

20

Multiple Groups

A network analysis for a single network is, at the network level, an $N = 1$ study: it tells us something about the rules of behavior followed by the actors in this particular network. We may have arguments for thinking that the network under study may be regarded as being representative for a larger population of networks, and these may even be good arguments, but the reasons for such a generalization – if any – are substantive and conceptual (cf. Snijders, 2011, p. 136), and cannot be internal to this particular network data set.

To obtain a statistical basis for a generalization to a (hypothetical) population of networks, one needs to analyze data for several networks that may be regarded, in some sense, as replications of each other. This point was raised and elaborated by Snijders and Baerveldt (2003) and Entwisle et al. (2007). The ‘replications’ could be network studies in several similar schools, several similar companies, etc. An early study of this kind was Coleman (1961) with an impressive study of friendship networks in 10 schools. More recent examples are the Add Health data (Harris et al., 2003), the PROSPER study (Moody et al., 2011), and the ASSIST study (Steglich et al., 2012). If the number of networks is large enough, such a study also permits the assessment of effects of variables at the network level.

This chapter treats four methods for such multiple group network data sets — in decreasing strictness of assumptions made. In all cases, the available data are $G \geq 2$ network panel data sets, with or without corresponding dependent behavior variables, all having the same variables available; the number of actors will most often differ across the G data sets; the number of waves could be the same or could be different between the data sets. The G data sets will for brevity be referred to as ‘networks’ or ‘groups’. Always, it will be assumed that there is stochastic independence between all the G networks, and that within each data set a stochastic actor-oriented model applies with the same specification, i.e., the same dependent and explanatory variables and the same set of effects.

The first method assumes that all networks are exact replicates of each other in the sense that the dynamics in all networks obeys the same model with the same parameter values; the second approach assumes that the networks obey the same model but with different parameters, following a normal distribution in the population of networks; and the third still assumes that the networks obey the same model, but now with different parameters that have an arbitrary unknown distribution in the population of networks. The fourth approach is still more parsimonious in its assumptions, and does not even assume that we can speak about a population of networks. The last section gives some closing remarks comparing these four approaches.

20.1 Replicate Groups

The most restrictive approach to analyzing multiple groups assumes that the dynamics is identical across the G networks, except for the basic rate parameters. Thus, not only the model specification but also the parameters (excepting the basic rate parameters) are assumed to be identical across groups. It is permitted that there are variables depending on the groups only, and including their effects and their interactions with other effects will still permit to represent differences between the groups. (In the extreme case, including dummy variables for all the groups and giving these interactions with all other effects would allow the groups to follow totally different specifications, bringing us to the general case of Section 20.3, but this would yield a self-defeatingly high number of parameters.) This situation is totally analogous to the analysis of M consecutive waves, with $M \geq 3$, with a common model, the basic approach of the SAOM whenever there are 3 or more waves. There also we assume identical parameters across the $M - 1$ periods, and period-dependent variables may be used to give some amount of difference in the model between periods.

The assumption of identical parameters across groups is highly restrictive, and should be tested because violations may lead to unreliable (perhaps even meaningless) results. This assumption should be made only when a priori it is plausible that the network dynamics in the various groups is quite the same; and making this assumption will be helpful especially when each of the groups is too small to be analyzed on its own using the model of interest. This approach could be meaningful, e.g., in cases like the following.

- Two or more relatively small groups are studied in the same wider setting: classes in the same school, departments in the same firm.
- The same network experiment is carried out in multiple small groups all having different participants.
- The data come from almost the same group of actors, but some kind of rearrangement of the actors has taken place. E.g., the actors are firms and at some moment some of the firms have merged into a smaller number of new firms. Then the first network data set refers to the dynamics before the date of mergers, and the second data set to the dynamics after this date. This is a particular kind of change of composition, different from the case treated in Section 9.1.

The assumption of independence between the groups, in this case where the groups are largely composed of the same actors, goes back to the assumption of a Markov chain for the overall dynamics.

This treatment of multiple groups can be carried out by stringing the waves for the different groups together in a consecutive sequence, allowing the number of actors to change at the transition from one group to the next and not analyzing these group-to-group transitions, as suggested by Figure 20.1. This is done by the function `sienaGroupCreate` in `RSiena`.

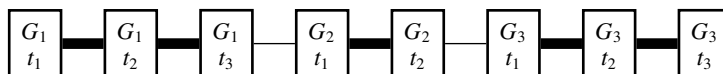


Figure 20.1 The waves for three groups strung in a sequence, where the first and third groups have 3 waves, and the second has 2 waves; the analyzed 2+1+2 periods are indicated as bold lines.

Another, more cumbersome, way to combine multiple groups is available if the number of waves is the same for all groups. Then an artificial data set is constructed, in which for each wave the G networks are combined into one big network, with the total number of actors equal to the sum of actors across the groups, and with structural zeros (Section 9.3.1) between the networks to represent that ties between different groups are not permitted. In this construction it is necessary also to include dummy variables for the groups and to make the rates of change dependent on these dummy variables.

20.1.1 Example

As an example, we consider network dynamics across two waves in seven classes in the same school. The data is part of a data set collected by Andrea Knecht (Knecht, 2008; Knecht et al., 2010), see Section 2.3. Here we study the seven school classes in the school with code number 03, for the second and third waves (here relabeled as 1 and 2). The students were about twelve years old, and were requested to list up to twelve classmates as ‘best friends’. Table 20.1 gives some basic descriptives of the data set. The average degrees are quite homogeneous, almost all between 4 and 5, and the Jaccard coefficients are quite high and likewise homogeneous.

Table 20.1 *Basic descriptives for multi-group data set from school 03 of Knecht (2008).*

		<i>Group</i>						
	<i>wave</i>	1	2	3	4	5	6	7
Classroom size		30	30	27	27	25	25	26
Average degree	1	3.88	4.63	4.26	4.20	4.04	4.52	4.08
	2	4.34	4.43	4.30	4.15	4.36	4.67	4.84
Missing fraction	1	0.13	0.10	0.00	0.07	0.00	0.00	0.04
	2	0.03	0.00	0.00	0.00	0.00	0.04	0.04
Jaccard coeff.	1–2	0.47	0.48	0.59	0.63	0.62	0.52	0.55

The estimation results of a model assuming the parameter values are identical across the seven classrooms are given in Table 20.2.

Because of high multicollinearity between the outdegree effect and the outdegree - activity ($\sqrt{\nu}$) effect, rather many iterations of the estimation procedure were required to achieve convergence with a satisfactorily low overall maximum convergence ratio (cf. Section 3.6). The parameter estimates are quite usual. The outdegree effect is rather high, but this has to compensate the strongly negative effects of outdegree-popularity and outdegree-activity.

These results were obtained under the assumption that the parameter values are identical for the seven groups, and that is quite a strong assumption. If the assumption does not hold, the results may be unreliable. The next section is about testing this assumption.

20.1.2 Testing the assumption of identical parameter values

The replicate-groups approach is based on the assumption that all groups have the same parameter values – quite a strong requirement. It will be important to test this assumption. Because of the representation of the multiple groups in RSiena as a sequence of periods (see Figure 20.1), this can be done by the same function `sienaTimeTest` that is used also to test time homogeneity across periods for one group. This test is a score-type test (see Section 3.7) which means that it does not require any further estimation.

The null hypothesis tested is that, while the basic rate parameters can be group-dependent, the other eight parameters are the same in each group, as estimated in Table 20.2. The alternative hypothesis is that these parameters also could differ between the groups; this can be represented as group by effect interactions. Since there are seven groups, if the groups would have different parameters, the total number of non-rate

Table 20.2 *Estimation results for seven classrooms under assumptions of complete homogeneity.*

Effect	par.	(s.e.)
Rate classroom 1	5.00	(0.83)
Rate classroom 2	6.72	(1.10)
Rate classroom 3	3.83	(0.67)
Rate classroom 4	3.21	(0.60)
Rate classroom 5	3.30	(0.61)
Rate classroom 6	6.53	(1.24)
Rate classroom 7	5.67	(1.15)
outdegree	0.29	(0.44)
reciprocity	1.40	(0.14)
transitive triplets	0.45	(0.04)
3-cycles	-0.28	(0.07)
indegree - popularity (\surd)	0.06	(0.11)
outdegree - popularity (\surd)	-0.67	(0.15)
outdegree - activity (\surd)	-0.38	(0.09)
sex similarity	0.59	(0.11)

all convergence t ratios < 0.04 .

parameters would be $7 \times 8 = 56$. The number of interaction parameters necessary to describe all these group differences would be this number minus the number of parameters for the null case of identical parameter values, $56 - 8 = 48$. Therefore we could say that we are testing that 48 group-by-effect interaction parameters are zero. The result given by `sienaTimeTest` is $\chi^2 = 77.6, d.f. = 48, p = 0.004$. This indicates that there is evidence for differences between groups.

