



# School friendship networks, homophily and multiculturalism: evidence from European countries

Nicola Campigotto<sup>1</sup> · Chiara Rapallini<sup>2,3</sup> · Aldo Rustichini<sup>4</sup>

Received: 13 December 2019 / Accepted: 8 December 2020 / Published online: 10 February 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

This paper investigates the determinants of school friendship networks among adolescents, proposing a model of network formation and estimating it using a sample of about 10,000 secondary school students in four countries: England, Germany, the Netherlands and Sweden. We test the idea that networks arise according to homophily along many characteristics (gender, school achievement and ethnic and cultural backgrounds), and assess the relative importance of each factor. In addition to gender, we find that country of origin, generational status and religion predict friendship for foreign-born students. For individuals born in the survey country, ties depend on a broader set of factors, including socioeconomic status and school achievement. In sum, homophilic preferences go considerably beyond ethnicity. Multiculturalism, which gives prominence to ethnic backgrounds, risks emphasising the differences in that dimension at the expense of affinity in others.

**Keywords** Friendship · Homophily · Immigration · Networks · Social cohesion

**JEL Classification** D85 · J15 · Z13

---

Earlier versions of this paper were presented at the 2nd Florence-Konstanz workshop on Behavioural and Experimental Social Sciences (University of Florence) and the 2nd International CILS4EU User Conference (Utrecht University). We thank participants at these meetings, Paolo Brunori, Editor Klaus F. Zimmermann and two anonymous referees for their comments and suggestions. Aldo Rustichini thanks the National Science Foundation for the grant SES 1728056, “Rules based policies, and Intelligence in Strategic Behavior”

*Responsible editor:* Klaus F. Zimmermann

✉ Chiara Rapallini  
chiara.rapallini@unifi.it

Extended author information available on the last page of the article.

## 1 Introduction

In recent years, European countries have witnessed a substantial increase in immigration. In 2017, foreign-born residents accounted for approximately 11% of the European population, with 38.2 million people born outside the Union living in EU Member States as of January 2018, and 20.4 million born in a Member State different from the one they reside in (Eurostat 2019). The largest volumes of non-national residents are recorded in Germany (9.7 million persons), the UK (6.3 million), Spain (5.3 million), France (3.7 million) and Italy (3.4 million); in proportional terms, the foreign-born presence is most pronounced in Austria (19.2% of the population), Sweden (18.5%), Ireland (16.8%) and Germany (16.6%). Changes have also occurred in the composition of schools, with more than one in ten 15-year-old students now a first- or second-generation immigrant (OECD 2017).

This increase in diversity has raised new social issues, including differences in educational achievement (Rangvid 2007; Böhlmark 2008; Dustmann et al. 2012; Frattini and Meschi 2019), changes in host nationals' preferences for redistribution (Senik et al. 2009; Alesina et al. 2019) and, at a more general level, a decrease in social cohesion and solidarity (Alesina and La Ferrara 2000; Putnam 2007; Algan et al. 2016).

Potential explanations for the effects of diversity on social relationships typically involve homophily (van der Meer and Tolsma 2014), namely the tendency of individuals to interact and associate with others who are similar to themselves (Lazarsfeld and Merton 1954; Rogers and Bhowmik 1970). Homophilic preferences have been documented in a variety of contexts and with respect to many individual characteristics (see McPherson et al. 2001 and the references therein), ethnicity being the most frequently examined (Currarini et al. 2009, 2010, 2016). Along this line of inquiry, the present paper addresses three questions:

1. Which homophily dimensions matter for school friendship network formation?
2. How do preferences for peer association change with generational status?
3. Do socialization patterns differ by age at migration?

Question 1, as simple as it may seem, has serious implications. The role that individual traits play in network formation is often considered *in isolation*, i.e. ignoring other characteristics. This point, made by Block and Grund (2014) and Rapallini and Rustichini (2016), is central to examining peer preferences: studies conducted under the lens of one-dimensional homophily fail to consider the concurrent influence of other individual attributes, with the risk of misinterpreting the role of the characteristic being examined.

A corollary to this question is whether school friendship networks are mainly shaped by ethnic backgrounds. From a policy-oriented perspective, this means testing the underlying implicit hypothesis of the multiculturalism approach that ethnicity is the key determinant of social bonds. Multiculturalism entered the public discourse as an approach aimed at incorporating, rather than assimilating, immigrants into host societies, and strongly encourages ethnic community associations. Questions 2 and 3 look further into this matter, asking whether the role of co-ethnicity and other characteristics varies with time-since-migration.

The paper presents a simple model of friendship formation that accounts for multiple dimensions of homophily and group composition, so as to take into consideration the simultaneous influences of individual preferences and mixing opportunities. We then estimate the model using a two-wave survey (CILS4EU) of about 10,000 secondary school students in four European countries: England, Germany, the Netherlands and Sweden. We assess the relative strength of different dimensions of similarity by simultaneously considering a wide array of individual characteristics, including gender, country of origin, socioeconomic status, religion, normative beliefs and school achievement. We compare the results for first-generation immigrants with those for natives, and examine the extent to which age at migration affects the likelihood of a friendship tie.

Studying homophilic preferences can pose the empirical challenge of distinguishing between peer selection and influence. Put simply, similarity among people can result from either or both of two mechanisms: while on the one hand ‘similarity breeds fellowship’ (McPherson et al. 2001, p. 428), on the other hand individuals may end up engaging in certain behaviors or developing certain individual traits as a consequence of socialization. This issue is related to the well-known Manski reflection problem (Manski 1993), which makes it difficult to disentangle homophilic pairing from the influence of an individual’s peers and the correlated effects caused by exposure to a common environment. To address this endogeneity concern, we adopt an empirical strategy which draws on the approach proposed by Bramoullé et al. (2009) to estimate peer effects, using the characteristics of individuals’ indirect friends to identify instruments.

This study aims to contribute to the literature on social cohesion along several lines. First, it investigates friendship formation using a two-wave panel of schools in European countries that are among the more ethnically diverse. The focus is on adolescents as they are going through a developmental phase that is crucial in shaping their identity. Second, it takes a perspective which allows us to move beyond ethnicity-centered views and assess the relative importance of multiple dimensions of homophily. Third, it tries to make progress in separating the effects of selection and influence.

The remainder of the paper is organized as follows. Section 2 provides some background on multiculturalism and the empirical literature on link formation. Section 3 introduces the model and estimation strategy. Section 4 presents the data and descriptive statistics. Section 5 discusses the main results. Section 6 gives some concluding remarks.

## 2 Multiculturalism and homophily dimensions

The main thesis of multiculturalism as a prescriptive idea (Kymlicka 1995, 2007) is that members of minority groups should be allowed, and even encouraged, to maintain their culture, religion, language and political and moral traditions. Multiculturalism is not only a theoretical stance; since the 1990s, it has entered the public vocabulary in reference to policies adopted by central states and local authorities to manage multi-ethnic societies (Rattansi 2011). According to Malik (2015)

‘multicultural policies seek to institutionalize diversity’ (p. 21) by defining people’s needs and rights according to their ethnic background. At the school level, which is the focus of the present study, this has resulted in ethnicity-centered educational practices and demands for students’ right to their own language (Zilliagus et al. 2017). For example, in 2011, a high school in Pennsylvania started separating students by race, language and gender for some minutes a day, claiming that this helped them develop strong bonds and improve their grades and self-esteem.<sup>1</sup>

As stated above, this paper tests the underlying hypothesis of the multicultural approach, i.e. that ethnicity is crucially involved in the formation of social ties. In doing this, it is important to consider homophily from a multidimensional standpoint. Previous studies in this direction include Goodreau et al. (2009), who use the Add Health survey of American adolescents and show that ethnic groups differ in terms of mixing. For example, blacks are generally the most cohesive category, although when whites are the minority group, they display stronger ethnic homophily than do blacks when they are the minority. The authors also find evidence of strong homophilic preferences over gender, and show that females form triangular links more often than males do. Wimmer and Lewis (2010) use information from the Facebook profiles of about 1600 students enrolled in an American private college, finding that co-residence and non-racial characteristics (such as the socioeconomic background and the US state of origin) are more important than ethnicity in determining friendship choices. Block and Grund (2014) use data from a survey of three American and British schools to assess the role of gender, ethnicity and socioeconomic status, showing that they are all significant predictors of friendship and that sharing more than one attribute has a diminishing effect on the probability of a tie. Smith et al. (2014) focus on directed (non-reciprocated) relationships, using data from the first wave of the CILS4EU survey to ask whether ethnic homophily is a by-product of similarity in socioeconomic conditions, leisure activities and cultural traits. Within this framework, they do not find evidence that ethnic homophily stands for homophily in other characteristics. Rapallini and Rustichini (2016, 2019) investigate the role of gender, ethnicity, personality traits and cognitive abilities in predicting relationships among classmates, using an original survey of children from eight Italian primary schools. The first paper uses cross-sectional data from 7-year-old students and finds that similarity in personality and cognitive skills matters as much as similarity in ethnic background, but less than gender; the second paper substantiates these results by incorporating an additional wave of data, collected at age 11, and estimating classroom peer effects to address the aforementioned reflection problem.

These results add to an earlier body of literature which documents that cleavages among peer groups tend to increase from childhood to adolescence (Hartup 1983; Epstein 1986; DuBois and Hirsch 1990). In the words of Hallinan and Teixeira (1987, p. 556), as children grow older, ‘friendships become more exclusive, the importance of similarity as a basis for friendship becomes more pronounced [...] and similar ascribed and achieved characteristics are assigned even more weight in the selection of friends.’

