Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Social Networks 29 (2007) 25–43
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Homogeneity of social networks by age and marital status: A multilevel analysis of ego-centered networks☆

Matthijs Kalmijn [a,*], Jeroen K. Vermunt [b]

[a] *Department of Social and Cultural Sciences, Tilburg University, P.O. Box 90.153, 5000 LE, Tilburg, The Netherlands*
[b] *Department of Methodology and Statistics, Tilburg University, P.O. Box 90.153, 5000 LE, Tilburg, The Netherlands*

Abstract

Is homogeneity in personal relationships in one trait the by-product of selection on another trait, or is it the result of direct selection on that trait? This question has often been analyzed in the context of marriage homogamy. We apply this issue to the question of whether there is selection in networks based on age on the one hand, and marital status on the other hand. The role of age has been documented before, but selection on the basis of marital status has not been documented. We analyze a representative survey containing data on contact and support networks. We use a novel analytical approach by adopting a latent class type random-effects approach to the multilevel structure of the network data which allows for simple descriptions of homogeneity in terms of odds ratio's. Our analyses show that age boundaries are strong and that they partly explain marital status boundaries. Nevertheless, even after controlling for age, we see important social boundaries between marital status groups. Moreover, we see a pattern of what we call clustered selection—the tendency of alters to be more similar to each other than one would expect from their similarity to ego.

Keywords: Marriage; Age; Homogeneity; Social networks; Latent class; Multilevel models

You've done too much, much too young
Now you're married with a kid
when you could be having fun with me
Too much too young – The Specials (1979).

☆ This paper was earlier presented at (a) the autumn meeting of the Social Sciences Section of The Netherlands Society for Statistics and Operations Research (NVVS), Tilburg, 1 October 2004, (b) the SISWO working group Social Inequality and the Life Course, Amsterdam, 10 November 2004.

* Corresponding author. Tel.: +31 13 466 2246; fax: +31 13 466 3002.
E-mail addresses: M.Kalmijn@uvt.nl (M. Kalmijn), J.K.Vermunt@uvt.nl (J.K. Vermunt).

doi:10.1016/j.socnet.2005.11.008

1. Introduction

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 (Lazarsfeld and Merton, 1954; Marsden, 1988, 1990; McPherson et al., 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 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 share their religion.

An important question this raises is whether the homogeneity that it is found in reality is based on explicit selection on that trait, or whether it is a by-product of selecting 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.

In the literature on homogamy and intermarriage, the by-product thesis has been examined empirically by applying multivariate log-linear models to marriage choices in two or more dimensions. Examples are analyses of education and class background homogamy (Kalmijn, 1991; Uunk et al., 1996), religious and ethnic intermarriage (Hout and Goldstein, 1994), religion and education (Hendrickx, 1998), race and education (Kalmijn, 1993; Qian, 1997), and education and unemployment status (Henkens et al., 1993; Ultee et al., 1988). In most of these studies, the degree of similarity declines when considering matching on the other trait, which shows that there is some truth in the by-product notion. However, the fact the similarity does not fully disappear in the multivariate model indicates that probably direct selection is involved as well.

In the field of network research, the issue of multidimensional homogeneity has rarely been analyzed. Our paper offers a new analysis of this problem by looking at the role of age and marital status in the composition of social networks. 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. That networks are homogeneous by age has been shown before (Burt, 1991; Hagestad and Uhlenberg, 2005; Louch, 2000; Marsden, 1988; McPherson et al., 2001), but marital status selection has not been demonstrated convincingly. In his classic American study of personal networks, Fischer (1982, p. 180–181) showed that married respondents more often named married associates, that never married more often named the never married, and that the divorced more often named the divorced. Fischer also demonstrated, like others, that age homogeneity is strong, but he did not analyze age and marital status simultaneously. Hence, his finding of marital status homogeneity may well be a by-product of age selection.

There are also many studies showing that the size and composition of personal networks change over the life course. Important life course transitions for networks are entering a cohabiting relationship or marriage (Hurlbert and Acock, 1990; Kalmijn, 2003; Milardo and Allan, 2000), becoming a parent (Bost et al., 2002; Munch et al., 1997), and experiencing a divorce or the death of a spouse (Booth et al., 1991; De Jong-Gierveld and Dykstra, 1993; Knipscheer et al., 1995). In this literature, it has not been examined, however, if and to what extent there is direct network selection on the basis of the stage in the life course that people are in.

There are several reasons why direct selection on the basis of marital status may exist. First, the contexts in which people meet others may be segregated by marital status, even after controlling for age. Examples of contexts that are important for network formation are schools, work places, neighborhoods, voluntary organizations, and leisure settings (Feld, 1981; Kalmijn and Flap, 2001). Most of these settings are homogeneous by age, but some can also be homogeneous with respect to marital status. For example, leisure settings and voluntary associations may be focused on specific marital status groups (e.g., parents, singles) and certain types of neighborhoods are heavily composed of married persons with children (e.g., suburbs). Second, people may have a preference for interacting with others in the same marital status category. People who are in the same marital status position may better understand each, they may have more relevant information for each other, and they may share a certain lifestyle which increases possibilities for joint activities. People may also want approval for the life course decisions they made and this may be obtained more easily from persons who went through the same transition. Third, marital status homogeneity may arise because people influence each other's life course decisions. If a person's friends start getting married, for example, this may speed up this person's decision to get married as well, thereby increasing the degree of marital status homogeneity.

