



CIRANO
Allier savoir et décision

UNDERSTANDING HOMOPHILY: FROM INTERACTIONS TO INTEGRATION

VINCENT BOUCHER
IRIS LAUGIER

CS

2026s-21
CAHIER SCIENTIFIQUE

Les **cahiers de la série scientifique** visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

The purpose of the Working Papers is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Le **CIRANO** est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO – CIRANO Partners

Partenaires Corporatifs - Corporate Partners

Autorité des marchés financiers
Banque de développement du
Canada
Banque du Canada
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
La Caisse
Énergir
Hydro-Québec
Intact Corporation Financière
Mouvement Desjardins
Power Corporation du Canada
Pratt & Whitney Canada
VIA Rail Canada

Partenaires gouvernementaux - Governmental partners

Ministère des Finances du Québec
Ministère de l'Économie, de
l'Innovation et de l'Énergie
Innovation, Sciences et
Développement Économique
Canada
Ville de Montréal

Partenaires universitaires - University Partners

École de technologie supérieure
École nationale d'administration
publique
HEC Montréal
Institut national de la recherche
scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.
CIRANO collaborates with many centres and university research chairs; list available on its website.

© Octobre 2026. Vincent Boucher, Iris Laugier. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©. *Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*

Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires. *The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partners.*

ISSN 2292-0838 (version en ligne)

Understanding homophily: from interactions to integration *

Vincent Boucher[†], Iris Laugier[‡]

Abstract

Homophily in networks can arise either from meeting opportunities or from preferences. We use the number of joint classes students have as a measure of exposure intensity to identify the relative contributions of meeting-based and preference-based racial homophily in teenagers' friendships in US secondary schools. We present robust evidence that friendship increases with exposure, but even more for same-race students. We develop and estimate an intuitive structural model of friendship formation to reproduce and decompose this stylized fact. We find that preference-based homophily explains the majority of overall racial homophily, but that meeting-based one explains more than a third of it. We argue that this still leaves room for policies targeting meetings in order to improve racial integration. We simulate such policies and discuss their effectiveness in increasing meetings, friendships, and reducing segregation at the school level.

Résumé

L'homophilie dans les réseaux peut résulter soit des occasions de rencontre, soit des préférences individuelles. Nous utilisons le nombre de cours suivis en commun par les élèves comme mesure de l'intensité de leur exposition mutuelle afin de déterminer les contributions relatives de l'homophilie raciale fondée sur les rencontres et de celle fondée sur les préférences dans les relations d'amitié entre adolescents des écoles secondaires américaines. Nous présentons des résultats robustes montrant que la probabilité de nouer une amitié augmente avec l'exposition, mais que cette augmentation est encore plus marquée entre

* We thank Luc Bissonnette, Caitlin Brown, Marion Goussé and Habiba Djebbari for their valuable comments. This research uses data from the AHAA study, which was funded by a grant (R01 HD040428-02, Chandra Muller, PI) from the National Institute of Child Health and Human Development, and a grant (REC-0126167, Chandra Muller, PI, and Pedro Reyes, Co-PI) from the National Science Foundation. This research was also supported by grant, 5 R24 HD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Health and Child Development. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies. This research uses data from Add Health, funded by grant P01 HD31921 (Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

[†] Department of Economics, Université Laval, CRREP, CREATE and CIRANO; vincent.boucher@ecn.ulaval.ca

[‡] Department of Economics, Université Laval, CRREP; iris.laugier.1@ulaval.ca

élèves appartenant au même groupe racial. Nous développons et estimons un modèle structurel intuitif de formation des amitiés afin de reproduire et de décomposer ce fait stylisé. Nous constatons que l'homophilie fondée sur les préférences explique la majeure partie de l'homophilie raciale globale, tandis que celle fondée sur les rencontres en explique plus du tiers. Nous soutenons que ce constat laisse néanmoins une marge de manœuvre pour des politiques visant à favoriser les occasions de rencontre afin d'améliorer l'intégration raciale. Nous simulons de telles politiques et examinons leur efficacité pour accroître les occasions de rencontre et les relations d'amitié, ainsi que pour réduire la ségrégation à l'échelle des établissements scolaires.

Mots-clés/Keywords : Homophilie raciale, réseaux d'amitié, interactions sociales, intégration raciale, ségrégation scolaire, formation des amitiés / Racial homophily, friendship networks, social interactions, racial integration, school segregation, friendship formation.

Pour citer ce document / To quote this document

Boucher, V., & Laugier, I. (2026). Understanding homophily: from interactions to integration (2026s-21, Working Papers, CIRANO.) <https://doi.org/10.54932/EWTZ6539>

1 Introduction

Social capital and social networks are strong determinants of social mobility, e.g. [Putnam \(2016\)](#); [Report \(2019\)](#); [Chetty et al. \(2022a,b\)](#). In the context of online social networks, for example, the diversity of friendship links has been shown to be one of the best predictors of economic mobility, e.g. [Chetty et al. \(2022a\)](#). Unfortunately, most friendship networks are not as diverse as they could, a feature known as *homophily*: similar people are more likely to be connected, everything else equal. Homophily is prevalent in most social networks and for a large set of observable characteristics such as ethnicity, education or gender, e.g. [McPherson et al. \(2001\)](#). Understanding the sources of homophily is therefore a crucial step towards a better understanding of social mobility.

In theory, homophily can either originate from meeting opportunities (similar people are more likely to meet, e.g. [Bramoullé et al. \(2012\)](#)) or from preferences (people prefer to have relationships with similar individuals, e.g. [Currarini et al. \(2009\)](#); [Boucher \(2015\)](#); [Mele \(2017\)](#)). While these two sources of homophily potentially result in very different public policies (see below), very little evidence on their relative prevalence exists.¹

In this paper, we focus on racial homophily, an important predictor of social mobility, e.g. [McPherson et al. \(2001\)](#); [Toomet et al. \(2013\)](#). Using data on students’ friendship networks, we exploit a measure of exposure intensity in order to identify the relative contributions of meetings and preferences to realized homophily. We document the robust empirical fact that increased exposure between students has a stronger impact on the probability of friendship for same-race students. We propose a structural model that replicates this empirical finding and that enables us to identify the main source of racial homophily observed in the data.

We find that preference-based homophily accounts for most of the racial homophily measured in the friendship data, but that meeting-based homophily also plays a significant role, accounting for more than a third of realized homophily in the data. We argue that this leaves room for policy makers to implement policies targeting meetings to improve racial integration. Our simulated public policies highlight an important tradeoff between reducing the level of racial homophily, and increasing the overall interactions between students. We also show that, with incomplete meeting opportunities, segregation indices ([Freeman, 1978](#); [Coleman, 1958](#); [Echenique](#)

¹A first set of policies targets meetings, promoting interactions between different-type individuals. In a school context, and even more particularly in the US context, desegregation policies ([Angrist and Lang, 2002](#); [Billings et al., 2014](#); [Mele, 2020](#)) are examples of such policies because they involve moving students across schools in order to decrease racial formal segregation. A second type of policies is based on preferences. Based on the *contact theory* ([Allport, 1954](#)), it is now well documented that exposure to diverse people can decrease biased preferences and discriminatory behaviors, e.g. [Pettigrew and Tropp \(2008\)](#); [Rao \(2019\)](#) ; [Boucher et al. \(2025\)](#).

and Fryer Jr, 2007) should be interpreted with caution.