What to do as a next step, after having found that the assumption for the replicate-groups approach is invalid? One possibility is to go on with the approaches of the next section, which make less stringent assumptions. Another possibility is to see whether small modifications can be made, resulting in assumptions that are more reasonable for this data set. Perhaps the parameter heterogeneity is restricted to only one parameter, and including interactions of group dummies with this parameter leads to an acceptable model. Or perhaps there is only one group different from the rest, and after dropping this groups the remaining groups are reasonably homogenous.

The `sienaTimeTest` results also contain information that is helpful to explore these possibilities. Table 20.3 present results testing the homogeneity for each of the parameters separately. This cuts up the 48 d.f. of the overall test in eight groups of 6 d.f. The null hypothesis is that the parameters for all effects are homogeneous across groups; the alternative hypothesis, that only for this particular parameter the groups are heterogeneous. The table shows that none of the effects are significantly heterogeneous when considered

by themselves. The sum of these chi-squared (d.f. = 6) values is about half of the overall (d.f. = 48) test statistic. This means that the heterogeneity is spread out over all the parameters, and strongly dependent on the correlations between the parameters. This gives little hope for trying to improve the plausibility of the assumptions by adding interaction effects to one or two parameters.

Table 20.3 *Effect-wise tests for homogeneity:chi-squared test statistics with d.f. = 6 and p-values.*

Effect	χ^2	p
outdegree	2.86	.83
reciprocity	3.35	.76
transitive triplets	5.87	.44
3-cycles	8.78	.19
indegree - popularity ($\sqrt{\quad}$)	3.01	.81
outdegree - popularity ($\sqrt{\quad}$)	2.54	.86
outdegree - activity ($\sqrt{\quad}$)	4.85	.56
sex similarity	5.39	.49

Conversely, table 20.4 present homogeneity tests for the seven groups, testing the alternative hypothesis that the other six groups are homogeneous but this single group deviates from the rest. Here we see that the test is significant ($p < 0.05$) for groups 3 and 7, and borderline significant for group 6. This suggests that many of the groups differ from each other, and again does not yield much hope for obtaining plausible assumptions by a small modification or by leaving out one group.

Table 20.4 *Group-wise tests for homogeneity:chi-squared test statistics with d.f. = 8 and p-values.*

Effect	χ^2	p
Group 1	12.90	.12
Group 2	10.38	.24
Group 3	17.30	.03
Group 4	12.94	.11
Group 5	6.64	.58
Group 6	13.62	.09
Group 7	16.61	.03

20.2 Random Coefficient Multilevel Network Models

When a similar dynamics is studied in multiple networks, it often can be meaningful to elaborate this idea of similarity by assuming that the set of networks under consideration all come from the same population of networks — thus in a sense represent the same social mechanisms —, but have different parameter values that determine their evolution — thus may represent the various social mechanisms to different extents. An example could be a study of friendship networks all showing a measure of sex homophily, but where the precise extent of sex homophily is stronger in some networks and weaker in others, depending on the group processes and social context of each group. The assumption that the set of networks studied all come from a common population will in the first place be based on prior conceptual considerations; e.g., all could be different school classes of a similar kind, or sets of stakeholders in similar decision processes but in different institutional units. We could say these are *parallel* rather than replicate networks.

This leads to a *multilevel longitudinal network analysis*, similar to multilevel regression analysis (cf. Raudenbush and Bryk, 2002; Snijders and Bosker, 2012), where the network processes take place at the lower level, the within-network level, while the population of networks constitutes the higher, or global level. This is a random coefficient approach because the parameters for the different networks are assumed to be randomly drawn from a global distribution of parameters. In the approach of this section, this distribution is assumed to be multivariate normal. Let θ be the vector of all parameters for the network dynamics (which also could include parameters for behavioral dependent variables as in Part Two of this book), and let $\theta^{(g)}$ be the value of this parameter vector for the g^{th} network; we denote the total number of parallel networks by G . The assumption made is that the G groups, and therefore the different parameter vectors $\theta^{(g)}$ are independent; and the parameters $\theta^{(g)}$ have a multivariate normal distribution with mean μ and covariance matrix Σ ,

$$\theta^{(g)} \sim \mathcal{N}(\mu, \Sigma). \quad (20.1)$$

If for one of the elements of the parameter, θ_k — which might represent, for example, reciprocity — the variance is 0, i.e., $\sigma_k^2 = 0$ where σ_k^2 is the k^{th} diagonal element of Σ , then this parameter would be the same for all networks, and not variable.

A crucial aspect of multilevel analysis is that the primary focus of the investigations is directed at the global parameters, μ and Σ . *We wish to say something about the population of networks, rather than about all the individual networks.* The individual networks are important because they give us information about the population of networks. After having satisfied our primary curiosity about the population, we may continue to a secondary part of the analysis, and also say something about the groupwise parameters $\theta^{(g)}$ for specific groups.

A great gain that can be achieved in this secondary part as a consequence of the assumption that the various networks all come from the same population of networks, is that this assumption — if it is valid — will strengthen and stabilize the parameter estimation for the individual networks. Non-systematic random deviations in a particular network will be redressed by the knowledge of the common population to which this network belongs. This is done by the application of a Bayesian or empirical Bayesian approach.

In Bayesian statistics (Gelman et al., 2004; Jackman, 2009) the parameters of the statistical model, of which the values are unknown, are assumed to have a probability distribution. In a purely Bayesian approach this distribution does not reflect relative frequencies in some population, but rather the beliefs and uncertainty of the researcher as to the likely values of the parameters. The probability distribution of the parameters reflecting the beliefs of the researcher before having seen the data, is called the *prior distribution*. Observing the data, and applying statistical procedures based on Bayes' rule — a direct application of the rules of conditional probability — transforms the prior into the *posterior distribution*. In an empirical Bayesian approach, on the other hand, the probability distribution of the parameters does reflect relative

frequencies, and it is estimated from data from a set of studies that may be regarded as parallel in some way.

Such an approach is very suitable for the case of multiple networks that are assumed to be drawn from a common population of networks. The statistical procedures integrate the parameter estimation for all networks in the observed set of networks, and estimate simultaneously the parameters characterizing the population and those corresponding to the individual networks. This is done by so-called *Markov chain Monte Carlo* ('MCMC') methods (Gelman et al., 2004). The methods employed for the random-coefficient estimation of multilevel actor-oriented models were developed by Snijders & Koskinen (

current

) and are built on the likelihood-based estimation approaches of Koskinen and Snijders (2007) and Snijders et al. (2010b).

20.2.1 A model with restricted variation

In the multilevel analysis of non-network data the hierarchical linear model is often used (Raudenbush and Bryk, 2002; Snijders and Bosker, 2012). From that type of modeling we know that not all level-1 variables need to be given random slopes. Similarly, for our model not all parameters need to be specified as being randomly differing between the groups. We assume that a subset of the parameters is fixed across groups, and the others are randomly varying. The fixed parameters are denoted by η and the randomly varying by $\theta^{(g)}$. In this way, we can represent the entire parameter vector for group g by

$$\theta^{(g)} = \begin{pmatrix} \theta^{(g1)} \\ \eta \end{pmatrix}. \quad (20.2)$$

We denote the number of parameters in $\theta^{(g1)}$ by p . The basic rate parameters are always assumed to be randomly varying. Some of the coefficients $\beta_k^{(g)}$ of the evaluation function can be fixed (part of η), the others randomly varying (part of $\theta^{(g1)}$). The distributional assumption (20.1) is restricted to the randomly varying parameters,

$$\theta^{(g1)} \sim \mathcal{N}_p(\mu, \Sigma). \quad (20.3)$$

For the rate parameters, which cannot assume negative values, the normal distribution must be truncated at 0.

20.2.2 Bayesian parameter estimation

The parameters in the model are μ , η , and Σ ; where η may or may not exist, depending on whether any of the parameters are supposed to be fixed.

The currently implemented method for the estimation of the global parameters μ , η and Σ is a Bayesian approach. This requires the researcher to specify the *a priori* likely values of the parameters μ and Σ in a prior distribution. For example, if for this type of network evolution the experience is that outdegree parameters of the stochastic actor-oriented model are usually between -3.0 and -1.0 and reciprocity parameters between $+0.9$ and $+2.5$, the prior distributions for these parameters could be assumed to be normal with expected value -2.0 and standard deviation 0.5 for the outdegree parameter, and with expected value $+1.7$ and standard deviation 0.4 for the reciprocity parameter. This represents distributions of which 95 % is positioned in the likely intervals. The data analysis then combines this prior information with the observations, and the posterior distribution represents the likely values of the parameters that results from this combination.

For the parameter η no such prior specification is required. Technically speaking, an 'improper prior

distribution' with constant density¹ is assumed. Informally, this means that a priori we think that the values of η could be just anything, and we specify no preference between different potential values; this can be done because all groups supply information about the value of η . This contrasts with the values of $\theta_k^{(g1)}$, for which information is supplied only by group g ; the global parameters μ and Σ are estimated from a multilevel network data set where the highest-level sample size is G , the number of groups. This implies that for μ and Σ much less information is available; in the Bayesian approach the data information is combined with the prior information.

