Research & Methods
ISSN 1234-9224, e-ISSN 2544-0799
Vol. 30 (1, 2021): 3–24
(published in 2022)
The Ohio State University Libraries
Columbus, Ohio, USA
Institute of Philosophy and Sociology
Polish Academy of Sciences, Warsaw, Poland
www.askresearchandmethods.org
https://doi.org/10.18061/ask.v30i1.0001

Influence of the source of information on homophily assessment: data provided by subjects vs data provided by their confidants[1]

Adam Kęska
Institute of Political Studies
Polish Academy of Sciences
Polna 18/20, 00-625 Warsaw, Poland
E-mail: akeska@isppan.waw.pl
https://orcid.org/0000-0002-1603-7815

Dariusz Przybysz
Institute of Philosophy and Sociology
Polish Academy of Sciences
Nowy Świat 72, 00-330 Warsaw, Poland
E-mail: przybysz@ifispan.edu.pl
https://orcid.org/0000-0001-9872-3140

Bogdan W. Mach
Institute of Political Studies
Polish Academy of Sciences
Polna 18/20, 00-625 Warsaw, Poland
E-mail: bmach@isppan.waw.pl
https://orcid.org/0000-0003-3440-7899

Abstract: The present article analyses data on ego-centred networks from a nationally representative sample of Polish citizens (egos) and their regular conversation partners (alters). In the study being used, apart from obtaining information from the respondent (the ego) about characteristics of their friends (alters), the respondent was additionally asked to provide contact details for his/her alters and the interviewers traced them. This allows to compare data concerning alters obtained from ego and from alters themselves. In the first part of the article, a comparison for three characteristics – age, education and occupational status – is provided as well as patterned differences are presented. In the second part, the issue whether the assessment of the strength of homophily depends on the source of

information about the alters is raised. The differences are present, although they are rather slight. The authors observe greater homophily when they rely on the responses of the ego. Based on the results, recommendations for future research on ego-centred networks are formulated.

Keywords: homophily, social networks, friendship networks, sociological methodology, log-linear models

INTRODUCTION

Homophily is a frequently considered and analysed phenomenon in social sciences (McPherson et al. 2001). In general, it is defined as establishing relationships more often between people similar to each other (e.g. belonging to the same socio-professional category) than in a hypothetical situation, if the relationships were concluded randomly (Blau 1977; Marsden 1988; Skvoretz 1991). Homophily is considered in reference to social relations of various types, relatively most often while referring to getting married (Smits et al. 1998; Mare 1991; Raymo and Xie 2000) and patterns of making friends (Lazarsfeld and Merton 1954; Chan and Goldthorpe 2004; Smith et al. 2014; Domański and Przybysz 2012, Leszczensky and Pink 2019, 2020). In recent years, this topic has also been raised in the context of online contacts, relationships between social media users (Wiemer and Lewis 2010; Bisgin et al. 2012; Boutyline and Willer 2017; De Salve et al. 2018; Ladhari et al. 2020; Khanam et al. 2020), and even from the perspective of the spread of COVID-19 (Kadelka and McCombs 2021).

More frequent establishing relationships between people similar to each other may be the result of their individual choices. These choices are related to phenomena, which are known in the literature as "attraction to similar others" and "repulsion from dissimilar others" (Skvoretz 1983; Karpiński and Skvoretz 2015). However, it may also be the result of structural constraints i.e. the size of individual social categories – which make a person more or less likely to meet a person from the same or from another category (Blau 1977). Other conditions related to various forms of social segregation, including spatial segregation, educational or work segregation, are also important. Due to these conditions, the chances of meeting a person with a similar status are much greater in some structural locations. On the other hand, the consequences of the processes described above seem vital too. The greater frequency of relationships between people similar to each other has a significant impact on the formation of the social structure and the formation of social barriers. Less frequent establishing relationships between people from different categories may lead to the formation of stereotypes, limiting cognitive horizons (Smith et al. 2014).

The methodological issues of measuring homophily are not without significance. In the present article, the authors cover one of them. Analyses concerning patterns of marriage or friendship generally concern the comparison of the people pairs characteristics. Usually, variables such as age, education, socio-occupational position, religion, race, and political views are taken into account. For practical reasons, this information is obtained from one of the people who makes up the pair. In other words, the starting point is a study of a random sample of individuals. During the questionnaire interview, respondents are asked about their characteristics, as well as the features of their relatives, spouses, partners or friends. It is indirectly assumed that the respondent is able to determine these features. While in the case of spouses this assumption seems legitimate, it can be argued that there are situations when the respondent does not know exactly all the features of his or her close or distant friends about which he or she is asked during the survey. In addition, there are also other sources of discrepancy between what the respondent says about their friends and how they describe themselves. Any errors in measurement resulting from this may overestimate or underestimate the researcher's estimates of the strength of homophily.

The present article is an attempt to address the following methodological problem: whether relying on the respondents' information is a good enough simplification or maybe getting the information directly from their confidants seems a worthy investment. In order to answer the question, the authors assess the size of the mentioned errors and then their consequences. In the *People in Networks Study* (Mach et al. 2018), apart from obtaining information from the respondent about the age, education and professional situation of their friends, the respondent was additionally asked to provide contact details to these people, the interviewers reached out to these people and then a questionnaire interview was conducted with them. This allows to compile information about the respondent's friends obtained from two sources. In the first part of the article, there is a comparison of three characteristics that are crucial in the homophily context: age, education and professional situation. In the second part, the authors check whether the assessment of the homophily strength depends on the source of information about the variables describing the respondent's acquaintance. Since only three characteristics are taken into consideration, some general conclusions are made with respect to these variables.