We use a weighted measure of the number of classes students have in common as a (quasi) exogenous variation in meeting probability, conditional on observables. This information is provided by the Adolescent Health and Academic Achievement Study (AHAA), which we can link to the friendship data available in the National Longitudinal Study of Adolescent to Adult Health (AddHealth), a widely-used dataset in the literature.² The resulting dataset comprises over one million directed pairs of students for which we observe their friendship status, the number of classes they have together (weighted by the frequency and sizes), as well as a series of sociodemographic characteristics.

Using a linear probability model, we first show that an increase in the weighted number of classes students have in common is strongly associated with the higher proportion of same-race and same-gender friendship relations. Of course, some of the effect may be due to students choosing the same classes as their friends, and not to the increased meeting opportunities for new friendship relations. To explore this issue, we proceed to various robustness analyses (including using instrumental variable estimators) and show that our baseline results are remarkably robust.

Next, we build a simple, but flexible, structural model. We assume that the probability of a friendship nomination is equal to the probability that the students have met in some meaningful way (meeting opportunity), and that the friendship was beneficial (preferences). Both probabilities are allowed to flexibly vary as a function of the observed socio-demographic characteristics. To improve identification, we assume that the (weighted) number of classes students have in common does not affect preferences. We show that the econometric model admits valid microfoundation following the approach in Hsieh et al. (2020).

This simple framework is sufficient to replicate the results from the linear probability model. When the probability of a meaningful meeting increases, the impact is larger for pairs of students with a higher probability of beneficial links. If there is preference-based homophily, this implies that an increase in the meeting probability has a stronger impact on the linking probability for similar students.

We estimate the model via a weighted maximum likelihood in order to account for data errors in friendship nominations. (Boucher and Houndetoungan, 2025) We find that the probability of a meaningful meeting increases with (weighted) number of joint classes for same-race students and for students of the same grade level, but decreases for same-gender students. Perhaps less surprisingly, the probability of a beneficial friendship is higher for students of the same race,

²See for example Boucher (2015), Boucher and Mourifié (2017), Currarini et al. (2009), Hsieh et al. (2020), Mele (2017), Mele (2020), Moody (2001).

gender or grade level. We therefore find racial homophily in both meetings and preferences. Using simulations, we decompose the overall marginal effect and find that racial homophily in preferences explains 63.7% of the total. This means that more than a third of the observed racial homophily is due to meeting opportunities.

We argue that meeting-based public policies are attractive since they are easier to implement: one can make it easier for two people to meet, but can hardly force them to become friends conditional on their preferences.³ We specifically study two potential policies affecting meetings as a way to foster inter-racial friendships. In the first policy, we consider a teacher who increases meetings, for example by assigning students to teamwork. This policy leads to an increase in the number of meaningful meetings and in the number of friendship nominations. However, if the teacher does not bias the policy to promote same-race meetings, the policy results in an increase of racial homophily.

For the second policy, we consider a school administrator allocating students to classes. We simulate a counterfactual distribution of the weighted number of joint classes in which same-race students are less likely to have classes together. The policy introduces a sharp trade-off: it succeeds in lowering racial homophily, but it leads to an overall decrease in the number of friendship links (because of homophilic preferences).

This paper contributes to the literature on homophily and segregation patterns in social networks (e.g., [McPherson et al. \(2001\)](#), [Moody \(2001\)](#), [de Martí and Zenou \(2017\)](#), [Boucher \(2015\)](#), [Jackson \(2008\)](#)). More particularly, this paper is directly linked to the literature on economic models of friendship that tries to disentangle the role of meeting opportunities and of preferences in homophily patterns. In their seminal contribution, [Currarini et al. \(2009\)](#) develop a model of friendship formation in which homophily can originate from preference biases or meeting biases. They show that the homophily patterns observed in the data can only be explained if both types of biases are present, but they do not quantify their relative contributions to observed homophily. Using the same dataset, we show that roughly two-thirds of the observed homophily is due to preference biases, and one third to meeting biases. We also explore the impact of policies aimed at reducing racial homophily by affecting students' meetings and highlight an important tradeoff between reducing homophily and increasing the number of friendship relations.

[Mele \(2017\)](#) builds a model of friendship nominations where utility is a function of friends'

³For example, results in [Chetty et al. \(2016\)](#) suggest that complete integration in a new environment takes time. In our context, integration is interpreted as changes in preferences following an increase in heterophilious meetings.

characteristics but also depends on mutual nominations, popularity, and indirect ties (i.e friends of friends are valued). While students’ random meetings can be biased, all meetings happen with strictly positive probability. As a result, realized homophily at the steady state can only be the result of preference biases. [Fantozzi Alvarez et al. \(2022\)](#) adapt the model in [Mele \(2017\)](#) so that it can be estimated on the path towards the steady-state. Using observations from a Brazilian network, at two points in time and assuming that the meeting process has not converged, they estimate homophily in students’ cognitive skills coming from meeting and preferences. They show that removing biases in the meeting process does not lead to lower observed homophily levels.

In this paper, we model the meeting probabilities, using the number of classes in common as a measure of intensity in students’ exposure. Assuming this measure does not affect preferences, we are able to disentangle homophily coming from meetings from that coming from preferences. Contrary to [Fantozzi Alvarez et al. \(2022\)](#), we find that removing biases in the meeting process does reduce observed homophily. However, we show that the implemented policy may result in overall lower levels of interactions between students.

[Mele \(2020\)](#) also present an analysis of homophily based on the model in [Mele \(2017\)](#). Using the same data as in this paper, he simulates changes in school compositions, switching students across originally completely black and white schools. When the size of the new minority in the school is small, segregation (measured by Freeman’s index) decreases. However, as the fraction of the minority group increases, it is able to self-segregate and this results in more segregation. While [Mele \(2020\)](#) considers policy simulations across different schools and assumes away within-school meeting biases, we study within-school policies affecting the heterogeneity of within-school meeting probabilities. We similarly find that preference biases can work against the meeting-based policy, potentially leading to fewer interactions and lower welfare.

At a much more macroscopic level, [Chetty et al. \(2022b\)](#) explore the relative contributions of exposure (potential meetings) and friendship biases (preferences) on socioeconomic homophily. Using aggregated friendship data from Facebook as their measure of friendship, they show that exposure and friendship biases have roughly equal contributions on observed socioeconomic homophily. Our data allow us to exploit heterogeneity in meetings within schools. Focusing on racial homophily, we find that preference biases account for roughly two thirds of the observed racial homophily in the data.

Finally, studying segregation draws from two key strands of literature: first, the seminal work on the economic impacts of segregation on minorities and the outcomes of desegregation policies

(e.g., [Cutler and Glaeser \(1997\)](#), [Billings et al. \(2014\)](#), [Angrist and Lang \(2002\)](#), [Boozer et al. \(1992\)](#), [Card and Krueger \(1992\)](#)); and second, the methodologies for measuring segregation within social networks ([Freeman \(1978\)](#), [Boucher et al. \(2025\)](#), [Currarini et al. \(2009\)](#)). Our paper is closer to the second strand, exploiting the detailed structure of the friendship networks observed in the data.

The remainder of the paper is organized as follows: Section 2 introduces the data and descriptive statistics. We present the reduced form results in Section 3. The model and its estimates are developed in Section 4. Then, we conduct our counterfactual policies in Section 5 and finally, Section 6 concludes with a discussion.

