

# Sources of Segregation in Social Networks: A Novel Approach Using Facebook

Bas Hofstra\*    Rense Corten†    Fran Van Tubergen‡  
Nicole B. Ellison§

2017

## 1 Data and Measurements

We use the second wave of survey data on adolescents in the Netherlands, which is part of a larger project titled “Children of Immigrants Longitudinal Survey in Four European Countries” (CILS4EU) (Kalter et al. 2013, 2015). Although data were collected in the Netherlands, Sweden, Germany, and England, the measures we are interested in are included only in the Dutch data. In the CILS4EU, adolescents age 14 to 15 years were followed for three years, starting in 2010, with a one-year time lag. The survey included data on many individual characteristics, attitudes, and information about the individuals with whom respondents associated with in their leisure time. The survey also contained sociometric data on friendships within classrooms ( $\sim 22$  pupils in a classroom). The sample was stratified by the proportion of non-Western immigrants within a school. Within these strata, schools were chosen with a probability proportional to their size (based on the number of pupils at the relevant educational level).

In wave 1 (2010 to 2011), two classes were randomly selected within the schools, which resulted in 118 schools, 252 classes, and 4.963 Dutch pupils participating in the survey. Because changes in class compositions between grades are common in the Netherlands, respondents were distributed among different classes in wave 2 (2011 to 2012) that were not part of the original sampling frame. To ensure that many wave 1 pupils also participated in wave 2, schools were asked to include more than the two classes initially sampled in wave 1 when respondents from wave 1 were in classes other than the previously sampled classes. Consequently, 2.118 new pupils were interviewed, and 3.803 of wave 1 respondents were surveyed again in wave 2 (76,6 percent; total  $N = 5.921$ ).

---

\*Utrecht University

†Utrecht University

‡Utrecht University and King Abdulaziz University

§University of Michigan

We used the second wave of the CILS4EU because it is the latest licensed data including sociometric classroom information.

## 1.1 The Dutch Facebook Survey

The Dutch Facebook Survey (DFS) enriched the Dutch part of the CILS4EU survey (Hofstra, Corten, and Van Tubergen 2015). Data were collected between June and September 2014. Of the 4.864 respondents who indicated Facebook membership in wave 3 (2012 to 2013;  $N = 3.423$ ) or 4 (2013 to 2014;  $N = 3.595$ ) of the CILS4EU, 4.463 were tracked on Facebook. For respondents who kept a public friend list, we downloaded their complete Facebook friend lists ( $N = 3.252$ ; 72,8 percent). There is selectivity in the downloaded friend lists: some respondents kept their lists private, others kept public friend lists. Girls, ethnic minority members, and unpopular adolescents are somewhat underrepresented, because they more often keep private friend lists (Hofstra, Corten, and Van Tubergen 2016b). Various Heckman selection-model specifications (Heckman 1979) show that our results are insensitive to these selection biases. The 3.252 respondents have a combined total of 1.158.227 friends, and 2.810 (86,4 percent) of the respondents whose complete friend list we downloaded also participated in wave 2 of the CILS4EU. This is the number of respondents for whom we present results. Table 1 summarizes the data sources and our method of arriving at the final number of respondents.

Table 1: Overview of the Relevant Data Sources and Selections

	<i>N</i>	%
Survey data (CILS4EU)		
Wave 2 total number of respondents	5.921	100
Wave 2 respondents participated in wave 1	3.803	64,2
Wave 2 respondents who are newcomers	2.118	35,8
Online network data (DFS)		
Respondents indicated being on Facebook in waves 3 or 4 of the survey	4.864	100
Respondents whose profiles were tracked on Facebook	4.463	91,8
Respondent kept a public Facebook friend list	3.252	66,9
Conditions for inclusion in the final number of cases to analyze		
Participation wave 2 + Tracked on Facebook + Kept a public Facebook friend list	2.810*	

\*Various Heckman selection model specifications show that our results are insensitive to selection biases

