RIETI Discussion Paper Series 12-E-048

New Regression Models with Egocentric Social Network Data: An analysis of political party preference in Japan

YAMAGUCHI Kazuo
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Papers Series 12-E-048
July 2012

New Regression Models with Egocentric Social Network Data: An analysis of political party preference in Japan[1]

YAMAGUCHI Kazuo
University of Chicago/RIETI

Abstract

This paper introduces some regression models for a categorical dependent variable with data from egocentric social networks as covariates to analyze the determinants of the outcome for the subject and those of the extent of agreement or disagreement between the outcomes for the subject and the persons to whom he/she is directly connected. The merits of those models are their wide applicability to surveys that collect data on egocentric social networks and their capacity to identify the determinants of agreement between the subject's and his/her friends' attitudinal or behavioral outcomes, controlling for the tendency to agree due to homophily in the choice of friends. An illustrative application using data from the 2004 Japanese General Social Survey shows that several substantively distinct characteristics of egocentric networks affect the subject's political party preference and the extent of agreement in the preference between the subject and that of his/her significant others.

Keywords: Discrete regression, Social network, Homophily, Voting behavior
JEL classification: C25, D72, D85

[1] This study was partly supported by the Research Institute of Economy, Trade and Industry in Japan where I was a visiting fellow during my sabbatical leave in 2011-12.

SOME REGRESSION MODELS WITH EGOCENTRIC SOCIAL NETWORK DATA AS COVARIATES

1. INTRODUCTION

Surveys sometimes collect data using the same set of variables to characterize subjects and their significant others such as their spouses and friends, and an additional set of variables that characterize the qualities of relationships between the subjects and those significant others. The survey may also collect information about whether those significant others are one another's friends or acquaintances. The form of connections and the substance of relations for a particular person and the people socially connected with him/her are characteristics of an *egocentric* social network.

Egocentric network data differ from global, or "sociocentric," network data, because their observations can be assumed to be independent across persons if they are randomly sampled in the survey. The only nonindependence we need to assume is interdependence among social choices, or social ties, of each person. On the other hand, for sociocentric data, even for asymmetric data on social choices, rather than symmetric data on social ties, we need to assume interdependence of social choices across persons because of the tendency toward reciprocity of social choices and the nonrandom presence of a direct linkage between persons connected indirectly through various forms of two-step linkages as they are modeled parametrically in the p* model (Wasserman and Pattison 1996) .

There are rich histories of substantive research based on the analysis of egocentric social network data. In his Detroit-area study, Laumann (1973) used a survey to analyze egocentric social networks to clarify the multidimensional structure of social stratification and differentiation by identifying "social distance" among "social positions" revealed from the survey subjects' choice of friends. There, to the extent to which the attributes of chosen friends, such as occupations, are similar, the social positions of choosing subjects were considered closer in social distance. In his Northern California study, Fischer (1982) also focused on egocentric networks to clarify how friendship networks differ, for example, depending on education, lifecycle stage, and gender and discussed the meaning of urbanism from the viewpoint of friendship ties, by showing differences in friendship networks between people living in cities and those living in small towns – though he acknowledged that the results are partly consequences of self-selective migrations to cities. By introducing his path-breaking notion of "the strength of weak ties," Granovetter (1973, 1974) clarified how the form of egocentric networks is related to different labor market opportunities, and Burt (1992) elaborated Granovetter's idea of the structural cause of nonredundancy in information inflow to the actor, and introduced the notion of "structural holes" as an advantaged characteristic of egocentric social networks in his analysis of markets. Burt also played a key role in introducing into the General Social Survey (GSS) items related to egocentric social networks in 1986, and the items were also included the GSS in some later years.

Despite the richness of those and other substantive studies based on the analysis of egocentric social networks, there is a striking lack of development in statistical methods for the analysis of egocentric social network data. For example, the most comprehensive textbook on the models and methods for the analysis of social networks, by Wasserman and Faust (1994), does not include any chapter dedicated to the analysis of egocentric social networks. Among the few examples of methods specifically developed for the analysis of egocentric social network data are Laumann's (1973) multidimensional scaling method based on social proximity measures derived from survey subjects' choice of friends, Marsden's (2002) study of centrality measures, including those of egocentric networks, Yamaguchi's (1990) elaborations of loglinear and log-multiplicative association models for the analysis of homophily and social distance in friendship choice, and Snijder and his associates' use of the multilevel model for analyzing relational outcomes (Snijders et al. 1995; Van Duijn et al. 1999). If we include the longitudinal data analysis of dyads, various cross-lagged structural equation models have also been employed for the analysis of selection and socialization (e.g. Yamaguhi and Kandel 1993; Giletta et al. 2011). While multilevel analysis is the analysis of egocentric networks most closely related to the method introduced in this paper, the form of its regression equation is very different. The dependent variable in the multilevel model is a relational outcome. On the other hand, the dependent variable of the models introduced in this paper is an individual outcome, while it simultaneously tries to identify the determinants of a particular relational outcome, namely, similarity between subjects and their friends in the outcome. This difference gives a unique advantage to the models introduced in this paper for controlling homophily in the choice of friends in identifying the determinants of the relational outcome, as described later.

When we consider an analysis of the extent and determinants of similarity in attitude or behavior between subjects and their friends, two kinds of important methodological issues exist for which statistical modeling is useful. One is the issue of the separation of causal effects from selection effects regarding similarity between the subject and his/her significant other (Yamaguchi and Kandel 1997). Selection effects on similarity in attitude or behavior may occur because people may choose similar others as friends based on education, race/ethnicity, sex, age, region of residence, and so forth, and those attributes may also affect their attitude and behavior. Similarity between subjects and their friends in attitude or behavior caused by such a selection mechanism differs from similarity generated by dyadic relational characteristics such as frequency of contact, closeness, duration of friendship, and the number of common friends. Similarity generated by homophily in the choice of friends reflects selection bias, while we may be interested in identifying the determinants of similarity net of such selection bias. We will refer to in this paper selection bias due to homophily based on observable attributes as *selection bias due to observable homophily*. However, people may also choose similar others based on such factors as values and tastes, which are difficult to measure exactly or completely, and those tastes and values will also affect attitude and behavior. Hence, *selection bias due to unobservable homophily* also exists.

This paper first describes models for data from a cross-sectional survey that collects information on egocentric social networks. Since items about egocentric social network data have typically been collected in a cross-sectional survey such as the GSS, those models will be applicable widely. They models permit control for selection bias caused by observable homophily but not that caused by unobservable homophily. Hence, their control for selection bias is limited. The paper, however, describes an extension of those models for panel data analysis that permits control for selection bias due to unobservable as well as observable homophily. An illustrative application is presented only for an analysis of cross-sectional survey data, however.

The second related issue is the problem of endogeneity in the use of friends' characteristics and relational characteristics in predicting the subject's attitudinal or behavioral outcome. The endogeneity of covariate X in the *linear* regression model is expressed as the presence of a correlation between X and the error term of the dependent variable Y. This correlation occurs from various causes including, but not limited to, (1) simultaneity, where an unpredicted change in Y, reflected by a change in the error term of Y, affects X, and (2) omitted-variable bias, which is selection bias in the effect of X on Y caused by uncontrolled common antecedents of X and Y. The use of the standard maximum likelihood estimation as well as the OLS yields inconsistency in the estimated effect of X on Y, and as a result in the effects of other covariates not independent of X as well. In econometric models, it is typical to employ an instrumental variable to solve this inconsistency issue.

On the other hand, the method and models introduced in this paper escape inconsistency in parameter estimates caused by the simultaneity problem – while the solution for the issue of omitted-variable bias depends on the availability of panel survey data and the use of extended models with such data. The models introduced in this paper with a binary outcome variable with logit link function assume symmetry between the effect of the subject's outcome on friends' outcomes and the effect of friends' outcomes on the subject's outcome, and express this symmetric effect in the logit equations as the association of outcomes, and this modeling of mutual effects escapes the simultaneous issue. Under this assumption, the interdependence, or association, between the subject's and friends' outcomes for a binary outcome is expressed by the partial odds ratio between the subject's outcome and friends' outcomes and its estimate is consistent. Hence, if the assumption of symmetry can be considered reasonable for a given data set, we can consider that the models introduced in this paper have a merit of escaping from one troublesome inconsistency issue in handling interdependent variables in the regression analysis.

In this paper, I first introduce regression models with a dichotomous dependent variable with characteristics of the subject's egocentric social network as covariates, extend them for a dependent variable with ordered categories, and then briefly describe an extension of the models with a dichotomous dependent variable for panel data analysis. As I will show, those models take the form of specific logit and adjacent logit regression models and employ a function of the outcomes for significant others and relational characteristics as covariates, and identify simultaneously (1) the determinants of individual outcome and (2) the determinants of the extent of agreement between the subject's and his/her significant others' outcomes. The paper also briefly discusses how those models differ from (1) the multilevel model for the analysis of egocentric data and (2) linear regression models with egocentric social network data as covariates.

The models introduced in this paper assume independence of observations among subjects. Hence, they should not be applied to data for which this assumption does not hold, such as the data on egocentric social networks extracted from sociocentric social network data.

An application with data from Japanese General Social Survey focuses on the determinants of the subject's support for the Liberal Democratic Party in 2004, and the determinants of the strength of agreement between the political attitudes of the subject and his/her significant others.

