DISCUSSION PAPER SERIES

DP16182

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

Costanza Biavaschi, Corrado Giulietti and Yves Zenou

ECONOMIC HISTORY

PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

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

Costanza Biavaschi, Corrado Giulietti and Yves Zenou

Discussion Paper DP16182
Published 24 May 2021
Submitted 22 May 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Economic History
- Public Economics

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

Abstract

This paper investigates the pathways through which immigrant communities (social networks) influence individual naturalization. Specifically, we examine the impact that a fraction of naturalized co-ethnics, residing in the same block as a new immigrant in New York City in 1930, have on the probability of said immigrant becoming a U.S. citizen in 1940. Our results indicate that the concentration of naturalized co-ethnics residing in the block positively predicts 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 Classification: J61, J62, N32, Z1

Keywords: Social Networks, Assimilation, Naturalization, migration

Costanza Biavaschi - costanza.biavaschi@ntnu.no
Norwegian University

Corrado Giulietti - c.giulietti@soton.ac.uk
University of Southampton

Yves Zenou - yves.zenou@monash.edu
Monash University and CEPR

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

Costanza Biavaschi† Corrado Giulietti‡ Yves Zenou§

May 22, 2021

Abstract

This paper investigates the pathways through which immigrant communities (social networks) influence individual naturalization. Specifically, we examine the impact that a fraction of naturalized co-ethnics, residing in the same block as a new immigrant in New York City in 1930, have on the probability of said immigrant becoming a U.S. citizen in 1940. Our results indicate that the concentration of naturalized co-ethnics residing in the block positively predicts 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 Sacha Becker 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

Are ethnic enclaves good or bad? Several studies have shown that living in an ethnic enclave positively impacts the economic outcomes of minorities (Edin et al., 2003, Munshi, 2003, Damm, 2009, 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, Eriksson, 2020). In this paper, we examine the impact of ethnic enclaves (or co-ethnic networks) on the long-run political assimilation of immigrants, which is measured based on their acquisition of U.S. citizenship more than ten years after their arrival. Indeed, naturalization is the cornerstone of immigrant 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, can obtain access to government benefits, can 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 of 1930–1940. This was a very 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. With as many as 12 million immigrants from all over the world having entered the city through Ellis Island during the first part of the 20^{th} century, NYC 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 of 1925–1930, for whom we observe exogenous network characteristics in 1930 and a 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 level – 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.[2] Our key measure of social networks is the fraction of naturalized immigrants from the same country of birth residing in the same block as newly-arrived immigrants.

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 committment, since it implied renouncing to own nationality and, thus, not been able to go back to the country of origin. In our econometric models, we account for several pre-determined characteristics of the networks and for changes in individual characteristics between 1930 and 1940. 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 they 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 of the blocks (i.e., the social networks) within 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 appartments, 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. One concern may be that newly arrived immigrants reside with someone from their family, so that the network would be endogenous. In a robustness check, we show this is not the case; families do not differ from other co-ethnics and excluding members of the same family does not affect our results. We, therefore, treat the composition of the neighborhood by birthplace within an enumeration district 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 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 the immigrant's likelihood of being naturalized in 1940. A 10% increase in the neighborhood's co-ethnic naturalization rate is associated with about a 0.55% increase in the probability that an immigrant of that neighborhood naturalizes. The magnitude of the network effect

[2]The block is very small. On average, there are 54 individuals in each block (Table 2).

differs substantially with the immigrants' origins, with a stronger impact for immigrants from Poland, Russia, and Scandinavia. The impact also depends non-linearly on the network size. For example, immigrants who were exposed to the highest rate of co-ethnic naturalization (the upper quintile) were 5% more likely to naturalize than those exposed to the the lowest rate (the bottom quintile).

To further substantiate the credibility of our results, we perform a series of robustness checks. First, we perform a placebo test in which, instead of using the block where the immigrant resides, we assign her a randomly selected block. We find that the effect disappears, thereby confirming that the peers from the same block form the relevant group that influences the naturalization process. Second, we employ a finer definition of what is referred to as a 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. The effect remains positive, albeit smaller in magnitude. Third, we exclude the blocks where immigrants live with people that have the same surname (who are very likely to be a family member) and the blocks where the immigrant is either a boarder, roomer, lodger, or a domestic worker. The pattern of the results remains unchanged.

We then investigate the heterogeneity of the network effect across two dimensions: gender and borough of residence. We find that both males' and females' citizenship acquisition decision is influenced by the presence of naturalized co-nationals of the same gender in the network, albeit the point estimate is larger for males than for females. We also find that the effects are insignificant for Manhattan, while they are quantitatively and qualitatively the same across the other boroughs. One of the explanations that we put forward for this result is that Manhattan had a high population density, which would make social interactions between individuals less personal and more anonymous.

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 having 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. 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 a 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 to apply 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 providing useful information on the naturalization process using a common language.

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 type of information that the network conveys might not be particularly related to overcoming migration restrictions, which vary for different countries of origin. Second, we find that networks have a stronger effect in areas where the concentration of people working in barred occupations (jobs for American citizens only) was larger – suggesting the importance of labor market-related information as an additional incentive to naturalize – and in areas where a Democrat (usually more pro-migration than a Republican) was elected.

1.1 Related literature

Our paper contributes to different strands of the literature.

First, we contribute 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).[3]

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.[4,5] One 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 because the geographical unit is the ward, which is quite large. Indeed, the average size of the potential electorate in these wards (men aged 21 and above, excluding

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

[4]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).

[5]This is an active research area in the non-economic literature, but it mostly documents correlations. See Abascal (2017) and the references therein.

immigrants who have lived in the U.S. for less than two years) is slightly over 10,500 men per ward; thus, selection is clearly an issue. As a comparison, we consider the block, and we have, on average, 54 individuals in each network (Table 2). Another exception is 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 the focus, the time period, the country, and the outcome are very different from ours.[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 (IV) approach based on friends of friends' characteritics (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). Following the seminal paper of Bayer et al. (2008), we contribute to this literature by measuring the network at the neighborhood level, using the block to determine the ethnic network of each immigrant. By doing so, we are able to examine the causal impact of networks on the outcomes, i.e. assimilation.

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), 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., 2019, Giuliano and Tabellini, 2020, Sequeira et al., 2020, Tabellini, 2020).[8]

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 empirical model and our identification strategy. Section 4 presents our main results as well as the non-linear and heterogenous effects and provides some robustness checks. In Section 5, we investigate the mechanisms behind our results. Finally, Section 6 concludes the paper.

[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).

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

2 Data

To measure the role that social networks play in immigrant political assimilation, we exploit the 1930 and 1940 American censuses. This source contains the records of individuals with details about their neighborhood of residence, household composition, and, consequently, neighborhood composition. We describe 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 and 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 two main reasons. First, individuals who are not U.S.-born and not U.S. citizens can apply for naturalization. Second, having just arrived, the immigrants are unlikely to be already naturalized by law. Indeed, to become a citizen, immigrants had to first reside continuously in the U.S. for at least two years, one of which was 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 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 present 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.

