Social Networks and (Political) Assimilation in the Age of Mass Migration*

Costanza Biavaschi† Corrado Giulietti‡ Yves Zenou§

February 2, 2022

Abstract

This paper investigates the causal pathways through which ethnic social networks influence individual naturalization. Using the complete-count Census of 1930, we digitize information on the exact residence of newly arrived immigrants in New York City. This allows us to define networks with a granularity detail that was not used before for historical data – the Census block – and therefore to overcome issues of spatial sorting. By matching individual observations with the complete-count Census of 1940, we estimate the impact that the exogenous fraction of naturalized co-ethnics in the network observed in 1930 has on the probability of immigrants to acquire citizenship a decade later. Our results indicate that the concentration of naturalized co-ethnics in the network positively affects individual naturalization and that this relationship operates through one main channel: information dissemination. Indeed, immigrants who live among naturalized co-ethnics are more likely to naturalize because they have greater access to critical information about the benefits and procedures of naturalization.

JEL codes: J61, J62, N32, Z1
Keywords: Social networks, assimilation, naturalization, migration.

*We thank Sascha Becker and Yannay Spitzer for very helpful comments.

†Norwegian University of Science and Technology, Norway. E-mail: costanza.biavaschi@ntnu.no

‡University of Southampton, UK; Centre for Population Change; Global Labor Organization. E-mail: c.giulietti@soton.ac.uk

§Monash University, Australia; CEPR; IZA. E-mail: yves.zenou@monash.edu

1 Introduction

How do (characteristics of) ethnic enclaves shape behavior? Several studies have shown that ethnic enclaves positively impact the economic outcomes of minorities (Edin et al., 2003, Munshi, 2003, Damm, 2009, Beaman, 2012, Patacchini and Zenou, 2012, Burgess et al., 2014), whereas other studies have reached an opposite conclusion (Borjas, 1995, Cutler and Glaeser, 1997, Lazear, 1999, Cutler et al., 2008, Dustmann et al., 2016, Abramitzky et al., 2020, Eriksson, 2020).

In this paper, we examine the causal impact of ethnic enclaves (or co-ethnic networks) on the political assimilation of immigrants, measured with their acquisition of U.S. citizenship a decade after their arrival. Indeed, naturalization is the cornerstone of assimilation because it affects immigrants' potential political influence in the host country (Portes and Cutis, 1987). Unlike legal permanent residents, U.S. citizens can vote in federal elections, obtain access to government benefits, secure government jobs, sponsor immediate relatives for visas, and are guaranteed the right to remain permanently in the United States protected from deportation (Abascal, 2017, Amuedo-Dorantes and Lopez, 2021). At a broader level, citizenship acquisition determines the extent to which immigrants are willing to become an integral part of the host society and abide and pledge to its laws; from the perspective of the host country, the degree to which naturalization is facilitated or restricted is an indication of its willingness and ability to integrate immigrants into the society. Considering the importance of citizenship acquisition for both immigrants and host societies, understanding its determinants is of paramount significance because personal identification with the host country is considered a key marker of assimilation (Gordon, 1964) and a precondition of naturalization (DeSipio, 2001).[1] As Banulescu-Bogdan (2012) puts it: "citizenship is a significant milestone for immigrants: 'a rite of passage' to signal that newcomers take their rights and responsibilities seriously, and are to be recognized as full members of the community."

We study the role of ethnic networks in the political assimilation of immigrants by focusing on the decade 1930–1940. This was an important period of time in the U.S. history of immigration, as 1930 was considered the "climax" of the melting pot process, with cultural mixing and clashes between the natives and immigrants. We selected New York City (NYC) as a case study since, with some 12 million immigrants from all over the world having entered through Ellis Island during the first part of the 20th century, it best represents the paradigm of a diverse and multi-ethnic "gateway" into the U.S. at the end of the age of mass migration.

Our analysis is based on the matched 1930 and 1940 Censuses. The sample consists of *newly-arrived immigrants*, i.e., foreign persons who entered the U.S. during the period 1925–1930, for whom we observe exogenous network characteristics in 1930 and subsequent naturalization status in 1940. To define our network, we exploit the universe of all individuals in 1930. This allows us to construct network variables at a very fine geographic scale – the

[1] See, e.g., Hainmueller et al. (2015) who provide causal evidence on the long-term effects of naturalization on political integration. They find that naturalization causes long-lasting improvements in political integration, with immigrants becoming likely to vote and attaining considerably higher levels of political efficacy and political knowledge. See, also, Hainmueller et al. (2017), who show that naturalization improves the long-term social integration of immigrants into the host country.

block – a granularity level that was not used before for studying networks with historical data.[2] Our key measure of social networks is the fraction of naturalized immigrants from the same country of birth residing in the same block of newly-arrived immigrants. The reason why we focus on the 1930-1940 decades is because the 1930 census is the first to report the block of residence of the migrants. In addition, we focus our attention on recent arrivals since for these migrants the network is arguably exogenous – an aspect around which our identification strategy revolves.

The core objective of our empirical work is to study the causal impact of the social networks measured at the time of the immigrants' arrival on their probability of becoming U.S. citizens in 1940 – a decade later. Since dual citizenship was not allowed at the time, naturalization was a big commitment, as it implied renouncing to own nationality and, thus, not being able to go back to the country of origin.[3] In our econometric models, we account for several pre-determined individual and network characteristics measured in 1930. Crucially, in our specifications, we include fixed effects for the neighborhoods (defined by aggregations of the blocks known as enumeration districts) and for the immigrants' countries of birth. The latter account for the differences in naturalization patterns across origin countries, that is, unobserved factors specific to the country where immigrants come from and that could be correlated with the probability of acquiring citizenship. The neighborhood fixed effects capture the unobserved differences in the average immigrant characteristics across the enumeration districts of NYC and, importantly, control for the spatial sorting of the immigrants. Indeed, a newly arrived immigrant may choose a particular neighborhood over another based on her observation or knowledge of the characteristics of that neighborhood. For example, immigrants who have just arrived from Italy would tend to settle in neighborhoods with a relatively higher number of previous Italian immigrants. However, conditioning on having chosen a specific neighborhood, immigrants are unlikely to select the particular block where to live. Following the seminal work of Bayer et al. (2008), our key argument is that immigrants do not have information regarding the *differences* in the characteristics between blocks (i.e., the social networks) in the same enumeration district. In the example used above, newly arrived Italian immigrants would choose a neighborhood that is densely populated by Italians, but, within this area, depending on the availability of apartments, they will choose a random block in which to live. In particular, it is unlikely that they will be aware about how the *proportion of naturalized* Italians differs across the blocks within a given enumeration district.

Therefore, we treat the composition of the neighborhood by birthplace as quasi-random and use this exogenous block-level variation to test the effect of networks on individual naturalization. The assumption that immigrants do not self-select across blocks underpins the

[2]The block is very small. On average, there are about 55 individuals in each block (Table 2).

[3]One might be concerned about the fact that the rates of naturalization were closely linked to the rates of return migration, so that, even if a migrant stayed for a few years, she might have had a tendency to return and this might have affected her inclination to naturalize. However, in the particular context of the 1930s, it is reasonable to argue that almost anyone who stayed between 1930 and 1940 could be considered as a permanent migrant because, after the imposition of the quotas in the early 1920s, return migration became rare. This is especially true in the late 1930s, since, given the situation in Europe and, in particular, in Germany, no one would have contemplated returning back to her home country.

identification of the causal effect of social networks on citizenship acquisition. We provide solid and exhaustive empirical corroboration for this postulation.

We find a positive and significant effect of the share of naturalized co-nationals in 1930 on immigrants' likelihood to naturalize in 1940. A 10% increase in the neighborhood's co-ethnic naturalization rate is associated with about a 0.4% increase in the probability that an immigrant of that neighborhood naturalizes. This impact is non-linear: immigrants exposed to the highest rate of co-ethnic naturalization (upper quintile) exhibit a propensity to naturalize that is 2.5 percentage points higher than immigrants exposed to the lowest rate of co-ethnic naturalization (bottom quintile). The magnitude of the network effect also differs substantially depending on the country of origin (an aspect that we explore in depth throughout our analysis) and on the borough of residence. On the other hand, there does not seem to be large gender differences in the estimated effect.

To further substantiate the credibility of our results, we perform a series of robustness checks. First, we conduct placebo tests through a randomization-based inference procedure where we reassign at random the share of naturalized co-ethnic and estimate 5,000 regressions using these permuted values. We then compare the number of times the estimated coefficients from these simulations is larger than our baseline coefficient, concluding that the estimate from our preferred specification lies in the rejection area of the distribution of coefficients generated by random chance. In a second test, we employ a finer definition of network – the *block-street* area – which is defined by all the immigrants hailing from the same area of origin and living in the *same block and on the same street* as the newly arrived immigrants. With this specification – which allows us to exploit block fixed effects – we rule out the possibility that immigrants choose a specific block because of other confounders (e.g., workplace, friends, presence of local amenities) unobserved to us. Remarkably, the effect remains positive and statistically significant even with this demanding specification. In the last set of tests, we exclude blocks where immigrants live with people that have the same surname (who are likely to be a family member) and blocks where the immigrant is either a boarder, roomer, lodger, or a domestic worker. The pattern of the results remains unchanged.

We then explore the mechanisms behind our results. We find that networks only positively impact the naturalization propensity of immigrants from *non-English*-speaking countries while have no impact on those coming from English-speaking countries, such as Ireland and the UK. Moreover, we find that the fraction of American-born citizens as well as naturalized immigrants from other countries in 1930 have no significant impact on the probability of being naturalized in 1940. Combined with our main findings, these two results imply that it is earlier immigrants – who previously underwent the complicated naturalization process – who are more likely to help their newly arrived co-nationals apply for citizenship. For example, they help them go through the bureaucratic steps of naturalization and serve as witnesses when the immigrant petitions for naturalization. Since English-speaking naturalized immigrants as well as naturalized immigrants from other countries have no impact on a newly arrived immigrant, we believe that *information*, rather than social norms, is the key mechanism through which political assimilation occurred. Let us take the example of a newly arrived immigrant from

Russia who needs help in terms of language and is going through the bureaucratic process of applying for U.S. citizenship. It seems obvious that earlier Russian immigrants who are residing nearby and who have been naturalized will be of great help for the new immigrant. We believe that this mechanism has two dimensions. First, a newly arrived immigrant is more likely to socially interact with people from the same country of origin as a result of the overlapping social ties and shared language. In our example, the recent immigrant will find it less "costly" to learn about the naturalization process from naturalized Russians (instead of, for example, Italians) living in her community because she can easily connect with a naturalized immigrant through friends or family members or by frequenting the same store or church. This is a reflection of homophily being stronger within the same ethnic group (see, for example, Currarini et al., 2009). Second, these interactions are particularly important when they occur between immigrants from non-English-speaking countries (such as Russia in our example) because naturalized immigrants can provide practical help to the newly arrived immigrant by supplying useful information on the naturalization process using the native language. Finally, these checks also reassure against the possibility that the results are mechanically driven by the rules of the naturalization process – which required to bring two witnesses at the time of the oath of allegiance. If our results were merely driven by this aspect, we would have found a similar effect across all nationalities, including those living close to American-born individuals.

After establishing that information is the key mechanism at work, we examine which type of information matters in the naturalization process. We first show that the impact of networks is stronger when there is a higher concentration of naturalized co-ethnics in jobs where non-citizens face entry barriers. This suggests the importance of labor-market related information as an incentive to naturalize. Second, we show that the impact of network is driven by countries who were most affected by the restrictions introduced by the Immigration Act of 1924. With more restrictive quotas, the gains of naturalizing increased (e.g., relatives of citizens could enter the U.S. as non-quota immigrants) and thus our result is consistent with the network playing a role in disseminating information about the benefits of naturalizing associated with the Immigration Act. Third, we provide some evidence that networks might have channeled information about the importance of political participation. We do so by showing that stronger network effects exist in areas where a representative to the Congress from the Democratic party – who is generally pro-migration.

1.1 Related literature

Our paper contributes to different strands of the literature. First, we add to the literature on the determinants of the *assimilation* of immigrants. Different studies have shown distinct, significant influences over the assimilation process for immigrants: the quality of immigrant cohorts (Borjas, 1985), the country of origin (Borjas, 1992, Beenstock et al., 2010, Chiswick and Miller, 2011), the time spent in the host country (Abramitzky et al., 2020), the ethnic concentration (Edin et al., 2003, Damm, 2009), and the importance of language skills (Chiswick

and Miller, 1995, Dustmann and Fabbri, 2003).[4]

In this paper, we focus on the role played by ethnic networks on the *long-term* assimilation of immigrants. To the best of our knowledge, very few papers have tested the *causal* impact of ethnic networks on citizenship acquisition.[5] One exception is the paper by Bratsberg et al. (2021). They study the impact of a quasi-exogenous placement policy of the Norwegian refugee resettlement program in the 1990s on refugees' future electoral participation in 2015. They find that the political engagement of peers within the arrival location is strongly linked to refugees' future electoral participation. Clearly, their focus, time period, country of study, and outcome are very different from ours. Another exception is the study conducted by Shertzer (2016), who examines the impact of the co-ethnic group's share of the local electorate at the ward level on an immigrant's likelihood of becoming politically mobilized, as measured by citizenship attainment. Apart from the fact that the focus is quite different, it is difficult to claim causality in their paper because the geographical unit is the ward, which is quite large. A last exception is the study by Abramitzky et al. (2020), who use a program that funded 39,000 Jewish households in New York City to leave enclave neighborhoods between 1900 and 1919. They show that men who left the enclaves tended to assimilate more into the broader U.S. society (by marrying spouses with less Jewish names) that those who stayed in the enclaves. We view our paper as complementary, since we focus on the impact of ethnic enclaves on political assimilation for a broader population and a different time period.[6]

Second, we contribute to the recent empirical and econometric literature on *social networks*.[7] Traditionally, the endogeneity of network formation has received rather limited attention, and researchers mostly used an instrumental variable approach based on friends of friends' characteristics (Bramoullé et al., 2009, Calvó-Armengol et al., 2009). The most recent literature uses a structural approach that explicitly models the formation of networks (Goldsmith-Pinkham and Imbens, 2013, Boucher, 2016, Mele, 2017, Hsieh et al., 2020, Badev, 2021). There are also some very recent papers that use an exogenous source for the variation in network formation (Algan et al., 2020, Boucher et al., 2020, Banerjee et al., 2021, Comola and Prina, 2021, Heß et al., 2021). In our paper, we follow the seminal approach of Bayer et al. (2008), and like them we use the Census block to identify the network. Differently from them, though, we focus on a historical period for which an analysis of network at such detailed geographic level was not done before. In addition, we also enrich the design with an ethnicity dimension, as our network is define within each block and country of birth of immigrants. This design provides us with the "exogenous assignment" of the co-ethnic network to each immigrant and allows us to estimate the causal impact of networks on assimilation.

