Discussion Paper Series

IZA DP No. 18931

September 2026

Networks, Diversity, and the Economic Performance of Migrants: Evidence from the Pontine Marshes (1932–1941)

Frédéric Docquier
LISER and IZA@LISER

Alice Dominici
IMT School, Bocconi University (INSPIRE) and IZA@LISER

Martín Fernández
LISER, IZA@LISER and RFBerlin

Fabio Mariani
IRES/LIDAM, UCLouvain and IZA@LISER

IZA DP No. 18931 September 2026

Networks, Diversity, and the Economic Performance of Migrants: Evidence from the Pontine Marshes (1932–1941)*

Abstract

This paper exploits a historical quasi-natural experiment to provide causal evidence on how networks and diversity affect migrants' economic outcomes in a rural setting. Between 1932 and 1941, Italy's Fascist government relocated about 4,000 families to the Pontine Marshes, a formerly malarial and sparsely inhabited area undergoing a massive reclamation project. The settlers came from different provinces and were quasi-randomly assigned to plots of land, providing plausibly exogenous variation in neighborhood composition. Using detailed data on the universe of settlers, we find that larger networks of common-origin neighbors significantly increase the likelihood of long-term settlement and land purchase. Greater neighborhood diversity leads to worse economic performance on average, largely by reducing the scope of network externalities, but it benefits isolated settlers who lack common-origin ties. Network effects operate, at least partly, through increased productivity, with no evidence of knowledge transmission as the underlying mechanism. Instead, our results point to cooperation and mutual support as the main channel. The effects are concentrated among families under the greatest strain and those from origins historically more reliant on bonding social capital. Network effects are also stronger when neighbors arrive at the same time and live nearby.

JEL classification

D85, J24, J61, N54, O13, O15, O40, R23

Keywords

migration, networks, diversity, productivity, agriculture

Corresponding author

Martín Fernández
martin.fernandez@liser.lu

* We thank Philipp Ager, Guido Alfani, Esther Arenas-Arroyo, Sam Bazzi, Sascha Becker, Francesco Ferlenga, Oded Galor, Andrea Ichino, Stelios Michalopoulos, Louis Putterman, Hillel Rapoport, Alexander Yarkin, and participants in the workshop on "Religion, Culture and Economic Growth in Historical Perspective" in Louvain-la-Neuve, the 15th Conference on Migration and Development, the 97th WEAI Annual Conference, the 43rd ASSET Annual Meeting, the 48th Meeting of the Spanish Economic Association, the 2024 SEHO Meeting, DATINI-ESTER Conference, as well as seminar participants in Brown University (Growth Lab), Clark University, ENS Paris-Saclay, LISER, LUISS, Sapienza University, University of Cagliari, University of Vienna, Paris School of Economics, University of Mannheim, University of Valencia, and THEMACergy for their helpful comments. We extend our gratitude to Giulio Alfieri for generously sharing his data with us and providing valuable insights on the historical context of our study. We also thank Sascha Becker and Mario Carillo for granting us access to their respective datasets. Fabio Mariani acknowledges the financial support of the FNRS (Crédit de Recherche J.0132.22).

1 Introduction

Migrants do not settle in isolation but often in groups. Facing challenging conditions, their economic success depends crucially on who their neighbors turn out to be and what they can provide, such as relevant knowledge, but also trust, reciprocal support, and insurance against adverse circumstances. A large literature documents positive effects of common-origin networks on migrant outcomes and emphasizes information transmission as a central mechanism (Massey 1985, Munshi 2003, Edin, Fredriksson, and Åslund 2003, Beaman 2012). At the same time, a wider mix of origins can raise productivity by broadening the range of skills and experience available locally (Ashraf and Galor 2012, Bazzi et al. 2016, Alesina, Harnoss, and Rapoport 2016, Posch, Schulz, and Henrich 2026). Because common-origin networks and diversity are intertwined by construction, which of these forces dominates, and through which mechanisms, remains an open question.

Understanding this trade-off is particularly relevant for contemporary displacement. More than 117 million people are forcibly displaced worldwide, and refugees are often resettled in rural areas (UNHCR 2026, Maystadt and Verwimp 2014, Ruiz and Vargas-Silva 2015). In regions facing increasing climate pressures, migration may increasingly involve the creation of new rural communities around agricultural land rather than movement into established urban labor markets (Clement et al. 2021). The social composition of these communities may therefore shape migrants' ability to adapt and prosper. Yet, despite the importance of this question, the evidence on the effects of neighborhood composition comes primarily from urban labor markets (Edin, Fredriksson, and Åslund 2003, Beaman 2012), where networks facilitate job matching and information transmission. Much less is known about rural economies, in which formal labor markets are thin and local interactions may instead shape economic performance through cooperation, resource sharing, and social capital. Moreover, common-origin networks and diversity have largely been studied separately, leaving their interconnection poorly understood.

This paper studies the causal effects of neighborhood composition (network size and diversity) on migrants' economic success using a historical quasi-natural experiment in interwar Italy. Between 1932 and 1941, the Fascist government resettled approximately 4,000 families to the Pontine Marshes, a formerly malarial and sparsely inhabited area situated about 60 km southeast of Rome. After extensive reclamation works, the available 51,000 hectares of land were progressively partitioned into plots and allotted to settlers from rural areas in various Italian provinces, who were attracted by the prospect of land ownership.

This setting is particularly appropriate for our research question. First, the settlement was centrally planned, and families were quasi-randomly assigned to farming plots, generating plausibly exogenous variation in neighborhood composition. Second, newly digitized and georeferenced administrative records covering the universe of settler families allow us to measure household outcomes and the composition and spatial structure of their neighborhoods at a highly granular level.[1] Third, the economic environment provides a unique opportunity to explore the channels through which networks may operate. Settlers worked under sharecropping arrangements, could not choose their crops or trade in local markets, and were paid in credits recorded in accounts held by the state agency rather than in cash. Because migration was exclusively internal, institutional

[1] These data come from the State Archives in Latina, and were partly digitized by the authors and partly obtained from Alfieri (2018).

and legal differences that sometimes affect migrants of certain origins differently are here absent, while the cultural distances that shaped social interaction remain. Although historical, the setting shares important features with contemporary rural migration: the establishment of new communities, the absence of formal labor markets, and the reliance on local interactions.

We measure economic performance with two distinct yet interrelated outcomes: whether settlers remained in the Pontine Marshes until 1941 (*i.e.*, the year the government started selling the plots), and whether they ultimately purchased their allotted plot. Remaining until 1941 captures economic resilience in a harsh and closely supervised environment, while land purchase captures long-term economic potential and investment capacity.[2] Of all settler families, 28% left or were expelled before 1941, and only 53% eventually purchased their plot. We measure network exposure using the share of nearby families originating from the same province and construct a measure of neighborhood diversity based on neighbors' origin provinces.

As the quasi-random assignment of settlers to plots is the cornerstone of our identification strategy, we support this assumption by gathering historical evidence on the characteristics of the allocation process, and conducting extensive tests on the available data. We establish that network size and diversity (as well as other neighborhood features) are uncorrelated with a wide array of family and plot characteristics. We also find that the *ex post* composition of a settler's neighborhood closely matches what would be expected from the origins of families arriving in that month, and deviations from it do not appear to correlate with observable household characteristics. We estimate the effects of networks and diversity using OLS and including fixed effects that absorb several dimensions of the settlement process, including year of arrival, region of origin, and administrative area.[3]

Our core results show strong evidence of positive network effects. Increasing the share of common-origin neighbors living within one kilometer from 0 to 1 (all of them) raises the probability of remaining in the Pontine Marshes until 1941 by 18.1 percentage points, and the probability of purchasing the land by 16.6 percentage points. These effects are substantial, representing improvements of 25% and 31% relative to the average outcomes. For comparison, settlers on the most fertile soils were 22 percentage points more likely to remain than those on the least fertile—soil quality being the most important observable determinant of success in this environment. Hence, the network effect is worth about four-fifths of that gap. The effects of diversity are more nuanced. Considered in isolation, diversity decreases the chances of economic success, but its average effect dissipates after accounting for network size. Yet, among the roughly 20% of settlers without same-origin neighbors, greater diversity is associated with better outcomes. An increase from zero to maximum diversity raises the probability of staying by 16.9 percentage points and of purchasing by 19.5 percentage points, magnitudes comparable to the network effects. Diversity may therefore partially compensate for the absence of a common-origin network, suggesting that the value of neighborhood composition depends on migrants' access to alternative sources of social support.

The granularity of our data allows us to investigate what neighbors actually supplied. We proceed in three steps, asking what made ties stronger, whether they operated through productivity rather than amenities, and what they transmitted. The evidence consistently points to cooperation

[2] These two outcomes are strongly correlated with soil quality and other measures of agricultural productivity.

[3] The results are stable across specifications that include plot and origin-municipality controls as well as successive layers of fixed effects. We also test the robustness of the results to alternative buffer sizes and definitions of common origin.

rather than knowledge transmission. First, shared early experiences with common-origin families appear important, as arriving in the same month as other families from the same province generates large positive effects.[4] This result is consistent with the idea that bonds of trust and cooperation can be established during critical moments of the settling experience. Second, the evidence suggests that networks operated through productivity rather than amenities alone. Families embedded in more extensive networks are significantly less likely to be expelled for failing to meet the production quotas. Third, we exploit detailed climatic and geological data and find no evidence of a transmission of agricultural knowledge, either from network families or from other neighbors.[5]

The evidence for cooperation and solidarity comes from several patterns. Exploring the spatial features of the network, we find that what matters most is physical proximity rather than how tightly interconnected the network is. In fact, distance to the nearest same-origin family predicts economic performance, especially when that family lives within walking distance, and having no occupied plot next door is harmful even conditional on network size. The value of a network also increases with a family's need for assistance, with larger effects among settlers facing greater demographic pressure and those less familiar with the local environment. Importantly, the positive effects of diversity among settlers with no same-origin neighbors are also concentrated in precisely these most vulnerable groups for which cooperation mattered the most. Finally, network effects are larger for settlers from origins historically reliant on bonding rather than bridging social capital. We find that network effects change discretely for origin municipalities surrounding the Po river, where a standard proxy for bonding ties (*i.e.*, the prevalence of consanguineous marriage) also changes sharply. In sum, while prior work on rural environments emphasizes the information value of networks (Foster and Rosenzweig 1995, Groeger and Zylberberg 2026), in a setting with little room for learning and knowledge diffusion, our findings indicate the fundamental role played by cooperation and mutual support.

Our paper intersects with three strands of existing literature. First, we contribute to a large body of research examining the effects of migrant networks on their economic success (Massey 1985, Edin, Fredriksson, and Åslund 2003, Munshi 2003, Comola and Mendola 2015, Battisti, Peri, and Romiti 2022). In particular, several studies leverage the quasi-random allocation of refugees across locations to uncover a positive impact of co-ethnic networks on migrant labor-market outcomes—mainly through a more efficient transmission of information (Edin, Fredriksson, and Åslund 2003, Damm 2009, Beaman 2012, Martén, Hainmueller, and Hangartner 2019).[6] Although not primarily concerned with migration, some research explores the informational role of social networks in driving technology adoption in rural contexts (Munshi 2004, Conley and Udry 2010, Maertens and Barrett 2013, Mekonnen, Gerber, and Matz 2018, Beaman et al. 2021). Our work bridges the gap between these two sets of papers by providing causal evidence on the consequences of migrant networks in a rural, subsistence-based agricultural economy with limited scope for market interactions. We also provide evidence that allows us to distinguish the cooperative role of networks from knowledge transmission, and shed light on the formation and spatial structure of migrant networks.

