Random Effects Models for Personal Networks

An Application to Marital Status Homogeneity

Jeroen K. Vermunt and Matthijs Kalmijn

Tilburg University, the Netherlands

Abstract. We propose analyzing personal or ego-centered network data by means of two-level generalized linear models. The approach is illustrated with an example in which we assess whether personal networks are homogenous with respect to marital status after controlling for age homogeneity. In this example, the outcome variable is a bivariate categorical response variable (alter's marital status and age category). We apply both factor-analytic parametric and latent-class-based nonparametric random effects models and compare the results obtained with the two approaches. The proposed models can be estimated with the Latent GOLD program for latent class analysis.

Keywords: multilevel analysis, latent class analysis, random effects models, personal networks, ego-centered networks, quasi-symmetry models, conditional logit models

Introduction

One of the main types of network data is personal network data, sometimes referred to as ego-centered network data. Since personal network data can be collected by means of survey methods, sample sizes are typically larger than for data sets on complete networks. The information obtained is, however, more limited: Rather than providing a full reconstruction of one or more complete networks, ego-centered network data sets list the network members (alters) for each respondent (ego) participating in the survey. In addition, one may have information on alter and tie characteristics, such as demographics and the type of relationship between ego and alter. In some cases (partial) information is also available on the presence of ties among the alters (the alter-alter dyads) in the personal networks (Gerich & Lehner, 2006).

Many research questions to be answered using ego-centered network data require no more than simple descriptive analyses, such as the computation of the average number of alters (or ties) with a certain characteristic (e.g., average number of married alters) or the distribution of alters (or relationships) across categories (e.g., marital status distribution), possibly conditional on characteristics of the ego (e.g., separately for men and women). For slightly more advanced (multivariate) research questions, one will, however, need to switch to regression analysis techniques.

In a regression analysis of personal network data, the dependent variable will usually be a characteristic of either the alter or the ego-alter dyad, whereas predictors can be characteristics of the ego, the alter, or the dyad. Because each ego has multiple alters—that is, alters are nested within egos—the data have the form of a two-level data set. A natural way to analyze personal network data is therefore to make use of multilevel or random-coefficients regression techniques (Van Duijn, Van Busschbach, & Snijders, 1999). Note that in order to have a two-level data structure, there should be (negligible) overlap between the personal networks of the egos involved in the study. This is, however, a realistic assumption in survey research based on representative samples. It should also be noted that our approach uses only ego-alter dyads, and thus does not make use of alter-alter dyads' information if available. In this article, we propose two alternative approaches for dealing with data on personal networks. These are (a) standard parametric random effects models that assume random coefficients to come from a (restricted) multivariate normal distribution and (b) nonparametric random effects models based on latent class methods. The latter approach has several advantages: It is more practical with categorical outcome variables, it makes no assumptions about the distribution of the random effects across egos, and it yields a clustering of egos based on dependencies between alters as a by-product.

The random-coefficients models for ego-centered network data are illustrated with an application on marital status homogeneity of network ties, controlling for their age homogeneity. To facilitate the introduction of random effects, we transform the well-known quasi-symmetry log-linear model for studying homogeneity in squared frequency tables into a conditional logit model.

The organization of the article is as follows. The next section introduces the application and the data set that is at our disposal. Then we present the two types of random-coefficients models and describe the results obtained in our application. The article ends with some final remarks.

DOI 10.1027/1614-1881.2.1.34

Description of Data Set and Motivation of Application

Background of the Study

The literature on the selection of marriage partners and the literature on social networks have both shown that personal relationships are homogeneous with respect to various social and cultural characteristics (Kalmijn, 1998; Lazarsfeld & Merton, 1954; Marsden, 1988, 1990; Miller McPherson, Smith-Lovin, & Cook, 2001). Examples are class and educational homogeneity, religious and ethnic homogeneity, and homogeneity by age. An important feature of the selection of friends or marriage partners is that people have to consider multiple characteristics simultaneously. Because traits are correlated within persons, a choice for a given characteristic in a friend or potential spouse often implies a choice for another characteristic as well. If someone looks for a friend who is highly educated, for example, the chances are good that the friend he or she will find is also relatively rich. Similarly, if someone prefers to marry someone who shares his or her national background, the chances are high that they will also be of the same religion.

An important question this raises is whether the homogeneity found in reality is based on explicit selection on that trait, or whether it is a by-product of selection on another trait. To establish that there is direct selection, one would need to show that the degree of similarity with respect to a certain trait is greater than one would expect from the similarity that exists in another trait. Because the problem is symmetric, this needs to be established the other way around as well, and hence, the traits need to be analyzed simultaneously.