[4]Brell et al. (2020) provide a recent survey on the factors that affect the success of refugees' economic integration.

[5]Conversely, there is a wealth of papers that have shown the causal impact of ethnic enclaves on the economic outcomes of ethnic minorities, especially in the labor market. See, e.g., Edin et al. (2003), Damm (2009), Lafortune and Tessada (2019), Eriksson (2020), and Battisti et al. (2021). There is an active research area in the non-economic literature, but it mostly documents correlations. See Abascal (2017) and the references therein.

[6]There is also a very less theoretical literature on the impact of networks on assimilation. See, e.g., Verdier and Zenou (2017).

[7]For an overview, see Graham (2015), Bramoullé et al. (2016), Jackson et al. (2017), and De Paula (2020).

Finally, we contribute to the recent and growing literature on the age of mass migration, which has studied the role of religious institutions on the assimilation of immigrants (Gagliarducci and Tabellini, 2021), the impact of networks on the economic outcomes of immigrants (Hatton and Williamson, 1998, Connor, 2018, Eriksson, 2020),[8] the selection and the assimilation of European immigrants (Abramitzky et al., 2012, 2014, 2020), and their economic and political effects and their impact on economic growth and political ideology (Abramitzky et al., 2021, Giuliano and Tabellini, 2020, Sequeira et al., 2020, Tabellini, 2020).[9]

We complement this literature by being the first to study the causal impact of ethnic networks in 1930 on individual assimilation (i.e., citizenship) in 1940. One of our contributions is the construction of a unique dataset in which we provide the exact identity of all the households located in the same block and on the same street for each new immigrant in 1930. This allows us to construct a social network for each immigrant based on the country of birth and the citizenship of all the individuals residing in the same block and/or on the same street.

The rest of the paper progresses as follows. In the next section, we describe our dataset. In Section 3, we explain our identification strategy. Section 4 presents our main results as well as heterogeneity analyses and robustness checks. In Section 5, we investigate the mechanisms behind our results. Finally, Section 6 concludes the paper.

2 Data

To measure the role that social networks play in immigrant political assimilation, we exploit the 1930 and 1940 complete-count Censuses. The census contains the records of individuals with details about their characteristics, household composition and attributes, and the address of residence. In particular, the 1930 Census is the first one to report the block of residence of respondents. Such information is essential for our research design, as further explained below. The same analysis, therefore, could not be performed neither for earlier periods, where block information was missing, nor for later periods, as for example at the time of writing the complete-count 1950 census has not been released yet.[10] We further detail below the sample of interest and our exact definition of neighborhood and network.

We restrict our attention to all foreign-born individuals residing in NYC in 1930, who are not U.S. citizens, who are not born in outlying American areas or territories, and who entered the U.S. during the period of 1925–1930. We focus on this group for three main reasons. First, individuals who are not U.S.-born and not U.S. citizens can apply for naturalization. Second, newly arrived immigrants are unlikely to be already naturalized by law. To become a citizen, immigrants had to first reside continuously in the U.S. for at least two years, one of which had to be spent in the state in which the application was submitted, before filing a declaration of intention for citizenship (also known as "first papers"). Immigrants with first papers who

[8] See also Stuart and Taylor (2021) who study the role of migration networks on long-run location decisions in the mid-twentieth century but for American citizens only.

[9] For an overview, see Abramitzky and Boustan (2017).

[10] Even if the data were available, the period 1940–1950 is complicated to analyze because the United States entered World War II in December 1941.

had resided in the U.S. continuously for at least five years, and within the county in which the petition was being filed for at least six months, could then file a petition for admission to citizenship. At this final hearing, the applicant had to bring two witnesses, citizens of the U.S., who could testify to their continuous residence and "good moral character". Dual citizenship was not allowed at the time. In 1929, the full procedure costed a fee of $20 (roughly $304 today), which was reduced in 1934 to $12.5.[11] At the same time, with naturalization immigrants could vote but also access professions legally barred to non-citizens. Third, it is more likely that the network is exogenous for recently arrived immigrants.

We created a matched dataset that tracks immigrants from the 1930 census to the 1940 census. By creating this matched sample that follows the same individuals over a decade, we are able to compare the citizenship acquisition patterns in 1940 based on the characteristics observed in 1930, including neighborhood attributes. In our baseline analysis, we link men and women over time by first and last name, age, and country of birth, and in heterogeneity checks we show the results for men and women separately.[12] Further details on the main linking procedure are provided in Appendix A.

Our main linked sample, after eliminating observations with missing values, contains 37,761 individuals, corresponding to about 16% of our initial sample for 1930. Our matching rate is in line with historical studies that have used matched samples (e.g., Ferrie, 1996, Abramitzky et al., 2014, Abramitzky and Boustan, 2017). In the Appendix B, we provide some analysis on the balancing of covariates between the matched and unmatched sample, and we also elaborate on the potential impact of selectivity by estimating Heckman selection models of our baseline estimates.

2.1 Empirical definition of networks

To define networks, we first combine the information about enumeration districts, blocks of residence, and the regions of birth of the immigrants. We postulate that, upon arrival, the immigrants choose the area (neighborhood) in which they wanted to live. In our settings, we use the enumeration districts to identify the neighborhoods.

Enumeration districts were geographic areas assigned to an individual census taker, representing a specific portion of a city. As an example, Panel (a) of Figure 1 shows the subdivision of lower Manhattan into enumeration districts, while Panel (b) zooms in on the Enumeration District number 1174, bordered by Canal Street, Hudson Street, and Beach Street.

In our framework, we hypothesize that, while immigrants choose a given neighborhood (i.e., enumeration district), they are unable to select a precise sub-area (i.e., within the neighborhood) in which they will settle and develop their social tie (i.e., their network). We provide exhaustive arguments to support this conjecture in the next section.

[11] For comparison, the US today allows dual citizenship, and immigrants can apply for naturalization after five years of permanent residency, showing good moral character, English and civic knowledge and attachment to the Constitution. Fees usually amount to about $725.

[12] Matching rates between men and women might differ, as women marrying between 1930 and 1940 will change surname. However, as shown in Table 5 our results are unaffected when we look at males and females separately. We refer the interested reader to Appendix A for further details on the linking procedure.

To operationalize our definition of network, we partition each neighborhood into sub-areas. In practice, we use the *block of residence* of the immigrants – available for the first time with the 1930 census – to identify the sub-areas. We chose blocks for several reasons. First, they are the smallest spatial unit that can be identified in the census, as explained below. Second, blocks are self-contained within the enumeration districts. Third, they are well-defined areas, formed by intersection of physical features, such as streets, roads, and rivers, and are quite homogeneous in terms of size, making them an ideal dimension for approximating the social space in which individuals establish social contacts. As an example, Panel (b) of Figure 1 highlights Block J in Enumeration District 1174.

Although the block of residence could serve as a straightforward definition of the network, accessing such information is challenging both with current and with historical data. Nowadays, the census reports the census blocks, which typically correspond to city blocks; but these are available only in the restricted version of the census (Bayer et al., 2008). Going back in time, the information on the block of residence was available in the original census manifests of the 1930 Census. One such example is shown in Panel (c) of Figure 1. Yet, such information has never been digitized before (as opposed to enumeration districts, which are instead available in the complete-count microdata from IPUMS). For this reason, we accessed the original census manifests for the 6.9 million inhabitants of NYC in 1930 through "ancestry.com" and manually transcribed the block information, as shown in Panel (c). To the best of our knowledge, no other paper has ever included information at such a disaggregated level about the individual city block of residence for each observation in the sample.[13]

With the block information available, we assume that immigrants interact with other individuals living in the same block or – perhaps more realistically – that social interactions within the block are more intense than across blocks even within the same neighborhood.

Finally, we define an immigrant's social network by all individuals from the *same country of birth* and *living in the same block* as the immigrant, excluding their own household members. Our choice to define networks by the same country of birth is motivated by various reasons. First, immigrants tend to sort in areas where other immigrants have settled in the past. Second, ethnic communities tend to be more socially cohesive – something already emerging during the age of mass migration (see, e.g., Eriksson, 2020) – and, thus, social networks are more easily established within the same ethnic group. Third, ethnic networks are relevant. Data from the study conducted by Biavaschi et al. (2017) show that, across all origin countries, immigrants primarily relied on co-nationals during their naturalization procedure.

In Figure 2, we report the share of co-nationals among those who served as the first witness for the immigrant during the naturalization process. The results suggest that the immigrants who arrive from the major sending countries in our sample (Germany, Ireland, and Italy), strongly relied on their networks during the naturalization process. Among these groups, in

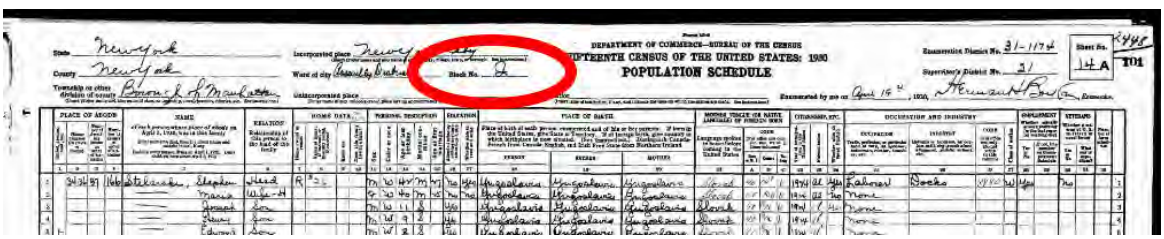
[13] Observe that the enumeration districts have been digitized and geolocalized for New York in 1930 and 1940 (see: https://s4.ad.brown.edu/Projects/UTP2/citymaps.htm and https://s4.ad.brown.edu/Projects/UTP2/ncities.htm). However, blocks, to the best of our knowledge, have never been digitized.

Figure 1: Enumeration districts and blocks

(a) Enumeration Districts in Lower Manhattan

(b) Enumeration district 1174 and Block J

(c) Block J in the Original Census Manifest



Source Panel 1a and 1b: United States enumeration district maps for the twelfth through the sixteenth US censuses, 1900–1940 images, FamilySearch, Roll 42, New York, New York City boroughs; Niagara-Rockland 1900–1940, image 675 of 875; citing NARA microfilm publication A3378 (Washington, D.C.: National Archives and Records Administration, 2003).
Source Panel 1c: Ancestry.com

fact, more than 80% brought a witness from their country of birth. These percentages are lower, but remain substantial, for other countries, such as Poland and Russia.[14]

To summarize, we assume that the network exposure of each immigrant i is given by the share of *naturalized* non-household members $ynet_{dbc,t}$ living at time t in the same enumeration district d and in the same block b and coming from the same country of birth c, that is,

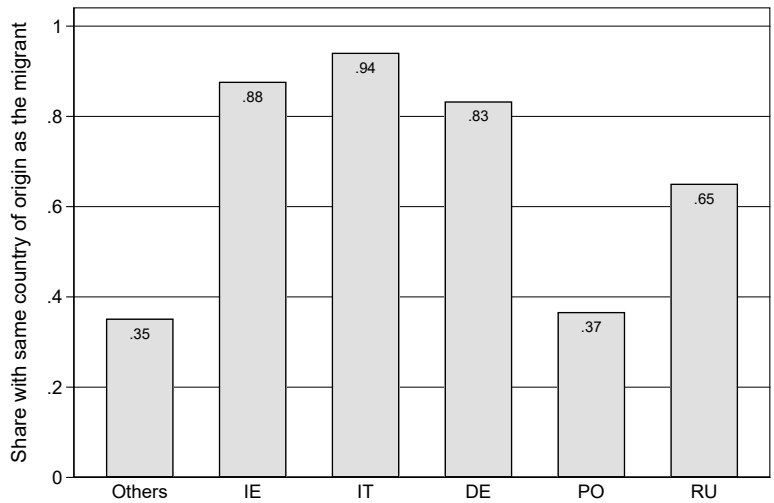
$$ynet_{dbc,t} = \frac{\sum_{j \neq i} y_{jdbc,t}}{N_{dbc,t} - 1}, \tag{1}$$

where $y_{jdbc,t}$ is a dummy variable that takes a value of 1 if an immigrant $jdbc, t$, i.e., an immigrant j outside i's household living in enumeration district d, block b at time t and born

[14] A part of the lower share among Polish and Russian immigrants might stem from the difference in classification between the census and the data collected by Biavaschi et al. (2017).

in the same country c as immigrant i, is a U.S. citizen. $N_{dbc,t}$ indicates the total number of network members of i, i.e., the total number of individuals $jdbc, t$, which includes both U.S. and non-U.S. citizens. For example, for an Italian immigrant, $ynet_{dbc,t}$ measures the fraction of U.S. citizens born in Italy (i.e., the number of U.S. citizens born in Italy divided by the total number of individuals born in Italy) residing in the same block and same enumeration district at time t as said Italian immigrant.

Figure 2: Share of co-nationals among the first witnesses during the naturalization process



Source: Data from the study of Biavaschi et al. (2017). The figure shows, conditioning on being a naturalized first witness, the share of witnesses with the same country of birth as that of the immigrant.

2.2 Descriptive Statistics

Table 1 shows the characteristics of immigrants in our sample as measured in the 1930 Census. About 48% of the immigrants are married, and the average household size is 5.[15] The average value of the log occupational score income is 1.5.[16] The table also reports network characteristics of our sample. In particular, our key variable of interest is $ynet_{dbc,t}$, the share of naturalized co-ethnics. This table shows that in 1930, about 49% of immigrants in the network are naturalized.

In terms of country of origin, 37% of the recent arrivals are from Germany, and 17% from Ireland. Other major sending countries are the UK (13%) and Italy (8%). This composition is unsurprising, given the restrictive measures adopted for migration in 1924, which imposed a tightening of the borders for immigrants coming from Italy, Russia, and Eastern Europe – which had been major sending countries during the first decades of the 1900s.

[15] About 20% are household heads, about 20% are spouses, 20% are children, 20% are family members with other relationships to the head (e.g., siblings, siblings-in-law, parents, etc.), and the remainder is composed of persons who were living in the same households but are not related to the household head (i.e., boarders, lodgers, roomers, and servants).

[16] This corresponds to about $10,000 in 2020.