Formally, in the Bayesian approach, as the prior distribution we assume an inverse Wishart distribution² for Σ , and conditional on Σ for μ a multivariate normal distribution, specified as follows:

$$\Sigma^{-1} \sim \text{wishart}_p(\Lambda_0^{-1}, \nu_0), \quad (20.4a)$$

and conditionally on Σ

$$\mu \mid \Sigma \sim N_p(\mu_0, \Sigma/\kappa_0). \quad (20.4b)$$

This is the so-called conjugate prior distribution given that the group-specific parameters are distributed according to the multivariate normal distribution with parameters μ and Σ ; see, e.g., Gelman et al. (2004), Section 3.6, or O'Hagan and Forster (2004), Chapter 14. Thus, the parameters of the prior are the p -vector μ_0 , the $p \times p$ matrix Λ_0 , the integer ν_0 , and the positive real number κ_0 . The central tendencies of the prior distribution are μ_0 for μ and, approximately, $\nu_0^{-1}\Lambda_0$ for Σ . We denote

$$\Sigma_0 = \nu_0^{-1}\Lambda_0 \quad (20.5)$$

as our 'prior guess' for this parameter. The degrees of freedom ν_0 can be regarded as the strength of the prior evidence for Σ , expressed in terms of equivalent sample size. However, the requirement here is $\nu_0 > p + 1$. Parameter κ_0 expresses (inversely) the prior uncertainty that we have about μ as compared to the variability between the groupwise parameters.

The group-specific parameters $\theta^{(g1)}$ are estimated using the estimated global parameters μ and Σ as the prior distribution. One could say that the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ provides the prior for the $\theta^{(g1)}$. Accordingly, (20.4) may be called the *hyperprior*.

20.2.3 Choice of the prior parameters

The logic of the Bayesian approach requires the researcher to specify the prior knowledge and uncertainty about the parameters in the form of a prior distribution. The influence of the prior distribution will be larger when the data set is smaller. How important is the influence of the prior is currently unknown, and some sensitivity analyses will be helpful to get an impression about this.

The summary of the circus expounded above, with respect to the global parameters, is the following:

- The researcher has to specify for each parameter in the model, whether it is fixed or randomly varying between groups.
- The fixed parameters are collected in the vector η and will be estimated without further prior input.
- For the global population means of the random parameters, we specify our prior knowledge in the vector μ_0 .
- Our 'prior guess' for the between-group variability of the random parameters is given by the covariance matrix Σ_0 ; the degrees of freedom ν_0 reflect the amount of information (number of groups) on which this guess is based.

¹ The term 'improper' reflects that this is not a probability distribution, because the integral of a constant function is infinity.

² The Wishart distribution is a multivariate version of the chi-squared distribution, often used for covariance matrices.

- Our uncertainty about μ , i.e., the variability of its prior distribution, is reflected by the covariance matrix $\kappa^{-1} \Sigma_0$.

It is a bit strange that the matrix Σ_0 occurs in two different ways. This is because of the mathematical properties of conjugate priors (O’Hagan and Forster, 2004) and allows simpler mathematical operations for estimation. For practical purposes this limitation has the pleasant consequence that it also limits the specifications required to be given by the researcher.

The following are some basic guidelines for the choice of the prior.

- (i) For the fixed parameters η , nothing needs to be done.
- (ii) The basic rate parameters are random, but the RSiena software (function `sienaBayes`) has a special data-dependent way to specify their priors, and usually this option can be employed.
- (iii) For the global mean vector μ , we wish to specify as little as possible, so as not to prejudge the results of our analysis. We know that in most cases the outdegree parameter in Stochastic Actor-Oriented Models is negative; therefore, we could specify a prior value $\mu_{0k} = -1$ for this parameter. For many relations we know that the reciprocity parameter is expected to be positive; if the expected value is, e.g., equal to 2, we could specify a prior value $\mu_{0k} = 2$ for this parameter. For other parameters we could just a prior value $\mu_{0k} = 0$ to avoid a positive or negative bias.
- (iv) The between-groups covariance matrix Σ_0 should be chosen not too large, because this determines how strongly the estimation for each of the groups g is helped by the information from the other groups. Given the usual parameter sizes of Stochastic Actor-Oriented Models, a between-groups standard deviation of about 0.1 seems a reasonable guess in most cases; of course, depending on the researcher’s knowledge about the case at hand, this can be modified. Although parameters in Stochastic Actor-Oriented Models are highly correlated, it will usually be too hard to specify prior ideas about this. The resulting matrix Σ_0 accordingly might have the values 0.01 on the diagonal, and off-diagonal values 0. Usually we wish to keep the influence of prior information small, and this leads to choosing the degrees of freedom $\nu_0 = p + 2$, the smallest permitted value. It cannot be stressed strongly enough that if the researcher has other information, perhaps distinguishing between the several parameters instead of treating them all equally as in this proposal, this information should be used for a more adequate specification of Σ_0 . In particular, scales of covariates should be taken into account, when these are not in the range where standard deviations of the covariates are roughly between 0.25 and 2.
- (v) The parameter κ , finally, represents the proportionality between our prior uncertainty about the global mean μ and our prior guess about the between-groups covariance matrix Σ ; on the scale of variances. It is fair to expect that we wish to give our prior guess for the global mean a relatively small weight. The covariance matrix $\kappa^{-1} \Sigma_0$ could, e.g., be the identity matrix, reflecting that parameters in the Stochastic Actor-Oriented Model are scaled normally so as to usually have absolute values less than 1 or 2, so that an uncertainty with prior standard deviation of 1 would be quite considerable already. In the preceding example where Σ_0 has diagonal elements 0.01, this would imply a value $\kappa = 0.01$.

20.2.4 Results from the analysis

The Bayesian MCMC analysis provides a sample from the posterior distribution of the global parameters. It also produces samples from the so-called predictive posterior distributions of the group-specific parameters. All these samples, obtained after a ‘burn-in’ period that should be long ago to let the process stabilize, are not independent but rather they are Markov chains. By inspecting traceline plots of the samples one may assess whether the stability is sufficient. One of the advantages of Bayesian MCMC procedures is that although the prior distributions may be assumed to be normal, the posterior distributions are not restricted in this way and one may make density plots of the posteriors exhibiting their shapes.

Example 20.1 *Multilevel analysis of friendship dynamics in 20 classrooms*

As an example, we use friendship networks in 20 school classes from the study by Knecht (2008) and Knecht et al. (2010). The data collection is described in both of these publications. These are the school classes from Knecht et al. (2010), from which classroom 14a was dropped because it seemed an outlier.

We consider a model for a longitudinal study of the co-evolution of friendship and delinquency with 2 waves, and with $p = 13$ parameters, as shown in Table 20.5.

	Post. mean	Post. s.d.
<i>Network Dynamics: Friendship</i>		
1. rate	5.4460	(0.5302)
2. eval outdegree	-2.1917	(0.2458)
3. eval reciprocity	1.5274	(0.2456)
4. eval transitive triplets	0.4026	(0.1880)
5. eval 3-cycles	-0.3122	(0.1999)
6. eval delinquency alter	-0.0023	(0.2051)
7. eval delinquency ego	-0.1353	(0.2996)
8. eval delinquency ego \times delinq. alter	-0.0917	(0.2428)
9. eval sex similarity	0.7146	(0.2318)
<i>Behavior Dynamics: Delinquency</i>		
10. rate	6.3915	(3.0968)
11. eval linear shape	-0.0843	(0.2178)
12. eval quadratic shape	-0.3272	(0.2117)
13. eval average alter	0.0642	(0.3375)

Table 20.5 *Posterior means and standard deviations for global parameters, Bayesian multilevel analysis of 20 classroom friendship networks from Knecht (2008).*

For the MCMC algorithm, we used:

- (i) groupwise number of MH iterations for sampling micro-steps varies between 75–500 depending on distance between observed networks;
- (ii) 10,000 iterations sampling $\theta^{(g)}, \mu, \Sigma$ for warmup
- (iii) 1,000,000 iterations sampling $\theta^{(g)}, \mu, \Sigma$ for estimation, with a thinning ratio of 1:100.

Two kinds of plot will be given:

- (i) *trace plots*, representing successive draws from the posterior distribution (after *thinning*),
- (ii) *density plots*, representing the plausible values of the parameters, given the observed data.

First we show some trace plots.