2. REGRESION MODELS WITH A DICHOTOMOUS DEPENDENT VARIABLE WITH EGOCENTRIC SOCIAL NETWORK DATA AS COVARIATES

2.1. A note on the equivalence of two models

Shrout and Kandel (1981) introduced the following special multinomial logit model for a simultaneous analysis of the determinants of the subject's behavior, his/her best friend's behavior, and the association between the two using a dichotomous distinction of marijuana use versus nonuse as the behavioral outcome. They first expressed the joint probability distribution P_{ij}^{SF} of a pair of dichotomous variables (y^S, y^F) for the subject's and his/her best friend's use versus nonuse of marijuana according to the loglinear saturated model such that

$$\log(P_{ij}^{SF}) = \lambda + 0.5\lambda_i^S + 0.5\lambda_j^F + 0.25\lambda_{ij}^{SF} . \qquad (1)$$

They assumed a deviation contrast for parameters here such that $\lambda_2^S = -\lambda_1^S$, $\lambda_2^F = -\lambda_1^F$, $\lambda_{12}^{SF} = \lambda_{21}^{SF} = -\lambda_{11}^{SF}$, and $\lambda_{22}^{SF} = -\lambda_{11}^{SF}$. With this parameterization, it follows that the estimate of parameter λ_1^S is equal to the average log odds of $y^S = 1$ versus $y^S = 2$ by holding y^F constant, that is, $\left(\log(P_{11}^{SF} / P_{21}^{SF}) + \log(P_{12}^{SF} / P_{22}^{SF})\right)/2$, the estimate of parameter λ_1^F, is equal to the average log odds of $y^F = 1$ versus $y^F = 2$ by holding y^S constant, that is, $\left(\log(P_{11}^{SF} / P_{12}^{SF}) + \log(P_{21}^{SF} / P_{22}^{SF})\right)/2$, and the estimate parameter λ_{11}^{SF} is equal to the log odds ratio for the association between the two variables, that is, $\log(P_{11}^{SF} P_{22}^{SF} / P_{12}^{SF} P_{21}^{SF})$.

Then they expressed those three parameters as functions of covariates such that

$$\lambda_1^S(\mathbf{x}) = \alpha_1 + \boldsymbol{\beta}_1'\mathbf{x}\,,\;\; \lambda_1^F(\mathbf{x}) = \alpha_2 + \boldsymbol{\beta}_2'\mathbf{x}\,, \text{ and } \lambda_{11}^{SF}(\mathbf{x}) = \alpha_3 + \boldsymbol{\beta}_3'\mathbf{x}\,, \qquad (2)$$

where $\boldsymbol{\beta}_1'$, $\boldsymbol{\beta}_2'$, and $\boldsymbol{\beta}_3'$ were row vectors of parameters, and $\mathbf{x}$ was the column vector of covariates. These simultaneous regressions identified the determinants of the subject's marijuana use independent of his/her friend's use, those of the friend's marijuana use independent of the subject's use, and those of the strength of association between the subject's and the friend's marijuana use. Although Shrout and Kandel (1981) employed a common set of covariates across three regression equations, they found that only the subject's attributes affected $\lambda_1^S(\mathbf{x})$, the corresponding friend's attributes affected $\lambda_1^F(\mathbf{x})$, and only characteristics of relations between the subject and the friend affected $\lambda_{11}^{SF}(\mathbf{x})$. We take those characteristics of covariate effects into account in the models introduced below.

It can be easily shown that the model of equations (1) and (2) is a reparameterization of the standard multinomial logit model for a dependent variable with four categories. Hence, there is no problem of inconsistent parameter estimates here – though we are making a symmetry assumption between the subject's effect on the friend's outcome and the friend's effect on the subject's outcome, in addition to the assumption of the independence of irrelevant alternatives (IIA), as in the case of the standard multinomial logit model.

However, this formulation becomes rather clumsy if we extend the model to data with two or more friends. For two friends, for example, we will need six simultaneous equations even if we ignore the three-factor interactions of y^S, y^{F_1}, and y^{F_2}, because we will have three equations for predicting the subject's outcome and each of the two friends' outcomes, and three additional equations, two for predicting the association between the outcomes of the subject and each friend and one for the association between the two friends' outcomes. The number of equations increases further for data with three friends.

However, we may be interested only in the determinants of the subject's outcome and the determinants of the association between the subject's and the friends' outcomes, and not interested either in the determinants of the friends' outcomes or in the determinants of the association among friends' outcomes. Our interests may be that specific because we can expect the determinants of a given friend's outcome to be similar in characteristics to those of the subject's outcome, except that the corresponding friend's attributes, rather than the subject's attributes, affect the friend's outcome. Similarly, we can expect that the determinants of the association in the outcome between two friends with a tie will be similar to those of the association between the outcomes of the subject and one of his/her friends.

Even though the determinants of the association between the outcomes of friends *without* a tie will be different, we can reasonably guess that the association will be explained either as selection bias due to similarity between the two friends in the attributes that affect individual outcomes or as the effect of indirect ties with their common friends including the subject, since no direct relation exists. This mechanism will be of much less interest to many researchers than the mechanism of determining similarity in the outcome between persons with a tie.

On the other hand, a restriction of the analysis to the determinants of the subject's outcome and those of the association between the subject's and friends' outcomes provides a great benefit of simplification, because it leads to the use of a single regression equation, as shown below, and the model can be extended easily for the analysis of data with two or more friends. Hence, we adopt this methodological strategy in what follows.

For data with a single friend, we can obtain a simple logistic regression model that retains only parameters λ_i^S and λ_{ij}^{SF} by taking the difference in equation (1) between $\log(P_{1|j}^{S|F}) \equiv \log\left(P_{1j}^{SF} / (P_{1j}^{SF} + P_{2j}^{SF})\right)$ and $\log(P_{2|j}^{S|F}) = \log(1 - P_{1|j}^{S|F})$ such that

$$\log\left(\frac{P_{1|j}^{S|F}(\mathbf{x})}{1 - P_{1|j}^{S|F}(\mathbf{x})}\right) = \lambda_1^S(\mathbf{x}) + 0.5\lambda_{1j}^{SF}(\mathbf{x}), \qquad (3)$$

where $\lambda_1^S(\mathbf{x}) = \alpha_1 + \boldsymbol{\beta}_1{}'\mathbf{x}$, $\lambda_{11}^{SF}(\mathbf{x}) = \alpha_3 + \boldsymbol{\beta}_3{}'\mathbf{x}$, $\lambda_{12}^{SF}(\mathbf{x}) = -\lambda_{11}^{SF}(\mathbf{x})$.

The multinomial logit model described above and the logit model of equation (3) are basically the same except that the latter model removes constraints imposed by the equation for the friend's outcome. Two important facts here are that (1) parameter $\lambda_{11}^{SF}(\mathbf{x})$ in equation (3) still characterizes the log odds ratio between y^S and y^F, and is in this sense symmetric between the subject and the friend, and this fact requires caveats in the use of its covariates, as discussed below, and (2) this characteristic of $\lambda_{11}^{SF}(\mathbf{x})$, despite the fact that it appears in the single regression form as if it represented the effect of y^F on y^S, indicates that this regression equation does not generate the problem of inconsitent parameter estimates – unlike the estimate for the effect of y^F on y^S in the linear regression when y^S affects y^F. In sections 2.3, 2.4, and 2.5, I describe models for cases with two or more friends.

2.2. A reformulation for the case of a dichotomous dependent variable with a single friend

First let's express equation (3) in a more conventional way. First we obtain the following logistic regression by replacing $\lambda_1^S(\mathbf{x})$ with $\alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S$, and $\lambda_{11}^{SF}(\mathbf{x})$ with $\beta_0 + \boldsymbol{\beta}_1{}'\mathbf{z}^{SF}$, assuming that only the subject's characteristics $\mathbf{x}^S$ affect $\lambda_1^S(\mathbf{x})$ and only the subject's and the friend's relational characteristics $\mathbf{z}^{SF}$ affect $\lambda_{11}^{SF}(\mathbf{x})$:

$$\log\left(\frac{P^{S|F}}{1-P^{S|F}}\right) = \alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \beta_0 D^F + \boldsymbol{\beta}_1{}'\mathbf{z}^{SF} D^F \,, \qquad (4)$$

where $D^F \equiv 1.5 - y^S$ is a standardized dummy variable with a value of 0.5 for $y^F = 1$ and −0.5 for $y^F = 2$, depending on the friend's outcome.

Equation (4) indicates that the effects of covariates on the log odds ratio between y^S and y^F can be expressed as the interaction effects of D^F and $\mathbf{z}^{SF}$ on the odds of $y^S = 1$ versus $y^S = 2$. Two characteristics of equation (4) are important here. First, equation (4) does not include the "main effects" of $\mathbf{z}^{SF}$. This is because covariates $\mathbf{z}^{SF}$ are assumed to affect only the association between the subject's and friend's outcomes, $\lambda_{11}^{SF}(\mathbf{x})$, and thereby modify only the effects of D^F. Due to the absence of main effects of $\mathbf{z}^{SF}$, coding D^F as 0.5 and -0.5 rather than 1 and 0 is essential here.[2] Second, since the coefficients of $\mathbf{z}^{SF}$, $\boldsymbol{\beta}_1{}'$, in equation (4) indicate the effects on the log odds ratio of y^S and y^F, $\mathbf{z}^{SF}$ must be variables that are symmetric in characterizing the subject and the friend, including variables that characterize the quality of relations between the subject and the friend such

[2] With a dummy variable coding, we would obtain a model that hypothesizes that $\mathbf{z}^{SF}$ affect the odds of y^S=1 versus y^S=2 only for $y^F = 1$, and this does not make sense.

as the extent of intimacy, frequency of contacts, and so on, variables $(x^S + x^F)/2$ or $|x^S - x^F|$ for a pair of corresponding interval scale variables x^S and x^F, and variable $\delta(x^S, x^F)$, which takes a value of 1 if and only of $x^S = x^F$ and takes a value of 0 otherwise for a pair of corresponding categorical variables. Note that since we control for the effect of x^S on the subject's outcome, and the effect of x^F on the friend's outcome is implicitly taken into account because of unrestricted relations among covariates, the effects of such variables as $(x^S + x^F)/2$, $|x^S - x^F|$, or $\delta(x^S, x^F)$ on $\lambda_{11}^{SF}(\mathbf{x})$ reflect their effects net of homophily in the choice of friends based on x.