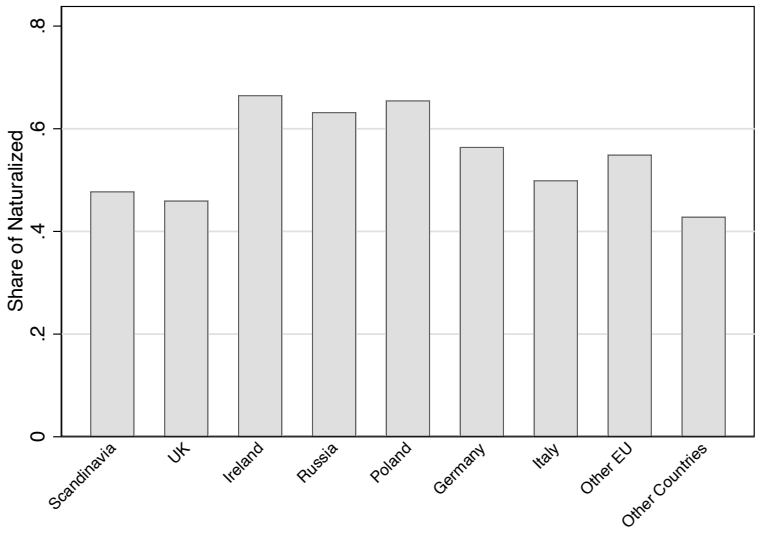
Table 1: Summary Statistics

Individual & Network Characteristics	Mean	St.Dev.	Countries of Birth	Mean	St.Dev.
Female	0.438	0.496	Scandinavia	0.051	0.221
Married	0.482	0.500	UK	0.128	0.334
Log Occupational Income Score	1.496	1.558	Ireland	0.167	0.373
Household Size	5.132	3.914	Russia	0.034	0.180
Share Naturalized in Network	0.493	0.222	Poland	0.047	0.212
Share of Females in Network	0.493	0.153	Germany	0.370	0.483
Share Married in Network	0.608	0.205	Italy	0.075	0.263
Avg. Log Occupat. Income Score in Network	1.469	0.494	Other EU	0.071	0.257
Avg. Household Size in Network	4.758	2.215	Other Countries	0.058	0.233

NOTES. Other countries include all other countries present in the dataset. Of these, the largest ones are Canada, New Foundland, Cuba, Turkey
A network are defined is defined as the group of immigrants from the same country of birth living in the same block of residence (excluding household members).
The sample is composed of 37,761 individuals and covers 17,488 networks across 4,305 neighborhoods.

Figure 3 provides some insight about the outcome of interest: naturalization in 1940. About 56% of the immigrants in our sample had naturalized throughout the decade after we observe the networks. Yet, as the figure shows, there is some variation in naturalization rates across countries of birth. Polish, Irish and Russian immigrants are the most likely to naturalize (with rates above 60%). On the other extreme, less than one out of two immigrants from Scandinavia and from the UK had naturalized by 1940.

Figure 3: Naturalization Rates by Country of Birth – 1940



The figure shows the percentage of immigrants from each country of birth who had naturalized by 1940.

Figure 4 provides some insight about the spatial distribution of the immigrants group in 1930 for selected countries of origin. The maps compare the relative size of the immigrant groups and their spatial concentration across neighborhoods (enumeration districts). They are constructed using the universe of immigrants in Manhattan and Brooklyn, the two counties for which geospatial information on enumeration districts currently exist.

This figure shows that immigrants from Ireland and Russia constitute relatively large groups but their spatial concentrations are quite different; the former is more concentrated in Manhattan, while the latter is overwhelmingly concentrated in Brooklyn. There are also plenty of immigrants from Italy, but they are spread throughout the two counties. Further, there are relatively less immigrants from Scandinavia, and they are mostly concentrated in certain areas of Brooklyn.

Last but not least, Table 2 provides descriptive statistics for the networks. These are constructed using all networks where our sample of immigrants reside. For comparison, Table C1 in the Appendix provides the same statistics for the whole New York City. Table 2 shows that in each neighborhood, there are on average 4 networks (i.e., block-country-of-birth combinations). In each network, there are on average 26 immigrant households and 55 immigrants, of which 26 are naturalized.

Table 2: Ethnic Networks Characteristics

	Mean	St.Dev.
N. Immigrant Networks in each Neighborhood	4.065	3.138
N. Immigrant Households in each Network	26.48	38.669
N. Immigrants in each Network	54.947	87.069
N. Naturalized Immigrants in each Network	26.707	43.804
N. Individuals in each Network	54.947	87.069

NOTES. Neighborhoods correspond to Enumeration Districts.
A network are defined is defined as the group of immigrants from the same country of birth living in the same block of residence (excluding household members).
The sample is composed of 37,761 individuals and covers 17,488 networks across 4,305 neighborhoods.

3 Empirical Framework

3.1 Econometric model

Our empirical model can be written as follows:

$$y_{idbc,1940} = \alpha + \beta ynet_{dbc,1930} + \eta_c + \eta_{d,1930} + \sum_{k=1}^{K} \theta^k x^k_{idbc,1930} + \sum_{j=1}^{J} \theta^j z^j_{dbc,1930} + u_{idbc,1940}, \quad (2)$$

where $y_{idbc,1940}$ is the citizenship status of individual *idbc* at time $t = 1940$, i.e., it is a dummy variable that takes a value of 1 if, in 1940, an immigrant *idbc* is a U.S. citizen, and 0 otherwise. The key explanatory variable is $ynet_{dbc,1930}$, the *share of naturalized co-ethnics in 1930*, as defined in equation (1). The term $x^k_{idbc,1930}$ contains the individual controls reported in Table 1. We also include a set of dummy variables for each one-year age group. The term $z^j_{dbc,1930}$ includes the same variables contained in $x^k_{idbc,1930}$, but constructed at network level (i.e., at the same level of our key regressor). These are also summarized in

Figure 4: Share of immigrants in Manhattan and Brooklyn – Enumeration Districts

Source: Enumeration district Shapefiles last retrieved in April 2019 at `https://s4.ad.brown.edu/Projects/UTP2/ncities.htm`. See Shertzer et al. (2016) for description. The immigrant shares are calculated from the 1930 US census as the number of people born in a given country in each enumeration district divided by the total population of each enumeration district.

Table 1. All control variables are measured in 1930.[17] Finally, we include η_c, country-of-birth fixed effects, and $\eta_{d,1930}$, enumeration district fixed effects (again defined in 1930).

[17]The fact that the outcome variable is measured ten years after the time the network is observed should naturally pre-empt any concern about the existence of the reflection problem (Manski, 1993).

The country-of-birth fixed effects account for the differences in naturalization patterns across origin areas. The enumeration district fixed effects control for any unobserved differences in the average immigrant characteristics across all areas in NYC, as well as for the aspects of the enumeration district that characterize all residents in that area.[18] Importantly, the country-of-birth and enumeration district fixed effects control for selection across enumeration districts. Thus, we treat the composition of the neighborhood by birthplace and naturalization within an enumeration district as quasi-random. We devote the next subsection to discuss and corroborate the plausibility of this assumption.

Our identification strategy builds on the local nature of our data and follows the research design of Bayer et al. (2008). In their paper, the authors identify the impact of neighborhood referrals by leveraging the exogenous block-level variation within the same neighborhood, as defined by the block group (which are aggregations of blocks and the smallest area for which census tabulations are available). In our approach, we use Enumeration Districts, which are the antecedents of the block groups and, similarly, consist of aggregations of blocks.[19] In practice, our identification is achieved by comparing blocks' attributes within narrowly defined neighborhoods using fixed effects to control for neighborhoods' unobservable characteristics. The key assumption is that, although immigrants from a given country of birth choose a given neighborhood to live in, they are unable to select a precise block in which to reside within the same neighborhood. Consequently, there will not be any correlation between the unobserved factors that affect the outcome (probability of becoming an American citizen) among the individuals living in the same block within the larger selected area.

Several arguments support the assumption that immigrants' sorting takes place across but not within neighborhoods (i.e., across blocks). The first argument is that, while newly arrived immigrants might have a realistic ex ante view of the characteristics of different neighborhoods (e.g., those located in Little Italy versus those in the Upper West Side), it is unlikely that they have enough information to identify the differences between these characteristics across the various blocks. This is particularly plausible in our case for three reasons. First, we purposely focus our analysis on recently arrived immigrants who, upon arrival, lack information about the blocks' characteristics. Second, our definition of social network is rather narrow, as it includes, on average, about 55 individuals; hence, knowing the characteristics of such small groups of individuals across blocks requires very detailed knowledge of the city (which, again, newly arrived immigrants arguably lack). Third, it is very unlikely that immigrants, at the time of making their residential choice, would be able to observe or have full information about the *citizenship status* of all individuals living in a neighborhood and, thus, are unlikely to be aware of how the share of naturalized co-ethnics varies across blocks within a given enumeration district.

[18] Enumeration district fixed effects would also capture all unobserved amenities of a neighborhood that might influence the propensity to stay in (or leave) the US and/or to apply for citizenship.

[19] The seminal approach of Bayer et al. (2008) has been used to identify neighborhood and network effects in different contexts; for example, to study the impact of the local concentration of foreign-born individuals on immigrants' employment outcomes (Boeri et al., 2015), to investigate the role of networks with regard to changing jobs (Schmutte, 2015), and to examine the relationship between unemployment rates and crime victimization at the neighborhood level (Hémet, 2020).

The second argument is that the housing market can be tight. Therefore, in order to be able to choose a specific block, at least one housing unit should be vacant in each block within a chosen neighborhood when the immigrant is searching for a place. Moreover, each housing unit in each block should satisfy the immigrants' preferences (for example, size or price).

Finally, enumeration districts do not follow administrative or official borders and even change over time. Immigrants do not know where the borders are and, more generally, do not even know what an enumeration district is because such a definition is only used as a sampling unit in the census.

For all these reasons, it is unlikely that immigrants, having chosen an enumeration district, can intentionally decide to live in a given block rather than in the next one. It is worth noting that while we have made the point that immigrants are unlikely to select into blocks based on naturalization, for the same reasons mentioned above, they are equally unlikely to select into blocks based on the correlates of naturalization. For example, an immigrant might choose a particular neighborhood because of the presence of work or schools, but he/she will not be able to choose a particular block – and hence a network – within that area.

These arguments support the validity of the assumption that there should be no correlation between the unobserved factors that affect the citizenship status among neighbors living in the same block within the same enumeration district. As a consequence, once we control for the sorting into a particular neighborhood selected by the individual (through enumeration district fixed effects), the remaining variability in the citizenship rates across the blocks within the neighborhood is as good as exogenous.[20]

Before moving to the results, we corroborate the validity of our identification strategy through a series of tests. Across all of them, we aim to show that there is little remaining sorting on the basis of the *observable* characteristics once we control for selection into the neighborhood. In fact, if immigrants systematically tend to reside in blocks in which the residents are similar to them in terms of a number of observable characteristics (such as income for instance) but do not pick a specific block within this neighborhood, we should not observe a strong systematic correlation in the neighbors' observable characteristics across blocks within an enumeration district (i.e., once the fixed effect $\eta_{d,1930}$ is accounted for). To validate this, we follow, in part, Bayer et al. (2008) and Hémet and Malgouyres (2018) by performing a series of statistical tests consisting of checking that the naturalization propensity of the network (block) is not strongly predictive of individual characteristics.

For each network in the sample, we randomly selected one immigrant and construct the naturalization rate of the other individuals who are a part of the same network (enumeration district–country of birth) but not of the same household.[21] In Table 3, we present evidence

[20] In a robustness check, to address the issue that families might sort into specific blocks and help their own family members to settle down, we exclude in our estimation immigrants who hold the same surnames as the main person in the household. Further, in another robustness check, we run a regression where we exclude an immigrant if he or she is either a boarder, roomer, lodger, or domestic worker. These results can be found in Table 8 in Section 4.4 below.

[21] As explained by Bayer et al. (2008), sampling only one individual per block is necessary to avoid any mechanical negative correlation that arises because each individual serves as a neighbor for all others in the same network.

supporting the argument that the citizenship status of the network is not a factor that determines the location decision of immigrants. We show that the naturalization rate in the network $ynet_{dbc,1930}$ is a poor predictor of any of the observable characteristics of the (randomly selected) immigrants. In practice, if the network's naturalization rates were a strong determinant of sorting across blocks, one would expect $ynet_{dbc,1930}$ to be a strong predictor of the individual's characteristics even when accounting for the enumeration district fixed effects. We perform these regressions with and without the enumeration district fixed effects. All regressions are weighted by the number of network members. In Table 3, we report the coefficient from the regression of $ynet_{dbc,1930}$ on the relevant observable characteristics, the p-value associated with it, and the R^2 from such a regression.

Table 3: Testing for the endogenous sorting at the block level

	Unconditional			Within ED		
Characteristic	Coeff.	P-value	R^2	Coeff.	P-value	R^2
Female	-0.004	0.850	0.000	-0.011	0.609	0.000
Married	0.049	0.039	0.000	0.005	0.831	0.000
Log Occupational Income Score	-0.069	0.320	0.000	-0.040	0.572	0.000
Household Size	-0.615	0.000	0.001	0.033	0.838	0.000
Number of Children in Household	0.253	0.000	0.002	0.061	0.262	0.000
Number of Siblings	0.013	0.762	0.000	-0.041	0.353	0.000
Speaks English at Home	0.119	0.000	0.003	-0.013	0.110	0.000
Rentinng	-0.115	0.000	0.007	-0.019	0.142	0.000
Value of the Rent	-32.787	0.030	0.001	-16.779	0.246	0.000
Value of the House	-16.032	0.001	0.014	-3.774	0.716	0.001

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses. The number of observations is 17,501, except for the regressions of "Is Renting" (16,634). "Value of the rent" (13,451) and "Value of the house" (768). All regressions are weighted as per the number of members in the network (see text for details).
We present the coefficient, p-value, and R^2 from a regression of the relevant characteristics of the first column on to the characteristics of its network members. In Column 1–3, we show the raw correlation. In Column 4–6, we include the birthplace and enumeration-district fixed effects.

The first three columns show the results when we do not include fixed effects, while in the remaining columns we include enumeration district fixed effects as well as country-of-birth fixed effects. We start by looking at the relationship between the key socio-economic outcomes and the share of naturalized co-nationals in the neighborhood. If the immigrants were able to choose their residence based on the naturalization rates and, for instance, their income level, we would expect to find a correlation between these two variables. In the OLS regressions without fixed effects, we find that the naturalization rates and that majority of immigrant characteristics co-vary, with the correlations being statistically significant, albeit not explaining much of the overall variation in the dependent variable. Remarkably, all these correlations disappear when we introduce fixed effects, supporting the argument that controlling for unobservable characteristics of the neighborhood and of the country of birth annihilates the issue of immigrants' spatial sorting. Specifically, the naturalization rate of a neighborhood is a "weak predictor" of the characteristics of the immigrant, both in terms of statistical significance and size. When we control for fixed effects all the estimated coefficients are smaller in size when compared to the OLS, and none of them is statistically significant.

Even though Table 3 provides solid evidence that our approach overcomes spatial sorting

issues, in Section 4.4 we provide two further tests to support the credibility of our identification strategy. First, we estimated placebo regressions through randomization inference by reassign at random the share of naturalized co-ethnics. Second, we re-define the network with even finer granularity and calculate the fraction of naturalized co-ethnics at the block-street level, allowing us to include a full set of block fixed effects in the regressions. This specification addresses any residual potential concern about unobserved *block* characteristics that correlate with the residential choices. We now proceed with describing the main results of our analysis.