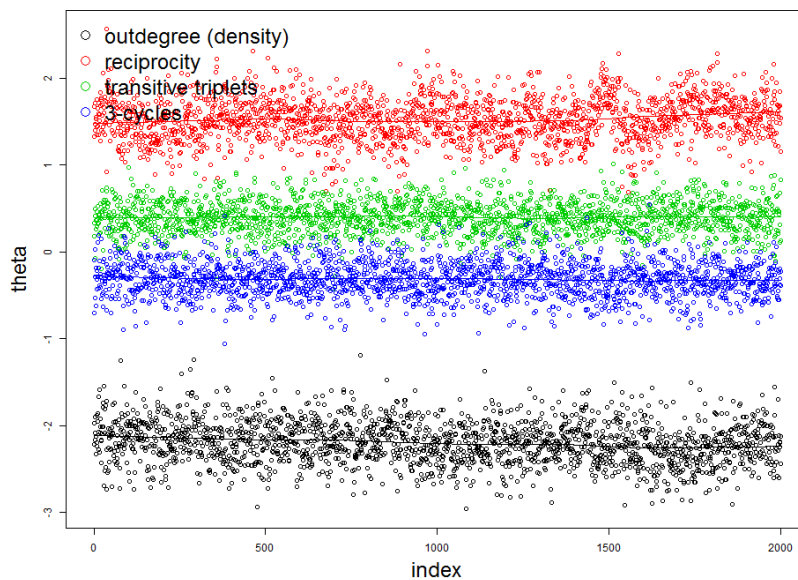


Figure 20.2 Trace plots for global means of structural network effects (evaluation function).

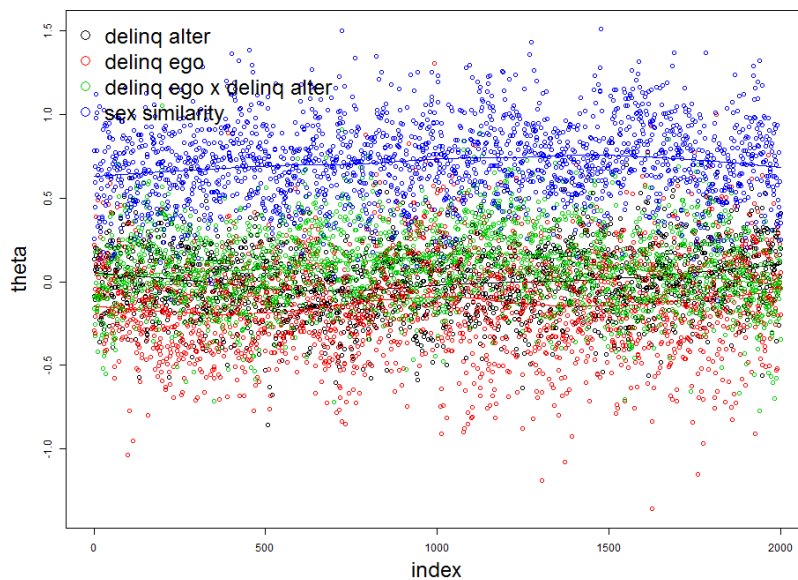


Figure 20.3 Trace plots for global means of covariate-related network effects (evaluation function)

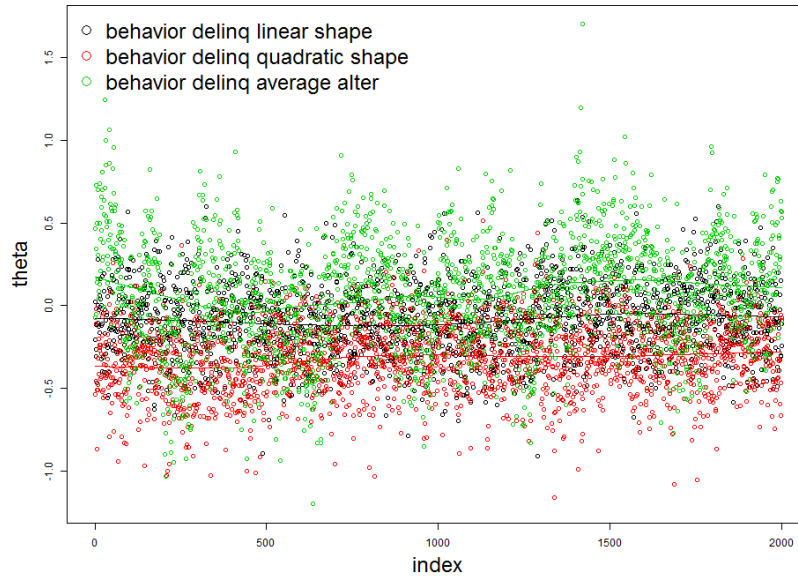


Figure 20.4 Trace plots for global means of behavior effects (evaluation function)

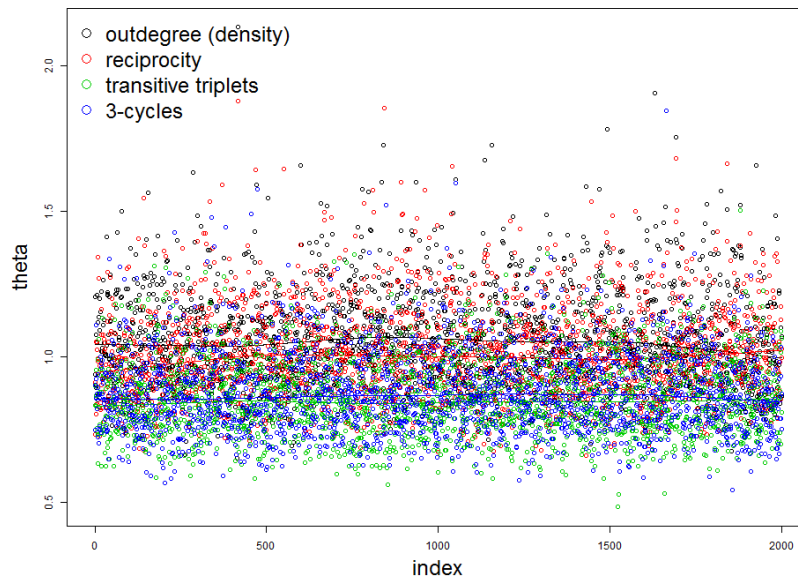


Figure 20.5 Trace plots for global S.D.s of structural network effects (evaluation function)

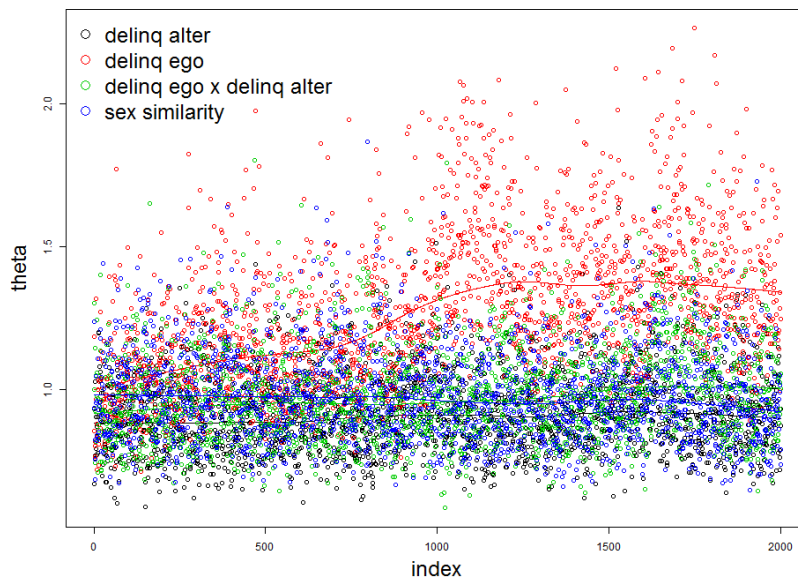


Figure 20.6 Trace plots for global S.D.s of covariate-related network effects (evaluation function)

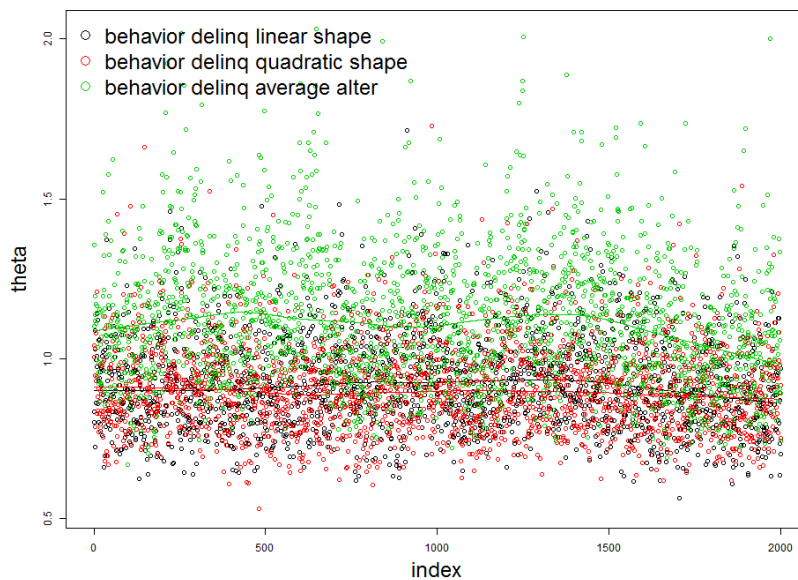


Figure 20.7 Trace plots for global S.D.s of behavior effects (evaluation function)