---

<sup>1</sup><http://edition.cnn.com/2011/US/01/27/pennsyation>

Regarding the relationship between immigrants' time of arrival and the structure of their network, cross-cultural adaptation theory predicts that the ratio of same-ethnic ties to all ties decreases with the time spent in the host country (Kim 2001). Comola and Mendola (2015) investigate co-ethnic networks using acquaintance information from a survey of Sri Lankans living in Milan. The authors find that newly arrived immigrants tend to know one another but are also likely to be acquainted with immigrants who arrived several decades earlier; the overall relationship between the probability of a tie and the difference in time of arrival is U-shaped, a result consistent with the hypothesis that newcomers contact more experienced members of their community to obtain useful information.

### 3 Modeling friendship formation

As in Rapallini and Rustichini (2019), we model how friendship links among individuals in a group are formed depending on their characteristics. Members of the group (e.g. students in a class) are indexed by  $i = 1, \dots, n$ . Each individual is described by a vector  $\theta^i \equiv (\theta_k^i)_{k=1, \dots, K}$  of characteristics; for example, the first coordinate may describe whether individual  $i$  is female or male, the second may describe whether she is foreign born or not and so on. The utility that an individual derives from having a link depends on her own characteristics and on the characteristics of the match. Assuming a simple additive form, the value for  $i$  of having a friendship link with  $j$  is:

$$w^{ij} = \sum_{k=1}^K \left( \lambda_k \mathbb{1}_{\theta_k^i = \theta_k^j} + \mu_k \mathbb{1}_{\theta_k^i \neq \theta_k^j} \right),$$

where  $\lambda$  and  $\mu$  are real-valued parameters and  $\mathbb{1}$  denotes the indicator function. Letting  $\sigma_{ij}^k \equiv \mathbb{1}_{\theta_k^i = \theta_k^j}$  and  $\gamma_k \equiv \lambda_k - \mu_k$ , the expression can be rewritten as:

$$w^{ij} = \sum_{k=1}^K \mu_k + \sum_{k=1}^K \gamma_k \sigma_{ij}^k, \tag{1}$$

and assuming preferences to be homophilic, then for every characteristic  $k$  we have that:

$$\lambda_k \geq \mu_k \geq 0,$$

implying  $\gamma_k \geq 0$ . Let  $C$  denote the set of individuals who are nominated by agent  $i$ . The utility of  $i$  is:

$$U^i(C) = \sum_{j \in C} w^{ij} - \sharp(C)^\alpha, \tag{2}$$

where  $\sharp(C)$  is the cardinality of set  $C$  and  $\alpha > 1$ . The convex cost associated with a bigger  $C$  represents the cost (network management, transportation and so on) of having a larger group of friends.

### 3.1 Matching

Suppose that individuals know the distribution of characteristics in the group. Each agent then sends a link to a set of peers to maximize the utility defined in Eq. 2. This choice is made simultaneously by all individuals, and the links are sent to all agents that one would like to have as a friend. If the vector of values is ordered in a decreasing order of the index, then the optimal size  $c_i^*$  of the group of links for agent  $i$  is the solution of the maximization problem:

$$c_i^* \equiv \max\{c : w^{ic} \geq (c + 1)^\alpha - (c)^\alpha\}, \tag{3}$$

meaning that agent  $c_i^*$  is the last for whom the marginal benefit from adding  $c_i^*$  to the list of friends,  $w^{ic^*}$ , is larger than the incremental cost of adding another friend to the list,  $(c^* + 1)^\alpha - (c^*)^\alpha$ . The optimal  $c^*$  exists and is unique: this follows from the definition in Eq. 3, the fact that the marginal cost only depends on the size  $c$ , and the ordering convention for the values.

### 3.2 Estimation

Let  $L_{ij}$  be a dichotomous variable equal to 1 when  $i$  sends a link to  $j$ , and 0 otherwise. Agent  $i$  chooses to send a link to  $j$  if the value she places on establishing the link,  $w^{ij}$ , is larger than a threshold  $w^{i,th}$  which is group and agent dependent. A simple statistics of the distribution of values in the group to agent  $i$  is the average value:

$$\begin{aligned} \bar{w}^i &\equiv \frac{1}{n-1} \sum_{l \neq i} w^{il} \\ &= \sum_k \mu_k + \sum_k \gamma_k \left( \frac{1}{n-1} \sum_{l \neq i} \sigma_{il}^k \right) \\ &= \sum_k \mu_k + \sum_k \gamma_k \bar{\sigma}_i^k, \end{aligned} \tag{4}$$

where  $\bar{\sigma}_i^k \equiv \frac{1}{n-1} \sum_{l \neq i} \sigma_{il}^k$  indicates how many agents in the group have characteristic  $k$  equal to  $i$ . The value  $\bar{\sigma}_i^k$  depends on both subject and group, and can be easily computed from the data. In view of Eq. 3 determining the optimal size of the group of friends, the threshold  $w^{i,th}$  is increasing if the vector of values increases point-wise. The threshold  $w^{i,th}$  is assumed to increase in the average value  $\bar{w}^i$  and, for simplicity, to satisfy:

$$w^{i,th} = \eta \bar{w}^i, \tag{5}$$

where  $\eta > 0$ . Thus, agent  $i$  sends a link to  $j$  if and only if  $w^{ij} > w^{i,th}$ .

Finally, the parameters  $(\alpha, (\beta\gamma_k)_{k=1,\dots,K})$  can be determined by specifying a logit model for the conditional probability of sending a link:

$$Pr\{L_{ij} = 1 \mid \theta^1, \dots, \theta^n\} = \Lambda\left[\alpha + \beta\left(w^{ij} - w^{i,th}\right)\right] \\ = \Lambda\left[\alpha + \beta(1-\eta)\sum_k \mu_k + \beta\sum_k \gamma_k\left(\sigma_{ij}^k - \eta\bar{\sigma}_i^k\right)\right], \quad (6)$$

with  $\Lambda(\cdot) = \exp(\cdot) / [1 + \exp(\cdot)]$ ,  $\beta\gamma_k$  the coefficient for the  $j$ -specific similarity  $\sigma_{ij}^k$ , and  $\beta\gamma_k\eta$  for the average similarity, that is the fraction of individuals in the group similar to  $i$  in characteristic  $k$ .

### 3.3 Empirical strategy

The model of friendship formation yields a simple equation amenable to estimation:

$$Pr\{Y_{ij,l,t} = 1 \mid \cdot\} = \Lambda\left(\beta_0 + \mathbf{X}'_{ij,l,t}\beta_1 + \bar{\mathbf{X}}'_{ij,l,t}\beta_2\right), \quad (7)$$

where the subscripts  $i$  and  $j, l$  and  $t$  denote individuals, class and time, respectively. The dependent variable,  $Y_{ij,l,t} = Y_{ji,l,t}$ , is a dichotomous variable that takes the value 1 if  $i$  indicates  $j$  and  $j$  indicates  $i$  as a friend at time  $t$ , and 0 otherwise. This is an undirected (or bilateral) link, which is what we consider a friendship relationship to be. On the contrary, a directed link obtains whenever  $i$  indicates  $j$  as a friend but  $j$  does not reciprocate, or vice versa. The equation for the directed link in Eq. 6 shows that the probability of a link from  $i$  (Ego) to  $j$  (Alter) is increasing in those characteristics which Ego and Alter share and for which both have a homophilic preference. If we assume that a bilateral link is formed if and only if the two directed links exist, the same conclusion will hold for the undirected case. The vector  $\mathbf{X}_{ij,l,t}$  captures similarities and differences in the characteristics of members of dyad  $ij$  (same gender, same country of origin and so on), whereas  $\bar{\mathbf{X}}'_{ij,l,t} \equiv [\bar{\mathbf{X}}'_{i,l,t}, \bar{\mathbf{X}}'_{j,l,t}]$  represents the shares of students in class  $l$  similar to  $i$  and  $j$  in each characteristic at time  $t$  —  $\bar{\mathbf{X}}_{i,l,t}$  and  $\bar{\mathbf{X}}_{j,l,t}$ , respectively. This allows us to control for how common a given characteristic is in a class, that is for differences in mixing opportunities. The number of observations used for estimation is  $\sum_{l=1}^L (\kappa_l(\kappa_l - 1))/2$ , where  $L$  and  $\kappa_l$  are, respectively, the number of classes and the number of children in class  $l$  who filled in the relevant sections of the questionnaire.

A well-known feature of friendship relationships, and of social networks in general, is cliquishness. For each triad  $ijm$ , the probability of a link between  $i$  and  $j$  is not independent of the presence of links between  $i$  and  $m$  and between  $j$  and  $m$ ; everything else being equal, the likelihood of a link is higher if both individuals belong to the same clique than if they do not. To capture this non-stochastic component of peer selection, all regressions control for the number of friends that  $i$  and  $j$  share.

The study of peer choices can also pose a simultaneity problem. To try and separate the role of selection from those of peer influence and unobserved group-level heterogeneity, we refer to the approach proposed by Bramoullé et al. (2009) for the estimation of peer effects (see also Calvó-Armengol et al. 2009; Lin 2010; de Melo

2014; Patacchini et al. 2017). The main idea is that each individual  $i$ 's reference group consists of her best friends. Then, whenever the structure of the social network is known and correlated effects affect all members of the group, the characteristics of  $i$ 's friends-of-friends, who themselves are not friends with  $i$ , can serve to instrument the behavior of  $i$ 's friends. The crucial assumption is that relevant unobservable characteristics are group-specific; as observed by Patacchini et al. (2017), this is most reasonable when networks are small—which is true in our case. Correlated effects at the group level, caused by exposure to a common environment, may be related to the influence of good or bad teachers, the availability and quality of school facilities and participation in education or vocational programs. Their presence is accounted for through inclusion of class fixed effects.