The coefficients $\boldsymbol{\beta}_1{}'$ should not be interpreted as the effects of the friend's outcome on the subject's outcome but should be interpreted as the effects on the association between the outcomes of the subject and the friend, and therefore a causal direction between y^S and y^F is not assumed and cannot be determined.

2.3. The case of a dichotomous dependent variable with two friends

Similar to the case with a friend, the saturated loglinear model for the joint probability distribution of outcomes for cases with two friends is given as

$$\log(P_{ijk}^{SF_1F_2}) = \lambda + 0.5\lambda_i^S + 0.5\lambda_j^{F_1} + 0.5\lambda_k^{F_2} + 0.25\lambda_{ij}^{SF_1} + 0.25\lambda_{ik}^{SF_2} + 0.25\lambda_{jk}^{F_1F_2} + 0.125\lambda_{ijk}^{SF_1F_2}, \qquad (5)$$

with a standard set of deviation contrasts for each factor of lambda parameters such that $\lambda_2^S = -\lambda_1^S$, $\lambda_2^{F_1} = -\lambda_1^{F_1}$, $\lambda_2^{F_2} = -\lambda_1^{F_2}$, $\lambda_{12}^{SF_1} = \lambda_{21}^{SF_1} = -\lambda_{11}^{SF_1}$, $\lambda_{22}^{SF_1} = \lambda_{11}^{SF_1}$, $\lambda_{12}^{SF_2} = \lambda_{21}^{SF_2} = -\lambda_{11}^{SF_2}$, $\lambda_{22}^{SF_2} = \lambda_{11}^{SF_2}$, $\lambda_{112}^{SF_1F_2} = \lambda_{121}^{SF_1F_2} = \lambda_{211}^{SF_1F_2} = \lambda_{222}^{SF_1F_2} = -\lambda_{111}^{SF_1F_2}$, and

$\lambda_{122}^{SF_1F_2} = \lambda_{212}^{SF_1F_2} = \lambda_{221}^{SF_1F_2} = \lambda_{111}^{SF_1F_2}$. By taking the log odds of $P_{1|jk}^{S|F_1F_2} \equiv \left(P_{1jk}^{SF_1F_2} \big/ (P_{1jk}^{SF_1F_2} + P_{2jk}^{SF_1F_2}) \right)$ versus $P_{2|jk}^{S|F_1F_2} = 1 - P_{1|jk}^{S|F_1F_2}$, we obtain

$$\log\left(\frac{P_{1|jk}^{S|F_1F_2}}{1 - P_{1|jk}^{S|F_1F_2}} \right) = \lambda_1^S + 0.5\lambda_{1j}^{SF_1} + 0.5\lambda_{1k}^{SF_2} + 0.25\lambda_{1jk}^{SF_1F_2} . \qquad (6)$$

By hypothesizing that parameters λ_1^S, $\lambda_{11}^{SF_1}$, and $\lambda_{11}^{SF_2}$ depend on covariates in the same way as the case with one friend such that

$$\lambda_1^S(\mathbf{x}^S) = \alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S \text{, and} \qquad (7)$$

$$\lambda_{11}^{SF_1}(\mathbf{z}^{SF_1}) = \beta_0 + \boldsymbol{\beta}_1\mathbf{z}^{SF_1} \text{ and } \lambda_{11}^{SF_2}(\mathbf{z}^{SF_2}) = \beta_0 + \boldsymbol{\beta}_1\mathbf{z}^{SF_2} \ , \qquad (8)$$

and by replacing parameter $\lambda_{111}^{SF_1F_2}$ with γ, we obtain

$$\log\left(\frac{P^{S|F_1F_2}}{1 - P^{S|F_1F_2}} \right) = \alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \beta_0(D^{F_1} + D^{F_2}) + \boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}D^{F_1} + \mathbf{z}^{SF_2}D^{F_2}) + \gamma D^{F_1}D^{F_2}, \quad (9)$$

where D^{F_1} and D^{F_2} are both coded 0.5 and −0.5, depending on friend 1's and friend 2's outcomes, respectively. By employing the same set of parameters $\boldsymbol{\beta}_1{}'$ for $\mathbf{z}^{SF_1}$ and $\mathbf{z}^{SF_2}$ in equation (8), we are making a simplifying assumption that the effects of those corresponding covariates on the association between the outcomes of the subject and a friend are the same for the two friends. Similar to equation (4), parameters $\boldsymbol{\alpha}_1$ represent the effects of the subject's characteristics on the outcome, parameter β_0 indicates the extent of baseline common association between the outcomes of the subject and each friend when $\mathbf{z}^{SF_1} = \mathbf{z}^{SF_2} = 0$, and parameters $\boldsymbol{\beta}_1{}'$ represent the effects of relations $\mathbf{z}^{SF}$ between the subject and friends on the extent of the association between the subject's outcome and friends' outcomes.

Parameter γ represents the interaction effects of friend 1's and friend 2's outcomes on the log odds of $y^S = 1$ versus $y^S = 2$. When γ >0, the agreement between two friends' outcomes increases the odds of $y^S = 1$ versus $y^S = 2$, and the opposite holds when γ <0. It seems that the estimate of γ is usually nonsignificant, and in such cases, parameter γ is better omitted from the model, because an inclusion of nonsignificant interaction effects in the model weakens the power to reveal significance in the main effects.

We may also hypothesize that the presence versus absence of a tie between friends may affect (a) the subject's outcome and (b) the extent of the association between the outcomes of the subject and each friend. The model that incorporates those two factors is as follows.

$$\log\left(\frac{P^{S|F_1F_2}}{1-P^{S|F_1F_2}}\right) = \alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \alpha_2 r^{F_1F_2} + \beta_0(D^{F_1} + D^{F_2}) + \boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}D^{F_1} + \mathbf{z}_m^{SF_2}D^{F_2})$$
$$+ \beta_2 r^{F_1F_2}(D^{F_1} + D^{F_2}) + \gamma D^{F_1}D^{F_2}. \qquad (10)$$

The inclusion of factor $\alpha_2 r^{F_1F_2}$ in equation (10) means that we regard $r^{F_1F_2}$, which takes a value of 1 when friends 1 and 2 have a tie and a value of zero if no such tie exists, as if it were one of the subject's individual attributes $\mathbf{x}^S$, because $r^{F_1F_2}$ is a measure of local network density around the subject. When $\alpha_2 > 0$ and is significant, those who have a tie between their two friends tend to have a greater probability of $y^S = 1$, and the opposite holds if $\alpha_2 < 0$ and is significant. The inclusion of $\beta_2 r^{F_1F_2}(D^{F_1} + D^{F_2})$ in equation (10) means that we also regard $r^{F_1F_2}$ as if it were a common characteristic of relations $\mathbf{z}^{SF_1}$ between the subject and friend 1 and of relations $\mathbf{z}^{SF_2}$ between the subject and friend 2, because $r^{F_1F_2}$ indicates the presence versus the absence of an indirect tie between the subject and each friend through ties with another friend. If $\beta_2 > 0$ and is significant, having an indirect tie between the subject and a friend increases the extent of the association between the outcomes of the subject and each friend, and the opposite holds if $\beta_2 < 0$ and is significant.

2.4. The case of a dichotomous dependent variable with three friends

Similarly, with three friends, we have the following equation:

$$\begin{aligned}\log\left(\frac{P^{S|F_1F_2F_3}}{1-P^{S|F_1F_2F_3}}\right) &= \alpha_0 + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \alpha_2(r^{F_1F_2} + r^{F_1F_3} + r^{F_2F_3})/3 \\ &+ \beta_0(D^{F_1} + D^{F_2} + D^{F_3}) + \boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}D^{F_1} + \mathbf{z}^{SF_2}D^{F_2} + \mathbf{z}^{SF_3}D^{F_3}) + \\ &+ \beta_2\{r^{F_1F_2}(D^{F_1} + D^{F_2}) + r^{F_1F_3}(D^{F_1} + D^{F_3}) + r^{F_2F_3}(D^{F_2} + D^{F_3})\} \\ &+ \gamma(D^{F_1}D^{F_2} + D^{F_1}D^{F_3} + D^{F_2}D^{F_3}). \qquad (11)\end{aligned}$$

Parameter α_2 in equation (11) generalizes the corresponding effect in equation (10) and tests whether a person with a higher local density of his/her egocentric network, that is, a higher proportion of each subject's friends' being friends among themselves, tends to have a higher probability of $y^S = 1$.

Since

$$\begin{aligned}&r^{F_1F_2}(D^{F_1} + D^{F_2}) + r^{F_1F_3}(D^{F_1} + D^{F_3}) + r^{F_2F_3}(D^{F_2} + D^{F_3}) \\ &= (r^{F_1F_2} + r^{F_1F_3})D^{F_1} + (r^{F_1F_2} + r^{F_2F_3})D^{F_2} + (r^{F_1F_3} + r^{F_2F_3})D^{F_3},\end{aligned}$$

parameter β_2 in equation (11) also generalizes the corresponding parameter in equation (10) and tests whether the number of indirect ties between the subject and each friend through the other two friends increases or decreases the extent of the association between the subject's and friends' outcomes.

