | (.181) | .000 | −.070 | (.105) | .000 | −.257 | (.149) | .000 | −.161 | (.095) | .000 |

*Note*: $\hat{\mu}_{WLS}$, estimated average effect size; (S.E.), standard error associated to estimated average effect size; $\hat{\sigma}^2$, estimated variance of the effect size between classes. Significant average effect sizes at $p < .05$ are printed in bold. The PL Markov results in this table were reprinted from Lubbers (2003), with permission from Elsevier.

[a] For the attribute 'different ethnicity', $N=23$ boys' networks and $N=24$ girls' networks.

or two-paths as part of cycles. In other words, when we focus on the average picture, we would conclude that relationships between students tend to be hierarchically structured. The positive effect of same primary school suggests that a tie is about 3 times more likely to be observed in a pair of students who have attended the same primary school than in a pair of students who attended different primary schools.

## 6.2. A comparison of two estimation methods of the Markov specification of the ERGM

We reanalyzed the Markov model with Monte Carlo Maximum Likelihood estimation. Table 1 shows that the estimation converged and led to a good fit for a number of networks that is considerably higher than in the PL approach ($N = 47$ for boys; $N = 50$ for girls). For all the networks for which we obtained PL estimates, we also obtained ML estimates. The results in Table 2 are based on these networks.

Similar to the results from the unconditional Markov model with PL estimation, the strongest effects were observed for reciprocity, 2-out-stars, transitivity, and same primary school attended, for boys and girls. When we compare the average effect sizes of the significant tendencies between PL and ML in Table 2, we can draw the same general conclusions. However, it is striking to note that the true variance of most of the coefficients is estimated to be much smaller when we used ML.

When we compared the estimates of the two approaches *per class* (Fig. 1), it appeared that they can yield substantially different estimates. For reciprocity (Fig. 1a), we saw that most of the coefficients have quite similar effect sizes for the two estimation methods, and that only a few classes have strongly deviating values. For same primary school (Fig. 1c), on the other hand, it appears that there is hardly a correlation between the PL and the ML estimates. However, it should be noted that the estimates were not significant for all of the classes (see for a comparison of *t*-values Fig. 2). So, some of the cases in the graphs of Fig. 1 are based on estimates that do not differ significantly from zero.

A comparison of the standard errors in Table 2 shows that for the majority of coefficients in the meta-analysis, the standard errors estimated with ML were smaller than the standard errors estimated with PL. In contrast to this, when we compare the standard errors at the class level, it appears that for two-thirds of the effects and for most of the classes, the standard errors estimated with PL were smaller than those estimated with ML. This suggests that PL underestimates the standard errors, but seems at variance with the tendency for ML to give smaller standard errors in the meta-analysis. Fig. 2 shows the *t*-values (see Eq. (17)) of PL versus ML estimation for the three substantively interesting coefficients that had significant meta-level effects (reciprocity, transitivity, and similarity in primary school). From this figure, it becomes clear that the *t*-values for reciprocity and transitivity are generally larger when estimated with PL than with ML, while the reverse is true for similarity in primary school. More specifically, the figure shows that for reciprocity and transitivity, there is quite a number of networks for which PL does and ML does not indicate a significant result ($t > 1.96$). The reverse is true for similarity in primary school. It is not clear why PL tends to underestimate some standard errors and overestimate others. Possible influences, such as average degree, structural effects as estimated in the ML Markov model, group size and group composition were not significant in an analysis that compared these two methods.