We draw on this method to fit our needs. Let  $\mathcal{Z}$  denote the set of students in a class, and for each dyad  $ij$  let  $\mathcal{Z}_{ind}^{i;j}$  and  $\mathcal{Z}_{ind}^{j;i}$  be the sets of  $i$ 's indirect friends through  $j$  and  $j$ 's indirect friends through  $i$ , respectively:

$$\mathcal{Z}_{ind}^{i;j} = \{m \in \mathcal{Z} \setminus \{i, j\} \mid Y_{im} = 0 \text{ and } Y_{jm} = 1\},$$

$$\mathcal{Z}_{ind}^{j;i} = \{m' \in \mathcal{Z} \setminus \{i, j\} \mid Y_{im'} = 1 \text{ and } Y_{jm'} = 0\}.$$

(Class and time subscripts are omitted for simplicity.) To make the ideas concrete, consider the friendship network shown in Fig. 1. In the case of dyad  $(i, j) = (2, 3)$ , we have  $\mathcal{Z}_{ind}^{2;3} = \{8\}$  and  $\mathcal{Z}_{ind}^{3;2} = \{1, 4, 5\}$ . Note that the label 'indirect friends' entails a slight abuse of terminology because the definitions of  $\mathcal{Z}_{ind}^{i;j}$  and  $\mathcal{Z}_{ind}^{j;i}$  do not require  $i$  and  $j$  to be friends. For example, in the case of individuals 1 and 9, we have  $\mathcal{Z}_{ind}^{1;9} = \{10\}$  and  $\mathcal{Z}_{ind}^{9;1} = \{2, 4, 5, 7\}$  but  $Y_{1,9} = 0$ . This choice is made for ease of exposition and with confidence that it will cause no confusion.

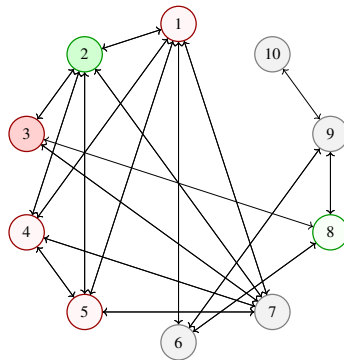


Fig. 1 An undirected friendship network



For each dyad in class  $l$ , instruments are identified by extracting and combining information from Alter’s (Ego’s) friends who are not friends of Ego (Alter). Exploiting the symmetry of bilateral friendship links, it is possible to identify a number of instruments that is twice the number of endogenous variables. Each endogenous regressor  $x_{ij} \in \mathbf{X}_{ij}$  is instrumented by measures of how the characteristics of  $i$  and  $j$  differ from the characteristics of their respective indirect friends. The underlying claim is that, on average, the influence exerted by Ego on Alter when a friendship link exists carries over to Ego’s indirect friends only in a residual, possibly negligible manner. In contrast, when  $Y_{ij} = 0$ , no influence between Ego and Alter is to be expected because neither belongs to the other’s reference group.

Formally, the real-valued variable  $x_{ij} = |x_i - x_j|$  is instrumented by:

$$z_{ij} = \left| x_i - \frac{1}{\#(\mathcal{Z}_{ind}^{i;j})} \sum_{m \in \mathcal{Z}_{ind}^{i;j}} x_m \right| \quad \text{and} \quad z'_{ij} = \left| x_j - \frac{1}{\#(\mathcal{Z}_{ind}^{j;i})} \sum_{m' \in \mathcal{Z}_{ind}^{j;i}} x_{m'} \right|,$$

while the dichotomous variable

$$\tilde{x}_{ij} = \begin{cases} 1 & \text{if } \theta_k^i = \theta_k^j = \tilde{\theta}_k \\ 0 & \text{otherwise} \end{cases}$$

is instrumented by:

$$\tilde{z}_{ij} = \begin{cases} 1 & \text{if } \theta_k^i = \tilde{\theta}_k \text{ and } \sum_{m \in \mathcal{Z}_{ind}^{i;j}} \mathbb{1}_{\theta_k^m = \tilde{\theta}_k} > \frac{\#(\mathcal{Z}_{ind}^{i;j})}{2} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{z}'_{ij} = \begin{cases} 1 & \text{if } \theta_k^j = \tilde{\theta}_k \text{ and } \sum_{m' \in \mathcal{Z}_{ind}^{j;i}} \mathbb{1}_{\theta_k^{m'} = \tilde{\theta}_k} > \frac{\#(\mathcal{Z}_{ind}^{j;i})}{2}, \\ 0 & \text{otherwise} \end{cases}$$

where the last definitions simply require binary instruments to take a value of 1 if and only if both Ego (Alter) and the majority of her indirect friends through Alter (Ego) share characteristic  $\tilde{\theta}_k$ . The endogeneity of regressors and the validity of instruments are assessed in Section 5 through the usual tests.

Paired observations also pose challenges related to the pattern of error correlations. For example, errors for dyad  $ij$  are correlated with errors for any other dyad featuring either  $i$  or  $j$  (Cameron and Miller 2014; Tabord-Meehan 2019). This makes two-way clustering inadequate. A dyadic-robust variance estimator is proposed in Fafchamps and Gubert (2007), but it relies on the assumption that errors for dyads which do not share a member are always uncorrelated. While this assumption may hold, for example, for country-pair data on international trade flows, it seems unlikely in the context of class friendship relationships. Moreover, dyadic clustering is nested in class clustering. The standard approach of clustering at the higher (class) level of aggregation is therefore adopted (Cameron et al. 2011).

## 4 Data and descriptive statistics

Peer preferences are assessed using the first two waves of the CILS4EU (Children of Immigrants Longitudinal Study for Four European Countries; Kalter et al. 2016) survey, which allows us to reconstruct friendship networks in nationally representative samples of secondary schools from England, Germany, the Netherlands and Sweden. This is convenient for the purposes of this study, as these countries are all among the most ethnically heterogeneous in Europe. The first wave of surveys (471 schools, 11,920 respondents) was administered during academic year 2011–2012; the second wave (457 schools, 9,839 respondents) was conducted the following year. The target population for wave 1 consisted of students who, at the time of the survey design, were in the grade where most students are aged 14. For each country, prior to sampling, schools were grouped into mutually exclusive strata depending on the proportion of students with a migration background. Within each stratum, schools were then selected with probabilities proportional to their size, and two classes per school were chosen at random to participate in the survey. The descriptive statistics and demographics of survey participants are summarized in Table 1, while questionnaire items are reported in Appendix A.2.

Information on students' socioeconomic background, academic performance and cultural attributes are used as regressors. Socioeconomic status is proxied by the number of books in an individual's house, the size of the house measured by the number of rooms and parental education (which is a binary variable equal to 1 if at least one of a student's parents has a university degree).<sup>2</sup> Data on these three variables was collected only at wave 1 but can reasonably be assumed not to have changed over 1 year. Academic performance is measured by school grades. The focus is on achievement in mathematics and the survey country's Language (SCL), which reflects analytical and communication abilities. The grades of English and Swedish students are only available for wave 2.

Cultural attributes include religion, normative beliefs about gender roles, attitudes towards cultural assimilation and the bond with the country of residence. These traits are intended to reflect aspects of individual and group identity that are related to culture but are not entirely reducible to ethnicity. Two questions that asked 'in a family, who should do the following?' capture gender-role orientations with respect to earning money and cleaning. We are particularly interested in attitudes indicative of traditionalist views, i.e. women should clean the house and men should be the breadwinners. Instead, normative beliefs about cultural assimilation and diversity are measured by agreement on a 5-point scale with two statements: 'natives should be open to immigrants' customs and traditions' and 'immigrants should do all they can to keep their customs.' Finally, the strength of the bond with the country of residence is scored on a 4-point scale; answers can be viewed as reflecting subjective integration or identification with the survey country.

---

<sup>2</sup>This way of operationalizing socioeconomic conditions takes inspiration from the PISA survey (OECD 2017), which collects information on students' parental education in years, highest parental occupation and number of home possessions (including books). The CILS4EU survey did not ask students the number of years spent by their parents in education but only their highest level of educational attainment.

**Table 1** Descriptive statistics

	Wave	England	Germany	Netherlands	Sweden	Total
<i>N</i>	1	2655	3079	3289	2897	11,920
	%	22.27	25.83	27.59	24.30	100.00
	2	2154	2578	2655	2452	9839
	%	21.89	26.20	26.98	24.92	100.00
Schools	1	107	135	100	129	471
	2	97	134	99	127	457
Classes	1	207	250	222	251	930
	2	187	242	220	247	896
Class size (avg±s.d.)	1	23.54±6.59	21.05±5.27	22.36±5.32	21.83±4.33	22.15±5.47
	2	20.61±6.33	19.20±5.90	18.97±5.28	19.63±4.33	19.55±5.52
% Males	1	48.15	50.95	48.45	48.20	48.97
	2	48.39	49.90	48.27	48.55	48.80
Age (avg±s.d.)	1	15.34±.49	15.79±.74	15.54±.63	15.02±.25	15.43±.63
	2	16.34±.48	16.75±.72	16.50±.61	16.01±.24	16.41±.61
% Natives	1	48.01	45.58	60.20	46.49	50.38
	2	46.73	46.75	61.67	47.51	50.97
% 1st gen. immigrants	1	12.22	8.71	5.94	10.99	9.28
	2	12.42	8.21	5.62	10.36	8.96
% 2nd gen. immigrants	1	25.69	34.50	23.17	29.93	28.30
	2	26.70	33.55	21.71	29.45	27.83
% 3rd gen. immigrants	1	14.08	11.22	10.69	12.58	12.04
	2	14.15	11.48	11.01	12.68	12.24
% Christians <sup>2</sup>	1	38.95	-	-	49.67	43.77 <sup>1</sup>
	2	39.93	-	-	49.76	44.27 <sup>1</sup>
% Catholics	1	-	31.02	14.96	-	-
	2	-	31.77	14.76	-	-
% Protestants	1	-	31.67	9.79	-	-
	2	-	30.80	10.21	-	-
% Muslims	1	12.09	20.69	14.26	15.91	15.84
	2	13.32	19.12	12.84	14.40	14.98
% Atheists/ agnostics	1	38.79	11.85	53.85	30.86	34.06
	2	38.67	12.99	55.93	32.18	34.98
% Others	1	10.17	4.77	7.14	3.55	6.33
	2	8.08	5.32	6.26	3.66	5.76