Age and marital status are closely related aspects of the life course, with age being the gradual component of the life course and marital status being the discrete and transitional component. Several studies have shown that personal networks tend to be homogeneous by age (Burt, 1991; Louch, 2000; Marsden, 1988; Miller McPherson et al., 2001). Moreover, in his classic American study on personal networks, Fischer (1982, pp. 180–181) demonstrated that networks are homogenous with respect to marital status: Married respondents tend to name more often married associates, those that never married more often never married, and divorced more often divorced. However, because Fischer did not correct for age homogeneity, we do not know whether the encountered marital status homogeneity is spurious (a by-product of age homogeneity) or caused by selection of network members based on their marital status.

The research goal within the application is threefold. First, we want to describe the degree of age and marital status homogeneity in personal networks in our data set. Second, we want to assess to what extent marital status homogeneity can be attributed to age homogeneity and vice versa. Third, by using a multilevel framework, we want to take into account explicitly the dependencies between the multiple alters of a respondent.

Description of Data Set

The ego-centered network data we analyze come from a survey based on face-to-face interviews with a random national sample of 902 individuals in the Netherlands (Fiselier, Vander Poel, & Felling, 1987). Network members were identified through a mix of the contact, exchange, and role methods (Broese van Groenou & Van Tilburg, 1996). For example, respondents had to list the people with whom they regularly went out (contact method) and the persons who had helped (or could have helped) with odd jobs around the house (support method). Respondents also had to list certain role relationships (e.g., the spouse, the children). The respondent was not only asked about actual support given and received (which is heavily dependent on needs), but also about *potential* support (i.e., persons from whom support was or could be expected and persons who could have asked for support). For each network member, several pieces of personal information were collected, including age and marital status. The marital status categories are (a) single and never married, (b) married or cohabiting, (c) divorced, and (d) widowed. Unfortunately, we do not have information on whether the alter has children living at home. We should therefore emphasize that the status of being married often combines the effects of having a partner and the effects of having children. Five age categories are distinguished: (a) under 30, (b) 30–39, (c) 40–49, (d) 50–59, and (e) 60 and over. By categorizing age, we are able to study age homogeneity and marital status homogeneity in a single categorical data analysis framework. Although the choice of the specific age categories is always somewhat arbitrary, these age categories are meaningful in the sense that they represent important life stages. As shown in other studies, conclusions on homogeneity are robust for how the specific (age) categories are formed (Van Poppel, Liefbroer, Vermunt, & Smeenk, 2001). The age range in the sample of respondents is 20–72.

In the analyses, some types of relationships were excluded because marital status or age differences are theoretically of a different order. More specifically, we excluded alters who are partners because this would lead to an overestimate of the degree of similarity by marital status. We excluded family relationships where age differences are extreme for reasons that have little do with choice (i.e., parents, children, parents-in-law, children-in-law, grandparents, and grandchildren). After these selections are made, the number of respondents is 875 (7,896 relationships).

Random Effects Conditional Logit Models

A Conditional Logit Multivariate Quasi-Symmetry Model

Let Z_j^M and Z_j^A denote the marital status and age of ego j and let Y_{ij}^M and Y_{ij}^A denote the marital status and age of alter i of ego j. A particular marital status will be denoted by r

and p, for egos and alters, respectively, and a particular age category by s and q. Let us first concentrate on the marital status variables. A well-established approach for studying homogeneity of pairs of actors with respect to a categorical outcome variable is the use of the log-linear quasi-symmetry model (Hout & Goldstein, 1994; Kalmijn, 1991; Uunk, Ganzeboom, & Róbert, 1996). This model has the following form:

$$\log[P(Y_{ij}^{M} = p, Z_{j}^{M} = r)] = \alpha_0 + \alpha_r^{M} + \beta_p^{M} - 0.5\beta_{pr}^{MM},$$

$$\text{with } \beta_{pr}^{MM} = \beta_{rp}^{MM} \text{ if } p \neq r, \text{ and } \beta_{pr}^{MM} = 0 \text{ otherwise.} \quad (1)$$

In this equation, α_0 represents a normalizing constant, α_r^{M} the main effect of the ego's status, β_p^{M} the main effect of the alter's status, and β_{pr}^{MM} the association parameter capturing the dependence between ego's and alter's status categories. By imposing a symmetric structure on the association parameters ($\beta_{pr}^{MM} = \beta_{rp}^{MM}$) one obtains what is known as the log-linear quasi-symmetry model (Agresti, 2002). The chosen parameterization in which β_{pr}^{MM} is multiplied by -0.5 and equals 0 for the diagonal elements yields association parameters with a very simple and useful interpretation; that is, $\exp(\beta_{pr}^{MM})$ is the odds ratio in the two-way table formed by egos' and alters' marital status categories p and r. For example, the coefficient for the single-married combination ($p = 1$ and $r = 2$ or $p = 2$ and $r = 1$) compares the odds of having a single rather than a married alter for single and married egos. A value for $\exp(\beta_{pr}^{MM})$ that is larger than 1 indicates that among p and r egos it is more likely (than can be expected based on the marginal distributions) that they will choose alters with the same marital status. Worded differently, large positive β_{pr}^{MM} values indicate a strong boundary between the two categories concerned and small positive values a weak boundary.