The fact that most of the standard errors associated to the *average* effect sizes estimated with ML are smaller than those estimated with PL (Table 2), despite the larger *individual* standard errors for ML, can be explained from the calculation of the standard error in the meta-
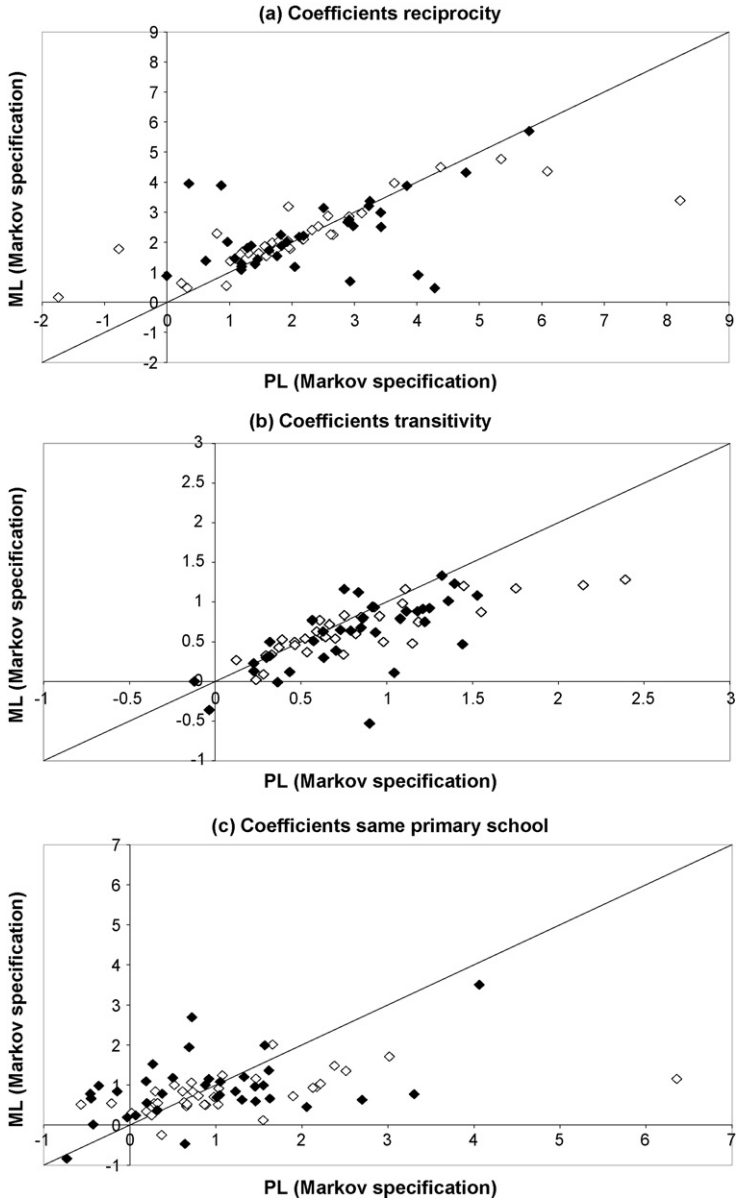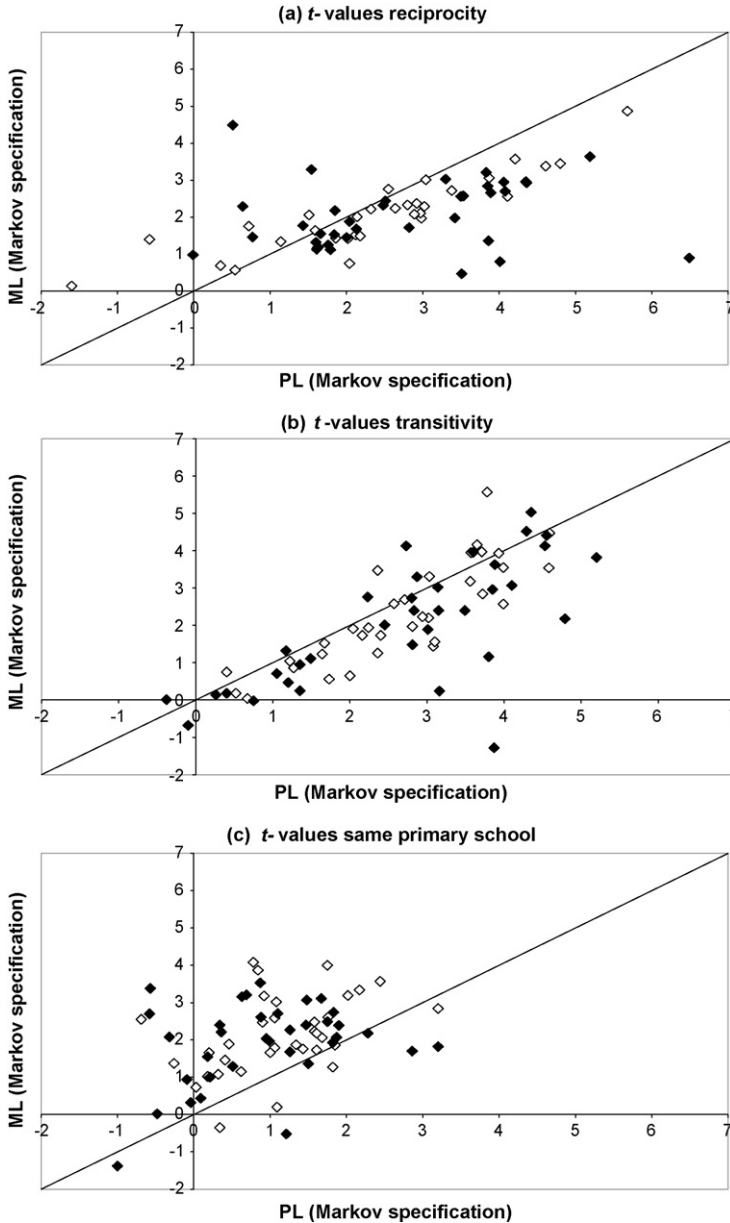
Fig. 1. A comparison of the coefficients in the Markov model estimated with PL vs. ML for 69 networks. *Note*: The black nodes represent the estimates for boys' networks ($N = 35$); the white nodes represent the estimates for girls' networks ($N = 34$).