If marital status selection exists, it has important implications for the dynamics of social networks. Many friends are made when people are young and at that age, age homogeneity is generally high (Allan, 1979). As the members of a certain age cohort become older, they experience life course events such as marriage, parenthood, divorce, and the death of a spouse. The timing of these events may differ among the members of an age cohort, and this creates a partial disconnection between age and the life course. Selection solely on the basis of age implies that not much network change would occur: the age cohort stays together as it ages. Selection on the basis of marital status, however, implies that networks will change. Friends who marry will become disconnected from friends who remain single, and friends who divorce will become disconnected from friends who remain married. In a sense, the members of an age cohort will be put on different tracks along the life course. The *Specials* song that is quoted in the beginning of this paper illustrates this effect.

Applying the concept of multiple homogeneity to social networks introduces issues that are not present when considering marriage choices. The most important analytical difference is that people have more than one network member at the same time, whereas they usually have only one spouse. This implies that we are in fact dealing with a two-level data structure, with network members nested within individuals. Because of the dependence of observations within an individual, standard log-linear models will no longer suffice. In models for continuous variables, solutions have already been offered (Van Duijn et al., 1999), but in the context of log-linear analyses of discrete characteristics, no standard practice exists.[1] Dependencies of observations within individuals are the result of statistical association among the network members themselves (Yamaguchi, 1990). Put differently, the network members can be more similar to each other than one would expect on the basis of their link to the observed respondent's characteristics. If such a pattern of what we call 'clustered selection' exists, it provides additional substantive information about the selection process. One probable cause of clustered selection is that selections are made within homogeneous settings. Another cause is that the selection of one network member is made through other network members. In a more general sense, the application of log-linear

[1] In an early paper, Marsden (1988) applied log-linear models to the analysis of multiple ego-alter dyads and applied statistical corrections for clustering to deal with the problem of dependencies.

association models to social networks provides new challenges and options that we will address in this paper.

The goal of this paper is three-fold. First, we want to describe the degree of age and marital status homogeneity in personal networks. Second, we want to assess whether marital status and age homogeneity are a function of each other. Third, we want to analyze the dependence between multiple alters of a respondent, thereby assessing whether there is clustered selection. To achieve these goals, we analyze an individual survey from The Netherlands that contains detailed data on personal contact and support networks. The data will be analyzed using a latent class discrete choice model (Kamakura et al., 1994; Vermunt and Magidson, 2003). More specifically, a conditional logit model is specified in which the network member's marital status and age serve as a joint dependent variable. Dependencies between the observations of the same individual are taken into account by adopting a latent class type random-effects approach (Vermunt and Van Dijk, 2001).

2. Data

We analyze a survey which is based on face-to-face interviews with a random national sample of 902 individuals in The Netherlands (Fiselier, Van der Poel & Felling 1987). Earlier analyses of the Dutch data can be found in Van der Poel (1993). The network we are studying can best be described as the *personal contact and support network* (Broese van Groenou and Van Tilburg, 1996). For example, respondents had to list the people with whom they regularly went out with (contact method) and the persons who had helped (or could have helped) with odd jobs around the house (support method). 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. There are no representative network data that we know of that contain information on the parent status of network members.

Five age categories are distinguished: (a) 20–30, (b) 30–39, (c) 40–49, (d) 50–59, and (e) 60–72. We also limited the ages of alters to 20 years and over since we are interested in relationships among adults.

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). These alters were excluded. After these selections were made, the number of respondents is 875 (7896 relationships).

Note that the analysis is cross-sectional: we see both old and young relationships in the data, and we cannot assess to what extent homogeneity is the result of people choosing each other and to what extent it is the result of people ending relationships selectively. It is plausible that both processes will operate but longitudinal data are needed to establish this.[2] Note that prospective longitudinal designs are now becoming more popular in network research (Suitor et al., 1997; Van Duijn et al., 1999), but such data currently still have important disadvantages. Prospective network data are often based on very small and selective samples (e.g., 20–100 respondents), and what is more problematic, they rarely cover a broad span of the life course.

3. Method: a latent class conditional logit model

We start with the well-established log-linear framework that has been used in analyses of (multidimensional) marriage tables. We subsequently translate these models into bivariate and multivariate latent class conditional logit models which take into account the network structure of the data.

3.1. A multivariate quasi-symmetry model and its conditional logit variant

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 and Goldstein, 1994; Uunk et al., 1996). This model has the following form:

$$\log[P(Y_{ij}^{\mathrm{M}} = p, Z_j^{\mathrm{M}} = r)] = \alpha_0 + \alpha_r^{\mathrm{M}} + \beta_p^{\mathrm{M}} - 0.5\beta_{pr}^{\mathrm{MM}}, \tag{1}$$

where α_0 is 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} is the association parameter capturing the dependence between ego's and alter's status categories. A quasi-symmetry model is obtained by imposing a symmetric structure on these association parameters, which can be achieved using different types of parameterizations.

We assume that $\beta_{pr}^{\mathrm{MM}} = \beta_{rp}^{\mathrm{MM}}$ if $p \neq r$, and that $\beta_{pr}^{MM} = 0$ otherwise. Under this parameterization, each of the free β_{pr}^{MM} parameters has a very simply and useful interpretation; that is, it is the log odds ratio in the two-way table formed by egos' and alters' marital status categories p and r. This log odds ratio is defined as the odds that, for example, a married person interacts with a married person rather than with a single person, divided by the odds that a single person interacts with a married person (rather than with a single person). A positive log odds ratio indicates that there is less interaction between the categories concerned (marital statuses p and r) than can be expected based on the marginal distributions; and the higher the β_{pr}^{MM} parameter, the stronger the boundary between the two categories. Negative values, on the other hand, indicate that there is more interaction between the two categories than can be expected from the marginal distributions.

In Eq. (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

[2] We will present some analyses in which the duration of the relationship is included in order to say something about the role of time in the network.

equation:

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

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 Eq. (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}.