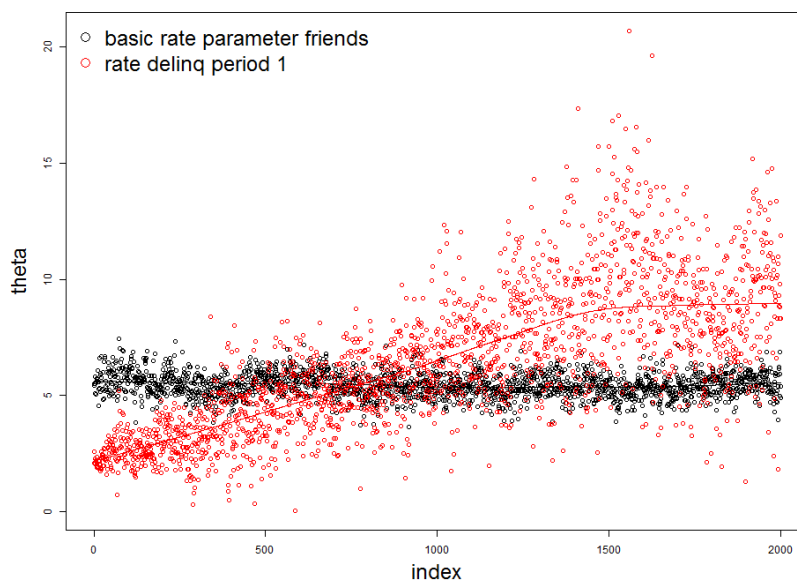
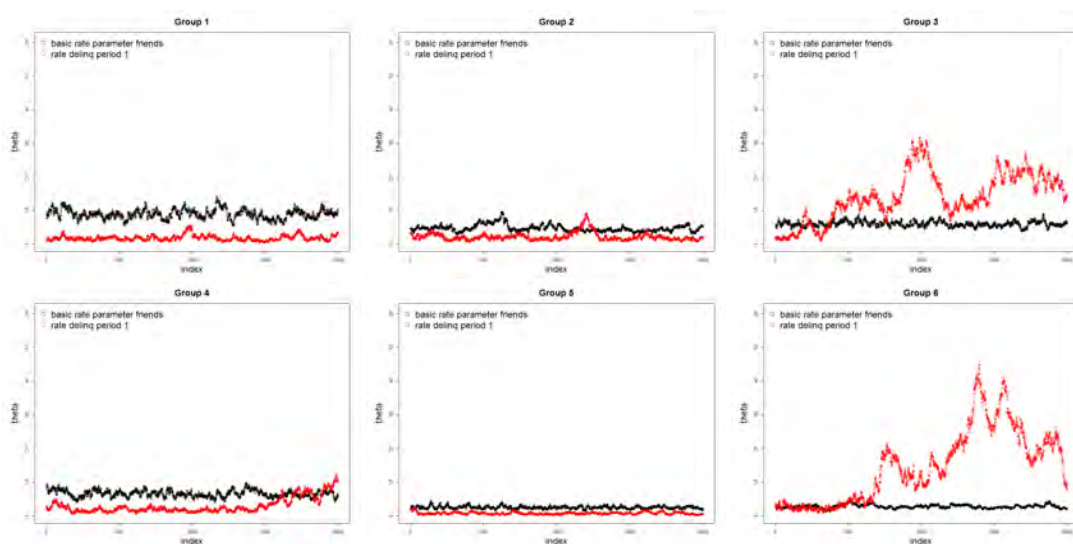


Figure 20.8 Trace plots for rate effects

The trace plot for the global mean of the rate effects, Figure 20.8, and also the trace plot for the standard deviation of the delinquency ego effect, Figure 20.6, show that, in spite of the large number of runs, the rate parameters and the delinquency ego parameters did not stabilize before the first half of the runs. This implies that these results cannot be used as firm conclusions; but they can be used as illustrations. Below we give the trace plots for the rate parameters of the first six groups. It appears that for some of the groups, the rate parameters diverge to values that are so high as to be quite unrealistic.

This shows that further tuning of the algorithm still is required.



Although the results are not satisfactory, as an illustration we present plots of the estimated densities for two parameters: reciprocity and average alter. The plots are distributed over two figures because of the large number of groups. The curve "Average theta" is the estimated distribution of the average group-specific parameter $G^{-1} \sum_g \theta_k^{(g)}$; the curve "Mu" is the distribution of the global parameter μ_k . It is to be expected that the uncertainty about the latter, being the mean of a population of which we have data only for twenty of its elements, is greater than the uncertainty about the former, being the mean of twenty group-specific parameters for groups for each of which we have data.