DATA

Empirical analyses in this article are based on the results of a survey carried out as a part of the project *People in Networks* (Mach et al. 2018). The study was conducted in 2016 (from May to July). The starting point was a random sample

drawn from the PESEL register[2], among which 1712 interviews were carried out (response rate was approx. 30.4%)[3]. During the interviews, the respondents were asked to name their closest friends, but they could not indicate more than five people. To be more precise, the respondents answered the question "From time to time, we talk about matters important to us – such as relationships with loved ones, difficult decisions, professional problems – with our friends or acquaintances. Please think of up to five people outside your immediate family with whom you most often talk to about matters important to you. On this card, please write their names, initials or other markings". Then, the respondent answered a number of questions regarding each of the indicated friends, including gender, age, education and occupational category.

Based on the data provided by the respondents, the interviewers made attempts to establish contact with these people and conduct an interview with them. In the present article, the main respondents drawn from the PESEL database are called the ego, while each of the friends indicated by the respondent is referred to as the alter. Interviews with 1201 alters were successfully completed[4]. The response rates may seem relatively low for both the main sample and the alter sample. It should be noted that this may be due to the nature of the study. The respondents knew from the advance letter that they would be asked to provide contact information to their friends. The analyses presented by the authors of the study (Mach et al. 2018)) indicate that the realized ego's sample has a similar structure to the population in terms of a number of socio-demographic variables (age, gender, place of residence, level of education). The distributions of selected variables for the alter's sample and the ego's sample were also similar (Mach et al. 2022).

In general, the interviews were carried out using the face to face technique. The questionnaires for egos and alters were very similar. A minor part of the interviews with the alters was carried out by an e-mail or internet questionnaire. From the point of view of the goals we set for ourselves in this article, it is important that for 1121 alters indicated by 746 egos, it was possible to compare information concerning the age, education and occupational status of the alter on the basis of the ego's and the alter's answers.

SOURCES OF DISCREPANCIES IN THE ANSWERS

First of all, it should be noted that the incompatibility of the values of the corresponding variables does not always mean an error in the description of the alter made by the ego or the alter. The specificity, uniqueness and level of procedural complexity of the study almost inevitably results in a certain level of an error in the data structure. It seems also plausible to assume that not all persons in the resulting dataset were properly matched (a person is flagged in the wrong

position in their ego list, or is incorrectly assigned to the wrong ego). This is evidenced by the occurrence of a statistical discrepancy between the sex declared by alters themselves and the sex assigned to them by the respective egos (about 5% of cases of discrepancy). At the same time, the situation in which the respondent does not know the gender of a person who belongs to his close friends would be so exceptional that we considered the assumption of a wrong matching of people in such cases as the most probable solution[5]. Such mismatched pairs were therefore excluded from most of the following analyses, although they also constitute an interesting point of reference[6].

It is worth considering what can be the reason for the discrepancy between the information provided by the ego and the alter if their mutual assignment is correct in regard of the data structure. Firstly, the most obvious case is when the ego has erroneous knowledge of the alter. The alter could, without the knowledge of his ego, change his profession or employment, or get higher education. They can also lie to friends, for example about their age. Then, even if the ego provides information that is fully consistent with his or her state of knowledge, the information is not true. Secondly, we may have to deal with the ego's lack of knowledge concerning the ego's characteristics. This often results in lack of data, but not necessarily. It may be difficult for the subject to admit ignorance, which, combined with the fact that he or she did not have a choice of "don't know", may have led him/her to guess or infer from other properties of the alter. He/she would not be without a chance in this guesswork, for example, knowing the profession, one can guess one's education, and the age can be judged "by eye", the more so because respondents were asked about the age range, not the exact value. However, if he or she was wrong, it results in the discrepancy in the question. Thirdly, the ego may have an accurate and truthful knowledge of the alter, but he/she may not be able to correctly (or in accordance with our expectations) put the alter into the right category. This applies both to education and professional situation. For example, in the minds of some people, a completed teacher training college may be considered a higher education, and some professions are difficult to classify unambiguously. "Theater manager" – does he/she belong to managers and directors or to people of culture and art? A similar problem also applies to the alters themselves – they may not know how to classify their own education or profession. It can also be noticed that this phenomenon may lead to consistent, however, untrue answers. An additional problem is the fact that the respondents (both the ego and the alters) defined their own education more precisely, while the alter's education was described by egos only as belonging to one of the four categories. In order to identify the above effect, the authors use different recodes into broader categories below. A final phenomenon that may be an important potential source of divergence is that the alters inflate their own status.Thereis

a known tendency of the respondents to show themselves in the best light possible. Therefore it is possible that the information from the ego is true, and the information provided by the alter – in some way overstated in status, e.g. a white collar worker defines himself/herself as a specialist, or an unskilled blue-collar worker defines himself as a skilled one.

Attention has also been paid to the role of motivation and the difficulty of the task imposed on ego (Stark and Stocke 2021) for the accuracy of the ego's imputations about the alter (as well as the interdependence of these two factors). The former mainly depends on how important and interesting are the issues to the ego they are asked about. This could also play a role in the present study, although it would be difficult to determine how "interesting" to the respondents is the age, education or profession of their friends. The difficulty of the task would matter for measurement of variables that account for less observable properties related to emotions or attitudes. The variables analysed here would be classified by the above-mentioned authors as "easy".

AGE

Before the comparison of discrepancies with respect to the information about the age of the alter provided by the ego and by the alter is presented, it is worth emphasizing that the first of them was to indicate the age range (i.e. one of the categories given in Table 1) – while alters were asked during the interview to indicate their age in years. The less detailed way of asking about the friend's age seems justified, one would hardly expect the respondent to know their exact value. It can be assumed that the necessity to indicate the exact value of a friend's age in many situations would cause them to answer "Don't know" or refuse to answer such a question. Nevertheless, the different way of asking the age question means that, in order to compare them, it is necessary to recode the detailed information provided by alters into more general categories, which are intervals of years (the same as indicated by the ego). As a consequence, the percentage of age discrepancies estimated by this type of analysis may be underestimated, i.e. lower than if a more precise age was asked.