In Equation 1, the quasi-symmetry model was specified as a restricted log-linear model for the joint distribution of ego's (Z_j^{M}) and alter's (Y_{ij}^{M}) marital status. It can, however, also be specified as a model for the conditional distribution of alter's status given ego's status, a formulation that will simplify the various extensions discussed below. This yields the following logistic regression equation:

$$P(Y_{ij}^{M} = p|Z_j^{M} = r)] = \frac{\exp(\beta_p^{M} - 0.5\beta_{pr}^{MM})}{\sum_{p'=1}^{4} \exp(\beta_{p'}^{M} - 0.5\beta_{p'r}^{MM})}. \quad (2)$$

As can be seen, the α_0 and α_r^{M} terms cancel from the equation because they do not depend on alter's status. Moreover, the constraints on and the interpretation of the β_p^{M} and β_{pr}^{MM} parameters remain exactly the same. Note that the model described in Equation 2 is not a standard multinomial logit model but a conditional logit model (McFadden, 1974) because parameters are constrained across categories of the dependent variable Y_{ij}^{M}.

The model described in Equation 2 does not take into account the mutual dependence between marital status and age homogeneity. To study the impact of age homogeneity on status homogeneity, we have to analyze the marital status and age variables simultaneously by means of a multivariate variant of the quasi-symmetry model. Using again the logistic regression form, we obtain the following conditional logit model for the joint distribution of alter's marital status and age given ego's marital status and age:

$$P(Y_{ij}^{M} {}^{T} p, Y_{ij}^{A} = q|Z_j^{M} = r, Z_j^{A} = s) \quad (3)$$
$$= \frac{\exp(\beta_p^{M} + \beta_q^{A} + \beta_{pq}^{MA} - 0.5\beta_{pr}^{MM} - 0.5\beta_{qs}^{AA})}{\sum_{p'=1}^{4}\sum_{q'=1}^{5} \exp(\beta_{p'}^{M} + \beta_{q'}^{A} + \beta_{p'q'}^{MA} - 0.5\beta_{p'r}^{MM} - 0.5\beta_{q's}^{AA})}.$$

Here, β_p^{M} and β_q^{A} are the intercepts corresponding to the two dependent variables and β_{pq}^{MA} captures their mutual dependency. For identification, we use the well-known effects-coding constraints $\Sigma_p \beta_p^{M} = \Sigma_q \beta_q^{A} = \Sigma_p \beta_{pq}^{MA} = \Sigma_q \beta_{pq}^{MA} = 0$. The other two terms—$\beta_{pr}^{MM}$ and β_{qs}^{AA}—describe the association between alter's and ego's marital statuses and ages, respectively, and are restricted to have the symmetric association structure that was already introduced above for β_{pr}^{MM}.

The parameters of main interest are the symmetric association parameters β_{pr}^{MM}, denoting the strength of the relationship between ego's and alter's marital status. The effect of age homogeneity on marital status homogeneity can be determined by comparing the results obtained with Equation 2 with the ones of a model in which the β_{pq}^{MA} terms are omitted or, equivalently, in which $\beta_{pq}^{MA} = 0$. Note that with this set of constraints, we obtain the same estimates for β_p^{M} and β_{pr}^{MM} as are obtained with the simpler model described in Equation 2 in which the age variables are fully omitted.

Modeling Dependencies Between Alters Using Random Effects

So far, we have ignored the fact that the multiple observations within individuals (the characteristics of the various alters within egos) cannot be assumed to be independent of each other, even after controlling for ego's characteristics. More specifically, alters may resemble each other in age and martial status in a way that is not completely explained by ego's own age and marital status. It is well known that standard errors are biased (usually downward) when dependencies between observations are not taken into account, which yields incorrect tests. In nonlinear regression models such as our conditional logit model, parameter estimates may also be biased, usually downward (Agresti, 2002). In other words, ignoring dependencies may seriously distort the results. Important to recognize, however, is that the dependencies between the alters' characteristics are not just a methodological problem: they also contain relevant information on individual differences with respect to the structure of their networks, that is, on the (unobserved) heterogeneity of preferences and opportunities. Ignoring this information would be a loss.

Yamaguchi (1990) proposed modeling and describing dependencies between alters' characteristics in friendship networks using a restricted log-linear model for a table cross-tabulating the ego's status with the combination of statuses of all ego's alters. When there are c status categories and n alters, the model is estimated for a c^{n+1} table. The associations between the alters' statuses are captured by two-variable log-linear association terms that can be assumed to be the same for each pair of alters. Despite the fact that this approach is elegant, conceptually simple, and fits very well within the log-linear modeling framework introduced above, it is not practical with more than a few alters per ego. In our data set, for example, the largest personal network consists of 31 alters, which means that—given that we deal with two characteristics simultaneously—we would have to set up a log-linear model for a frequency table consisting of $(4 \times 5)^{32}$ cells, which is, of course, impossible.