Model (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 the ego's marital status and age:

$$\begin{aligned} &P(Y_{ij}^{\mathrm{M}} = p, Y_{ij}^{\mathrm{A}} = q|Z_j^{\mathrm{M}} = r, Z_j^{\mathrm{A}} = s) \\ &\quad = \frac{\exp(\beta_p^{\mathrm{M}} + \beta_q^{\mathrm{A}} + \beta_{pq}^{\mathrm{MA}} - 0.5\beta_{pr}^{\mathrm{MM}} - 0.5\beta_{qs}^{\mathrm{AA}})}{\sum_{p=1}^{4}\sum_{q=1}^{5}\exp(\beta_p^{\mathrm{M}} + \beta_q^{\mathrm{A}} + \beta_{pq}^{\mathrm{MA}} - 0.5\beta_{pr}^{\mathrm{MM}} - 0.5\beta_{qs}^{\mathrm{AA}})}. \end{aligned} \tag{3}$$

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. The other two terms – β_{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 model (2) with the ones of a model in which the β_{pq}^{MA} terms are omitted or, equivalently, in which $\beta_{pq}^{\mathrm{MA}} = 0$. 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 Eq. (2) in which the age variables are fully omitted.

3.2. Taking into account dependencies, a latent class approach

So far, we 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. For example, the alters of a certain ego may have a tendency to be relatively old or disproportionately married, regardless of ego's own age and marital status. It is well known that standard errors are biased (usually downwards) when dependencies between observations are not taken into account, which yields incorrect tests. In nonlinear regression models like our conditional logit model, parameter estimates may also be biased, usually downwards (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 (Yamaguchi, 1990). When there are c status categories and n alters, the model is estimated for a $c \times c^n$ table. The associations between the alters' statuses are captured by two-variable log-linear association terms which can be assumed to be the same for each pair of alters. Despite of the fact that this approach is elegant, conceptually simply, and that it 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 Dutch data set, for example, the largest personal network consist 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, Busschbach, and Snijders (1999) proposed using linear regression models with random effects for the analysis of personal networks with tie information—for example, distance—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 Eq. (3), that is, a model in which the β_p^{M} and β_q^{A} parameters, and possibly also the β_{pq}^{MA} parameters, are specified to be random effects. Estimation of such a random-effects conditional logit model can be extremely complicated and computationally intensive if we would like to make the standard assumption that the random effects come from a multivariate normal distribution.[3] Moreover, interpretation of the parameters associated with the random effects may become difficult with more than a few random effects.

Because of the computational and conceptual difficulties associated with such a parametric random-effects conditional logit model, we decided 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 and Rabe-Hesketh, 2004; Vermunt and Van Dijk, 2001). This yields a model that is called a latent class or mixture conditional logit model (Kamakura et al., 1994). It should be noted that the use of latent class models for describing dependencies between categorical observed variables has a long tradition in sociology (Goodman, 1974; Lazarsfeld, 1950; Yamaguchi, 2000). As pointed out by Aitkin (1999), the proposed 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 made about the distribution of the random effects (Aitkin, 1999).

The relevant latent class variant of the conditional logit described in Eq. (3) has the following form:

$$P(Y_{ij}^{\mathrm{M}} = p, Y_{ij}^{\mathrm{A}} = q | Z_j^{\mathrm{M}} = r, Z_j^{\mathrm{A}} = s, X_j = t)$$
$$= \frac{\exp(\beta_{pt}^{\mathrm{M}} + \beta_{qt}^{\mathrm{A}} + \beta_{pqt}^{\mathrm{MA}} - 0.5\beta_{pr}^{\mathrm{MM}} - 0.5\beta_{qs}^{\mathrm{AA}})}{\sum_{p=1}^{4}\sum_{q=1}^{5} \exp(\beta_{pt}^{\mathrm{M}} + \beta_{qt}^{\mathrm{A}} + \beta_{pqt}^{\mathrm{MA}} - 0.5\beta_{pr}^{\mathrm{MM}} - 0.5\beta_{qs}^{\mathrm{AA}})}.$$

[3] Computationally more efficient approximate maximum likelihood estimation methods have been developed for some types of nonlinear random-effects models (and implemented in software such as HLM and MLwin), but these are not available for the conditional logit model.

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^{M}_{pt}, β^{A}_{qt}, and $\beta^{\mathrm{MA}}_{pqt}$ now contain an index t, indicating that these terms may differ across latent classes; that is, that these terms are random effects. The novelty of this approach is that the effects (and in this case, the intercepts referring to alter's age and marital status) are not random among individuals, but random among classes. In other words, rather than assuming that each individual has its own specific selection of alters, it is assumed that there are groups (or classes) of individuals who have a specific selection of alters.

We also use the less complex model specification in which the association between alter's marital status and alter's age is assumed to be class independent; that is, in which $\beta^{\mathrm{MA}}_{pqt} = \beta^{\mathrm{MA}}_{pq}$. To explain the difference between these two specifications, we can use an example. The simpler model assumes that ego's in a certain class disproportionally choose old people and disproportionally choose people who are widowed. The more complex specification assumes that the people in a certain class disproportionally choose older widows. In the latter case, the alters will more often be widowed than would be expected on the basis of the alter's age and they will more often be old than expected on the basis of alter's marital status. In other words, in the former case, alters are chosen based on one characteristic at a time, in the latter case, they are chosen on the combination of their traits. Note finally, that—although it is technically possible within the latent class regression framework—it does not make sense to assume that the effects of interest (the association between ego's and alter's age and marital status) are class-specific (differ across ego's). In multilevel terminology, ego's age and marital status are level-2 predictors, and the effects of such predictors are not allowed to vary across level-2 units.