2 Data

We use data from the National Longitudinal Study of Adolescent to Adult Health (AddHealth), which is a panel dataset collected in the United States since 1994 from a nationally representative sample of secondary schools (public and private ones). It is one of the richest available panel surveys on adolescents and includes information on a wide range of sociodemographic characteristics, health conditions and behaviors. We are interested in two particular features of these data.

First, the data contains information on students’ friendship nominations. In particular, students can nominate up to ten friends, five male and five female. We focus on within-school friendship nominations from the first wave of the survey. It was conducted in 144 schools and collected during the 1994-1995 school year. Table 1 displays the descriptive statistics at the student level for a few selected variables. The sample is well balanced across gender. 7% of the students report being Asian or Pacific islander, 17% report being Hispanic, 19% report being black or African American, 61% report being white and 14% of the students report being from another racial background (including American Indians, Native Americans, other and missing information). A third of students’ mothers completed high school and another third went to college.

Friendship data come from students’ nominations. Every student in each school is given a list of all students in their school and they then nominate up to five female and five male friends. On average, students declare a little more than six friends, although the data contain many errors which prevent us from identifying the nominated friends in an important number of cases. ([Boucher and Houndetoungan, 2025](#)) On average only 4.6 friendships are identified,

which means that around a third of friends' nominations are not identified.

Variable	Mean	St. Dev.
Male	0.4982	0.5000
Age	14.9898	1.7196
Race declared		
Asian	0.0701	0.2552
Black	0.1905	0.3927
Hispanic	0.1722	0.3775
White	0.6057	0.4887
Mother's education		
Less than high school	0.0969	0.2958
High school or General Educational Development (GED)	0.2899	0.4537
Business, trade, or vocational school (BTV)	0.0498	0.2175
Went to college	0.0887	0.2844
College graduate	0.1802	0.3843
Beyond college	0.0687	0.2529
Number of friends		
Declared	6.4330	3.6899
Identified	4.6645	3.5285

Notes: The sample consists of the first wave of the in-school AddHealth survey, including 89,941 students. The table reports the mean and standard deviation of selected variables used in the subsequent estimations.

Table 1: Descriptive statistics at the individual level

Second, the data also include the Adolescent Health and Academic Achievement Study (AHAA), an administrative dataset providing information on the number of classes students took together, for a subset of students. These data were retrospectively collected by the Population Research Center, University of Texas Austin, during the third wave of the Add Health survey (2001-2002). Its main goal is to fill the gap in the information about Add Health respondents' academic performance and background. The main objective of this database is to collect information on the grades and the courses taken in secondary school by all wave 3 respondents, which is a subset of the wave 1 respondents for which we have detailed friendship data. The AHAA obtained consent from more than 87% of the wave 3 respondents to collect their transcripts in their former secondary schools. From these transcripts, the AHAA built the variation in students' exposure that we use in this paper: the number of classes in common between pairs of students. The number of classes in common is weighted by the number of students in the classroom and by the frequency of the class.⁴

This administrative data contain information on 74 out of the 144 schools from the first wave

⁴Frequencies are measured in Carnegie Units, CU: a class of one hour weekly during the entire school year equals 1 CU.

of the AddHealth data. Our final sample is built for all the pairs of students for whom we have information in the AHAA as well as in the wave 1 survey questionnaire. We linked the AHAA with the survey data in wave I at the students’ pair level. The final dataset contains 1,083,660 directed pairs between 6061 distinct individuals. In the Supplementary Appendix, we present evidence that the final dataset is representative of the wave 1 survey population.

Variable	Mean	St. Dev.
Same-race	0.5809	0.4934
Same-sex	0.4981	0.5000
Same-grade	0.2860	0.4519
Number of classes in common	0.1559	0.4840
Total number of friends identified	5.1211	3.4200
Total number of friends declared	6.8182	3.4899
Fraction of nominations	0.0058	0.0760

Notes: The sample consists of the first wave of the in-school AddHealth survey merged with the Adolescent Health and Academic Achievement Study (AHAA), which provides the number of joint classes for a subset of pairs of students. The sample includes 1,083,660 directed pairs among 6,061 students. The table reports the mean and standard deviation of selected pair-level variables used in the subsequent estimations.

Table 2: Descriptive statistics at the pair level

Table 2 displays the characteristics of the pairs of individuals in our final dataset. The variable *Same race* is equal to one if student i and j have declared belonging to at least one common racial group, and 0 otherwise. 58% of the pairs are made of same-race students, 49% are same-sex pairs and 29% student pairs are in the same grade level. On average, a student in a pair declares having almost 7 friends, and 5 of them are identified in the school.

3 Stylized facts

We present a preliminary analysis of the determinants of friendship. It is based on a simple linear probability model, which has the advantage of allowing for the inclusion of student-level fixed effects. Specifically, our main specification is the following:

$$a_{ij} = \alpha_1 w_{ij} + \mathbf{x}_{ij} \boldsymbol{\alpha}_2 + w_{ij} \mathbf{x}_{ij} \boldsymbol{\alpha}_3 + \lambda_i + \mu_j + \varepsilon_{ij} \quad (1)$$

where a_{ij} equals 1 if i nominates j as a friend, $\mathbf{x}_{ij}$ is a vector of binary pair characteristics (same-race, same-sex, same-grade), w_{ij} is the weighted number of classes between i and j , λ_i and μ_j are individual fixed effects, and ε_{ij} is an exogenous error term. The fixed effect λ_i captures individual-specific factors that increase the probability of nominating friends. The fixed effect μ_i captures individual-specific factors that increase the probability of receiving friendship nominations. We assume the following:

Assumption 1. For all i, j , we have: $\mathbb{E}\varepsilon_{ij}|w_{ij}, \mathbf{x}_{ij}, \lambda_i, \mu_j = 0$.

Assumption 1 implies that our main variable of interest, w_{ij} is conditionally exogenous. In particular, it assumes that students do not coordinate and do not choose their classes as a function of the friends they already have.

While this is *a priori* a strong assumption, we argue that it is nonetheless a useful one. *First*, recall that w_{ij} is weighted by class frequency and size, two characteristics over which students have little influence. Moreover, elective courses are less frequent and as a result, students' choices have a weaker impact on w_{ij} . *Second*, students' choices (e.g. selecting a specific elective course) operate at a higher dimension: Even if a student i decides to enroll in a class only because their friend j is also enrolling, that choice does not only impact w_{ij} , but also w_{ik} , for all the other students k in the class. We therefore expect that the impact of students' choices has a limited impact on Assumption 1.

In any case, we also provide an extensive analysis of the issue in the Supplementary Appendix. We show that our main messages are robust to alternative specifications. In particular, we consider alternative measures of exposure that are less (and more) prone to manipulation and present instrumental variables regressions.

In what follows, we therefore use w_{ij} as a source of plausibly exogenous variation in students' within-classroom interaction. Table 3 presents the estimation results. The first column presents estimates without any fixed effects, the second column presents estimates with school-level fixed effects and the third column presents estimates with student-level two-way fixed effects as in Equation (1).

