Advances in Methodology, Data Analysis, and Statistics
Anuška Ferligoj (Editor), Metodološki zvezki, 14
Ljubljana: FDV, 1998

Quality of Scales Measuring Complete Social Networks

Anuška Ferligoj and Valentina Hlebec*

Abstract

This paper evaluates the validity and reliability of complete network measurement instruments. The authors present and discuss the results of two experiments which were designed to systematically analyze the quality of the following three measurement instruments each of which are defined by different scales: the binary scale, the ordinal scale and the line production scale as used to estimate the strength of interpersonal relations. Reliability and validity were assessed with the true score MTMM approach. Similar results emerged from from both experiments despite the reverse ordering of the measurement scales. The binary scale had the lowest reliability in both experiments.

1 Introduction

Social network data consist of one or more relations measured among a set of units. Not surprisingly, there are many issues concerning the measurement and collection of network data. Wasserman and Faust (1994, p. 56) emphasized that very little work has been done on the issues of validity, reliability, and measurement error in social network data. The principal concern of this paper, therefore, is the quality of network data obtained by survey.

Wasserman and Faust (1994, p. 58) reported that the following three approaches have been used to assess the reliability of social network data: test-retest comparison, comparison of alternative questionnaire formats, and the reciprocity of sociometric choices. Hammer (1984), Sudman (1985, 1988), and Hlebec (1993) used test-retest comparisons to assess the differences between recall and recognition methods for listing members of egocentric networks. Bien et al. (1991) and Neyer et al. (1991) developed network related test-retest reliability measures at individual and aggregated levels (size of individual network, stability of name generators and relations, spatial distance and contact frequencies). Lairetier (1993) used rank

* Faculty of Social Sciences, University of Ljubljana, P.O. Box 2547, 1001 Ljubljana, Slovenia.

correlation and correspondence to estimate the test-retest reliability of the egocentric network measures. He also computed aggregate measures with regard to roles, relations, and support resources. The reliability at the level of egocentric network density and composition measures was examined by Marsden (1993), among others. All of these researchers concluded that reliability of aggregated measures is higher than the reliability of individual choices and that strong or intimate relations are more likely to be reciprocated than less intense ones.

Even less work has been done to assess the quality of complete network measurements. The validity of the informal leadesrhip concept was explored in a paper written by Macur and Hlebec (1996). In their study, the results of in-depth interviews as well as informant reports about the informal leadership of a choir were compared with the sociometric choices made by all members of the choir.

Ferligoj and Hlebec (1993) analyzed the (in)sensitivity of centrality measures (in-degree point centrality indices, in-closeness global centrality indices (Sabidussi, 1966), and Freeman's betweenness index (1979) using test-retest approach and discovered that complex measures of centrality tend to be more affected by measurement error. In fact, measurement error seemed to be related to the prominence of the actors involved. Namely, other actors listed more central actors with fewer errors. Calloway and colleagues (1993) analyzed the reliability of complete interorganizational, self-reported networks. The percentage of mutually confirmed relations (as being present or absent) between respondents was used to estimate reliability. This approach is less appropriate for asymmetrical relations where the absence of one report should not be interpreted as unreliability. Ferligoj and Hlebec (1995) analyzed the reliability of complete interpersonal network measurements using traditional reliability estimates: Crombach's alpha (Crombach, 1951), the theta coefficient (Armor, 1974), and the true score MTMM approach (Saris and Andrews, 1991) on vectorized relational matrices. Reliability on the level of complete networks prooved to be quite high. Comparing the three measurement methods defined by different scales for measuring the strength of relations, they discovered that the binary scale was the least reliable. These result could be explained in a number of ways. Because the binary scale was the first one used, its low reliability might be due merely to its position in the questionnaire. The higher reliability of the second and the third method similarly might be the consequence of the memory effect since the repetitions took place within one interview that lasted, on average, only 23 minutes.

In order to control for the position of the method and memory effect, the second experiment was designed with reverse method ordering and longer time intervals between interviews. In this paper, the authors compare the results of the second experiment, with the experiment of Ferligoj and Hlebec (1995). The authors of this paper used similar to previous experiment, the true score MTMM approach (Saris and Andrews, 1991) to estimate reliability and validity of a single question as well as the method for measuring complete interpersonal networks on vectorized relational matrices without diagonal elements.