This formulation can be easily extended for cases with four or more friends by using a similar parameterization, and, therefore, their expressions are omitted.

2.5. The model for data pooled across different numbers of friends

So far, the model is described separately for each given number of friends. While the models are simpler when they are applied to data with the same number of friends, the power of statistical analysis is greater for the use of pooled data across different numbers of friends if we can assume some commonalities in parameters across equations for different number of friends. We may reasonably assume that the effects of the subject's characteristics $\mathbf{x}^S$, $\boldsymbol{\alpha}_1'$, are the same across different numbers of friends, including the case with no friend. Whether other parameters can be made common across different numbers of friends can be determined by the relative goodness of fit with the data achieved with alternative models.

Alternatively, since the model with three friends (equation (11)) does not include any substantively new parameter that is absent from the model with two friends (equation (10)), we may randomly choose two friends from subjects with three or more friends and analyze data with those two friends. We may also add the number of friends as an element of covariates $\mathbf{x}^S$. However, we should replace the effect of $r^{F_1F_2}$ in equation (10) with the effect of the average proportion of ties among each subject's friends rather than a dichotomous distinction between the presence and the absence of a tie between two randomly chosen friends, because the former rather than the latter variable adequately characterizes the subject's local network density. We should also replace the effect of $r^{F_1F_2}(D^{F_1}+D^{F_2})$ for two randomly chosen friends with the variable that indicates the number of indirect ties between the subject and each of those two friends, which is equal to $\sum_{j\neq i} r^{F_iF_j}$ between the subject and friend F_i, as an element of covariates $\mathbf{x}^{SF}$. Hence, a model for the data of two randomly selected friends F_1 and F_2 can be

$$\log\left(\frac{P^{S|F_1F_2}}{1-P^{S|F_1F_2}}\right)=\alpha_0+\boldsymbol{\alpha}_1{}'\mathbf{x}^S+\alpha_2 M_{i\neq j}(r^{F_iF_j})+\alpha_3(n_f-2)+\beta_0(D^{F_1}+D^{F_2})$$
$$+\boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}D^{F_1}+\mathbf{z}_m^{SF_2}D^{F_2})+\beta_2\left((\textstyle\sum_{i\neq 1} r^{F_1F_i})D^{F_1}+(\sum_{i\neq 2} r^{F_2F_i})D^{F_2}\right)+\gamma D^{F_1}D^{F_2}, \quad (12)$$

where n_f indicates the number of friends, and $M(r^{F_iF_j})$ stands for the mean of $r^{F_iF_j}$.

However, if the survey specifies an order in listing significant others, such as that significant others be listed in the order of closeness to the subject, we may apply regression equation (11) with the data for the closest two friends.

3. REGRESSION MODELS WITH A DEPENDENT VARIABLE HAVING ORDERED CATEGORIES

3.1. On the use of uniform association in characterizing covariate effects on association

An extension of the models described above to cases with a dependent variable having ordered categories is straightforward if we assume (a) proportional odds for the effects of $\mathbf{x}^S$ on the outcome, (b) a uniform association (Goodman 1979; Clogg 1982) for covariate effects on the association of y^S and y^F, and (c) the absence of interaction effects of the two friends' outcomes.[3] Then we can retain a similar symmetry between the

[3] An extension to the use of a symmetric row-and-column association model with egocentric network data (Yamaguchi 1990) is possible, but I omit such an extension because it will require specialized software to handle log-multiplicative covariates.

treatment of y^S and y^F for their association. The model, however, may retain the main association between y^S and y^F , rather than covariate effects on the association, to be saturated in order not to make an unnecessarily strong assumption. As an expression of the loglinear model with a group variable, the three-way model for the joint probability distribution of the subject's ordered outcomes, the friend's ordered outcomes, and a group as a common covariate can be described, under the three simplifying assumptions made above as follows:

$$\log(P_{ijk}^{SFG}) = \lambda + \lambda_i^S + \lambda_i^F + \lambda_k^G + \lambda_{ij}^{SF} + s(i)^S \mu_k^G + s(j)^F \nu_k^G + s(i)^S s(j)^F \omega_k^G, \quad (13)$$

where I is the number of categories of y^S or y^F , $s(i)^S \equiv i-(I+1)/2$ when $y^S = i$ and $s(j)^F \equiv j-(I+1)/2$ when $y^F = j$ are standardized integers, and we assume the deviation contrast for λ parameters and parameters μ^G, ν^G, and ω^G.

In the adjacent logit form, this model can be simplified, since $s(i+1)^S - s(i)^S = 1$, as

$$\log(P_{i+1,jk}^{SFG} / P_{i,jk}^{SFG}) = (\lambda_{i+1}^S - \lambda_i^S) + (\lambda_{i+1,j}^{SF} - \lambda_{i,j}^{SF}) + \mu_k^G + s(j)^F \omega_k^G. \quad (14)$$

The four sets of parameters in the right-hand side of equation (14) respectively represent the threshold-specific intercept at the i-th threshold, the association between y^S and y^F at the i-th threshold of y^S , proportional covariate effects on the log odds of having one higher level of y^S , and covariate effects on the uniform association between y^S and y^F .

3.2. The case with a single friend

Suppose that the covariates that affect the adjacent odds, u_k^G in equation (14), are the subject's characteristics, $\mathbf{x}^S$, and the covariates that affect the uniform association between the subject and his/her friend, w_k^G in equation (14), are symmetric characterizations of relations between the subject and the friends, $\mathbf{z}^{SF}$. Then the model of equation (14) can be expressed in a more conventional form with covariates as follows:

$$\log\left(\frac{P_{i+1|j}^{S|F}}{P_{i|j}^{S|F}}\right) = \alpha_{0i} + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \beta_{0,ij} + \boldsymbol{\beta}_1{}'\mathbf{z}^{SF} s(j)^F, \qquad (15)$$

for i =1,…, I-1, and j=1,…, I, $\beta_{0,iI} = -\sum_{j=1}^{I-1}\beta_{0,ij}$ for the last category I of y^F, and $s(j)^F \equiv j-(I+1)/2$ when $y^F = j$.

3.3. The case with two friends

An extension for the case of two friends with common effects of the two friends and without interaction effects of the two friends' outcomes can be given as

$$\log\left(\frac{P_{i+1|jk}^{S|F_1F_2}}{P_{i|jk}^{S|F_1F_2}}\right) = \alpha_{0i} + \boldsymbol{\alpha}_1{}'\mathbf{x}^S + \alpha_2 r^{F_1F_2} + (\beta_{0,ij} + \beta_{0,ik})$$
$$+\boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}s(j)^{F_1} + \mathbf{z}^{SF_2}s(k)^{F_2}) + \beta_2 r^{F_1F_2}(s(j)^{F_1} + s(k)^{F_2}), \qquad (16)$$

for i=1,…, I-1, and j, k = 1,..., I, and $\beta_{0,iI} = -\sum_{j=1}^{I-1}\beta_{0,ij}$. Note that parameters α_2 and parameter β_2 are added in equation (16) for data with two friends as if variable $r^{F_1F_2}$ were one of $\mathbf{x}^S$ and one of $\mathbf{z}^{SF}$, respectively, as in equation (10).

3.4. Ordered categories with three friends and a model with pooled data

For the case with three friends, we obtain by a straightforward extension of equation (16)

$$\log\left(\frac{P_{i+1|jkl}^{S|F_1F_2F_3}}{P_{i|jkl}^{S|F_1F_2F_3}}\right)=\alpha_{0,i}+\boldsymbol{\alpha}_1{}'\mathbf{x}^S+\alpha_2(r^{F_1F_2}+r^{F_1F_3}+r^{F_2F_3})/3+$$
$$+(\beta_{0,ij}+\beta_{0,ik}+\beta_{0,il})+\boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}s(j)^{F_1}+\mathbf{z}^{SF_2}s(k)^{F_2}+\mathbf{z}^{SF_3}s(l)^{F_3})+$$
$$+\beta_3\left\{(r^{F_1F_2}+r^{F_1F_3})s(j)^{F_1}+(r^{F_1F_2}+r^{F_2F_3})s(k)^{F_2}+(r^{F_1F_3}+F^{F_2F_3})s(k)^{F_2}\right\}. \quad (17)$$

As in the dichotomous case, we can analyze data pooled across different numbers of friends with parameters made common across different number of friends if they do not differ significantly. Alternatively, we may randomly choose two friends for cases with three or more friends and add the number of friends to $\mathbf{x}^S$, replace the effect of $r^{F_1F_2}$ as an element of $\mathbf{x}^S$ with the proportion of ties among friends, and replace the effect of $r^{F_1F_2}$ as an element of $\mathbf{z}^{SF_1}$ with the effect of the number of indirect ties between the subject and each friend, such that

$$\log\left(\frac{P_{i+1|jk}^{S|F_1F_2}}{P_{i|jk}^{S|F_1F_2}}\right)=\alpha_{0i}+\boldsymbol{\alpha}_1{}'\mathbf{x}^S+\alpha_2M_{i\neq j}(r^{F_iF_j})+\alpha_2(n_f-2)+(\beta_{0,ij}+\beta_{0,ik})$$
$$+\boldsymbol{\beta}_1{}'(\mathbf{z}^{SF_1}s(j)^{F_1}+\mathbf{z}^{SF_2}s(k)^{F_2})+\beta_2\left((\textstyle\sum_{i\neq 1}r^{F_1F_i})s(j)^{F_1}+(\sum_{i\neq 2}r^{F_2F_i})s(k)^{F_2}\right). \quad (18)$$

4. AN EXTENSION FOR A FIXED-EFFECTS MODEL WITH PANEL SURVEY DATA

I stated in the introduction that models for cross-sectional data permit controls only for selection bias due to observable homophily. However, if we have data from a panel survey that collects information on egocentric social networks, and assume that (1) omitted variables causing selection bias in covariate effects are constant over time, and (2) the outcome variables are dichotomous and conditionally independent over time, controlling for person-specific fixed effects and time-dependent covariates included in the model, we can extend the models introduced above to models with person-specific fixed effects with the conditional likelihood estimation, which enables the estimation of covariate effects controlling for selection bias due unobservable as well as observable confounders.