analysis: This standard error contains a term that is proportional to the variance of the parameter between the classes (see Snijders and Baerveldt, 2003, Eq. (14)). As the between-class variance is smaller for the ML estimates, so is the estimated standard error associated with the average estimate.

Fig. 2. A comparison of the *t*-values in the Markov model estimated with PL vs. ML for 69 networks. *Note*: The black nodes represent the estimates for boys' networks (*N* = 35); the white nodes represent the estimates for girls' networks (*N* = 34).

## 6.3. A comparison of three model specifications of the ERGM, with ML estimation

Next, we analyzed the dyadic independence model and the partial conditional dependence model, both with ML estimation, and compared the outcomes with the Markov model that was also
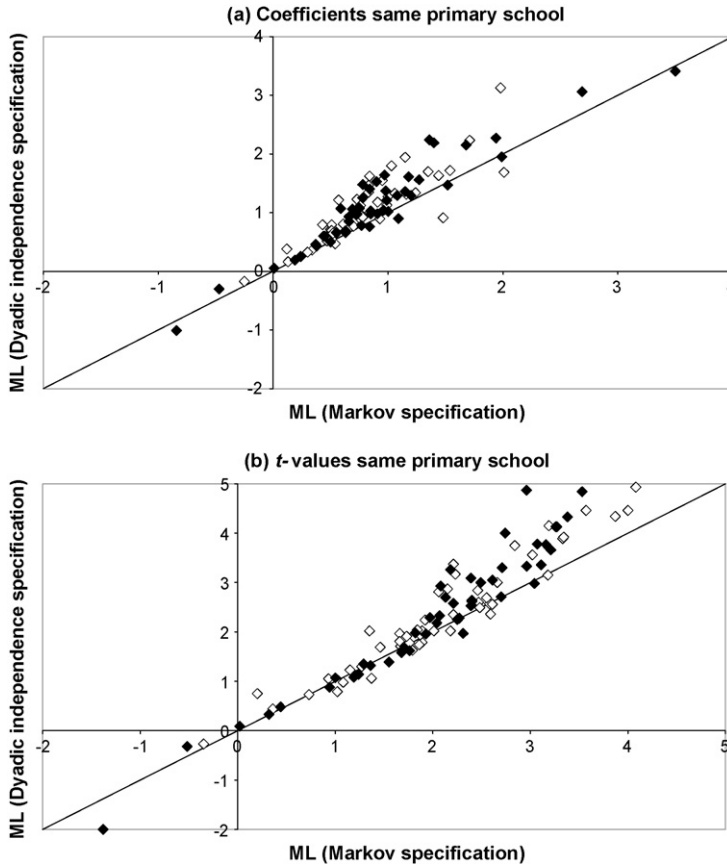
Fig. 3. A comparison of the estimates of the attribute 'same primary school attended' (coefficients and *t*-values) for 97 networks, estimated with ML: Markov specification versus dyadic independence specification. *Note*: The black nodes represent the estimates for boys' networks ($N = 47$); the white nodes represent the estimates for girls' networks ($N = 50$).

estimated with ML. We did not encounter any convergence problems for the dyadic independence model. For the partial conditional dependence model, the estimation converged and led to a good fit for 45 networks for boys and 50 networks for girls (see Table 1). These 45 boys' networks and 48 of the girls' networks showed good convergence for all three model specifications estimated with ML. Table 3 is based on these networks.