We created a matched dataset that tracked 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. We link men and women over time by first and last name, age, and country of birth. Further details on the main linking procedure are provided in Appendix A. In Table A1, we report the summary statistics of the matched and unmatched observations, while, in Table A2, we provide the correlation between being matched and being naturalized. Our main linked sample contains 46,175 individuals, which corresponds 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).

Table 1 shows the summary statistics of the matched sample, including the immigrants' naturalization status in 1940 and their characteristics measured in 1930. By 1940, about 57% of the immigrants in our sample had naturalized. There is some variation in the naturalization propensities across the countries of birth. The naturalization rates are the highest among Polish immigrants (71%), followed by Russian (68%) and Irish immigrants (66%). On the other extreme, only 50% of the Scandinavian and 48% of British immigrants had naturalized by 1940.

Looking at the household characteristics for 1930, 45% of the immigrants were married,

Table 1: Summary statistics of the sample

	Mean	SD
Naturalized, 1940	0.575	0.494
Naturalized among Irish, 1940	0.664	0.472
Naturalized among Italians, 1940	0.556	0.497
Naturalized among Germans, 1940	0.567	0.495
Naturalized among Polish, 1940	0.711	0.453
Naturalized among Russians, 1940	0.681	0.466
Naturalized among British, 1940	0.481	0.500
Naturalized among Scandinavians, 1940	0.500	0.500
Naturalized among other Europeans, 1940	0.595	0.491
Naturalized among all others, 1940	0.444	0.497
Marital status, 1930	0.450	0.498
Family size, 1930	3.533	2.167
N. children, 1930	0.524	1.091
Ireland	0.155	0.362
Italy	0.097	0.296
Germany	0.332	0.471
Poland	0.061	0.240
Russia	0.039	0.194
UK	0.127	0.334
Scandinavia	0.050	0.218
Other EU	0.078	0.268
Other country	0.060	0.237
Occupational income score, 1930	10.056	12.807
House value (in 10,000), 1930	0.786	3.166
Is renting, 1930	0.897	0.304
Monthly rent contract, 1930	73.989	353.145
Observations	46175	

Other country includes all other countries present in the dataset. Of these, the largest ones are Canada, New Foundland, Cuba, and Turkey.

with an average family size of 3.5 individuals.[9] About 30% of the immigrants were from English-speaking countries, and the vast majority could read and write. In terms of country of origin, 33% of the recent arrivals were from Germany, and 15% were from Ireland. The other major sending countries were the UK (12%), Italy (10%), Poland (6%), Scandinavia (5%), and Russia (4%). 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. Table 1 also reports the measures of the socio-economic background of immigrants. The average occupational income score was $1,000, with almost 90% of the immigrants renting their living premises and paying an average rent of $73. Among the homeowners, the average value of a house amounted to $84,339.[10]

Figure 1 provides some insight about the spatial distribution of the immigrants group in 1930 for selected countries of origin. These different maps compare the relative size of the immigrant groups and their spatial concentration across the enumeration districts. Figure 1 shows that the 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 NYC. Further, there are relatively less immigrants from Scandinavia, and they are mostly concentrated in certain areas of Brooklyn.

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 setting, 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 2 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. To operationalize our definition of network, we partition each neighborhood into sub-areas. In practice, we use the

[9] 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).

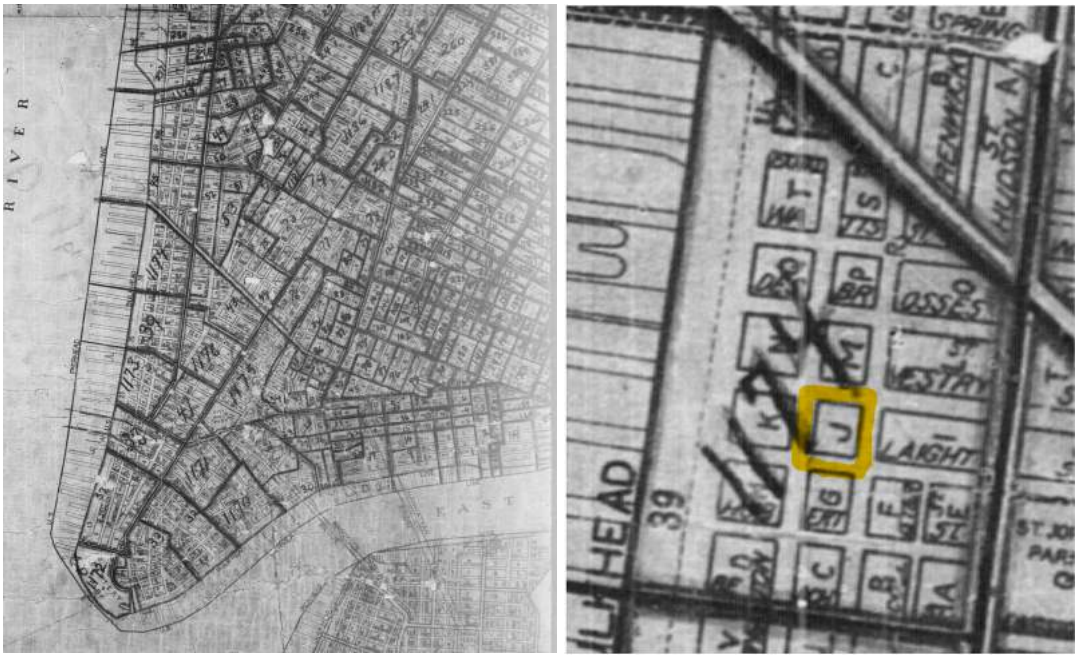
[10] The occupational income score is measured as per the value of the dollar in 1950. The value of $1,000 corresponds to about $10,060 in 2020. Likewise, the average value of a house would correspond to about $894,000 in 2020. It should be noted that the value of the house and the value of the rent are reported only for 7,237 and 37,399 observations, respectively.

Figure 1: Share of immigrants in Manhattan and Brooklyn – Enumeration Districts

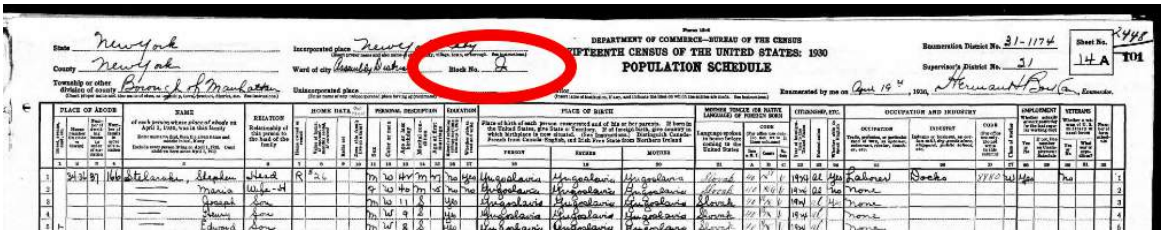
Source: 1930 Census. The immigrant shares are calculated as the number of people born in a given country in each enumeration district divided by the total population of each enumeration district.

