Sociometric Fieldwork Report

Children of Immigrants Longitudinal Survey in 4 European Countries

Wave 1

Hanno Kruse and Konstanze Jacob

Contents

Introduction

The sociometric module in the first wave of the CILS4EU data aims at capturing the social structure within respondents' classrooms. Asking all students who participated in the main questionnaire[1] about different types of relations within the classroom provides a detailed image of their classroom networks. These network data do not only serve as information about students' most significant education-related context, they can also be a convenient indicator for their levels of social embeddedness, the relative amount of cross-ethnic class friendships being one example.

Compared to other network-related measurements, like ego-centered networks, sociometric questionnaires are able to capture only those social relations that are held within one predefined context, in this case the classroom. This restriction, however, provides certain advantages: All members of the context can be surveyed, thus providing information about the complete structure of the network, including relations present between actors, those that are absent and the position of each actor in the network relative to the others. Also, information about characteristics of the alters (those that are nominated) rely on self-reports and do not depend on reports of ego (the nominating person), thus providing higher data validity.

In order to fully capture adolescents' social embeddedness in their classrooms respondents reported different types of relations:

- best friends
- most popular classmates
- classmates that they would not want to sit by
- classmates that they often spend time with outside of school
- classmates that they are sometimes mean to
- classmates that are sometimes mean to them
- classmates that they sometimes do homework with
- classmates that live within a five minute walk from their home
- classmates whom their parents know
- classmates whose parents get together with / call their own parents

Data collection

The sociometric module was issued in the classroom context, implying that students could only account for the relations to their classmates. The survey procedure was as follows: First, classroom-specific nomination lists were defined containing respondents' nomination possibilities; with personal socio-ids being assigned to each classmate. In the Netherlands and Sweden, all students in the sampled classes, in other words the complete sociometric gross sample[2], could be nominated. In various German and English schools non-participating classmates (refusals or absent students) were crossed off the list, so that students could nominate participating classmates only. For more information refer to the subsection "Nomination of absent students".

Respondents received a sociometric questionnaire (see figure 1 for the first five questions) and were then asked to report all those socio-ids from the nomination list that corresponded to classmates for whom the questioned type of relation applied. If the question did not refer to any of their classmates they were asked to write down a big "X" in the empty space.

---

[1] For a descriptive overview of sample sizes and response rates regarding the main questionnaire refer to the technical report of wave 1.

[2] For more information about the sociometric gross sample, refer to the section "A note on the sociometric gross sample".

Figure 1: Sociometric Questionnaire Wave 1, Questions 1-5

Structure of the sociometric data

For each item in the sociometric questionnaire, the dataset contains several nomination variables indicating the youthids of the classmates that were nominated. The maximum number of these nomination variables is either restricted to one ($y1\_bf\_1$, see question 2 in figure 1), to five ($y1\_bfs\_*, y1\_sit\_*, y1\_pos\_*,$ see questions 1, 3 and 5 in figure 1) or it is unrestricted and thus determined by the empirical maximum number of nominations (applies to all other questions, for example question 5 in figure 1). If students made more nominations than allowed for we kept only the first answers they reported.

To illustrate the structure of the nomination variables we provide an example of a classroom consisting of seven students in figure 2. Panel (a) displays the variable structure of the five best friend nominations, the other panels indicate how these data would translate into a sociomatrix (panel (b)) and into a sociograph (panel (c)).
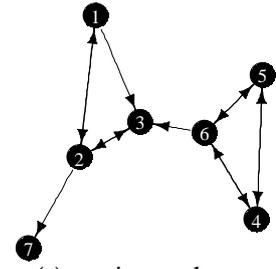
Figure 2: Exemplary Classroom Network Data

youthid	y1_bfs_1	y1_bfs_2	y1_bfs_3	y1_bfs_4	y1_bfs_5
01	02	03	-88	-88	-88
02	01	03	07	-88	-88
03	02	-55	-88	-88	-88
04	05	06	-55	-88	-88
05	04	06	-88	-88	-88
06	03	04	05	-88	-88
07	-77	-77	-77	-77	-77

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(a) variable structure   (b) sociomatrix   (c) sociograph

For each sociometric question we added another variable which indicates whether a respondent explicitly reported to have no relation of the specific kind asked about. This was the case if the respondent wrote down a big "X" in the respective empty space. The names of these variables all end with "_0". For example, y1_bfs_0 indicates if a respondent reported to have one or more best friends in the classroom (actual nomination(s) made), to have no best friends ("X" in the empty space) or whether it was a case of item nonresponse (nothing filled in). The variables ending with "_0" are coded as follows:

    0 =   at least one classmate nominated
    1 =   relation does not apply to any classmate
  -88 =   no answer

Please note, that for England category -88 includes both item non-response cases and those who explicitly nominated no classmate in the respective question.