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 case j; that is,

$$\begin{aligned} P(Y^{\mathrm{M}}_j, Y^{\mathrm{A}}_j | Z^{\mathrm{M}}_j, Z^{\mathrm{A}}_j) &= \sum_{t=1}^{T} P(X_j = t) P(Y^{\mathrm{M}}_j, Y^{\mathrm{A}}_j | Z^{\mathrm{M}}_j, Z^{\mathrm{A}}_j, X_j = t) \\ &= \sum_{t=1}^{T} P(X_j = t) \prod_{i=1}^{N_j} P(Y^{\mathrm{M}}_{ij}, Y^{\mathrm{A}}_{ij} | Z^{\mathrm{M}}_j, Z^{\mathrm{A}}_j, X_j = t). \end{aligned}$$

As in a standard latent class model, the joint distribution of the observed variables, $P(Y^{\mathrm{M}}_j, Y^{\mathrm{A}}_j | Z^{\mathrm{M}}_j, Z^{\mathrm{A}}_j)$, is obtained as a weighted average of the class-specific distributions, $P(Y^{\mathrm{M}}_j, Y^{\mathrm{A}}_j | Z^{\mathrm{M}}_j, Z^{\mathrm{A}}_j, X_j = t)$, were 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. This assumption is similar to the local independence assumption in a standard latent class model (Goodman, 1974). Different from a standard latent class model is that variable pairs $(Y^{\mathrm{M}}_{ij}, Y^{\mathrm{A}}_{ij})$ serve as joint indicators instead of single variables. Another difference is that the number of indicators (observed responses) varies across cases instead of having a fixed number of response variables or items for each case.

Allowing parameters to vary across latent classes of individuals is not only a manner 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 do not only capture dependencies, but can also be given meaningful labels in terms of types of personal networks.

Because of its bivariate dependent variable and its symmetric association structures, it is not possible to estimate the proposed model with standard multinomial logistic regression analysis procedures. Whereas the model without random effects can be estimated with either standard log-linear analysis or conditional logit procedures, for the latent class variant we need specialized software. We used the Latent GOLD Choice software package (Vermunt and Magidson, 2003). This program provides maximum likelihood estimates of latent class conditional logit models using a hybrid EM and Newton–Raphson algorithm.

4. Results

4.1. Descriptive results

Table 1 presents the cross-tabulation for the marital status of the respondent and his or her alter. Table 2 presents the cross-tabulation for the age of the respondent and that of his or her alter. The top part of these two tables have dyads as the unit of analysis. In the bottom part of the table we use ego's as the unit of analysis and add information on aggregate characteristics, i.e., the percentage of ego's who have at least one person of a specified age or marital status in their network. The discussion below refers to dyads.

The vast majority of the relationships of married respondents are with other married persons (83%). For single persons, the pattern is a little different. A large number of the relationships of singles are with other singles (39%), but there is also a large group of relationships that are with married or divorced persons. We further see that both widowed and divorced persons most often have married persons in their network. Interesting is that the relationships of the widowed often are with other widowers. The relationships of divorced persons do not appear to be often with other divorcees. The bottom part of the table shows the percentages of ego's having at least one specified category in their network. These numbers show that virtually all categories of ego's have at least one married person in the network.

While these tables are interesting for descriptive purposes, the patterns also reflect the relative sizes of the various groups in society as a whole (Blau and Schwartz, 1984). Due to the relatively

Table 1
Crosstabulation of marital status of ego and alter: row percentages

Alter	Single	Married	Divorced	Widowed	Total	Percentage	N
Ego							
Single	38.5	55.7	2.0	3.8	100.0	21.7	1715
Married	9.5	82.5	3.1	4.9	100.0	71.8	5667
Divorced	17.4	65.0	8.7	9.0	100.0	4.2	334
Widowed	10.0	63.3	6.7	20.0	100.0	2.3	180
Percentage	16.2	75.5	3.2	5.1	100.0	100.0	7896
Ego							
Single	87.1	95.2	14.5	21.0		21.3	186
Married	47.7	99.2	21.5	30.1		71.7	627
Divorced	65.8	92.1	50.0	44.7		4.3	38
Widowed	56.6	97.9	21.6	29.9		2.7	24

Note: married includes cohabiting. Top part has dyads as the unit of analysis. Bottom part has ego's as the unit, where the numbers represent the percentages of ego's with at least one alter of the specified column category.

Table 2
Crosstabulation of age category of ego and alter: row percentages

Alter	<30	30–39	40–49	50–59	60–72	Total	Percentage	*N*
Ego								
<30	60.7	22.8	7.8	4.7	4.0	100.0	21.2	1675
30–39	19.9	52.2	18.6	4.9	4.4	100.0	26.7	2112
40–49	6.8	29.6	39.8	14.2	9.6	100.0	20.9	1653
50–59	4.4	12.6	29.9	31.4	21.8	100.0	16.9	1333
60–72	3.0	9.0	12.9	23.0	52.1	100.0	14.2	1123
Percentage	20.8	28.4	21.8	13.8	15.1	100.0	100.0	7896
Ego								
<30	96.3	71.6	40.0	28.4	25.6		21.7	190
30–39	65.6	97.8	72.2	31.3	31.3		25.9	227
40–49	36.0	84.3	94.2	61.0	47.1		19.7	172
50–59	26.2	53.8	88.3	93.8	75.2		16.6	145
60–72	14.9	39.7	53.2	68.8	92.2		16.1	141

Note: top part has dyads as the unit of analysis. Bottom part has ego's as the unit, where the numbers represent the percentages of ego's with at least one alter of the specified column category.

large numbers of married persons in the population, most groups will have a tendency to include other married persons in their network. This may explain, for example, why singles choose other singles less often than that married persons choose other married persons. Logit models yield a better understanding of the boundaries that exist between the groups because the implied (log) odds ratios are independent of the effects of the marginal distributions.