Figure 2: 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 2a and 2b: 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 2c: Ancestry.com

block of residence of the immigrants to identify the sub-areas. We chose to use blocks for several reasons. First, they are the smallest geographic 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 spatial unit for approximating the social space in which individuals establish social contacts. As an example, Panel (b) of Figure 2 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 as challenging with the current data as it is with the historical data. Nowadays, the census reports the census blocks, which typically, albeit not always, 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 present in the original census manifests of the 1930 Census. One such example is shown in Panel (c) of Figure 2. Yet, such information has never been digitalized. For this reason, for all the 2.3 million inhabitants of NYC in 1930, we accessed the original census manifests 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.

With the block information available, we assume that the 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 the individuals who come from the *same country of birth* and *live in the same block* as the immigrant (excluding their own household's members). Our choice to define networks by the same country of birth is motivated by various reasons. First, immigrants tend to settle in areas in which other immigrants have settled in the past. Second, ethnic communities tend to be more socially cohesive – something that was 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. In fact, the 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 3, 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, namely Germany, Ireland, and Italy, strongly relied on their networks during the naturalization process. Among these groups, in fact, more than 80% brought a witness from their country of birth. These percentages are lower, but remain substantial, for the other countries, such as Poland and Russia.[11]

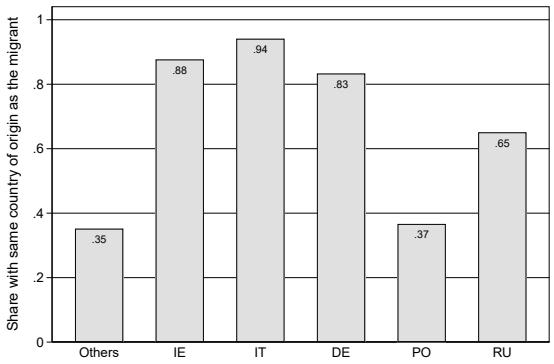
To summarize, we assume that the network exposure of each immigrant i is given by the share of *naturalized* non-household members $ynet_{dbc,t}$ living at time t in the same enumeration district d and in the same block b and coming from the same country of birth c. This is equal to:

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

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 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

[11] 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).

Figure 3: 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.

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.

Table 2 shows the descriptive statistics for our networks. In each enumeration district, we observe, on average, 35 block-country-of-birth combinations, with 54 individuals in each of them, 10 of who are naturalized, i.e., the average value of $ynet_{dbc,t}$ is equal to $10/54 = 18.5\%$. There are 22 distinct households in a typical network.

Table 2: Summary statistics of the networks

	Mean	SD
Distinct networks in ED	35.608	21.800
N. individuals in network	54.658	123.757
N. distinct households	22.557	48.871

3 Econometric model and identification

3.1 Econometric model

Our empirical model can be written as follows:

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

where, as above, $y_{idbc,1940}$ 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, at $t = 1940$, a immigrant $idbc$ is a U.S. citizen, and 0 otherwise. The key explanatory variable is the *network exposure* $ynet_{dbc,1930}$, which is defined in equation (1) at time $t = 1930$; it is the share of naturalized co-nationals (same

country of birth c) who live in the same enumeration district d and block b as immigrant $idbc$ in 1930. We include controls for changes between 1930 and 1940 in individual characteristics x^k, i.e., $\Delta x^k_{idbc} = x^k_{idbc,1940} - x^k_{idbc,1930}$ as well as controls for other network characteristics in 1930, i.e., $z^j_{dbc,1930}$. Finally, we include η_c, a country-of-birth fixed effect, and $\eta_{d,1930}$, an enumeration district fixed effect in 1930. 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. 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 discussing the plausibility of this identification strategy.

3.2 Identification

Our identification strategy builds on the local nature of our data and follows the research design of Bayer et al. (2008). In the latter, 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 an aggregation of blocks.[12] In practice, the identification in our analysis is achieved by comparing blocks' attributes within narrowly defined neighborhoods using fixed effects to control for neighborhood's 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). First, 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). Second, 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

[12]The seminal approach of Bayer et al. (2008) has been used to identify neighborhood and network effects in different contexts; for example, it has been used 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).

recently arrived immigrants, who are likely to lack information about the blocks' characteristics. Second, our definition of social network is rather narrow, as it includes, on average, about 22 households and 54 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 would 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 the 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. Finally, these areas (enumeration districts) do not follow administrative or official borders and even change over time. The 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.

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 the enumeration district fixed effects), the remaining variability in the citizenship rates across the blocks within the neighborhood is as good as exogenous.

Yet, before moving to the results, we investigate 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 select, at random, a single 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.[13] In Table 3, we present evidence 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 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

[13] 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.

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
Earnings	0.755	0.160	0.000	0.876	0.159	0.000
Value of the house	-27674.070	0.000	0.023	-4178.883	0.249	0.001
Is renting	-0.121	0.000	0.008	-0.029	0.033	0.000
Value of the rent	-47.498	0.034	0.001	-14.611	0.356	0.000
Married	0.030	0.176	0.000	0.013	0.587	0.000
Family size	0.598	0.000	0.004	-0.058	0.548	0.000
Number of children	0.141	0.003	0.001	0.020	0.701	0.000

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses. 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.

In the first three columns, we do not include the neighborhood fixed effects, while, in the last three columns, we include the enumeration district fixed effects as well as the country-of-birth fixed effects, i.e., we control for sorting across the blocks and ethnic groups. 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 simple OLS regressions without fixed effects (first three columns), we find that, indeed, the naturalization rates and immigrant characteristics co-vary, with the correlations being statistically significant, albeit, reassuringly, not explaining much of the overall variation in the dependent variable. Interestingly, in the last three columns, these findings disappear. Indeed, once we include the enumeration district fixed effects and country-of-birth fixed effects, the immigrants do not sort across blocks within a chosen neighborhood based on nationality. The naturalization rate of a neighborhood is always a "weak predictor" of the characteristics of the immigrant, both in terms of statistical significance and size: Most of the coefficients are greatly reduced as compared to their OLS counterparts. The only characteristic that could potentially be worrying is that the immigrants living in larger networks of naturalized co-nationals were less likely to be renting their housing unit. Our analysis will control for *changes* in this characteristic, and, therefore, arguably net out any individual-specific heterogeneity that could explain the raw correlation in 1930. Even if sorting does not seem to take place as per the observable characteristics in 1930, we will control for demographic (married, family size, and number of children) and economic characteristics (income and home ownership) in several of

our specifications below, demonstrating that the results do not change.[14]

We now proceed with the main results of our analysis. Section 4.4 provides further evidence in favor of the credibility of our identification strategy.

4 Results

4.1 Main results

Table 4 reports the estimation results for Model (2). Column (1) shows the results without controlling for any of the fixed effects, whereas birthplace and enumeration district fixed effects are added to the specification in Column (2) and maintained throughout the table.

Starting from Column (1), we find a positive and significant causal effect created by the share of the co-nationals who have acquired citizenship in 1930 on the immigrant's likelihood of being naturalized in 1940. A 10% increase in the neighborhood's naturalization rate in 1930 is associated with an increase of 1.37% in the likelihood of the immigrant acquiring citizenship in 1940. 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 a neighborhood 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 the enumeration district fixed effects partly reduces the coefficient of the network variable: A 10% increase in the neighborhood's naturalization rate is now associated with about a 0.56% 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.