Below, I will show the extension only to the case with one friend because further extensions for cases with two or more friends are quite straightforward. The extension of the model of equation (4) with person-specific fixed effects and time effects is given as:

$$\log\left(\frac{P_{it}^{S|F}}{1-P_{it}^{S|F}}\right) = \alpha_i + \alpha_t + \boldsymbol{\alpha}_1 ' \mathbf{x}_{it}{}^S + \beta_0 D_{it}^F + \boldsymbol{\beta}_1 ' \mathbf{z}_{it}{}^{SF} D_{it}^F \qquad (19)$$

where $\alpha_{t=1} = 0$, *i* indicates a subject, and *t* indicates time. As is well known for the Rasch model (Rasch 1960), the conditional likelihood estimation using the sum of outcomes over time, $\sum_t y_{it}$, where y_{it} is equal to 0 or 1, as the sufficient statistics gives consistent estimates of structural parameters independent of person-specific effects. In case where we have only two time points, using the sample with $\sum_{t=1}^{2} y_{it} = 1$, the conditional likelihood is given as

$$\begin{aligned} CL &= \prod_i \Pr(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid y_{i1} + y_{i2} = 1) \\ &= \prod_{i|y_{i1}+y_{i2}=1} \frac{\Pr(Y_{i1}=1, Y_{i2}=0)^{y_{i1}} \Pr(Y_{i1}=0, Y_{i2}=1)^{y_{i2}}}{\Pr(Y_{i1}=1, Y_{i2}=0) + \Pr(Y_{i1}=0, Y_{i2}=1)} \\ &= \prod_{i|y_{i1}+y_{i2}=1} \left[\frac{\prod_{t=1}^{2} \exp\left\{ y_{it} (\alpha_t + \boldsymbol{\alpha}_1 ' \mathbf{x}_{it}{}^S + \beta_0 D_{it}^F + \boldsymbol{\beta}_1 ' \mathbf{z}_{it}{}^{SF} D_{it}^F) \right\}}{\sum_{t=1}^{2} \exp(\alpha_t + \boldsymbol{\alpha}_1 ' \mathbf{x}_{it}{}^S + \beta_0 D_{it}^F + \boldsymbol{\beta}_1 ' \mathbf{z}_{it}{}^{SF} D_{it}^F)} \right]. \qquad (20) \end{aligned}$$

Similarly, if we have three time points, the conditional likelihood is given as

$$
\begin{aligned}
CL &= \prod_i \Pr(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}.Y_{i3} = y_{i3} \mid y_{i1} + y_{i2} + y_{i3} = 1) \\
&\times \prod_i \Pr(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3} \mid y_{i1} + y_{i2} + y_{i3} = 2) \\
&= \prod_{i|y_{i1}+y_{i2}+y_{i3}=1} \frac{\Pr(Y_{i1}=1, Y_{i2}=0, Y_{i3}=0)^{y_{i1}} \ \Pr(Y_{i1}=0, Y_{i2}=1, Y_{i3}=0)^{y_{i2}} \ \Pr(Y_{i1}=0, Y_{i2}=0, Y_{i3}=1)^{y_{i3}}}{\Pr(Y_{i1}=1, Y_{i2}=0, Y_{i3}=0) + \Pr(Y_{i1}=0, Y_{i2}=1, Y_{i3}=0) + \Pr(Y_{i1}=0, Y_{i2}=0, Y_{i3}=1)} \\
&\times \prod_{i|y_{i1}+y_{i2}+y_{i3}=2} \frac{\Pr(Y_{i1}=1, Y_{i2}=1, Y_{i3}=0)^{y_{i1}y_{i2}} \ \Pr(Y_{i1}=1, Y_{i2}=0, Y_{i3}=1)^{y_{i1}y_{i3}} \ \Pr(Y_{i1}=0, Y_{i2}=1, Y_{i3}=1)^{y_{i2}y_{i3}}}{\Pr(Y_{i1}=1, Y_{i2}=1, Y_{i3}=0) + \Pr(Y_{i1}=1, Y_{i2}=0, Y_{i3}=1) + \Pr(Y_{i1}=0, Y_{i2}=1, Y_{i3}=1)} \\
&= \prod_{i|y_{i1}+y_{i2}=1} \left[\frac{\prod_{t=1}^{3} \exp\left\{ y_{it} (\alpha_t + \boldsymbol{\alpha}_1 ' \mathbf{x}_{it}^{S} + \beta_0 D_{it}^{F} + \boldsymbol{\beta}_1 ' \mathbf{z}_{it}^{SF} D_{it}^{F}) \right\}}{\sum_{t=1}^{3} \exp(\alpha_t + \boldsymbol{\alpha}_1 ' \mathbf{x}_{it}^{S} + \beta_0 D_{it}^{F} + \boldsymbol{\beta}_1 ' \mathbf{z}_{it}^{SF} D_{it}^{F})} \right] \\
&\times \prod_{i|y_{i1}+y_{i2}=2} \left[\frac{\prod_{(t,s)=(1,2),(1,3),(2,3)} \exp\left\{ y_{it} y_{is} (\alpha_t + \alpha_s + \boldsymbol{\alpha}_1 '(\mathbf{x}_{it}^{S} + \mathbf{x}_{is}^{S}) + \beta_0 (D_{it}^{F} + D_{is}^{F}) + \boldsymbol{\beta}_1 '(\mathbf{z}_{it}^{SF} D_{it}^{F} + \mathbf{z}_{is}^{SF} D_{is}^{F})) \right\}}{\sum_{(t,s)=(1,2),(1,3),(2,3)} \exp((\alpha_t + \alpha_s + \boldsymbol{\alpha}_1 '(\mathbf{x}_{it}^{S} + \mathbf{x}_{is}^{S}) + \beta_0 (D_{it}^{F} + D_{is}^{F}) + \boldsymbol{\beta}_1 '(\mathbf{z}_{it}^{SF} D_{it}^{F} + \mathbf{z}_{is}^{SF} D_{is}^{F}))} \right].
\end{aligned}
$$

(21)

More generally, the conditional likelihood estimation for those fixed-effect logit models can be applied by using XTLOGIT in STATA.

5. NOTES ON SOME RELATED REGRESSION MODELS

5.1 Multilevel models for egocentric social network data

Snijders and his associates (Snijders et al. 1995, Marijtje et al. 1999) introduced the use of multilevel analysis with egocentric network data. This multilevel method considers individuals as the higher level, and individuals' relations with others as the lower level.

Formally, the method consists of the following three equations (Van Duijn et al. 1999):

$$Y_{ij} = \beta_{0j} + \boldsymbol{\beta}_{1j} ' \mathbf{x}_{ij} + \varepsilon_{ij}, \quad \beta_{0j} = \gamma_{00} + \boldsymbol{\gamma}_{01} ' \mathbf{z}_j + \varepsilon_{0j} \ \text{ and } \ \boldsymbol{\beta}_{1j} = \boldsymbol{\gamma}_{10} + \Gamma_{\mathbf{11}} ' \mathbf{z}_j + \boldsymbol{\varepsilon}_{1j}$$

where j indicates an individual, i indicates a relation, and Y_{ij} is the dependent variable that indicates a relation-specific outcome, $\mathbf{x}_{ij}$ is a set of relational-level predictors, and $\mathbf{z}_j$

is a set of individual-level predictors. Greek letters indicate parameters. Although their analysis considered an interval-scale variable for *Y*, the model is easily modified for a dichotomous outcome variable by introducing a logit link function for the multi-level analysis.

The major merit of this model is that it includes error terms at two levels, at the individual level and the relational level, and therefore, compared with a model without the individual-level error term, the model is not likely to underestimate the standard error of individual-level predictors.

The multilevel model, however, is not an alternative to the models introduced in this paper. The dependent variable in the multilevel model is a relational-level outcome, while the dependent variable for the models introduced in this paper is an individual-level outcome and relational-level variables appear in the equation as covariates even though those models simultaneously estimate the effects of covariates on a particular relational variable, namely, the extent of agreement between the subject's and friends' outcomes. The use of the individual-level dependent variable in the models implies that the equation has intrinsically only a one-level error term and therefore cannot be formulated as the multilevel model. In the models introduced in this paper as well as in the multilevel models, however, the interdependence among multiple relations of each person is taken into account. In the case with two friends, the interdependence in the models introduced in this paper is expressed parametrically by certain commonalities between λ^{SF_1} and λ^{SF_2}, while it is expressed by the individual-level error term in multilevel analysis.

Another major difference is that the models introduced in this paper try to eliminate the effects of homophily on the similarity between the subject's and friends outcomes by controlling for the effects of individual attributes on the subject's outcome. The multilevel model, however, considers the effects of individual-level variables only on the relational outcome, and does not control for the effects of individual-level variables on the individual outcome and, therefore, is not intended to eliminate the selection bias due to homophily. Instead, the multilevel model intends to assess the effects of individual-level covariates $\mathbf{z}_j$ on $\overline{Y}_j$, the mean of Y_{ij} averaged across multiple relations of each person *j*, and the interaction effects of individual-level and relational-level variables, $\mathbf{z}_j\mathbf{x}_{ij}$, on Y_{ij}.