**Table 1** (continued)

	Wave	England	Germany	Netherlands	Sweden
Biggest minority		Pakistani (~7%)	Turkish (~15.5%)	Turkish (~6%)	Finnish (~6.5%)
		Indian (~7%)	Russian (~6%)	Moroccan (~5.5%)	ex-Yugoslav (~6%)
		Irish (~3.5%)	Polish (~5.5%)	Surinamese (~5%)	Iraqi (~4%)

<sup>1</sup> Christian denominations not assessed in England and Sweden

<sup>2</sup> Includes Catholics and Protestants

The answers are summarized in Table 2. Normative beliefs of third-generation immigrants roughly match those of natives, and the figures for first-generation immigrants are similar to those for second-generation immigrants. These similarities should not be understood as necessarily depending on the time spent by non-native households in a country; the data presented here reflect a particular social context, and might well have been different under different social and economic conditions (e.g. in the presence of substantially higher migration flows and ethnic diversity). Table 3 reports intervariable correlations, which are significant but not so strong as to raise concerns of collinearity.

#### 4.1 Directed and undirected links

From the perspective of network formation, the most relevant feature of the dataset is the availability of information on within-class friendship relationships. Students were asked to nominate their five closest friends in the class and to specify who their very best friend was. In both cases, an undirected friendship link exists whenever a nomination by Ego is reciprocated by Alter.<sup>3</sup> Information on directed and undirected links is reported in Table 4. The difference in the average number of links by gender is nonsignificant at wave 1 ( $p$  value = 0.614) but significant at wave 2 ( $p$  value = 0.008), while the number of links of country-born students significantly exceeds that of first-generation immigrants at both waves ( $p$  values < 0.001).

About 88% of friendship nominations of both males and females were towards students of the same gender as themselves. Calculations do not consider mixing opportunities (e.g. single-sex or highly unbalanced classes), but nevertheless suggest strong homophilic preferences over gender. Similar though less pronounced patterns

<sup>3</sup>The ceiling on the number of nominations is a possible limitation of the data since it may prevent identification of an individual's entire reference group. It also means that students' weak ties (i.e. ties that reflect less frequent, often instrumental relationships) are not considered. About 94% and 97% of students were linked to less than five friends at wave 1 and wave 2, respectively; 62% and 69% indicated fewer than five names.

**Table 2** Normative beliefs and subjective integration

	Natives	1st gen. immigrants	2nd gen. immigrants	3rd gen. immigrants
<i>In a family... who should earn money? (%)</i>				
Mostly the man	31.76	40.45	38.93	30.02
Mostly the woman	1.01	0.98	1.33	0.81
Both about the same	67.23	58.57	59.74	69.18
<i>... who should clean the house? (%)</i>				
Mostly the woman	34.62	37.81	39.82	29.92
Mostly the man	0.74	1.69	0.98	0.77
Both about the same	64.64	60.49	59.20	69.32
<i>[Survey country members] should be open to immigrants' customs and traditions (%)</i>				
Strongly agree	13.46	31.54	33.68	17.07
Agree	37.21	40.95	40.01	40.52
Neither agree nor disagree	31.38	21.72	19.40	28.85
Disagree	11.68	3.41	4.53	8.54
Strongly disagree	6.26	2.38	2.38	5.02
<i>Immigrants should do all they can to keep their customs and traditions (%)</i>				
Strongly agree	5.62	24.48	24.95	6.66
Agree	21.78	35.12	33.09	24.96
Neither agree nor disagree	45.11	31.46	31.55	46.05
Disagree	18.65	6.30	7.48	15.25
Strongly disagree	8.84	2.63	2.93	7.08
<i>How strongly do you feel [survey country member]? (%)</i>				
Very strongly	71.40	14.23	24.49	55.70
Fairly strongly	24.70	44.82	48.18	35.60
Not very strongly	3.05	25.75	18.66	7.12
Not at all strongly	0.85	15.20	8.67	1.57

can be observed with respect to generational status and religion. More than 75% of nominations made by natives and third-generation immigrants were directed to other natives and third-generation immigrants, whereas first- and second-generation immigrants nominated other first- and second-generation immigrants 61% of the time. Same-faith nominations of Christians, Muslims and atheists amount to 60%, 54% and 56%, respectively. Similarities can also be observed from the perspectives of cultural traits and school achievement; as shown in Table 5, the percentages of dyads that share potentially endogenous characteristics are always higher for friends than for non-friends. Tests of proportions show that, even when small, the differences are statistically significant at the 95% level or higher.

**Table 3** Inter-variable correlations

	Male	Native	1st gen. immigrant	Catholic	Protestant	Muslim	Atheist/ agnostic	Parent with univ. degree
Subjective integration: weak	.01*	-.30*	.23*	-.02*	-.06*	.27*	-.15*	-.02*
Men should be the breadwinners	.21*	-.06*	.04*	.03*	.04*	.14*	-.07*	-.11*
Women should clean the house	.16*	-.02*	.01	.04*	.03*	.10*	-.03*	-.13*
Natives should be open to immigrants' customs	-.02*	-.19*	.08*	-.04*	-.05*	.22*	-.13*	.08*
Immigrants should do all to keep their customs	.04*	-.22*	.11*	-.04*	-.07*	.31*	-.15*	.01
Grade: mathematics	.03*	.05*	-.02*	.05*	.05*	-.06*	.00	.06*
Grade: SC language	-.17*	.04*	-.06*	.05*	.06*	-.06*	-.02*	.05*

\* Significant at 5 percent level or better

**Table 4** Friendship nominations and links

		Wave 1 (avg±s.d.)	Wave 2 (avg±s.d.)
Ego-to-alter nominations	overall	3.35±1.59	2.43±1.77
	Males	3.49±1.58	2.51±1.83
	Natives	3.43±1.55	2.52±1.74
	Country-borns <sup>1</sup>	3.38±1.58	2.45±1.76
	1st gen. immigrants	3.10±1.68	2.25±1.83
Undirected links	overall	2.23±1.46	1.63±1.43
	Males	2.22±1.48	1.60±1.45
	Natives	2.30±1.45	1.73±1.43
	Country-borns <sup>1</sup>	2.26±1.45	1.66±1.43
	1st gen. immigrants	1.95±1.47	1.43±1.41
Reciprocity index	.65±.14	.67±.15	

<sup>1</sup> Students born in the survey country, either with or without a migration background

## 5 Results

This section examines the determinants of friendship ties. Research question 1 (which homophily dimensions affect school friendship choices?) is addressed in Sections 5.1 and 5.2; research questions 2 and 3 (do peer selection patterns change with generational status and time-since-migration?) are dealt with in the second part of Section 5.1. Section 5.3 provides a consistency check.

**Table 5** Percent of dyads with similar characteristics

	$Y_{ij} = 0$ (not friends)	$Y_{ij} = 1$ (friends)
Subjective integration: strong	28.78	33.03
Subjective integration: weak	3.13	4.34
Men should be the breadwinners	13.04	15.18
Women should clean the house	14.22	16.44
Natives: open to immigrants' customs	5.38	6.04
Immigrants: keep own customs	2.40	3.14
Math grade higher than median	12.94	14.21
Math grade lower than median	12.48	14.01
SCL <sup>1</sup> grade higher than median	10.29	12.06
SCL <sup>1</sup> grade lower than median	10.19	11.03

<sup>1</sup> Survey country language

## 5.1 Gender, country of origin, religion and socioeconomic background

Table 6 shows maximum likelihood estimates for the whole sample and for selected subsamples. In column 1, the probability of forming a link is regressed against the country of origin alone, while in column 2 the covariates related to generational status, socioeconomic background, religion and age are included, as well as the full set of controls for confounding factors (described below). Aside from the improvement in goodness-of-fit, several points can be made. First, the coefficient on students' country of origin remains highly significant but drops by more than 50%, indicating that the single-variable model of column 1 picks up the effect of all variables correlated with country of origin. Second, gender proves to be the strongest predictor of friendship—a well-established result in the literature on adolescent relationships (Karweit and Hansell 1983; Epstein 1986), which confirms the observation made in the previous section. Third, on average, common generational status matters for first- and second-generation immigrants but not for third-generation immigrants. The interaction term between generational status and country of origin is also nonsignificant. Fourth, there is evidence of homophily with respect to socioeconomic status, but the coefficients are considerably smaller than those on country of origin and religion.<sup>4</sup> Broadly similar findings are obtained when very best friendship ties are used as dependent variable (column 6).

These results carry over when restricting the analysis to dyads with one or more natives (column 3) and dyads with one or more country-borns with a migration background (column 4), but differ for dyads with at least one foreign-born (column 5). First-generation immigrants show no evidence of homophilic mixing with respect to socioeconomic characteristics, whereas other generational status groups do. On the other hand, the coefficients on country of origin and religion are highly significant in all specifications.