In the remaining columns, we progressively add control variables. Overall, the size and significance of our key estimate is largely unaffected by the inclusion of additional covariates. Specifically, Column (3) includes age fixed effects to capture the potentially different naturalization gradients between young and old immigrants. Column (4) includes a set of time-varying individual characteristics, such as changes in marital status, income, home ownership, family size, and number of children, to control for how incentives to naturalize have evolved over time. Column (5) introduces controls for additional network characteristics in 1930, such as the share of married individuals in the network, average family size, and number of children as well as average income levels and home ownership rates. Finally, in Column (6), we include the full set of control variables.

The estimated coefficient of the share of naturalized immigrants decreases slightly (from

[14] Due to the large number of missing values and questions about data quality for the variables that capture the value of the house and the rent, we will not control for these variables.

Table 4: The effect of networks on political assimilation in 1940, by country

	Without controls			With controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Share naturalized in 1930	.137***	.056***	.058***	.063***	.054***	.055***
	(.013)	(.016)	(.016)	(.016)	(.017)	(.018)
Change in earnings				.003***		.003***
				(.000)		(.000)
Change in home ownership				–.060***		–.060***
				(.007)		(.007)
Change in marital status				.106***		.106***
				(.006)		(.006)
Change in the number of children				.001		.001
				(.003)		(.003)
Change in family size				–.005***		–.005***
				(.002)		(.002)
Earnings network in 1930					.002**	.001
					(.001)	(.001)
Home ownership network in 1930					.005	–.010
					(.018)	(.019)
Share of married network in 1930					–.011	–.005
					(.021)	(.021)
Size of fam. network in 1930					–.002	–.001
					(.006)	(.006)
N. of children network in 1930					.016*	.015*
					(.009)	(.009)
ED f.e.	No	Yes	Yes	Yes	Yes	Yes
Age f.e.	No	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes	Yes
R^2	.00	.13	.14	.16	.14	.16
N	45111	44607	44604	41610	44266	41584

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.

The dependent variable is an indicator that equals to one if the immigrant is a U.S. citizen in 1940. *Share naturalized 1930* measures the share of the co-nationals who are naturalized and live in the same block as the immigrant, in 1930. *Change in earnings* measures the change in the (log) occupational income score between 1940 and 1930. *Change in marital status* measures the change in marital status between 1940 and 1930. *Change in the number of children* measures the change in the number of children between 1940 and 1930. *Change in family size* measures the change in the size of the family betweeen 1940 and 1930. *Earnings network in 1930* measures the average occupational income score in 1930 among the co-nationals living in the same block as the immigrant. *Home ownership network in 1930* measures the home ownership rate in 1930 among the co-nationals living in the same block as the immigrant. *Share of married network in 1930* measures the share of the co-nationals who are married in 1930 and live in the same block as the immigrant. *N. children network in 1930* measures the average number of children in 1930 among the co-nationals of the immigrants who live in the same block.

* $p < .10.$; ** $p < .05.$; *** $p < .01.$

0.063 to 0.055) as we progressively add control variables to our specification; but the difference between the two coefficients is statistically insignificant, and the estimate itself remains statistically significant at the 1% level across columns. The estimate in Column (6) implies that a 10% increase in the size of the network raises the probability of naturalization by 0.55%. Other estimates suggest that marriage and income as well as children in the network and average income in the network are positively associated with a immigrant's probability

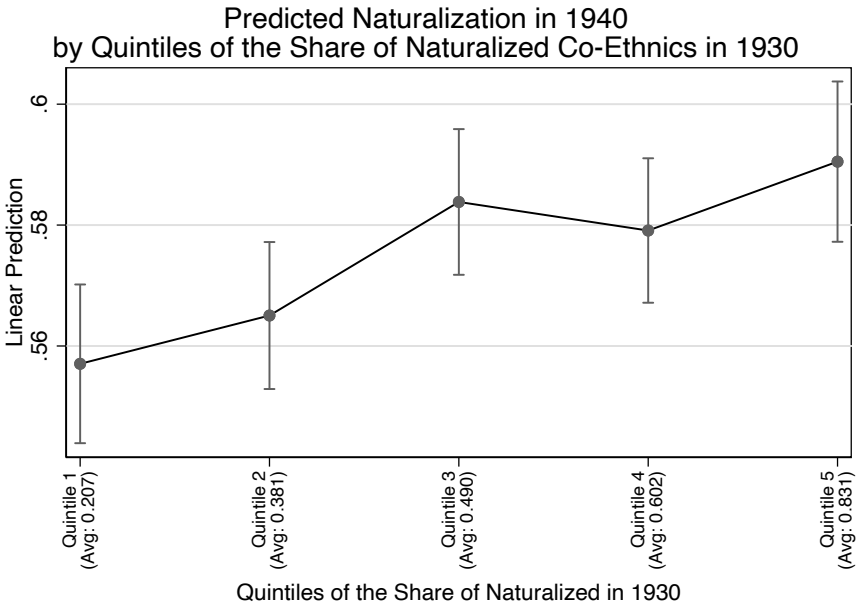
of acquiring citizenship.

4.2 Non-linear effects

We now explore the possible sources of heterogeneity in the social network effect. In particular, we examine whether the impact of ethnic networks varies with the intensity of exposure. In Table 5, we include four dummy variables for the four top quintiles of the share of naturalized co-ethnic peers distribution. This table has the same structure as Table 4, in which Column (6) is our preferred estimate.

We find that all the coefficients are positive and increasing along with the quintiles, with the third, fourth, and fifth quintiles being statistically distinguishable from the first. Having a proportion of naturalized co-ethnic peers in the fifth quintile in 1930 is associated with an increase of 2.9 percentage points in the likelihood of the immigrant being naturalized in 1940 relative to the first quintile. Figure 4 displays the predicted probability of naturalization in 1940 by the quintiles of the share in 1930 and report the average share of the naturalized co-ethnic peers in each quintile. This implies that, if we were to move a recently arrived immigrant in 1930 from a block peer group in which 83.1% of her co-ethnic peers are naturalized (average share in fifth quintile) to one in which 20.7% of her co-ethnic peers are naturalized (average share in first quintile), her likelihood of being naturalized in 1940 will be reduced by more than 3.3 percentage points, which corresponds to a 16% reduction in the average likelihood of naturalization in that quintile.

Figure 4: Predicted probability of naturalization



Notes: The plot represents coefficients β and confidence intervals from separate regressions of equation (2) for each quintiles of the variable $ynet_{dbc,1930}$.