An alternative approach for dealing with dependent observations involves introducing random effects. Van Duijn et al. (1999) proposed using linear regression models with random effects for the analysis of personal networks with tie information—in their example, change in contact frequency—that can be treated as a continuous outcome variable. In our application, the tie outcome is clearly not a continuous variable, which implies that we cannot apply such a standard hierarchical linear model. What is needed is a random effects variant of the conditional logit model described in Equation 3, that is, a model in which (at least) the main effect parameters β_p^M and β_q^A are specified to be random effects; that is,

$$P(Y_{ij}^M = p, Y_{ij}^A = q | Z_j^M = r, Z_j^A = s) = \frac{\exp(\beta_{pj}^M + \beta_{qj}^A + \beta_{pq}^{MA} - 0.5\beta_{pr}^{MM} - 0.5\beta_{qs}^{AA})}{\sum_{p'=1}^{4}\sum_{q'=1}^{5}\exp(\beta_{p'j}^M + \beta_{q'j}^A + \beta_{p'q}^{MA} - 0.5\beta_{p'r}^{MM} - 0.5\beta_{q's}^{AA})}. \tag{4}$$

The fact that the two main effects now contain an index j indicates these parameters may vary across egos. In total, this model contains 7 (= 3 + 4) random effects. Under the standard assumption that random effects come from an unrestricted multivariate normal distribution, the log-likelihood function will thus contain a seven-dimensional integral. Because this integral cannot be solved analytically but only approximated by numerical or Monte Carlo integration methods, maximum likelihood estimation of the above random effects conditional logit model is very computationally intensive. Another problem is that the interpretation of the parameters associated with the random effects may become difficult with so many random effects. A possible way out of these two problems is to make use of more parsimonious factor-analytic structures for the random effects, as described by Skrondal and Rabe-Hesketh (2004) and used by Hedeker (2003) and Vermunt (2005) in the context of random effects multinomial logit models. The following factor-analytic structures may be of interest for restricting the term $\beta_{pj}^M + \beta_{qj}^A$: a one-factor model

$$\beta_p^M + \beta_q^A + \lambda_{p1}^M F_{j1} + \lambda_{q1}^A F_{j1}, \tag{5}$$

a simple structure two-factor model

$$\beta_p^M + \beta_q^A + \lambda_{p1}^M F_{j1} + \lambda_{q2}^A F_{j2}, \tag{6}$$

with either correlated $(\sigma_{F_1F_2} \neq 0)$ or uncorrelated $(\sigma_{F_1F_2} = 0)$ factors, or a full two-factor model

$$\beta_p^M + \beta_q^A + \lambda_{p1}^M F_{j1} + \lambda_{q1}^A F_{j1} + \lambda_{p2}^M F_{j2} + \lambda_{q2}^A F_{j2}, \tag{7}$$

with uncorrelated factors $(\sigma_{F_1F_2} = 0)$. In each of these specifications, F denotes a normally distributed continuous factor with a variance equal to 1 $(\sigma_{F_1}^2 = \sigma_{F_2}^2 = 1)$ and λ are factor loadings, which for identification purposes are assumed to sum to 0 across categories

$(\Sigma_p \lambda_{p1}^M = \Sigma_p \lambda_{p2}^M = \Sigma_q \lambda_{q1}^A = \Sigma_p \lambda_{q1}^A = 0)$.

An alternative approach that does not have the computational and conceptual difficulties associated with the parametric random effects conditional logit model is to use a nonparametric specification for the random effects in which individuals are assumed to belong to one of T latent classes that differ with respect to the model parameters of interest (Skrondal & Rabe-Hesketh, 2004; Vermunt & Van Dijk, 2001). In our application, this yields a model that is called a latent class or mixture conditional logit model (Kamakura, Wedel, & Agrawal, 1994). As pointed out by Aitkin (1999), such a latent-class-based random effects approach is not only more practical, it is also much less restrictive than the standard approach in the sense that no arbitrary a priori assumptions need to be made about the distribution of the random effects.

The relevant latent class variant of the conditional logit described in Equation 3 has the following form:

$$P(Y_{ij}^M = p, Y_{ij}^A = q | Z_j^M = r, Z_j^A = s, X_j = t) = \frac{\exp(\beta_{pt}^M + \beta_{qt}^A + \beta_{pq}^{MA} - 0.5\beta_{pr}^{MM} - 0.5\beta_{qs}^{AA})}{\sum_{p'=1}^{4}\sum_{q'=1}^{5}\exp(\beta_{p't}^M + \beta_{q't}^A + \beta_{p'q'}^{MA} - 0.5\beta_{p'r}^{MM} - 0.5\beta_{q's}^{AA})}. \tag{8}$$

Here, the term $X_j = t$ indicates that we condition the logit on ego j's membership of latent class t. As can be seen, the parameters β_{pt}^M and β_{qt}^M contain an index t, indicating that these terms may differ across latent classes; that is, that these terms can be viewed as random effects. More specifically, rather than assuming that each individual has its own specific selection of alters, it is assumed that there are groups of individuals who have a specific selection of alters.