Table 2 presents the table of the ages of the respondents and alters. The table confirms the high degree of homogeneity. The correlation between the ages is $r = 0.63$ (for the continuous version of the age variables), which is a substantial correlation. Important to note is that the correlation between ego's and alter's ages is partly due to people growing old together. We can check the degree to which this is true by calculating the correlation for different stages of the relationship. The correlation is $r = 0.45$ for people who had known each other for 1 year. This correlation is lower than it is for all dyads, showing that the 'ageing' of networks contribute to age homogeneity in society.

4.2. Model selection

The fit results of the latent class conditional logit models are presented in Table 3. Model A is a one-class model—comparable to a conventional conditional logit model—in which we do not control for the association between alter's age and alter's marital status. Model B1 is also equal to a conventional conditional logit model but compared to Model A, it adds parameters for the association between alter's age and marital status. The large increase of the log-likelihood value and the lower BIC value show that the fit improves considerably, which is as expected. The remaining models are latent class models with more than one class. These models take into account that the alters may be correlated within ego's, or more precisely, that there are latent classes of egos that differ with respect to the age and marital status distributions of alters. There are two versions of this model, one in which only the main effects of age and marital status are assumed to be class specific (Models B2, B3, etc.), and one in which also the association between age and marital status differs across classes (Models C2, C3, etc.).

Table 3
Fit of latent class conditional logit models

Model	Description	Log likelihood	BIC	Number of parameters
A	Age alter + stage alter + age ego − age alter + stage ego − stage alter	−15917	31990	23
B1/C1	+Age alter − stage alter	−15074	30385	35
B2	+Age alter random + stage alter random; 2 class	−14891	30074	43
B3	+Age alter random + stage alter random; 3 class	−14807	29960	51
B4	+Age alter random + stage alter random; 4 class	−14758	29915	59
B5	+Age alter random + stage alter random; 5 class	−14725	29904	67
B6	+Age alter random + stage alter random; 6 class	−14699	29907	75
C2	+Age alter − stage alter random; 2 class	−15074	30385	55
C3	+Age alter − stage alter random; 3 class	−14888	30149	75
C4	+Age alter − stage alter random; 4 class	−14790	30089	95
C5	+Age alter − stage alter random; 5 class	−14735	30114	115
C6	+Age alter − stage alter random; 6 class	−14692	30163	135
	Number of cases	875		
	Number of replications	7896		

Note: stage means marital status. See text for formal details of models.

The fit measures indicate that the latent class models are an improvement over the one-class or conventional conditional logit model. The optimal number of latent classes according to the BIC criterion is five. The models containing class specific age and marital status associations apparently fit better than the models without such class specific effects.

4.3. Boundaries between age and marital status groups

In Table 4, we present the parameters describing the boundaries between categories (i.e., the log odds ratio's). 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, and the preferred multiple class model (Model B6).

We start by discussing 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. When looking at Model B1, we see that 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, however. The exact reduction depends on the parameter we look at, but the change is in most cases considerable. While 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 to a large part a function of selection by age, whereas age selection is not a function of selection by marital status. More importantly, there is an independent tendency to select alters within marital status groups. Stage in the life course thus has a direct effect on network selection.

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

Table 4
Log odds ratios of boundaries between age and marital status groups

	Model A		Model B1		Model B5	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
One age class difference						
20–29 vs. 30–39	1.95	0.08	1.75	0.08	1.95	0.12
30–39 vs. 40–49	1.32	0.08	1.32	0.08	1.46	0.11
40–49 vs. 50–59	1.04	0.10	1.04	0.10	1.11	0.12
50–59 vs. 60–72	1.20	0.11	1.17	0.11	1.26	0.14
Two age classes differences						
20–29 vs. 40–49	3.81	0.14	3.57	0.14	3.94	0.18
30–39 vs. 50–59	3.25	0.13	3.25	0.13	3.69	0.18
40–49 vs. 60–72	2.81	0.13	2.78	0.13	2.97	0.16
Three age classes differences						
20–29 vs. 50–59	4.54	0.18	4.29	0.18	4.94	0.24
30–39 vs. 60–72	4.23	0.15	4.20	0.15	4.84	0.22
Four age class differences						
20–29 vs. 60–72	5.45	0.21	5.32	0.21	6.03	0.26
Marital status differences						
Single – married	1.78	0.07	1.22	0.07	1.27	0.09
Married – divorced	1.27	0.21	1.22	0.21	1.17	0.24
Divorced – widowed	1.19	0.40	0.76	0.41	0.69	0.43
Single – divorced	2.21	0.28	1.20	0.29	1.26	0.30
Single – widowed	3.06	0.29	0.79	0.30	0.88	0.32
Married – widowed	1.66	0.20	0.91	0.20	0.88	0.22

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%. The boundary between divorced and married people, finally, cannot be explained at all by age selection.

Model B5 takes into account that alters may be dependent within egos. When comparing parameters of the multiple-class model with the ones of the one-class model, we see that the standard errors increase for virtually all parameters. This shows that the efficiency is lower when alters are clustered within ego's. At the same time, however, we see that in most cases, the parameters increase in magnitude. These 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 standard errors are biased when dependencies are not taken into account, but also 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 B5). 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 (i.e., 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). 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.

Table 4 also tells us how strong age boundaries are in network formation. We see that all log odds ratios are strong and significant. Relationships crossing two age categories are less common than relationships crossing one age category. The same applies when comparing three and two age categories. This may suggest that simpler models for odds ratios, such as uniform association, may be more parsimonious, but there are also deviations from a symmetrical pattern. Especially interesting to observe is that the age boundaries are weaker when people become older. The boundary between adjacent age categories, for example, declines from 7.0 for people in their twenties and thirties, to 3.5 for people in their sixties and fifties. This corresponds to the observation made elsewhere that age is socially more salient when people are young. A similar finding has been obtained in analyses of age homogamy in marriage (Van Poppel et al., 2001).