4 Results

4.1 Main results

Table 4 reports regression estimates for Equation (2). Column (1) shows the bivariate correlation, whereas birthplace and enumeration district fixed effects are added to the specification in Column (2) and maintained throughout the table. Individual controls are added in Column (3), while in Column (4) we include network characteristics. Column (5), our preferred specification, has all the fixed effects and control variables. Throughout the tables, we also report the standardized coefficient for the share of naturalized co-ethnics, in order to facilitate the comparison of the magnitude of the effect across various specifications and tests.

Starting from Column (1), we find a positive and significant effect of the share of naturalized co-ethnics on the immigrant's likelihood of to acquire citizenship in 1940. A 10% increase in the neighborhood's naturalization rate in 1930 is associated with a 1.13% increase in the propensity to naturalize. Yet, this specification does not control for the unobservable neighborhood characteristics and, thus, for the potential spatial sorting of newly arrived immigrants. Therefore, the estimated coefficient of interest is likely biased due to the non-random sorting of individuals within NYC. Immigrants who are more willing to assimilate or who are more educated might sort into neighborhoods populated by culturally integrated and more economically successful co-ethnics. This issue is directly tackled in the specification in Column (2), where – by introducing enumeration district fixed effects – we exploit the quasi-random variation in the share of naturalized co-ethnics within the neighborhood. The inclusion of fixed effects partly reduces the coefficient of the network variable: A 10% increase in the neighborhood's naturalization rate is now associated with a 0.48% increase in the probability that the immigrants will naturalize. The comparison of the estimates from Columns (1) and (2) suggests that a part of the correlation found in Column (1) is driven by the sorting of immigrants across neighborhoods. Yet, even after controlling for this sorting, we find that the network coefficient remains significant, positive, and economically relevant.

Overall, the size and significance of our key estimate is largely unaffected by the inclusion of additional covariates in Columns (3) to (5). The estimated coefficient of the share of naturalized co-ethnics is 0.043, somewhat smaller than Column (2) but still statistically significant at the 1% level. Other estimates suggest that being female and being married are negatively related to the probability of naturalizing, while we find a positive association for income. Additional network variables appear to be only weakly related to the outcome.

Table 4: The Effect of Networks on Political Assimilation

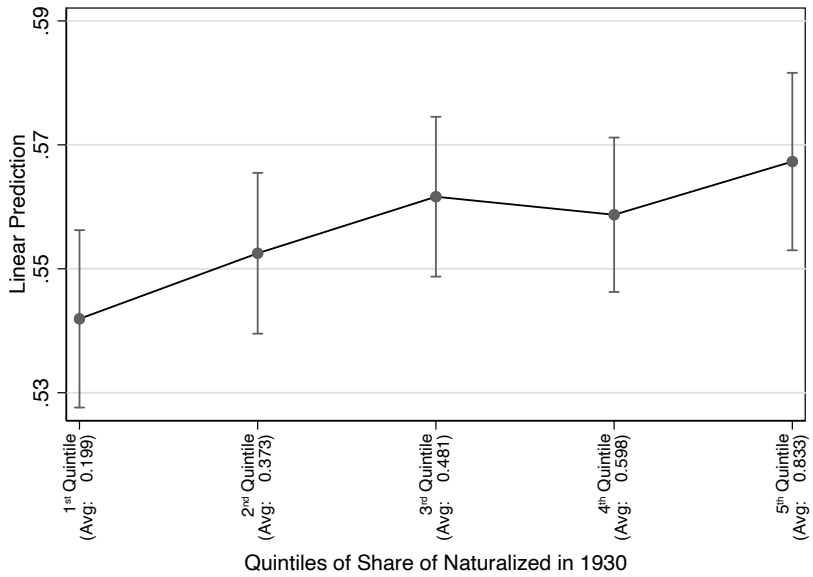
	(1)	(2)	(3)	(4)	(5)
Share Naturalized in 1930	.113***	.048***	.039**	.051***	.043***
	(.011)	(.016)	(.016)	(.016)	(.016)
	[.051]	[.021]	[.017]	[.023]	[.019]
Female			–.197***		–.198***
			(.006)		(.006)
Married			–.021***		–.020***
			(.007)		(.007)
Log Occupational Income Score			.012***		.012***
			(.002)		(.002)
Household Size			–.000		–.000
			(.001)		(.001)
Share of Females in Network				.013	.053**
				(.025)	(.024)
Share Married in Network				.010	–.008
				(.020)	(.020)
Average Log Occupational Income Score in Network				.004	.010
				(.008)	(.007)
Average Household Size in Network				.002	.001
				(.002)	(.002)
ED f.e.	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes
N	37,761	37,761	37,761	37,761	37,761
R^2	.00	.14	.19	.15	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant in 1940 is a U.S. citizen. *Share Naturalized in 1930* measures the share of co-nationals who are naturalized and live in the same block of the migrant in 1930.
All control variables are measured in 1930.
The network variables are calculated within the network, e.g., the *Share Married in Network* measures the share of co-nationals who are married in 1930 and live in the same block of the migrant.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

4.2 Non-linear effects

We now examine whether the impact of ethnic networks varies with the intensity of exposure. We construct quintiles of ethnic exposure and examine whether the impact of the share of naturalized co-ethnics impacts non-linearly the propensity to naturalize in 1940. We plot the predicted probabilities of regression that includes the quintiles of the share of naturalized co-ethnics and all controls of our preferred specification in Figure 5. Table C2 in the Appendix contains the regression results of this analysis, which we conduct for specifications (2), (3) and (5) of Table 4. Albeit not monotonically, the estimated effect increases with the magnitude of the share of naturalized co-ethnics in the network. Having a large proportion of naturalized co-ethnics in 1930 (i.e., 83.3% as in the fifth quintile) is associated with an increase of 2.5 percentage points in the likelihood of the immigrant being naturalized in 1940 relative to having a low proportion of naturalized co-ethnics (i.e., 19.9% as in the first quintile), corresponding to a 14.5% increase in the average likelihood of naturalization in that quintile.

Figure 5: Predicted probability of naturalization



Notes: The plot represents coefficients and 5% confidence intervals from a regression of equation (2) using quintiles of the variable $ynet_{dbc,1930}$. The regression includes all controls of the specification in Col (5) of Table 4

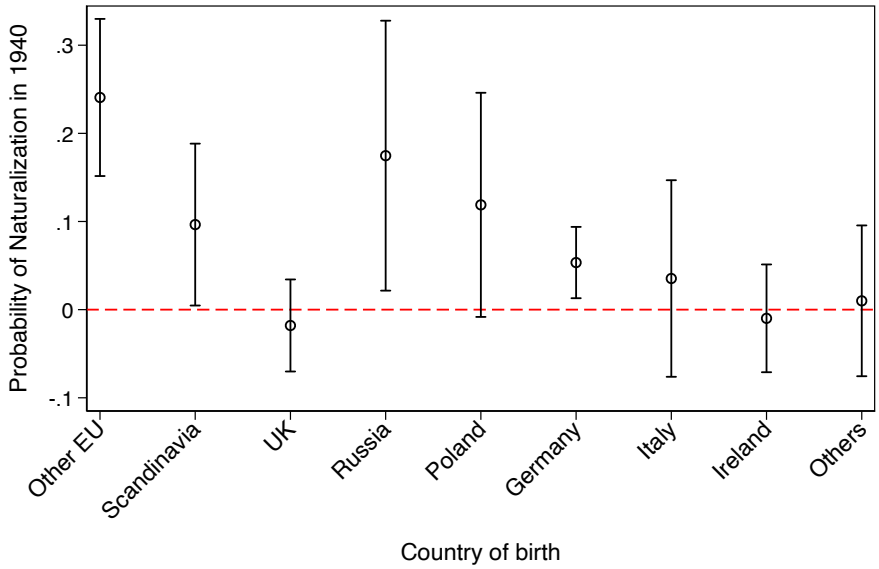
4.3 Heterogenous effects

Next, we investigate whether our results vary with the immigrants' origin. Immigrants differ substantially depending on the country they are from, for example, in terms of time of arrival, level of education, language skills, and socio-economic background. All these factors, in turn, might have an impact on the way social networks influence the propensity to acquire citizenship. To explore this angle, we estimate a model where we allow the variable $ynet_{dbc,1930}$ to have a differential effect for each country of origin. Figure 6 reports the predicted marginal effects for each country associated with the shares of the naturalized co-nationals on the likelihood of acquiring citizenship. Although, in terms of statistical significance, the effects are only marginally different across countries of birth, the magnitude of the network effect differs substantially, with a stronger effect for immigrants from Russia, Poland, Scandinavia, and other European countries. For example, a 10% increase in the share of naturalized co-ethnics raises the probability that an immigrant from Russia naturalizes by 1.75%, while the same level for exposure will result in a 1.19% increase in naturalization for an immigrant from Poland. However, for British and Irish immigrants, such effect is essentially inexistent.[22] The result that there is no network impact for immigrants from two large English-speaking countries is a remarkable one – and an angle we will explore in length in the analysis of mechanisms in Section 5.

Let us try to provide a possible explanation for why the effect of social networks on naturalization is quite different between Russians and Italians. Many Russian immigrants, who

[22] A test of joint significance for the interaction terms yields an F-value of 2.91, with a p-value of 0.003.

Figure 6: Effect by Country of Birth



Notes: We estimate a model where we interact the share of naturalized in the network $y_{dbc,1930}$ with country of birth indicators. The figure shows the estimated marginal effects and 10% confidence interval of the share of naturalized in the network on the probability of naturalization in 1940.

were Jews fleeing pogroms from Russia, were very eager to assimilate and become American citizens.[23] Their social network reinforced the eagerness to assimilate because it helped them to efficiently deal with the naturalization process, which was bureaucratic, needed a good knowledge of the English language, as well as required to have American witnesses (Figure 2 shows that 65% of Russian who applied for American citizenship had a naturalized Russian as a first witness). In the case of the Italians, there was an active movement by the Italian Catholic church that was trying to reduce the social assimilation of Italian immigrants by raising the frequency of interactions among fellow Italians (Gagliarducci and Tabellini, 2021). This possibly changed the social norm that assimilation and becoming an American citizen were not desirable, leading to the social networks of naturalized neighbors having a much lower effect on naturalization. This explanation is compatible with the data in Figure 3 that shows that naturalization rate for Italians was 50%, while for Russians was 63% in 1940.

[23]Spitzer (2018) shows that a district that experienced at least one pogrom during the period of 1903–1906 led to 10–20% more Russian immigrants arriving at Ellis Island during the years 1906–1914 as compared to a similar district that did not experience a pogrom. Soyer (1997) also documents that Russian Jews formed very strong close-knit groups in New York. In particular, he shows how a Jewish immigrant hometown associations ("landsmanshaftn") transformed old-world communal ties into vehicles for integration into American society in NYC. These associations assisted newly arrived immigrants by handling deportation cases, placing workers in jobs, providing shelters, and conducting English and citizenship classes. By 1920–1930, however, the shifting demographics and the cessation of mass migration – which reduced the pool of recent arrivals from which to draw new members – reduced the role played by home associations in favor of "family circles" (Soyer, 1997). Our evidence, which comes from the end of this "golden period" suggests that the neighborhood played an additional important role. While home associations are historically known to have played a role in immigrant assimilation (Soyer, 1997), we show that having a network of co-nationals at the neighborhood level was equally important for the assimilation of these ethnic minorities.

We subsequently explore the heterogeneity of the network effect across two additional dimensions: gender and borough of residence. Table 5 shows the results separately for male and female immigrants. Here, we estimate the impact that co-nationals who are of the same gender as the immigrant have on her naturalization propensity. In other words, our network measure is the same as in Equation (1), but constructed separately for males and females. The estimates suggest that both males' and females' citizenship acquisition decision is influenced by the presence of naturalized co-nationals of the same gender in their network, albeit the point estimate is somewhat larger for females than for males.

Table 5: The Effect of Networks on Political Assimilation – by gender

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Share Naturalized in 1930 (by gender)	.037*	.042**	.039*	.044*	.043*	.053**
	(.020)	(.020)	(.020)	(.023)	(.023)	(.024)
	[.019]	[.021]	[.019]	[.021]	[.021]	[.026]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	21,216	21,216	21,216	16,545	16,545	16,545
R^2	.19	.21	.21	.21	.22	.22

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
In column (1)-(3) the sample is restricted to males only. In column (4)-(6) the sample is restricted to females only. The dependent variable is an indicator that equals one if the migrant is a citizen in 1940. *Share Naturalized in 1930 (by gender)* measures the share of co-nationals of the migrant with the same sex, who are naturalized and live in the same block of the migrant in 1930.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

In Table 6, we explore whether the results differ based on the borough of residence. NYC boroughs differ substantially with regard to characteristics, especially in terms of immigrant settlements (Figure 4) and this could affect the importance that networks have on the process of naturalization. In the table, we split our data in three parts: Manhattan, Brooklyn and the the remaining boroughs (Bronx, Richmond and Queens). The results of our analyses show that the impact of the share of naturalized co-ethnics is relatively weaker in Manhattan, while estimates are larger and statistically significant throghout most of the specifications for the other boroughs. This result is likely driven by several factors, including population density and immigrants spatial segregation across the boroughs. For example, as Figure 4 shows, Russian immigrants – whose impact of co-ethnic networks was one of the largest – are highly concentrated in Brooklyn. On the contrary, Irish immigrants – for whom we found a zero effect – are relatively more concentrated in Manhattan.

To sum up, we find a strong positive effect of living closer to politically mobilized co-nationals on political assimilation, and such effects remain unchanged regardless of the gender or of the borough of residence of immigrants. Hence, even at the end of mass migration, political integration occurred through the use of networks and across ethnic lines.