The connection between the above model and a standard latent class model becomes clearer if we write down the model for the joint probability density function associated with the full network of ego j; that is,

$$\begin{aligned} P(\mathbf{Y}_j^M, \mathbf{Y}_j^A | Z_j^M, Z_j^A) &= \sum_{t=1}^{T} P(X_j = t) P(\mathbf{Y}_j^M, \mathbf{Y}_j^A | Z_j^M, Z_j^A, X_j = t) \\ &= \sum_{t=1}^{T} P(X_j = t) \prod_{i=1}^{N_j} P(Y_{ij}^M, Y_{ij}^A | Z_j^M, Z_j^A, X_j = t). \end{aligned} \tag{9}$$

As in a standard latent class model, the joint distribution of the observed response variables given external variables, $P(\mathbf{Y}_j^M, \mathbf{Y}_j^A | Z_j^M, Z_j^A)$, is obtained as a weighted average of T class-specific distributions, $P(\mathbf{Y}_j^M, \mathbf{Y}_j^A | Z_j^M, Z_j^A, X_j = t)$, where the class sizes $P(X_j = t)$ serve as weights. As can be seen, the N_j observations of case j (alters' responses of ego j) are assumed to be independent given the class membership of case j, which is equivalent to the local independence assumption in a standard latent class model (Goodman, 1974). Note that also in the parametric random effects model we make the assumption of local independence; that is, responses are assumed to be independent given the random effects. Different from a standard latent class model are that variable pairs (Y_{ij}^{M}, Y_{ij}^{A}) serve as joint indicators instead of single variables and that the number of indicators (observed responses) varies across cases.

Allowing parameters to vary across latent classes of individuals is not only a way to take into account dependencies between observations, it also provides us information about the marital status and age homogeneity of networks, controlling for alter and ego characteristics. As will be shown when presenting the results obtained with our analysis, latent classes not only capture dependencies, but can also be given meaningful labels in terms of types of personal networks.

Whereas the model without random effects can be estimated with either standard log-linear analysis or conditional logit procedures, our two random effects variants require specialized software. We used the Latent GOLD Choice 4.0 software package (Vermunt & Magidson, 2005), a program that provides maximum likelihood estimates for standard and latent-class-based random effects conditional logit models using a hybrid EM and Newton-Raphson algorithm.

Results

Tables 1 and 2 present the cross-tabulations for the respondents' and alters' marital status and age, respectively. As can be seen from the bottom row of Table 1, the overall marginal distribution of alters' marital status is far from uniform (large overrepresentation of married), which should be taken into account when interpreting the numbers in this table. Comparison of the row percentages with the overall marginal distribution provides evidence for marital status homogeneity: Married egos have more alters with the same married status in their network than overall (82.5% versus 75.5%), and the same applies for egos that are single (38.5% versus 16.2%), divorced (8.7% versus 3.2%), and widowed (20.0% versus 5.1%). Similarly, Table 2 shows that there is a high degree of age homogeneity in this sample, which is confirmed by the large correlation ($r = .63$) between the respondents' and alters' age (computed using the original noncategorized age variables). Note that the logit models presented in this article yield an easier understanding of the boundaries that exist between the groups because (log) odds ratios are not affected by marginal distributions.

Table 1. Cross-tabulation of marital status of ego and alter: row percentages.

	Alter					
Ego	Single	Married	Divorced	Widowed	Total	Percentage
Single	38.5	55.7	2.0	3.8	100.0	21.7
Married	9.5	82.5	3.1	4.9	100.0	71.8
Divorced	17.4	65.0	8.7	9.0	100.0	4.2
Widowed	10.0	63.3	6.7	20.0	100.0	2.3
Total	16.2	75.5	3.2	5.1	100.0	100.0

Table 2. Cross-tabulation of marital status of ego and alter: row percentages.

	Alter						
Ego	<30	30–39	40–49	50–59	60+	Total	Percentage
<30	60.7	22.8	7.8	4.7	4.0	100.0	21.2
30–39	19.9	52.2	18.6	4.9	4.4	100.0	26.7
40–49	6.8	29.6	39.8	14.2	9.6	100.0	20.9
50–59	4.4	12.6	29.9	31.4	21.8	100.0	16.9
60+	3.0	9.0	12.9	23.0	52.1	100.0	14.2
Total	20.8	28.4	21.8	13.8	15.1	100.0	100.0

The fit results of the estimated conditional logit models are presented in Table 3. Model A is a conventional conditional logit model—in which we do not control for the association between alter's age and alter's marital status ($\beta_{pq}^{MA} = 0$). Model B1 is also equal to a conventional conditional logit model, but compared to Model A, it adds a set of parameters for the association between alter's age and marital status. Both the large increase of the log-like-

Table 3. Fit measures for the estimated conditional logit models.