Estimates are obtained controlling for mixing opportunities, cliquishness, unobserved heterogeneity at the class level and time-specific effects. A control for geographical proximity is also included, which checks whether or not Ego and Alter live within a 5-min walking distance. The size of the coefficients is in line with other studies on bilateral network formation (e.g. Comola and Fafchamps 2013, 2015). The results indicate a substantial degree of segregation among groups and suggest that first-generation immigrants bond mostly on the basis of country of origin, generational status and religion, while country-borns exhibit homophilic preferences over a broader set of dimensions.

To further investigate immigrants' peer choices, we follow Comola and Mendola (2015) and examine whether differences in the probability of a link are accounted for by differences in the time spent in the country of residence (third research question). Table 7 shows estimates for dyads with at least one first-generation immigrant. The absolute difference in years of residence and its square are added to the model in

<sup>4</sup>Additional estimates by gender are show in Appendix A.1.1. The role of socioeconomic factors is stronger for males than for females, and the coefficient on second-generation immigration status is significant for males only.



**Table 6** Benchmark estimates

Logit estimates · Marginal effects, clustered SEs in parentheses

	(1)	(2)	(3)	(4)	(5)	(6)
	All dyads	All dyads	Natives	2nd/3rd gen. immigrants	1st gen. immigrants	All dyads
Dependent variable	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (very best friends)
Same country of origin	.054*** (.004)	.026*** (.003)	.032*** (.005)	.026*** (.004)	.033*** (.009)	.010*** (.002)
Same gender		.130*** (.003)	.127*** (.003)	.130*** (.004)	.145*** (.008)	.054*** (.002)
Both 1st gen. immigrants		.019** (.009)			.027*** (.010)	.014*** (.004)
Both 2nd gen. immigrants		.013*** (.004)		.011** (.004)		.003 (.003)
Both 3rd gen. immigrants		-.002 (.008)		-.003 (.008)		.003 (.005)
Same country of origin × both 1st gen. immigrants		.023 (.016)			.019 (.019)	-.008 (.009)
Same religion		.018*** (.002)	.012*** (.002)	.029*** (.003)	.029*** (.007)	.008*** (.001)
Both have a parent with a university degree		.005** (.002)	.007** (.003)	.007** (.003)	.009 (.007)	.003* (.002)

**Table 6** (continued)

Logit estimates · Marginal effects, clustered SEs in parentheses					
(1)	(2)	(3)	(4)	(5)	(6)
All dyads	All dyads	Natives	2nd/3rd gen. immigrants	1st gen. immigrants	All dyads
Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (5 best friends)	Undirected link (very best friends)
Difference: number of books	-.004*** (.001)	-.004*** (.001)	-.004*** (.001)	-.001 (.003)	.001 (.001)
Difference: number of rooms	-.001** (.000)	-.001 (.001)	-.001** (.000)	-.001 (.001)	-.001** (.000)
Difference: age	-.005*** (.002)	-.004** (.002)	-.006** (.002)	-.002 (.005)	-.002** (.001)
Class and time fixed effects	✓	✓	✓	✓	✓
Shares of students with same characteristics	✓	✓	✓	✓	✓
Number of common friends	✓	✓	✓	✓	✓
Five minutes distance	✓	✓	✓	✓	✓
Socioeconomic characteristics jointly nonsignificant (p-value)				.559	
N (dyads)	108,193	80,837	59,363	14,242	102,026
Pseudo- <i>R</i> <sup>2</sup>	.005	.370	.382	.394	.221

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

**Table 7** Estimates (difference in years of residence, Ego or Alter is a first-generation immigrant)

Logit estimates · Marginal effects, clustered SEs in parentheses

Dependent variable: undirected link (five best friends)

	(1)	(2)	(3)	(4)	(5)
Same country of origin	.036*** (.008)	.036*** (.008)	.037*** (.008)	.038*** (.008)	.037*** (.008)
Same religion	.030*** (.007)	.030*** (.007)	.030*** (.007)	.030*** (.007)	.030*** (.007)
Both 1st gen. immigrants	.023** (.010)	.023** (.010)	.025** (.010)	.049*** (.012)	.025** (.013)
Difference: years of residence	-.003*** (.001)	.002 (.003)			
Difference: years of residence, squared		-.001* (.000)			
Difference in years of residence: 3–5 years			-.002 (.008)		
Difference in years of residence: 6–10 years			-.010 (.009)		
Difference in years of residence: ≥11 years			-.034*** (.013)		
Both 1st gen. immigrants × difference in years of residence: 3–5 years				-.026 (.017)	
Both 1st gen. immigrants × difference in years of residence: 6–10 years				-.017 (.019)	
Both 1st gen. immigrants × difference in years of residence: ≥11 years				-.081 (.057)	
Different gen. status × difference in years of residence: 3–5 years					.005 (.009)
Different gen. status × difference in years of residence: 6–10 years					-.006 (.010)

**Table 7** (continued)

Logit estimates · Marginal effects, clustered SEs in parentheses

Dependent variable: undirected link (five best friends)

	(1)	(2)	(3)	(4)	(5)
Different gen. status × difference in years of residence: $\geq 11$ years					-.026** (.012)
Abs. difference and quadratic term jointly nonsignificant ( $p$ value)		.002			
$N$ (dyads)	13,292	13,292	13,292	13,292	13,292

One, two and three asterisks denote significance at the 10, 5 and 1 percent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

columns 1 and 2, respectively. When the two variables are entered together, they are jointly significant, indicating a nonlinear, decreasing relationship. The specification in column 3 includes a set of dummies corresponding to different time intervals. The omitted baseline category is that of dyads where the difference in years of residence is at most 2; compared to them, dyads with a difference greater than 11 years have a significantly lower probability of having a link, whereas no significance is found for other categories. Simply put, students who migrated between birth and the end of primary school have similar probabilities of forming ties with other students, while later-age immigrants (who migrated after primary school) are considerably less likely to do so.

In columns 4 and 5, the time interval dummies are interacted with generational status. The omitted category in column 4 consists of dyads where both Ego and Alter are first-generation immigrants and the difference in years-since-migration is less than 3. The coefficients on the interaction terms are all nonsignificant, indicating that, on average, first-generation immigrants bond with one another more than they do with country-borns, regardless of whether they migrated at the same age or not. Column 5 focuses on dyads where Ego and Alter are of different generational status; the results are similar to those in column 3.

## 5.2 Cultural traits and school achievement

Next we consider cultural traits.<sup>5</sup> As shown in Table 8, the endogeneity test rejects the null hypothesis of exogeneity for all regressors (a conservative significance level of 20% was chosen); the predictive power of normative beliefs and the bond with the country of residence is therefore estimated using IV-GMM methods (Baum et al. 2003, 2007).

<sup>5</sup>Variations in sample size are due to unanswered survey questions.

**Table 8** Estimates (cultural traits, all dyads)

IV-GMM second stage estimates · Clustered SEs in parentheses						
Dependent variable: undirected link (five best friends)						
	(1)	(2)	(3)	(4)	(5)	(6)
Same country of origin	.015*** (.003)	.015*** (.003)	.016*** (.003)	.016*** (.003)	.015*** (.003)	.015*** (.003)
Same religion	.012*** (.002)	.012*** (.002)	.012*** (.002)	.012*** (.002)	.012*** (.002)	.012*** (.002)
Subjective integration: both strong <sup>1</sup>	.005 (.006)					
Subjective integration: both weak <sup>1</sup>		.027* (.014)				
Both think men should be the breadwinners <sup>1</sup>			.018*** (.007)			
Both think women should clean the house <sup>1</sup>				.014** (.006)		
Both think immigrants should keep their customs <sup>1</sup>					.035*** (.014)	
Both think natives should open to immigrants' customs <sup>1</sup>						.040** (.021)
First stage <i>F</i> -statistic	394.200	149.491	600.986	679.661	256.720	116.258
Sargan-Hansen <i>J</i> test ( <i>p</i> value)	.811	.958	.749	.795	.504	.328
Underidentification test ( <i>p</i> value)	<.001	<.001	<.001	<.001	<.001	<.001
Endogeneity test ( <i>p</i> value)	.050	.119	.036	.131	.014	.043
<i>N</i> (dyads)	72,937	72,937	73,253	73,118	72,304	72,492

One, two and three asterisks denote significance at the 10, 5 and 1 percent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

<sup>1</sup> Instrumented

The first stage *F*-statistics on all specifications exceed the rule-of-thumb level of 10, suggesting that weakness of instruments is not a concern (Staiger and Stock 1997), and the underidentification test (Kleibergen and Paap 2006) always rejects the hypothesis that the excluded instruments are uncorrelated with the endogenous regressors. Importantly, since the model is overidentified, it is possible to test for the validity of instruments using the Sargan-Hansen test of overidentifying restrictions. The null hypothesis that the excluded instruments are orthogonal to the residuals is never rejected, which suggests that the instruments are valid.

Estimated coefficients on endogenous variables are based on first stage estimates (reported in Appendix A.1.2), and standard errors are larger than those of other covariates. This makes it difficult to assess the relative predictive power of instru-

**Table 9** Estimates (school achievement, all dyads)

Logit estimates · Marginal effects, clustered SEs in parentheses					
Dependent variable: undirected link (five best friends)					
	(1)	(2)	(3)	(4)	(5)
Same country of origin	.027*** (.004)	.026*** (.004)	.027*** (.004)	.027*** (.004)	.026*** (.004)
Same religion	.017*** (.002)	.017*** (.002)	.017*** (.002)	.017*** (.003)	.017*** (.003)
Math grade: both higher than class median	.010*** (.003)				.011*** (.003)
Math grade: both lower than class median		.011*** (.003)			.012*** (.003)
SCL grade: both higher than class median			.016*** (.003)		.015*** (.003)
SCL grade: both lower than class median				.004 (.003)	.006* (.003)
Endogeneity test ( <i>p</i> value)	.763	.406	.817	.623	.730
<i>N</i> (dyads)	76,755	76,755	77,349	77,349	75,595

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

EN and SW: wave 2 only

mented regressors. Nevertheless, the coefficients and significances support the idea that culture plays an important role in shaping friendship patterns.