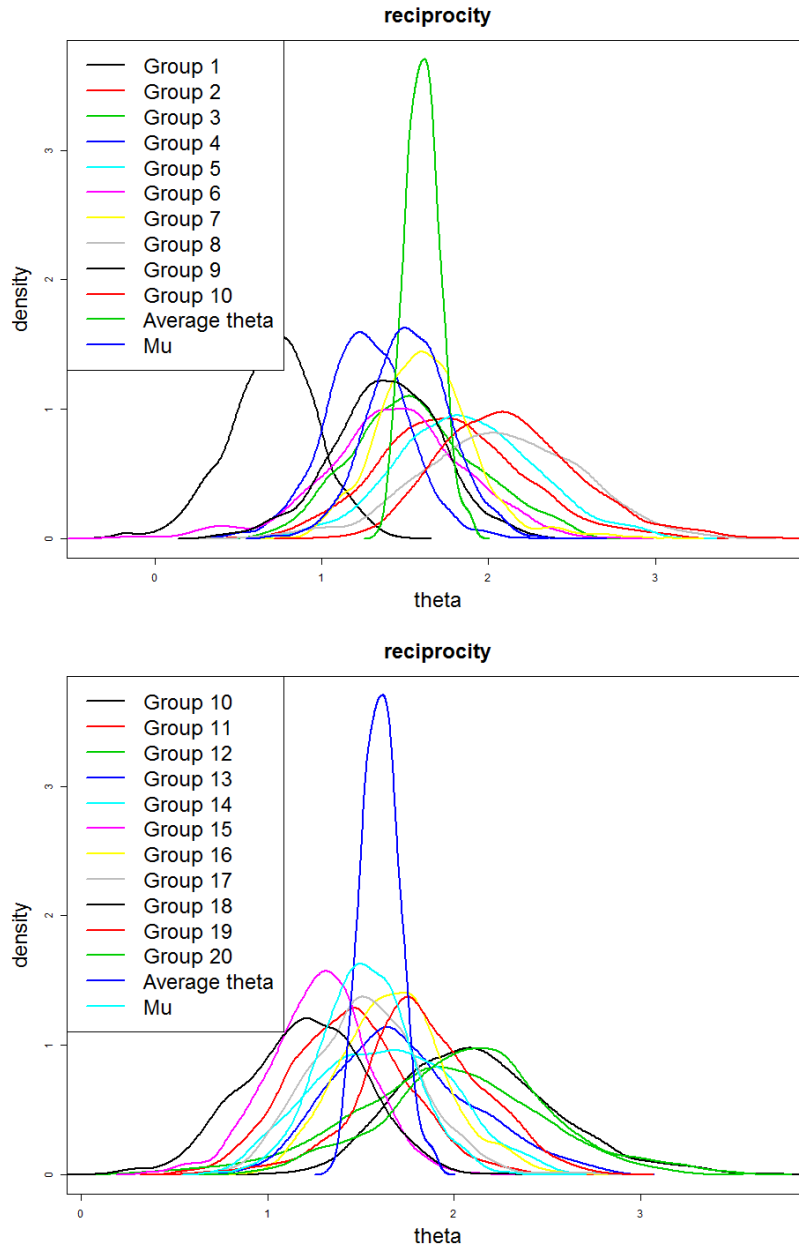


Figure 20.9 Density plots for reciprocity effects

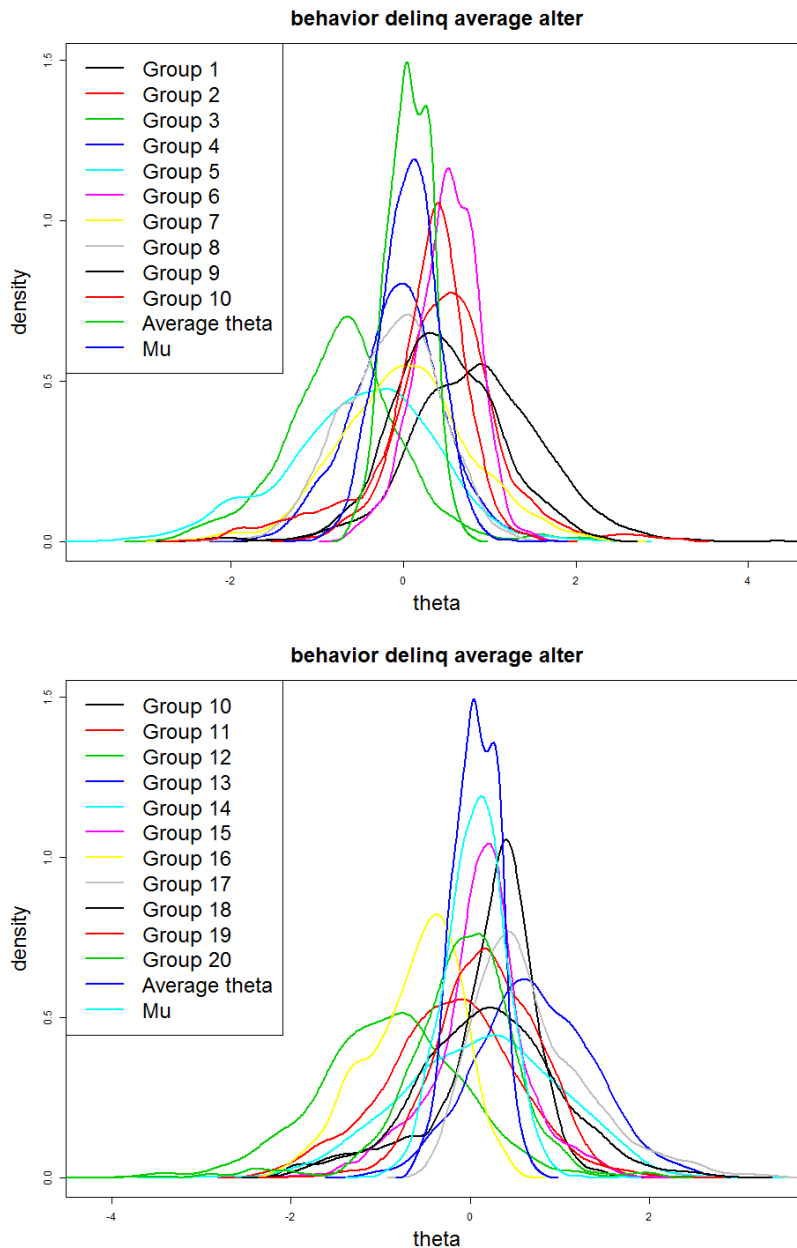


Figure 20.10 Density plots for average alter effects

20.3 Two-stage Meta Analysis in a Population Model

The meta analysis approach has the same basis as the random coefficient approach: all multiple networks are supposed to follow a similar dynamics, i.e., the same model specification but with potentially different parameters. They are *parallel* networks as it was called in the preceding section. Here also it is assumed that the networks come from a population of networks, and we wish to say something primarily about this population. In this two-stage approach, first the parameters for each group must be estimated separately. They

are not simultaneously estimated like in the preceding section. Note that the usual Bayes or empirical Bayes simultaneous estimation procedures for multilevel analysis, of which the preceding section provides an example, have the advantage that the parameter estimation for one group draws strength from the estimation for the other groups; we do not have that advantage here.

The advantage of the two-stage procedure is that we do not need to make the assumption of normal distributions. It is assumed that the parameter vector for the g^{th} group is $\theta^{(g)}$, and these parameter vectors are independent across groups. The population mean of the $\theta^{(g)}$ still is called μ , and their covariance matrix Σ . In the approach now available in RSiena, the meta analysis is done for all parameters separately, and the correlations between the parameters do not receive consideration. Suppose that we are considering one parameter coordinate, with index k , which generally can be denoted θ_k . For example, the parameter could be a homophily parameter for a given actor covariate. This parameter has true value $\theta_k^{(g)}$ in the g^{th} group, for which we have the estimate $\hat{\theta}_k^{(g)}$ and standard error $s_k^{(g)}$. (Note the distinction between ‘true value’ and ‘estimated value’, in line with the general statistical approach that we know the data but we do not precisely know the true parameters.) In the example where θ_k is a homophily parameter, $\theta_k^{(g)}$ is the tendency to homophily in group g ; the average tendency toward homophily on this covariate in the population is μ_k , which denotes the population mean of $\theta_k^{(g)}$; and the variability between groups in the population with respect to this type of homophily is the standard deviation σ_k . Note that σ_k and $s_k^{(g)}$ are totally different things: σ_k says something about the population of networks, viz., how much different networks may differ with respect to this type of homophily, whereas $s_k^{(g)}$ indicates how precisely this parameter was estimated specifically for group g .

The meta analysis approach has the general aim to make conclusions about the population mean μ_k and standard deviation σ_k for any of the parameters θ_k . In the first place, we can pose the question whether there is any effect at all. This question is addressed by the null hypothesis

$$H_0^{\text{overall}} : \theta_k = 0 \text{ everywhere in the population} : \mu_k = 0 \text{ and } \sigma_k = 0. \quad (20.6a)$$

Given the availability of the data for G groups, what we test effectively is whether the effect is nil in all groups under study:

$$H_0^{\text{overall}} : \theta_k^{(1)} = \theta_k^{(2)} = \dots = \theta_k^{(G)} = 0. \quad (20.6b)$$

For this question, there is a test statistic T^2 having a chi-squared null distribution with G degrees of freedom.

If the effect is not generally nil, the question is how it is distributed over the population of networks. For this question we have two ways of tackling it: with or without making the assumption that this distribution is approximately normal. Like is often the case in statistical modeling, the analysis is more precise and also more nicely coherent when based on the normality assumption; but if we make this assumption and it is grossly incorrect, then our results may be misleading.

The normality assumption can be made at two levels: we may assume that the true parameters $\theta_k^{(g)}$ have approximately a normal distribution in the population of networks, and also that the estimation errors $\hat{\theta}_k^{(g)} - \theta_k^{(g)}$ have approximately a normal distribution. What is needed here is a slightly weaker assumption: the estimates $\hat{\theta}_k^{(g)}$ should have a normal distribution with a common mean μ_k and variances given by $\sigma_k^2 + (s_k^{(g)})^2$. In many cases this is likely to be reasonable. Since the estimates are the sum of the true parameter values and the estimation errors, and taking the sum leads towards normality, this will often be a plausible assumption. It may be checked³ by considering the distribution of the estimates $\hat{\theta}_k^{(g)}$, e.g., by looking at the shape of their histogram or a QQ plot. The use of the normality assumption here is less intensive than in section 20.2 because the estimation of the parameters $\theta_k^{(g)}$ is not dependent on this assumption.

Under this assumption, the mean μ_k and the standard deviation σ_k can be estimated by the maximum

³ A formal normality test will have only an approximate nature here, because under the assumptions tested the estimates $\hat{\theta}_k^{(g)}$ do not have constant variance.

likelihood principle, and asymptotic confidence intervals can be constructed for them based on the profile likelihood. The ‘asymptotic’ means that large- G approximations are used for the critical values, and the intervals will be a bit too tight for small G .

If we do not wish to make the assumption of normal distributions, we still can estimate and test the mean as well as the standard deviation, using a method that was pioneered by Cochran (1954) and discussed for dynamic network analysis by Snijders and Baerveldt (2003). This is based on an iterated⁴ weighted least squares method. This provides us with the following statistical tools:

- (i) An estimator for the population mean, $\hat{\mu}_k$.
- (ii) A standard error s.e. ($\hat{\mu}_k$).
Equipped with these two, we can make the usual kind of test and confidence interval for μ_k , based on the t -distribution with G degrees of freedom.
(The use of the t -distribution is based on the approximation of the distribution of $\hat{\mu}_k$ by a normal distribution, which is a weaker assumption than the normality of the true parameters $\theta_k^{(g)}$ as well as the estimation errors $\hat{\theta}_k^{(g)} - \theta_k^{(g)}$.)
- (iii) An estimator $\hat{\sigma}_k$ for the standard deviation.
- (iv) A test for the null hypothesis that $\sigma_k = 0$, i.e., the true parameter for all the groups $\theta_k^{(g)}$ are equal, and therefore are the same as their expected value $= \mu_k$.
This test is carried out by means of a chi-squared test statistic Q with $G - 1$ d.f.

A disadvantage of the iterative weighted least squares method is that results are possible where the estimated standard deviation $\hat{\sigma}_k$ is zero while, in contrast, the test for the standard deviation based on test statistic Q is significant at a usual level of significance such as $\alpha = 0.05$. This paradoxical inconsistency is possible because the test and the estimator in this approach are not directly related (cf. Snijders and Baerveldt, 2003). The likelihood-based method under the normality assumption does not suffer from this problem because the maximum likelihood estimate always is contained in the confidence interval based on the profile likelihood.

In both cases, it is assumed that the standard errors $s_k^{(g)}$ are so precise that we can treat them as if they are the true standard deviations of estimation:

$$E\{\hat{\theta}_k^{(g)}\} = \theta_k^{(g)}, \text{ var}\{\hat{\theta}_k^{(g)}\} = (s_k^{(g)})^2 \text{ in group } g.$$

In both cases, the further assumption is made that the true parameter values

$$\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(G)}$$

and the estimation errors

$$\hat{\theta}_k^{(1)} - \theta_k^{(1)}, \hat{\theta}_k^{(2)} - \theta_k^{(2)}, \dots, \hat{\theta}_k^{(G)} - \theta_k^{(G)}$$

are uncorrelated.