The multilevel model with egocentric network data can also be applied to the analysis of many different kinds of relational outcomes, while the models introduced in this paper can be applied only to the analysis of agreement/disagreement in attitude or behavior between the subject and his/her significant others.

5.2 Linear regression models with egocentric network data as covariates

Linear regression models with egocentric social network data as covariates differ from models with a categorical dependent variables because (1) a *symmetric* effect of y^F on y^S and of y^S on y^F should not be interpreted as an association (or correlation) between the two, and (2) the endogeneity problem for the effect of y^F on y^S and the effect of y^S on y^F exists even when we assume a symmetric effect. Although those models are not of central concern in this paper, I will briefly describe their characteristics below.

Let us consider the simplest case, the case with a single friend. For that we have the following pair of equations: $y^S = \alpha + \beta y^F + \boldsymbol{\gamma}'\mathbf{x}^S + \varepsilon_1$ and $y^F = \alpha + \beta y^S + \boldsymbol{\gamma}'\mathbf{x}^F + \varepsilon_2$.

Since y^F is not independent of ε_1 in the first equation and is an endogenous variable in this sense, and, similarly, since y^S is not independent of ε_2 in the second equation, the maximum likelihood estimates of parameters for this simultaneous model are inconsistent. However, we can obtain the following pair of reduced-form equations:

$$y^S = \frac{1}{1-\beta^2}\left\{\alpha(1+\beta) + \boldsymbol{\gamma}'\left(\mathbf{x}^S + \beta\mathbf{x}^{\mathbf{F}}\right) + \varepsilon_1 + \beta\varepsilon_2\right\} \text{ , and}$$

$$y^F = \frac{1}{1-\beta^2}\left\{\alpha(1+\beta) + \boldsymbol{\gamma}'\left(\beta\mathbf{x}^S + \mathbf{x}^{\mathbf{F}}\right) + \beta\varepsilon_1 + \varepsilon_2\right\} \text{ .}$$

While parameters are not linear even after a reparameterization because of the multiplicative form between β and $\boldsymbol{\gamma}$, we can obtain estimates of the parameters consistently by maximum likelihood estimation. A major problem for this reduced-form equation, however, is that when we hypothesize that the symmetric effect of y^F on y^S and of y^S on y^F, β, depends on relations between the subject and the friend, such that $\beta = \beta_0 + \boldsymbol{\beta}_1\mathbf{z}^{SF}$, then the reduced-form equations become very complicated, because of a resulting nonlinearity both in the regression component and in the heteroskedastic error term. Hence, we should consider a different method.

A more tractable approach is to use an instrumental variable for y^F in the equation $y^S = \alpha + \beta y^F + \boldsymbol{\gamma}'\mathbf{x}^S + \varepsilon_1$, and that can be done relatively easily, because a friend's characteristic, say x_m^F, that strongly affects y^F and can be assumed to be independent of ε_1 can be used as an instrumental variable. If the effect of y^F on y^S depends on covariates such that $\beta = \beta_0 + \boldsymbol{\beta}_1 \mathbf{z}^{SF}$, then the equation becomes

$y^S = \alpha + \beta_0 y^F + \boldsymbol{\beta}_1 ' y^F \mathbf{z}^{SF} + \boldsymbol{\gamma} ' \mathbf{x}^S + \varepsilon_1$, and variables $y^F \mathbf{z}^{SF}$ are also endogenous covariates. However, we can also employ $x_m^F \mathbf{z}^{SF}$ as the set of instrumental variables for $y^F \mathbf{z}^{SF}$ when we use x_m^F as the instrumental variable for y^F. As in the case of a categorical dependent variable, the model and the use of the IV method for parameter estimation can be extended for the case with three or more friends. However, those models for linear regression still differ from the models for a categorical dependent variable, because, unlike the case of the categorical dependent variable, the effect of y^F on y^S should be interpreted as one of possibly symmetric causally bidirected effects without a simultaneous modeling of $y^F = \alpha + \beta y^S + \boldsymbol{\gamma} ' \mathbf{x}^F + \varepsilon_2$. On the other hand, parameter λ^{SF} in equation (4) and its extensions in the following equations are intrinsically symmetric characteristics of the log odds ratio between y^S and y^F.

6. APPLICATION

6.1. Data and descriptive statistics

Data employed for an application come from Japanese General Social Survey (JGSS) 2004, with special supplements for egocentric social network data. The survey asked subjects to identify at most four significant others under the following specification for relations: “Among people with whom you frequently talk, who are the people with whom you talk about your important things or about your personal problems?” After the identification of those people, the survey asked the subject several questions about each of these significant others: questions about their major attributes, such as sex, age, and educational attainment, a few questions about their attitudes, including the political party they supported, and several questions about the characteristics of their relations, including the social category of the relationship (such as spouse, kin, friend, colleague, work-related other, neighbor), extent of closeness, duration of relation, frequency of communication, and whether politicians, politics, or elections were among the topics of the subject's talk with those people. Then the survey also asked whether those significant others were acquainted with one another. The portion of the survey questionnaire in Japanese that is related to question items described above is attached in the Appendix.

Our analysis is concerned with (1) the determinants of the subject's support for the Liberal Democratic Party (hereafter the LDP), and (2) the determinants of the strength of the association between the subject's and significant others' support for the LDP.

The LDP is the political party that has long dominated as the majority party in both the upper and lower houses of congress in Japan. It formed the cabinet for a long time, under a political system in which the majority party of the lower house governs, until its historic defeat by the Democratic Party of Japan (DPJ) in the lower house election in 2009. Despite its name, the LDP is a conservative political party. At the time when the survey was conducted, the LDP was still a stable, dominant party under Prime Minister Koizumi, who was among the most popular prime ministers in the post–World War II history of Japan. Nevertheless, in the JGSS data we analyze, only 31.5% of subjects expressed support for the LDP. While this figure may seem small, the DPJ, the second largest political party at that time, had the support of only 10.6% of the survey subjects. The largest group (41.0%) responded that there was no particular political party they supported.

The following analysis is restricted to subjects of ages 20-79 and those who identified 2–4 people as significant others according to the above-specified criteria. Among 1,201 such subjects, 33.1% (N = 397) identified two significant others, 28.1% (N = 337) identified three significant others, and 38.9% (N = 467) identified four significant others.

Although the survey did not specify any specific order among significant others when the subject identified two or more, the results indicate that the first significant other is significantly closer to the subject (with the score of 2.84 versus 2.73 on a scale of 1 = “not very close”, 2 = “close”, and 3 = “very close”), and is much more likely to be the subject’s spouse (45% versus 6% in the 1,201 total subjects or 61% versus 8% among the married) than the second significant other, both with a 0.1% level of significance.

Among 804 subjects who identified three or more significant others, there is also evidence at the 5% level of significance that the second significant other is closer to the subject than the third significant other (with a score of 2.74 versus 2.66). However, there is no statistically significant tendency for the second significant other to be the spouse more often than the third significant other (5% versus 4%). Among 467 subjects who identified four significant others, there is a marginally significant tendency (with a 10% level of significance) for the third significant other to be closer to the subject than the fourth significant other (with a score of 2.70 versus 2.60), but the proportions of those groups that are the spouse do not differ (both are 3%).

The results indicate that there is a tendency for subjects to name closer significant others earlier in the list – though we cannot claim that the first two are the two closest significant others. Second, if one of the significant others is a subject’s spouse, the subject has a strong tendency to place the spouse first on the list. However, about 22% of married subjects with two or more significant others did not specify their spouse as one of the significant others. From those observations, we conclude that significant others are not randomly ordered. Hence, the choice of the first two significant others differs from the random choice of two significant others – among subjects who identified three or more significant others. Hence, we present two regression analyses. First, we apply the regression model of equation (10) to the data for the first two significant others the subject identified, with the number of friends as the additional covariate. Second, we apply the model of equation (12) to the data for two randomly chosen friends. Both regression models are applied to the data for the 1,165 of the 1,201 sample subjects with two or more friends without covariates that cannot be defined due to missing data. For certain dichotomous variables, however, missing cases are combined with cases in the baseline category and are therefore not omitted from the analysis. In particular, when the political party that a significant other supports is missing, the case is classified as a case where he/she does not support the LDP, and when the information about a tie between significant others is missing, it is classified as a case without a tie.

As covariates to characterize the subject, $\mathbf{x}^S$, we employ the following nine: (1) the number of friends, (2) the local network density among significant others (or, alternatively, the presence versus absence of a tie between the two significant others in cases in which the first two significant others are chosen), (3) whether or not the spouse is among the significant others identified, (4) sex, (5) marital status (2 categories), (6) age (6 categories), (7) educational attainment (4 categories), (8) employment status (4 categories), and (9) the size of the residential municipality (3 categories). Table 1 presents the descriptive statistics of those covariates for the sample of 1,165 included in the regression analyses.

(Table 1 About Here)

As covariates on the association between the political attitudes of the subject and each friend, we employ (1) the number of indirect ties in the egocentric network through other significant others (or, alternatively, the presence versus absence of an indirect tie through the other significant other in the case in which the first two significant others are chosen), (2) the closeness of the relation (an interval scale variable based on three ordered categories), (3) the duration of the relation in years, (4) whether or not the relation is spousal, (5) the frequency of communication (an interval scale variable based on five ordered categories), and (6) whether or not the subject talks with this significant other about politicians, politics, or elections. As shown in equations (10) and (12), these covariates do not directly enter as the covariates of the dependent variables, but the sums of their interactions with the standardized dummy variable of the outcome of each significant other enter as covariates.