[4] This should not be taken to suggest that these families knew each other beforehand, as most came from different and distant municipalities within the province.

[5] We use granular historical rainfall and temperature data, together with other geographical variables, to construct measures of relevant agricultural knowledge, such as agroclimatic similarity and origin suitability for various crops.

[6] See also Blumenstock, Chi, and Tan (2025), Patacchini and Zenou (2012), and Elsner, Narciso, and Thijssen (2018).

Second, we contribute to the literature on the economic consequences of diversity. At the aggregate level, seminal work by Ashraf and Galor (2012, 2013a, 2013b) and subsequent research on migrants' diversity relate the origin composition of populations to long-run development, with the trade-off between skill complementarities and cooperation at their core (Alesina, Harnoss, and Rapoport 2016, Bazzi et al. 2016, Minale, Rocha, and Vigna 2024). At the local level, some studies highlight the effects of neighborhood diversity on public goods and labor-market outcomes (Algan, Hémet, and Laitin 2016, Hémet and Malgouyres 2018). We connect these two levels, measuring diversity and networks locally within a single framework and in a subsistence economy. Observing both simultaneously allows us to distinguish the effects of diversity from those of the absence of common-origin ties and to examine how diversity shapes outcomes when settlers lack such ties. Our findings support the hypothesis that diversity can hinder economic performance in subsistence-based, Malthusian economies, as speculated by the Unified Growth Theory literature (Ashraf and Galor 2012, Mariani, Mercier, and Pensieroso 2023).[7] We also unveil a novel benefit of diversity: for settlers without any common-origin neighbors, a diverse neighborhood partly substitutes for the missing network.

Lastly, our study relates to the growing literature on how the composition of newly founded communities can shape their economic and social trajectories. Prior work has traced the formation and persistence of norms on the US frontier (Bazzi, Fiszbein, and Gebresilasse 2020, Haddad 2026, Bazzi et al. 2026), the emergence of national identity in new rural settlements in Indonesia (Bazzi et al. 2019), and political attitudes in the settlements created by the Fascist government in Italy, including those of the Pontine Marshes (Carillo 2022, Belmonte 2026).[8] Instead of looking at communities as a whole, we exploit the quasi-randomness in who a settler's immediate neighbors were, at the moment the community was formed, to assess what this was worth to them in terms of economic resilience in a harsh environment.

The remainder of the paper is organized as follows. Section 2 provides historical background. Sections 3 and 4 describe the data and the empirical strategy. Section 5 presents the main findings, and Section 6 provides evidence on the potential mechanisms. Section 7 concludes.

2 Historical Background

The Pontine Marshes (*Agro Pontino*) are located southeast of Rome, in the Lazio region. The area is mostly in the modern province of Latina and extends 45 km along the coast, covering approximately 80,000 hectares. For centuries, the marshes were an inhospitable region, plagued by malaria and unsuitable for agriculture due to their swampiness. From the Roman Empire to the Papal States, every attempt to eradicate malaria and drain the marshes had failed, as the available technology could not sustain drainage in the long term or support permanent human settlements (de Haas and Schepers 2022). At the beginning of the 20th century, the marshes were sparsely populated, with only a few hundred isolated shepherd families and nomadic groups (Gentilcore 1970).[9]

[7] This result stands somewhat in contrast with Baggio and Coşgel (2026), who identify a U-shaped effect of racial diversity on economic performance in a specific pre-industrial sector (*i.e.*, whaling).

[8] Relatedly, Muñoz-Blanco (2026) studies the long-term political consequences of internal displacement under the Spanish dictatorship.

[9] The absence of an indigenous population in our context is a feature shared by Brown and Cuberes (2023).

In 1924, the Fascist regime launched a large-scale reclamation project known as *Bonifica Integrale*. This program had two goals: to showcase an unparalleled technological achievement for propaganda purposes, and to expand wheat production as part of Mussolini's *"Battle for Grain"*. By 1930, the marshes had been drained, the terrain transformed into arable land, malaria systematically combated, and an intricate network of canals and roads put in place.[10] The available land, amounting to approximately 51,000 hectares, was then divided into plots. Each plot, referred to as *podere*, featured a two-story farmhouse intended to accommodate the forthcoming settlers. From 1932 to 1941, approximately 4,000 families occupied the plots that were progressively made available, creating the quasi-natural experiment at the heart of our study. The settlers initially worked under sharecropping arrangements.[11] From 1941 onward, as the government sought to raise funds for the war effort, they were granted the option to redeem the land, the alternative being a rental contract or leaving.

2.1 Settler Recruitment and Quasi-Random Plot Assignment

The resettlement program in the Pontine Marshes, with its centralized assignment of plots, provides a quasi-natural experiment in neighborhood composition. The selection of settlers by province of origin resulted from a combination of supply and demand forces.[12] On the demand side, the Fascist regime recruited people with significant experience as rural laborers under sharecropping contracts. In line with the government's pro-natalist stance, the inflow of high-fertility families was strongly encouraged. On the supply side, the settlement of the Pontine Marshes coincided with a period of economic turmoil in the aftermath of the Mussolini government's *Quota 90* policy, which involved a revaluation of the Italian lira against the British pound, leading to lower agricultural wages and higher rural unemployment. Consequently, the economic conditions of agricultural workers deteriorated substantially, with many moving from sharecropping to day labor. This combination of supply and demand explains why the Po Valley emerged as the main origin of prospective settlers. Figure 1 illustrates the main origins of settlers based on data from Alfieri (2018). Veneto and Emilia-Romagna together account for 68% of the families. Both regions had a high historical prevalence of malaria and were severely affected by the rural crisis. Despite the predominance of settlers from Northern Italy, hundreds of families also originated from areas surrounding the Pontine Marshes and from other parts of the country.[13]

The *Commissariato per le Migrazioni e la Colonizzazione Interna* (CMCI), the government institution responsible for domestic migration, oversaw the logistics of the selection process and plot allocation. Prospective settlers began the process by applying to the CMCI office in their province of origin. The only eligibility criterion was that the head of the household needed to be a war veteran. This was in line with the Fascist propaganda, which depicted settlers as soldiers and the

[10] The construction works employed over 100,000 temporary workers under arduous conditions and persistent malaria exposure. Despite the drainage, malaria remained present until after WWII, when U.S.-led reconstruction efforts eradicated it using DDT. As malaria-related mortality was not reported on death certificates, we use malaria prevalence at origin as a proxy for environmental suitability and malaria resistance in our regressions.

[11] Under sharecropping, settlers inhabited and cultivated the plot in exchange for a share of the produce. While retaining this standard feature, contracts in the Pontine Marshes had indefinite duration, unlike typical sharecropping arrangements, which lasted between one and three years.

[12] The number of families interested in participating in the program generally exceeded the availability of plots. This excess demand varied by year and with the pace of new plot construction. While our preferred specifications always include year fixed effects, their omission does not affect the results.

[13] A smaller share of settlers, though Italian-born, arrived from abroad, chiefly from Yugoslavia and Romania, where they had previously participated in other state-sponsored migration programs.

(a) Arrivals by region

(b) Total families across provinces

Figure 1: Origin of settler families, 1932-1940

Note: own elaboration based on data from Alfieri (2018).

settlement itself as a war effort; affiliation with the Fascist party, however, was not mandatory, and most settlers were not party members (Folchi 2000).

The CMCI reviewed applications on a rolling basis in Rome and assigned families to plots. This centralized decision-making process ensured that plot allocation was not shaped by local favoritism (Boccini and Ciccozzi 2007). The allocation rather followed the logic of the so-called "Tommasi method" with larger plots assigned to families with more working units. [14] As shown in Figure 2, plots became available gradually so that the selection and settlement process occurred in multiple waves, with CMCI authorizing the migration of a predetermined number of families. The settlement initially concentrated in the core of the Pontine Marshes, and later expanded outward toward the periphery as successive waves of settlers arrived – see Figure A4 in the Appendix. If a family was expelled or withdrew, its plot was immediately reassigned to an eligible newcomer family, which typically settled within a year.

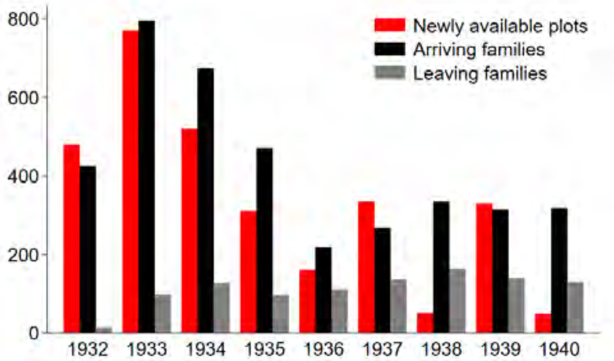


Figure 2: The settlement process, 1932-1940

Note: own elaboration based on data from Alfieri (2018).

The CMCI organized settlers' travel to the Pontine Marshes by train, often grouping families from the same province on specific dates.[15] The journey to the Pontine Marshes was among the shared experiences that fostered strong social bonds, especially given the uncertainty and early

[14] Giuseppe Tommasi, the agronomist who oversaw the agricultural transformation of the Pontine Marshes, proposed that adult men be counted as one working unit, women as 0.5, and children and the elderly as 0.3 (Alfieri 2014).

[15] The journey from Northern Italy to the Pontine Marshes could take as long as 30 hours (*e.g.*, from the station of Verona).

hardships experienced by settlers, as described by Alfieri (2018) and Folchi (2000). Upon arrival, settlers were checked in by the *Organizzazione Nazionale Combattenti* (ONC), the war veterans' agency responsible for supervising activities in the Pontine Marshes, and accompanied to their designated plots.[16]

2.2 Settlers' Living Conditions

The ONC was responsible for organizing and monitoring agricultural production in the Pontine Marshes. It supplied seeds, fertilizers, productive assets, and provided bookkeeping, technical aid, and training. ONC agents also enforced strict monitoring of daily activities in the plots. For administrative convenience, the 3,000 plots were grouped into 18 agricultural *aziende*. Additionally, 28 rural settlements (*borgate rurali*) were created to house ONC agents and provide schooling, healthcare, as well as administrative and religious services to the settlers. Consistent with Mussolini's emphasis on de-urbanization and the promotion of rural living, the settlers could not live in these villages, and formal market interactions were heavily restricted (Gentilcore 1970).[17] The colonization program was designed with production from the plot as the sole source of sustenance, consistent with sharecropping customs.