Table 5: The effect of networks on political assimilation in 1940, by quintiles of the share of naturalized immigrants in 1930

	Without controls			With controls		
	(1)	(2)	(3)	(4)	(5)	(6)
2nd quintile in 1930	.032***	.007	.008	.006	.006	.004
	(.009)	(.010)	(.010)	(.010)	(.010)	(.010)
3rd quintile in 1930	.073***	.027***	.027***	.030***	.025**	.027**
	(.009)	(.010)	(.010)	(.011)	(.010)	(.011)
4th quintile in 1930	.084***	.021**	.022**	.026**	.019*	.022*
	(.009)	(.010)	(.010)	(.011)	(.011)	(.011)
5th quintile in 1930	.093***	.031***	.033***	.035***	.030**	.029**
	(.009)	(.011)	(.011)	(.011)	(.012)	(.012)
Change in earnings				.003***		.003***
				(.000)		(.000)
Change in home ownership				–.060***		–.061***
				(.007)		(.007)
Change in marital status				.106***		.106***
				(.006)		(.006)
Change in n. children				.001		.001
				(.003)		(.003)
Change in family size				–.005***		–.005***
				(.002)		(.002)
Earnings network in 1930					.002**	.001
					(.001)	(.001)
Home ownership network in 1930					.002	–.014
					(.018)	(.019)
Share of married network in 1930					–.009	–.003
					(.021)	(.021)
Size of fam. network in 1930					–.002	–.001
					(.006)	(.006)
N. of children network in 1930					.017*	.017*
					(.009)	(.009)
ED f.e.	No	Yes	Yes	Yes	Yes	Yes
Age f.e.	No	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes	Yes
R^2	.01	.13	.14	.16	.14	.16
N	45111	44607	44604	41610	44266	41584

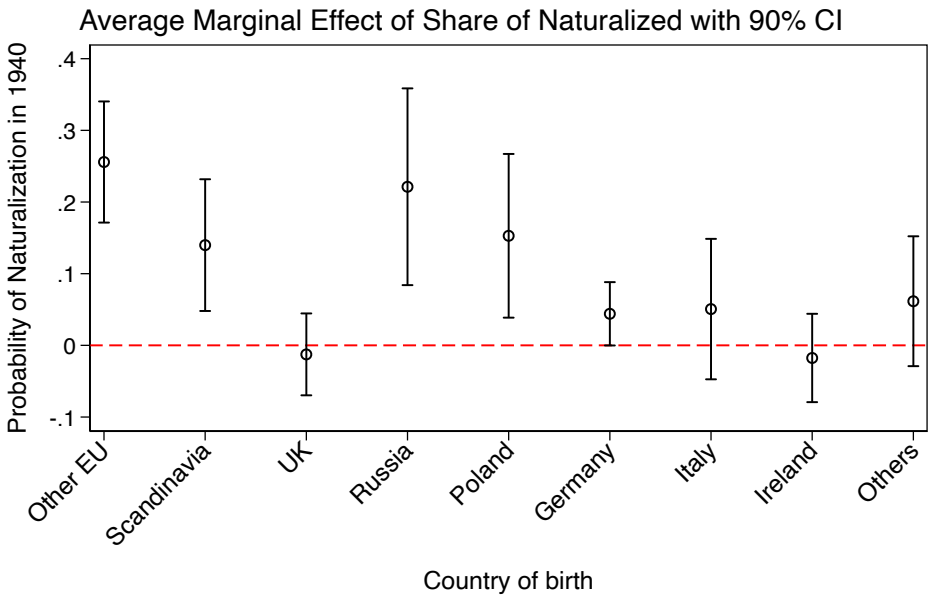
NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
The dependent variable is an indicator that equals to one if the immigrant is a U.S. citizen in 1940. The *2nd–5th quintile in 1930* measures whether our key independent variable is in the 2nd–5th quintile of its distribution in the sample. For all other variable descriptions, see Table 4.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

4.3 Heterogenous effects

Next, we investigate whether these results vary with the immigrants' origin. Immigrants differ substantially depending on the country they are from, for example, in terms of their 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 in which we allow the

variable $ynet_{dbc,1930}$ to have a differential effect for each country of origin. Figure 5 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 the 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 size of the network raises the probability that an immigrant from Russia naturalizes by 2.2%, while the same for an immigrant from Poland will result in a 1.52% increase in probability. However, for British and Irish immigrants, as well as Italians, such effect is mainly zero.[15]

Figure 5: 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 of the share of naturalized in the network on the probability of naturalization in 1940.

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 were Jews fleeing pogroms from Russia,[16] were very eager to assimilate and become an

[15] A test of joint significance for the interaction terms yields an F-value of 2.21, with a p-value of 5.06%.

[16] 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

American citizen. Their social network did reinforce their eagerness to assimilate because it helped them efficiently deal with the naturalization process, which is bureaucratic and needs a good knowledge of the English language as well as American witnesses (see Figure 3, which 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 presented in Table 1 that shows that the rate of naturalization for Italians was 55.6%, while that for Russians was 68.1% in 1940.

We explore the heterogeneity of the network effect across two additional dimensions: gender and borough of residence. Table 6 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. 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 larger for males than for females. One potential explanation for this could be that females could also acquire citizenship through means other than the direct application, including marriage, and, thus, the role played by their networks of naturalized neighbors might be less relevant.

Table 6: The effect of networks on political assimilation in 1940, by sex

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of naturalized population in 1930 (by sex)	.098***	.068***	.067***	.131***	.035*	.039**
	(.015)	(.021)	(.021)	(.014)	(.019)	(.018)
ED f.e.	No	Yes	Yes	Yes	Yes	Yes
Age f.e.	No	No	Yes	Yes	Yes	Yes
BPL f.e.	No	No	No	No	Yes	Yes
R^2	.00	.21	.22	.00	.19	.20
N	19706	18928	18925	24719	23984	23981

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses. In Columns (1)–(3), the sample is restricted to males only. In Columns (4)–(6), the sample is restricted to females only. The dependent variable is an indicator that equals to one if the immigrant is a citizen in 1940. *Share of naturalized population in 1930 (by sex)* measures the share of the co-nationals of the immigrant who are of the same sex, naturalized, and who reside in the same block as the immigrant in 1930.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

In Table 7, 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 1). Furthermore, one might expect that, in a densely populated area such as Manhattan, social interactions with neighbors might be more intense than those in less densely populated areas, such as Brooklyn or Queens. At the same time, the network might play a larger role in the less populated areas, where it is harder to interact with

important for the assimilation of these ethnic minorities.

others. The point estimates of the network effect in Manhattan are similar to those in our benchmark specification and, yet, statistically insignificant. Overall, however, the effects seem quantitatively and qualitatively the same across the other boroughs.

Table 7: The effect of networks on political assimilation in 1940, across the boroughs

	Manhattan			Brooklyn			Bronx, Richmond, Queens		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of naturalized population in 1930	.205***	.047	.051	.143***	.068**	.068**	.137***	.056***	.058***
	(.031)	(.036)	(.036)	(.021)	(.028)	(.028)	(.013)	(.016)	(.016)
ED f.e.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age f.e.	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
BPL f.e.	No	No	No	No	Yes	Yes	Yes	Yes	Yes
R^2	.01	.13	.14	.00	.15	.16	.00	.13	.14
N	14022	13942	13939	14311	14153	14152	45111	44607	44604

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
In Columns (1)–(3), the sample is restricted to immigrants whose county of residence in 1930 was NYC. In Columns (4)–(6), the sample is restricted to immigrants whose county of residence in 1930 was Kings. In Columns (7)–(9), the sample is restricted to immigrants whose county of residence in 1930 were Queens, Richmond, or the Bronx.
Share of naturalized population in 1930 measures the share of the co-nationals of the immigrant who are naturalized and live in the same block as the immigrant in 1930.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

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 the place of residence of the immigrants. Hence, even at the end of mass migration, political integration occurred through the use of networks and across ethnic lines.