6.2. Analytical results

Table 2 presents four results of logistic regressions models, two models for each of the two data sets. One data set is based on the data for the first two significant others whom the subject identified, and the other data set is based on the data for two randomly chosen significant others. Because of the difference in the characteristics of the data, two variables, the subject's local network density and the number of indirect ties though other significant others, are calculated differently in the two data sets, as described in more detail in a footnote to Table 2.

(Table 2 About Here)

Model 1 of the each model includes six covariates for the log odds ratio between the subject's and significant other's supporting versus not supporting the LDP, but no individual-level covariates for the log-odds of supporting versus not supporting the LDP, and model 2 adds to model 1 individual-level covariates for the log odds of supporting versus not supporting the LDP. Those nested models are applied to see how the effects of relational covariates on the agreement in the outcome between the subject and his/her significant others can be explained as a results of homophily in the choice of significant others based on the individual-level attributes considered in the model.

Despite differences in the choice of two significant others, results from model 1 of table 2 consistently indicate that the determinants of the association between the subject's and significant others' support of the LDP are basically the same between the two data sets. Only three factors showed statistically significance effects on the strength of the associations between the political attitudes of the subject and significant others. First, the baseline intercept of the association is positive and strongly significant. Hence, there is a strong tendency for the agreement about the presence/absence of support for the LDP between the subject and his/her significant others – though we cannot establish from this cross-sectional data analysis whether this is a result of social influence, of unobserved selection bias in the choice of significant others, or of response to the uncontrolled common social context they share. Second, there is a stronger tendency of agreement in the presence/absence of support for the LDP if the relation is spousal. Although the analysis itself cannot differentiate causes, we consider it very likely that spouses are similar in this political attitude either because of their influence on each other or because of common lives and life chances they shared under the LDP government, rather than because of self-selective marriages based on a common political attitude. Third, although the frequency of communication does not have an effect, the presence of communication about "politicians, politics, or elections" increases the extent of agreement between the political party preferences of the subject and significant others. Hence, only when an exchange of information and opinions on things related to the specific substantive content of the attitude in question exists between the subject and his/her significant others do their attitudes become more similar.

The results from model 2, which includes the effects of individual-level covariates on the subject's support of the LDP, reveal several things. First, while the covariate effects on the agreement in political party preference between the subjects and their significant others do not change their characteristics, the intercept for the extent of agreement is reduced considerably compared with the results from model 1, and more so for the data set with the first two significant others than for the data set with randomly chosen significant others. Hence, homophily seems to explain a part of the strength of association between the subject's and the significant others' outcomes, especially in the case of the data set with the first two friends.

Second, regarding the effects of individual-level covariates related to egocentric network on the subject's outcome, all three covariates are shown to affect the odds of supporting the LDP. First, a greater number of significant others leads to a higher probability of supporting the LDP. Second, having a more "interlocking" rather than a “radial” egocentric social network (Laumann 1973) leads to a higher probability of supporting the LDP. Third, not having the spouse among the significant others while being married leads to a higher probability of supporting the LDP. The qualification about marital status in the third finding comes from the fact we control for marital status by another covariate. These findings indicate that the LDP supporters are likely to have more expansive and denser social ties than others and are in this regard better integrated into the society, but at the same time they have poorer informal support relations with the spouse than others.

The effects of other individual attributes indicate that LDP supporters tend to be older and are less likely to be college graduates. Although it has been known that the LDP has a stronger basis of support from nonurban areas, this does not hold any more, and the municipality size had no significant effect during the time of the Koizumi government, which reduced the government’s financial aid to farmers.

7. CONCLUSION AND DISCUSSION

This paper introduced some novel formulations of logit and adjacent logit models for an analysis of a categorical dependent variable with egocentric social network data as covariates in order to identify the determinants of outcomes for the subjects and the determinants of the extent of agreement/disagreement between the outcomes of the subject and his/her significant others. The method has the advantage of being able to control for homophily in the choice of significant others in assessing the effects of covariates on the agreement in attitude or behavior between the subject and his/her significant others.

Fischer (1982) identified three social relations (a) “formal” relations based on social roles such as spousal, kinship, or collegial roles, (b) “sentimental” relations typically represented by the extent of intimacy or subjective closeness, and (c) “exchange” relations, including those based on economic exchange and those based on social exchange, such as informal social support relations. In addition, we can also characterize relations by certain quantitative aspects such as the frequency of contacts or communication, or the duration of a relation. Such information is relatively simple to collect for egocentric social networks in a survey, and by examining how those distinct aspects of relations function in affecting people’s attitude and behavior and the extent of agreement in the attitude and behavior between subjects and their significant others, the method introduced in this paper will enrich the findings about interdependence in social attitude and social behavior from studies based on analyses of sample survey data.

REFERENCES

Burt, Ronald S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Clogg Clifford C. 1982. “Some Models for the Analysis of Association in Multi-way Cross-Classifications Having Ordered Categories.” *Journal of the American Statistical Association* 77:803-15.

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

Giletta, Mattero, Ron H.J. Scholte, William J. Park, Rutger C.M.E. Engels, Junilla K. Larsen, Mitchell J. Princtein, and Silvia Ciairano. 2011. “Similarity in Depressive Symptoms in Adorescents’ Friendship Dyads: Selection ot Socialization?” *Developmenmtal Psychology* 47: 1804-14.

Goodman Leo A. 1979. “Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories.” *Journal of the American Statistical Association* 74:537-52.

Granovetter, Mark. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78:1360-80.

Granovetter, Mark. 1974. *Getting a Job: A Study of Contacts and Careers.* Cambridge, MA: Harvard University Press.

Laumann, Edward O. 1973. *Bonds of Pluralism*: *The Form and Substance of Urban Social Networks.* New York: Wiley.

Marsden, Peter. 2002. “Egocentric and Sociocentric Measures of Centrality.” *Social Networks* 24:407-22.

Rasch, Georg. 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Paedagogiske Institut.

Shrout, Patrick and Denise B. Kandel 1981. “Analyzing the Property of the Dyads: Determinants of Similarity in Marijuana Use in Adolescents Friendship Dyads.” *Sociological Methods and Research* 9:363-74.

Snijders, T.A.B., M. Spreen, and R. Zwaagstra. 1995. “The Use of Multilevel Modelling for Analyzing Personal Networks. *Journal of Quantitative Anthropology* 5: 85-105.

Van Duijn, M.A.J., J.T. Van Busschbach, and T.A.B. Snijders. 1999. “Multilevel Analysis of Personal Networks as Dependent Variable.” *Social Networks* 21: 187-209.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Wasserman. Stanley and Phillipa Pattison 1996. “Logit Models and Logistic Regressions for Social Networks I. An Introduction Markov Graphs and *p**.” *Psychometrika* 61:401-25.

Yamaguchi, Kazuo. 1990. "Homophily and Social Distance in the Choice of Multiple Friends: An Analysis Based on Conditionally Symmetric Log-Bilinear Association Models." *Journal of the American Statistical Association* 85:356-66.

Yamaguchi, Kazuo and Denise B. Kandel. 1993. "Marital Homophily and Illicit Drug Use Among Young Adults: Assortative Mating or Marital Influence?" *Social Forces* 72: 505-28.

Yamaguchi, Kazuo and Denise B. Kandel. 1997. "The Influences of Spouses' Behavior and Marital Dissolution on Marijuana Use: Causation or Selection?" *Journal of Marriage and the Family* 59:22-36.

Table 1. Descriptive Statistics of the Dependent Variable and Covariates

I. Dependent Variable		
Supports the LDP	0.315	
Does not support the LDP	0.685	
II. Covariates: Categorical (in %)		
(1) Whether the Spouse is One of the Significant Others		
Yes	0.553	
(2)Sex		
Women	0.616	
(3) Marital Status		
Married	0.746	
(4) Age		
20-29	0.129	
30-39	0.203	
40-49	0.165	
50-59	0.211	
60-69	0.186	
70-79	0.106	
(5) Educational Attainment		
Junior high school	0.148	
High School	0.487	
Junior college	0.178	
College or more	0.236	
(6) Employment Status		
Regular employment	0.333	
Temporary employment	0.179	
Self-Employed	0.157	
Non-employed	0.397	
(7) Size of Municipality		
Large Cities	0.208	
Other cities	0.572	
Tows/villages	0.221	
III. Interval Scale Variables		
	Mean	S.D.
(8) Number of Significant Others	3.062	0.845
(9a) Local Network Density Among All Significant Others	0.777	0.329
(9b) Local Network Density Between the First Two Significant Others	0.849	0.358