The settlers entered into strict sharecropping agreements that required them to cultivate specific crops and meet production targets set by the ONC. Despite the soil not being particularly favorable for such crops, the assigned crop mix heavily emphasized wheat and other cereals, as well as forage intended to support livestock.[18] Once the production quotas were met, the settlers could keep the remainder of the harvest for personal consumption and receive compensation for any surplus.[19] ONC agents monitored settlers daily to ensure compliance with the rules, with violations resulting in contract termination and the expulsion of all family members.[20]

The strict contractual conditions, the poor quality of the land, and the eventual resurgence of malaria made life in the Pontine Marshes extremely difficult. In this challenging environment, support from neighboring settlers was essential. Historical evidence underlines the strong solidarity among settlers, who regularly exchanged labor, tools, commodities, childcare, and essential aid, as documented by Alfieri (2014, 2018) and Folchi (2000). Notably, despite the ONC's discouragement, families often "borrowed" labor from their neighbors to improve the efficiency of sowing and harvesting. These accounts suggest that social bonds and reciprocal trust were fundamental conditions for coping with hardship and achieving economic resilience, and motivate our analysis of cooperation in Section 6.

[16] Luigi Barzini's account in *Il Popolo d'Italia* (October 11, 1933) vividly describes the settlers' arrival in Latina: "When the settlers arrive at the station, [...] an ONC officer checks the composition of the families and provides them the number of their plot. [...] Then, big trucks take the families and their belongings to the farms." See also Folchi (2000).

[17] At first, Mussolini was reluctant to allow urban centers to develop in the Pontine Marshes. Over time, however, "new towns" were founded: Littoria in 1932, Sabaudia in 1933, Pontinia in 1934, Aprilia in 1936, and Pomezia in 1938.

[18] Del Bravo (1956) reports that volcanic and alluvial land was allocated 40% to wheat and other cereals, 40% to legumes and forage, and 20% to set-aside, while for sandy or clay soils the shares were 33% wheat, 50% lucerne, and 17% set-aside.

[19] Rather than being paid in cash, these compensations were recorded in settlers' debt and credit books, resulting in limited monetary circulation.

[20] In particular, controls had to ascertain that (i) all family members worked in their plot, (ii) no outsiders lived or worked in the plot, and (iii) no resources, especially fertilizers, milk, and crops, were hidden or stolen (Folchi 2000).

3 Data and Measurement

We combine archival sources with newly digitized and georeferenced historical maps to construct settler-level measures of neighborhood composition at arrival. This data allows us to link migrants' long-term performance to characteristics of their municipalities of origin and to the features (namely, network and diversity) of their neighborhood in the Pontine Marshes.

3.1 Archival Records and Spatial Mapping

Our primary dataset, which combines information from Alfieri (2018) and ONC records in the State Archives of Latina, covers the universe of settler families. It includes information on settlers' municipality of origin, arrival date, plot number, plot size, family size, and whether they ultimately secured land ownership after 1941. We complement this data with additional ONC archival material documenting voluntary departures, expulsions (including their reasons), and the *azienda* to which each plot belonged. Figure A1 in the Appendix provides an example of the archival data.[21]

We digitize and georeference historical maps of the Pontine Marshes, which display the location of each farmhouse and the characteristics of the terrain (see Figures A2–A3b in the Appendix). The unique plot identifier allows us to link the maps to the available information on the settlers. The house coordinates also enable us to calculate the distances between plots and their proximity to rural settlements. By overlaying the georeferenced plots with the historical ONC map of soil composition, we can classify each plot into one of nine soil types and four levels of soil fertility.

We then combine this data with the geographical and historical characteristics of the 726 municipalities of origin. We use these variables as controls and to shed light on the mechanisms:

- *Geography:* longitude, latitude, average elevation, and wheat suitability (FAO index).
- *Transferability of skills:* following Bazzi et al. (2016), we leverage topographic and climatic data to construct an index of agroclimatic similarity between the municipality of origin and the assigned plot, as well as a wheat-specific agroclimatic similarity index.[22]
- *Adaptability:* since farmers from malaria-endemic areas were particularly fit for the Pontine Marshes and could have developed resilience to the disease, we collect data on historical exposure to malaria from Torelli (1882) (see Figure A5 in the Appendix).
- *Cultural distance:* linguistic similarity between Italian dialects from *AIS Dialectometry* as a proxy for cultural distance between settlers of different origin.[23]

3.2 Economic Outcomes of Settlers

Given the context of the Pontine Marshes, we measure migrants' economic performance using two distinct yet complementary outcomes: (i) staying in the marshes until 1941 and (ii) eventually buying the land. Although the contracts did not specify a precise timeline for land redemption, the

[21] Unfortunately, aside from family size, detailed information on settlers' demographic characteristics and agricultural output is unavailable due to the partial destruction of the ONC archives after World War II.

[22] These indices are computed using average yearly temperature and precipitation data covering 1900-30 from the Climatic Research Unit (CRU) TS4.07, raster data from the FAO Harmonized Soil Database (HWSD), on several soil characteristics, FAO's Global Agro-Ecological Zones (GAEZ) database, as well as data on altitude, altitude variability, and distance from rivers. Further details are provided in Section E of the Appendix.

[23] Based on Jaberg and Jud (1940), available at https://dialektkarten.ch/dmviewer/ais/index.en.html.

prospect of land ownership was the central incentive for settlers to migrate and their ultimate goal (Del Bravo 1956). As mentioned above, sharecropping agreements were terminated in 1941, allowing settlers either to redeem their plot or to remain under new rental contracts. Staying in the marshes until 1941 (*i.e.*, until becoming eligible to purchase land) is a meaningful proxy for economic resilience given the difficulty of meeting production quotas. This is especially true given the military-like oversight imposed by the ONC and the widespread rhetoric suggesting that "leaving meant being a deserter" (Folchi 2000).

Land acquisition constitutes our second measure of economic success. Starting in 1941, settlers could redeem their plot at a quasi-fixed price of 10,200 lire per hectare, plus an additional amount reflecting the value of physical capital and livestock (Rossetti 1994). Although settlers could opt for a rental agreement and purchase the land at a later stage, the uncertainty involved likely encouraged early acquisition. This measure has an important limitation. In some cases, the purchase decision came after the war and was likely shaped by wartime confounders affecting both plots and settler families.[24]

Although settlers had few attractive alternatives (*e.g.*, they lacked access to land elsewhere), 28% left before 1941. A few hundred families were expelled for falling short of production targets or for keeping their plots in poor condition, information that we exploit to shed light on the productivity effects of networks and diversity.[25]

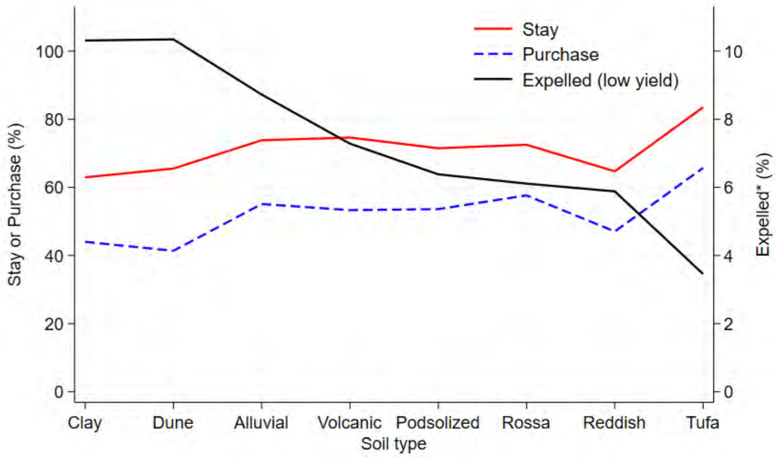


Figure 3: Soil types and economic performance

Notes: The figure shows the different soil types (from least to most fertile) and the average performance of settlers in three different dimensions: (i) whether they stayed until 1941, (ii) whether they purchased (left axis), and (iii) whether they were expelled due to low productivity (right axis).

Figure 3 illustrates the relationship between our proxies for economic performance and soil quality, an arguably fundamental determinant of economic success in the agricultural economy of the Pontine Marshes that is unrelated to network or diversity effects. Among settlers assigned to plots with the least fertile soils (*i.e.*, clay and dune), over 10% were expelled, 61% remained

[24]During World War II (1941–1945), much of the population was temporarily displaced from their homes. Many young men left the Pontine Marshes to join the Italian army. Moreover, between 1943 and 1944, German troops flooded several hectares of land in an attempt to slow the Allied advance. Other plots were bombed or mined, so that 26% of them were heavily damaged.

[25]A small number of families were expelled for immoral or violent behavior, or for suspected communist leanings.

until 1941, and 42% eventually purchased the land. In contrast, among those settling on the most fertile soils (*i.e.*, tufa), fewer than 3% were expelled, 83% remained until 1941, and 65% purchased the land. As expected, soil fertility is a decisive predictor of settlers' success. Substantial residual heterogeneity nonetheless remains, which neighborhood composition may help explain (see Figure 4).

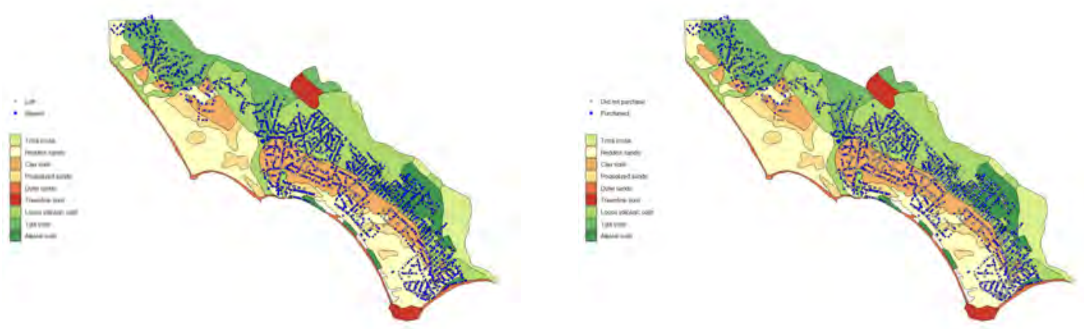


Figure 4: Spatial distribution of settler outcomes and soil types

Notes: The maps show the settler families who (i) remained until 1941 and (ii) purchased the land. Areas shaded in yellow-red tonalities represent poorer soil conditions, while those in blue-green tonalities represent better soil conditions. The sample is restricted to the first occupiers of each plot.

3.3 Neighborhood Composition by Origin

In our analysis, we focus on two distinct yet interrelated dimensions of neighborhood composition: the proportion of neighbors with the same origin (*Network*) and an origin fractionalization index (*Diversity*). We measure network and diversity within one kilometer of each settler family's farmhouse and define origin at the province level (corresponding to NUTS-3 administrative units). As Section 3.3 explains, both choices (*i.e.*, buffer size and provincial origin) are motivated by the specific context under study and by the characteristics of the data.[26]

To obtain measures that are homogeneous and comparable across families, we consider only neighbors present in the year of arrival. For example, if nearby plots are initially empty but are assigned to incoming settlers in subsequent years, these "new" neighbors are not included in our network and diversity measures, and the settler is therefore considered to be initially isolated.

We measure the network of family i as follows:

$$\text{Network}_i = \sum_{j \in B_i^{1km}} \frac{\mathbb{1}(\text{origin}_j = \text{origin}_i)}{N_i},$$

where i identifies the settler family, j identifies other families within the 1-km buffer (B_i^{1km}), and N_i is the total number of families within the buffer excluding family i itself.