Can we compare the strength of age and marital status boundaries? As is clear from the table, this comparison will work out differently, depending on which categories one looks at. Boundaries spanning two age categories are stronger than the marital status boundaries and age boundaries between adjacent categories are more or less of the same magnitude as the marital status boundaries. Although we should be cautious in drawing general conclusions here, the results do suggest that age is a more important factor in network formation than marital status.

In Table 5, we present the odds ratios for men and women separately. Network change and composition differ considerably between the genders, which makes it important to explore whether age and marital status have different effects on men and women. When we focus on age first, we see that age boundaries are stronger for women than for men. The marital status boundaries also reveal interesting differences between men and women. The most important difference is that single men are more segregated from the other marital status groups than single women. The other pairs of groups do not reveal clear differences.

4.4. Labeling the latent classes

We now turn to the interpretation of the latent classes. A latent class can be interpreted as a group of ego's who have a tendency to choose alters of a certain kind, independently of their own age and marital status. This tendency is responsible for the association that can exist between alters within egos. In other words, the alters of an ego can be more similar in terms of age and marital status than what one would expect on the basis of ego's age and marital status. The latent classes thus give us information about what we call, 'clustered selection'. The fact that we identified multiple classes is the first and most important conclusion of the latent class analysis because it provides evidence that clustered selection indeed exists. More detailed analyses of this clustering are provided in Tables 6 and 7.

In Table 6, we present the age and marital characteristics of alters in each latent class. These calculations are from the preferred model (Model B5). The first class in the data consists of ego's who have a tendency to pick middle-aged and older persons, independent of their own age and marital status. The second class is the 'thirty-something class:' these are alters who are disproportionally in their thirties, without having any specific marital status profile. The third class

Table 5
Log odds ratios of boundaries between age and marital status groups for men and women

	Men		Women	
	Parameter	S.E.	Parameter	S.E.
One age class differences				
20–29 vs. 30–39	1.88	0.21	1.96	0.15
30–39 vs. 40–49	1.24	0.17	1.66	0.15
40–49 vs. 50–59	1.00	0.15	1.28	0.17
50–59 vs. 60–72	1.13	0.18	1.51	0.20
Two age classes difference				
20–29 vs. 40–49	3.33	0.26	4.84	0.28
30–39 vs. 50–59	3.02	0.23	4.57	0.28
40–49 vs. 60–72	3.10	0.25	3.00	0.21
Three age classes difference				
20–29 vs. 50–59	4.31	0.30	5.92	0.37
30–39 vs. 60–72	4.62	0.31	5.07	0.29
Four age class difference				
20–29 vs. 60–72	6.59	0.43	6.33	0.37
Marital status differences				
Single – married	1.38	0.13	1.19	0.13
Married – divorced	1.13	0.37	1.10	0.32
Divorced – widowed	0.88	0.76	1.03	0.56
Single – divorced	1.46	0.48	1.08	0.42
Single – widowed	1.58	0.62	0.65	0.37
Married – widowed	1.37	0.46	0.75	0.25

Note: based on Model B6.

Table 6
Characteristics of latent classes of ego's

	Class 1	Class 2	Class 3	Class 4	Class 5	Average proportion
Alter's age group						
20–29	0.05	0.19	0.18	0.02	0.25	0.14
30–39	0.09	0.39	0.10	0.19	0.20	0.19
40–49	0.24	0.21	0.07	0.19	0.22	0.18
50–59	0.28	0.09	0.13	0.09	0.16	0.15
60–72	0.34	0.12	0.51	0.52	0.17	0.33
Total	1.00	1.00	1.00	1.00	1.00	
Alter's marital status						
Single	0.13	0.12	0.06	0.21	0.31	0.17
Married	0.83	0.83	0.84	0.72	0.59	0.76
Divorced	0.02	0.02	0.07	0.04	0.08	0.05
Widowed	0.02	0.03	0.03	0.02	0.03	0.03
Total	1.00	1.00	1.00	1.00	1.00	
Size of class	0.25	0.25	0.20	0.17	0.13	

Table 7
Implied log odds ratios of boundaries between alters' ages and marital statuses while controlling for ego's age and marital status

One age class differences	
20–29 vs. 30–39	0.153
30–39 vs. 40–49	0.129
40–49 vs. 50–59	0.084
50–59 vs. 60–72	0.191
Two age classes differences	
20–29 vs. 40–49	0.239
30–39 vs. 50–59	0.406
40-49 vs. 60-72	0.229
Three age classes differences	
20–29 vs. 50–59	0.351
30–39 vs. 60–72	0.439
Four age class differences	
20–29 vs. 60–72	0.400
Marital status differences	
Single – married	0.154
Married – divorced	0.167
Divorced – widowed	0.108
Single – divorced	0.145
Single – widowed	0.127
Married – widowed	0.007

has a disproportionate number of elderly alters and also relatively few single alters. The fourth class looks like the third in terms of age – many elderly – but it also has many single alters. This class is somewhat difficult to interpret. The fifth class is the most special of all the classes in terms of marital status: this clearly is the 'singles' class.' In this class, there is also an overrepresentation of divorced alters.