Table 2. The Results from Logistic Regression Models

	The first two Significant others		Two Randomly Chosen Significant Others	
	Model 1	Model 2	Model 1	Model 2
I. Covariate effects on the log odds of supporting the LDP versus not supporting the LDP				
1. Covariates related to egocentric network				
(1) Number of Significant Others	-----	0.195*	-----	0.267**
(2) Local Network Density[1]	-----	0.898**	-----	0.654*
(3) Whether the spouse is among the significant others	-----	-0.519*	-----	-0.463*
2. Other attributes of the subject				
(4) Sex (versus men)				
women	-----	-0.189	-----	-0.171
(5) Marital Status (versus single)				
Married	-----	0.257	-----	0.263
(6) Age: versus 20-29				
30-39	-----	-0.098	-----	0.007
40-49	-----	0.787*	-----	0.823*
50-59	-----	0.724*	-----	0.725*
60-69	-----	1.386***	-----	1.330***
70-79	-----	1.162**	-----	1.242**
(7) Educational Attainment (versus high school)				
Junior high school	-----	-0.231	-----	-0.112
Junior college	-----	-0.210	-----	-0.210
College or more	-----	-0.527*	-----	-0.529*
(8) Employment Status (versus regular employment)				
Temporary employment	-----	-0.187	-----	-0.237
Self-Employed	-----	0.191	-----	0.162
Non-employed	-----	0.216	-----	0.225
DK	-----	0.300	-----	0.268
(9) Size of Municipality (versus large cities)				
Other cities	-----	0.070	-----	0.121
Towns/Villages	-----	0.369	-----	0.452
(10) Intercept	-0.475***	-2.037***	-0.441***	-1.937***
II. Covariate effects on the log odds ratio between the subject’s and each significant other’s supporting versus not supporting the LDP				
(1) Intercept	1.590***	1.224**	1.119**	0.884*
(2) Number of indirect ties through other significant Others[2]	-0.439	0.091	-0.169	-0.014
(3) Closeness of relation	-0.087	-0.057	-0.054	-0.014
(4) Duration of relation	0.001	0.001	0.001	0.001
(5) Spousal relation (versus other relations)	0.958***	0.999***	1.152***	1.106***
(6) Frequency of communication	0.024	-0.024	0.052	0.013
(7) Presence of communication about politicians, politics or elections	0.245*	0.236*	0.254*	0.239*

***p<.001, **p<.01, *p<.05.

[1]A dummy variable for the presence versus absence of a tie between the two significant others in the case of the first two significant others, and the proportion of ties among significant others in the case of two significant others chosen randomly.

[2]The presence versus the absence of an indirect tie with the other significant others in the case of the first two significant others, and the number of indirect ties through all other significant others in the case of two randomly-chosen significant others.

[DO NOT PRINT \JGSSSEMINAR2010\SPSS2\EGONETNEW6,EGONETNEW8,EGONETNEWS5,EGONETNEWS6]

APPENDIX: Survey items of egocentric social networks in the 2003 Japanese General Social Survey

面接：調査員記入

問１「これから、あなたがよく話をする人たちについておうかがいします。まず、あなたが重要なことを話したり、悩みを相談する人たちを思い浮かべてください。」
ここで回答者にメモ用紙を渡す。

「ご自分で後から見て誰かわかるように、このメモ用紙のＡからＤの四角の中に、お名前あるいは頭文字（イニシャル）、愛称・ニックネームなどを、書き入れてください。
４人いなければ、思い浮かぶ人数だけで結構です。誰も思い浮かばなければ何も書かなくて結構です。」
回答者がメモに記入したことを確かめる。

問 1-1 「何人の名前　（頭文字など）を書きましたか？」
該当する人数に○をつける。

Z1NUM　０ 人　１ 人　２ 人　３ 人　４ 人

（０〜３人 → 問 1-3へ、４人 → 問 1-2へ）

問 1-2　Z1ELSE
「それ以外に、あなたが重要なことを話したり悩みを相談する人は何人いますか？」　□□ 人

問 1-3 「ＡからＤのうち、名前が記入されていないのは どの欄ですか？」
記入のない欄の四角の中に×をつける。×の数と問 1-1 の人数との合計が「４」にならない場合は、問 1-1 と問 1-3 を見直す。

A	B	C	D
□	□	□	□
Z1XXA	Z1XXB	Z1XXC	Z1XXD

「これからは、Ａ欄に書いた人をＡさん、Ｂ欄に書いた人をＢさん、Ｃ欄に書いた人をＣさん・・というように呼びます。」

問 1-4 「今メモ用紙に書いた人たちは、お互いに知り合いですか？」
問 1-1 で「０人」「１人」の場合、「４ 該当する人はいない」にすべて○をつける。5 人以上いた場合は、A,B,C,D の 4 人についてのみ尋ねる。

Z1KNOW**　**には下記のｱﾙﾌｧﾍﾞｯﾄが入る

		知り合い	知り合いではないと思う	わからない	該当する人はいない
AB	①「ＡさんとＢさんは、知り合いですか？」	1	2	3	4
AC	②「ＡさんとＣさんは、知り合いですか？」	1	2	3	4
AD	③「ＡさんとＤさんは、知り合いですか？」	1	2	3	4
BC	④「ＢさんとＣさんは、知り合いですか？」	1	2	3	4
BD	⑤「ＢさんとＤさんは、知り合いですか？」	1	2	3	4
CD	⑥「ＣさんとＤさんは、知り合いですか？」	1	2	3	4

あなたが**重要なことを話したり、悩みを相談する人たち**（Aさん・Bさん・Cさん・Dさん）についてうかがいます。お手元のメモ用紙を見ながらお答えください。

問5-1　その人たちは、あなたにとってどのような間柄ですか。あてはまるものすべてに○をつけてください。

		*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
ZSS*	配偶者（夫または妻）	1	1	1	1
ZKIN1*	親または子ども	2	2	2	2
ZKIN2*	兄弟姉妹・その他の家族・親せき	3	3	3	3
ZJOBREL*	職場の上司または部下	4	4	4	4
ZJOBCOL*	職場の同僚（上司・部下以外）	5	5	5	5
ZJOBETC*	その他の仕事関係	6	6	6	6
ZTEAM*	同じ組織や団体に加入している人	7	7	7	7
ZNEIB*	近所の人	8	8	8	8
ZFRIE*	友人	9	9	9	9
ZRELETC*	その他	10	10	10	10
ZRLNOMK*	（いずれも選択していない）				

問5-2　その人たちは男性ですか、女性ですか。

ZSEX*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
男性	1	1	1	1
女性	2	2	2	2

問5-3　その人たちの年齢を、以下に記入してください。だいたいの年齢で結構です。
ZAGE*

*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
＿＿歳	＿＿歳	＿＿歳	＿＿歳

問5-4　その人たちと知り合ったのは、いまから何年前でしたか。だいたいの年数で結構です。
ZKNOWYR*

*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
＿＿年前	＿＿年前	＿＿年前	＿＿年前

問5-5　その人たちとあなたは、通常どのくらいの頻度で話をしますか（電話やメールも含みます）。
ZFQTALK*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
ほとんど毎日	1	1	1	1
週に数回	2	2	2	2
週に1回程度	3	3	3	3
月に1回程度	4	4	4	4
年に数回	5	5	5	5

問5-6　その人たちが最後に通学した（または現在通学している）学校は、次のどれにあたりますか。なお、中退の場合も、その学校をお答えください。

ZLSTSCH*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
中学校（旧制小学校）	1	1	1	1
高校（旧制中学校・高等女学校・実業学校・師範学校）	2	2	2	2
短大・高専	3	3	3	3
専門学校	4	4	4	4
大学（旧制高校・大学）・大学院	5	5	5	5
わからない	6	6	6	6

問5-7　その人たちは、現在どのようなかたちで仕事をしていますか。学生でアルバイトをしている場合は「学生」を選んでください。

ZTP7JOB*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
自営業主・自由業者・家族従業員	1	1	1	1
経営者・役員	2	2	2	2
正規の職員・社員	3	3	3	3
公務員	4	4	4	4
パート・アルバイト・嘱託・臨時・派遣	5	5	5	5
学生	6	6	6	6
仕事をしていない（専業主婦、退職者など）	7	7	7	7
わからない	8	8	8	8

問5-8　その人たちが現在行なっている仕事の内容は、以下のどれにあたりますか。前問で「学生」と「仕事をしていない」を選んだ人については回答の必要はありません。

ZJOB*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
上級管理職（経営者、役員、部長など）	1	1	1	1
中間管理職（課長、店長など）	2	2	2	2
専門・技術（技術者、教員、弁護士など）	3	3	3	3
事務（総務、経理、企画、営業事務など）	4	4	4	4
販売（小売店主、店員、外交員など）	5	5	5	5
サービス（理美容、調理、家事サービスなど）	6	6	6	6
運輸・通信（運転手、船員、通信員、郵便外務など）	7	7	7	7
保安・警備（守衛、警官、自衛官など）	8	8	8	8
製造・建設（工場作業・建築作業者など）	9	9	9	9
農林漁業・鉱業	10	10	10	10
わからない	11	11	11	11

問5-9　あなたは、その人たちとどのくらい親しいですか。

ZCLOSE*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
とても親しい	1	1	1	1
親しい	2	2	2	2
それほど親しくない	3	3	3	3

問5-10　その人たちとはこのところ政治家や選挙・政治についてどのくらい話題になりましたか。

ZPLTALK*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
話題になった	1	1	1	1
あまり話題にならなかった	2	2	2	2

問5-11　その人たちは、国政選挙ではどの政党の候補者に投票する(あるいは投票した)と思いますか。
１つだけ○をつけてください。

ZVOTE*

	*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
自民党	1	1	1	1
民主党	2	2	2	2
公明党	3	3	3	3
~~自由党~~	~~4~~	~~4~~	~~4~~	~~4~~
共産党	5	5	5	5
社民党	6	6	6	6
保守新党	7	7	7	7
その他の政党	8	8	8	8
わからない	9	9	9	9
投票に行かないと思う	10	10	10	10
選挙権がない	11	11	11	11

問5-12　その人たちについて、次のうちあてはまるものすべてに○をつけてください。

		*A Aさんは？	*B Bさんは？	*C Cさんは？	*D Dさんは？
ZCOHOBB*	共通の趣味や娯楽を持っている	1	1	1	1
ZCOUTG*	最近6ヶ月間に、遊びや食事のためにいっしょに出かけたことがある	2	2	2	2
ZCOMONY*	まとまったお金を借りることができる	3	3	3	3
ZCONOMK*	（いずれも選択していない）				