Table 1 summarises information concerning the age of alter indicated by ego and alter. As it may be noticed, the vast majority of cases are placed on the diagonal of the table, i.e. alter and ego indicate the same category. Such cases account for 82.3% of instances. The inconsistencies usually concern the indication of the adjacent age category: 10.1% of alters indicated that they were older than age indicated by their respective egos, and the opposite situation is observed in 4.4% of cases. Discrepancies greater than a single age bracket concerned about 3% of the alters. It should be added that the basis for percentage are alters, whose

information about their age we have, both from the alter and the ego. In about 7% of cases, there is lack of data, among which 3.8% constitute situations in which the ego was not able or did not want to determine the age of the alter. It is worth adding that respective Kendall's tau-b rank correlation coefficient is 0.89.[7]

Table 1. Alter's age – comparison of answers given by ego and by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

	Alter's age – provided by ego							
Alter's age – provided by alter	Younger than 18 years old	18-29 years old	30-39 years old	40-49 years old	50-59 years old	60-69 years old	70 years old and older	Total
Younger than 18 years old	8	0	0	0	0	0	0	8
18-29 years	1	255	13	3	1	1	0	274
30-39 years	0	20	170	10	4	0	1	205
40-49 years	0	2	20	123	4	1	0	150
50-59 years	0	2	5	21	115	11	2	156
60-69 years	0	1	3	3	27	110	6	150
70 years old and older	0	0	0	0	2	11	33	46
Total	9	280	211	160	153	134	42	989

EDUCATION

In the case of ego's evaluation of alter's education, they were asked to describe it by selecting one of the four categories listed in Table 2, while alter had a more detailed classification at his disposal. To compare this information, it was necessary to group it into four more general categories that were administered to the ego. Footnotes below Table 2 provide details on the rules according to which this grouping was performed. However, it should be highlighted that it is difficult to clearly indicate the rules of transition between the two classifications. For example, the respondent, even if he/she knew exactly that their friend had an incomplete higher education, could indicate either the category of "higher" or "secondary". This is an additional source of error that may increase the number of discrepancies between the row and column variables in Table 2. Therefore, a control analysis was performed with different grouping rules adopted. They are presented in Table A1 in the Appendix.

Table 2. Alter's education – comparison of answers given by ego and by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

	Alter's education – provided by ego				
Alter's education – provided by alter	Elementary	Vocational	Secondary	Higher	Total
Elementary[a]	56	18	10	2	86
Vocational[b]	13	165	30	7	215
Secondary[c]	19	62	278	25	384
Higher[d]	1	4	55	244	304
Total	89	249	373	278	989

[a]Collapsed categories "incomplete elementary", "elementary" "incomplete gymnasium" and "gymnasium".
[b]Collapsed categories "incomplete basic vocational or agricultural training school" and "basic vocational or agricultural training school".
[c]Collapsed categories "incomplete secondary", "secondary school with final exams" and "post-secondary school".
[d]Collapsed categories "teacher training college", "incomplete university", "bachelor or engineer level", "master's degree", "incomplete doctoral studies" and "doctorate".

The percentage of cases in which the ego and alter information about the alter's education agree (cells on the diagonal of Table 2) is 75.1%[8]. In 13.1% of cases, the ego assesses the education of the alter one category lower, in 7.4% of cases, the situation is the opposite. Ego's and alter's responses differ by at least two categories for 4.3% of cases. Table 2 does not take into account 7% of situations in which the respondent did not provide information about the education of a friend or the alter did not provide information concerning him/her. Kendall's rank correlation coefficient is 0.74. Additionally, the results for the alternative grouping (Table A1 in the Appendix) lead to similar conclusions. However, the percentage of consistent cases is slightly lower and amounts to 72.9%, with the tau-b measure being 0.70. An important difference is that less often the ego assesses the alter's education one category lower than the alter himself did (8.6%), the opposite case appears slightly more often (9.7%). This is because the grouping in table A1 is more "conservative" than in Table 2, i.e., people who have incomplete vocational education or incomplete secondary education (among them those who decided not to take exams needed for "maturity diploma") are classified as having primary education only.

OCCUPATIONAL POSITION

The last of the analysed variables concerns the occupational position of the respondent. The ego respondent described their friend's situation by indicating one of the 19 categories presented in Table 3. They describe the alter's occupation or

labour force position if a given person is not working, unemployed, studying or on a pension. When describing own position, alter had the same categories to choose from. It is worth noting that in stratification studies, more often the occupational position is determined by means of open questions, i.e. questions about the name of the job, activities performed at work, whether it is an employed or self-employed job, and whether a given person supervises work of others, etc. Based on this information, the occupation is coded according to a detailed classification[9]. The advantage of this procedure is that individual occupations are grouped into more general categories according to specific theoretical and methodological assumptions. It can be assumed that in the case of a pre-categorised question, two respondents working on the same job may indicate two different general categories, one of them for whatever reason selects the category "skilled worker", while the other "unskilled worker". It should be noted, however, that coded open-ended questions is a costly and time-consuming endeavor. Moreover, it seems unrealistic that when asked about the occupation of a friend, the respondent would be able to answer detailed open-ended questions. For this reason, asking about the occupation of a friend by means of a pre-categorized question may be justified and useful in many situations.