Model	Description[a]	log-likelihood	BIC	# parameters
A	Equation 3 with $\beta_{pq}^{MA} = 0$	−15917	31990	23
B1	Equation 3	−15074	30385	35
B2	Equations 8 and 9, 2 classes	−14891	30074	43
B3	Equations 8 and 9, 3 classes	−14807	29960	51
B4	Equations 8 and 9, 4 classes	−14758	29915	59
B5	Equations 8 and 9, 5 classes	−14725	29904	67
B6	Equations 8 and 9, 6 classes	−14699	29907	75
C1	Equations 4 and 5, one factor	−14876	30036	42
C2	Equations 4 and 6, simple structure and two correlated factors[b]	−14807	29906	43
C3	Equations 4 and 6, simple structure and two uncorrelated factors	−14807	29899	42
C4	Equations 4 and 7, two uncorrelated factors	−14792	29916	47

[a]See text for a detailed explanation of the models. [b]Whereas all other models where estimated with Latent GOLD Choice 4.0, this model was estimated with an experimental version of Latent GOLD Choice that allows including factor correlations.

lihood value and the lower Bayesian Information Criterion (BIC) value compared to Model A show that the fit improves considerably when taking into account the dependency between marital status and age homogeneity, which is as expected.

Models B2–B6 are random effects models with 2 to 6 latent classes, and Models C1–C4 are factor-analytic random effect models based on the various specifications described in Equations 5 to 7. The fit measures indicate that the random effects models fit much better than the conventional conditional logit model, indicating that even after controlling for ego characteristics there is substantial dependence between alters' marital statuses and ages. The optimal number of latent classes according to the BIC criterion is 5. According to the BIC criterion, Model C3 is the preferred parametric model, indicating that heterogeneity in marital status and age are different and even uncorrelated dimensions. Note that despite the fact that the best latent class model (Model B5) yields a much lower log-likelihood value than the best parametric models (Model C3), the latter has a slightly lower BIC value because it contains a smaller number of free parameters.

In Table 4, we present the quasi-symmetry parameters describing the boundaries between categories (i.e., the log odds ratios). Positive log odds ratios indicate that there is less interaction between categories than one would expect under the independence model and the more positive the parameter, the stronger the boundary between the two categories concerned. We present the estimates for Model A, Model B1, the preferred latent class model (Model B5), and the preferred parametric random effects model (Model C3).

Let us first have a look the differences between Models A and B1. In Model A, all age and marital status parameters are strong and positive, confirming that age and marital status serve as boundaries in interaction. In Model B1 the age parameters hardly change. The reduction varies somewhat across parameters but is 10% at the highest. The parameters for marital status selection change considerably, where the exact reduction depends on the parameter we look at. Whereas these findings clearly support the by-product hypothesis, this explanation is not sufficient since the marital status parameters remain positive and statistically significant (in most cases). Hence, our first conclusion is that marital status selection is in large part a function of selection by age, whereas age selection is not a function of selection by marital status. More important, there is an independent tendency to select alters within marital status groups.

The degree to which age selection is responsible for marital status selection depends on the type of marital status boundary we look at. For the degree of interaction between widows and the other categories, age selection is very important. For the combination of single and widowed, for example, the reduction is 74%. This is also plausible, given the older ages of most widows and widowers. For the boundary between single and married people, age selection is comparatively less important. After controlling for age, the relevant parameter declines by 31%. According to this model, the boundary between divorced and married people, finally, cannot be explained at all by age selection.

Models B5 and C3 take into account that alters may show dependencies within egos. When comparing parameters of the two random effects models with the ones of the conventional conditional logit model, we see that the standard errors increase for virtually all parameters. This shows that the efficiency is lower when alters are clustered within egos. At the same time, however, we see that in most cases, the parameters increase in magnitude. These

Table 4. Log odds ratios (beta parameters) for boundaries between age and marital status groups.

	Model A		Model B1		Model B5		Model C3	
	Beta	SE	Beta	SE	Beta	SE	Beta	SE
One age class difference								
<30 vs. 30–39	1.95	0.08	1.75	0.08	1.95	0.12	2.01	0.11
30–39 vs. 40–49	1.32	0.08	1.32	0.08	1.46	0.11	1.40	0.09
40–49 vs. 50–59	1.04	0.10	1.04	0.10	1.11	0.12	1.10	0.11
50–59 vs. 60+	1.20	0.11	1.17	0.11	1.26	0.14	1.22	0.12
Two age class difference								
<30 vs. 40–49	3.81	0.14	3.57	0.14	3.94	0.18	4.27	0.19
30–39 vs. 50–59	3.25	0.13	3.25	0.13	3.69	0.18	3.48	0.16
40–49 vs. 60+	2.81	0.13	2.78	0.13	2.97	0.16	3.02	0.16
Three age class difference								
<30 vs. 50–59	4.54	0.18	4.29	0.18	4.94	0.24	5.29	0.25
30–39 vs. 60+	4.23	0.15	4.20	0.15	4.84	0.22	4.69	0.21
Four age class difference								
<30 vs. 60+	5.45	0.21	5.32	0.21	6.03	0.26	6.97	0.31
Marital status differences								
Single - married	1.78	0.07	1.22	0.07	1.27	0.09	1.38	0.11
Married - divorced	1.27	0.21	1.22	0.21	1.17	0.24	1.15	0.24
Divorced - widowed	1.19	0.40	0.76	0.41	0.69	0.43	0.67	0.43
Single - divorced	2.21	0.28	1.20	0.29	1.26	0.30	1.28	0.30
Single - widowed	3.06	0.29	0.79	0.30	0.88	0.32	0.89	0.32
Married - widowed	1.66	0.20	0.91	0.20	0.88	0.22	1.03	0.22