<i>Dependent variable: i nominates j as a friend</i>			
	(1)	(2)	(3)
Same race	0.0026 (0.0002)	0.0031 (0.0002)	0.0048 (0.0003)
Same sex	0.0001 (0.0002)	0.0001 (0.0002)	0.0001 (0.0002)
Same grade	0.0081 (0.0003)	0.0077 (0.0003)	0.0082 (0.0003)
Joint classes	0.0038 (0.0013)	0.0053 (0.0013)	0.0032 (0.0013)
Same race $\times$ joint classes	0.0181 (0.0015)	0.0180 (0.0015)	0.0191 (0.0015)
Same sex $\times$ joint classes	0.0090 (0.0013)	0.0088 (0.0013)	0.0093 (0.0012)
School fixed effects	$\emptyset$	$\checkmark$	$\emptyset$
Student fixed effects	$\emptyset$	$\emptyset$	$\checkmark$

Notes: $N = 1,083,660$ for all specifications. Regressions are estimated at the i, j pair level using a linear probability model. The dependent variable is a binary indicator that is equal to one if student i nominates student j as a friend. Column (1) excludes school and student fixed effects. Column (2) includes school fixed effects. Column (3) includes student fixed effects. The constant is estimated but not reported in specification (1). Standard errors are clustered at the student i level in all specifications.

Table 3: Reduced form estimations

The results show the presence of race-based homophily. This is anticipated and consistent with the literature (e.g. Currarini et al. (2009); Boucher (2015); Mele (2020)). Unsurprisingly, the probability that i nominates j increases if they belong to the same grade level and if they have many classes in common.

Perhaps less obviously however, Table 3 shows that race-based (and gender-based) homophily also affects the relationship between the number of classes taken together and the probability of friendship nominations. In particular, increasing the number of classes students have in common

increases *more* the probability of friendship nomination for same-race students.

We show that this finding has important implications for the comparison of segregation indices across different populations, and on the effectiveness of public policies aimed at decreasing race-based segregation in students' friendship networks. In particular, populations (e.g. schools) with higher levels of interaction between students are more likely to result in higher segregation levels, even conditional on similar preferences.

In order to discuss this point more formally, we propose a simple structural model that allow us to explicitly disentangle the role of meetings and preferences in friendship formation. We present the model in the next section and policy simulations in Section 5.

4 Structural model

In this section, we present a structural model that rationalizes the results found in the previous section. We first describe the data generating process, and then discuss the relevant microfoundations.

We assume that friendship creation requires two elements. First, it requires that students have met in some meaningful way. Second, it requires that friendship is valuable. Importantly, we do not observe meetings, and we only observe whether i nominated j as a friend, or not.

Specifically, define the (unobserved) binary random variable m_{ij} such that $m_{ij} = m_{ji} = 1$ if i and j have met, and 0 otherwise. Define also the (unobserved) binary random variable v_{ij} such that $v_{ij} = 1$ if i would *prefer* to create a friendship relation with j , and 0 otherwise. Finally, define the (observed) binary random variable a_{ij} such that $a_{ij} = 1$ if i nominated j as a friend and 0 otherwise. There are four possible observable states:

$$\begin{aligned} P_{ij}(a_{ij} = 1, a_{ji} = 1) &= P_{ij}(v_{ij} = 1, v_{ji} = 1 | m_{ij} = 1) P_{ij}(m_{ij} = 1), \\ P_{ij}(a_{ij} = 1, a_{ji} = 0) &= P_{ij}(v_{ij} = 1, v_{ji} = 0 | m_{ij} = 1) P_{ij}(m_{ij} = 1), \\ P_{ij}(a_{ij} = 0, a_{ji} = 1) &= P_{ij}(v_{ij} = 0, v_{ji} = 1 | m_{ij} = 1) P_{ij}(m_{ij} = 1), \end{aligned}$$

and $P_{ij}(a_{ij} = 0, a_{ji} = 0)$ with complementary probability.

Note that in principle, there is nothing that prevents the probabilities for v_{ij} and v_{ji} to depend on meeting realizations, i.e. $P_{ij}(v_{ij}, v_{ji} | m_{ij} = 1) \neq P_{ij}(v_{ij}, v_{ji} | m_{ij} = 0)$. However, this dependence cannot be identified from data. This is because preferences are inferred from

friendship nominations, and one cannot observe friendship nominations for students who have not met. We will therefore note $P_{ij}(v_{ij}, v_{ji}|m_{ij} = 1) = P_{ij}(v_{ij}, v_{ji})$ to simplify the notation, but v_{ij} should still be interpreted as being conditional on the students having met each other.

Although simple, this framework is sufficient to reproduce the patterns observed in Table 3. For example, let x_{ij} be some real-valued variable that only affects meeting probabilities. We have (assuming differentiability):

$$\frac{\partial P_{ij}(a_{ij} = 1, a_{ji} = 1|x_{ij})}{\partial x_{ij}} = P_{ij}(v_{ij} = 1, v_{ji} = 1) \frac{\partial P_{ij}(m_{ij} = 1|x_{ij})}{\partial x_{ij}}.$$

Even if $\frac{\partial P_{ij}(m_{ij}=1|x_{ij})}{\partial x_{ij}}$ is constant for all ij , the marginal impact of increasing x_{ij} is larger for pairs for which $P_{ij}(v_{ij} = 1, v_{ji} = 1)$ is larger (i.e. more valuable links). In the presence of homophily in students' preferences, the value of a link between similar students is higher, and therefore *even a homogeneous increase in meeting probabilities will increase homophily*. In reality, $\frac{\partial P_{ij}(m_{ij}=1|x_{ij})}{\partial x_{ij}}$ is likely to vary among pairs, but we show that the intuition carries through to a more general setting.

We now introduce our main structural assumptions:

$$P_{ij}(m_{ij}) \propto \exp\{m_{ij}(\mathbf{x}_{ij}\boldsymbol{\theta} + \delta w_{ij})\}, \quad (2)$$

$$P_{ij}(v_{ij}, v_{ji}) \propto \exp(v_{ij}\mathbf{x}_{ij}\boldsymbol{\beta} + v_{ji}\mathbf{x}_{ji}\boldsymbol{\beta} + \gamma v_{ij}v_{ji}), \quad (3)$$

where $\mathbf{x}_{ij}$ is again a vector of observable characteristics.

We assume that the probability of a meaningful meeting between i and j depends on their observable characteristics as well as on the (weighted) number of classes they have together. The value of a friendship between i and j depends on their observable characteristics, on whether the value is positive for both students (reciprocity), but not on the number of classes they have together.⁵

While the structural model for meeting probabilities in Equation (2) is fairly standard, the structural model for the probability of valuable friendship nomination in Equation (3) depends on the preferences of both students involved and deserves some discussion.

Equation (3) belongs to the class of exponential random graph models (ERGM).⁶ ERGM

⁵This exclusion assumption is in essence an identification assumption: we assume that exposure does not affect preferences. We believe that this is a reasonable assumption given that we are looking at short term effects. Note that the model would technically still be identified without this exclusion restriction, but in that case, identification would heavily rely on the functional form chosen (i.e. Logistic).

⁶Uses of ERGM in economics can be found e.g. in [Mele \(2017\)](#), [Badev \(2021\)](#) and [Hsieh et al. \(2020\)](#).

were introduced in the statistical literature on network formation as a way to introduce additional correlations between links in social networks. Indeed, conditional on observables, networks typically feature more links than what would be predicted assuming that links are independently created. (Jackson, 2008)

In the context of this paper, a large fraction of friendships are reciprocal (i.e. i nominates j and j nominates i). Since meetings are symmetric, the dependence between friendship nominations must therefore come from preferences.⁷ Similar strategies have been used recently in economics by e.g. Mele (2020).

The dependence between individual decisions implies that standard microfoundations (i.e. the random utility framework in McFadden (1974)) do not apply to Equation (3). Indeed, since students' choices are interdependent, the underlying microeconomic framework leads to a *game*. ERGM can nonetheless be rationalized, using (at least) three types of observationally equivalent microeconomic models.