2 Estimating reliability and validity of complete networks

It is possible to repeat different network generators several times, each time with a different method that measures the presence or absence of ties as well as the strength of present ties. Therefore, several relational matrices can be produced by different measurement procedures.

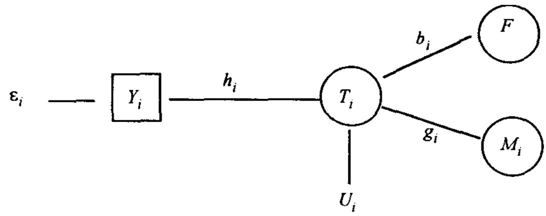

Figure 1: True score measurement model

There are only a few proposed procedures for estimating the reliability of egocentric network measures and even fewer for estimating the reliability of complete network measures -- e.g., matching procedure which counts the number (or percentage) of cells with different values in two matrices. In this paper, the reliability measures that were designed to assess the reliability of variables are also used on complete network data. This is done by vectoring of relational matrices.

In the following analysis, one unit of the analysis is a dyad and each network (i.e. vector) is treated as a variable. There are different methods which could be used to assess reliability. Ferligoj and Hlebec (1995) used several traditional approaches to estimate the reliability of a composite and single variable. In this paper, the authors focus exclusively on the reliability and validity of a single variable provided by the true score measurement model as conceived by Saris and Andrews (1991: 576-583) and presented in Figure 1.

This measurement model can be expressed by the following equations:

$$Y_i = h_i\, T_i + \varepsilon_I$$
$$T_i = b_i F + g_i M_i + U_i$$

where:

Y_i is the response or observed variable corresponding to the question measued with the method i;

T_i is the stable component when the same question is repeated under exactly the same conditions;
ε_i is the random error in the observed variable Y_i;
F is the unobserved variable of interest, assumed to be independent of the measurement procedure used;
M_i is a method specific component;
U_i is the unique disturbance, representing the interaction between the trait (question) and the method.

In this model it is assumed that:

$$E(\varepsilon_i) = 0;\quad E(U_i) = 0;\quad cov(F, U_i) = 0;\quad cov(M_i, U_i) = 0;\quad cov(M_i, \varepsilon_i) = 0;$$
$$cov(F, \varepsilon_i) = 0;\quad cov(U_i, \varepsilon_i) = 0;\quad cov(F_i, M_i) = 0.$$

In this measurement model, reliability is defined as the proportion of the variance in Y_i that remains stable across repetitions of the same measure, or:

$$reliability = \frac{var(T_i)}{var(Y_i)} = h_i^2$$

Validity [1] is defined as the percentage of the variance of the true score explained by the variable of interest, or:

$$validity = b_i^2$$

Invalidity ($1 - b_i^2$) can be interpreted as method variance (g_i^2), if $U_i = 0$. Otherwise, invalidity is defined as follows:

$$invalidity = g_i^2 + var(U_i)$$

In this model (using one measurement), the reliability, validity and invalidity coefficients can not be estimated. Therefore, several different approaches with repeated measurements were suggested. In this paper, the authors utilize the true score MTMM approach proposed by Saris and Andrews (1991) to assess the coefficients. The true score MTMM model allowed us to estimate the reliability and validity of each variable separately.

[1] These are not the only possible definitions of reliability and validity (see Saris and Andrews, 1991: 581-582).

3 Experiments

The results of the first experiment showed that the binary scale, which is the scale used most frequently to measure personal networks, was less reliable than the ordinal 11 point scale and the line production scale. In the first experiment, the binary scale was positioned first in the questionnaire and one of the possible explanations for the lower quality of the binary scale could certainly be method ordering. Morover, one of the possible explanations for higher quality of the second and the third method could be memory effect due to the short time intervals between the questions posed in the interview. In order to control for method ordering and to prevent the memory effect a new experiment was designed with reversed method ordering and longer time intervals between questions.