Cases where both ego and alter indicated the same occupational category accounted for 64.3% of the alters listed in Table 3. In addition to this, 3.9% of alters were people with missing data, i.e., alter or ego did not choose any of the categories listed. It is worth paying attention to the types of non-consistency which occurred most frequently. Relatively often, alters who described themselves as skilled manual workers were classified by the ego as unskilled. The opposite happened somewhat less frequently. Likewise, persons describing themselves as "pensioners" were classified by the ego as retirees. Quite often, those who indicated that they were farmers, entrepreneurs, or traders were classified by the ego as housekeepers.

Table 3. Alter's occupational *People in Networks Study* (Mach et al. 2018)) situation – comparison of answers given by ego and by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Alter's occupational situation – provided by alter	Alter's occupational situation – provided by ego																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total
1. Pupils or students	112	0	1	1	4	0	0	0	0	3	4	0	1	1	0	0	0	1	0	128
2. Retirees	2	128	2	3	1	3	2	0	0	3	0	0	0	2	2	4	1	2	0	155
3. Not employed pensioners	1	10	12	1	0	3	1	1	0	0	4	0	0	1	0	0	0	0	0	34
4. Housekeepers	1	2	3	33	7	2	1	0	0	1	1	1	2	3	0	0	0	0	0	57
5. Unemployed	5	3	1	7	26	1	0	0	0	4	1	0	0	0	0	0	0	0	0	48
6. Self-employed farmers	0	1	0	7	0	24	1	0	0	3	0	0	0	0	0	0	0	0	0	36
7. Entrepreneurs without employees	1	1	0	1	0	0	7	2	0	1	1	0	2	0	0	0	0	0	1	17
8. Entrepreneurs employing at least one employee	0	2	0	0	0	1	3	8	0	3	1	0	1	1	0	0	0	1	0	21
9. Representatives of the uniformed services	0	0	0	0	0	0	0	0	5	0	1	0	1	1	0	0	0	0	0	8
10. Manual workers performing simple tasks	1	1	0	2	2	2	2	1	0	49	9	0	5	1	1	0	0	2	0	78
11. Skilled manual workers	0	1	1	0	2	3	0	3	1	27	63	0	8	1	3	1	1	3	0	118
12. Female or male nurses, physiotherapists	1	0	0	0	0	0	0	0	0	0	0	9	0	1	0	0	0	0	0	11
13. Sellers or service workers	2	3	3	6	2	1	2	1	0	6	8	1	41	3	2	1	0	5	0	87
14. Officials, administration staff, secretaries	2	1	0	4	0	0	0	1	0	3	4	2	0	47	1	3	0	4	1	73
15. Technicians	1	0	0	1	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	6
16. Teachers, educators	1	0	0	1	0	0	1	0	0	0	0	0	2	0	0	50	0	1	0	56
17. People of art and culture	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	5	0	0	7
18. Specialists with higher education	1	2	0	1	0	0	0	0	1	0	1	0	1	8	0	2	0	32	1	50
19. Directors, presidents, managers	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	1	0	0	5	9
Total	131	156	23	68	44	40	20	19	7	103	100	13	64	73	10	62	7	51	8	999

Table 4. Alter's occupational situation - comparison of answers given by ego and by alter – grouped categories (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Alter's occupational situation – provided by alter	Alter's occupational situation – provided by ego									
	1	2	3	4	5	6	7	8	9	Total
1. Pupils or students	112	1	5	0	7	1	0	1	1	128
2. Retirees or pensioners	3	152	5	6	7	0	4	9	3	189
3. Unemployed	6	9	73	3	7	2	1	4	0	105
4. Self-employed farmers	0	1	7	24	3	0	1	0	0	36
5. Manual workers	1	3	6	5	148	14	6	7	6	196
6. Service workers	2	6	8	1	15	47	3	8	5	95
7. Entrepreneurs	1	3	1	1	6	3	20	1	2	38
8. White-collar workers	5	1	6	0	8	2	3	115	6	146
9. Specialists, people of culture or directors	1	3	1	0	2	2	1	13	43	66
Total	131	179	112	40	203	71	39	158	66	999

Many analyses using measures of job status apply more general categories, especially while analysing social mobility or marital homogamy. The job/occupation is then treated as an indicator of social position and it is important to use it in identifying the most important social barriers. Grouped categories for both ego and alter responses are presented in Table 4. For data in this form, the percentage of agreeing responses was 69%.

INFLUENCE OF THE INFORMATION SOURCE ON THE MEASUREMENT OF HOMOPHILY

In this section, we will analyse whether the source of information (ego or alter) concerning friend's traits influences patterns and strength of homophily measures. In the Appendix we have included tables A2a-A4b in which information about ego traits and alter traits are summarised, while the variables related to a friend are based either on ego's responses (tables A2a, A3a, A3c, A4a) or alter's responses (tables A2b, A3b, A3d, A4b).

Table 5 summarises the age-related homophily indices[10]. As can be noted, the friends indicated by the respondents usually came from the same age group. The percentage of cases on the main diagonal of tables A2a and A2b is 54.9% and 56.0%, respectively – so they hardly depend on whether the information about the acquaintance's age comes from the ego or from the alter. If the selection

of friends was random, this percentage would be 18.8%[11]. The situation of the ego being older than the alter is slightly more common than the reverse, and the difference is slightly greater when the age information comes from the ego than from the alter. The rank correlation measure has practically the same value, it is slightly higher when we use information from the ego. Likewise, there are no major differences when the strength of homophily for successive age categories is measured using the interaction parameters of the log-linear model typically used to measure the strength of homophily (Marsden 1988). These parameters are given in the logarithmic metrics, positive values indicate that a given combination of the values of both variables occurs more often than in the hypothetical situation of their independence. For example, a value of 2.91 indicates that the situation when both alter and ego are less than 28 years old is 18.3 (i.e. exp (2.91)) times more often than if the egos' and alters' ages were fully independent.