First, Mele (2017) proposes a model in which Equation (3) is interpreted as the steady-state distribution of a random meeting process in which pairs of students periodically have the occasion of reassessing their friendship decisions. This narrative is less adapted to our context, in which we explicitly model meetings. Second, Equation (3) can be interpreted as the likelihood of the unique Nash equilibrium that maximizes the potential function of the friendship nomination game. This is a natural equilibrium refinement as it features some robustness properties. (Carbonell-Nicolau and McLean, 2014; Ui, 2001) More details can be found in Supplementary Appendix.

Third, Hsieh et al. (2020) present a microfoundation for Equation (3) using a game with transferable utilities, which we adopt here. We briefly expose the intuition here and refer the interested reader to Hsieh et al. (2020) for a formal treatment. Assume that individuals have utility

$$u_i(\mathbf{A}) = \sum_{j \neq i} a_{ij}(\mathbf{x}_{ij}\boldsymbol{\beta} + \gamma \frac{a_{ji}}{2} + \eta_{ij}),$$

where the η_{ij} are random utility shocks, and where $\mathbf{A}$ is the matrix of friendship decisions a_{ij} for all i, j . Since friendship decisions are conditional on meetings, individuals' decisions are constrained so that $m_{ij} = 0$ implies that $a_{ij} = a_{ji} = 0$.

If $\gamma \geq 0$, the value for i of a friendship with j increases if j decides to befriend i . This

⁷An alternative is to assume correlation in unobservables (Hoff et al., 2002). For example, the unobserved component of the utility functions for i and j could be positively correlated for any i and j , e.g. $P_{ij}(v_{ij}, v_{ji}) = \Phi(\mathbf{x}_{ij}\boldsymbol{\beta}, \mathbf{x}_{ji}\boldsymbol{\beta}, \rho)$. The poor performance of this model in our context led us to choose the ERGM for our main specification.

externality incentivizes i to make a contingent “utility transfer” to j in order to convince j to befriend them (and vice versa). As usual in transferable utility games (see e.g. [Chiappori \(2017\)](#); [Dutta and Jackson \(2000\)](#)), the presence of these transfers is sufficient to guarantee efficiency, i.e. the equilibrium network $\mathbf{A}^*$ is such that

$$\mathbf{A}^* = \arg \max_{a_{ij} \in \{0, m_{ij}\}} \sum_i \sum_{j \neq i} a_{ij} \mathbf{x}_{ij} \boldsymbol{\beta} + \gamma a_{ij} a_{ji} / 2 + \varepsilon_{\mathbf{A}},$$

where $\varepsilon_{\mathbf{A}} = \sum_i \sum_{j \neq i} a_{ij} \eta_{ij}$. Thus, if $\varepsilon_{\mathbf{A}}$ follows a Type 1 extreme value distribution (iid across all the possible realisations of $\mathbf{A}$), then the distribution of the equilibrium network $\mathbf{A}^*$ can be represented by Equation (3).

To summarize, the likelihood of observing the network of friendship nominations $\mathbf{A}$, conditional on meetings, is given by the likelihood that this friendship network is the (unique) outcome of the associated game with transferable utility.

4.1 Data Errors

As documented by [Boucher and Houndetoungan \(2025\)](#), the Add Health friendship dataset contains many missing links: roughly a third of friendship relationships are missing due to coding errors. Specifically, we may observe that student i nominated someone as a friend, but that individual cannot be associated with a valid identifier in the dataset. This is problematic for the estimation of the model since the data will feature more missing links than what would be observed without data errors. In order to account for this, we treat the problem as one of selection on outcomes, following [Manski and Lerman \(1977\)](#).

Specifically, we assume that links are missing at random (conditional on students’ characteristics), which allows us to compute the probability that a link between i and j is absent due to data errors, conditional on i and j ’s characteristics. Using this probability, we can weight each observation in our sample. For example, links ij such that $a_{ij} = a_{ji} = 0$ have smaller weights since they are potentially affected by data errors (they are too frequent in the sample, as opposed to the sample we would have obtained without data errors). Details of the construction of these weights are presented in Supplementary Appendix.

In the next section, we present the results for the estimation of $\boldsymbol{\theta}$, δ , $\boldsymbol{\beta}$ and γ via a (weighted) maximum likelihood estimator.

4.2 Results

Estimation results are displayed in Table 4 and discussed below.

Meeting probability The first block of panel A presents estimates for the meeting probabilities. The number of joint classes strongly increases the probability of meaningful meeting. Same-race students and students of the same grade level have a higher probability of meeting. However, same-sex students have a lower meeting probability.

Positive payoffs The second block of panel A displays estimates for friendship decisions conditional on having had a meaningful meeting. Preferences for friendship are affected by homophily. Estimates show that friendships are significantly more valuable for pairs of students of the same race, sex or grade level. As expected, the reciprocity parameter γ is also positive and significant, meaning that reciprocal relationships are valued by students.

Baseline predictions Panel B of Table 4 shows the predicted probabilities for meaningful meetings and for friendship links. The average probability of meaningful meeting is roughly 13%. Note that the probabilities of friendship nominations are very low due to the small number of nominations relative to the number of potential friendship relations.

Average marginal effects Table 5 presents the average marginal effects obtained through simulations, derived from random draws from the distribution of our estimates. This enables us to compare the magnitudes of the effects of joint classes and pairs' characteristics on our model probabilities and to quantify the role of homophily in meetings and in preferences in friendship nominations.

Looking at meaningful meeting probabilities in Panel A, our measure of exposure through joint classes is the main driver of meetings. Regarding the effect of pairs' characteristics, we find that switching to a pair of same-race students increases by 7.4 percentage points, on average, the probability of meaningful meeting, relative to that for not same-race students. Being in the same grade level increases the probability of having met in a meaningful way by 2.8 percentage points, on average. In contrast, our marginal effects estimates indicate that being of the same sex decreases the meaningful probability of meeting by 4.7 percentage points. Nevertheless, the overall effect of being of the same sex on the probability of mutual friendship nominations is positive, as shown in Panel C.

Panel A: Estimates		
Meeting probability	Estimate	Standard error
Constant	-3.2034	0.2546
Number of joint classes	3.4463	0.4813
Same Race	1.0275	0.1604
Same Sex	-0.5735	0.1222
Same Grade	0.3306	0.1648
Utility - <i>SFE</i> included but not shown		
Constant	-4.6823	0.4150
Same Race	0.5195	0.1053
Same Sex	0.4639	0.0724
Same Grade	0.8551	0.0922
Reciprocity (γ)	3.3589	0.1393
Mother's education:		
Less than HS	0.0568	0.0886
BTV school	0.0759	0.0641
Went to college	0.0509	0.0810
College graduate	-0.0192	0.0669
Beyond college	0.0609	0.1126
Unknown	-0.1514	0.0842
Panel B: Baseline predicted probabilities		
	Mean	Standard error
Meaningful meeting	0.1256	0.1677
Friendship nominations:		
P(00)	0.9922	0.0191
P(10)	0.0026	0.0053
P(01)	0.0026	0.0053
P(11)	0.0026	0.0088

Notes: $N = 1,083,660$. Panel A reports estimates of the friendship formation model along with standard errors. The model is estimated by weighted maximum likelihood using a bootstrap with 500 iterations. For each iteration, we resample individuals with replacement at the school level, predict the fraction of missing links based on observables, construct the corresponding sample of pairs, estimate the associated weights to correct for missing links, and then re-estimate the structural model. School fixed effects are included in the equation of utility. Panel B reports predicted probabilities evaluated at the estimated parameters.