In both experiments, we measured the social support exchange relations among students using the same questions and the same scales though in different order. The first experiment is described in detail in the paper of Ferligoj and Hlebec (1995) and various elements of the expeiment, the questionnaire and other characteristics, will be presented bellow. The first analyzed network consisted of social support exchange relations among thirteen students of the Social Science Informatics second year class (1992/1993) at the Faculty of Social Sciences, University of Ljubljana. Data were collected by CAPI (Computer Assisted Personal Interviewing) which is supported by the program INTERV (de Pijper and Saris, 1986). Interviews were carried out in May 1993. Network generators were repeated three times, each time with a different scale. The first was the binary scale in which respondents listed by heart the colleagues with whom they have the strongest relation. The second was the line production scale in which respondents received a list of all members of the group and indicated the strength of their relation with each memeber of the group using length of a line. The third was the ordinal 11 point scale in which respondents received a list of all members of the group and indicated the strength of their relation with each member using a number between 0 and 10. All three repetitions took place in the same interview that lasted on average of 23 minutes. In order to prevent memory effect, several other questions were placed between the repetitions of the network questions.

The second analyzed network consisted of social support exchange relations among thirty four students of the first year class (1994/1995) of secondary vocational school (Srednja šola tehničnih strok in osebnih storitev) in Ljubljana. Data were collected via a self administered questionnaire. Interviews were carried out in May 1995. Network generators were again repeated three times and each time with different scale. This time, however, the order was reversed in relation to the ordering in the first experiment. The first was the ordinal 11 point scale, the second was line production scale and the last was the binary scale. Respondents received a list of all the members of the class at each time. The three repetitions took place with one week intervals in between.

Social support was measured by four network questions. The questions measuring social support were identical in both experiments except for minor changes made to accomodate differences between the groups.

Introduction: You have taken several exams (tests) since you are second (first) year in scool. Students usually borrow studying material from their classmates.

1. List the names of the classmates from whom you have most frequently borrowed studying materials. (The number of listed persons is not limited.)
2. List the names the classmates whos have most often borrowed studying material from you. (The number of listed persons is not limited.)

Introduction: Suppose you were ill at the beginning of May and you had to stay in hospital for a month. Therefore, you need to obtain studying materials as weel as information about important school events.

3. Wich of your classmates would you most likely ask for help? (The number of listed persons is not limited.)
4. Which of your classmates would most likely ask you for help in a similar situation? (The number of listed persons is not limited.)

Going back to correct answers was not allowed during the first and second interview though it was possible to correct the length of line or the number expressing the strength of a relation within each individual question. The maximum length of the line was 20 points in the first experiment and 10 points in the second experiment.

The exchange of studying materials and help was measured in both directions, i.e. giving and receiving. First, a respondent reported about the studying materials (s)he borrowed from others (questions 1 and 3 - the original questions) and then about the studying materials (s)he lent to the others (questions 2 and 4 - the reversed questions). In order to arrive at the same type of relations all four times, the reversed matrices were transposed. In other words, the lent studying material perceived by lenders was attributed to borrowers as if reported by borrowers themselves.

All together, twelve different social support relations among respondents were measured in each experiment, i.e. four different questions or traits within each of the three methods.

We excluded the diagonals from the matrices because they have been set to 0 due to the nature of the measured relations. The matrices were then vectorized. In the following discussion we will refer to these vectorized matrices as variables.

4 Results

The univariate statistics for each individual relation are presented in Table 1.

Table 1: Means, standard deviations, coefficients of variation

Measurement Scale	Material	M. Reversed	Illness	I. Reversed
First Experiment				
	Mean			
Binary (0-1)	.21	.19	.17	.19
Line (1-20)	3.68	2.76	4.85	4.57
11 p.s. (0-10)	2.22	2.09	2.50	2.33
	Standard Deviation			
Binary (0-1)	.41	.39	.38	.40
Line (1-20)	4.64	3.22	6.07	5.70
11 p.s. (0-10)	2.91	2.50	3.20	2.98
	Variation Coefficient			
Binary (0-1)	1.95	2.05	2.24	2.11
Line (1-20)	1.26	1.17	1.25	1.25
11 p.s. (0-10)	1.31	1.20	1.28	1.28
Second Experiment				
	Mean			
Binary (0-1)	.20	.20	.15	.14
Line (1-20)	.76	.80	1.16	1.15
11 p.s. (0-10)	.80	.80	1.68	1.39
	Standard Deviation			
Binary (0-1)	.40	.40	.36	.35
Line (1-20)	1.89	1.93	2.71	2.67
11 p.s. (0-10)	1.93	1.82	3.05	2.67
	Variation Coefficient			
Binary (0-1)	2.00	2.00	2.40	2.50
Line (1-20)	2.49	2.41	2.34	2.32
11 p.s. (0-10)	2.41	2.28	1.82	1.92

The means in the first experiment are higher than those in the second experiment, most likely as a result of the number of members in each group. In order to arrive at comparable measures of variability, we also calculated the coefficient of variation.