We measure diversity around family i using a fractionalization index based on the province of origin within a 1-km radius buffer. The index measures the probability that two randomly drawn

[26] We nonetheless perform extensive robustness analyses considering possible alternatives along both margins.

families come from different provinces (*i.e.*, it is equal to one minus the Herfindahl–Hirschman index of concentration), and is given by:

$$\text{Diversity}_i = 1 - H_i.$$

Here H_i is the sum, across origin provinces p, of the squared shares of family i's neighbors originating from p

$$H_i = \sum_p \left(\frac{\sum_{j \in B_i^{1km}} \mathbb{1}(\text{origin}_j = p)}{N_i} \right)^2.$$

The fractionalization index is the most widely used measure of diversity in the related literature (see, for example, Ashraf and Galor 2013b, Alesina, Harnoss, and Rapoport 2016, Bazzi et al. 2016, and Docquier et al. 2020). However, we also verify the robustness of our results using polarization as an alternative measure (Montalvo and Reynal-Querol 2005). We also calculate an additional index ($\tilde{H}_i$) to measure diversity among non-network members within the neighborhood using the same formulas described above.

We measure neighborhood composition using 1-km-radius buffers, as this choice best reflects the structure of social interactions in the Pontine Marshes. Smaller buffers would overlook the ONC's deliberate design of sparse settlements, while larger buffers would overestimate the scope of feasible social interactions in a low-mobility environment. On average, a 1-km buffer contains 12 neighboring families. A related paper on rural networks (Groeger and Zylberberg 2026) provides evidence that most interactions occur among households living within one kilometer of one another.

The choice of province as our benchmark indicator of origin is motivated by three main considerations. First, settlers typically self-identified by province, reflecting common sociocultural norms and a sense of shared destiny.[27] Cultural proximity was closely tied to dialect similarity, and dialect boundaries often coincided with provincial borders.[28] Second, provinces were, at the time, Italy's most salient administrative units, with the Fascist government creating 17 new provinces to meet demands for local identity.[29] Third, from a measurement standpoint, using provinces as the identity marker avoids extreme values (0 or 1) that arise when networks are defined at the municipal or regional level.

As will be detailed later in the text, we also assess the robustness of our results to alternative buffer sizes (ranging from 500 m to 1.5 km) and to alternative definitions of origin (regions, municipalities, and artificial fixed-size areas). These tests confirm that a 1-km buffer best captures the relevant social interactions in our setting and that the province constitutes the primary identity marker, yielding remarkably stable estimates.

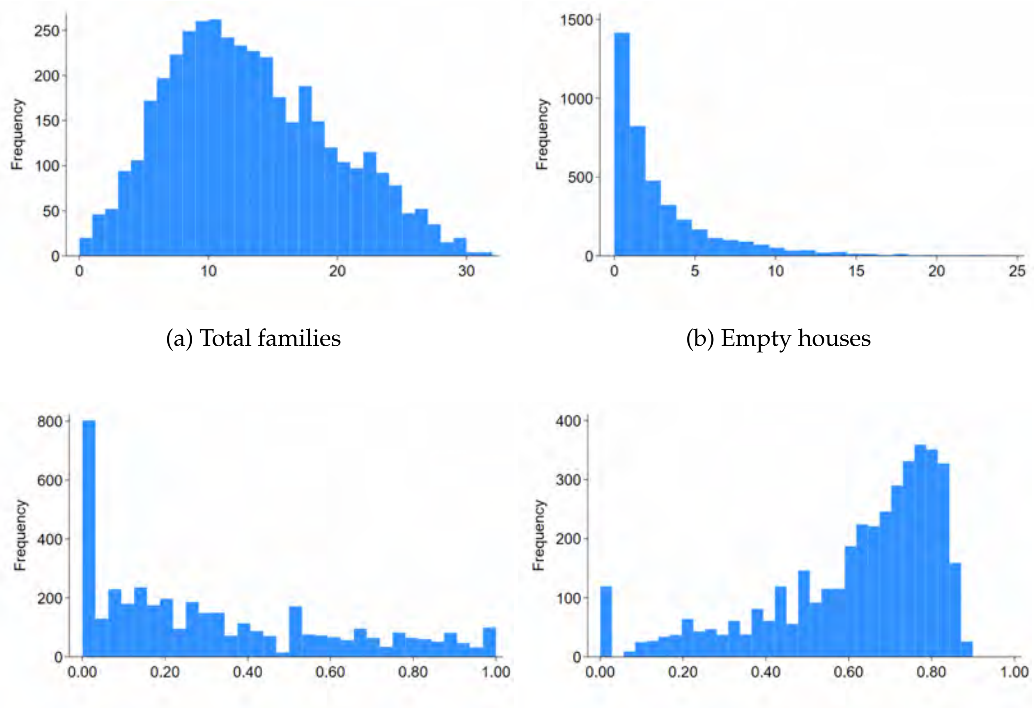
[27] Rossetti (1994, 2008) documents extensively the case of the *Ferraresi*, *i.e.*, settlers originating from the province of Ferrara.

[28] In this respect, we find that, once we control for neighbors' provinces of origin, linguistic distance has no additional influence on our outcomes.

[29] The ONC recruitment program emphasized provincial affiliation. For example, while families could bring in additional members, these members could not originate from a different province (Folchi 2000).

3.4 Sample and Descriptive Statistics

Our sample comprises the universe of settlers who arrived between 1932-1940, namely 3,784 families. Table C1 reports descriptive statistics for the sample.[30] Around 40% of settlers came from areas with high malaria prevalence. They originated from highly diverse environments, yet with ecological conditions generally unsuitable for wheat cultivation. The climatic and soil characteristics of the Pontine Marshes differ, on average, by 1.72 standard deviations from those of settlers' municipalities of origin. This indicates a substantial mismatch between settlers' familiar environments and the new agroclimatic conditions.[31]



(a) Total families

(b) Empty houses

(c) Network

(d) Diversity

Figure 5: Neighbors in the buffer, networks, and diversity

Notes: The figure shows the distribution of the number of families, empty houses, network size, and diversity, within a 1-km radius at the time of arrival. "Network" represents the proportion of families in the buffer from the same province of origin. "Diversity" represents the fractionalization index of the provinces of origin in the buffer.

Neighborhood composition exhibits substantial heterogeneity. Figure 5 shows the distribution at arrival of key variables measured within a 1-km buffer, including the number of families, empty houses, and our measures of network and diversity. The average number of families is 12, with an inter-quartile range of 8–17. Upon arrival, 10% of houses within the buffer were empty, and 15% of settlers had an empty house as their closest neighbor. Network size varies widely. On average, 33% of neighbors come from the same province, but nearly 20% of families had no neighbors from the same province, indicating complete isolation. Conversely, fewer than 3% of families lived in neighborhoods where all neighbors shared their provincial origin. Diversity also displays a broad

[30] We do not include the few hundred settler families who arrived during 1941 or afterwards because many of them could enter different rental agreements and already purchase the land directly.

[31] The agroclimatic similarity index is the average standardized difference between the municipality of origin and the assigned plot across several dimensions, including soil composition, average temperature, and precipitation. See Section E.

range, with a mean of 0.61 and an inter-quartile range of 0.50–0.77. The right-skewed distribution suggests that many settlers lived in highly diverse neighborhoods, often surrounded by families from different parts of Italy.

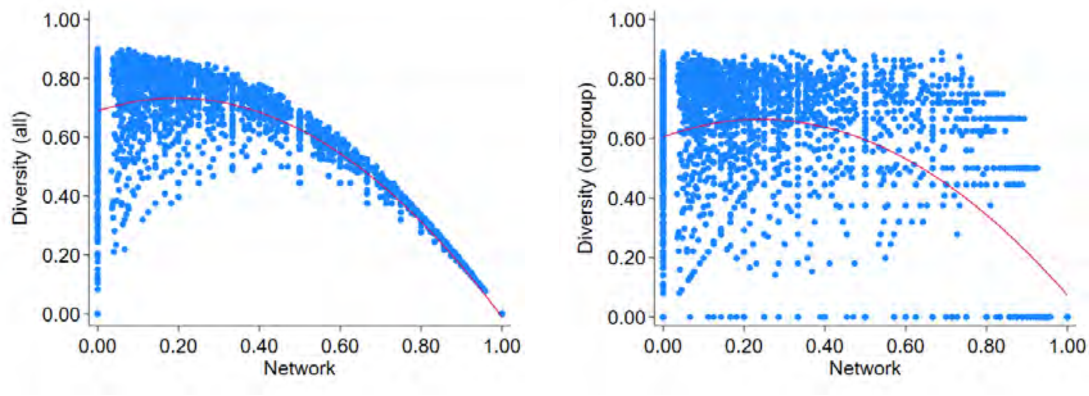


Figure 6: Relationship between network and diversity

Notes: Network represents the share of families from the same province of origin in a 1-km radius buffer. Diversity represents the fractionalization index of origin provinces in the buffer, considering all families on the left, and only the outgroup (non-network) families on the right.

Network size and diversity are mechanically intertwined. As the proportion of neighbors from the same origin increases, a family's exposure to diversity decreases, unless the initial size of the network is very small and network expansion contributes to diversity. The left panel of Figure 6 illustrates this pattern. In particular, as the network approaches 100%, overall diversity $(1 - H_i)$ converges to zero by construction. The Pearson correlation coefficient between network and diversity is -0.73, confirming that the two measures capture related but distinct dimensions of neighborhood composition. Although weaker, this pattern also applies to diversity measured only among outgroup families $(1 - \tilde{H}_i)$, as depicted in the right panel of Figure 6. In fact, a larger network implies fewer outgroup families in the buffer, and thus less scope for diverse origins among them. When the network is small, overall and outgroup diversity almost coincide (the correlation is 0.94 below the median network size).

Overall, these descriptive patterns show that settlers entered highly heterogeneous local environments spanning sparsely populated areas, strong provincial clusters, and highly diverse social settings. They also underscore the importance of distinguishing between network size and diversity to better understand the mechanisms operating in a subsistence agricultural context.

4 Empirical Strategy

We are interested in estimating the effect of neighborhood composition at arrival (network size and diversity) on settlers' economic outcomes. Our empirical approach leverages the quasi-random allocation of settlers to plots, which creates exogenous variation in the origins of nearby neighbors. We start by presenting our estimating equation, and then discuss the identification assumptions.

4.1 Estimating the Impact of Neighborhood Composition

We estimate the following equation by OLS:

$$y_{irta} = \gamma Z_i + X_i'\Omega + \alpha_r + \alpha_t + \alpha_a + \epsilon_{irta}, \quad (1)$$

where subscripts i, r, t, and a refer to a family i, originating from region r, arriving in year t, and locating in a plot in *azienda* a.

Our main outcomes, denoted as y_{irta}, are binary variables indicating whether family i stayed in the Pontine Marshes until 1941 ("Stay"), and whether they purchased the plot ("Purchase"). For our exploration of possible mechanisms, we will rely on the third outcome variable, a binary variable equal to 1 if the family is expelled due to low productivity before 1941 ("Expelled").

The key explanatory variable, Z_i, captures neighborhood composition at arrival and is measured either by the share of same-origin families or by the diversity index described above, based on the settler's province.