Missing values in the nomination variables

Within the nomination variables three different types of missing values can occur: "no answer" (-88), "not applicable" (-77) and "other missing" (-55). They are described in further detail in the following.

No answer (-88)

In many cases respondents did not nominate the maximally possible number of classmates. For example, instead of nominating their five best friends in the classroom they only nominated two (see student "01" in figure 2). For these cases all "empty" nomination variables are assigned to the missing category "no answer" (-88).

Not applicable (-77)

As mentioned before, there are respondents who nominated no classmates at all in one or more questions, either because they stated not having such relations as asked ("_0"-variable has value of 1) or because of item non-response ("_0"-variable has value of -88). In both cases, the missing category of all their nomination variables is then "not applicable" (-77).

In the exemplary classroom from figure 2 student "07" reported to have no best friends in the classroom (wrote down an "X" in the respective empty space), which leads to a value of -77 in his own nomination variables. Since he participated in the survey he himself could be nominated by others, as was done by student "02".

Other Missing (-55)

Several answers were considered as implausible and therefore assigned to the "other missing" category (-55). This category includes:

- all nominations that do not correspond with an existing youthid in the sociometric gross sample: Several IDs were nominated that are not part of the gross sample, which consists of all students attending one of the classrooms that is part of the survey sample.[3]
- all self-nominations (exception: question about most popular students): Some respondents nominated themselves, which is logically impossible for most of the sociometric questions.
- all nominations that a respondent made more than once: Some respondents nominated the same classmate more than once within the same question. Only the first of these multilple nominations was kept.
- further nominations that could not be linked to a valid youthid: In some cases, respondents did not report a socio-ID but real names or nicknames that could not be linked to actual classmates.

Additional variables

Besides the nomination variables outlined above, the dataset contains several variables that provide additional information about the structure of the survey units or about respondents' nomination behavior. In the following both types of variables are laid out in greater detail.

Structure of the survey units

Survey unit ID (classid_yc, individual level)

The variable classid_yc identifies the survey unit in which respondents could make their nominations. All respondents who could nominate each other in the sociometric module share the same value. Figure 3 displays different ways of how survey units (grey-shaded areas) overlap with classroom boundaries, exemplified by two classes consisting of three students each. For the vast majority of the sample, survey unit definitions correspond to the case displayed in panel (a) where the survey units are equal to the classroom context. Here, nominations are restricted to classmates only (as indicated by the arrows).

In 12 English schools and one German school, however, classid_yc differs from respondents' regular classid, which indicates that for these cases nomination possibilities were not exclusively bounded to the classroom context. This could be the case due to several reasons:

- merged nomination list of two classrooms (see figure 3b): Two classrooms were surveyed together and respondents could therefore nominate not only their own classmates but also those in the other class.
- respondents moved to another survey unit (see figure 3c): Some students took part in the sociometric survey in another classroom than their own. They are therefore part of a different survey unit than their classmates, which implies that they can only nominate and be nominated by students of the other survey unit.

---

[3] For country-specific information about the sociometric gross sample refer to the section „A note on the sociometric gross sample".
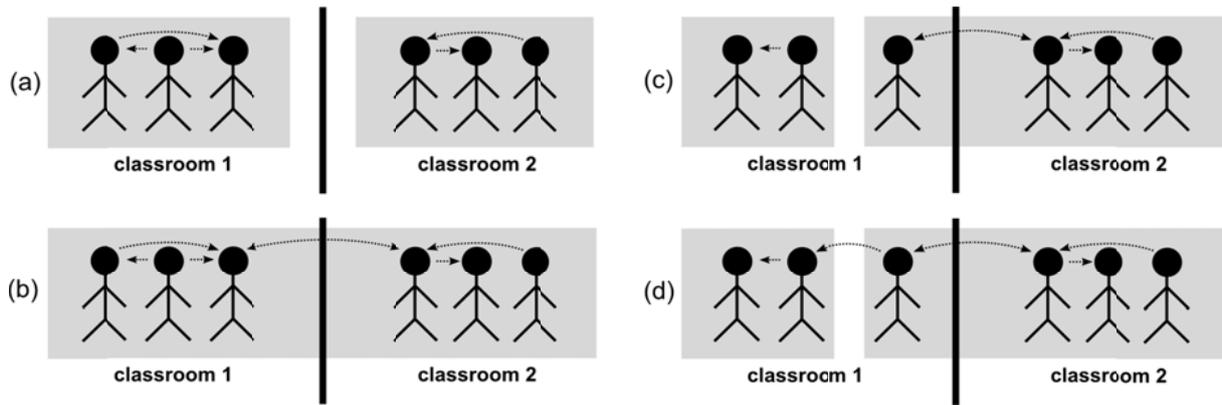
Figure 3: Exemplary Survey Unit Structures