Table 4: Estimation results

Panel B highlights that the marginal effect of being a same-race pair of students affects the probability of deriving a positive utility from a mutual nomination in the same order of magnitude as whether or not this is a same-sex pair of students, by roughly 1 percentage point.

Panel C displays the overall average marginal effects of exposure and pairs' characteristics on mutual nomination probability, without differentiating the effect solely coming through meetings and that from preferences. It is interesting to note that joint classes have roughly the same effect on friendship nomination as the type of pair, based on race and grade level.

Meetings and preferences in racial homophily We now quantify the mechanisms through which racial homophily affects the probability of mutual friendship nominations, distinguishing between two channels: meeting opportunities and preference-driven utility.

The total marginal effect of racial homophily is of 0.00297, which we decompose into two channels. The first one is the average marginal effect of switching to same-race pairs on meetings weighted by the mean probability of deriving a positive utility. This is what we define as *meeting-based* racial homophily. The second one is *preference-based* and is the average marginal effect of switching to same-race pairs on utility, weighted by the mean probability of meetings.

We find that the estimated average effect of *preference-based* racial homophily is 0.0019, while the *meeting-based* effect is 0.0011. Hence, racial homophily resulting from preferences is 1.75 times bigger than that coming from meeting opportunities. In other words, we find that *preference-based* racial homophily explains 63.7% of the total effect of racial homophily on mutual friendship nominations, while *meeting-based* one explains the remaining 36.3%.

As pointed out in the introduction, preferences are much harder to change through policies than meeting opportunities in the short run. Hence, even if the role of preferences is the main driver of race-based homophily in our friendship networks, meeting opportunities also play a significant role in explaining these patterns. This still leaves room for policies aiming to promote inter-racial integration through higher exposure of different types of students. We study two potential policies targeting meetings in the next section.

Panel A - Marginal effects on		
Meeting probability	Mean	Std. Err.
Joint classes	0.283	0.0175
Same Race	0.074	0.0081
Same Sex	-0.047	0.0071
Same Grade	0.028	0.0091
Panel B - Marginal effects on		
Payoffs - <i>Mutual nominations</i>		
Same Race	0.00986	0.0011
Same Sex	0.0099	0.0011
Same Grade	0.0221	0.0014
Panel C - Marginal effects on		
Mutual nomination probability		
Joint classes	0.004	0.0002
Same Race	0.003	0.0001
Same Sex	0.001	0.0001
Same Grade	0.003	0.0002

Notes: This table reports the average marginal effects of joint classes and pair-level characteristics on the probability of a meaningful meeting (panel A), the probability of deriving positive utility (panel B), and the probability of mutual nomination (panel C). Average marginal effects and their standard errors are computed by simulation ($N_{\text{reps}} = 100$), drawing parameter values from the estimated distribution $\mathcal{N}(\hat{\beta}, \hat{\Omega})$.

Table 5: Average marginal effects

5 Policy simulations

In order to gain more insight from the results of the structural model, we present two policy simulations aiming to increase inter-racial friendships. As discussed in the introduction, policy makers cannot directly affect students’ preferences, at least in the short run.

We present their effects on meetings and friendship links by pair type. We also simulate the segregation measures resulting from the policies, relative to the baseline. We focus on Freeman’s segregation index (Freeman, 1978), Coleman’s homophily index Coleman (1958) and

the Spectral Segregation Index (Echenique and Fryer Jr, 2007).⁸

5.1 Increased meetings

We assume that the policy maker is able to generate meetings. Consider for example a teacher who decides to assign students to team work. In order to mimic such a policy, we simulate an higher probability of meaningful meetings starting from the meeting probability Equation (2), evaluated at the estimated values of the parameters (see Table 4). We present two versions of the policy, targeting an increase of the probability of meaningful meetings by a third (i.e. from 12.56% to 16.7%). In the first one (“Unbiased”), we increase meetings through an increase in the constant term in Equation (2). In the second one (“Biased”), we increase the mean meeting probability through an increase in the constant term, combined with a decrease in the parameter associated with racial homophily. Then, conditional on the resulting meetings, we simulate friendship nominations using Equation (3), again evaluated at the estimated values of the parameters (see Table 4).

	Baseline	Unbiased	Biased
Total number of meetings	67 490	90 073	90 441
Same-race meetings ($\#SR_{meet}$)	52 638	69 322	66 285
Not same-race meetings ($\#NSR_{meet}$)	14 852	20 751	24 156
$\frac{\#SR_{meet}}{\#NSR_{meet}}$	3.54	3.34	2.74
Total number of links	5641	7055	7025
Same-race links ($\#SR$)	4940	6118	5978
Not same-race links ($\#NSR$)	701	937	1047
$\frac{\#SR}{\#NSR}$	7.05	6.53	5.71

Notes : This table reports the number of meetings and friendship links resulting from counterfactual scenarios where the probability of meaningful meetings is increased by one third relative to the baseline. In the “Unbiased” policy, the increase comes from a higher constant in the meeting equation, while in the “Biased” policy it is combined with a lower racial homophily parameter. Friendship nominations are then simulated using the estimated payoff equation.

Table 6: Increased meetings - Counterfactual results

Table 6 displays the results of both versions of the policy. The overall number of meaningful meetings consistently increases by a third relative to the baseline. Because of homophily in meetings, the relative prevalence of same-race meetings compared to meetings among different-race students decreases more sharply in the “Biased” version of the policy than in the “Unbiased”

⁸Our results regarding the index in Echenique and Fryer Jr (2007) should be interpreted with caution. Indeed, the index heavily relies on transitive triads. Since our structural model (ERGM) only features reciprocity, we underpredict the number of transitive triads compared with the observed network.

one.

Turning to friendship links conditional on meetings, the overall number of links increases less in the “Biased” version than in the “Unbiased” one because of homophilic preferences conditional on meetings. Both policies succeed in reducing the ratio of intra- to inter-racial meetings and friendships relative to the baseline. Nevertheless, the “Biased” one does better because part of racial homophily occurring through meetings is mitigated in the scenario where homophily in meetings is reduced.

Figure 1 displays the distribution of the segregation measures. The baseline segregation distribution is displayed in green, that for the “Unbiased” increase in meetings in orange and the “Biased” one in purple. The Freeman segregation index and Coleman’s homophily index distributions are coherent with the results of Table 6 as both policies reduce segregation. The spectral segregation index however implies an *increase* in segregation for both policies, relative to the baseline.

Two particular features of the SSI explain this finding. First, the SSI assumes a constant “time” constraint: by increasing a student’s number of friends, the index assumes that students decrease the time they spend with each individual friend (thus it also reduces the strength of the relationship). Second, and perhaps more importantly, the SSI assumes that a student’s segregation index is a linear function of their friends’. Thus, segregation indices are compounded. Increasing the number of friendships among racially segregated students therefore increases segregation.

5.2 Class composition

We assume that the policy maker can affect class assignments. For example, school administrators could use information about students’ race when assigning students to classes, hoping to foster inter-racial interactions. In the context of this paper however, we do not know the students’ classroom allocation or the constraints faced by the school administrators. The only information we have access to is the weighted number of classes that students have in common, i.e. w_{ij} .