The vector X_i' includes three types of controls: (i) characteristics of the plot such as size (in hectares), a dummy for poor soil, distance to the nearest rural settlement, and geographic coordinates (longitude and latitude), (ii) characteristics of the origin-municipality, including geographic coordinates, altitude, historical prevalence of malaria, wheat suitability (FAO index), and (iii) an index of agroclimatic similarity between the municipality of origin and the assigned plot.

The specification introduces fixed effects for region of origin, year of arrival, and *azienda*, denoted as α_r, α_t, and α_a respectively. The reason why we opt for region instead of province as origin fixed effects is to have sufficient variation when doing heterogeneity analyses and when exploiting additional origin information for the exploration of mechanisms. Standard errors are clustered at the *azienda* level, since the latter comprised geographically close plots and provided technical training, productive means (such as seeds, specific machinery, and fertilizers), and monitoring. We demonstrate however that our results are robust to alternative fixed effects and clustering choices.[32]

The coefficient γ captures the causal effect of increasing the share of common-origin neighbors in the buffer (or the diversity index) from 0 to 1. Given the fixed effects, identification comes from comparing families that arrived in the same year, came from the same region, and were assigned to a plot in the same *azienda*, but were exposed to different neighborhood composition.

As will be shown later, the coefficients are more precisely estimated when including fixed effects and controls, but they are of similar magnitude when obtained from regressions without controls. This is consistent with the idea that settlers were assigned to plots largely at random.

4.2 Identification: Quasi-Random Allocation of Neighbors

Historical accounts indicate that the assignment of plots to incoming settlers was conducted without any consideration for settlers' geographic origins or for a desired composition of neighborhoods (Folchi 2000, Alfieri 2014). The settlement proceeded sequentially as land was reclaimed, from the

[32]There are only 18 aziende. Given the small number of clusters, Table D2 reports alternative clustering levels and Conley spatial standard errors, which leave inference unchanged.

center of the Pontine Marshes towards the periphery. As plots became available, the CMCI assigned them to eligible families and organized the resettlement. The sole allocation criterion was the correspondence between the plot's size and the family labor capacity. There was no deliberate effort to place settlers from similar backgrounds together, nor to assign families from specific origins to plots with particular characteristics. Importantly, settlers had no prior knowledge of the location or characteristics of their plot, and were allowed to subsequently relocate only under very special circumstances.[33]

We complement this historical evidence with two empirical exercises to assess the quasi-random nature of plot assignment.

Balance on neighborhood composition measures. We assess balance between observable covariates and four measures of neighborhood composition. These include our main network and diversity measures within a 1-km buffer around the settler's house, as well as indicators for whether the closest plot is empty, and whether it is occupied by a common-origin settler. Covariates include plot characteristics (size, soil fertility, distance to the nearest rural settlement, and geographic coordinates) and origin-municipality attributes (coordinates, altitude, historical malaria prevalence, wheat suitability, and agroclimatic similarity with the assigned plot). Table 1 reports coefficients from regressions of neighborhood composition measures on this full set of controls mentioned in Section 4.1.

Consistent with the absence of strategic location choices, the estimated coefficients are generally close to zero and statistically insignificant. The few significant correlations are plausibly explained by mechanical features of the settlement process, such as the arrival of families from the same areas in consecutive waves. Our baseline specifications nonetheless control for all these characteristics, thereby accounting for any specific patterns related to the settlement process and the geographical origin of settlers. Moreover, whether one controls for our demanding set of fixed effects or not, the estimated coefficients are very similar, providing additional support to our identification assumption (see Figure C1 in the Appendix).

Expected vs. observed local common-origin exposure. As settlement occurred in waves, we can exploit variation in the monthly composition of incoming settlers to test whether proximity to common-origin neighbors reflects cohort composition rather than selective placement. For settlers arriving in each month t from province of origin p, we compute the *expected* common-origin exposure as the share of common-origin families arriving in the same month. For settlers arriving from p in t, this variable measures the probability of having a family from the same province in the closest plot, if (i) plots were randomly assigned within that arrival cohort, and (ii) the closest plot was empty. For example, if in a given month one third of settlers originate from a given province, random assignment implies that those settlers face an expected common-origin exposure of $1/3$. We then compute the corresponding *observed* common-origin exposure: the share of settlers from province p arriving in month t whose closest neighbor is a settler from the same province (irrespective of their month of arrival).

Based on the above definitions, the data should generate a linear correlation between our two measures of common-origin exposure across all groups of settlers arriving from different provinces

[33]See Folchi (2000) for more details on the link between labor units and plot size (p.123), the assignment to available plots (p.64), and the independence of soil quality from family characteristics (p.117).

Table 1: Balance test

	(1)	(2)	(3)	(4)
	Network	Diversity	Closest plot empty	Closest plot network
Plot size	-0.019	0.004	0.002	-0.028*
	(0.012)	(0.011)	(0.014)	(0.015)
Plot poor soil	-0.008	0.017	-0.050*	0.034
	(0.027)	(0.031)	(0.026)	(0.036)
Plot distance to borgo	0.018	-0.022*	0.002	0.015
	(0.012)	(0.011)	(0.007)	(0.013)
Plot longitude	0.097*	-0.083	0.085*	-0.041
	(0.056)	(0.060)	(0.048)	(0.075)
Plot latitude	0.117	-0.079	0.039	0.048
	(0.085)	(0.090)	(0.044)	(0.094)
Origin longitude	0.146***	-0.043	-0.026	0.162***
	(0.044)	(0.027)	(0.023)	(0.039)
Origin latitude	0.062	0.002	-0.024	0.047
	(0.073)	(0.039)	(0.038)	(0.060)
Origin altitude	-0.005	0.005	-0.009	-0.008
	(0.009)	(0.005)	(0.006)	(0.011)
Origin malaria exposure	0.024	-0.009	0.007	-0.002
	(0.016)	(0.007)	(0.012)	(0.020)
Origin wheat suitability	0.015	-0.000	-0.008	0.017*
	(0.010)	(0.005)	(0.005)	(0.009)
Agro-climatic similarity	0.012	0.008	-0.023	0.030
	(0.021)	(0.024)	(0.024)	(0.028)
Year FE	Yes	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes
Azienda FE	Yes	Yes	Yes	Yes
Mean dep. variable	0.33	0.61	0.15	0.34
Std. dev. dep. variable	0.29	0.21	0.35	0.47
R^2	0.30	0.36	0.10	0.11
Observations	3,784	3,784	3,784	3,784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in different months. With the observed variable on the vertical axis, we would also expect the regression line to have a slope below one and a positive intercept, because some of the settlers arriving in month t had families from earlier cohorts settled on the closest plot. Note that the expected measure can be interpreted as a *flow* variable, describing potential common-origin exposure generated by the composition of the current arrival cohort, whereas the observed measure captures the *stock* of local common-origin exposure resulting from both current and previous settlement waves.[34]

The left panel of Figure 7 plots observed against expected common-origin exposure across all settlers in our sample. Consistent with the hypothesis that the institutional design of the whole settlement process generated a quasi-random assignment of settlers to available plots, the graph displays a remarkably tight and linear relationship, with a slope below (but close to) one, and a positive intercept. Also consistent with quasi-random assignment, the distribution of the difference between observed and expected common-origin exposure is centered near zero, as shown in the right panel of Figure 7. At times, settlers ended up in areas that had more/fewer common-origin neighbors than expected, but such deviations average out. Moreover, regressing this gap on our large set of controls yields coefficients that are uniformly small and statistically insignificant, indicating that deviations from expected exposure are not systematically related to observable characteristics (see Figure C2 in the Appendix).[35]

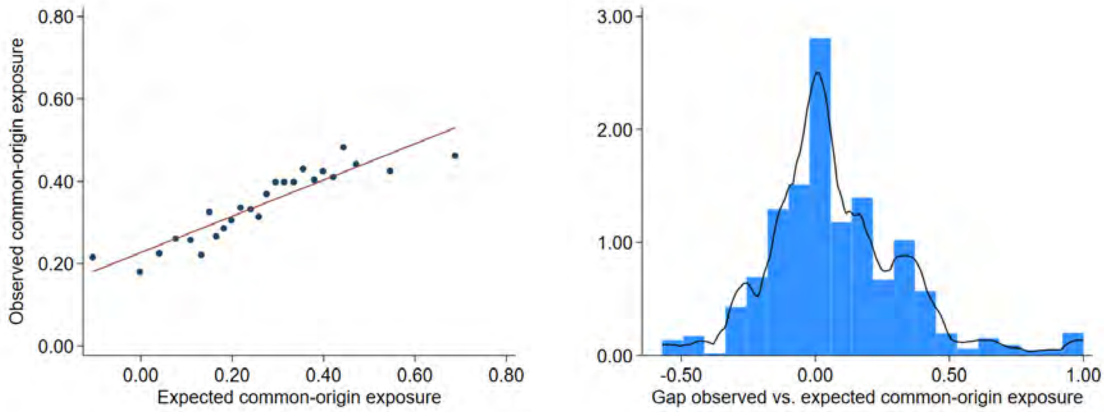


Figure 7: Expected and observed common-origin exposure

Notes: The left panel plots *observed* against *expected* common-origin exposure in our baseline specification. The right panel shows the distribution of the observed–expected exposure gap across all settlers in our sample.

5 Results

This section presents the main findings of our empirical analysis. Sections 5.1 and 5.2 show how settlers' outcomes are impacted by network size and diversity, respectively. Section 5.3 presents

[34] One can formalize these patterns as follows. Define $\mu_{p,t}$ as the share of settlers from p arriving in t who have a common-origin neighbor in the closest plot (*i.e.*, observed exposure), and $\delta_{p,t}$ as the share of common-origin neighbors arriving in the same month t (expected exposure). Accordingly, $\mu_{p,t-1}$ denotes the share of common-origin neighbors arriving before t ("past" exposure). Define $\sigma_{p,t} \in [0,1]$ as the share of settlers whose closest plot is not occupied prior to their arrival. We can then express observed exposure as a linear function of expected and past exposure: $\mu_{p,t} = (1-\sigma_{p,t})\mu_{p,t-1} + \sigma_{p,t}\delta_{p,t}$. If plot assignment is entirely random, but some incoming settlers locate in areas that are not totally empty (making $\sigma_{p,t} < 1$), the linear relationship exhibits a slope smaller than 1 and has a positive intercept.

[35] This pattern holds true even when excluding the first waves of settlers (*i.e.*, Veneto) or when looking separately at plots that are occupied for the first time or resettled.

additional results concerning the spatial features of the neighborhood. Section 5.4 discusses several robustness checks.

5.1 Network

Table 2 presents our results on the impact of network (*i.e.*, the share of common-origin neighbors within a 1-km radius buffer) on our main outcomes: *stay* in Panel A and *purchase* in Panel B. Controls and fixed effects are introduced sequentially: Column (1) displays the unconditional correlations without any controls, column (2) introduces the control variables listed in Table 1, while columns (3)-(5) introduce arrival year, region of origin, and *azienda* fixed effects one at a time. Finally, column (6) presents our preferred and most demanding specification, which includes the entire set of controls and fixed effects in a single regression.