In neither experiments were there substantial differences in coefficient of variation between the original and reversed questions. However, there are some differences between the two experiments. In the first experiment, the coefficients of variation were the largest on all questions when the binary scale was used whereas in the second experiment they were the largest only on those questions regarding illness. When we compare all values of the coefficients of variation across experiments, it is clear that the values are systematically higher in the second experiment. Possibly, this this is due simply to the larger number of network members in the second experiment.

5 Correlation matrix

The correlation coefficients between the twelve variables (relations) are presented in Table 2a (first experiment) and Table 2b (second experiment).

Table 2a: Correlation coefficients

	MB	RMB	IB	RIB	ML	RML	IL	RIL	MN	RMN	IN	RIN
MB	1.00											
RMB	.44	1.00										
IB	.39	.48	1.00									
RIB	.30	.48	.64	1.00								
ML	.75	.43	.43	.44	1.00							
RML	.47	.65	.45	.62	.60	1.00						
IL	.50	.55	.82	.70	.57	.52	1.00					
RIL	.22	.49	.62	.75	.33	.61	.68	1.00				
MN	.78	.56	.53	.53	.90	.63	.68	.44	1.00			
RMN	.46	.68	.66	.67	.56	.81	.68	.72	.64	1.00		
IN	.49	.52	.79	.68	.53	.52	.93	.66	.66	.69	1.00	
RIN	.21	.47	.66	.70	.30	.61	.69	.91	.40	.77	.68	1.00

The labels consist of two characters:

The first character denotes the question label:

- M borrowing studying materials,
- RM borrowing studying materials, reversed,
- I informal help in the case of illness,
- RI informal help in the case of illness, reversed;

The second character denotes the method used:

- B binary scale,
- L line production scale,
- N numerical scale.

The triangles at the edge show the correlation among variables measured with the same method, i.e. heterotrait - monomethod blocks (see also Campbell and Fiske, 1959). The rectangles show the correlation among the variables measured with different methods, i.e. heterotrait - heteromethod blocks. Within these rectangles, the diagonals, which show the correlation measured between the same variable measured with two different methods (the monotrait - heteromethod diagonals) are especially important. If these values are high, then the variable measured with different methods show high level of convergent validity and measure the same theoretical construct.

The correlation within the heterotrait - monomethod triangles differ from each other. This demonstrates that the measurements of different variables (traits) with the same method are not parallel. The correlation coefficients between the original questions (e.g., IB) and the reversed ones (e.g., RIB) are usually higher than others, but never higher than 0.68. The correlation coefficients are the lowest in the top triangle (binary scale), higher in the middle triangle (line scale) and the highest in the bottom triangle (11 point scale). It is difficult to distinguish whether the higher correlation coefficients in later methods are due to better quality of the specific method or to the method order.

The correlation coefficients in the heterotrait - heteromethod rectangles are low for the combination of the binary and the line scales and for the combination of the binary and the 11 point scales. They are higher for the combination of the line production and the 11 point scales. Possibly, the higher correlation coefficients between the line and the 11 point scales are due to the greater similarity of the two scales temselves.

As expected, the monotrait diagonals have the highest level of correlation coefficients. The monotrait-heteromethod diagonal correlation coefficient is consistently higher than the correlation coefficients lying in the same column and row of the heterotrait-heteromethod triangles. The monotrait-heteromethod correlation is also consistlenty higher than the corresponding heterotrait-monomethod correlation coefficients. To some extent the patterns of trait interrelationships follow the same pattern. Therefore, we can say that some of Campbell and Fiske's (1959) criteria for convergent and discriminant validity are met in the data from our first experiment.