Table 6: The Effect of Networks on Political Assimilation – across boroughs

	Manhattan			Brooklyn			Bronx, Richmond, Queens		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share Naturalized in 1930	.028	.023	.026	.058**	.045*	.043	.046**	.035	.043*
	(.036)	(.036)	(.038)	(.028)	(.027)	(.028)	(.023)	(.023)	(.023)
	[.010]	[.008]	[.009]	[.027]	[.021]	[.020]	[.022]	[.017]	[.021]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes	No	No	Yes
N	12,119	12,119	12,119	11,530	11,530	11,530	14,112	14,112	14,112
R^2	.13	.17	.17	.16	.22	.22	.13	.20	.20

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
In column (1)-(3) the sample is restricted to migrants whose county of residence in 1930 was New York City (Manhattan). In column (4)-(6) the sample is restricted to migrants whose county of residence in 1930 was Kings. In column (7)-(9) the sample is restricted to migrants whose counties of residence in 1930 were Queens, Richmond or the Bronx.
Share Naturalized in 1930 measures the share of co-nationals of the migrant who are naturalized and live in the same block of the migrant in 1930.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

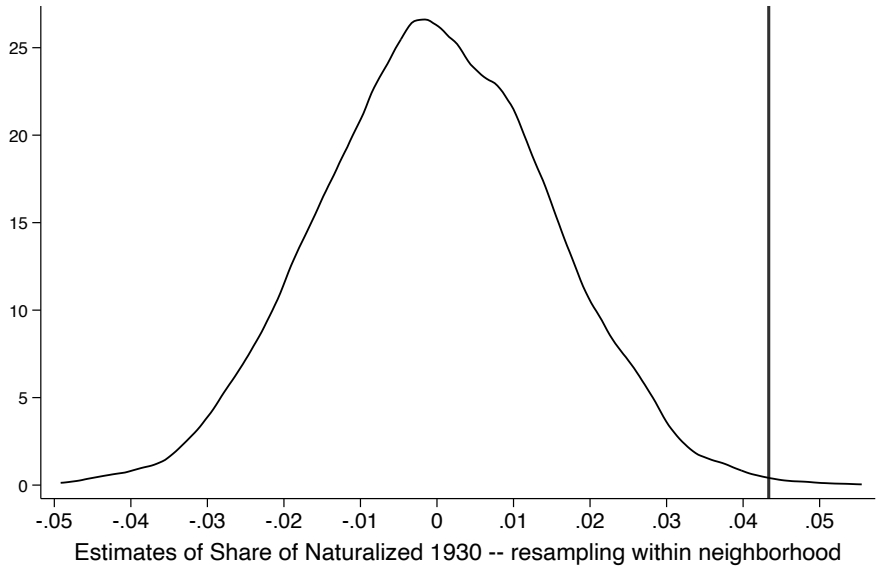
4.4 Robustness checks

In this subsection, we report robustness checks that we performed in support of our main analysis.

First, we estimate placebo tests. These are conducted through a randomization-based inference procedure. The core idea is to simulate the network effect using "fake" values of the share of naturalized co-ethnics. These values are obtained by permuting at random the share of naturalized co-ethnics within our sample. We perform two separate placebo tests, one where we do not impose any strata, i.e., permutations occur within the full sample of observations, and one where we allow the resampling to take place within each neighborhood, i.e., enumeration district. For each test, we estimate 5,000 regressions using our preferred specification, with the important difference that in each regression the share of naturalized co-ethnics is permuted. Under the null hypothesis of no effect, these simulations are expected to generate a distribution of coefficients and our aim is to compare where in this distribution the coefficient obtained from our preferred specification lies. The ultimate goal is to obtain a p-value for the randomization-based inference exercise. This is calculated by looking at how many times, out of the 5,000 simulations, we find a coefficient that is in absolute value larger than our baseline estimate. The simulation results reveal that in only 2 cases out of 5,000 (i.e., $p = 0.0004$) and 18 cases out of 5,000 (i.e., $p = 0.0036$) – respectively for the placebo without strata and for the placebo with strata – were the simulated coefficients larger than our baseline estimate. To put it in simple words, these tests show that our baseline results are "extreme/unusual" with respect to a distribution of coefficients generated by chance, providing further support that we are identifying a causal effect. Figure 7 show the results of the placebo test with randomization within the neighborhood. The graphs displayed in this figure show the distribution of the simulated coefficients and of our baseline estimate (for the

latter, see the vertical line). For reference, the placebo test for the randomization within the whole sample is reported in Figure C1 in Appendix C.

Figure 7: Placebo Tests



NOTES: The figure plots the kernel density of coefficient estimates of 5,000 regression specifications as in Col (5) of Table 4 where the share of naturalized co-ethnics is reassigned at random. Permutations are performed within each enumeration district. The vertical line represents the coefficient estimate from the specification in Col (5) of Table 4. Out of 5,000 simulations, 18 generated a coefficient greater than the baseline (0.043), implying a randomization inference p-value=18/5,000=0.0036

In the next robustness check, we propose a more demanding identification strategy, whereby we define the network of an immigrant as the group of people living within an even smaller spatial unit – the *block-street* area. We do so by leveraging the availability of full address of people in the 1930 Census, which allows us to code the street names and map them along the block or residence of immigrants. We then define our key regressor as the fraction of naturalized co-ethnics living in the *same block* and in the *same street* as the newly arrived immigrant. Individuals living in the same block-street are likely to share the same facade and entrance of the building. Using this strategy, we are able to include a full set of block fixed effects. Hence, the reference group in this regression consists of immigrants who are observed to be living not just in the same nieghborhood but also in the same block and street. The arguments in favor of this tight identification strategy are similar to before, with the exception that here we allow for the possible sorting within the block but not within the streets of the same block. In other words, this demanding specification controls for all the sorting that might still occur and that is based on block characteristics, such as the presence in a block of (extended) family and friends or closeness to a specific workplace or a school. Similarly to our argument made in Section 3, to invalidate this strategy, one would need to think that multiple housing units are available to the migrant at the time of choosing a residence among all the streets in a block. More importantly, it would require to suppose that the immigrants

are ex-ante knowledgeable of the characteristics and, particularly, of the share of the naturalized co-nationals living on all the streets of a block. Table 7 displays the results. Note that the construction of the block-street network requires the harmonization of the addresses in the Census in order to create consistent street names. This was possible for the majority of observations in our sample (35,305). However, for the remaining observations, the addresses were incomplete, missing or with a street that could not be mapped in the street list of NYC. Consequently, for completeness, we also show in the table (columns (1), (2), and (3)) the estimates of our baseline regression on the subsample of observations for which block-street networks are available.

The first remark is that, for this subsample, while the pattern of results of our benchmark analysis is confirmed in the first three columns of the Table, the estimates are somewhat smaller and slightly more noisy than those in Table 4. More importantly, though, the estimates from the block-street analysis in the last three columns show that the causal effect of networks on naturalization remains even in this highly demanding, saturated specifications. Lastly, if sorting within the block occurred, one would have expected the estimates between the two definitions of network (block level and block-street level) to differ. Instead a comparison of the coefficients of columns (3) and (6) show that the implied effects are remarkably similar, demonstrating that our main identification strategy, based on block level definition of network and on using neighborhood and country of birth fixed effects is able to effectively annihilate sorting issues.

A last set of robustness checks is presented in Table 8. There are two remaining potential concerns that cannot be addressed even with the inclusion of a rich set of fixed effects as we did in Table 7. First, an immigrant might decide to reside in a specific block (or block-street) because their relatives (i.e., beyond their own household, which are already excluded in the computation of the network) are living there and can offer a place to stay regardless of housing or rental market tightness. Second, an immigrant might decide to room or board with individuals who might not be family members but could be co-nationals or friends of co-nationals. For these to be valid selectivity concerns, social ties in a given block have to correlate with the share of the naturalized co-nationals in that block. To eliminate any doubt, we exploit the richness of our data and provide indicative evidence that these concerns are actually unfounded. To address the fact that families might sort into specific blocks and help their own family members to settle down, we identify the share of individuals outside the household and residing in the same block who have the same surname. In other words, we assume that persons who live in the same block and have the same surname are related.[24] We then include in our estimation only those immigrants who hold surnames that are unique within the network. The results are reported in columns (1)–(3) of Table 8. We find that focusing only on immigrants with unique surnames does not affect our main conclusion. To address the second concern, namely that an immigrant who arrived earlier might decide to

[24]To avoid concerns of small transliterations and errors, we use the the New York State Identification and Intelligence System Phonetic Code (NYSIIS) algorithm. With this algorithm, surnames that sound the same but are spelled differently are transliterated in a similar way. Therefore, we are able to purge possible misspelling errors made in the original record by the census enumerators.

Table 7: The Effect of Networks on Political Assimilation in 1940 – Block Fixed Effects

	Benchmark			Block-street		
	(1)	(2)	(3)	(4)	(5)	(6)
Share Naturalized in 1930	.037**	.027	.032*	.042***	.031**	.025*
	(.018)	(.018)	(.018)	(.015)	(.015)	(.015)
	[.016]	[.011]	[.013]	[.022]	[.017]	[.014]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Block f.e.	No	No	No	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	35,305	35,305	35,305	35,305	35,305	35,305
R^2	.14	.19	.19	.24	.28	.28

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant is a U.S. citizen in 1940.
Columns (1) to (3) report benchmark specifications of Table 4 estimated on the subsample for which block-street information is available.
Share Naturalized in 1930 measures the share of co-nationals of the migrant who are naturalized and live in the same block of the migrant in 1930 (Columns (1), (2) and (3)) and the share of co-nationals of the migrant who are naturalized and live in the same block and street of the migrant in 1930 (Columns (4), (5) and (6)).
For a definition of all other variables, see table notes of Table 4.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

become a boarder, roomer, lodger, or domestic worker following the suggestions of their close ties, we run a set of regressions in which we exclude these categories of immigrants. These regressions are shown in columns (4)–(9) of Table 8. Excluding these observations does not affect our main results.

5 Investigating the mechanisms

5.1 Information or social norms?

In this section, we perform additional analyses with the aim of pinpointing the mechanisms behind our results. Two possible roles for the social networks could be at play in our study. First, the co-ethnic social network may help the immigrant to acquire U.S. citizenzhip through *information dissemination*. Indeed, individuals pursuing citizenship require specialized information about the naturalization process. This information may come in two forms: (*i*) information about the benefits of citizenship and (*ii*) information about the process of acquiring it. For example, a migrant aspiring to naturalize may seek advice about filling out paperwork or studying for the citizenship examination (Abascal, 2017). Moreover, a newly arrived immigrant may have a poor command of the English language and may have difficulty in understanding the bureaucratic naturalization process. Research has shown that the poor

Table 8: The Effect of Networks on Political Assimilation – excluding groups

	Unique Surnames			Excluding Roomers, Lodgers and Boarders			Excluding Domestic Workers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share Naturalized in 1930	.041**	.032**	.038**	.042**	.032*	.038**	.040**	.038**	.042**
	(.017)	(.016)	(.017)	(.018)	(.017)	(.018)	(.017)	(.016)	(.017)
	[.019]	[.015]	[.017]	[.019]	[.014]	[.017]	[.018]	[.017]	[.018]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes	No	No	Yes
N	34,574	34,574	34,574	31,067	31,067	31,067	35,619	35,619	35,619
R^2	.15	.19	.19	.16	.21	.21	.15	.19	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant is a U.S. citizen in 1940.
Columns (1)-(3): includes in our estimation sample only those migrants who hold surnames that unique within their block.
Columns (4)-(6): excludes from the estimation all migrants who are boarders, roomers, lodgers.
Columns (7)-(9): excludes all migrants who are domestic workers.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

language skills of immigrants not only negatively affect their labor-market outcomes (Auer, 2017, Arendt et al., 2020) but also the health of their new-born child (Auer and Kunz, 2021). Indeed, newly arrived immigrants do not understand clearly or immediately the way the labor market or the health system operates in the new country and, often, communication barriers might determine immigrants' slow assimilation in the host society. This might be particularly true during the 1930s in the U.S., especially for immigrants coming from non-English speaking countries. As highlighted in Section 2, another reason why information dissemination might have been a key factor can be explained by the naturalization process itself. Indeed, to petition an admission to citizenship, immigrants needed to have two American citizens to act as witnesses and could testify the immigrant's continuous residence and "good moral character." In Figure 2, we showed that most immigrants provided the names of their co-ethnic connections who had already become citizens; thus, the co-ethnic social network may have helped them find one or two witnesses.

A second potential mechanism has to do with *social norms*. Indeed, there is a large segment of the literature showing that peer pressure and social norms are important factors in assimilation, as immigrants do not want to deviate from the social norm of their ethnic peers. A society may have many social groups –"American," "Hispanic," "Italian," and so on –but, in any given situation, individuals identify with only one of these and pay a cost for deviating from the social norm of the group that she identifies with (Akerlof, 1997, Shayo, 2009, Sato and Zenou, 2020, Ushchev and Zenou, 2020). As a result, if someone is "randomly" exposed to a network with a large fraction of naturalized co-nationals, he or she may feel compelled to also naturalize in order to conform to the social norm of the group.

To disentangle the role played by information vis-à-vis social norms, we perform a heterogeneity test using the language spoken by the immigrant at the time when the network is

observed (i.e., in 1930). Our argument is that, while information acquisition through social contacts is predominantly relevant for non-English speaking immigrants – because the bureaucratic process of naturalization involves plenty of technical terms in English, paperwork, and navigating a complex bureaucratic process, which is very difficult with low English proficiency – the pressure of the social norms with regard to naturalization by co-ethnics is real for all immigrants. We explore this by splitting our sample into immigrants from non-English-speaking countries (such as Russia, Poland, Germany, etc.) and English-speaking countries (such as UK, Ireland, Canada, Australia, etc.).[25] Therefore, if naturalization social norms were the sole driver behind our results, we would expect a positive effect of the networks on naturalization to be present for both groups.

The regression results for the two groups are reported in the first panel of Table 9. Once sorting is accounted for (see columns (3) and (6)), we see that networks positively impact the naturalization propensity of immigrants from *non-English-speaking countries only*. This result, therefore, suggests that the *acquisition of information* regarding the naturalization process through close contacts was more important than social norms for immigrants who faced stronger barriers. In other words, immigrants who faced language and cultural obstacles – such as Russians – were likely to be helped by other Russians who had arrived before them and who were already naturalized in 1930 and thus could provide them with information about the naturalization process. As discussed in Section 2, the naturalization process is complicated because it involves two steps – a declaration of intention and a petition for citizenship – with several residency requirements to be met and documents to be produced. Consequently, it could be too bureaucratic and complicated for newly arrived immigrants without a social network, especially for those coming from non-English-speaking countries.

To further substantiate the argument that the network primarily served as a way to acquire information, we show two additional tests in the second panel of Table 9. In columns (1)–(3), we start by relating the probability of naturalization in 1940 with the share of the American-born individuals living in the same block as the immigrant in 1930. In practice, Americans could act as witnesses in the naturalization process and, therefore, provide naturalization referrals; if this was the case, we would find a positive and significant impact of the share of Americans living on the block on the likelihood of naturalization. On the other hand, Americans did not have to undergo the process of naturalization and are unlikely to have had any knowledge on it; therefore, if information was a driving mechanism of our main finding, we would find that the probability that an immigrant naturalizes by 1940 should be unrelated to the share of American-born individuals living in her block in 1930. Our results confirm this conjecture, namely Americans play no role in determining citizenship acquisition. This finding is compatible with the idea that the network provides useful information about the naturalization process because naturalized immigrants have already experienced the process, while American-born individuals have not. In addition, this test also excludes the fact that the network works only as an incentive to assimilate to an American social norm, since, in

[25] Observe that the naturalization process is exactly the same for English- and non-English-speaking individuals.

this case, we would have expected that immigrants surrounded by American-born individuals might face stronger pressure to Americanize.