4.4 Robustness checks

In this last section, we report the robustness results from a set of additional analyses.

First, we estimate Model (2) by replacing the actual composition of the block of residence of immigrant i with the composition of a randomly selected block. If our fixed effect strategy controls for the unobserved characteristics, then the composition of the other blocks should not have any effect on the citizenship patterns in this *placebo* regression. Table 8 reports the results. We find *no evidence* across all the specifications of a significant impact on the citizenship decisions of the share of naturalized co-nationals in a randomly selected block. This indicates that the share of naturalized co-nationals residing in the same block are the relevant *peers* for a newly arrived immigrant.

Next, we propose a more constraining alternative strategy, whereby we define the network of an immigrant as the group of people living within an even finer zone–a *block-street* area. We define a block-street network as the fraction of naturalized co-ethnic individuals living in the *same block* and on 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 only in the same enumeration district but also on 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 sorting within the block but not within the streets of the same block. As explained in

Table 8: The effect of networks on political assimilation in 1940, placebo

	Without controls			With controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of naturalized population in 1930, from random block	.016	.014	.014	.015	.014	.015
	(.010)	(.011)	(.011)	(.011)	(.011)	(.011)
Change in earnings				.003***		.003***
				(.000)		(.000)
Change in home ownership				–.060***		–.059***
				(.007)		(.007)
Change in marital status				.105***		.105***
				(.006)		(.006)
Change in n. children				.002		.002
				(.003)		(.003)
Change in family size				–.004***		–.004***
				(.002)		(.002)
Share of married network in 1930					.008	.121
					(.085)	(.097)
Size of fam. network in 1930					–.030*	–.022
					(.016)	(.016)
N. of children network in 1930					.088	.081
					(.079)	(.082)
Home ownership network in 1930					–.027	–.029
					(.031)	(.033)
Earnings network in 1930					.003	.002
					(.003)	(.003)
ED f.e.	No	Yes	Yes	Yes	Yes	Yes
Age f.e.	No	No	Yes	Yes	Yes	Yes
BPL f.e.	No	Yes	Yes	Yes	Yes	Yes
R^2	.00	.14	.14	.16	.14	.16
N	45111	44591	44589	41598	44348	41575

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
The dependent variable is an indicator that equals to one if the immigrant is a U.S. citizen in 1940.
The independent variable *Share of naturalized population in 1930, random network* was constructed by randomly assigning to each immigrant any other network.
For the definition of all other variables, see descriptions in Table 4.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

Section 3.2, 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 9 displays the results of this test. To make the results comparable to those of our benchmark model, we start by re-estimating the model in Column (3) of Table 4, which included birthplace, age, and enumeration district fixed effects, with standardized variables. We find that a one standard deviation increase in the share of the naturalized co-ethnics increases the likelihood of citizenship acquisition by 0.013 standard deviations. Next, we estimate the new model in which the network is defined at the block-street level, so that we can include the block fixed effects. Column (2) of Table 9 shows that the key results are unchanged. In the last two columns of this table, we add additional controls and show that the main findings remain stable, albeit statistically weaker. Allowing for sorting within block does not affect our main conclusions.

Table 9: The effect of networks on political assimilation in 1940, including block fixed effects

	Benchmark	Block f.e.	Benchmark	Block f.e.
	(1)	(2)	(3)	(4)
Share of naturalized population in 1930	.013***	.011***	.012***	.008*
	(.004)	(.004)	(.004)	(.005)
Change in earnings			.003***	.003***
			(.000)	(.000)
Change in home ownership			–.060***	–.060***
			(.007)	(.008)
Change in marital status			.106***	.107***
			(.006)	(.007)
Change in n. children			.001	.000
			(.003)	(.003)
Change in family size			–.005***	–.005***
			(.002)	(.002)
Earnings network in 1930			.001	.001
			(.001)	(.001)
Home ownership network in 1930			–.010	.002
			(.019)	(.021)
Share of married network in 1930			–.005	.021
			(.021)	(.019)
Size of fam. network in 1930			–.001	.001
			(.006)	(.005)
N. of children network in 1930			.015*	.011
			(.009)	(.009)
ED f.e.	Yes	Yes	Yes	Yes
Block f.e.	No	Yes	No	Yes
Age f.e.	Yes	Yes	Yes	Yes
BPL f.e.	Yes	Yes	Yes	Yes
R^2	.14	.24	.16	.26
N	44604	39079	41584	36190

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
The dependent variable is an indicator that equals to one if the immigrant is a U.S. citizen in 1940.
For comparison purposes, Column (1) reports our benchmark specification of Column (3) of Table 4, and Column (3) reports our benchmark specification for Column (6) of Table 4, but where the share of naturalized co-nationals has been standardized.
Columns (2) and (4) report a specification in which the independent variable is defined at the block-street level, and, therefore, it measures the (standardized) share of co-nationals who were naturalized in 1930 and lived on the same block and street as the immigrant.
For a definition of all other variables, see table notes of Table 4.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

A last set of robustness checks are presented in Table 10. There are two remaining concerns in this framework that cannot be addressed with the inclusion of a rich set of fixed effects, such as those in our benchmark specification or in Table 9. First, an immigrant might decide to reside in a specific block (or street-block) because her relatives 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 be co-nationals or friends of co-nationals. Notwithstanding the fact that, for these to be valid concerns, the existence of these perhaps stronger family ties in a given block has to correlate with the share of the naturalized co-nationals in that block. We exploit the richness of our data to provide indicative evidence that these effects are unlikely to drive our results. To address the fact that families might sort into specific blocks and ease the residence of their own family

members, we identify the share of individuals outside the household and residing in the same block who have the same surname.[17] We then include in our estimation sample only those immigrants who hold surnames that are unique within the network. In other words, as long as family members share the same surname, we focus on the immigrants who do not reside in the same block as any other family member. The results are reported in Columns (1)–(3) of Table 10, where we increasingly control for a richer set of fixed effects. We find that excluding the blocks where immigrants live with people who have the same surname as the immigrant herself does not affect the main conclusions. To address the second concern, namely that an immigrant who arrived earlier might decide to 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 migrants. These regressions are shown in Columns (4)–(9) of Table 10. Excluding these migrants does not affect our main results.