Table 5. Latent class parameters (transformed to probabilities) for Model B5 and factor loadings for Model C3.

	Model B5					Model C3
	Class 1	Class 2	Class 3	Class 4	Class 5	Loadings
Alter's age group						
<30	0.05	0.19	0.18	0.02	0.25	0.86
30–39	0.09	0.39	0.10	0.19	0.20	0.34
40–49	0.24	0.21	0.07	0.19	0.22	−0.07
50–59	0.28	0.09	0.13	0.09	0.16	−0.35
60+	0.34	0.12	0.51	0.52	0.17	−0.77
Alter's marital status						
Single	0.13	0.12	0.06	0.21	0.31	0.38
Married	0.83	0.83	0.84	0.72	0.59	−0.43
Divorced	0.02	0.02	0.07	0.04	0.08	0.17
Widowed	0.02	0.03	0.03	0.02	0.03	−0.12
Size of latent class	0.25	0.25	0.2	0.17	0.13	

increases are not large, but it is interesting that they more than compensate the increase in the standard errors. This is a common phenomenon in nonlinear random effects models. Not only are standard errors biased when dependencies are not taken into account, but the parameters estimates themselves may be biased downwards. See, for example, the discussion on the difference between marginal and subject-specific effects by Agresti (2002).

After controlling for age selection and for dependencies between alters, we see positive and statistically significant marital status parameters. How strong are these marital status boundaries and where in the life course are they strongest? The parameter for single and married persons is 1.27 (Model B1). This shows that the odds that a single person picks a single person (rather than a married person) are $e^{1.27} = 3.6$ times higher than the odds that a married person picks a single person. This is a substantial boundary. There is also a strong boundary between married and divorced persons (the odds ratio is 3.2), and a strong boundary between single and divorced persons (3.5). While these three boundaries are more or less comparable in magnitude, the boundaries involving widows are much weaker (2.0 for interaction with divorced persons and 2.4 for interaction with either married or single persons). Together these results show that life course events that involve choice (i.e., marriage, divorce) produce stronger boundaries in social life than life course events that do not (i.e., widowhood). One could have expected divorce to produce the strongest boundaries, because divorce is often normatively disapproved of by others or can be considered a threat to others. This is not the case, however. The boundaries between divorced and married persons are as strong as the boundaries between (never-married) single persons and married persons.

We now turn to the interpretation of the latent classes and the continuous factors in the two types of random effects models. A latent class is a cluster of egos having a tendency of choosing alters of a certain kind, independently of their own ages and marital statuses. As a result of this clustered selection, alters are more similar in terms of age and marital status than can be explained by ego's age and marital status. The (partial) conditional probabilities obtained with the β^M_{pt} and β^A_{rt} parameters (e.g., $P(Y^M_{ij} = p|X_j = t) = \exp(\beta^M_{pt})/\Sigma^4_{p'=1}\exp(\beta^M_{p't})$) reported in Table 5 show the (ego) cluster differences with respect to the ages and marital statuses of their personal network members. The first class consists of egos having a tendency to pick age categories 50–59 and 60+ more often than average (averages appear in the last row of Table 2), Class 2 age category 30–39, Classes 3 and 4 age category 60+, and Class 5 age category 18–29. Classes 1 to 3 as well as Classes 4 and 5 are similar with respect to their marital status preferences. Egos in Classes 4 and 5 have more single people in their network compared to the average (see last row, Table 1) and egos in Classes 1 to 3 more married people. The classes of egos are clearly different with respect to the distribution of alters' age and marital status, which means that there is evidence for what we called clustered selection. The labeling of the classes is, however, less clear than in standard latent class cluster or scaling applications. One may, however, also treat the latent class parameters of our model as nuisance parameters; that is, as parameters that are only included in the model to correct for dependencies between the multiple alters within an ego.

The factor loadings for the parametric random effects model (Model 3c) are reported in the last column of Table 5. As can be seen, the category-specific loadings for the age-related factor show a clear monotonic pattern, indicating that there is a positive residual association between alters' ages. The loadings for the marital status factor indicate that the main contrast is between networks with an overrepresentation of either married alters or singles, which is in agreement with what we saw in the latent class solution. Widowed and divorced take an intermediate position, where in terms of the ego preferences divorced alters are more similar to singles and widowed to married alters.