Contrary to what was found for cultural traits, the null hypothesis of the endogeneity test for school grades is never rejected. The coefficients in Table 9 are therefore estimated by maximum likelihood. Note that this does not necessarily rule out the possibility of peer effects in education, which are generally intended as influence-related spillovers from group average outcomes and not just from Alter to Ego. Estimates show that proficiency in mathematics and the survey country's language both have significant predictive power. In contrast, the evidence for grades below the class median is mixed. Some caution is warranted in interpreting coefficients, as they do not necessarily reflect a preference for interacting with peers who are similar in terms of school interests and abilities. Instead, what matters may be affinity in terms of other unobservable determinants of grades, such as effort.

Finally, Table 10 checks whether similar results hold for foreign-borns. Even without investigating the direction of causation, it can be seen that no significant relationship exists between similarity in normative beliefs and the probability of a friendship tie. The only significant coefficients are those on subjective integration and high language and math skills, which is to say that, all other things being equal,

**Table 10** Estimates (cultural traits and school achievement, Ego or Alter is a first-generation immigrant)

	(1)	(2)	(3)	(4)	(5)	(6)	(7) <sup>1</sup>	(8) <sup>1</sup>	(9) <sup>1</sup>
Logit estimates · Marginal effects, clustered SEs in parentheses Dependent variable: undirected link (five best friends)									
Same country of origin	.033*** (.009)	.032*** (.009)	.032*** (.009)	.033*** (.009)	.032*** (.009)	.033*** (.009)	.033*** (.012)	.034*** (.011)	.034*** (.012)
Same religion	.028*** (.006)	.027*** (.006)	.030*** (.007)	.029*** (.006)	.028*** (.007)	.028*** (.007)	.034*** (.008)	.033*** (.009)	.034*** (.010)
Subjective integration: both strong	.021** (.010)								
Subjective integration: both weak		.021** (.008)							
Both think men should be the breadwinners			.001 (.006)						
Both think women should clean the house				.004 (.008)					
Both think immigrants should keep their customs					.001 (.008)				
Both think natives should be open to immigrants' customs						-.001 (.012)			
Math grade: both higher than class median							.026*** (.010)		.023** (.010)
Math grade: both lower than class median							-.010 (.011)		-.008 (.011)

**Table 10** (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7) <sup>1</sup>	(8) <sup>1</sup>	(9) <sup>1</sup>
Logit estimates · Marginal effects, clustered SEs in parentheses Dependent variable: undirected link (five best friends)									
SCL grade: both higher than class median								.031*** (.010)	.028** (.011)
SCL grade: both lower than class median								.013 (.010)	.016 (.010)
<i>N</i> (dyads)	14,028	14,028	14,177	14,080	13,873	13,917	8,617	8,659	8,435

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

<sup>1</sup> EN and SW: wave 2 only



proficient first-generation immigrants have a higher probability of bonding with other proficient students.

### 5.3 A consistency check

The CILS4EU dataset provides information on the students' parents, some of whom were interviewed during the first wave of the survey. Ideally, this could allow us to use parental cultural traits as instruments, based on the assumption that Ego's (Alter's) opinions do not influence the opinions of Alter's (Ego's) parents. However, the data on parents suffer from several limitations. First, as already mentioned, they were collected only at wave 1 and only for approximately 50% of students. Second, and importantly, identification of respondents was not random but voluntary. This leads to self-selection problems and differences between students and parents; for example, the sample is over-representative of mothers (79%) and natives (81%). However, in the case of subjective integration, the distributions of students' and parents' answers are similar and the instruments are found to be relevant. The estimates for wave 1 data, reported in Table 11, confirm the results previously obtained: the coefficient on weak subjective integration remains significant at the 10 per cent level, while the coefficient on strong subjective integration is still nonsignificant. The outcomes of the Sargan-Hansen test indicate that the instruments are uncorrelated with the error terms.

**Table 11** Consistency check (all dyads)

IV-GMM second stage estimates · Clustered SEs in parentheses			
Dependent variable: undirected link (five best friends)			
	(1)	(2)	(3)
Same country of origin	.026*** (.006)	.025*** (.006)	.025*** (.006)
Same religion	.019*** (.004)	.018*** (.004)	.018*** (.004)
Subjective integration: both strong <sup>1</sup>	.014 (.010)		.014 (.010)
Subjective integration: both weak <sup>1</sup>		.035* (.019)	.034* (.020)
First stage <i>F</i> -statistic	182.881	238.199	90.258
Sargan-Hansen J test ( <i>p</i> value)	.158	.525	.306
Underidentification test ( <i>p</i> value)	<.001	<.001	<.001
<i>N</i> (dyads)	34,027	34,027	34,027

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

<sup>1</sup> Instrumented

## 6 Concluding remarks

This paper offers support for the view that peer selection operates across a wide range of individual traits. Using a two-wave survey of schools in four European countries, we examined which characteristics matter for school friendship formation (research question 1). The results confirm that homophilic preferences go considerably beyond ethnicity: gender, socioeconomic characteristics, academic achievement and cultural traits all have distinct roles. A second research question asked whether preferences for peer choices vary with generational status. Foreign-born students were found to bond mostly on the basis of country of origin, generational status and religion, whereas the mixing choices of country-borns depend on a broader set of factors. This finding is in line with Kim's (2001) argument that newly arrived immigrants tend to be drawn to co-ethnics, and then over time they focus more on non-co-ethnic characteristics and increase the proportion of host nationals in their network. Finally, we studied the extent to which immigrants' friendship ties are influenced by the time spent in the host country (research question 3). First-generation immigrants are likely to befriend one another regardless of their time of arrival, while the probability of a link between a foreign- and a country-born decreases nonlinearly as the difference in the length of residence increases. Contrary to Comola and Mendola (2015), no evidence was found of a U-shaped relationship; this may be due to differences in the type of network considered, the age of respondents and the ethnic heterogeneity of the sample.

Overall, the picture that emerges is one of high group fragmentation. Students' country of origin and religious affiliation both have a strong effect on network formation (second only to that of gender), and this effect is persistent over time. Yet as shown in this and earlier studies (Rapallini and Rustichini 2016, 2019), other characteristics, the support of which runs across ethnic lines, are significantly associated with the likelihood of a friendship tie. This body of research involves examining individual behaviors (peer choices) as a premise for policy appraisal, and the results suggest that the emphasis placed on ethnicity by multicultural policies is misguided.

Similarities between individuals of different groups may act as a powerful force in gradually removing the barriers that separate groups. For example, if two students become friends on the basis of common personality traits (such as curiosity, intelligence, openness and extroversion), as well as more general traits such as school and leisure interests and abilities, then ties between them are likely to go across ethnic, religious and heritage boundaries. In contrast, if ties are heritage-based, a mechanism perpetuating the differences is in place. If this is the case, then an emphasis or even a positive attitude bent on preserving such differences (as in the multicultural idea) may have profoundly adverse effects in the long run. In other words, multiculturalism-inspired policies based on co-ethnicity run the risk of further entrenching segregation and cleavages among groups; the same goals of interpersonal and academic development may be achieved through programs based, for example, on individual interests and attitudes.

The study is not without limitations. First, we only had information on students' five closest friends, that is on their strongest ties. Weak ties, the structure of which

we could not examine, can bridge otherwise disconnected cliques and play an important role in social cohesion (Granovetter 1973, 1983). Second, the availability of only two waves of data, separated by a small interval of time, did not allow us to investigate the co-evolution of similarity and fellowship. Third, despite their importance in terms of time spent with others, individuals' school networks are not necessarily representative of their personal networks as a whole. This is especially true of foreign-born students, who may have considerably different cross-ethnic interaction opportunities in their in-school and out-of-school lives. Fourth, in spite of the reassuring outcomes of the tests for instrument validity, the instrumentation method proposed here is not generally applicable. As we have maintained, on average, the influence exerted by Ego on her indirect friends through Alter can be deemed limited. However, it is also possible that Alter may influence both Ego and Ego's indirect friends in a similar manner. To smooth out this latter effect, instruments were constructed by averaging over the characteristics of several indirect friends. This seems to have worked in our case; nevertheless, the possibility that instruments may be affected by the influence affecting endogenous covariates cannot be ruled out a priori.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Appendix**

**A.1 Additional estimates**

**A.1.1 Estimates by gender**

**Table 12** Benchmark estimates by gender

	Logit estimates · Marginal effects, clustered SEs in parentheses			
	Dependent variable: undirected link (five best friends)			
	(1)	(2)	(3)	(4)
	Ego or Alter is a female	Ego and Alter are females	Ego or Alter is a male	Ego and Alter are males
Same country of origin	.020*** (.003)	.041*** (.007)	.021*** (.003)	.044*** (.008)
Same gender	.099*** (.002)		.110*** (.003)	
Both 1st gen. immigrants	.018** (.009)	.044** (.019)	.005 (.010)	.015 (.023)