There may be reasons to distrust the estimates which are large with also a large standard error. (This is known as the Donner-Hauck phenomenon in logistic regression, discussed in Section 3.7.1.) Unfortunately, it is impossible to say in general what is to be regarded as a large standard error, as this will depend on the parameter in question. A threshold of 5 for the standard error often is reasonable for most effects; if a tested parameter has a standard error larger than 5, then it is advisable to redo the analysis in a specification where this parameter only is fixed to 0 and a score test is carried out for this parameter. However, for some effects, in any case for the ‘average similarity’ effect for behavior dynamics, parameters and standard errors tend to be larger, and a larger threshold (e.g. 10) is appropriate. The same holds for effects of covariates with small variances (less than .1).

⁴ The cited publications made only two iteration steps, but the current implementation in `siena08C` continues iterating until convergence is reached.

20.3.1 Example

We continue the example of Section 20.1.1. Table 20.6 presents the parameter estimates obtained for all seven groups separately. We see that the differences between the groups are considerable, confirming the

Table 20.6 *Separate estimation results for seven classrooms.*

Effect	Group 1		Group 2		Group 3	
	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
Rate	4.71	(0.78)	6.49	(1.04)	3.59	(0.61)
outdegree	0.78	(1.18)	0.08	(1.21)	3.67	(2.58)
reciprocity	0.61 [†]	(0.35)	1.69***	(0.37)	1.06*	(0.44)
trans. triplets	0.48***	(0.13)	0.25 [†]	(0.14)	0.46***	(0.12)
3-cycles	0.05	(0.23)	0.23	(0.25)	0.20	(0.23)
indegree - pop. ($\sqrt{}$)	-0.30	(0.29)	0.13	(0.33)	0.62*	(0.32)
outdegree - pop. ($\sqrt{}$)	-0.56 [†]	(0.32)	-1.19***	(0.35)	-2.00*	(0.89)
outdegree - act. ($\sqrt{}$)	-0.32	(0.22)	-0.16	(0.25)	-1.16*	(0.49)
sex similarity	0.41	(0.28)	0.90**	(0.31)	0.18	(0.33)

	Group 4		Group 5		Group 6		Group 7	
	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
rate	3.75	(0.78)	3.08	(0.53)	5.89	(1.04)	5.19	(0.92)
o.d.	2.17	(1.81)	-1.22	(1.39)	-1.54	(1.12)	-0.66	(1.20)
rec.	2.00***	(0.59)	0.80 [†]	(0.44)	1.78***	(0.39)	1.61***	(0.47)
t.t.	0.70***	(0.15)	0.45**	(0.16)	0.33**	(0.10)	0.60***	(0.12)
3-c.	-0.49*	(0.24)	-0.47	(0.35)	-0.54*	(0.22)	-0.70**	(0.26)
i.p.	-0.08	(0.41)	-0.36	(0.59)	0.14	(0.39)	0.29	(0.31)
o.p.	-0.87*	(0.39)	0.31	(0.79)	-0.10	(0.56)	-0.45	(0.40)
o.a.	-0.99*	(0.50)	-0.16	(0.26)	-0.11	(0.22)	-0.37	(0.27)
s.sim.	0.05	(0.31)	1.39*	(0.60)	0.93**	(0.33)	0.78*	(0.32)

[†] $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

conclusion of the homogeneity test in Section 20.1.1.

The results of the meta analysis under the assumption of normally distributed parameter estimates are given in Table 20.7, and those without the normality assumption in Table 20.8.

Table 20.7 *Meta-analysis for seven classrooms: with normality assumptions.*

Effect	T^2	(p_{T^2})	$\hat{\mu}_k$	(conf. int.)	$\hat{\sigma}_k$ (conf. int.)
outdegree	6.9	(0.333)	-0.13	(-1.13, 1.09)	0.00 [0, 2.02)
reciprocity	77.2	(<0.001)	1.31	(0.89, 1.77)	0.30 [0, 0.86)
trans. triplets	95.4	(<0.001)	0.46	(0.34, 0.58)	0.05 [0, 0.23)
3-cycles	20.7	(0.002)	-0.23	(-0.55, 0.08)	0.27 [0, 0.64)
indegree - pop. ($\sqrt{\cdot}$)	6.5	(0.372)	0.11	(-0.20, 0.40)	0.03 [0, 0.53)
outdegree - pop. ($\sqrt{\cdot}$)	26.2	(<0.001)	-0.71	(-1.06, -0.32)	0.00 [0, 0.71)
outdegree - act. ($\sqrt{\cdot}$)	14.4	(0.025)	-0.29	(-0.52, -0.09)	0.00 [0, 0.36)
sex similarity	29.9	(<0.001)	0.58	(0.28, 0.92)	0.15 [0, 0.63)

T^2 has d.f. = 7. Confidence intervals calculated for confidence level $1 - \alpha = 0.95$.

Table 20.8 *Meta-analysis for seven classrooms: without normality assumptions.*

Effect	T^2	(p_{T^2})	$\hat{\mu}_k$	(s.e.)	$\hat{\sigma}_k$	Q	(p_Q)
outdegree	6.9	(0.333)	-0.13	(0.51)	0.00	6.8	(0.36)
reciprocity	77.2	(<0.001)	1.31	(0.20)	0.30	10.2	(0.13)
trans. triplets	95.4	(<0.001)	0.46	(0.05)	0.05	8.1	(0.23)
3-cycles	20.7	(0.002)	-0.23	(0.14)	0.27	15.0	(0.02)
indegree - pop. ($\sqrt{\cdot}$)	6.5	(0.372)	0.11	(0.13)	0.03	5.8	(0.44)
outdegree - pop. ($\sqrt{\cdot}$)	26.2	(<0.001)	-0.71	(0.17)	0.00	7.7	(0.28)
outdegree - act. ($\sqrt{\cdot}$)	14.4	(0.025)	-0.29	(0.10)	0.00	6.4	(0.38)
sex similarity	29.9	(<0.001)	0.58	(0.14)	0.15	9.1	(0.17)

T^2 has d.f. = 7, Q has d.f. = 6.

For the outdegree effect and the indegree-popularity effect the overall null hypothesis is not rejected, so as a next step we could estimate a model without the indegree-popularity effect (the outdegree effect should not be removed from the model).

The confidence intervals for the population standard deviation in Table 20.7 all contain the point zero, implying that the standard deviations are not significantly greater than zero. This is in line with the effect-wise homogeneity tests of Table 20.3. The test that does not rely on the normality assumption, using statistic

Q in Table 20.8, on the other hand, signals that for the three-cycles effect the parameters are significantly different across the population. This is, unfortunately, the kind of inconsistency that may arise whenever we apply two or more different statistical tests to the same null hypothesis: they are likely to coincide in most, but not all cases. Recall that in Section 20.1.2 we obtained the result that in a multivariate analysis, consider all eight effects jointly, there is clear evidence for heterogeneity ($p = 0.004$). This illustrates that the approach of the current section, which considers all effects separately, does not tell the whole story.

20.4 Meta Analysis Without a Population Model

Finally we treat a method that makes the least assumptions. No population of networks is assumed. This last inference method considers only the G groups studied, again for one particular parameter, say θ_k , for which the true value in group g is $\theta_k^{(g)}$ and the estimated value $\hat{\theta}_k^{(g)}$. The tested null hypothesis is that the parameter always is zero:

$$H_0^{\text{overall}} : \theta_k^{(1)} = \theta_k^{(2)} = \dots = \theta_k^{(G)} = 0. \quad (20.7)$$

For example, suppose θ_k is a parameter expressing social influence, then we wish to test whether there is evidence for social influence in any group. Answering this question by checking whether the parameter differs significantly from 0 in any group would not be a valid approach, because it ignores the problem of multiple testing: the type-one error probability for this approach would be the probability, if the overall null hypothesis is true, of finding a significant result in any group, which will be much higher than the nominal significance level α .

To achieve a combined statement about the significance of the parameter in any group, we can use Fisher's method for combining independent p -values; this is a technique in meta analysis, proposed by Fisher (1932) and described in texts about meta analysis such as Hedges and Olkin (1985) and also (briefly) in Snijders and Bosker (2012, Chapter 3). The procedure operates as follows: for each group we obtain a p -value; denote this by p_g . Intuitively, some very low p -values, or a lot of rather low p -values, would give combined evidence against the null hypothesis. It turns out that a good combination is obtained by adding minus twice the logarithms:

$$X^2 = -2 \sum_{g=1}^G \ln(p_g).$$

(Recall that the logarithm of a number between 0 and 1 is negative, so that X^2 always is positive.) This can be tested in a chi-squared distribution on $2G$ degrees of freedom.