A second and equally important way to interpret these classes is to look at the partial associations among pairs of alters, given ego's age and marital status. Within classes, the ages and marital status categories of pairs of alters are independent of one another, and as a result, an artificial two-way cross-tabulation of the ages of two alters will reveal independence. However, aggregating such artificial bivariate tables over the classes using the class sizes as weights will reveal the strength of the association between pairs of alters captured by the latent classes, while controlling for ego's age or marital status. These cell entries in the symmetric two-way tables are obtained as follows:

$$\sum_{t=1}^{T} P(X_j = t)\frac{\exp(\beta^{\mathrm{M}}_{pt})}{\sum_{p=1}^{4}\exp(\beta^{\mathrm{M}}_{pt})}\frac{\exp(\beta^{\mathrm{M}}_{p't})}{\sum_{p'=1}^{4}\exp(\beta^{\mathrm{M}}_{p't})},$$

$$\sum_{t=1}^{T} P(X_j = t)\frac{\exp(\beta^{\mathrm{A}}_{qt})}{\sum_{q=1}^{5}\exp(\beta^{\mathrm{A}}_{qt})}\frac{\exp(\beta^{\mathrm{A}}_{q't})}{\sum_{q'=1}^{5}\exp(\beta^{\mathrm{A}}_{q't})},$$

for marital status and age, respectively. Because we are primarily interested in the question of whether there is an association between alters, we present the log odds ratios in these tables in Table 7. Focusing first on age, we see that all log odds ratios are positive, indicating that the ages of alters are positively related. Most log odds ratios are small, however, especially in comparison with the odds ratios in Table 4. In other words, there is a weak residual association between alters' ages. A similar conclusion can be drawn for marital status categories. Alters are more alike in marital status than would be expected solely based on their link to ego, but the association is small.

5. Conclusion

The most important substantive finding of this paper is that marital status categories serve as boundaries in social networks. We made a distinction between single persons, married (or cohabiting) persons, divorced persons, and widowed persons, and showed that these groups interact less with each other than one would expect. These boundaries are to some extent due to the role of age – confirming the by-product hypothesis – but even independent of age, we find significant marital status boundaries. The role of marital status in the formation of networks has been suggested by several authors in the past (Gerstel, 1988; Milardo and Allan, 2000) but to our knowledge, it has not been demonstrated convincingly. Classic studies have looked at marital status (Fischer, 1982) but have not analyzed age and marital status simultaneously, thereby ignoring the important role of the by-product hypothesis.

There are several possible explanations of marital status homogeneity. First, there is the distinction between selection and causation. Selection produces homogeneity because people select others of the same marital status and because they end relationships with people who have a different marital status (positive and negative selection). Causation also produces homogeneity because people may influence each other's life course transitions. If the demographic choices a person makes are influenced by the demographic transitions that occur in his or her network – if demographic transitions are contagious – this will lead to network homogeneity by marital status. Second, there is the important distinction between preferences and constraints in network formation (Feld, 1981; Marsden, 1990). Network homogeneity in part may arise from the fact that some of the local settings in which people are embedded are segregated by marital status. Examples are suburban neighborhoods, voluntary organizations, churches and outgoing places. But people may also have preferences for interacting with persons of the same marital status. Interacting with equals may lead to mutual confirmation of one's life course decisions, it may lead to better mutual understanding, and it may also be an important way to share information about one's social role. We have empirically established that there is marital status homogeneity. Future research can focus on the question of why this occurs.

Our second substantive conclusion is that age boundaries are strong as well. Although this is not a new conclusion, it is an important point to establish again since the topic of age segregation in society has received relatively little attention recently (Hagestad and Uhlenberg, 2005). Our analyses also provides some new insights into age segregation. There is some tendency for age boundaries to be stronger when persons are relatively young. Moreover, we have shown that age similarity is also caused to some extent by the fact that network members grow old together. For more recently formed dyads, the correlation between ego's and alter's age is weaker than for all dyads. In addition, we find that age selection is hardly a by-product of marital status selection. In other words, the role of age in the formation of networks is hardly related to the life course stage that people are in. Finally, we find that age boundaries are stronger for women than they are for men.

Third, we have shown that there is clustered selection. Using latent class models, we were able to show that there is more similarity between alters than one would expect on the basis of ego's own characteristics. Operationally, this meant that there is a discrete variable that has an effect on the ages and marital status positions of alters, independent of the effect of ego on alter's ages and marital status positions. This discrete variable represents latent classes of respondents who tend to have a certain type of network. Although it is difficult to label all these clusters, the analyses indicated that there are people who have a disproportionate number of married persons in their thirties in their network (regardless of their own age and marital status). There is also a group of people who have a high proportion of married people over 60 in their network.

Clustered selection of age and marital status can point to unmeasured preferences that are not captured by the age and marital status of ego. Certain people may have a tendency to prefer interacting with older persons, regardless of their own age. Another, and we think more plausible interpretation of clustered selection lies in the role of opportunities. If people operate in certain homogeneous settings, it is likely that their friends and acquaintances will be more alike. Suburban neighborhoods may for instance, produce a network of young married persons, and this corresponds well with the second latent class in the data. If networks are formed along these lines, there will be an association between the age and marital status of alters, independent of ego's characteristics. A similar form of clustered selection is getting to know the friends of your friends. This will also produce an association between alter's characteristics independent of ego (Yamaguchi, 1990).

The substantive results discussed above were obtained by using new methods on network data that build on the well-established log-linear framework that has been used extensively in research on marital homogamy (Hout, 1982; Kalmijn, 1991; Mare, 1991). We reformulated the model of quasi-symmetry to conditional logit models, thereby obtaining a more flexible model while retaining the meaningful description of marriage boundaries in terms of odds ratios. We applied these models to both bivariate and multivariate data, thereby allowing us to test the by-product hypothesis in a systematic fashion. And most importantly, the model was estimated by a random effects approach which takes into account that in personal network data, network members are clustered within individuals. In addition, the random effects approach was estimated by a latent class model, which currently is the only available method for applying random effects to nonlinear models. The latent class approach yields new substantive results because it translates the dependencies between alters in classes that can be looked at empirically. Although the classes themselves were somewhat difficult to interpret, the implication of clustered selection is a substantive novelty. Hence, our novel approach to random effects has clear advantages to random effects models that merely treat such dependencies as a nuisance.