As Table 3 shows, we again found significant average effects of mutuality and similarity in primary school for the dyadic independence model, whereas the average effects of same ethnicity and similar prior performance level were not significant. The average effect sizes of mutuality and similarity in primary school were clearly higher in the dyadic independence model than they were in the Markov model (around 21% and 36% higher, respectively); in the latter model, the structural effects take over part of the dyadic similarity effects. At class level, Fig. 3a shows that the effect sizes of same primary school were indeed larger in the dyadic independence model, and Fig. 3b shows that the *t*-values for similarity in primary school also tend to be larger in the dyadic independence model than in the Markov model. This shows the importance of controlling for network structure.

Table 3
Meta-analysis comparison of method specifications (with ML estimation)

| Variables | Dyadic independence | | | Markov | | | Partial conditional dependence | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\mu}_{\text{WLS}}$ | (S.E.) | $\hat{\sigma}^2$ | $\hat{\mu}_{\text{WLS}}$ | (S.E.) | $\hat{\sigma}^2$ | $\hat{\mu}_{\text{WLS}}$ | (S.E.) | $\hat{\sigma}^2$ |
| **Boys' networks ($N=45$)** | | | | | | | | | |
| Network effects | | | | | | | | | |
| Density | – | – | – | – | – | – | – | – | – |
| Mutuality | **2.365** | (.114) | .054 | **2.093** | (.163) | .000 | **2.104** | (.141) | .000 |
| 2-in-star | – | – | – | −.009 | (.038) | .005 | −.111 | (.098) | .000 |
| 2-out-star | – | – | – | **−.958** | (.106) | .032 | **−1.256** | (.189) | .213 |
| 2-mixed-star | – | – | – | **−.238** | (.052) | .000 | **−.209** | (.040) | .000 |
| Transitivity | – | – | – | **.747** | (.051) | .017 | **.874** | (.069) | .020 |
| Cyclicity | – | – | – | **−.522** | (.143) | .000 | **−.469** | (.123) | .000 |
| Dyadic attributes | | | | | | | | | |
| Same school | **1.075** | (.089) | .133 | **.777** | (.057) | .000 | **.779** | (.056) | .000 |
| Dif. perform. | .005 | (.109) | .000 | −.100 | (.113) | .000 | −.084 | (.112) | .000 |
| Dif. ethnic.[a] | −.026 | (.077) | .000 | −.083 | (.089) | .000 | −.073 | (.087) | .000 |
| **Girls' networks ($N=48$)** | | | | | | | | | |
| Network effects | | | | | | | | | |
| Density | – | – | – | – | – | – | – | – | – |
| Mutuality | **2.458** | (.115) | .180 | **1.970**[b] | (.154) | .000 | **1.939** | (.133) | .000 |
| 2-in-star | – | – | – | −.005 | (.029) | .001 | **−.229** | (.091) | .005 |
| 2-out-star | – | – | – | **−1.270** | (.147) | .215 | **−1.964** | (.294) | 1.095 |
| 2-mixed-star | – | – | – | **−.358** | (.048) | .000 | **−.290** | (.038) | .000 |
| Transitivity | – | – | – | **.743** | (.041) | .000 | **.874** | (.055) | .000 |
| Cyclicity | – | – | – | **−.515** | (.134) | .061 | **−.513** | (.109) | .009 |
| Dyadic attributes | | | | | | | | | |
| Same school | **.953** | (.063) | .027 | **.683** | (.046) | .000 | **.686** | (.046) | .000 |
| Dif. perform. | −.170 | (.100) | .000 | **−.273** | (.102) | .000 | **−.261** | (.103) | .000 |
| Dif. ethnic.[a] | −.060 | (.071) | .000 | −.141 | (.080) | .000 | −.119 | (.080) | .000 |

*Note*: $\hat{\mu}_{\text{WLS}}$, estimated average effect size; (S.E.), standard error associated to estimated average effect size; $\hat{\sigma}^2$, estimated variance of the effect size between classes. Significant average effect sizes at $p < .05$ are printed in bold.