Table 5. Indicators of the relationship between ego's age and alter's age (source: *People in Networks Study* (Mach et al. 2018))

Indicator	Alter's responses	Ego's responses	Expected value of an indicator with random matching of acquaintances
Same age bracket of ego and alter (%)	54.9	56.0	18.8
Alter is younger than ego (%)	23.5	25.1	40.6
Alter is older than ego (%)	21.5	18.9	40.6
Rank correlation coefficient	0.66	0.67	0
Diagonal parameters of the quasi-symmetry model[a]			
29 years old and younger	2.91	2.91	0
30-39 years old	1.87	1.80	0
40-49 years old	0.95	0.97	0
50-59 years old	0.65	0.89	0
60-69 years old	1.33	1.31	0
70 years old or older	2.45	2.55	0

[a]The fit of the quasi-symmetry model is satisfactory. If we rely on alter's responses regarding his age the statistic L^2=4.1, with a number of degrees of freedom of df=10, p-value= 0.9425. When we rely on ego's responses: L^2=10.3, p-value=0.4186.

Table 6. Indicators of the relationship between ego's education and alter's education (source: *People in Networks Study* (Mach et al. 2018))

Indicators	Alter's responses	Ego's responses	Expected value of an indicator with random matching of acquaintances
Same level of education of alter and ego (%)	51.8	53.3	28.9
Alter has a lower level of education than ego (%)	19.0	21.5	35.6
Alter has a higher level of education than ego (%)	29.1	25.2	35.6
Rank correlation coefficient	0.48	0.52	0
Diagonal parameters of the quasi-symmetry model[a]			
Elementary	1.43	1.74	0
Vocational	0.59	0.63	0
Secondary	0.31	0.33	0
Higher	1.93	1.95	0

[a]The fit of the quasi-symmetry model is satisfactory. If we rely on the alter's responses regarding his education the statistic L^2=2.92, df=6, p-value=0.4041. When relying on ego's responses: L^2=0.79, p-value=0.8521.

In the case of education (Table 6), the share of cases where ego and alter do not differ is slightly over 50%. This indicator is a little higher when the information about a friend's education comes from the ego. Relatively more respondents indicate that their friend is better educated than they are, and the discrepancy is smaller if the acquaintance's education is based on ego responses. A stronger tendency to homophily, both measured by the rank correlation and the parameters of the log-linear model, occurs when the education of the alter is indicated by the ego than by the alter itself. However, it should be noted that these discrepancies are not very large. They are the greatest when we compare the strength of homophily among people with primary education[12].

Table 7 shows the rates of occupational homophily. In 35.8% of cases, ego and alter indicate the same of 19 categories. This percentage is higher when the information about the situation of a friend is based on the ego's answers. The results are similar when we consider the occupational situation in a more general sense, i.e. after distinguishing 9 categories. Table 7 also shows the diagonal parameters estimated in the quasi-symmetry model. Their values are (with one exception) positive, which indicates that relationships between people similar to each other occur more often than in the hypothetical situation of no relationship between the ego and alter's status situation: this applies in particular to students, farmers, white-collar workers and specialists. In general, the power of homophily is greater

if information about a friend's profession is based on the ego's responses rather than alter's own declarations. It is worth emphasising, however, that the basis of the analysis is a 9 by 9 table in which many cells have very small numbers. This is mainly the case when pupils or students indicate people from other categories as friends, it happens very rarely. This may indirectly affect the accuracy of the estimated parameters. For this reason, an additional analysis was made from which pupils and students were excluded. The homophily parameters based on the 8 by 8 table are shown in parentheses in Table 7. While there are some differences between the two analyses, the main conclusions remains the same: the strength of homophily for most categories is greater when based on ego responses. This applies in particular to the categories: farmers, entrepreneurs, white-collar workers and specialists.

Table 7. Indicators of the relationship between ego's and alter's occupational situation (source: *People in Networks Study* (Mach et al. 2018))

Indicators	Alter's responses	Ego's responses	Expected value of an indicator with random matching of acquaintances
Same category for ego and alter – 19 categories (%)	35.8	39.8	8.8
Same category for ego and alter – 9 categories (%)	43.9	46.6	14.3
Diagonal parameters of the quasi-symmetry model[a] and diagonal parameters of the quasi-symmetry model after the exclusion of pupils and students (in brackets)[b]			
Pupils or students	3.85 (N/A – excluded)	3.94 (N/A – excluded)	0
Retirees and pensioners	1.88 (1.58)	1.95 (1.67)	0
Housekeepers and unemployed	1.09 (1.28)	0.77 (0.96)	0
Self-employed farmers	2.11 (1.88)	2.65 (2.47)	0
Manual workers	0.88 (1.00)	0.95 (1.02)	0
Service workers	1.34 (1.49)	1.21 (1.36)	0
Entrepreneurs	-0.10 (-0.19)	1.02 (0.86)	0
White-collar workers	1.20 (1.08)	1.62 (1.49)	0
Specialists, people of culture, directors	2.03 (1.85)	2.39 (2.24)	0

[a] The fit of the quasi-symmetry model is satisfactory. If we rely on alter''s responses regarding his age: statistic L^2=30.52, df=28, p-value= 0.3389. When we rely on ego's answers: L^2=29.13, p-value=0.4058.

[b] Using alter responses the statistics of model fit to data are L^2=18.13, df=21, p-value=0.6408, with ego responses L^2=20.83, p-value=0.4694.