The correlation coefficients between the twelve variables are presented in Table 2b (second experiment). Again, the level of correlation within the heterotrait - monomethod triangles are different from the others. The correlation coefficients between the original questions (e.g., IB) and the reversed ones (e.g., RIB) are generally higher than others, appearing in the blocks but never higher than 0.60. The Correlation coefficients are the lowest in the bottom (11 point scale), the highest in the middle triangle (line production scale) and medium in the top triangle (the binary scale). This time the order of the methods (ordinal 11 point scale, line production scale and binary scale) does not correspond to the values of the correlation coefficients as was the case in the first experiment. For the first method used in the second experiment, the correlation coefficients were actually lower (the ordinal 11 point scale) than for the second scale used (line production scale). This was also the case in the first experiment.

The correlation coefficients in the heterotrait - heteromethod rectangles are low for the combination of the binary and the line production scales and even lower for the combination of the binary and the 11 point scale. They are higher for the combination of the line production and the 11 point scales. This result resembles that of the first experiment and makes it even more likely that the higher correlation coefficient between the line production and the ordinal 11 point scales are due to the greater similarity of these two scales.

Once again, themonotrait diagonals show the highest levels of correlation. The monotrait-heteromethod diagonal correlation is consistently higher than the correlation coefficients which lie in the same column and row in the heterotrait-heteromethod triangles. The monotrait-heteromethod correlation is, however, not consistently higher than the corresponding heterotrait-monomethod correlations. To some extent, the patterns of trait interrelationships are similar. At the very least it can be said that some of the criteria for convergent and discriminant validity are also met in the second experiment data.

Table 2b: Correlation coefficients (second experiment) (labels are given in Table 2a)

	MB	RMB	IB	RIB	ML	RML	IL	RIL	MN	RMN	IN	RIN
MB	1.00											
RMB	.58	1.00										
IB	.51	.43	1.00									
RIB	.53	.59	.53	1.00								
ML	.65	.49	.58	.48	1.00							
RML	.51	.60	.48	.55	.51	1.00						
IL	.45	.33	.65	.40	.60	.47	1.00					
RIL	.42	.50	.56	.63	.50	.62	.59	1.00				
MN	.61	.50	.53	.42	.76	.55	.55	.46	1.00			
RMN	.51	.58	.43	.53	.47	.70	.40	.56	.50	1.00		
IN	.44	.33	.59	.42	.54	.42	.75	.56	.58	.40	1.00	
RIN	.42	.44	.48	.55	.42	.48	.48	.61	.41	.60	.45	1.00

6 MTMM true score model coefficients

The reliability coefficients, validity coefficients and method effects obtained from the true score MTMM model are presented in Table 3. The effects were estimated by the ML procedure of LISREL VI program (Joreskog and Sorbom, 1986). Please note that the reliability coefficients h_{ij} are presented in the table though not the reliabilities h_{ij}^2 as defined in a previous section.

The results in Table 3 indicate that the reliability coefficients for the binary scale are lower than the reliability coefficients for the other two scales. This is true for all four traits in the first experiment and the second experiment. Reversed questions have lower reliability coefficients compared than those of the original questions in both experiments. One would intuitively expect lower reliability coefficients on the reversed questions due to asymmetric nature of relations. The reversed questions measure respondents' perception of the social support (s)he is providing. These are compared with the amount of support actually provided (or at least as it is perceived by receivers of support).

Validity coefficients are very high and almost the same for all variables and all methods in the first experiment. Validity coefficients are high and almost the same for all variables and for two of the scales (line production and numeric estimation) in the second experiment. Validity coefficients are the highest for the binary scale in the first experiment and the lowest in the second experiment.

Table 3: MTMM true score model reliability, validity and method effect coefficients

Measurement Scale	Variable (Relation) Material	M. Reversed	Illness	Reversed
First Experiment				
	Reliability Coefficients h_{ij}			
Binary (0-1)	.80	.72	.84	.77
Line (1-20)	.92	.87	.98	.95
11 p.s. (0-10)	.99	.95	.95	.95
	Validity Coefficients b_{ij}			
Binary (0-1)	1.000	1.000	1.000	1.000
Line (1-20)	.998	.998	.998	.998
11 p.s. (0-10)	.996	.995	.995	.995
	Method Effect Coefficients g_{ij}			
Binary (0-1)	.046	.051	.044	.047
Line (1-20)	.061	.065	.057	.059
11 p.s. (0-10)	.093	.097	.097	.097
Second Experiment				
	Reliability Coefficients h_{ij}			
Binary (0-1)	.80	.80	.74	.79
Line (1-20)	.90	.84	.93	.82
11 p.s. (0-10)	.85	.82	.84	.79
	Validity Coefficients b_{ij}			
Binary (0-1)	.919	.918	.905	.916
Line (1-20)	.990	.989	.991	.988
11 p.s. (0-10)	.984	.981	.983	.981
	Method Effect Coefficients g_{ij}			
Binary (0-1)	.393	.396	.425	.401
Line (1-20)	.141	.150	.136	.154
11 p.s. (0-10)	.181	.187	.183	.194