## 1.2 Measuring Ethnic and Gender Homogeneity in Online Networks

There is no direct measure for friends ethnic background and gender in the Facebook network. We predicted friends gender and ethnic background based on their first names, using the Dutch Civil Registration data (hereafter, DCR) for the entire Dutch population in 2010 ( $N = 15.785.208$ ; Bloothoof and Schraagen 2011). We obtained (1) the fraction of the name carriers and (2) the fraction of the name carriers fathers and (3) mothers who were born in the Netherlands,

Turkey, Morocco, the Dutch Caribbean, other Western countries, or other non-Western countries. Additionally, we obtained the percentage of women among the name carriers.

We matched first names in the DCR to first names in the second wave of the CILS4EU survey as a training dataset. In the CILS4EU, we measured respondents ethnic background by classifying them into one of the six largest ethnic groups in the Netherlands (Castles, De Haas, and Miller 2013): Dutch majority, Turkish, Moroccan, Dutch Caribbean, other Western (European or English speaking), and other non-Western. Moroccan and Turkish adolescents are children of immigrants from the low-educated labor force that was recruited by the Netherlands in the 1950s and 1960s. Dutch Caribbean adolescents originate from post-colonial countries in the Dutch Caribbean (e.g., Aruba and Suriname). Western and non-Western adolescents originate from neighboring countries such as Germany or conflict areas such as Afghanistan; these immigrant groups are relatively similar across Western European countries (Smith, Maas, and Van Tubergen 2014).

We classified respondents according to their biological parents country of birth, which is standard practice in research on Dutch ethnic minority groups (cf. Smith, Maas, and Van Tubergen 2014; Stark and Flache 2012; Vermeij et al. 2009). When students have one parent who was born in the Netherlands, the student is classified into the ethnic background of the parent who was not born in the Netherlands; if a students parents were born in different non-Dutch countries, the student is classified according to the mothers birth country. This definition is regularly applied and used by Statistics Netherlands (Statistics Netherlands 2012).

Combining the DCR and the CILS4EU, we developed an algorithm to estimate gender and ethnic segregation based on peoples first names, which yields high correlations between the predicted and actual ethnicity and gender (this method is outlined in Part A of the online supplement). We calculated the percentage of women and the percentage of each of the six ethnicities in respondents online networks. For each respondent, we assigned the percentage of same-gender friendships (i.e., the percentage of women for girls and percentage of men for boys) in their online networks.<sup>11</sup> Finally, we assigned each respondent the percentage of co-ethnic ties in their online networks (e.g., the percentage of Dutch majority members among online network friends for the Dutch majority adolescents).

### 1.3 Homogeneity in Core Friendship Networks

Wave 2 of the CILS4EU has two measures that capture ethnic homogeneity and one measure that captures gender homogeneity in core friendship networks: a name generator for the five best friends *in general* (only for ethnicity), and a name generator for the five best friends *in class* (not necessarily the same friends as the former).

First, we measured the actual number of friends of Dutch, Turkish, Moroccan, Dutch Caribbean, or another ethnic background using a name-generator

question. Respondents could nominate their best friends (with a maximum of five) and provide ethnic background information. From these data, we calculated the percentage of co-ethnic friends among all the close friends (co-ethnicFRIENDS IN GENERAL). We consider *ethnically similar* friends among best friends in general. Respondents may be more accurate in reporting ethnicities of ethnically similar than ethnically dissimilar friends. Furthermore, respondents were asked to report the ethnicities of their *best friends*. Respondents may more accurately report the ethnicities of their best friends than those of acquaintances. Therefore, respondents misreporting of alter characteristics is likely reduced to a minimum.

Second, we measured the number of best friends in a class (with a maximum of five) who were girls (which is the only core-network measure available for measuring gender homogeneity) and those who were of Dutch, Turkish, Moroccan, Dutch Caribbean, other Western, or other non-Western ethnic backgrounds. We calculated the percentage of co-ethnic friends and the percentage of same-gender friends among all friends in a class (co-ethnicFRIENDS IN CLASS and same-genderFRIENDS IN CLASS). Because these friends themselves were respondents in the survey, they self-reported their gender and ethnicity. We constructed gender and ethnic homogeneity with respect to best friends in a class with these self-reports, and hence they do not suffer from respondents misperceptions in alter characteristics.