References

Agresti, A., 2002. Categorical Data Analysis. Wiley, New York.

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

Allan, G., 1979. A Sociology of Friendship and Kinship. Allen and Unwin, London.

Blau, P.M., Schwartz, J.E., 1984. Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations. Academic Press, New York.

Booth, A., Edwards, J.N., Johnson, D.R., 1991. Social integration and divorce. Social Forces 701, 207–224.

Bost, K.K., Cox, M.J., Burchinal, M.R., Payne, C., 2002. Structural and supportive changes in couples' family and friendship networks across the transition to parenthood. Journal of Marriage and Family 64, 517–531.

Broese van Groenou, M.I., Van Tilburg, T.G., 1996. Network analysis. In: Birren, L.E. (Ed.), Ecyclopedia of Gerontology: Age, Aging, and the Aged. Academic Press, San Diego, pp. 197–210.

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

De Jong-Gierveld, J., Dykstra, P.A., 1993. Life transitions and the network of personal relationships: methodological and theoretical issues. In: Jones, W.H., Perlman, D. (Eds.), Advances in Personal Relationships, vol. 4. Kingsley, London, pp. 195–227.

Feld, S.L., 1981. The focused organization of social ties. American Journal of Sociology 86, 1015–1035.

Fischer, C.S., 1982. To Dwell among Friends: Personal Networks in Town and City. University of Chicago Press, Chicago.

Gerstel, N., 1988. Divorce, gender and social integration. Gender and Society 2, 343–367.

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.

Hagestad, G., Uhlenberg, P., 2005. The social separation of old and young: a root of ageism. Journal of Social Issues 61, 343–360.

Hendrickx, J., 1998. Religious and educational assortative marriage patterns in The Netherlands 1940–1985. Netherlands' Journal of Social Sciences 34, 5–22.

Henkens, K., Kraaykamp, G., Siegers, J., 1993. Married couples and their labour market status: a study of the relationship between the labour market status of partners. European Sociological Review 9 (1), 67–78.

Hout, M., 1982. The association between husbands' and wives' occupations in two-earner families. American Journal of Sociology 88 (2), 397–409.

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 (1), 64–82.

Hurlbert, J.S., Acock, A.C., 1990. The effects of marital status on the form and composition of social networks. Social Science Quarterly 71, 163–174.

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

Kalmijn, M., 1993. Trends in black/white intermarriage. Social Forces 72, 119–146.

Kalmijn, M., 2003. Friendship networks over the life course: a test of the dyadic withdrawal hypothesis using survey data on couples. Social Networks 25, 231–249.

Kalmijn, M., Flap, H., 2001. Assortative meeting and mating: unintended consequences of organized settings for partner choices. Social Forces 79, 1289–1312.

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.

Knipscheer, C.P.M., De Jong Gierveld, J., Van Tilburg, T.G., Dykstra, P.A. (Eds.), 1995. Living Arrangements and Social Networks of Older Adults. VU University Press, Amsterdam.

Lazarsfeld, P.F., 1950. The logical and mathematical foundation of latent structure analysis and the interpretation and mathematical foundation of latent structure analysis. In: Stouffer, S.A. (Ed.), Measurement and Prediction. Princeton University Press, Princeton, NJ, pp. 362–472.

Lazarsfeld, P.F., Merton, R.K., 1954. Friendship as social process: a substantive and methodological analysis. In: Berger, M. (Ed.), Freedom and Control in Modern Society. Van Nostrand, New York, pp. 18–66.

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

Mare, R.D., 1991. Five decades of educational assortative mating. American Sociological Review 56, 15–32.

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: Calhoun, C., Meyer, M.W., Richard, W. (Eds.), Structures of Power and Constraint. Cambridge University Press, Cambridge, pp. 397–410.

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

McPherson, J.M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annual Review of Sociology 27, 415–444.

Milardo, R.M., Allan, G., 2000. Social networks and marital relationships. In: Milardo, R.M., Duck, S. (Eds.), Families as Relationships. John Wiley and Sons, Chichester, pp. 117–134.

Munch, A., Miller McPherson, J., Smith-Lovin, L., 1997. Gender, children, and social contact: the effects of childrearing for men and women. American Sociological Review 62, 509–520.

Qian, Z., 1997. Breaking the racial barriers: variations in interracial marriage between 1980 and 1990. Demography 34, 263–276.

Skrondal, A., Rabe-Hesketh, S., 2004. Generalized latent variable modeling: multilevel. In: Longitudinal and Structural Equation Models. Chapman & Hall/CRC, London.

Suitor, J.J., Wellman, B., Morgan, D.L., 1997. It's about time: how, why, and when networks change. Social Networks 19, 1–7.

Ultee, W., Dessens, J., Jansen, W., 1988. Why does unemployment come in couples? an analysis of (un)employment and (non)employment homogamy tables for Canada, The Netherlands and the United States in the 1980s. European Sociological Review 4, 111–122.

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 der Poel, M.G.M., 1993. Delineating personal support networks. Social Networks 15, 49–70.

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., 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., Magidson, J., 2003. Latent GOLD Choice User's Guide. Statistical Innovations, Belmont, MA.

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.

Yamaguchi, K., 2000. Multinomial logit latent-class regression models: an analysis of the predictors of gender-role attitudes among Japanese women. American Journal of Sociology 105, 1702–1740.