Method effects are small although they are somewhat higher for reversed questions and the 11 point scale in the first experiment. Method effects are small for line production and ordinal 11 points scales but are considerably higher for the binary scale in the second experiment. This is exactly the opposite of the findings of the first experiment but corresponds to method ordering. The last scale used produces the highest method effects and the first scale used produces the lowest method effects.

Reliability coefficients were the lowest for binary scale in both experiments regardless of scale ordering. the ordinal 11 points scale produced the highest value for the reliability coefficient in the first experiment and the second highest value in the second experiment. Possibly the results in both experiments are, at least in part, influenced by scale ordering.

Several problems are present in this approach (Saris and van Meurs, 1990). Some of them (e.g., the presence of unique variance, instability of the method effects, convergence problems of the program, identification problems of different design testing) are related uniquely to the MTMM approach and were discussed by several authors (e.g., Saris and van Meurs, 1990; Saris and Munnich, 1995). Others (e.g., memory effect, changes of opinion during the interview, question wording, method order effect) are due to survey data collection and to measurement repetitions which were discussed by many survey methodologists, also e.g., Schuman and Presser, 1996; Krosnick and Fabrigar, 1997.

7 Conclusions

By vectorising network matrices we were able to aply traditional approaches to the evaluation of measurement instruments. We used MTMM true score model coefficients to estimate reliability, validity and method effects. The results we obtained demonstrate that the reliability of each single relational matrix is the lowest when measured by the binary scale. This finding can not be explained by the position of the binary scale in the experiment because the scale ordering was reversed in the second experiment. Morover, in order to prevent memory effect we lengthened the time intervals between interviews in the second experiment. Nevertheless, the low quality of the binary scale could still be partly due to the high level of similarity between other two methods. Reliability coefficients were the highest for the ordinal 11 point scale in the first experiment and for line production scale in the second experiment. The ordinal 11 point scale was the last scale used in the first experiment and the first one used in the second experiment. Yet, it is still possible that the scale ordering had some influence on the estimated quality of the scale because one would expect the same pattern in both experiments regardless of scale ordering. The results of both experiments indicate that reversed questions have lower reliability coefficients when compared with the original questions.

The method effects were considerably higher in the second experiment. The method effects coefficients were the highest for the ordinal 11 point scale in the first experiment and for the binary scale in the second experiment. In both experiments, the scales which produced the highest method effects was the last scales used. Therefore, these might be an artificial pattern resulting from scale ordering.

Certain problems emerged from the experimental design. The number of cases in the first experiment was rather small and therefore we had only 156 dyads to work with. Three repetitions were carried out within one interview lasting, on average, only of 23 minutes, which is probably too short for repeated measurements (memory effect). The distribution of variables on both numerical scales is not normal but according to the findings of Satorra (1989) the used procedure is tends to be robust to non-normality.

Lastly it should be noted that certain problems occurred as a result of the particular approach while some were specificaly related to the network. For example, individual measurements (dyads), due to network specific data, are not mutually independent. This problem becomes more serious when standard errors and other statistical tests are applied. However, this was not done in our analysis. Nevertheless, there is a definite need for further experiments to study the stability of the results and there are certainly many other effects on network data quality which merit further study.

8 References

[1] Armor, D.J. (1974): Theta reliability and factor scaling. In H.L. Costner (Ed.): *Sociological Methodology 1973/74,* San Francisco: Jossey-Bass, 17-50.

[2] Bien, W., Marbach, J., and Neyer, F. (1991): Using egocentered networks in survey research. A methodological preview on an application of social network analysis in the area of family research. *Social Networks*, **13**, 75-90.

[3] Calloway, M., Morrissey, J.P., and Paulson, R.I. (1993): Accuracy and reliability of self-reported data in interorganizational networks. Social Networks, **15**, 377-398.