**Table 12** (continued)

	Logit estimates · Marginal effects, clustered SEs in parentheses			
	Dependent variable: undirected link (five best friends)			
	(1)	(2)	(3)	(4)
	Ego or Alter is a female	Ego and Alter are females	Ego or Alter is a male	Ego and Alter are males
Both 2nd gen. immigrants	.007 (.005)	.014 (.011)	.013*** (.004)	.036*** (.010)
Both 3rd gen. immigrants	-.006 (.008)	-.024 (.017)	.003 (.009)	.008 (.024)
Same country of origin × both 1st gen. immigrants	-.001 (.017)	-.029 (.035)	.054*** (.016)	.108*** (.039)
Same religion	.014*** (.002)	.026*** (.005)	.015*** (.002)	.034*** (.006)
Both have a parent with a university degree	.001 (.003)	.006 (.006)	.005* (.003)	.019*** (.006)
Difference: number of books	-.002** (.001)	-.005* (.002)	-.004*** (.001)	-.008*** (.003)
Difference: number of rooms	-.001 (.001)	-.001 (.001)	-.001** (.000)	-.002** (.001)
Difference: age	-.005** (.002)	-.008** (.004)	-.004** (.002)	-.008* (.004)
Socioeconomic characteristics jointly nonsignificant ( <i>p</i> value)	.197	.194		
<i>N</i> (dyads)	78,773	32,265	74,546	28,042
Pseudo- <i>R</i> <sup>2</sup>	.413	.340	.378	.287

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively  
All specifications include the controls in the benchmark specification (see Table 6 for details)

## A.1.2 First-stage regressions

**Table 13** Estimates (cultural traits, first stage)

	IV-GMM first stage estimates · Clustered SEs in parentheses					
	(1)	(2)	(3)	(4)	(5)	(6)
Same country of origin	.093*** (.007)	.012*** (.003)	-.008** (.004)	-.003 (.004)	.003 (.003)	-.001 (.001)
Same gender	.002 (.002)	-.001 (.001)	.007*** (.002)	.004* (.002)	-.001 (.001)	.001 (.001)
Both 1st gen. immigrants	-.013 (.014)	.072*** (.018)	.008 (.016)	.017 (.017)	.016 (.012)	.012 (.011)

**Table 13** (continued)

IV-GMM first stage estimates · Clustered SEs in parentheses						
	(1)	(2)	(3)	(4)	(5)	(6)
Same religion	.009*** (.003)	.006*** (.002)	.001 (.003)	.005* (.003)	.007*** (.002)	.007*** (.002)
Both have a parent with a university degree	-.014*** (.005)	-.001 (.002)	-.002 (.004)	-.010** (.004)	.012*** (.004)	.001 (.002)
Subjective integration: strong (Ego and Ego's IFs) <sup>1</sup>	.257*** (.011)					
Subjective integration: strong (Alter and Alter's IFs) <sup>2</sup>	.261*** (.010)					
Subjective integration: weak (Ego and Ego's IFs) <sup>1</sup>		.292*** (.021)				
Subjective integration: weak (Alter and Alter's IFs) <sup>2</sup>		.312*** (.021)				
Men should be the breadwinners (Ego and Ego's IFs) <sup>1</sup>			.299*** (.011)			
Men should be the breadwinners (Alter and Alter's IFs) <sup>2</sup>			.330*** (.012)			
Women should clean the house (Ego and Ego's IFs) <sup>1</sup>				.309*** (.010)		
Women should clean the house (Alter and Alter's IFs) <sup>2</sup>				.349*** (.011)		
Immigrants should keep their customs (Ego and Ego's IFs) <sup>1</sup>					.258*** (.014)	
Immigrants should keep their customs (Alter and Alter's IFs) <sup>2</sup>					.228*** (.013)	
Natives should open to immigrants' customs (Ego and Ego's IFs) <sup>1</sup>						.227*** (.020)
Natives should open to immigrants' customs (Alter and Alter's IFs) <sup>2</sup>						.231*** (.019)
<i>N</i> (dyads)	72,937	72,937	73,253	73,118	72,304	72,492

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

<sup>1</sup> Instrument constructed using information from Ego and Alter's friends who are not friends of Ego

<sup>2</sup> Instrument constructed using information from Alter and Ego's friends who are not friends of Alter

**Table 14** Consistency check (cultural traits, first stage)

	IV-GMM first stage estimates · Clustered SEs in parentheses	
	(1)	(2)
Same country of origin	.064*** (.007)	.010*** (.003)
Same gender	.003 (.007)	-.001 (.001)
Both 1st gen. immigrants	.025 (.020)	.036 (.022)
Same religion	.013*** (.004)	.004** (.002)
Both have a parent with a university degree	.001 (.008)	-.001 (.002)
Subjective integration: strong (Ego and Alter's parent)	.231*** (.013)	
Subjective integration: strong (Alter and Ego's parent)	.241*** (.014)	
Subjective integration: weak (Ego and Alter's parent)		.322*** .024
Subjective integration: weak (Alter and Ego's parent)		.355*** (.023)
<i>N</i> (dyads)	34,027	34,027

One, two and three asterisks denote significance at the 10, 5 and 1 per cent levels, respectively

All specifications shown include the other covariates and controls in the benchmark specification (see Table 6 for details)

## A.2 Questionnaire items

The CILS4EU questionnaire is structured in several sections. Our variables were constructed from information assessed in the sections named 'Youth Main' (YM), 'Youth Classmates' (YC) and 'Youth Parents' (YP).

- YM-1 Are you a boy or a girl?  
 YM-2 When were you born?  
 YM-3 In which country were you born?  
 YM-4 How old were you when you moved to <survey country>?  
 YM-19 Which grades did you get in the last school year in the following subjects?  
     (a) Mathematics  
     (b) <Survey country language>
- YM-27 Did your mother complete university? [Yes; No; Don't know]  
 YM-33 Did your father complete university? [Yes; No; Don't know]

- YM-63 In a family, who should do the following?
- (a) Earn money [Mostly the man; Mostly the woman; Both about the same]
  - (b) Clean the house [Mostly the man; Mostly the woman; Both about the same]
- YM-66 How strongly do you feel <survey country member>? [Very strongly; Fairly strongly; Not very strongly; Not at all strongly]
- YM-71 What is your religion? [No religion; Buddhism; Christianity; Christianity: Catholic; Christianity: Protestant; Hinduism; Islam; Judaism; Sikh; Other: specify]
- YM-75 How much do you agree or disagree with each of these statements?
- (a) The <survey country> people should be open to the customs and traditions of immigrants [Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree]
  - (b) Immigrants should do all they can to keep their customs and traditions [Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree]
- YM-95 How many rooms are there in your home (not counting kitchen and bathroom)?
- YM-96 About how many books are there in your home? [0-25; 26-100; 101-200; 201-500; More than 500]
- YC-1 Who are your best friends in class? Here you may write down no more than five names.
- YC-2 Who is your best friend in class? Here you may write down no more than one number.
- YC-9 Which classmates live within a 5 minute walk from your home?
- YP-9 How strongly do you feel <survey country member>? [Very strongly; Fairly strongly; Not very strongly; Not at all strongly]

### A.3 Variables description

#### Dependent variables and controls for confounding factors

- *Undirected link (five best friends)*. Binary: 1 if a friendship nomination from Ego to Alter is reciprocated, as assessed by question YC-1.
- *Undirected link (very best friends)*. Binary: 1 if a friendship nomination from Ego to Alter is reciprocated, as assessed by question YC-2.
- *Shares of students with same characteristics*. Share of Ego's (Alter's) classmates similar to her in each characteristic considered.
- *Five minutes distance*. Binary: 1 if Ego and Alter live within a five-minute walking distance, as assessed by question YC-9.
- *Number of common friends*. Number of friends that Ego and Alter share, as assessed by question YC-1.

## Demographic characteristics

- *Same country of origin; Both first/second/third-generation immigrants.* The information used to construct these variables was retrieved from previous work by Dollmann and Konstanze (2016).
- *Same gender; Same religion; Difference: age; Difference: years of residence.* Variables were constructed from the answers to questions YM-1, YM-2, YM-3, YM-4, and YM-71.

## Socioeconomic characteristics

- *Both have a parent with a university degree.* Binary: 1 if both Ego and Alter have at least one parent who completed tertiary education, as assessed by questions YM-27 and YM-33.
- *Difference: number of books.* Absolute difference in the number of books in Ego's and Alter's houses, as assessed by question YM-95 .
- *Difference: number of rooms.* Absolute difference in the number of rooms in Ego's and Alter's houses, as assessed by question YM-96.

## Normative beliefs and subjective integration

- *Both think men should be the breadwinners.* Binary: 1 if both Ego and Alter answered 'mostly the man' to question YM-63a.
- *Both think women should clean the house.* Binary: 1 if both Ego and Alter answered 'mostly the woman' to question YM-63b.
- *Both think natives should open to immigrants' customs.* Binary: 1 if both Ego and Alter answered 'strongly agree' to question YM-75a.
- *Both think immigrants should keep their customs.* Binary: 1 if both Ego and Alter answered 'strongly agree' to question YM-75b.
- *Subjective integration: both strong.* Binary: 1 if both Ego and Alter answered 'very strongly' to question YM-66.
- *Subjective integration: both weak.* Binary: 1 if both Ego and Alter answered 'not very strongly' or 'not at all strongly' to question YM-66.