[a] For the attribute 'different ethnicity', $N=30$ boys' networks and $N=32$ girls' networks.

[b] The mutuality parameter in the Markov model for girls' networks was estimated for $N=46$ networks. Two networks appeared to be outliers, which strongly affected both the average effect and the variance (For $N=48$, $\hat{\mu}_{\text{WLS}} = 1.801$, S.E. $= .196$, $\hat{\sigma}^2 = .750$).

For the partial conditional dependence model, we also found that the strongest average effects were observed for reciprocity, alternating *k*-out-stars, alternating *k*-triangles, and same primary school attended, for boys and girls. The direction of these average effects was also similar. Again, the negative effect of alternating *k*-out-stars merely reflects the design of the sociometric questionnaire, indicating that students did not tend to nominate many classmates. The other effects reflect tendencies toward reciprocity, transitive closure, and toward relating to students from the same primary school. The interpretation of the parameters in the partial conditional dependence model deviates slightly from that of the Markov parameters (see Robins et al., 2007; Snijders et al., 2006). Most importantly, the positive alternating *k*-triangle statistic, in the presence of the alternating independent 2-paths, does not only indicate that students who were related via multiple independent 2-paths tended to form direct relations, but also inform us about the higher order structure in the data: students tended to form triangulated, rather than a degree-based (as can be deduced from the negative parameter for *k*-in-stars), core–periphery structures. In short however, we can conclude that the different types of analyses all supported the earlier conclusions at the meta-level.
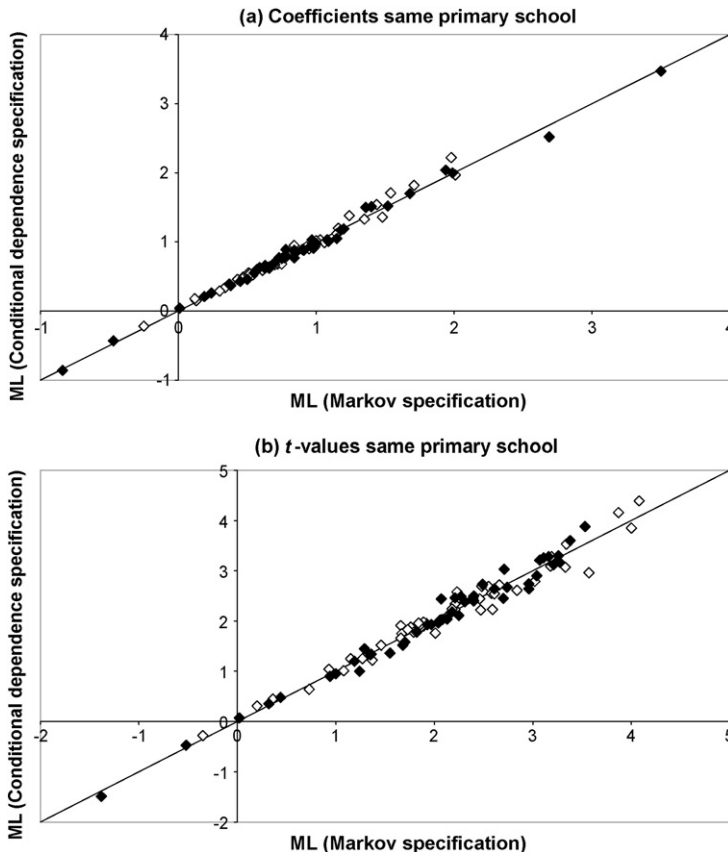


Fig. 4. A comparison of the estimates of the attribute 'same primary school attended' (coefficients and *t*-values) for 93 networks, estimated with ML: Markov specification versus partial conditional dependence specification. *Note*: The black nodes represent the estimates for boys' networks (*N* = 45); the white nodes represent the estimates for girls' networks (*N* = 48).

When we compare the Markov model with the partial conditional dependence model, both estimated with ML, it is interesting to note that both the coefficients and the standard errors of the dyadic attributes hardly seem to be affected by the precise specification of the structural part of the model. This is observed at the meta-level (see Table 3) and at the class level. As an illustration of the latter, Fig. 4 shows that the estimated coefficients and the *t*-values for the dyadic attribute 'same primary school' have very similar sizes for the two specifications. This pattern was also observed for the other two dyadic covariates, even though their effects were in both approaches insignificant.