SUMMARY

It is not easy to assess the overall reliability of information provided by the respondents about their friends and reliability of comparing that information with what friends say themselves. When making such an assessment, one should bear in mind how different the sources of discrepancies may be. They include also the level of complexity of the study, which results in an incorrect matching of egos and alters, which would undermine the statistical relationship between the variables "by ego" and "by alter" even if both sources of information were perfect. Nevertheless, especially since we tried to eliminate this influence by elimination of pairs with gender incompatibility, it can be stated with strong conviction that we did not deal with full reliability here. The proportion of agreeing responses ranged from only 64% for the most problematic variable with 19 categories reporting alter's occupational situation to 82% for the age category. A positive trait of consistency in the responses is the fact that, among the diverging responses, the discrepancies were not often great. In the case of ordinal variables (age and education), we observed that the answers most often differed by at most one category, with greater discrepancies in only a few percent of cases.

The question of the influence of the information source on the measurement of homophily is closely related to the above. Since we obtain slightly different information about the alter from the ego than from the alter themselves, it was expected that the results of analyses taking into account one or the other source of information would differ. As can be seen from Tables 5-7, in the case of measuring homophily, these differences are present, although they are rather small. More importantly, everything indicates that these are not random exceptions. The authors observe greater homophily when they rely on the responses of the ego. All homophily indices based on the percentage of dyads belonging to the same category, as well as the measures of rank correlation, are higher for the ego response. The only deviations from this rule were obtained for some parameters of the quasi-symmetry model, i.e. those relating to specific categories of the variable. We are probably dealing here with the phenomenon of perceiving our friends as more similar to us than they really are. Various group homogeneity theories are called here; empirical analyses of simultaneous data obtained from ego and alter (Mach et al. 2022) have to acknowledge all of them. Thus, the use of information from respondents about their friends is burdened with the risk of distorting the results of the analyses. It is recommended to reach out to these friends and get information directly from them.

NOTES

1 Empirical analyses in this article are based on the results of a survey carried out as a part of the project "People in networks: the impact of the social context on an individual and its role in shaping the structure of society". The project was carried out at the Institute of Political Studies of the Polish Academy of Sciences as part of a grant financed by the National Science Center (No. 2013/10 / M / HS6 / 00526) in 2014-2018. Bogdan W. Mach was the head of the study. The research was carried out by the Public Opinion Research Center. More details on the study can be found in: Mach, Manterys, Sadowski 2018.

2 PESEL (Universal Electronic System for Registration of the Population) is a register maintained by the Polish government. The PESEL identification number is mandatory for all permanent residents of Poland and for temporary residents living in Poland for over 2 months. The advantage of using PESEL register as a sampling frame is that the register is updated on a daily basis and is highly accurate. PESEL register contains a lot of information relevant to sampling like sex and date of birth, the registered address, which, although not always the same as actual place of residence, is the starting point for establishing contact with the sampled respondent.

3 More precisely, 1,587 interviews from a random sample from the PESEL registry were realised, with the sample being clustered when it came to smaller towns, especially villages. In the last phase of the survey, the sample was supplemented by 125 people selected by the quota selection method.

4 Respondents from the main sample indicated a total of 4,535 acquaintances, so the percentage of completed interviews was about 26.5%. However, it should be noted that respondents drawn from the PESEL sample often refused to provide contact details of their friends (they were obtained for about 70% of the alters).

5 There are of course some exceptions. Gender mismatch may be related to alter's being transgender or transsexual. Gender could also be miscoded. Both phenomena should be considered rare enough that they could not explain the number of cases of gender mismatch we have dealt with.

6 It should also be noted that, most likely, among the remaining alters for which the gender is consistent with the ego-indicated, there is a similar number of erroneous alter-ego matches. Since we have identified cases where the ego indicates a gender other than the alter himself/herself, there may have been cases where the gender information "coincidentally" matches. These cases would be very difficult to identify and are still included in the data set. This results in a slight underestimation of the reliability of information about the alter coming from the ego.

7 As indicated above, the analysis excluded cases for which the information on the gender of the alter was inconsistent. We checked that for these people the percentage of consistent ascribing information about age by ego and alter was 39.1%, while the rank correlation coefficient was 0.55. The presence of further cases of inappropriate alter-ego matches in the remaining data set must therefore undermine the overall level of consistency of the responses.

8 A very similar result was obtained by Stark and Stocke (2021), the corresponding proportion of education compliance in their study was 73%.

9 In Poland, the Social Classification of Occupations is used (Pohoski et al. 1974; Domański et al. 2009).

10 A random sample included people over 18 years of age. Some of the respondents named people under the age of 18 as friends, therefore the variable "Age of alter" has one

category more than the variable "Age of ego". There were only a few such cases in the dataset, therefore, when analysing age homophily, we combined such cases with the 18–29 years category. This was necessary for the estimation of log-linear models.

11 It is the sum of the percentages on the diagonal of the table, assuming stochastic independence in a situation where the distributions of the marginal row and column variable were the same and corresponded to the distribution of the variable "ego age category".

12 Larger discrepancies in the conclusions regarding the measurement of homophily occur when the detailed classification of education is grouped according to "formal" rules (Tables A3c and A3d and A5 in the Appendix).

REFERENCES

Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. Nowy Jork: Free Press.