Table 10: The effect of networks on political assimilation in 1940, excluding groups

	Different surnames			Excl. roomers,lodgers, boarders			Excl. domestic workers		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of naturalized population in 1930	.132***	.050***	.052***	.129***	.049***	.052***	.137***	.048***	.050***
	(.013)	(.017)	(.017)	(.014)	(.018)	(.018)	(.013)	(.017)	(.017)
ED f.e.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Age f.e.	No	No	Yes	No	No	Yes	No	No	Yes
BPL f.e.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
R^2	.00	.14	.15	.00	.15	.16	.00	.14	.15
N	40809	40255	40252	38080	37552	37549	42878	42383	42380

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
The dependent variable is an indicator that equals to one if the immigrant is a U.S. citizen in 1940.
For Columns (1)–(3), we include in our estimation sample only those immigrants who hold surnames that are unique within their block.
For Columns (4)–(6), we exclude from the estimation all immigrants who are boarders, roomers, and lodgers.
For Columns (7)–(9), we exclude all immigrants who are domestic workers.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.

5 Investigating the mechanisms

In this section, we perform an additional analysis with the aim of pinpointing the channels behind our results. Two possible mechanisms could be at play in this study. First, the co-ethnic social network may help the immigrant become a U.S. citizen because of *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 potential citizen 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 language skills of immigrants not only negatively affect their labor-market outcomes (Auer, 2017, Arendt et al., 2020) but also

[17] 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.

the health of their new-born child (Auer and Kunz, 2021). Indeed, newly arrived immigrants do not clearly understand the way the labor market or the health system operates in the new country, and, therefore, communication barriers are the main explanations for these negative outcomes. We believe that something similar could have happened for the newly arrived immigrants in the U.S. in 1930, especially for those who came from non-English speaking countries. Further, as explained in Section 2, to petition an admission to citizenship, one needed to have two American citizens as witnesses who could testify the immigrant's continuous residence and "good moral character." In Figure 3, 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.

Second, it is possible that *social norms* play an important role in the naturalization process. 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 in which the fraction of naturalized co-nationals is high, 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 regards 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, etc.).[18] 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 11. Once sorting is accounted for (Columns (3) and (6)), we see that networks positively impact the naturalization propensity of immigrants from *non-English-speaking countries only*. Moreover, for immigrants from English-speaking countries, we find that not only is the relationship between networks and naturalization statistically insignificant but also is much smaller in magnitude. This test, therefore, suggests that the *acquisition of information* regarding the naturalization process through close contacts was more important than social norms for im-

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

migrants who faced stronger barriers. In other words, the immigrants that faced language and cultural barriers (such as the Russians) were likely to be helped by other Russians who had arrived before them and were already naturalized in 1930; these co-nationals provided them with information about the naturalization process. As discussed in Section 2, it is a complicated process that involves two steps – a declaration of intention, and a petition for citizenship – with several residency requirements to be met and documents to be produced; it could therefore be too bureaucratic and complicated for newly arrived immigrants without a social network, especially for those coming from non-English-speaking countries.

Table 11: The effect of networks on political assimilation in 1940, information channel

	Probability of naturalization in 1940					
	non-English speakers			English speakers		
	(1)	(2)	(3)	(4)	(5)	(6)
Share of naturalized population in 1930	.163***	.102***	.104***	.085***	–.050	–.046
	(.015)	(.021)	(.021)	(.022)	(.034)	(.034)
R^2	.01	.16	.17	.00	.24	.25
N	31846	31247	31243	13265	12510	12509
	Probability of Naturalization in 1940					
Share of Americans in 1930	–.040	–.013	–.015			
	(.027)	(.065)	(.065)			
Share of naturalized in 1930, from other countries				.028	–.023	–.029
				(.018)	(.032)	(.032)
ED f.e.	No	Yes	Yes	No	Yes	Yes
Age f.e.	No	Yes	Yes	No	Yes	Yes
BPL f.e.	No	No	Yes	No	No	Yes
R^2	.00	.13	.14	.00	.13	.14
N	46137	45635	45633	45986	45485	45483

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses. In both panels, the dependent variable is an indicator that equals to one if the im migrant is a U.S. citizen in 1940. English-speaking countries include Canada, the UK, Ireland, Australia and New Zealand. *Share of naturalized population in 1930* measures the share of the naturalized co-nationals living in the same block as the immigrant in 1930. *Share of Americans in 1930* measures the share of American-born individuals living in the same block as the immigrant in 1930. *Share of naturalized population in 1930, from other countries* measures the share of the naturalized individuals in 1930 who live on the same block as the immigrant and are from all countries except for that of the immigrant.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

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 11. For 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. This is what we find: 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 this case, we would have expected that immigrants surrounded by American-born individuals might face stronger pressure to Americanize.

As a last test to substantiate our argument that co-ethnic networks provide information, we show in the second panel of Table 11, 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 Italians have no effect on the naturalization propensities of Russian immigrants. This is because naturalized Russian immigrants can explain in the same language and with the same cultural reference the complicated process of naturalization to a newly arrived Russian immigrant residing on the same block. Further, they are more likely to be friends and to socially interact with each other than with immigrants from other countries.

To conclude, the results in Table 11 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.

Having found suggestive evidence that information is the key mechanism behind our results, we conclude this analysis by showing *which type* of information might have mattered. We perform three heterogeneity tests to identify the groups of migrants for which we think information played a more essential role in becoming an American citizen. Our conjecture is that there exist groups that either faced stronger barriers to integration or had greater incentives to integrate once they became aware of the possible gains of naturalization.

First, we hypothesize that information matters more for immigrants whose mobility was reduced by the introduction of the immigration restrictions. In particular, by changing the quota allocations by country, the Immigration Act of 1924 lengthened the stay of immigrants (Greenwood and Ward, 2015) and, consequently, might have impacted their probability to naturalize.[19] We test whether immigrants with larger networks are more likely to naturalize when they come from countries that were affected by the migration restrictions. If networks provide useful information on naturalization, one would expect that immigrants who interact with a larger number of naturalized contacts might have gained more awareness about these

[19]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).

constraints. 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. We report the results of the two regressions in the first panel of Table 12. The point estimates suggest that networks had a larger impact on the naturalization rates for immigrants who were more severely restricted by the quotas. At the same time, these estimates are not statistically dissimilar from the less restricted group, suggesting that the type of information that the network conveys might not be necessarily related to overcoming migration restrictions.

Table 12: The effect of networks on political assimilation in 1940, different incentives

	To circumvent immigration restrictions					
	Low restrictions			High restrictions		
Share of naturalized population in 1930	.125***	.048***	.050***	.180***	.088**	.094**
	(.014)	(.018)	(.018)	(.027)	(.044)	(.044)
R^2	.00	.15	.16	.01	.30	.32
N	35148	34557	34556	9963	8983	8977
	To circumvent labor market restrictions					
	Low share in barred occupation			High share in barred occupations		
Share of naturalized population in 1930	.115***	.092***	.092***	.195***	.206***	.210***
	(.015)	(.018)	(.019)	(.025)	(.035)	(.039)
R^2	.00	.20	.18	.00	.20	.18
N	24820	24820	24085	20291	20291	19674
	To activate politically					
	Elected a Republican			Elected a Democrat		
Share of naturalized population in 1930	.060	.059	.066	.179***	.174***	.176***
	(.072)	(.067)	(.072)	(.018)	(.020)	(.023)
ED f.e.	No	Yes	Yes	No	Yes	Yes
Age f.e.	No	Yes	Yes	No	Yes	Yes
BPL f.e.	No	No	Yes	No	No	Yes
R^2	.00	.12	.14	.01	.13	.13
N	3246	3246	3224	25080	25080	24859