Survey unit bounded to classroom? (classid_match, school level)

The flag variable classid_match indicates whether all survey units in the respective school are class-specific (corresponding to the situation in panel (a)). The variable is coded as follows:

0 = Survey units in school are not all bounded to classrooms
1 = All survey units in school are bounded to classrooms (classid=classid_yc)

In the Netherlands and Sweden, all survey units are class-specific, in England there are 12 schools where survey units deviate from the classroom context, and in Germany there is one school.

Different nomination list (outnoms, individual level)

For a few respondents in two English survey units nomination lists differed from those of the others in their survey unit. Besides the students within their own survey unit they could also nominate students participating in other survey units. The flag variable outnoms indicates all survey units where this is the case. It is coded as follows:

0 = Nomination list bounded to survey unit
1 = Nomination list exceeds survey unit

The reason for this deviation is that a number of those respondents who moved to another survey unit different from their own classroom context were able to nominate within their survey unit as well as their own classroom members. The resulting survey unit structure is displayed in panel (d) of figure 3.

Nomination of absent students (absnoms, individual level)

As mentioned before, in the Netherlands and Sweden, all students in the sampled classes could be nominated even if they were absent during the in-school surveys or opted-out to participate. For England and Germany, the procedure differed between schools or even survey units within schools, implying that in some cases (ENG: N(survey units)=43, GER: N(survey units)=96) absent students were excluded from nomination lists.

We assume that respondents were allowed to nominate absent students if in any of the sociometric questions an absent student was actually nominated. The flag variable absnoms contains this information. It is coded as follows:

0 = absent classmates cannot be nominated
1 = absent classmates can be nominated

Table 1: Country differences in the survey unit structure

	England	Germany	The Netherlands	Sweden
Survey unit ≠ classroom context (classid_match=0)	Yes, in 12 schools	Yes, in 1 school	No	No
Different nomination lists in survey unit (outnoms=1)	Yes, in 2 survey units	No	No	No
Nomination of absent students possible (absnoms=1)	Not in 43 survey units	Not in 96 survey units	Yes	Yes

Nomination behavior

Share of participants (p_participants, class level)

The validity of classroom network data highly depends on the share of network members who actually report their nominations. As participation rates differ substantially between classes we included a variable to identify the share of classroom members participating in the sociometric module, p_participants.

The share is calculated as the number of classroom members who participated in the sociometric survey divided by the total number of classroom members in the gross sample.

Share of invalid nominations (p_invalidnoms, class level)

In order to get an impression of the relative frequency of invalid nominations in a classroom we included the variable p_invalidnoms. It informs about the relative share of invalid nominations out of all nominations made in a classroom that had to be recoded to "other missing" (-55).

Please note, that invalid nominations do not necessarily suggest that there have been mistakes in the surveying process. Rather, these may also be the result of self-nominations or double nominations.

Respondent never nominates (nnominator, individual level)

Some respondents did not make any nominations throughout any of the 11 sociometric questions, and also did never explicitly report to have no ties at all. They are referred to as never-nominators, with the corresponding flag variable nnominator identifying these respondents. It is coded as follows:

0 = at least one nomination made
1 = no nomination made in any sociometric question

Respondent is never nominated (nnominee, individual level)

Furthermore, some students were never nominated by their classmates in any of the questions (never-nominees). The flag variable nnominee can be used in order to identify these cases. It is coded as follows:

0 =   at least nominated once
1 =   never nominated by a classmate in any sociometric question

A note on the sociometric gross sample

In general, the four country-specific sociometric gross samples consist of all participants in the sociometric survey as well as those students who either refused to participate or who were absent at the date of the survey. As mentioned before, every person in the sociometric gross sample is being assigned an individual socio-ID.