Bisgin, Halil, Nitin Agarwal, and Xiaowei Xu. 2012. "A study of homophily on social media." *World Wide Web* 15(2): 213–232, https://doi.org/10.1007/s11280-011-0143-3

Boutyline, Andrei, and Robb Willer. 2017. "The social structure of political echo chambers: Variation in ideological homophily in online networks." *Political Psychology* 38(3): 551–569, https://doi.org/10.1111/pops.12337

Chan, Tak Wing, and John H. Goldthorpe. 2004. "Is there a status order in contemporary British society? Evidence from the occupational structure of friendship." *European Sociological Review* 20(5): 383–401, https://doi.org/10.1093/esr/jch033

Domański, Henryk, and Dariusz Przybysz. 2012. Sociological Tools. "Friendship patterns and social inequality." *International Journal of Sociology* 42(1): 31–59, https://doi.org/10.2753/IJS0020-7659420102

Domański, Henryk, Zbigniew Sawiński, and Kazimierz M. Słomczyński. 2009. *Sociological Tools. Measuring Occupations. New Classifications and Scales*. Warszawa: Wydawnictwo Instytutu Filozofii i Socjologii PAN.

Kadelka, Claus, and Audrey McCombs. 2021. "Effect of homophily and correlation of beliefs on COVID-19 and general infectious disease outbreaks." *PloS One* 16(12): e0260973, https://doi.org/10.1371/journal.pone.0260973

Karpiński, Zbigniew, and John Skvoretz. 2015. "Rejected by the 'Other': Integrating Theory with Method in the Study of Intergroup Association." *Sociological Theory* 33 (1): 20–43, https://doi.org/10.1177/0735275115572687

Khanam, Kazi Zainab, Gautam Srivastava, and Vijay Mago. 2020. "The homophily principle in social network analysis." *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 37(4), Article 111.

Ladhari, Riadh, Elodie Massa, and Hamida Skandrani. 2020. "YouTube vloggers' popularity and influence: The roles of homophily, emotional attachment, and expertise." *Journal of Retailing and Consumer Services* 54: 102027, https://doi.org/10.1016/j.jretconser.2019.102027

Lazarsfeld, Paul F., and Robert K. Merton. 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis." Pp. 18-66 in *Freedom and Control in Modern Society*, edited by Morroe Berger. New York: Van Nostrand.

Leszczensky, Lars, and Sebastian Pink. 2019. "What drives ethnic homophily? A relational approach on how ethnic identification moderates preferences for same-ethnic friends." *American Sociological Review* 84(3): 394–419, https://doi.org/10.1177/0003122419846849

Leszczensky, Lars, and Sebastian Pink. 2020. "Are birds of a feather praying together? Assessing friends' influence on Muslim youths' religiosity in Germany." *Social Psychology Quarterly* 83(3): 251-271, https://doi.org/10.1177/0190272520936633

Mach, Bogdan W., Aleksander Manterys, and Ireneusz Sadowski. 2018. *Individuals and Their Social Contexts*. Warszawa. Institute of Political Studies.

Mach, Bogdan W., John, E. Jackson, and Ireneusz Sadowski. 2022. "Estimating peer political influence with large N observational data on ego-centered social networks." *Social Networks* 70: 198–207, https://doi.org/10.1016/j.socnet.2022.01.002

Mare, Robert D. 1991. "Five decades of educational assortative mating." *American Sociological Review* 56(1). 15–32, https://doi.org/10.2307/2095670

Marsden, Peter V. 1988. "Homogeneity in confiding relations." *Social Networks* 10(1): 57–76, https://doi.org/10.1016/0378-8733(88)90010-X

Pohoski, Michał, Kazimierz M. Słomczyński, and Krystyna Milczarek. 1974. *Społeczna Klasyfikacja Zawodów*. Warszawa: IFiS PAN.

McPherson, Miller, Lynn Smith-Lovin, James M. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* (27): 415–444, https://doi.org/10.1146/annurev.soc.27.1.415

Raymo, James M., and Yu Xie. 2000. "Temporal and regional variation in the strength of educational homogamy." *American Sociological Review* 65(5): 773–781, https://doi.org/10.2307/2657546

De Salve, Andrea, Barbara Guidi, Laura Ricci, and Paolo Mori. 2018. "Discovering homophily in online social networks." *Mobile Networks and Applications* 23(6): 1715–1726, https://doi.org/10.1007/s11036-018-1067-2

Skvoretz, John. 1983. "Salience, Heterogeneity, and Consolidation of Parameters: Civilizing Blau's Primitive Theory." *American Sociological Review* 48: 370–85, https://doi.org/10.2307/2095228

Skvoretz, John. 1991. "Theoretical and methodological models of networks and relations." *Social Networks* 13: 275–300, https://doi.org/10.1016/0378-8733(91)90009-I

Smith, Jeffrey A., Miller McPherson, and Lynn Smith-Lovin. 2014. "Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004." *American Sociological Review* 79: 432–456, https://doi.org/10.1177/0003122414531776

Smits, Jeroen, Wout Ultee, and Jan Lammers. 1998. "Educational homogamy in 65 countries: An explanation of differences in openness using country-level explanatory variables." *American Sociological Review* 63(2): 264–285, https://doi.org/10.2307/2657327

Stark, Tobias H., and Volker Stocké. 2021. "Predicting data quality of proxy reports in egocentric network studies." *Social Networks* 66: 38–49, https://doi.org/10.1016/j.socnet.2021.01.007

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116 (2): 583–642, https://doi.org/10.1086/653658

APPENDIX

Table A1. Alter's education – comparison of answers given by ego and by alter – formal grouping (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Alter's education – provided by alter	Alter's education – provided by ego				
	Elementary	Vocational	Secondary	Higher	Total
Elementary[a]	56	18	10	2	86
Vocational[b]	13	165	30	7	215
Secondary[c]	19	62	278	25	384
Higher[d]	1	4	55	244	304
Total	89	249	373	278	989

[a] Collapsed categories "incomplete elementary", "elementary" "incomplete gymnasium", "gymnasium", "incomplete basic vocational or agricultural training school" and "incomplete secondary".
[b] Complete "basic vocational or agricultural training school" only.
[c] Collapsed categories "secondary school with final exams", "post-secondary school", "teacher training college" and "incomplete university".
[d] Collapsed categories "bachelor or engineer level", "master's degree", "incomplete doctoral studies" and "doctorate".