#### 1.4 Homogeneity in Meeting Opportunities and Number of Online Network Friends

We constructed various measures to capture ethnic and gender homogeneity in two adolescent opportunity structures, the class and the school. First, using the CILS4EU, we measured the number of classmates who are female and those with the six ethnic backgrounds mentioned above, excluding best friends who are mentioned in the class and respondents themselves. We calculated the percentage of same-gender and co-ethnic classmates, and we excluded the respondent and the number of best friends who are mentioned. We excluded best friends because they are included in the core-network measure, and we do not want to double-count best friends across variables. With this approach, we are better able to separate the effects between variables (same-genderIN CLASS and co-ethnicIN CLASS).

Second, we measured the number of female pupils in a school (aggregated from the classes surveyed) and the number of pupils in the school from a Dutch, Turkish, Moroccan, Dutch Caribbean, other Western, or other non-Western ethnic background (measured from secondary data obtained from the Dutch inspectorate), excluding best friends who are mentioned, other classmates, and the respondent. We calculated the percentage of same-gender schoolmates (excluding the respondent, the number of best friends who are mentioned, and the number of classmates) ( same-genderIN SCHOOL). We also measured the percentage of co-ethnic schoolmates (excluding the respondent, the number of best friends, and the number of classmates) (co-ethnicIN SCHOOL). We measured

these two variables using the CILS4EU.

We also calculated the number of online network friends from respondents Facebook friend lists using the DFS. The distribution of the number of online network friends, if it is plotted, strongly resembles the distribution plot reported by DiPrete and colleagues (2011:1254) of the number of acquaintances reported by Americans.

### 1.5 Kinship Ties in Online Networks as a Confounding Factor

An issue with online versus offline friendship networks is that we restricted respondents to name *friends* in their self-reported core networks offline, whereas Facebook networks likely include *kin*. Therefore, when we contrast core networks with online networks, we compare two data sources of different sampling frames. Kinship ties in online networks might pull ethnic and gender homogeneity in different directions. Kin likely have a similar ethnicity as the respondent, whereas the gender distribution in families is likely to be 50/50. On the one hand, the presence of kin in online networks overestimates ethnic homogeneity; on the other hand, the presence of kin among Facebook friends might lead us to underestimate gender homogeneity (see McPherson et al. 2001:431).

We identify kinship ties in online networks in two ways, using the DFS. First, Facebook allows members to show kinship tags on their profiles. We tracked the number of kinship tags on Facebook profiles and calculated the percentage of kinship tags in the Facebook network (mean = 1.1 percent). We considered realistic tags (e.g., no granddaughters, given that we study adolescents). Individuals might not tag each family member on Facebook. Therefore, we calculated the percentage of friends in the Facebook network who share a surname with the respondent (mean = 1.7 percent). Non-kin friends may have a similar surname, which makes our analyses more conservative, because individuals with similar surnames are likely of the same ethnicity.

Nevertheless, we may miss kin in online networks who are not tagged and who have different surnames. We mention where we remove kin from the online networks to avoid sampling mismatches (descriptive comparisons between core networks and the larger online networks) and where we control for these two variables (statistical tests of the hypotheses).

Table 2 shows the descriptive statistics for ethnic and gender homogeneity in the large online networks (including kin), core networks, opportunity structures, and kinship ties in online networks, along with the distributions of boys and girls and ethnic groups in the data.