In most countries sociometric gross samples were known. In England, however, we constructed the gross sample based on all nominations that respondents made. Nominated persons were assumed to be part of the gross sample, if one of the following conditions applied:

- The nominated person's socio-ID corresponds with an existing youthid.
- The nominated person's socio-ID is within the range of other existing socio-IDs: For example, let us assume there are four students in a classroom who participated in the sociometric module with their socio IDs being "01", "02", "04", "05". If one of the participants nominated "03", this would be assumed to be a valid nomination from the gross sample, as it lies within the range of existing socio-IDs.
- The nominated person's socio-ID equals the highest existing socio-ID number +1: Returning to the example, if any classmate nominated "06", this would be assumed to be a valid nomination from the gross sample.
- The person's socio-ID is nominated more than once: Referring again to the previous example, if respondents "01" and "04" both nominated "08", the latter would count as a valid nomination from the gross sample.

All socio-IDs in the English nomination variables that did not meet any of these conditions were assumed to be not corresponding with a valid youthid in the gross sample. In accordance with the general procedure (see section "Missing values in the nomination variables") they were therefore recoded to "other missing" (-55).

We are left with two types of valid nominations in all four countries; those that correspond to an existing youthid in the realized sample and those that correspond to a person from the gross sample whose interview could not be realized. About the latter type of nomination we do not have any further background information. Nevertheless, the nominations themselves carry some information, for example when analyzing respondents' outdegree (= nomination activity).

Regardless of the type of analysis carried out, we suggest to always keep these differences between gross samples in mind.

Accounting for measurement differences in analyses

The previous sections laid out that there can be differences in the survey unit structure and that nomination behavior might differ between schools and even between classes within a school. We strongly suggest to account for these differences when analyzing the sociometric data.

One way to do so would be by selecting a subsample of networks that is suitable for one's analytical purposes and, at the same time, complies with certain data measurement rules. Please

note that depending on the type of analysis to be carried out different restrictions might be appropriate, sometimes combining several of the measures introduced before.

Examples for classroom selections

In the following, we give three short examples for classroom selections, the first explicitly accounting for differences in the survey unit structures and the second and third for differences in nomination behavior. For each example we present sample sizes for different scenarios. Our starting point for each example is the complete sample of wave 1. It looks as follows:

	ENG N(classes)	GER N(classes)	NET N(classes)	SWE N(classes)	TOTAL N(classes)
N(classes)	208	271	222	251	952

Example 1: Accounting for differences in survey unit structures

As laid out before, throughout the sample the sociometric survey units have been defined in slightly different ways. To control for potential nomination biases due to these differences we could therefore restrict our sample to those survey units whose definitions are more similar; (1) those that are strictly bounded to the classroom context, (2) those in which all nomination lists are identical, (3) those in which absent students can be nominated:

	ENG N(classes)	GER N(classes)	NET N(classes)	SWE N(classes)	TOTAL N(classes)
(1)	182	267	220	249	918
(1)∧(2)	181	267	220	249	917
(1)∧(2)∧(3)	141	173	220	249	783

Example 2: Share of participants

Normally, we want the sociometric data to reflect most of the classroom interactions. We should therefore exclude those classrooms in which many of the students did not participate in the sociometric survey. If we excluded classrooms with participation shares lower than 60/75/90% this would result in:

	ENG N(classes)	GER N(classes)	NET N(classes)	SWE N(classes)	TOTAL N(classes)
>60% partic.	203	243	220	250	916
>75% partic.	192	201	211	235	839
>90% partic.	154	97	158	172	581

Restricting the sample this way leaves us with classes in which students could both nominate the vast majority of their classmates and be nominated by the vast majority of classmates. Whether or not absent students could also be nominated (indicated by absnoms) would therefore not make a big difference.

Example 3: Invalid nominations

In some classrooms we observe high numbers of invalid nominations. Following a conservative analysis approach, we could exclude these classrooms. Excluding all classrooms in which the proportion of invalid nominations out of all nominations exceeded 30/20/10% results in the following case numbers:

	ENG	GER	NET	SWE	TOTAL
	N(classes)	N(classes)	N(classes)	N(classes)	N(classes)
<30% inv. noms.	205	270	220	244	939
<20% inv. noms.	203	270	218	243	934
<10% inv. noms.	197	270	216	235	918