### 6.4. Convergence of the partial conditional dependence model

As indicated earlier, convergence of the model using the new specifications was not always achieved. To have a sense of why the model did not always converge, we compared the groups that did convergence ($N = 95$) under the partial conditional dependence model and the groups that did not converge under this model ($N = 7$). The two sets of groups differed significantly in average degree (for the converged groups, $M = 2.55$; for the groups that did not converge, $M = 2.91$; $t = 7.684$, $df = 20$, $p < .001$), but not in group size or amount of structure in the data.

However, a high average degree did not necessarily lead to non-convergence. There were 24 classes that did converge with equally high average degrees, including a few cases where all students nominated exactly three students. It is likely that among the classes with high degrees, factors such as group size and the amount of structure in the data did play a role. A comparison between the non-converging groups and the 24 converging groups with equally high degrees (degrees $\geq 2.80$, leading to the same average degree in the converging set as in the non-converging set) only showed a bivariate difference in the proportion of reciprocal dyads ($t = -2.866$, $df = 2$, $p < .01$), indicating that reciprocity was somewhat larger in the non-converging groups. So, we concluded that with a fixed maximum choice design of three nominations, convergence was not achieved for a small set of groups that had average agrees close to the maximum. Why some groups with high average degrees did and other groups did not converge, also seems to depend on other factors such as the amount of structure observed in the data.

## 7. Discussion

In this paper, we compared four approaches to the Exponential Random Graph Model. For this aim, we returned to the data that we analyzed earlier (Lubbers, 2003) by Pseudolikelihood estimation of a Markov model. The data concern 102 networks of students within school classes, of sizes ranging from 10 to 20. Pseudolikelihood estimators have been criticized as being unreliable, and recently, Monte Carlo Maximum Likelihood procedures were proposed as an alternative estimation method with better properties. Furthermore, new model specifications have been proposed in the literature to avoid the model degeneracy that was observed in Markov models; these new specifications satisfy the assumption of partial conditional dependence, which is less stringent than the classical assumption of Markov dependence.

Three specifications of the ERGM were tested, using the Monte Carlo Maximum Likelihood procedure: (1) a Markov model that was comparable to the model used in Lubbers, (2) a dyadic independence model, and (3) a partial conditional dependence model. For each variation, we conditioned on the number of observed ties in the estimation, as without this step, the models in specifications (1) and (3) converged for only a few classes.

At the meta-level (i.e., averaged over the classes), the results show that the four sets of analyses generate roughly the same conclusions about networks of students in school classes, although the effect sizes of parameters vary. All analyses suggest strong tendencies toward reciprocity and transitivity[2] and a tendency to pick classmates whom students knew from primary school. Having the same ethnic status (i.e., ethnic majority or minority) and similar prior performance level did not affect the probability that a student liked a classmate. Hence, the results we reported earlier were also supported by different model specifications and a different estimation method.

However, when we compared the results at the class level instead of at the meta-level, we found some important differences between the two estimation methods and (for ML) between the three model specifications. First, with respect to the estimation method, we found that the PL estimates of the Markov model could deviate quite substantially from the ML estimates. Also, for two-thirds of the effects, PL estimation led for the majority of the classes to smaller standard errors at class-level than ML for the majority of classes, and as a result, more class-level effects were reported as significant in PL than in ML. So, earlier conjectures that PL might underestimate the standard errors (in comparison with ML) are supported for most but not all of the coefficients. For some coefficients, PL even tends to overestimate the standard errors. Therefore, we advise against the use of PL estimation.

Second, it appeared that Monte Carlo Maximum Likelihood gave a good convergence for these small networks, regardless of the model specification, whereas Pseudolikelihood led to non-convergence for one third of the networks. The latter is presumably due to the near collinearity of the change statistics and related to the maximum allowed number of three nominations in the network data collection.