Table 2: Impact of network

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Stay			
Network	0.128***	0.115***	0.114***	0.113***	0.172***	0.181***
	(0.039)	(0.027)	(0.023)	(0.025)	(0.015)	(0.024)
Mean dep. variable	0.72	0.72	0.72	0.72	0.72	0.72
Observations	3,784	3,784	3,784	3,784	3,784	3,784
			Panel B: Purchase			
Network	0.138**	0.132**	0.130***	0.111**	0.192***	0.166***
	(0.060)	(0.047)	(0.039)	(0.044)	(0.027)	(0.027)
Mean dep. variable	0.53	0.53	0.53	0.53	0.53	0.53
Observations	3,784	3,784	3,784	3,784	3,784	3,784
Controls	No	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Origin FE	No	No	No	Yes	No	Yes
Azienda FE	No	No	No	No	Yes	Yes

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The variable network takes values in the interval [0, 1]. Controls include plot characteristics (size in hectares, a dummy for poor soil, distance to the closest village, longitude and latitude), municipality of origin characteristics (longitude, latitude, altitude, malaria exposure, wheat suitability), and an index of agroclimatic similarity between the assigned plot and the municipality of origin. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Looking at column (6), transitioning from a network of 0 (no common-origin neighbors in the buffer) to 1 (all neighbors from the same province as settler *i*) translates into an increase of 18.1 percentage points in the probability that settler *i* remains in the Pontine Marshes until 1941. This effect is substantial, considering that the average likelihood of staying is 72%. A 0 to 1 increase in network leads to an increase of 16.6 percentage points in the probability of buying the land, against an average purchase rate of 53%. In other words, considering that the mean share of neighbors from the same origin stands at 0.33 with a standard deviation of 0.29, the estimates imply that a one standard deviation increase in the network measure leads to increases of about 5 percentage points in the probabilities of staying until 1941 and eventually purchasing the land.[36]

The estimates remain stable with and without controls, and across different specifications with demanding fixed effects, consistent with the hypothesis of quasi-random neighborhood composition.[37] The estimates remain quantitatively comparable with alternative combinations of fixed effects for both settlers' origins (*e.g.*, provinces instead of regions) and their plot location (*e.g.*, municipalities instead of *aziende*) as reported in Table D1. Consistent with the quasi-random assignment of plots, altering comparison groups does not significantly change our estimates. We also examine how standard errors vary under different clustering strategies. Table D2 reports regression results when clustering at the level of destination municipalities, origin provinces, and year of arrival, as well as allowing for spatial correlation at distances of 5 or even 10 kilometers (Conley 1999). The significance levels remain very similar across the board. Further robustness checks are discussed in Section 5.4.

These results pertain to the universe of settlers, including those who left before being offered the opportunity to redeem the land in 1941. To assess the extent to which the effects on land purchases are linked to settlers' likelihood of remaining in the Pontine Marshes, Table D3 reports estimates accounting for this channel. Columns (1) and (4) control for a "stay" dummy, columns (2) and (5) restrict the sample to settlers who stayed until 1940, and columns (3) and (6) focus on those who left and consider their duration of stay as outcome. The results show that an increase in network size from zero to one, or having a common-origin neighbor next door upon arrival, is associated with an increase of 3.6–4.9 percentage points in the likelihood of purchasing the land (*i.e.*, 6.7% of the baseline mean). These results suggest that the effects on land purchases documented above are not merely driven by a higher likelihood of staying in the Pontine Marshes, but reflect a direct effect of common-origin networks on settlers' capacity and willingness to redeem the land.

5.2 Diversity

Table 3 presents the effect of diversity (as fractionalization) on our key outcomes, stay and purchase. We present different specifications depending on: (i) whether we compute diversity using all neighbors in the 1-km buffer, or only among those who are not part of the network (*i.e.*, originating from a different province, what we refer to as "outgroup"); (ii) whether we include only diversity or we also condition on network size in the regression; (iii) whether we allow the effect to vary by degree of isolation through interaction terms.

As shown in column (1), overall neighborhood diversity has a negative and significant effect, estimated with no consideration of network size: moving from zero diversity (all the families in the buffer share the same origin) to maximum diversity (each family from a different origin) decreases the likelihood of long-term stay by 11.6 percentage points. The effect on purchase is comparable in size and marginally insignificant (p-value of 0.12), implying a reduction in the likelihood of buying the land of 9.6 percentage points.

These negative coefficients are not surprising in light of the inverse relationship between network and diversity displayed in Figure 6, and the large positive effects of network. Indeed, when

[36] As shown in Figure 5, network size displays substantial variation and extreme values close to 0 and 1 are common.

[37] The estimates across columns (1) to (6) are not statistically different from each other. The magnitude increases when introducing *azienda* fixed effects, which restricts the comparisons to fine administrative areas.

Table 3: Impact of diversity

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Stay			
Diversity (all)	-0.116**	0.094			-0.146**	-0.157**
	(0.055)	(0.084)			(0.053)	(0.054)
Network		0.223***		0.141***		
		(0.039)		(0.029)		
Diversity (outgroup)			-0.022	0.070**		
			(0.038)	(0.030)		
Small network					-0.224***	
					(0.060)	
Small network x Diversity					0.255**	
					(0.084)	
Zero network						-0.293***
						(0.077)
Zero network x Diversity						0.326***
						(0.084)
			Panel B: Purchase			
Diversity (all)	-0.096	0.104			-0.123*	-0.141**
	(0.059)	(0.096)			(0.055)	(0.051)
Network		0.213***		0.142**		
		(0.051)		(0.049)		
Diversity (outgroup)			-0.007	0.028		
			(0.040)	(0.041)		
Small network					-0.207**	
					(0.074)	
Small network x Diversity					0.234*	
					(0.104)	
Zero network						-0.288***
						(0.073)
Zero network x Diversity						0.336***
						(0.099)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,784	3,784	3,784	3,784	3,784	3,784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The variables network and diversity take values in the interval [0,1]. Small network refers to values below the median (0.20) while zero network to absence of common-origin settlers in the buffer. Controls include plot characteristics (size in hectares, a dummy for poor soil, distance to the closest village, longitude and latitude), and municipality of origin characteristics (longitude and latitude, malaria exposure, wheat suitability), and an index of agroclimatic similarity between the assigned plot and the municipality of origin. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

both are included in the regression, as in column (2), the coefficient on diversity becomes positive and non-significant, suggesting that the negative effect from column (1) could be entirely driven by the mechanical reduction in network size as diversity increases. This pattern is consistent with the notion, highlighted by Ashraf and Galor (2012) among others, that cultural homogeneity may be beneficial to economic performance for relatively unsophisticated modes of production, with agents operating within the existing technological frontier.

Furthermore, columns (3) and (4) show that diversity measured only among non-network neighbors does not have significant effects when considered in isolation, but may exhibit additional beneficial effects when controlling for network size. A settler exposed to a higher degree of outgroup diversity appears more likely to stay in the Pontine Marshes until 1941 although such an effect is not visible on the purchase outcome. A limitation of this measure is that even though it considers only outgroup members, it is still strongly negatively correlated with network size.[38]

One may speculate that settlers with a smaller network, or none at all, could benefit from diversity, as the absence of a dominating tight-knit group can create opportunities for interactions with different-origin neighbors. Columns (5) and (6) focus the attention on settlers with a network size below the median (*i.e.*, under 0.20) or with no common-origin families within a 1-km radius (*i.e.*, zero network). We introduce dummies capturing these dimensions as well as their interaction with the diversity measure. While diversity has a strong negative effect for individuals above the median network, its effect becomes net positive for those below the median (column 5).

Similarly, among the most isolated settlers, diversity has an even stronger beneficial effect. Raising the index from 0 (all other families from the same origin) to 1 (each family from a different origin) raises the probability of staying by 16.9 percentage points (-0.157+0.326) and of buying the plot by 19.5 percentage points (-0.141+0.336). These findings provide original insights into the individual effects of diversity, highlighting a fundamental heterogeneity that the literature has so far overlooked. In fact, while diversity appears to be harmful to average economic performance, isolated settlers have better outcomes if exposed to a more diverse environment.

Overall, our results on diversity appear to be consistent with the specific features under study. As in most pre-industrial economies, the settlers in the Pontine Marshes had very limited scope for specialization and market interactions, which can help rationalize the lack of generalized gains from diversity. Limited specialization often resulted in economic hardship, especially in the first years of the colonization. Our results present the first micro-level evidence in a quasi-experimental setting supporting the theoretical conjecture of Ashraf and Galor (2012, 2013b) by which the advantages of homogeneity exceed those of diversity in the Malthusian stage of development. We cannot however separate a distinct cost of heterogeneity from the mechanical reduction in network size.

In the absence of common-origin neighbors, isolated settlers appear to have relied on outgroup neighbors instead. Section 6 examines whether this substitution operated mostly through cooperation or through information sharing, in particular, knowledge transmission and learning from each other.

In Table D4 in Appendix D, we replicate our analyses of diversity using a measure of polarization. The results parallel those of our benchmark, suggesting that in our context both measures

[38]The explanation behind this negative correlation is that by increasing network size, the total number of outgroup members in the buffers falls, which in turn makes outgroup diversity become lower on average.

have similar effects on economic outcomes. This finding contrasts with other settings where polarization and fractionalization have been shown to have different effects, for instance, on identity formation (Bazzi et al. 2019).

5.3 Spatial Features of the Network

One distinct feature of our data, relative to the existing literature, is that it enables us to investigate how the specific location of neighbors, together with their characteristics, generates economically relevant externalities that affect migration success.[39]

We now turn our attention to the spatial features of the neighborhood and for each settler consider the following dimensions at arrival: (i) whether the closest plot is empty, (ii) whether the closest plot is occupied by a family from the network (*i.e.*, same province of origin), (iii) the distance (in km) to the closest network family, (iv) the average distance to network families, (v) the average distance between families in the network, reflecting network tightness.

In Table 4, we present the results obtained by using each of the above variables as the main regressor, under the most demanding specification that includes all fixed effects (corresponding to column (6) of Table 2). In Panel A, these spatial features are considered in isolation, while in Panel B, we also control for network size, to test whether a specific spatial feature explains the overall effect of networks and to quantify, for a given available network size, the additional effect of each specific spatial feature on migration success.

Without conditioning on network size, geographic proximity to network families appears to matter for both the closest neighbors and the entire 1-km radius buffer. Network tightness, instead, does not affect the probability of staying. Rather than an interconnected network, these unconditional results would suggest that closer network neighbors from the point of view of the individual settler had the largest positive impact. This observation concerns both same-province neighbors and having any close neighbor at all: columns (1) and (2) of Panel A show that the cost associated with having no neighbor in the closest plot is close in magnitude to the benefits enjoyed when the closest plot neighbor is from the same province. Conditional on network size, the estimates become less precisely estimated, although they convey a similar message. The effect of having no neighbor in the closest plot and the average distance to network families remain negative, stable in magnitude, and significant. The effect of having a network family in the closest plot falls to zero, while the distance to the first network family remains negative and large, although marginally insignificant. Note that the overall network effect remains unchanged at 18 percentage points while accounting for these spatial features of the neighborhood.