[4] Campbell, D.T., and Fiske, D.W. (1959): Convergent and discriminant validation by the multitrait multimethod matrices. *Psychological Bulletin*, **56**, 81-105.

[5] Cronbach, L.J. (1951): Coefficient alpha and internal structure of test. *Psychometrika*, **16**, 297-334.

[6] Freeman, L.C. (1997): Centrality in social networks, conceptual clarification. *Social Networks*, **1**, 215-282.

[7] Ferligoj, A., and Hlebec, V. (1993): (In)sensitivity of Centrality Measurements in Social Networks. In: Ferligoj, A., and Kramberger, T. (Eds.): Developments in Statistics and Methodology, *Metodološki zvezki*: **9**, 139-145. Ljubljana: FDV.

[8] Ferligoj, A., and Hlebec, V. (1995): Reliability of Network Measurements. In: Ferligoj, A., and Kramberger, T. (Eds.): Contributions to Methodology and Statistics, *Metodološki zvezki*: **10**, 219-232. Ljubljana: FDV.

[9] Hammer, M. (1984): Explorations into the meaning of social network interview data. *Social Networks* **6**, 341-371.

[10] Hlebec, V. (1993): Recall versus Recognition: Comparison of Two Alternative Procedures for Collecting Social Network Data. In: Ferligoj, A., and Kramberger, T. (Eds.): Developments in Statistics and Methodology, *Metodološki zvezki*: **9**, 121-128. Ljubljana: FDV.

[11] Joreskog, K.G., and Sorbom, D. (1986): *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods*. Uppsala: University of Uppsala.

[12] Krosnick, J.A., and Fabrigar, L.R. (1997): Designing rating scales for effective measurement in surveys. In Lyberg, L.et al. (Eds.): *Survey Measurement nd Process Quality*. New York: John Wiley & Sons, INC.

[13] Laireiter, A.R. (1993): The interview on social network and social support "SONET": Studies on it's reliability and validity. Paper presented at 3rd European Conference on Social Network Analysis. Munich, June 1993.

[14] Marsden, P.V. (1993): The reliability of network density and composition measures. *Social Networks*, **15**, 399-421.

[15] Macur, M., and Hlebec, V. (1996): A Combined Qualitative and Quantitative Approach for Analysing the Informal Leadership of a Choral Society. In: Ferligoj, A., and Kramberger, T. (Eds.): Developments in Data Analysis, *Metodološki zvezki*: **12**, 185-194. Ljubljana: FDV.

[16] Neyer, F., Bien, W., Marbach, J., and Tempelton, R. (1991): Obtaining reliable network data about family life: A methodological examination concerning reliability of egocentered networks in survey research. *Connections*, **3**, 14-26.

[17] de Pijper, N.M., and Saris, W.E. (1986): *The Formulation of Interviews Using the Program INTERV.* Amsterdam: Sociometric Research Foundation.

[18] Sabidussi, G. (1966): The centrality index of a graph. *Psychometrica*, **31**, 581-603.

[19] Saris, W.E., and Andrews, F.M. (1991): Evaluation of measurement instruments using a structural modeling approach. In P.P. Biemer et al. (Eds.): *Measurement Errors in Surveys.* New York: Wiley, 575-599.

[20] Saris, W.E., and van Meurs, A. (1990): Memory effects in MTMM studies. In W.E. Saris and A. van Meurs (Eds.): *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies.* Amsterdam: North Holland, 134-147.

[21] Saris, W.E., and Munnich, A. (1995): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments.* Budapest: Eotvos University Press.

[22] Satorra, A. (1990): Robustness issues in the analysis of MTMM and RIM models. In W.E. Saris and A. van Meurs (Eds.): *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies.* Amsterdam: North Holland , 93-118.

[23] Schuman, H., and Presser, S. (1996): Questions and answers in attitude surveys. Thousand Oaks: Sage.

[24] Sudman, S. (1985): Experiments in the measurement of the size of social networks. *Social Networks*, **7**, 127-151.

[25] Sudman, S. (1988): Experiments in Measuring Neighbour and Relative Social Networks. *Social Networks*, **10**, 93-108.

[26] Wasserman, S. and Faust K. (1994): *Social Network Analysis - Methods and Applications.* Cambridge, England:Cambridge University Press.