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses. *High restrictions*: Immigrants from the countries whose quotas' tightness is above the top 75th percentile of the distribution. *Low restrictions*: Immigrants from the countries whose quotas' tightness is below the top 75th percentile of the distribution.
The *low share in barred occupations* is a subsample of immigrants who, in 1930, were living in a network where the employment share in barred occupations was below the median. The *high share in barred occupations* is a subsample of immigrants who, in 1930, were living in a network in which the employment share in barred occupations was above the median.
Elected a Republican is a subsample of immigrants living in enumeration districts that elected a Republican to the congress in 1930. *Elected a Democrat* is a subsample of the immigrants living in enumeration districts that elected a Democrat to the congress in 1930.
* $p < .10.$; ** $p < .05.$; *** $p < .01.$

Another possible area in which the acquisition of information is important concerns the possibility of overcoming the barriers accessing of the labor market. We identify occupations that were barred for non-citizens throughout the studied time period. The barred occupations include not only professional occupations, as one might expect, but also jobs such as

chauffeurs, steam boiler operators, hunters, peddlers, or bank directors. We obtained the list of barred occupations from Fields (1933). Next, we split the sample among immigrants who lived in a block in 1930 with an above-median share of naturalized co-nationals employed in barred occupations and those who lived in 1930 in a block with a below-median share of naturalized co-nationals employed in barred occupations. The rationale for this test is similar to the one above: If networks provide information on possible labor market opportunities that are reserved for citizens, we should find stronger effects only for those immigrants who, in 1930, were surrounded by neighbors who held such occupations and were aware of such benefits. The results are displayed in the second panel of Table 12. While networks are positively associated to larger probabilities to naturalize among both groups, we do find that networks have stronger effects for those immigrants who belong to a network in which the concentration of people working in barred occupations was larger. This finding is indicative of the importance of labor market information as an additional incentive to naturalize.

Last, networks might provide information regarding the importance of political participation, especially in those areas where, in the previous congressional election, Democratic representatives have been elected, who are traditionally more pro-migration than their Republican counterparts. 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 12. As expected, we find that networks have a stronger and statistically significant impact only where a Democrat was elected.

To summarize, we started this section with the aim of providing a mechanism behind our main results. The evidence in Table 11 points to the provision of information being a key reason why a larger share of naturalized co-nationals induces higher naturalization rates. In particular, networks only positively impact the naturalization propensity of migrants from non-English-speaking countries. We also did not find any effect of neither the share of American-born individuals nor of the naturalized citizens from other countries on the probability of becoming a U.S. citizen. This suggests that only immigrants from the same country of birth of the newly arrived, who went through the tedious naturalization process, are more likely to help their co-nationals to apply for U.S. citizenship. Next, we have shown that this mechanism is particularly important for groups for which information about specific barriers or opportunities was more apparent. Our results suggest that the naturalized co-nationals might have turned to naturalization as a way of overcoming barriers in the labor market and accessing jobs that were otherwise barred for immigrants. At the same time, the naturalized co-nationals 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.

6 Conclusion

Are ethnic enclaves good or bad? This is the question we asked at the beginning of this paper, which is of paramount importance for policies. Here, we focus on assimilation as the outcome[20] and measure the assimilation of immigrants by their naturalization. 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.[21] 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 "hot" and controversial debates, in particular from immigration opponents 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 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 first five 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 them 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 complicated, 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 "quality" of ethnic enclaves on the assimilation of newly arrived

[20] In the Introduction, we have seen that other outcomes have been considered with regard to the impact of ethnic enclaves, in particular, on labor-market outcomes.

[21] 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).

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 (2019). The effects of immigration on the economy: Lessons from the 1920s border closure. NBER Working Paper No. 26536.

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

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,* forthcoming.

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 U.S. naturalization patterns. *Review of Economics of the Household,* forthcoming.

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. Kunz (2021). Communication barriers and infant health: Causal evidence from the swiss language roulette. Unpublished manuscript, Monash University.

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.

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,* forthcoming.

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.

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 effects accounting for network changes. *Review of Economics and Statistics,* forthcoming.

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. *Review of Economic Studies,* forthcoming.

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.

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.

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.

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.

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.

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.

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 interms 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.

In Table A1, we report the summary statistics of the matched and unmatched observations. Several interesting patterns emerge. Immigrants in our sample tend to come from a more disadvantage socio-economic background. For example, they are more likely to be married and be living in larger families and, at the same time, have slightly lower occupational income scores and a higher probability of renting their place for a lower monthly rent. The share of Italian and German immigrants is comparable across the two groups, while, in our sample, we tend to include a larger share of British and Scandinavian immigrants. The differences between the matched and unmatched samples are unsurprising. The risk of introducing measurement error in the matching procedure will tend to induce attenuation bias in our main results. Yet, if anything, our conclusions could be a lower bound of the true effect.

Finally, to further substantiate whether sample selection is likely to be a big concern in our context, we perform an additional check in Table A2. Here, we correlate the probability of being matched with the share of naturalized immigrants in the same network, and we look at this correlation across NYC and within the enumeration district and birthplace. Interestingly, as we compare people of the same nationality and same neighborhood, high shares of

naturalized co-nationals are associated with a lower probability of being matched. This could imply that the results in our sample might be underestimated once we include immigrants with larger networks and better socioeconomic outcomes.

Table A1: Summary statistics of matched and unmatched observations

	(1) Matched obs		(2) Unmatched obs	
	Mean	SD	Mean	SD
Marital status, 1930	0.450	0.498	0.369	0.483
Family size, 1930	3.533	2.167	3.045	2.213
N. of children, 1930	0.524	1.091	0.416	0.991
Ireland	0.155	0.362	0.181	0.385
Italy	0.097	0.296	0.095	0.293
Germany	0.332	0.471	0.222	0.415
Poland	0.061	0.240	0.058	0.234
Russia	0.039	0.194	0.058	0.233
UK	0.127	0.334	0.077	0.266
Scandinavia	0.050	0.218	0.065	0.246
Other EU	0.078	0.268	0.107	0.309
Other country	0.060	0.237	0.138	0.345
Occupational income score, 1930	10.056	12.807	10.512	12.102
House value (in 10,000), 1930	2.122	5.984	3.074	10.150
Is renting, 1930	0.897	0.304	0.879	0.326
Monthly contract rent, 1930	73.989	353.145	89.251	396.602
Observations	46175		233358	

Table A2: Matching probability and networks

Share naturalized in 1930	.050***	–.017***	–.012***
	(.004)	(.005)	(.005)
ED f.e.	No	Yes	Yes
BPL f.e.	No	Yes	Yes
Age f.e.	No	No	Yes
R^2	.00	.05	.06
N	276026	276026	275886

NOTES: The robust standard errors clustered at the enumeration district level are presented in the parentheses.
* $p < .10$.; ** $p < .05$.; *** $p < .01$.