In many cases we are interested in whether there is evidence specifically for a positive effect, or specifically for a negative effect. In such cases Fisher's combination method can be applied simultaneously for a right-sided and a left-sided test. We then have two null hypotheses and two alternative hypotheses:

- (i) for detecting if there are any networks with a positive parameter value, the null hypothesis tested is $H_0^+ : \theta_k^{(g)} \leq 0$ for all networks $g = 1, \dots, G$;
with the alternative hypothesis;
 $H_1^+ : \theta_k^{(g)} > 0$ for at least one network g .
- (ii) for detecting if there are any networks with a negative parameter value, the null hypothesis tested is $H_0^- : \theta_k^{(g)} \geq 0$ for all networks $g = 1, \dots, G$;
with the alternative hypothesis
 $H_1^- : \theta_k^{(g)} < 0$ for at least one network g .

For both of these one-sided testing problems, we can then obtain the p -value; the first based on combining right-sided groupwise p -values, the second based on combining left-sided groupwise p -values.

It is advisable to use for both the significance level of $\alpha/2$ (e.g., 0.025 if $\alpha = 0.05$) which yields an overall combined test at significance level α . Note that four different overall results are possible. Indicating the right-sided and the left-sided p -values by p_r and p_l , respectively, these possible results are:

- (a) $p_r > \alpha/2$, $p_l > \alpha/2$:
no evidence for any nonzero parameter values;
- (b) $p_r \leq \alpha/2$, $p_l > \alpha/2$:
evidence that some networks have a positive parameter value, no evidence for any negative parameter values;
- (c) $p_r > \alpha/2$, $p_l \leq \alpha/2$:
evidence that some networks have a negative parameter value, no evidence for any positive parameter values;
- (d) $p_r \leq \alpha/2$, $p_l \leq \alpha/2$:
evidence that some networks have a negative parameter value, and some others have a positive parameter value.

If all networks have a zero true parameter value, i.e., under the overall null hypothesis that $\theta_n^{(g)} = 0$ for all g , the probability of result (a) is less than or equal to α ; this is the way in which this combined test respects the overall probability of an error of the first kind.

20.4.1 Example

We continue the example of Sections 20.1.1 and 20.3.1 and combine the one-sided tests for the seven groups by Fisher's method. The results are in Table 20.9. The results are clear: for outdegree and indegree-

Table 20.9 *Fisher's combination of right-sided and left-sided p -values for testing the null hypothesis that the mentioned effect is nil in all seven groups.*

Effect	p (right-sided)	p (left-sided)
outdegree	0.381	0.526
reciprocity	<0.001	1.000
trans. triplets	<0.001	1.000
3-cycles	0.851	0.001
indegree - pop. ($\sqrt{\quad}$)	0.256	0.733
outdegree - pop. ($\sqrt{\quad}$)	0.997	<0.001
outdegree - act. ($\sqrt{\quad}$)	1.000	0.002
sex similarity	<0.001	1.000

popularity, there is no evidence that any of the groups have a non-zero parameter. For reciprocity, transitive triplets, and sex similarity there is evidence that at least for some of the groups this parameter is positive, and there is no evidence for any groups with a negative parameter. For three-cycles, outdegree-popularity, and outdegree-activity it is the converse: there is evidence for at least some negative values, and no evidence for any positive values.

Note the distinction from the results about the population means of the parameters that can be drawn from Tables 20.7 and 20.8. For the three-cycles effect, the groups give evidence for at least some negative and not for any positive values, but the confidence interval for the population mean covers the value 0, so the data still is consonant with a positive population mean; in that case it is likely that there are both positive and negative three-cycle parameters in the population. This is further explored in the ‘funnel plot’ of Figure 20.11, which shows a plot of the parameter estimates of the three-cycles effect and their standard errors, with lines (forming a funnel) demarcating the region of significance at $\alpha = 0.05$. The four groups inside the

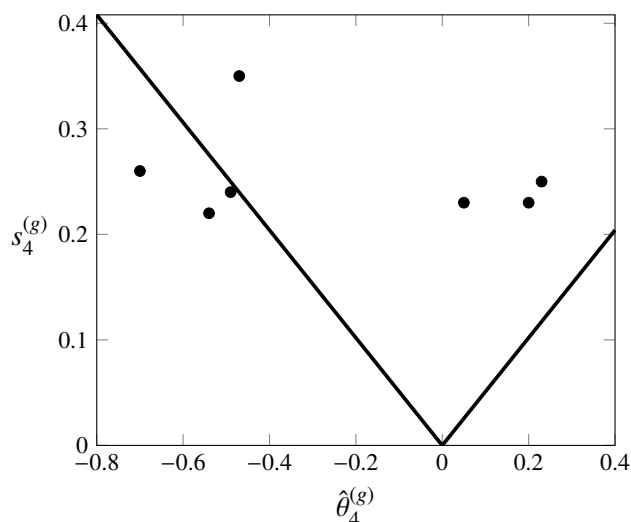


Figure 20.11 Estimates of three-cycle parameters plotted against their standard errors, with lines demarcating the region of significance ($\alpha = 0.05$).

funnel have individually non-significant parameters, the three groups outside the funnel have significantly negative parameters. This figure supports the conclusion that there is evidence for some negative parameter values, but if a population of networks is assumed, the mean three-cycles effect in this population still might be negative, zero, or positive.

20.5 Contrasting the multiple-group approaches to each other

In this chapter we treated three conceptually different ways for obtaining a combined analysis of several groups, of which one came in two varieties.

Replicate groups. First, an analysis was discussed based on the assumption that the parameter values are the same in all groups. For this purpose the data must be combined using the function `sienaGroupCreate` in `RSiena`.

The assumption made here is very strong, and if it does not hold the results are likely to be misleading; in particular, the standard errors are likely to be too small. This approach should be used only if on prior grounds it is reasonable to make this assumption, and if a homogeneity test (which can be executed using `sienaTimeTest`) does not signal that the assumption is violated. This means, in practice, it should almost never be used.

Networks are sampled from a population. Second, the assumption was treated that the networks all have the same model specification, but the parameter values differ across networks, with the auxiliary assumption that the networks are a sample from a population and the aim of the statistical inference is to draw

conclusions about the population.

This approach was elaborated in two different ways.

Integrated approach. In an Bayesian approach, with the further assumption of normally distributed parameters, the groupwise parameters can be estimated simultaneously with the population parameters, using the RSiena function `sienaBayes`.

Doing the estimation simultaneously implies that the estimation for each group can borrow strength from the other groups, like in other multilevel models, so that convergence difficulties of single groups may be less problematic. A practical disadvantage (which may be overcome in the future) is that the analysis, based on likelihoods, is time-consuming.

Two-stage approach. Another option here is first to estimate the model for each of the groups separately, and then to combine the results in a population-based kind of meta-analysis, for each parameter coordinate separately, using the RSiena function `siena08`.

The groupwise parameters here are estimated independently. This has the advantage of greater robustness and the disadvantage that the groups do not borrow strength from each other for the estimation of groupwise parameters.

Here again there are two variants. A more coherent approach is possible under the assumption that the estimated parameters have approximately normal distributions. The alternative is what may be called a semi-parametric approach, where no assumptions are made about the distribution in the population but the statistical inference still focuses on the mean and variance of the parameter in the population of networks.

Combination without population. Another way of synthesizing the information from multiple groups is by testing the combined hypothesis whether a parameter coordinate is non-zero (or positive, or negative) in *at least one* of the groups studied. This again is a two-step approach, where first the parameters are estimated for all groups separately, and then the results are combined using Fisher's method for combining independent p -values. This also can be done by the RSiena function `siena08`.

Since this approach involves only the inferential step from the observed data to the true parameters of the groups, not the inference from the observed set of groups to the population, this approach may yield significant results more easily than the population-based approaches. This difference is not a matter of power, but of answering a different question, as is further explained below.

To understand the contrast between the method following a population-based (random coefficient) approach and the Fisher approach for combining independent tests, the following may be helpful. Inferring about a population always adds some uncertainty; this is more serious when the sample size (here: number of combined networks) is smaller. In an extreme case, consider the combination of $G = 2$ networks, with estimates for the first group $\hat{\theta}_k^{(1)} = 1$, standard error $s_k^{(1)} = 0.1$, and for the second group $\hat{\theta}_k^{(2)} = 5$, $s_k^{(2)} = 0.1$. Then for both of the groups the t -statistic $\hat{\theta}_k^{(g)} / s_k^{(g)}$ is very large, leading to the conclusion that parameters $\theta_k^{(1)}$ and $\theta_k^{(2)}$ are very likely to be positive. This will lead to a significant result for Fisher's combination of tests. On the other hand, the mean in the population of networks, given that there is available a sample of size as low as $G = 2$, cannot be determined with any degree of precision, so the confidence interval for this mean μ_k must be huge, and the result for testing the null hypothesis that the mean in the *population of networks* is zero cannot be significant. However, the results for testing the null hypothesis that the effect is zero in *all observed networks* will be significant.