Table 9: The Effect of Networks on Political Assimilation – Information Channel

	non-English Speakers			English Speakers		
	(1)	(2)	(3)	(4)	(5)	(6)
Share Naturalized in 1930	.094***	.088***	.091***	–.050	–.062*	–.041
	(.022)	(.021)	(.022)	(.033)	(.032)	(.034)
	[.040]	[.038]	[.039]	[–.023]	[–.028]	[–.019]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	26,103	26,103	26,103	11,658	11,658	11,658
R^2	.16	.22	.22	.25	.30	.30
	Network: Americans			Network: non co-Nationals		
Share Americans in 1930	–.052	–.035	–.040			
	(.067)	(.066)	(.066)			
	[–.012]	[–.008]	[–.009]			
Share Naturalized in 1930, from Other Countries				–.024	–.032	–.031
				(.034)	(.033)	(.033)
				[–.007]	[–.009]	[–.009]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	37,761	37,761	37,761	37,686	37,686	37,686
R^2	.14	.19	.19	.14	.19	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
In both panels, the dependent variable is an indicator that equals one if the migrant is a U.S. citizen in 1940.
English-speaking countries include Canada, the UK, Ireland, Australia and New Zealand. *Share Naturalized in 1930* measures the share of naturalized co-nationals living in the same block of the migrant in 1930. *Share Americans in 1930* measures the share of American-born individuals living in the same block of the migrant in 1930. *Share Naturalized 1930 from Other Countries* measure the share of naturalized individuals in 1930 from all countries except the country of birth of the migrant, who live in the same block of the migrant.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

As a last test to substantiate our argument that co-ethnic networks provide information, we show in the second panel of Table 9, columns (4)–(6), that the naturalization probability in 1940 is unrelated to the share of immigrants who are from other countries and, thus, are not co-ethnics, although they are citizens and live in the same block as newly arrived immigrants. In other words, as an example, we show that the share of naturalized non-Russian immigrants have no effect on the naturalization propensity of recent immigrants from Russia.

To conclude, the results in Table 9 suggest that information was a key mechanism through which political assimilation occurred. In particular, the fact that only non-English speakers are affected by co-ethnic networks is an indication that information, rather than social norms, is at work. In addition, the fact that American-born individuals as well as immigrants from other countries have no impact suggest that it is a two-step mechanism: First, a newly arrived immigrant is more likely to socially interact with immigrants from the same country; second, these interactions are particularly important for future naturalization when they occur between immigrants from non-English-speaking countries because the naturalized ones can

explain and provide useful information about how the process works.

5.2 Naturalization as a stepping-stone for integration

Having found suggestive evidence that information is the key mechanism behind our results, we deepen the analysis of mechanisms by showing *which type* of information matters. Our conjecture is that naturalization could have been helpful to immigrants who either faced stronger barriers to integration or had greater incentives to integrate. Hence, in the following, we perform additional tests to characterize features of the network that were helpful to provide essential information on the benefits of naturalization, and to detect the groups of immigrants for which information might have played a more essential role in becoming an American citizen.

5.2.1 To Circumvent Labor Market Barriers

First, we hypothesize that acquisition of information could have been important for overcoming occupational barriers in the labor market. Research has demonstrated the existence of a "citizenship premium" in the labor market during the age of mass migration. For example Catron (2019) found that immigrants who acquired citizenship experienced better occupational outcomes than observationally equivalent non-citizen immigrants. Often barriers were also imposed by law; for example, there were occupations that were legally accessible only to citizens. Fields (1933) provides a list of such occupations, which included not only professional occupations, as one might expect, but also jobs such as chauffeurs, steam boiler operators, hunters, or bank directors. If networks provide information on possible labor market opportunities that are reserved for citizens, one would expect to find stronger naturalization impacts for those immigrants who, in 1930, were surrounded by naturalized neighbors who work on such difficult-to-access occupations and were aware of such benefits.

To operationalize this hypothesis, we conduct two type of tests. In the first one, we use the list provided by Fields (1933) to derive two network variables: the share of immigrants who are naturalized and work in barred occupations and the share of immigrants who are naturalized and work in non-barred occupations. Based on our hypothesis, we would expect that only the network of naturalized co-nationals in barred occupations would be producing an impact on naturalization, since these are the contacts who are more likely to inform about the benefits related to these occupations. The first three columns of the top panel of Table 10 display the results of the regression using these two measures of social networks. The results confirm that the impact on naturalization is driven by the share of immigrants who are naturalized and work in barred occupations.

In the second test, we expand the definition of difficult-to-access occupations, and define them based on the existence of observed occupational gaps between citizen and non-citizen immigrants. From the 1930 Census, we selected the full sample of immigrants living in New York City (2,406,549 observations) and calculated "conditional occupational gaps" by broadly following the approach of Hsieh et al. (2019) and Ortega and Hsin (2018). We modelled the probability of being in a given occupation as function of a rich set of covariates and include

an indicator for whether the immigrant was a citizen or not.[26] The coefficient of this dummy variable provides a measure of the existence of an "entry barrier" between two observationally equivalent immigrants, one naturalized and one not naturalized. We estimated such regression model for each of the 255 occupations. We subsequently computed the predicted values for each individual in a given occupation and, based on these, calculated the occupational shares, i.e., the percentage of individuals in each occupations – separately for naturalized immigrants and non-naturalized immigrants. Occupations with a relatively higher number of naturalized immigrants were classified as "high barrier occupations" and the remaining as "low barrier occupations".[27] As we did before, we derived two network variables: the share of immigrants who are naturalized and work in high barrier occupations and the share of immigrants who are naturalized and work in low-barrier occupations and use them to estimate our baseline regression models. Results are presented in the last three columns of the top panel of Table 10. Using this alternative classification for identifying difficult-to-access occupation produces a pattern of results that is remarkably similar to the one seen in first three columns. That is, the share of naturalized immigrants and in high barrier occupations positively impact naturalization rates, while the network of naturalized immigrants working in low barrier occupations exhibits a smaller, statistically insignificant coefficient. In summary, the results from the top panel of Table 10 are consistent with the hypothesis that the social network has provided concrete information and help about occupations for which immigrants might have experienced entry barriers.

5.2.2 To Overcome and Leverage Changes in Immigration Policy

A second hypothesis that we explore is that information might matter more for immigrants affected by recent changes in immigration policy introduced by the Immigration Act of 1924. This policy introduced restrictions and imposed quotas for migrants from certain country of origins.[28] By changing the quota allocations by country, the Act lengthened the stay of immigrants (Greenwood and Ward, 2015) and, consequently, might have impacted their probability to naturalize. Furthermore, the Act introduced also some provisions that might have increased the incentives to naturalize. For example, after 1924, relatives of citizens could enter the U.S. as non-quota immigrants. If networks provide useful information on naturalization, one would expect that the impact of the network would be more important for recent migrants coming from countries affected by the restrictions of the 1924 Immigration Act, since these migrants have stronger incentives to naturalize.

[26] The other covariate included in the regression are: gender, age and age squared, years in the US and years in the US squared, marital status, number of children, family size, race dummies, level of spoken English and of literacy, dummies for each borough and dummies for each broad country of origin (the same groups used to define our networks).

[27] The set of occupations for which non-citizen faced barriers includes nearly all barred occupations identified by Fields (1933) – the only exception being "Hucksters and peddlers" – and a number of other occupations, that in general, were relatively up high in the wage ladder. The average occupational score in barred occupations was 50% higher than in the remaining occupations (30.74 vs 20.47).

[28] Indeed, as the quota laws were intended to do, they greatly restricted immigration from the "new" source countries (such as Italy, Spain, Poland, and Greece), while having a limited influence on immigration from the "old" source countries (such as Germany, England, Ireland, and Sweden).

To operationalize our test, we construct a measure of restrictiveness for the quotas as in Greenwood and Ward (2015). For each country of birth, we take the difference between the number of admitted immigrants in year t and the quota limit for that year. The ratio between the admitted immigrants over the quota provides a proxy for the tightness of the quotas. Next, we split the sample between the immigrants from the countries whose quotas' tightness is below the top 75th percentile of the distribution and immigrants in the top quartile of the distribution. This cutoff produces two subsamples of similar size, one related to immigrants from countries facing "low restrictions" and one from countries experiencing "high restrictions". We report the results of the two regressions in the second panel of Table 10. The point estimates suggest that networks had a larger impact on the naturalization rates for immigrants who were more severely restricted by the quotas.

5.2.3 To be Active Politically

The last dimension we consider is political participation. Considering that the right to vote is embedded in the acquisition of citizenship, we hypothesize that social networks might provide information regarding the importance of political participation. This might have occurred especially in those areas where, in the previous congressional election, Democratic representatives – who are traditionally more pro-migration than their Republican counterparts – have been elected. To perform our test, we first merge the information on electoral districts with our data and then perform the analysis separately for the areas that elected a Republican and those that elected a Democrat (based on the 1930 elections). We report the results in the last panel of Table 10.[29] As expected, we find that networks have a stronger and statistically significant impact only where a Democrat was elected.

To summarize, we have shown that information is particularly important for groups for which barriers or opportunities were more apparent. Our results suggest that the naturalized co-ethnics might have provided useful information to overcome labor-market barriers and migration restrictions. At the same time, the naturalized co-ethnics might have also shown the positive effects that naturalization could have brought at the political level, leading to the election of a Democratic representative to the Congress.

5.3 Other Outcomes

To conclude our analysis, we explore whether living in a network with naturalized co-ethnics might have affected other outcomes – besides naturalization – in the long run. To do so, we estimate regression models where the dependent variables include labor-market outcomes, educational attainment, migration and housing – all measured in 1940. It is important to note, that the aim of this analysis is purely descriptive, since we do not have a particular conjecture on how these further outcomes are linked to naturalization. For example, some of these outcomes could be contextual to naturalization or a consequence of it. But they could also be mediating the impact of the network on naturalization. With this limitation

[29] Note that these regressions are performed on a subset of observations for which we could match the 1930 election results.

Table 10: The Effect of Networks on Political Assimilation – Mechanisms

	To Circumvent Labor Market Barriers					
	Using Fields (1933)			Conditional Occupational Gaps		
Share naturalized & in High Barrier Occupations in 1930	.091**	.091**	.096***	.047***	.036**	.039**
	(.037)	(.036)	(.036)	(.017)	(.016)	(.017)
	[.015]	[.015]	[.016]	[.020]	[.015]	[.016]
Share naturalized & in Low Barrier Occupations in 1930	.023	.017	.019	.003	.006	.008
	(.015)	(.014)	(.015)	(.019)	(.018)	(.019)
	[.011]	[.008]	[.009]	[.001]	[.002]	[.003]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	36,850	36,850	36,850	36,850	36,850	36,850
R^2	.14	.19	.19	.14	.19	.19
	To Overcome and Leverage Changes in Immigration Policy					
	Low Restrictions			High Restrictions		
Share Naturalized in 1930	–.001	–.018	–.013	.088***	.085***	.092***
	(.023)	(.023)	(.023)	(.027)	(.026)	(.027)
	[–.001]	[–.008]	[–.006]	[.037]	[.036]	[.039]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	18,825	18,825	18,825	18,724	18,724	18,724
R^2	.21	.25	.25	.20	.25	.25
	To be Active Politically					
	Elected a Republican			Elected a Democrat		
Share Naturalized in 1930	–.016	–.009	–.005	.053**	.043*	.043*
	(.074)	(.074)	(.078)	(.023)	(.022)	(.023)
	[–.006]	[–.003]	[–.002]	[.022]	[.018]	[.018]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Network Controls	No	No	Yes	No	No	Yes
N	2,819	2,819	2,819	20,823	20,823	20,823
R^2	.14	.18	.18	.15	.19	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
Share naturalized & in High Barrier Occupations in 1930 Low share in barred occupations represents the share of immigrants who are naturalized and work in one of the occupations classified as barred (according to Fields (1933) or as high barrier (according to the conditional occupational gap definition.)
High restrictions: immigrants from countries whose bite of the quota is above the 75th percentile. Low restrictions: immigrants fom countris whose bite of the quota is below the 75th percentile. See text for explanation.
Elected a Republican is a subsample of migrants living in enumeration districts that elected a Republican to the congress in 1930. Elected a Democrat is a subsample if migrants living in enumeration districts that elected a Democrat to the congress in 1930.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

in mind, it is still relevant to explore these outcomes, since this could shed further light on the understanding of how naturalization might have contributed to the pathway of economic, political and social assimilation of immigrants and the role of the social network in these outcomes. We present the results of this further analysis in Table 11.

The results show that the share of naturalized co-ethnics positively correlate with some of the outcomes. While there is no correlation with income and unemployment, the network of naturalized immigrants is positively correlated with the probability of being in one of the

Table 11: The Effect of Networks on Political Assimilation – Other Outcomes

	Occupat. Score	Barred Occup.	Unempl-oyment	Changed County	Household Education	Owning a House
Share Naturalized in 1930	–.035	.019***	.004	.033**	.254***	.010
	(.035)	(.006)	(.008)	(.016)	(.083)	(.013)
	[–.005]	[.030]	[.004]	[.015]	[.021]	[.006]
ED/BPL f.e.	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Network Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	37761	37761	37761	37211	37761	37761
R^2	.59	.16	.14	.20	.29	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
All outcomes are measured in 1940. *Occupat. Score* is the log of Occupational Income Score. *Barred Occup.* refers to the probabilty of being in one of the occupations listed by Fields (1933). *Changed County* is the probability of living in a different county than in 1930 and is constructed comparing answers about county of residence in the 1930 and 1940 Censuses. *Household Education* is the average highest grade in the household.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

barred occupations listed by Fields (1933).[30] There is also a positive association also with internal mobility (as measured by the probability of changing county of residence between 1930 and 1940) and education (as measured by the average educational attainment in the household). At the same time, there is no association with the probability of owning a house.