Table 2: Descriptive Statistics of Ethnic and Gender Homogeneity in Large Online Networks, Opportunity Structures, Kinship Ties on Facebook, and the Distribution of Boys and Girls and Ethnic Background

	Min	Max	Mean	SD	N
Online networks*					
Co-ethnicFACEBOOK	0	100	76,577	32,099	2.810
Same-genderFACEBOOK	0	100	56,087	9,745	2.809
% Female	0	100	49,453	11,475	2.810
% Dutch	0	100	86,200	15,670	2.810
% Turkish	0	100	2,304	7,649	2.810
% Moroccan	0	59,460	1,729	5,015	2.810
% Dutch Caribbean	0	54,237	1,347	3,097	2.810
% Other Western	0	57,142	3,176	2,566	2.810
% Other non-Western	0	75,676	4,234	5,025	2.810
Core networks					
Co-ethnicfriends in general	0	100	76,218	33,730	2.810
Co-ethnicfriends in class	0	100	67,525	38,249	2.677
Same-genderfriends in class	0	100	83,175	30,227	2.677
Opportunity structures					
Co-ethnicin class	0	100	65,710	31,743	2.690
Co-ethnicin school	0	100	67,038	30,926	2.763
Same-genderin class	0	100	50,212	22,163	2.690
Same-genderin school	0	100	47,428	18,328	2.638
Number of online network friends	1	1.067	336,853	177,702	2.810
Kinship ties on Facebook					
% Kinship ties declared	0	20	1,081	1,555	2.794
% Similar surname on Facebook	0	100	1,689	3,303	2.794
Girl	0	1	,515		2.809
Ethnic background					2,810
Dutch	0	1	,804		2.258
Turk	0	1	,020		57
Moroccan	0	1	,015		42
Dutch Caribbean	0	1	,023		65
Other Western	0	1	,088		247
Other non-Western	0	1	,050		141

\*These estimates of homogeneity in Facebook networks include kin.

## 1.6 Additional Confounding Factors

We adjust for the year in which respondents joined Facebook using the DFS (median = 2010). Respondents who were members for shorter periods may have been more selective in their online network friendships. Facebook membership duration and the number of Facebook friends are positively correlated ( $r = .250$ ;  $p < .001$ ).

Using the CILS4EU, we also control for educational track in high school,

because such a track may be related to ethnic prejudice (Lancee and Sarrasin 2015). When adolescents transition to high school in the Netherlands, they are placed into different tracks, which differ in their level and type of education. We measured this categorization using three dummy variables: preparatory vocational education ( $N = 1,358$ ; Dutch: VMBO), senior general ( $N = 750$ ; Dutch: HAVO), and university preparatory education ( $N = 586$ ; Dutch: VWO). We also control for respondents social attractiveness, which may be correlated with ethnicity (Wimmer and Lewis 2010). We measured social attractiveness by popularity (i.e., incoming popularity nominations from other classmates) (mean = 9,357; SD = 14,566). We calculated popularity by dividing the total number of classmates received nominations for popularity by the total number of students in the class minus one multiplied by 100.

We adjust for ethnic out-group attitudes because they may be related to ethnic homogeneity in online networks. With the survey question, “Please rate how you feel towards the following groups...” respondents used a scale ranging from 0 (negative) to 100 (positive), with 10-point intervals, to rate how positively they feel toward groups of Dutch, Turkish, Moroccan, and Dutch Caribbean ethnic backgrounds. We constructed ethnic outgroup attitudes by taking the mean positivity score on a scale from 0 to 10 of respondents answers to this question while excluding the respondents own ethnic group (mean = 5,011; SD = 1,997). This variable is significantly negatively related to the percentage of co-ethnic friends online ( $r = -.213$ ;  $p < .001$ ).

We accounted for respondents attitudes toward gender roles when we considered gender homogeneity in online networks. We captured respondents progressiveness toward gender roles by counting (from zero to four) how many times respondents indicated that both men and women (instead of men or women) should take care of children, cook, earn money, and clean (Davis and Greenstein 2009) ( $\alpha = .73$ ; mean = 2,689; SD = 1,352). This variable is significantly negatively related to the percentage of same-gender friends online ( $r = -.072$ ;  $p < .001$ ).