With respect to the specification of the model, the results indicated that the estimates of the *structural effects* differ of course between the Markov model and the partial conditional dependence model, which is due to the different parameterizations, but the two specifications hardly differed in their estimates of average effects and standard errors of the *dyadic attributes*. In comparison with these two models, the dyadic independence model gave strongly deviating estimates for the average effects as well as standard errors of the dyadic attributes. This implies that for testing covariates, it is important to control for structural effects, but the precise specification of the structural part (i.e., Markov specification or partial conditional dependence specification) seems to matter less. This inspires confidence in the control for network structure offered by the partial conditional dependence model as well as by the Markov model (when it yields converging estimates) in case where the focus is primarily on estimating effects of covariates.

As stated before by Snijders et al. (2006) and Robins et al. (2007), further development and experimenting with the new specifications is required to attain a well-balanced methodology for the statistical modeling of social networks. The current paper contributes to the experience with applying Exponential Random Graph Models by presenting results for a large number of small friendship networks. Actor or dyadic covariates may be among the elements that could improve empirical models, as well as interactions between such covariates and structural elements. Moreover, it is likely that additional higher-order structural effects (see for examples Snijders et al., 2006) could further improve the models. It is likely that such improvements depend in part on the network variable that is studied.

---

[2] Obviously, as the effect for transitivity was not included in the dyadic independence model, this result was not supported by the dyadic independence model.

## Acknowledgement

## References

Boer, P., Huisman, M., Snijders, T., Zeggelink, E., 2003. StOCNET: An open software system for the advanced analysis of social networks. Version 1.4. ProGAMMA/ICS, Groningen.

Borgatti, S.P., Everett, M.G., Freeman, L.C., 1999. UCINET 5 for Windows: Software for Social Network Analysis. Analytic Technologies, Natick.

Bryk, A.S., Raudenbush, S.W., 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage, Newbury Park, CA.

Frank, O., Strauss, D., 1986. Markov graphs. Journal of the American Statistical Association 81, 832–842.

Frank, O., 1991. Statistical analysis of change in networks. Statistica Neerlandica 45, 283–293.

Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo maximum likelihood calculations (with discussion). Journal of the Royal Statistical Society, Series C 54, 657–699.

Goldstein, H., 1995. Multilevel Statistical Models. Edward Arnold, London.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G., Healy, M., 1998. A User's Guide to MLwiN. Multilevel Models Project. Institute of Education, University of London.

Handcock, M.S., 2003. Assessing degeneracy in statistical models of social networks. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 39.

Holland, P.W., Leinhardt, S., 1981. An exponential family of probability densities for directed graphs. Journal of the American Statistical Association 76, 33–51.

Kuyper, H., Lubbers, M.J., Van der Werf, M.P.C., 2003. VOCL'99-1: Technisch rapport (VOCL'99-1: Technical report). GION, Groningen, The Netherlands.

Lubbers, M.J., 2003. Group composition and network structure in school classes: a multilevel application of the p* model. Social Networks 25, 309–332.

Pattison, P.E., Robins, G.L., 2002. Neighbourhood based models for social networks. In: Stolzenberg, R.M. (Ed.), Sociological Methodology, vol. 22. Blackwell Publishing, Boston, MA, pp. 301–337.

Robins, G., Snijders, T.A.B., Wang, P., Handcock, M., Pattison, P., 2007. Recent developments in Exponential Random Graph (p*) Models for social networks. Social Networks 29, 192–215.

Snijders, T.A.B., 2002. Markov Chain Monte Carlo estimation of exponential random graph models. Journal of Social Structure 3, Article 2. Retrieved from http://www.cmu.edu/joss/content/articles/volume3/Snijders.pdf.

Snijders, T.A.B., Baerveldt, C., 2003. A multilevel network study of the effects of delinquent behavior on friendship evolution. Journal of Mathematical Sociology 27, 123–151.

Snijders, T.A.B., Pattison, P.E., Robins, G.L., Handcock, M., 2006. New specifications for exponential random graph models. Sociological Methodology 36, 99–153.

Snijders, T.A.B., Steglich, C., Schweinberger, M., Huisman, M., 2005. Manual for SIENA Version 2.1. University of Groningen, Groningen.

Straus, D., Ikeda, M., 1990. Pseudolikelihood estimation for social networks. Journal of the American Statistical Association 5, 204–212.

Van Duijn, M.A.J., Snijders, T.A.B., Zijlstra, B.H., 2004. P2: a random effects model with covariates for directed graphs. Statistica Neerlandica 58, 234–254.

Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks. I: An introduction to Markov graphs and p*. Psychometrika 61, 401–425.