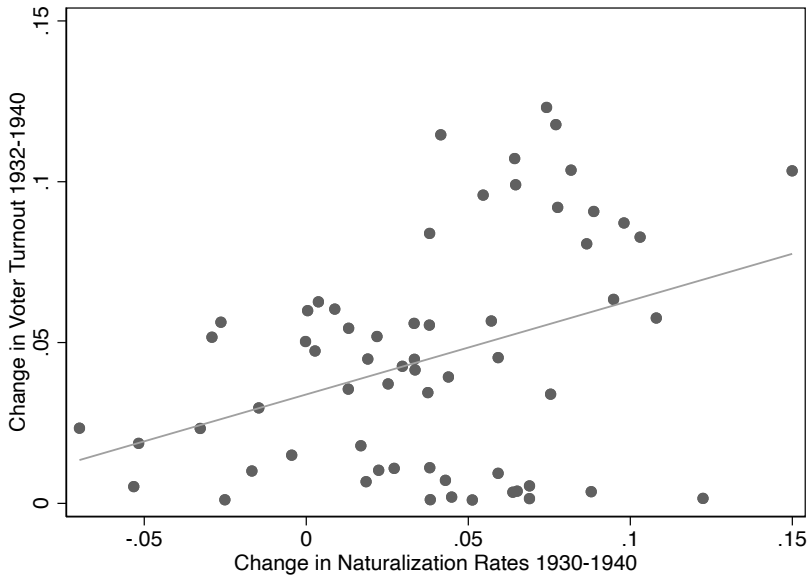
An additional key outcome worth exploring is political participation. Unfortunately the 1940 Census does not contain this information. To provide some insight on whether the share of naturalized co-ethnics relates to political mobilization in 1940, we collected data on voter turnout for all NYC counties for the two Presidential elections in 1932 and 1940. Our aim is to construct a measure of change in voter turnout and correlate it with the change in naturalization rates in each county between 1930 and 1940. To obtain the latter, we use the complete count Censuses of 1930 and 1940. The scatterplot in Figure 8 clearly shows the existence of a positive relationships between the (change) in the relative size of the naturalized immigrant community and the (change) in the propensity to vote (the R^2 of the related bivariate regression is 0.13). Remarkably, this association is still large and statistically significant when we estimate regression models were we include, besides the share of naturalized immigrants, also the (changes in) average age, share of non-white population, unemployment rates, and occupational income score.

6 Conclusion

How do (characteristics of) ethnic enclaves shape behavior? This is the question we asked at the beginning of this paper, which is of paramount importance for policies. Here, we measure the behavior of immigrants by their naturalization and the characteristics of ethnic enclaves by

[30] Remarkably this result persists also when we add as control variable the probability of being in such occupations in 1930.

Figure 8: Voter Turnout



Notes: Each observation represents a county in New York State. The Y-Axis represents the difference between voter turnout in 1940 and 1932. Voter turnout is calculated as the number of voters in the Presidential Elections over the total number of people living in the county. The X-Axis represents the difference in naturalization rates in each county between 1940 and 1930.

co-ethnic networks. The answer is not trivial because it is possible that immigrants are more likely to naturalize when they live among co-ethnics; but the reverse is also true. Indeed, on the one hand, immigrants who reside among naturalized co-ethnics may have greater access to information about the benefits and procedures of naturalizing. On the other hand, coethnic communities may reinforce in-group solidarity by limiting contact with mainstream society. This is referred to as the *ethnic enclosure hypothesis* (Liang, 1994), according to which ethnic communities impede incorporation by limiting contact with mainstream society.[31] With respect to naturalization, the enclosure hypothesis would predict that immigrants who are insulated from mainstream society are "less likely to use citizenship acquisition as a strategy for the purpose of self-protection and less likely to identify themselves as Americans" (Yang, 1994). This has led to many controversial debates, in particular from immigration critics such as Huntington (2004), who postulates that geographic concentration "retards other forms of assimilation."

In this paper, we show that residing with naturalized co-ethnics has a *positive* impact by increasing the individual chance of becoming a U.S. citizen, and this relationship operates through one main channel: information dissemination. Indeed, individuals pursuing citizenship require information about the benefits of citizenship and about the process of acquiring it. This is a complicated process, especially for non-English speakers.

We believe that this result provides some answers to the policy debate about the impact of

[31] This hypothesis is closely related to the spatial assimilation perspective, which views residential integration as both a cause and a consequence of cultural and economic assimilation (Massey and Denton, 1985).

ethnic enclaves on the integration and assimilation of immigrants. Prior research has argued that the period after arrival represents an "integration window," in which immigrants may be open to habit change (Hainmueller et al., 2015, 2017, Ferwerda et al., 2020). Our results indicate that the years after arrival may be critical and that, during this adjustment period, social interactions with co-ethnic neighbors may be of paramount importance for assimilation. In terms of policy implications, our findings suggest that, when immigrants arrive in a new country, mixing and interacting with other immigrants from the same country may help their long-term assimilation only if these co-ethnics are themselves well assimilated in the host country. In other words, assimilation can be achieved by targeting interventions and optimizing the geographic placement of newly arrived immigrants.

The debate on ethnic enclaves is complex, and we are aware that our results are specific to the studied time period and city, which is between 1930 and 1940 in NYC. We hope to see more research in the future on these important issues with more recent data that examines the causal impact of the composition of ethnic enclaves on the assimilation of newly arrived immigrants.

References

Abascal, M. (2017). Tu casa, mi casa: Naturalization and belonging among Latino immigrants. *International Migration Review 51*(2), 291–322.

Abramitzky, R., P. Ager, L. Boustan, E. Cohen, and C. Hansen (2021). The effects of immigration on the economy: Lessons from the 1920s border closure. *American Economic Journal: Applied Economics,* forthcoming.

Abramitzky, R. and L. Boustan (2017). Immigration in American economic history. *Journal of Economic Literature 55*(4), 1311–1345.

Abramitzky, R., L. Boustan, and D. Connor (2020). Leaving the enclave: Historical evidence on immigrant mobility from the industrial removal office. NBER Working Paper No. 27372.

Abramitzky, R., L. Boustan, and K. Eriksson (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review 102*(5), 1832–1856.

Abramitzky, R., L. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122*(3), 467–506.

Abramitzky, R., L. Boustan, and K. Eriksson (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights 2*(1), 125–141.

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Perez (2021). Automated linking of historical data. *Journal of Economic Literature 59*(3), 865–918.

Akerlof, G. (1997). Social distance and social decisions. *Econometrica 65*, 1005–1027.

Algan, Y., N. Dalvit, Q.-A. Do, A. Le Chapelain, and Y. Zenou (2020). Friendship networks and political opinions: A natural experiment among future french politicians. IZA Discussion Paper No. 14005.

Amuedo-Dorantes, C. and M. Lopez (2021). Recent changes in immigration policy and us naturalization patterns. *Review of Economics of the Household 19*(3), 843–872.

Arendt, J. N., I. Bolvig, M. Foged, L. Hasager, and G. Peri (2020). Language training and refugees' integration. NBER Working Paper No. 26834.

Auer, D. (2017). Language roulette – the effect of random placement on refugees' labour market integration. *Journal of Ethnic and Migration Studies 44*(3), 341–362.

Auer, D. and J. S. Kunz (2021). Communication barriers and infant health: Intergenerational effects of randomly allocating refugees across language regions. *GLO Discussion Paper N. 867*.

Badev, A. (2021). Nash equilibria on (un)stable networks. *Econometrica 89*(3), 1179–1206.

Banerjee, A., E. Breza, A. Chandrasekhar, E. Duflo, M. Jackson, and C. Kinnan (2021). Changes in social network structure in response to exposure to formal credit markets. Unpublished manuscript, MIT.

Banulescu-Bogdan, N. (2012). Shaping citizenship policies to strengthen immigrant integration. Migration Policy Institute paper, Washington.

Battisti, M., G. Peri, and A. Romiti (2021). Dynamic effects of co-ethnic networks on immigrants' economic success. *Economic Journal,* forthcoming.

Bayer, P., S. L. Ross, and G. Topa (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy 116*(6), 1150–1196.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US. *Review of Economic Studies 79*(1), 128–161.

Beenstock, M., B. Chiswick, and A. Paltiel (2010). Testing the immigrant assimilation hypothesis with longitudinal data. *Review of Economics of the Household 8*, 7–27.

Biavaschi, C., C. Giulietti, and Z. Siddique (2017). The economic payoff of name Americanization. *Journal of Labor Economics 35*(4), 1089–1116.

Boeri, T., M. De Philippis, E. Patacchini, and M. Pellizzari (2015). Immigration, housing discrimination and employment. *The Economic Journal 125*(586), F82–F114.

Borjas, G. J. (1985). Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics 3*(4), 463–489.

Borjas, G. J. (1992). Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics 107*, 123–150.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human capital externalities. *American Economic Review 85*(3), 365–390.

Boucher, V. (2016). Conformism and self-selection in social networks. *Journal of Public Economics 136*, 30–44.

Boucher, V., S. Tumen, M. Vlassopoulos, J. Wahba, and Y. Zenou (2020). Ethnic mixing in early childhood. CEPR Discussion Paper No. 15528.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41 – 55.

Bramoullé, Y., B. Rogers, and A. Galeotti (2016). *The Oxford Handbook of the Economics of Networks*. Oxford: Oxford University Press.

Bratsberg, B., J. Ferwerda, H. Finseraas, and A. Kotsadam (2021). How settlement locations and local networks influence immigrant political integration. *American Journal of Political Science 65*(3), 551–565.

Brell, C., C. Dustmann, and I. Preston (2020). The labor market integration of refugee migrants in high-income countries. *Journal of Economic Perspectives 34*(1), 94–121.

Burgess, S., F. Andersson, and J. Lane (2014). Do as the neighbors do: Examining the effect of residential neighborhoods on labor market outcomes. *Journal of Labor Research 35*, 373–392.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Catron, P. (2019). The citizenship advantage: Immigrant socioeconomic attainment in the age of mass migration. *American Journal of Sociology 124*(4), 999–1042.

Chiswick, B. and P. Miller (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics 13*, 246–288.

Chiswick, B. and P. Miller (2011). The "negative" assimilation of immigrants: A special case. *Industrial and Labor Relations Review 64*, 502–525.

Comola, M. and S. Prina (2021). Treatment effect accounting for network changes. *Review of Economics and Statistics 103*(3), 597–604.

Connor, D. (2018). The cream of the crop? Geography, networks, and irish migrant selection in the age of mass migration. *Journal of Economic History 79*, 139–175.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Cutler, D. M. and E. L. Glaeser (1997). Are ghettos good or bad? *Quarterly Journal of Economics 112*, 827–872.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (2008). When are ghettos bad? Lessons from immigrant segregation in the United States. *Journal of Urban Economics 63*(3), 759–774.

Damm, A. P. (2009). Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence. *Journal of Labor Economics 27*(2), 281–314.

De Paula, A. (2020). Econometric models of network formation. *Annual Review of Economics 12*, 775–799.

DeSipio, L. (2001). Building America, one person at a time: Naturalization and political behavior of the naturalized in contemporary American politics. In: G. Gerstle and J. Mollenkopf (Eds.), *E Pluribus Unum? Contemporary and Historical Perspectives on Immigrant Political Incorporation,* New York: Russell Sage Foundation, 67–106.

Dustmann, C. and F. Fabbri (2003). Language proficiency and labour market performance of immigrants in the UK. *Economic Journal 113*, 695–717.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016). Referral-based job search networks. *Review of Economic Studies 83*(2), 514–546.

Edin, P.-A., P. Fredriksson, and Aslund (2003). Ethnic enclaves and the economic success of immigrants–Evidence from a natural experiment. *Quarterly Journal of Economics 118*, 329–357.

Eriksson, K. (2020). Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration. *European Review of Economic History 24*(3), 427–446.

Ferrie, J. P. (1996). A new sample of males linked from the public use microdata sample of the 1850 us federal census of population to the 1860 us federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 29*(4), 141–156.

Ferwerda, J., H. Finseraas, and J. Bergh (2020). Voting rights and immigrant incorporation. *British Journal of Political Science 50*(2), 713–730.

Fields, H. (1933). Where shall the alien work? *Social Forces 12*(2), 213–221.

Gagliarducci, S. and M. Tabellini (2021). Faith and assimilation: Italian immigrants in the US. Einaudi Institute for Economics and Finance (EIEF) Working Paper 21/02.

Giuliano, P. and M. Tabellini (2020). The seeds of ideology: Historical immigration and political preferences in the United States. Harvard Business School Working Paper No. 20-118.

Goldsmith-Pinkham, P. and G. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business and Economic Statistics 31*(3), 253–264.

Gordon, M. (1964). *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics 7*(1), 465–485.

Greenwood, M. J. and Z. Ward (2015). Immigration quotas, world war i, and emigrant flows from the united states in the early 20th century. *Explorations in Economic History 55*, 76–96.

Hainmueller, J., D. Hangartner, and G. Pietrantuono (2015). Naturalization fosters the long-term political integration of immigrants. *Proceedings of the National Academy of Science of the USA 112*(41), 12651–12656.

Hainmueller, J., D. Hangartner, and G. Pietrantuono (2017). Catalyst or crown: Does naturalization promote the long-term social integration of immigrants? *American Political Science Review 111*(2), 256–276.

Hatton, T. and J. Williamson (1998). *The Age of Mass Migration: Causes and Economic Impact*. Oxford: Oxford University Press.

Hémet, C. (2020). Unemployment and crime victimization: A local approach. CEPR Discussion Paper No. 14947.

Hémet, C. and C. Malgouyres (2018). Diversity and employment prospects neighbors matter! *Journal of Human Resources 53*(3), 825–858.

Heß, S., D. Jaimovich, and M. Schündeln (2021). Development projects and economic networks: Lessons from rural Gambia. *The Review of Economic Studies 88*(3), 1347–1384.

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*(4), 1349–1390.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019). The allocation of talent and U.S. economic growth. *Econometrica 87*(5), 1439–1474.

Huntington, S. (2004). *Who Are We? The Challenges to America's National Identity*. New York: Simon and Schuster.

Jackson, M. O., B. W. Rogers, and Y. Zenou (2017). The economic consequences of social-network structure. *Journal of Economic Literature 55*(1), 49–95.

Koneru, K., V. S. V. Pulla, and C. Varol (2016). Performance evaluation of phonetic matching algorithms on english words and street names. In *Proceedings of the 5th International Conference on Data Management Technologies and Applications*, pp. 57–64.

Lafortune, J. and J. Tessada (2019). Smooth (er) landing? The dynamic role of networks in the location and occupational choice of immigrants. Unpublished manuscript, Universidad Catolica de Chile.

Lazear, E. P. (1999). Culture and language. *Journal of Political Economy 107*, S95–S126.

Liang, Z. (1994). Social contact, social capital, and the naturalization process: Evidence from six immigrant groups. *Social Science Research 23*(4), 407–437.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Massey, D. and N. Denton (1985). Spatial assimilation as a socioeconomic outcome. *American Sociological Review 50*(1), 94–106.

Mele, A. (2017). A structural model of dense network formation. *Econometrica 85*(3), 825–850.

Munshi, K. (2003). Networks in the modern economy: Mexican immigrants in the US labor market. *Quarterly Journal of Economics 118*(2), 549–599.

Ortega, F. and A. Hsin (2018). Occupational barriers and the labor market penalty from lack of legal status. *IZA Discussion Paper N. 11680*.

Patacchini, E. and Y. Zenou (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics 42*(6), 938–949.

Portes, A. and J. Cutis (1987). Changing flags: Naturalization and its determinants among mexican immigrants. *International Migration Review 21*(2), 352–371.

Sato, Y. and Y. Zenou (2020). Assimilation patterns in cities. *European Economic Review 129*, 103563.

Schmutte, I. M. (2015). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics 33*(1), 1–32.