Table A2a. Ages of ego and alter with alter's age provided by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's age	Alter's age – provided by alter						
	29 years old and younger	30-39 years old	40-49 years old	50-59 years old	60-69 years old	70 years old and older	Total
29 years old and younger	233	46	13	8	6	1	307
30-39 years old	27	97	19	12	8	1	164
40-49 years old	14	35	62	33	8	3	155
50-59 years old	8	22	46	72	42	8	198
60-69 years old	4	14	10	30	80	14	152
70 years old and older	1	0	5	10	16	21	53
Total	287	214	155	165	160	48	1029

Table A2b. Ages of ego and alter with alter's age provided by ego (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's age	Alter's age – provided by ego						
	29 years old and younger	30-39 years old	40-49 years old	50-59 years old	60-69 years old	70 years old and older	Total
29 years old and younger	243	42	15	8	4	1	307
30-39 years old	30	108	18	10	5	2	164
40-49 years old	13	35	67	26	10	3	155
50-59 years old	9	23	50	70	31	3	198
60-69 years old	3	15	11	37	66	15	152
70 years old and older	2	0	5	6	18	19	53
Total	287	214	155	165	160	48	1029

Table A3a. Ego and alter education with alter's education provided by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's education	Alter's education – provided by alter				
	Elementary	Vocational	Secondary	Higher	Total
Elementary	42	36	35	9	122
Vocational	26	107	97	29	259
Secondary	23	64	201	94	382
Higher	1	17	65	183	266
Total	92	224	398	315	1029

Table A3b. Ego and alter education with alter's education provided by ego (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's education	Alter's education – provided by ego				
	Elementary	Vocational	Secondary	Higher	Total
Elementary	47	36	27	5	115
Vocational	25	118	81	22	246
Secondary	15	85	194	87	381
Higher	2	16	77	186	281
Total	89	255	379	300	1023

Tabela A3c. Ego and alter education with alter's education provided by alter – formal grouping (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's education	Alter's education – provided by alter				
	Elementary	Vocational	Secondary	Higher	Total
Elementary	82	64	58	20	224
Vocational	48	94	72	25	239
Secondary	45	40	164	82	331
Higher	10	14	57	154	235
Total	185	212	351	281	1029

Table A3d. Ego and alter education with alter's education provided by ego – formal grouping (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's education	Alter's education - provided by ego				
	Elementary	Vocational	Secondary	Higher	Total
Elementary	58	76	66	19	219
Vocational	22	107	77	21	227
Secondary	7	57	180	85	329
Higher	2	15	56	175	248
Total	89	255	379	300	1023

Table 4a. Alter's and ego's occupational situation based on responses given by alter (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's occupational situation	Alter's occupational situation – provided by alter									
	1	2	3	4	5	6	7	8	9	Total
1. Pupils or students	99	3	7	1	7	7	2	5	1	132
2. Retirees or pensioners	0	124	13	14	20	8	5	21	7	212
3. Unemployed	10	12	37	4	28	9	7	10	3	120
4. Self-employed farmers	0	6	4	10	7	2	3	4	1	37
5. Manual workers	14	19	27	4	85	18	8	19	4	198
6. Service workers	3	3	7	0	9	20	2	10	2	56
7. Entrepreneurs	1	9	5	2	20	9	3	10	6	65
8. White-collar workers	1	15	7	2	15	22	5	62	20	149
9. Specialists, people of culture or directors	1	12	2	2	7	4	5	11	26	70
Total	129	203	109	39	198	99	40	152	70	1039

Table 4b. Alter's and ego's occupational situation based on responses given by ego (frequency distribution, source: *People in Networks Study* (Mach et al. 2018))

Ego's occupational situation	Alter's occupational situation – provided by ego									
	1	2	3	4	5	6	7	8	9	Total
1. Pupils or students	102	2	4	0	9	5	1	4	1	128
2. Retirees or pensioners	1	113	14	16	19	3	3	24	6	199
3. Unemployed	14	11	34	4	28	11	7	9	1	119
4. Self-employed farmers	0	6	5	12	8	0	1	2	1	35
5. Manual workers	8	15	28	4	89	18	9	19	2	192
6. Service workers	3	2	12	1	12	12	1	6	3	52
7. Entrepreneurs	1	8	6	1	19	6	8	7	6	62
8. White-collar workers	1	10	9	1	14	13	5	75	15	143
9. Specialists, people of culture or directors	1	11	3	1	4	3	4	10	31	68
Total	131	178	115	40	202	71	39	156	66	998

Table A5. Indicators of the relationship between ego's education and alter's education – formal grouping (source: *People in Networks Study* (Mach et al. 2018))

Indicators	Alter's responses	Ego's responses	Expected value of an indicator with random matching of acquaintances
Same level of education of alter and ego (%)	48.0	50.8	25.8
Alter has a lower level of education than ego (%)	20.8	15.5	37.1
Alter has a higher level of education than ego (%)	31.2	33.6	37.1
Rank correlation coefficient	0.44	0.50	0
Diagonal parameters of the quasi-symmetry model[a]			
Elementary	0.82	1.35	0
Vocational	0.71	0.70	0
Secondary	0.38	0.46	0
Higher	1.54	1.81	0

[a]The fit of the quasi-symmetry model is satisfactory. If we rely on the alter's responses regarding his education the statistic L^2=3.59, df=6, p-value= 0.3083. When relying on ego's responses: L^2=3.48, p-value=0.3239.