## Normative beliefs and subjective integration (instruments)

- *Men should be the breadwinners (Ego and Ego's IFs); Men should be the breadwinners (Alter and Alter's IFs).* Binary: 1 if both Ego (Alter) and the majority of Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'mostly the man' to question YM-63a.
- *Women should clean the house (Ego and Ego's IFs); Women should clean the house (Alter and Alter's IFs).* Binary: 1 if both Ego (Alter) and the majority of



- Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'mostly the woman' to question YM-63b.
- *Natives should open to immigrants' customs (Ego and Ego's IFs); Natives should open to immigrants' customs (Alter and Alter's IFs)*. Binary: 1 if both Ego (Alter) and the majority of Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'strongly agree' to question YM-75a.
  - *Immigrants should keep their customs (Ego and Ego's IFs); Immigrants should keep their customs (Alter and Alter's IFs)*. Binary: 1 if both Ego (Alter) and the majority of Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'strongly agree' to question YM-75b.
  - *Subjective integration: strong (Ego and Ego's IFs); Subjective integration: strong (Alter and Alter's IFs)*. Binary: 1 if both Ego (Alter) and the majority of Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'very strongly' to question YM-66.
  - *Subjective integration: weak (Ego and Ego's IFs); Subjective integration: weak (Alter and Alter's IFs)*. Binary: 1 if both Ego (Alter) and the majority of Alter's friends who are not friends of Ego (Ego's friends who are not friends of Alter) answered 'not very strongly' or 'not at all strongly' to question YM-66.
  - *Subjective integration: strong (Ego and Alter's parent); Subjective integration: strong (Alter and Ego's parent)*. Binary: 1 if both Ego (Alter) and Alter's parent (Ego's parent) answered 'very strongly' to questions YM-66 and YP-9.
  - *Subjective integration: weak (Ego and Alter's parent); Subjective integration: weak (Alter and Ego's parent)*. Binary: 1 if both Ego (Alter) and Alter's parent (Ego's parent) answered 'not very strongly' or 'not at all strongly' to questions YM-66 and YP-9.

## School grades

- *Math grade: both higher than class median*. Binary: 1 if both Ego's and Alter's grades in mathematics (as assessed by question YM-19a) are above the class median value.
- *Math grade: both lower than class median*. Binary: 1 if both Ego's and Alter's grades in mathematics (as assessed by question YM-19a) are below the class median value.
- *SCL grade: both higher than class median*. Binary: 1 if both Ego's and Alter's grades in <survey country language> (as assessed by question YM-19b) are above the class median value.
- *SCL grade: both lower than class median*. Binary: 1 if both Ego's and Alter's grades in <survey country language> (as assessed by question YM-19b) are below the class median value.

## References

Alesina A, La Ferrara E (2000) Participation in heterogeneous communities. *Q J Econ* 115(3):847–904

- Alesina A, Murard E, Rapoport H (2019) Immigration and preferences for redistribution in europe. NBER Working Paper no. 25562
- Algan Y, Hémet C, Laitin DD (2016) The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *J Polit Econ* 124(3):696–733
- Baum CF, Schaffer ME, Stillman S (2003) Instrumental variables and gmm: Estimation and testing. *Stata J* 3(1):1–13
- Baum CF, Schaffer ME, Stillman S (2007) Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata J* 7(4):465–506
- Block P, Grund T (2014) Multidimensional homophily in friendship networks. *Netw Sci* 2(2):189–212
- Böhlmark A (2008) Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Econ* 15(6):1366–1387
- Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. *J Econ* 150(1):41–55
- Calvó-Armengol A, Patacchini E, Zenou Y (2009) Peer effects and social networks in education. *Rev Econ Stud* 76(4):1239–1267
- Cameron AC, Gelbach JB, Miller DL (2011) Robust inference with multiway clustering. *J Bus Econ Stat* 29(2):238–249
- Cameron AC, Miller DL (2014) Robust inference for dyadic data. working paper
- Comola M, Fafchamps M (2013) Testing unilateral and bilateral link formation. *Econ J* 124(579):954–976
- Comola M, Mendola M (2015) Formation of migrant networks. *Scand J Econ* 117(2):592–618
- Currarini S, Jackson MO, Pin P (2009) An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* 77(4):1003–1045
- Currarini S, Jackson MO, Pin P (2010) Identifying the roles of race-based choice and chance in high school friendship network formation. *Proc Natl Acad Sci* 107(11):4857–4861
- Currarini S, Matheson J, Vega-Redondo F (2016) A simple model of homophily in social networks. *Eur Econ Rev* 90:18–39
- de Melo G (2014) Peer effects identified through social networks: Evidence from uruguayan schools. no. 2014/05
- Dollmann J, Konstanze J (2016) Examining the diversity of youth in europe: A classification of generations and ethnic origins using cils4eu data. a wave 2-update. Working Paper, Mannheim University
- DuBois DL, Hirsch BJ (1990) School and neighborhood friendship patterns of blacks and whites in early adolescence. *Child Dev* 61(2):524–536
- Dustmann C, Frattini T, Lanzara G, Algan Y (2012) Educational achievement of second-generation immigrants: an international comparison. *Econ Policy* 27(69):143–185
- Epstein JL (1986) Friendship selection: Developmental and environmental influences. In: Mueller EC, Cooper CR (eds) *Process and Outcome in Peer Relationships*. Academic Press, pp 129–160
- Eurostat (2019) Migration and migrant population statistics. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration\\_and\\_migrant\\_population\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics). Retrieved: 03/2019
- Fafchamps M, Gubert F (2007) The formation of risk sharing networks. *J Dev Econ* 83(2):326–350
- Frattini T, Meschi E (2019) The effect of immigrant peers in vocational schools. *Eur Econ Rev* 113:1–22
- Goodreau SM, Kitts JA, Morris M (2009) Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography* 46(1):103–125
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
- Granovetter MS (1983) The strength of weak ties: A network theory revisited. *Sociol Theory* 1:201–233
- Hallinan MT, Teixeira RA (1987) Students' interracial friendships: Individual characteristics, structural effects, and racial differences. *Am J Educ* 95(4):563–583
- Hartup WW (1983) Peer relations. In: Hetherington EM, Mussen PH (eds) *Handbook of child psychology*, vol. 4: Socialization, Personality, and Social Development. Wiley, pp 129–160
- Kalter F, Heath AF, Hewstone M, Jonsson JO, Kalmijn M, Kogan I, van Tubergen F (2016) Children of immigrants longitudinal survey in four european countries (cils4eu) – reduced version. <https://doi.org/10.4232/cils4eu.5656>. 2.3.0
- Karweit N, Hansel S (1983) Sex differences in adolescent relationships: Friendship and status. In: Epstein JL, Karweit N (eds) *Friends in School: Patterns of Selection and Influence in Secondary Schools*. Academic Press, pp 115–130
- Kim YY (2001) *Becoming intercultural: An integrative theory of communication and cross-cultural adaptation*. Sage Publications

- Kleibergen F, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. *J Econ* 133(1):97–126
- Kymlicka W (1995) *Multicultural citizenship: A liberal theory of minority rights*. Clarendon Press
- Kymlicka W (2007) Multicultural odysseys. *Ethnopolitics* 6(4):585–597
- Lazarsfeld PF, Merton RK (1954) Friendship as a social process: A substantive and methodological analysis. In: Berger M, Abel T, Page CH (eds) *Freedom and Control in Modern Society*. Van Nostrand, pp 18–66
- Lin X (2010) Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *J Labor Econ* 28(4):825–860
- Malik K (2015) The failure of multiculturalism: Community versus society in europe. *Foreign Aff* 94:21
- Manski CF (1993) Identification of endogenous social effects: The reflection problem. *Rev Econ Stud* 60(3):531–542
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 427:415–444
- OECD (2017) Pisa 2015 technical report. <http://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>. Retrieved: 03/2019
- Patacchini E, Rainou E, Zenou Y (2017) Heterogeneous peer effects in education. *J Econ Behav Organ* 134:190–227
- Putnam RD (2007) E pluribus unum: Diversity and community in the twenty-first century. the 2006 johan skytte prize lecture. *Scand Polit Stud* 30(2):137–174
- Rangvid BS (2007) Sources of immigrants-underachievement: Results from pisa-copenhagen. *Educ Econ* 15(3):293–326
- Rapallini C, Rustichini A (2016) Elective affinities matter as much as ethnicity in multi-ethnic schools. *J Econ Behav Organ* 131:243–262
- Rapallini C, Rustichini A (2019) Personality and cognitive skills in network of friends, for multi-ethnic schools. *Res Econ* 73(1):1–14
- Rattansi A (2011) *Multiculturalism: A very short introduction*. Oxford University Press
- Rogers EM, Bhowmik DK (1970) Homophily-heterophily: Relational concepts for communication research. *Publ Opinion Quart* 34(4):523–538
- Senik C, Stichnoth H, van der Straeten K (2009) Immigration and natives-attitudes towards the welfare state: Evidence from the european social survey. *Soc Indic Res* 91:345–370
- Smith S, Maas I, Van Tubergen F (2014) Ethnic ingroup friendships in schools: Testing the by-product hypothesis in england, germany, the netherlands and sweden. *Soc Netw* 39:33–45
- Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–586
- Tabord-Meehan M (2019) Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *J Bus Econ Stat* 37(4):671–680
- van der Meer T, Tolsma J (2014) Ethnic diversity and its effects on social cohesion. *Annu Rev Sociol* 40:459–478
- Wimmer A, Lewis K (2010) Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *Am J Sociol* 116(2):583–642
- Zilliacus H, Holm G, Sahlström F (2017) Taking steps towards institutionalising multicultural education—the national curriculum of finland. *Multicult Educ Rev* 9(4):231–248

## Affiliations

Nicola Campigotto<sup>1</sup>  · Chiara Rapallini<sup>2,3</sup>  · Aldo Rustichini<sup>4</sup>

Nicola Campigotto  
nicola.campigotto@univr.it

Aldo Rustichini  
aldo.rustichini@gmail.com

- <sup>1</sup> Department of Economics, University of Verona, via Cantarane 24, 37129 Verona, Italy
- <sup>2</sup> Department of Economics and Management, University of Florence, via delle Pandette 9, 50127 Firenze, Italy
- <sup>3</sup> GLO Network, Lagos, Nigeria
- <sup>4</sup> Department of Economics, University of Minnesota, Hanson Hall, MN 55455, USA