In order to capture the effect of what would happen if school administrators would bias the allocation of students to classes in order to increase inter-racial meetings, we adopt the following strategy. We (randomly) increase w_{ij} for students of different races, keeping the distribution of w_{ij} constant. Specifically, let s be the share of same-race pairs of students ($s = 58\%$), $\tilde{w}_{ij}$ be a random draw from the sample distribution of w_{ij} , and $\mu > 0$ be a policy shock. We generate an

alternative variable

$$w_{ij}^{\mu} = \begin{cases} \tilde{w}_{ij} - \frac{(1-s)}{s}\mu & \text{if } i \text{ and } j \text{ are of the same race.} \\ \tilde{w}_{ij} + \mu & \text{if } i \text{ and } j \text{ are of different races.} \end{cases}$$

We then draw friendship links at the baseline and after the policy implementation using Equation (3) and using either the predicted meeting probabilities at the baseline or the ones from w_{ij}^{μ} .

	Baseline	Policy
Total number of meetings	67 590	70 850
Same-race meetings ($\#SR_{meet}$)	52 638	39 371
Not same-race meetings ($\#NSR_{meet}$)	14 852	31 479
$\frac{\#SR_{meet}}{\#NSR_{meet}}$	3.54	1.25
Total number of links	5641	3735
Same-race links ($\#SR$)	4940	2670
Not same-race links ($\#NSR$)	701	1065
$\frac{\#SR}{\#NSR}$	7.05	2.51

Notes: This table reports the number of meetings and friendship links at the baseline and those resulting from a counterfactual scenario where we increase inter-racial meetings through an increase in the number of joint classes between students of different races (by one half of the joint classes standard deviation, i.e. 0.242) while keeping the overall distribution constant.

Table 7: Changes in classes composition - Counterfactual results

Table 7 displays the results of this policy. While the total number of meetings slightly increases, the total number of friendship relations decreases sharply. Since the average student with whom a student interacts is now more likely to be of a different race, the average value of prospective friendship is lower. Yet, the policy has the desired effect on the ratio of intra-to inter-racial number of friendships.

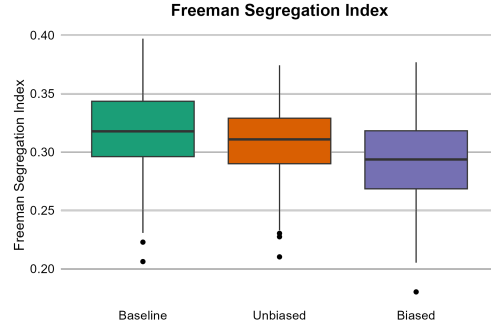
This policy therefore introduces a tradeoff. On the one hand, students' friendship network is less segregated (as confirmed by both the Freeman, Coleman and SSI indices, see Figure 2), but on the other hand, students have fewer friends.

6 Discussion

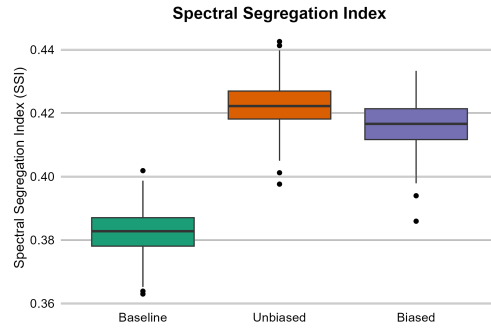
We analyze the relative contributions of meeting opportunities and preferences on the realized levels of racial homophily in friendship networks. We present evidence that race-based homophily increases the positive effect of exposure on the probability of a friendship relation. Building on the exponential random graph literature, we develop an intuitive structural model that reproduces this phenomenon and, most importantly, that allow us to disentangle the role of meeting opportunities from that of preferences. We find that most of the realized homophily acts through preferences, but that meetings still account for more than a third of realized homophily. These findings, together with those of [Chetty et al. \(2022b\)](#), confirm the importance of exposure to a diversity of individuals for homophily and social mobility.

However, our policy simulations highlight that, while exposure to a more diverse set of individuals has the potential to reduce homophily, it may be at the cost of a reduction in the number of connections. Since the overall number of connections is also a predictor of social mobility, ([Belot and Ermisch, 2009](#); [Ho, 2016](#)) our results indicate that special care is required to ensure that policies aimed at increasing social integration do not weaken social connectedness.

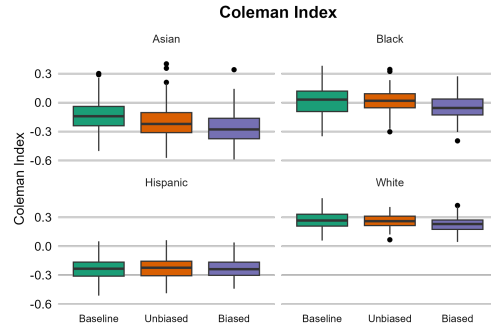
In summary, our results complement those of [Chetty et al. \(2022b\)](#) in showing that friendship relations are the result of both homophilous meetings and homophilous preferences. Policies aimed at increasing social mobility by promoting or incentivizing inter-ethnic interactions must ensure that they are not inadvertently reducing the number of friends individuals have. For example, policies increasing the overall number of meetings individuals have, even if these meetings are biased towards diversity, are likely more effective. Finally, our results imply caution in the interpretation of classical segregation indices, implicitly assuming a fixed level of meeting opportunities across groups.



(a) Freeman (1978)



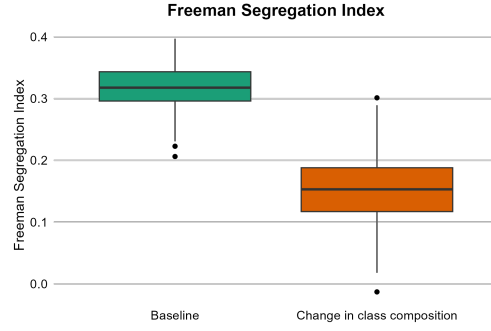
(b) SSI - Echenique & Fryer (2007)



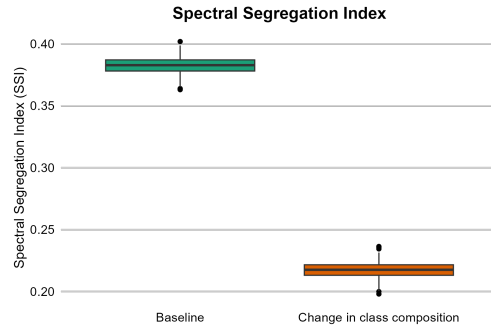
(c) Coleman's homophily index (1958)

Notes: The figures display segregation distributions at the baseline and under a counterfactual scenario where the probability of meaningful meetings is raised by one third. In the “Unbiased” policy, this increase comes from a higher constant in the meeting equation, while in the “Biased” policy it is combined with a lower racial homophily parameter. Segregation distributions are obtained by simulation (150 repetitions): in each scenario, we draw friendship nominations, construct the implied school-level adjacency matrix, and compute segregation measures, storing the mean index for each iteration.