In Appendix Table D5, we present the effects of the spatial features of the network on the ”purchase” outcome. Table D5 reinforces the findings of Table 4, highlighting the importance of distance to common-origin neighbors even after controlling for network size.

Overall, our results point to the importance of the spatial structure of the network: having access to common-origin neighbors within a short distance is a key determinant of migrants’ success, and can be well understood in the framework of the Pontine Marshes settlement process, given the strong restrictions imposed on migrants’ mobility and interactions.

[39] A related paper with such features is Groeger and Zylberberg (2026), which focuses however on crop choices and the role of networks on knowledge transfers.

Table 4: Effects of spatial features of the network on the probability of staying until 1941

	(1)	(2)	(3)	(4)	(5)
	Stay				
	Panel A: not controlling for network size				
Closest plot empty	-0.051*				
	(0.026)				
Closest plot network (same province)		0.041***			
		(0.011)			
Distance to the first network family			-0.108**		
			(0.039)		
Avg. distance to network families				-0.117**	
				(0.054)	
Avg. distance among network families					-0.053
					(0.081)
Observations	3,784	3,784	2,795	2,795	2,795
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
	Panel B: controlling for network size				
Closest plot empty	-0.049*				
	(0.025)				
Closest plot network (same province)		-0.007			
		(0.015)			
Distance to the first network family			-0.062		
			(0.045)		
Avg. distance to network families				-0.108*	
				(0.055)	
Avg. distance among network families					-0.023
					(0.084)
Network	0.180***	0.187***	0.122**	0.147***	0.149***
	(0.025)	(0.030)	(0.054)	(0.048)	(0.050)
Observations	3,784	3,784	2,795	2,795	2,795
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: Distances are expressed in kilometers. The network variable takes values in the interval [0,1]. The sample comprises all the settlers who arrived between 1932 and 1940 in columns (1) and (2), and those with at least a network member in the buffer in columns (3) to (5). Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.4 Robustness

The analysis presented so far is necessarily based on some arbitrary choices, namely with respect to the definition of common origin, and neighborhood size at destination. In this Section, we briefly discuss the robustness of our results to alternative definitions and measurement choices. In the interest of space, the Tables and Figures cited here appear in Appendix D.5.

Alternative definitions of common origin. In Section 3.3, we outlined the historical reasons that led us to define a common origin at the provincial level. Here, we discuss how our results are robust to experimenting with a number of other plausible measures.

We start by considering alternative administrative units, namely, regions and municipalities. From columns (1) and (2) of Table D6, we can see that a more extensive regional or municipal network has a positive impact on the main outcome variables. Importantly, column (3) shows that in a horse race involving all the measures, the baseline province-level indicator is the only one that remains significant and retains a sizable coefficient, thus supporting our choice to use the province as a marker of common identity.

We can also adopt a more flexible approach and construct ”artificial provinces” as buffers of variable radius centered around the settler's municipality of origin. Table D7 shows that, when we define common-origin networks with respect to increasingly larger areas, the impact of network size tends to decay. For instance, when considering as network members the families originating from municipalities located within 10 km *vs* 100 km from the settler's place of origin (column (1) *vs* column (5)), the estimated coefficient halves. When performing a horse race among multiple ”artificial provinces”, as in column (6), the largest effect appears to be related to buffers of 25- and 50-km radius, which roughly correspond to the size of Italian provinces.[40] Column (7) further shows that our benchmark network measure computed using actual provinces outperforms all the alternatives based on ”artificial” ones.

Two additional alternatives are tested in Table D8. First, we consider a standardized score of geographic distance, measured in standard deviations with respect to the entire matrix of distances among all settlers, rather than fixed buffers. Second, we turn to cultural distance, as proxied by differences in the dialects of origin (measured using the *AIS Dialectometry* from Jaberg and Jud 1940). Similar to previous exercises, while the standardized geographic measure suggests that neighbors from more distant origins are detrimental to the success of settlers, the effects do not survive the inclusion of the province-based network measure in the regression. Linguistic distance, instead, is never statistically significant and has a zero-magnitude coefficient, suggesting that identity was not formed within linguistic borders and/or that easier communication is not a key mechanism of network formation in our setting.

Alternative buffer sizes. Our benchmark variables of neighborhood composition are computed considering all the families located in a buffer with a 1-km radius and centered around the farmhouse of family i. Here, we look at two types of alternative neighborhoods: (i) buffers of different

[40] Based on the average and median area of Italian provinces, the radius of equivalent circles would fall in the range between 28 and 30 km.

radius (500, 1000, and 1500 meters), and (ii) donut-type areas which exclude the inner space (so that we compute our indicators by considering families living within 500-to-750 meters, 750-to-1000 meters, 1000-to-1250 meters of settler i).

The results in Table D9 provide systematic evidence that our benchmark network measure generates the most statistically meaningful results, outperforming all the alternative indicators in horse races. As shown in columns (1) to (3), the network measured in a 1-km radius buffer remains large and significant when adding equivalent measures at 500 or 1500 meters, which do not show up as significant. The same is true when introducing donut-type measures, as shown in columns (4) to (8).

Absolute *vs* relative network. One may argue that the number, rather than the share of common-origin neighbors, matters for economic success. To explore this possibility, we build an alternative network measure based on the number of neighbors from the same province living within one kilometer of the settlers in our sample. The results are reported in Table D10. Columns (1)-(3) and (5)-(7) show that each additional network family increases the probability of staying in the Pontine Marshes until 1941 and eventually buying the land by about 1.1 percentage points. Interestingly, columns (4) and (8) show that the absolute measure of same-province network loses the horse race against the relative measure, thus supporting our choice of the latter as the benchmark explanatory variable.

Sensitivity to specific groups. Some groups of settlers faced special circumstances that could influence the results. For instance, those who occupied a plot previously farmed by another family may have inherited different starting conditions, or formed different ties with their neighbors. Those who came from nearby areas may have had different outside options or been able to maintain a social network beyond the Pontine Marshes. Finally, those who came from outside of Italy may have been enrolled in the program in a different fashion. Table D11 shows that our baseline results are robust to excluding these types of settlers.

6 Mechanisms

After establishing the effects of networks and diversity on settlers' economic performance, we now exploit the data at our disposal to study the possible underlying mechanisms.

Our analysis proceeds in three steps. First, we assess what made ties stronger, using the timing of arrival to identify when social bonds were formed (Section 6.1). Second, using data on expulsions for failing to meet production quotas, we examine whether networks operated through productivity or merely through the amenity value of common-origin neighbors (Section 6.2). Third, we ask what neighbors transmitted. Here we consider two potential channels: the transmission of agricultural knowledge from neighbors, which we test directly using measures of their expertise and environmental fit (Section 6.3), and solidarity and cooperation, which we test through its implications for where network effects should be largest (Section 6.4).

6.1 Formation of Social Bonds

As previously argued, life in the Pontine Marshes was extremely arduous and challenging, particularly in the first few months after arrival. Although social interactions were discouraged by the ONC, anecdotal evidence indicates that support and help from neighbors at this particular moment were essential. We take advantage of information on the month of arrival of settlers to shed light on the formation of social bonds and their importance for economic success.

In particular, we measure the share of network and non-network neighbors who arrived in the same month as a given settler i, and see how this correlates with our outcome variables. Columns (1) and (4) of Table 5 show that a higher share of network families arriving in the same month makes settlers significantly more likely to stay until 1941 and eventually redeem the land. Most importantly, these effects remain sizable (although smaller) and statistically significant even after controlling for the benchmark size of the network (*i.e.*, regardless of arrival date), as shown in columns (2) and (5).

Table 5: Bonding: Month of arrival and network effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Stay			Purchase		
Network arriving same month	0.078***	0.046*		0.080**	0.052*	
	(0.021)	(0.022)		(0.029)	(0.028)	
Network		0.160***	0.182***		0.140***	0.165***
		(0.027)	(0.026)		(0.027)	(0.032)
Non-network arriving same month			0.012			-0.003
			(0.031)			(0.053)
Mean dep. variable	0.72	0.72	0.72	0.54	0.54	0.54
R^2	0.09	0.10	0.09	0.08	0.09	0.08
Observations	3,137	3,137	3,137	3,137	3,137	3,137
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample comprises all the settlers who arrived between 1932 and 1940 and for which data on the month of arrival is available. The network variable takes values in the interval [0,1]. The variables of network (non-network) arriving in the same month take values in the interval [0,1], interpreted as 0-100%. See Table 2 for the list of controls and fixed effects. Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

These results suggest that sharing early struggles with people one could trust, namely those from the same province, led to a "strong bonding" that later manifested in better economic prospects, regardless of the overall size of the network. Moreover, historical evidence indicates that settlers who arrived in the same month might also have shared the train journey to the Pontine Marshes, which could take between 15 and 30 hours from most provinces in Northern Italy (Folchi 2000, Alfieri 2018). This long journey may have been an opportunity to meet settlers from nearby origins and start building relationships of mutual trust. In this perspective, our findings are reminiscent of Battiston (2019), who shows how (brief) social interactions during a migratory journey are key to the future outcomes of migrants. They instead stand somewhat in contrast to Beaman (2012), who finds a negative effect of year-of-arrival migrant networks on economic success, possibly due to the competition and the lack of relevant labor market information from newly arrived migrants—something which does not concern our setting.

Finally, columns (3) and (6) of Table 5 show that arriving alongside non-network families does not affect settlers' economic success and leaves the benchmark network effects substantially unaltered.

6.2 Productivity

The positive effect of the network on staying in the marshes up to 1941 and ultimately redeeming the land could operate through two main and possibly overlapping channels. On the one hand, the possibility of relying on common-origin neighbors could enhance settlers' current and future productivity, thereby making them more resilient and willing to redeem the land. On the other hand, the presence of neighbors with a similar socio-cultural background may provide social amenities and increase the utility of living in the Pontine Marshes, regardless of its impact on productivity.

The data does not allow us to disentangle the two channels, as we lack information on social amenities and overall productivity.[41] This notwithstanding, we can identify settlers whose sharecropping contracts were terminated due to low productivity, and were subsequently expelled from the Pontine Marshes.[42]

Table 6 reports how common-origin families (in the buffer or in the closest house) affect the likelihood of being expelled from the land, differentiating by the reason for expulsion. A more extensive network significantly reduces expulsions and such reduction is entirely driven by expulsions due to low productivity. As shown in columns (2) and (5), both network size and having a common-origin family in the closest plot have a negative and sizable effect on the likelihood of expulsion for lack of productivity. Overall, increasing the network from 0 to 1 is associated with a 3.8 percentage point reduction in the probability of expulsion while a common-origin family in the closest house reduces expulsion by 2 percentage points (relative to a mean of 8%). Columns (3) and (6) show no effects on expulsions for other reasons. These findings provide support for the idea that the presence of neighbors of common origin helped prevent settlers' productivity from falling below the threshold that would have exposed them to the risk of expulsion, thereby underscoring the potential importance of solidarity, which we explore further in the following subsections.[43]

[41] Harvest registers were destroyed during the German invasion of the Pontine Marshes in 1944.