Sequeira, S., N. Nunn, and N. Qian (2020). Immigrants and the making of America. *Review of Economic Studies 87*(1), 382–419.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review 103*, 147–174.

Shertzer, A. (2016). Immigrant group size and political mobilization: Evidence from European migration to the United States. *Journal of Public Economics 139*, 1–12.

Shertzer, A., R. P. Walsh, and J. R. Logan (2016). Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 49*(4), 187–197.

Soyer, D. (1997). Jewish Immigrant Associations and American Identity in New York, 1880-1939. Cambridge, Mass.: Harvard University Press.

Spitzer, Y. (2018). Pogroms, networks, and migration. The Jewish migration from the Russian empire to the United States 1881–1914. Unpublished manuscript, The Hebrew University of Jerusalem.

Stuart, B. A. and E. J. Taylor (2021). Migration networks and location decisions: Evidence from US mass migration. *American Economic Journal: Applied Economics 13*(3), 134–175.

Tabellini, M. (2020). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. *Review of Economic Studies 87*, 454–486.

Ushchev, P. and Y. Zenou (2020). Social norms in networks. *Journal of Economic Theory 185*, 104969.

Verdier, T. and Y. Zenou (2017). The role of social networks in cultural assimilation. *Journal of Urban Economics 97*, 15–39.

Yang, P. (1994). Explaining immigrant naturalization. *International Migration Review 28*(3), 449–77.

A Appendix: Linking procedure

The analysis in our paper is based on the matched 1930 and 1940 censuses. The sample consists of newly arrived immigrants, i.e., foreign persons who entered the U.S. during the period of 1925–1930, for whom we observe exogenous network characteristics in 1930. We search for these individuals in the 1940 census. By creating this matched sample, which links the same individuals over a decade, we are able to compare the citizenship acquisition patterns in 1940 depending on the characteristics observed in 1930, including neighborhood attributes. Note that some observations might not be matched for several reasons: a change in name (especially for women, who might marry) or transliteration mistakes of the enumerator, mistakes in reporting the age, outmigration from the tri-state area for which we have data.

Nonetheless, the matching procedure closely follows the automated record linkage algorithm most recently discussed by Abramitzky et al. (2021). We start with all persons – men and women who entered the U.S. during the period of 1925–1930 and lived in NYC, for whom we observe exogenous network characteristics in 1930, and all foreign-born individuals observed in 1940, residing in either New York State, Connecticut, or New Jersey. Next, we "clean" the names in both datasets, removing any non-alphabetic characters and accounting for common misspellings and nicknames. As a third step, we restrict the sample to people who are unique in terms of first and last name, age, and country of birth in the 1930 census. Finally, we match these details to those in the 1940 census. A pair is considered to be an exact match if there is a unique match, while, in case of multiple matches, the observation is discarded. If no match is found, we look for unique matches within one and two years of the reported time period. At last, if no match is found, we look for unique matches within a two-year duration period for people whose first and last name are the same based on the New York State Identification and Intelligence System Phonetic Code. This phonetic code has been shown to result in fewer false positives than alternative algorithms (Koneru et al., 2016). In each of these steps, only unique matches are accepted. If none of these attempts produces a unique match, the observation is discarded. Finally, we use the updated version of this procedure, which also matches the 1940 census to the 1930 census and considers only the intersection of the two procedures.

B Appendix: Matching, attrition and selection

In Table B1, we analyze how our sample compares to the sample of unmatched individuals. The purpose of this exercise is to assess the extent to which our matched sample is representative of the universe of recently arrived immigrants living in New York City and whether our inference needs to take into account potential selection issues. Before presenting evidence about this, it is important to pinpoint the potential reasons why individuals might not be matched between the 1930 and the 1940 Censuses. Having a positive match depends on the ability of the algorithm described in Section A to successfully link two individuals based on their first and last name, age, and country of birth. In this context, a potential mismatch could occur "at random", i.e., irrespective of immigrant characteristics. While it is difficult

to quantify the extent of this type of mismatch, the only consequence that it will have is to potentially affect the representativeness of our sample. Considering that we match about 1/6 of the universe of observations, we think that this type of mismatch is not a cause of concern.

Table B1: Matching Probability and Naturalization

	(1)	(2)	(3)	(4)	(5)
Share Naturalized in 1930	.051***	.011**	.010**	.008*	.009**
	(.005)	(.004)	(.004)	(.004)	(.004)
	[.031]	[.007]	[.006]	[.005]	[.006]
Female			–.049***		–.049***
			(.002)		(.002)
Married			.068***		.068***
			(.002)		(.002)
Log Occupational Income Score			–.007***		–.007***
			(.001)		(.001)
Household Size			–.000**		–.000**
			(.000)		(.000)
Share of Females in Network				–.011*	.004
				(.006)	(.006)
Share Married in Network				.011**	.003
				(.005)	(.005)
Average Log Occupational Income Score in Network				–.005**	–.002
				(.002)	(.002)
Average Household Size in Network				.001*	.001*
				(.000)	(.000)
ED f.e.	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes
N	239,223	239,223	239,223	239,223	239,223
R^2	.00	.06	.08	.07	.08

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant in 1940 is a U.S. citizen. *Share Naturalized in 1930* measures the share of co-nationals who are naturalized and live in the same block of the migrant, in 1930.
All control variables are measured in 1930.
The network variables are calculated within the network, i.e. the *Share Married in Network* measures the share of co-nationals who are married in 1930 and live in the same block of the migrant.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

However, mismatches could also occur in a more systematic way, i.e., depending on observable and/or unobservable characteristics of migrants, which would generate sample "attrition". As an example, if immigrants from a given country of birth had a higher propensity or returning to their home country during the 1930s, they will be less likely to be matched. In fact, return migration could be a potential factor for a mismatch. Another, similar, reason would be internal mobility. If by 1940 some migrants moved out of the states of New York, New Jersey and Connecticut (the three states for which we have obtained 100% of the data), then they will not be matched. In other cases, the mismatch could occur in a more subtle manner. For example, if between 1930 and 1940 immigrants changed their name – and this change is phonetically substantial – then the algorithm would not be able to match these individuals. Changing name, often as part of the naturalization process, was much

more common than thought. As Biavaschi et al. (2017) show, in their sample about 30% of immigrants who naturalized decided to "Americanize" their name, with a peak for over 50% for immigrants from Central Europe and Russia. Name changes occurred more often among low-skilled migrants, who saw this as a way to overcome labour market barriers. For this reason, if in our sample there is a mismatch driven by name changes, it is possible that the matched individuals would have a different level of skills than the full sample of 1930.

To learn more about this point, we modelled the probability of being matched as a function of the key covariates from the 1930 Census that we use in our analysis – including the share of naturalized immigrants in the network. Table B1 shows that demographic characteristics such as being married and having a larger household are important determinant of being matched. Also some network characteristics – including the share of naturalized co-nationals – are associated with the probability of a match. To put things in perspective, a married person observed in 1930 has a 6.8% higher chances of an unmarried person observed in the same year of being observed in the 1940 Census, and therefore be part of our sample. The impact of the share of naturalized co-nationals seems to be rather modest: for example, individuals who live in areas with the highest share of naturalized co-nationals (i.e., 5th quintile in Figure 5) would have a 0.006% higher probability of being matched than individuals living in areas with the lowest share of naturalized co-nationals (1st quintile in Figure 5).

To further substantiate whether sample selection is a big concern in our context, we perform Heckman correction models of our baseline estimates. In order to model the selection equation, i.e., the probability of a match, we use a variable that can be plausibly excluded from the second stage equation (where we estimate the impact of the share of naturalized co-nationals on the probability of naturalization). This variable is the Scrabble points of the first and last names of the migrants. This is obtained by summing the scores attributed to each letter in the popular board game Scrabble. The rationale of using this particular variable is that Scrabble points quantify both the length and complexity of words, and thus provide a measure of the graphemic and phonemic characteristics of names. Foreign names tend to have high Scrabble points, because the structure of foreign words contains a combination of letters that is on average more uncommon (and thus bear higher points) when compared to the American English vocabulary – on which Scrabble points are defined. As such, this variable is a functional predictor of migrants' name complexity. How can name complexity affect the probability of a match (our selection equation)? There are at least two reasons. First, as Biavaschi et al. (2017) show, migrants with higher Scrabble points are more likely to change their name. If this is the case also in our sample, the consequence would be that higher Scrabble points would be associated with a lower probability of being matched. Second, name complexity (i.e., having a difficult to write name) is likely to generate mistakes in transcribing data by enumerators, which arguably would determine a lower probability of a match.

In Table B2, we present results from the same regressions as in Table 4 but with the Heckman correction, using Scrabble points as exclusion restriction. The selection equation confirms that name complexity is associated with a lower probability of a match. The positive sign of the Mills ratio seems to indicate the presence of positive selection bias. At the same

time, an inspection of the estimates of the share of naturalized co-nationals, especially in the most saturated specification (Column 6) indicates that the extent of the OLS bias might is rather limited and that, if anything, failing to correct for sample selection would result in our OLS estimates to produce a lower bound of the true effect.

Table B2: The Effect of Networks on Political Assimilation in 1940 – two-step Heckman estimates

	(1)	(2)	(3)	(4)	(5)
	Second Stage Regression				
Share Naturalized in 1930	0.113***	0.048**	0.039**	0.051**	0.043**
	(0.011)	(0.015)	(0.015)	(0.014)	(0.015)
	[0.088]	[0.038]	[0.030]	[0.039]	[0.034]
Female			-0.197***		-0.198***
			(0.006)		(0.006)
Married			-0.021**		-0.020**
			(0.007)		(0.007)
Log Occupational Income Score			0.012***		0.012***
			(0.002)		(0.002)
Household Size			0.000		0.000
			(0.001)		(0.001)
Share Females in Network				0.013	0.053*
				(0.024)	(0.023)
Share Married in Network				0.010	-0.008
				(0.019)	(0.019)
Average Log Occupational Income Score in Network				0.004	0.010
				(0.007)	(0.007)
Average Household Size in Network				0.002	0.001
				(0.002)	(0.002)
	Selection Equation				
Scrabble Points of Name and Surname	-0.409***	-0.397***	-0.372***	-0.398***	-0.372***
	(0.051)	(0.051)	(0.372)	(0.398)	(0.372)
Inverse Mills Ratio	0.106***	0.091***	0.051	0.092***	0.052
	(0.027)	(0.029)	(0.040)	(0.028)	(0.040)
ED f.e.	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes
N	239,223	239,223	239,223	239,223	239,223

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant in 1940 is a U.S. citizen. *Share Naturalized in 1930* measures the share of co-nationals who are naturalized and live in the same block of the migrant, in 1930.
All control variables are measured in 1930.
The network variables are calculated within the network, i.e. the *Share Married in Network* measures the share of co-nationals who are married in 1930 and live in the same block of the migrant.
Scrabble points of the names and surnames are calculated by summing the points of each letter composing names and surnames of migrants using the score of the Scrabble game.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

C Appendix: Additional Tables and Figures

Table C1: Ethnic Networks Characteristics – whole New York City

	Mean	St.Dev.
N. Immigrant Networks in each Neighborhood	25.117	19.327
N. Immigrant Households in each Network	8.658	19.083
N. Immigrants in each Network	17.022	41.484
N. Naturalized Immigrants in each Network	10.408	22.715
N. Individuals in each Network	41.389	108.802

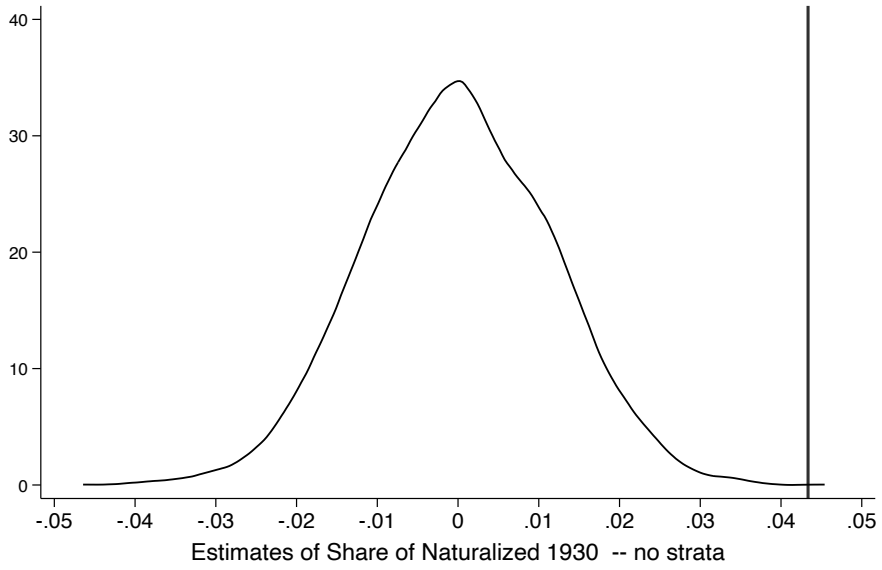
NOTES. Neighborhoods correspond to Enumeration Districts.
A network are defined is defined as the group of immigrants from the same country of birth living in the same block of residence (excluding household members).
The statistics refers to whole New York City and cover 133,896 networks across neighborhoods.

Table C2: The Effect of Networks on Political Assimilation – by Quintiles of Naturalization

	(1)	(2)	(3)
Share Naturalized in 1930 – 2nd Quintile	.012	.009	.011
	(.010)	(.010)	(.010)
	[.010]	[.008]	[.009]
Share Naturalized in 1930 – 3rd Quintile	.020*	.018*	.020*
	(.011)	(.010)	(.010)
	[.016]	[.015]	[.016]
Share Naturalized in 1930 – 4th Quintile	.018*	.015	.017
	(.011)	(.010)	(.011)
	[.015]	[.012]	[.014]
Share Naturalized in 1930 – 5th Quintile	.028**	.023**	.025**
	(.011)	(.011)	(.011)
	[.022]	[.018]	[.020]
ED/BPL f.e.	Yes	Yes	Yes
Individual Controls	No	Yes	Yes
Network Controls	No	No	Yes
N	37,761	37,761	37,761
R^2	.14	.19	.19

NOTES. Robust standard errors in parenthesis. Standardized estimates are presented in square brackets.
The dependent variable is an indicator that equals one if the migrant in 1940 is a U.S. citizen.
Each quintile refers to the quintile of the distribution of *Share Naturalized in 1930* in the sample. For all other variable descriptions, see Table 4.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

Figure C1: Placebo Tests



NOTES: The figure plots the kernel density of coefficient estimates of 5,000 regression specifications as in Col (5) of Table 4 where the share of naturalized co-ethnics is reassigned at random. Permutations are performed within the whole sample. The vertical line represents the coefficient estimate from the specification in Col (5) of Table 4. Out of 5,000 simulations, 2 generated a coefficient greater than the baseline (0.043), implying a randomization inference p-value=2/5,000=0.0003