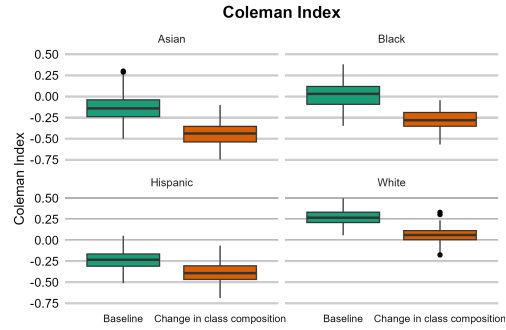
Figure 1: Segregation measures - Increased meetings



(a) Freeman (1978)



(b) SSI - Echenique & Fryer (2007)



(c) Coleman's homophily index (1958)

Notes : The figures display segregation distributions at the baseline and under a counterfactual scenario where we increase inter-racial meetings through an increase in the number of joint classes between students of different races (by one half of the joint classes standard deviation, i.e. 0.242) while keeping the overall distribution constant.. Segregation distributions are obtained by simulation (150 repetitions): in each scenario, we draw friendship nominations, construct the implied school-level adjacency matrix, and compute segregation measures, storing the mean index for each iteration.

Figure 2: Segregation measures - Changes in class composition

References

- ALLPORT, G. W. (1954): *The nature of prejudice*, Addison-Wesley.
- ANGRIST, J. D. AND K. LANG (2002): “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program,” *An exception is Jonathan Guryan*.
- BADEV, A. (2021): “Nash Equilibria on (Un)Stable Networks,” *Econometrica*, 89, 1179–1206.
- BELOT, M. AND J. ERMISCH (2009): “Friendship ties and geographical mobility: Evidence from Great Britain,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, 172, 427–442.
- BILLINGS, S. B., D. J. DEMING, AND J. ROCKOFF (2014): “School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg,” *Quarterly Journal of Economics*, 129, 435–476.
- BOOZER, M. A., A. B. KRUEGER, S. WOLKON, J. C. HALTIWANGER, AND G. LOURY (1992): “Race and School Quality Since Brown v. Board of Education,” *Source: Brookings Papers on Economic Activity. Microeconomics*, 1992, 269–338.
- BOUCHER, V. (2015): “Structural Homophily,” *International Economic Review*, 56, 235–264.
- BOUCHER, V. AND E. A. HOUNDETOUNGAN (2025): “Estimating peer effects using partial network data,” *The Review of Economics and Statistics*, Forthcoming.
- BOUCHER, V. AND I. MOURIFIÉ (2017): “My friend far, far away: a random field approach to exponential random graph models,” *Econometrics Journal*, 20, S14–S46.
- BOUCHER, V., S. TUMEN, M. VLASSOPOULOS, J. WAHBA, Y. ZENOU, W. THANK, S. MOUSA, L. MINALE, S. SEQUEIRA, J. GARCIA-HOMBRADOS, C. AKSOY, O. SAKA, P. OZEL, S. POLAT, M. LOWE, AND M. PELLIZZARI (2025): “Ethnic Mixing in Early Childhood: Evidence from a Randomized Field Experiment and a Structural Model,” *Working paper*.
- BRAMOULLÉ, Y., S. CURRARINI, M. O. JACKSON, P. PIN, AND B. W. ROGERS (2012): “Homophily and long-run integration in social networks,” *Journal of Economic Theory*, 147, 1754–1786.

- CARBONELL-NICOLAU, O. AND R. P. MCLEAN (2014): “Refinements of Nash equilibrium in potential games,” *Theoretical Economics*, 9, 555–582.
- CARD, D. AND A. B. KRUEGER (1992): “School Quality and Black-White Relative Earnings: A Direct Assessment,” *Source: The Quarterly Journal of Economics*, 107, 151–200.
- CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): “The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment,” *American Economic Review*, 106, 855–902.
- CHETTY, R., M. O. JACKSON, T. KUCHLER, J. STROEBEL, N. HENDREN, R. B. FLUEGGE, S. GONG, F. GONZALEZ, A. GRONDIN, M. JACOB, ET AL. (2022a): “Social capital I: measurement and associations with economic mobility,” *Nature*, 608, 108–121.
- (2022b): “Social capital II: determinants of economic connectedness,” *Nature*, 608, 122–134.
- CHIAPPORI, P.-A. (2017): “Matching under Transferable Utility: Theory,” *Matching with Transfers*.
- COLEMAN, J. S. (1958): “Relational analysis: The study of social organizations with survey methods,” *Human organization*, 17, 28–36.
- CURRARINI, JACKSON, AND PINE (2009): “An Economic Model of Friendship: Homophily, Minorities, and Segregation,” *Econometrica*, 77, 1003–1045.
- CUTLER, D. M. AND E. L. GLAESER (1997): “Are ghettos good or bad?” *The Quarterly Journal of Economics*, 112, 827–872.
- DE MARTÍ, J. AND Y. ZENOU (2017): “Segregation in Friendship Networks,” *Scandinavian Journal of Economics*, 119, 656–708.
- DUTTA, B. AND M. O. JACKSON (2000): “The stability and efficiency of directed communication networks,” *Review of Economic Design*, 5, 251–272.
- ECHENIQUE, F. AND R. G. FRYER JR (2007): “A measure of segregation based on social interactions,” *The Quarterly Journal of Economics*, 122, 441–485.
- FANTOZZI ALVAREZ, L. A., C. CAMPOS, X. PINTO, AND V. P. PONCZEK (2022): “Homophily in preferences or meetings? Identifying and estimating an iterative network formation model *,” Tech. rep.

- FREEMAN, L. C. (1978): “Segregation in Social Networks,” *Methods & Research*, 6.
- HO, C. Y. (2016): “The relationship from friendship links to educational achievement,” *The BE Journal of Economic Analysis & Policy*, 16, 1563–1572.
- HOFF, P. D., A. E. RAFTERY, AND M. S. HANDCOCK (2002): “Latent space approaches to social network analysis,” *Journal of the american Statistical association*, 97, 1090–1098.
- HSIEH, C.-S., L.-F. LEE, AND V. BOUCHER (2020): “Specification and estimation of network formation and network interaction models with the exponential probability distribution,” *Quantitative Economics*, 11, 1349–1390.
- JACKSON, M. O. (2008): “Social and Economic Networks | Princeton University Press,” .
- MANSKI, C. F. AND S. R. LERMAN (1977): “The estimation of choice probabilities from choice based samples,” *Econometrica: Journal of the Econometric Society*, 1977–1988.
- McFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Conditional Logit Analysis of Qualitative Choice Behavior*. In P. Zarembka (ed.), *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.
- MCPHERSON, M., L. SMITH-LOVIN, AND J. M. COOK (2001): “Birds of a feather: Homophily in social networks,” *Annual review of sociology*, 27, 415–444.
- MELE, A. (2017): “A Structural Model of Dense Network Formation,” *Econometrica*, 85, 825–850.
- (2020): “Does school desegregation promote diverse interactions? An equilibrium model of segregation within schools,” *American Economic Journal: Economic Policy*, 12, 228–257.
- MOODY, J. (2001): “Race, school integration, and friendship segregation in America,” *American journal of Sociology*, 107, 679–716.
- PETTIGREW, T. F. AND L. R. TROPP (2008): “How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators,” *European Journal of Social Psychology*, 38, 922–934.
- PUTNAM, R. D. (2016): *Our kids: The American dream in crisis*, Simon and Schuster.
- RAO, G. (2019): “Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools,” *American Economic Review*, 109, 774–809.

- REPORT, S. (2019): “The Wealth of Relations: Expanding Opportunity by Strengthening Families, Communities and Civil Society.” *Social Capital Project*, 3–19.
- TOOMET, O., M. VAN DER LEIJ, AND M. ROLFE (2013): “Social networks and labor market inequality between ethnicities and races,” *Network Science*, 1, 321–352.
- UI, T. (2001): “Robust Equilibria of Potential Games,” *Econometrica*, 69, 1373–1380.