Final Remarks

Using an example on network homogeneity, we showed that random effects models provide an elegant manner to deal with regression-like problems for ego-centered networks. Whereas in our application, we had to use a rather complex regression model—a conditional logit model for a bivariate response variable—in other applications, the regression model may be much simpler, for example, a Poisson regression for counts (number of events), a binary logit model for a dichotomous outcome variable, or a linear regression model for a continuous response variable.

Various existing extensions of the random effects models described in this article may be of interest for analyzing more complex personal network data. One useful extension is the possibility to include ego-level covariates in the model affecting either the latent class membership or the continuous factors. This would be a way to explain why egos differ with respect to their choice of alters. Another extension involves the possibility to deal with longitudinal personal network data. This could be done either by means of a growth model, which in our case would imply specifying a three-level random effects model, or by a transition model such as a latent class Markov model. A last interesting extension we would like to mention is relevant when egos are nested with groups of egos, for example, within regions or organizations. In such a situation, one would need a three-level random effects model, either a standard parametric model or a latent-class-based model.

References

Agresti, A. (2002). *Categorical data analysis*. New York: Wiley.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55,* 218–234.

Broese van Groenou, M. I., & Van Tilburg, T. G. (1996). Network analysis. In L. E. Birren (Ed.), *Encyclopedia of gerontology: Age, aging, and the aged* (pp. 197–210). San Diego, CA: Academic Press.

Burt, R. S. (1991). Measuring age as a structural concept. *Social Networks*, *13,* 1–34.

Fischer, C. S. (1982). *To dwell among friends: Personal networks in town and city*. Chicago: University of Chicago Press.

Fiselier, A. A. M., Van der Poel, M. G. M., & Felling, A. (1987). *Primary relations and social support*. Nijmegen, the Netherlands: University of Nijmegen.

Gerich, J., & Lehner, R. (2006). Collection of ego-centered network data with computer-assisted interviews. *Methodology*, *2,* 1–15.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. *Part I*: A modified latent structure approach. *American Journal of Sociology*, *79,* 1179–1259.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, *22,* 1433–1446.

Hout, M., & Goldstein, J. R. (1994). How 4.5 million Irish immigrants became 40 million Irish Americans: Demographic and subjective aspects of the ethnic composition of white Americans. *American Sociological Review*, *59,* 64–82.

Kalmijn, M. (1991). Status homogamy in the United States. *American Journal of Sociology*, *97,* 496–523.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, *24,* 395–421.

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for the external analysis of choice data. *International Journal of Research in Marketing*, *11,* 451–464.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as social process: A substantive and methodological analysis. In M. Berger, T. Abel, & C. H. Page (Eds.), *Freedom and control in modern society* (pp. 18–66). New York: Van Nostrand.

Louch, H. (2000). Personal network integration: Transitivity and homophily in strong-tie relations. *Social Networks*, *22,* 45–64.

Marsden, P. V. (1988). Homogeneity in confiding relations. *Social Networks*, *10,* 57–76.

Marsden, P. V. (1990). Network diversity, substructures, and opportunities for contact. In C. Calhoun, M. W. Meyer, & W. Richard (Eds.), *Structures of power and constraint* (pp. 397–410). Cambridge: Cambridge University Press.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In I. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Miller McPherson, J., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27,* 415–444.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* London: Chapman Hall/CRC.

Uunk, W. J. G., Ganzeboom, H. B. G., & Róbert, P. (1996). Bivariate and multivariate scaled association models: An application to homogamy of social origin and education in Hungary between 1930 and 1979. *Quality and Quantity*, *30,* 323–343.

Van Duijn, M. A. J., Van Busschbach, J. T., & Snijders, T. A. B. (1999). Multilevel analysis of personal networks as dependent variables. *Social Networks*, *21,* 187–209.

Van Poppel, F., Liefbroer, A. C., Vermunt, J. K., & Smeenk, W. (2001). Love, necessity and opportunity: Changing patterns of marital age homogamy in the Netherlands, 1850–1993. *Population Studies*, *55,* 1–13.

Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed outcome variables. *Multivariate Behavioral Research*, *40,* 281–301.

Vermunt, J. K., & Magidson, J. (2005). *Technical guide for Latent GOLD Choice 4.0: Basic and advanced*. Belmont, MA.: Statistical Innovations.

Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, *13,* 6–13.

Yamaguchi, K. (1990). Homophily and social distance in the choice of multiple friends. *Journal of the American Statistical Association*, *85,* 356–366.

Jeroen K. Vermunt

Tilburg University
Faculty of Social & Behavioral Sciences
Department of Methodology & Statistics
Postbus 90153
NL-5000 LE Tilburg
The Netherlands
Tel. +37 13 466 2748
Fax +37 13 466 3002
E-mail J.K.Vermunt@uvt.nl