[42] Archival sources indicate the following causes of expulsion (*i.e.*, *disdetta*): (a) political "indignity"; (b) insubordination or violence against ONC; (c) any criminal conviction; (d) stealing, or damaging the property; (e) abandoning, lending or selling cattle; (f) stealing or selling food, stocks, means of production, etc.; (g) leaving the plot unattended, or reducing the workforce, so that the standards of efficiency (productivity) were diminished; (h) not fulfilling contractual obligations (including failing to be a "good", productive settler). We consider the last two reasons (g-h) to be productivity-related. For instance, De Nardis (2010) describes (p. 43) the case of a family expelled because they "*neglected farm work, [and had] no care for products and their use. In two years, they have shown ill will and neglected land works. The plot is in the same condition or worse than when they arrived.*"

[43] An alternative hypothesis is that larger co-ethnic networks may allow individuals to organize and resist expulsions. This channel cannot explain the effects of a common-origin family in the closest plot (columns (4)-(5) in Table 6). Moreover, such episodes of resistance or political opposition would have been reflected in historical accounts and we have found no evidence of them.

Table 6: Productivity: Effects on probability to be expelled

	(1)	(2)	(3)	(4)	(5)	(6)
	Expelled					
Network	-0.032*	-0.038**	0.006			
	(0.017)	(0.014)	(0.009)			
Closest plot same province				-0.017**	-0.020***	0.003
				(0.008)	(0.006)	(0.005)
Mean dep. variable	0.09	0.08	0.01	0.09	0.08	0.01
R^2	0.04	0.04	0.02	0.04	0.04	0.02
Observations	3,784	3,784	3,784	3,784	3,784	3,784
Reason for expulsion	All	Low yield	Other	All	Low yield	Other
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The dependent variable varies across columns, capturing different definitions of being expelled: columns 1 and 4 include all reasons, columns 2 and 5 low yield, and columns 3 and 6 other reasons. The network variable takes values in the interval [0,1]. See Table 2 for the list of controls and fixed effects. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Learning

Among the possible channels through which neighborhood composition can influence productivity, and therefore individual success, a natural candidate is the opportunity to learn from others. In our context, this would mean that a settler's productivity may be enhanced by a transfer of agricultural expertise or skills from other settlers. In particular, while learning from common-origin neighbors might be easier, outgroup settlers may hold more valuable and complementary knowledge, because they come from provinces with different characteristics.

To explore the relevance of the learning mechanism in explaining the importance of neighbors, we rely on three proxies of agricultural expertise.[44] First, we construct a measure of agroclimatic similarity following Bazzi et al. (2016), aiming to capture the fit between the agricultural characteristics of the settlers' place of origin and their assigned plot in the Pontine Marshes. Second, since the ONC imposed wheat as the main crop as part of Mussolini's *"Battle for Grain"*, we use the FAO index of wheat suitability to measure how suitable the settlers' origin municipalities were for wheat cultivation. Finally, we compute the difference between wheat suitability at the origin and in the destination plot, as a proxy for crop-specific (as opposed to general) agroclimatic similarity. Wheat-specific similarity is arguably the more meaningful variable for the stay outcome, and general agroclimatic similarity for the purchase outcome. During the sharecropping period settlers had to follow the ONC's wheat-heavy crop assignments, while only after redeeming the plot were they free to choose what to grow.

[44] Appendix E provides a detailed description of how these measures are constructed.

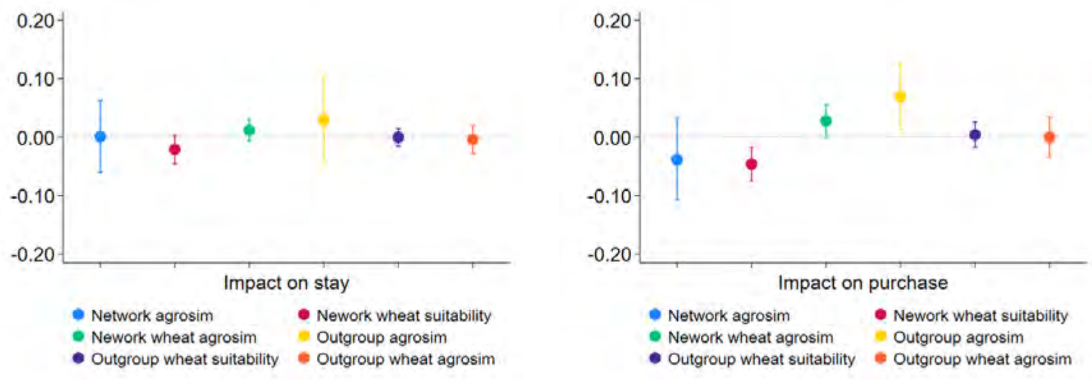


Figure 8: Impact of neighbors' agroclimatic measures

Notes: The figure plots the estimated coefficients and 95% confidence intervals of separate regressions of our baseline specification with various measures of neighbors' agricultural potential. Section E in the Appendix reports more details on the computation of these indicators.

We standardize the three measures described above and average them for all the neighbors in the buffer, distinguishing between network and non-network (outgroup) families. The regression coefficients from the most demanding specification, including all controls and the complete set of fixed effects, are shown in Figure 8. We do not find any effect of neighbors' agricultural knowledge or experience on the probability of staying until 1941, regardless of their origin. We read this as evidence against learning from network neighbors as a source of positive network externalities. Although not significant, the effect of the network's expertise with wheat has a negative coefficient, which becomes significant when the outcome is purchasing the land. One reading is that settlers accustomed to growing wheat in a favorable environment struggled in an area as unsuitable as the Pontine Marshes. Outgroup agroclimatic similarity, by contrast, has a modest beneficial effects only on the probability of purchasing the land. Therefore, learning appears to play a limited role overall, as we can only detect some possible effect after the end of the colonization phase (1941), and originating from non-network neighbors.

These results are better understood in light of the context. Rather than being able to experiment with crops and learn from neighbors with different agricultural expertise, settlers in the Pontine Marshes were expected to meet strict production targets and comply with specific crop mixes imposed by the *ONC*. These quotas left little room to adopt neighbors' techniques, or any practice that might be seen as riskier.

6.4 Solidarity and cooperation

We now examine solidarity as a key mechanism through which common-origin networks and diversity may have affected settlers' outcomes. If the influence of neighbors operated through mutual help and cooperation, the effects should be more pronounced for families for whom the marginal benefit of such help was largest, namely those under greater material stress and those with a stronger cultural reliance on in-group ties. We consider three dimensions of heterogeneity that map to these channels.

The first is demographic pressure on resources, defined as the ratio of family size to plot size. While larger families were assigned to larger plots, the Tommasi rule for plot allocation did not take into account the presence of dependent individuals (*e.g.*, children and the elderly), generating variation in mouths to feed per hectare that is plausibly unrelated to agricultural skill or plot quality. The second is prior exposure to malaria in the origin municipality. While the ONC had attempted to eradicate the disease, malaria remained a serious health concern in the Pontine Marshes in the 1930s. Settlers without prior exposure were therefore more prone to disease and less familiar with malarial land, and stood to benefit more from mutual help and cooperation with their neighbors. The third is cultural background, in particular social capital, which we explore by comparing settlers originating from the two main regions surrounding the Po River, Emilia-Romagna to the south and Veneto to the north. Emilia-Romagna was historically exposed to institutions that fostered the development of more open, civic-oriented social structures, based on generalized trust and participation in cooperative associations. By contrast, Veneto has historically been characterized by more "closed" social structures, with cooperation being mediated by kinship, parish, and local community.[45] Emilia-Romagna has traditionally relied on *bridging* social capital while Veneto has specialized in *bonding* social capital (Crescenzi, Gagliardi, and Percoco 2011, Putnam 1995, Sabatini 2009).[46] In our case, this suggests that settlers from Veneto would have been more inclined to seek help almost exclusively from neighbors of common origin, thereby benefiting relatively more from network externalities.

Figure 9 reports the estimated network coefficients across these sub-samples. Network effects are larger among families facing above-median demographic pressure and among those originating north of the Po. The estimates also point in the predicted direction for malaria exposure, although the differences are smaller; we therefore regard this cut as suggestive.

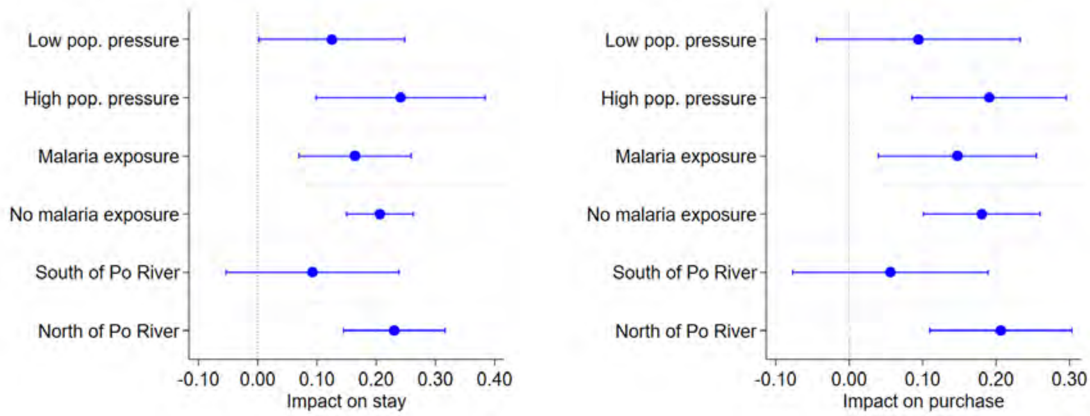


Figure 9: Networks and solidarity

Notes: The figure plots the estimated coefficients and 95% confidence intervals of network effects for different sub-samples of families: (i) below/above median family to plot size, (ii) exposure/non-exposure to malaria in origin municipality, and (iii) origin municipality to the south/north of Po River.

[45] Cultural variation across Italy may be traced back to deep historical and geographical factors, which have been shown to have long-lasting consequences on a wide range of outcomes, such as social and civic capital (Putnam, Nanetti, and Leonardi 1994, Guiso, Sapienza, and Zingales 2016, Buonanno et al. 2022), voting outcomes (Buccione and Knight 2023, Dominici 2024), firm structure (Trigilia 1986), as well as marriage and secularization patterns (Zei et al. 2005, De la Croix, Mariani, and Mercier 2023).

[46] Putnam (1995) and Sabatini (2009) interpret religiosity as a proxy for bonding social capital in the particular context of Italy, and highlight that Veneto is traditionally more religious than Emilia-Romagna.

Figure 10 reports the analogous heterogeneity for diversity among isolated families (*i.e.*, having no network in the buffer).[47] The compensating effect of diversity among isolated families is concentrated in the same sub-samples in which networks themselves matter most: families under higher demographic pressure, families without prior malaria exposure, and families from north of the Po. Diversity substitutes for the network precisely where the network is most valuable, which is what cooperation (rather than knowledge transfer) would predict. A further piece of evidence in the same direction comes from soil quality. Diversity has essentially no effect on isolated families assigned to good-soil plots and a sizable negative effect on those assigned to poor-soil plots (Tables F4 and F5 in the Appendix). Because settlers were required to deliver half of their harvest to the ONC, families on poor soil operated closer to subsistence output, making consumption-smoothing through neighbors and cooperation particularly valuable (Munshi and Rosenzweig 2016, Litina 2016).

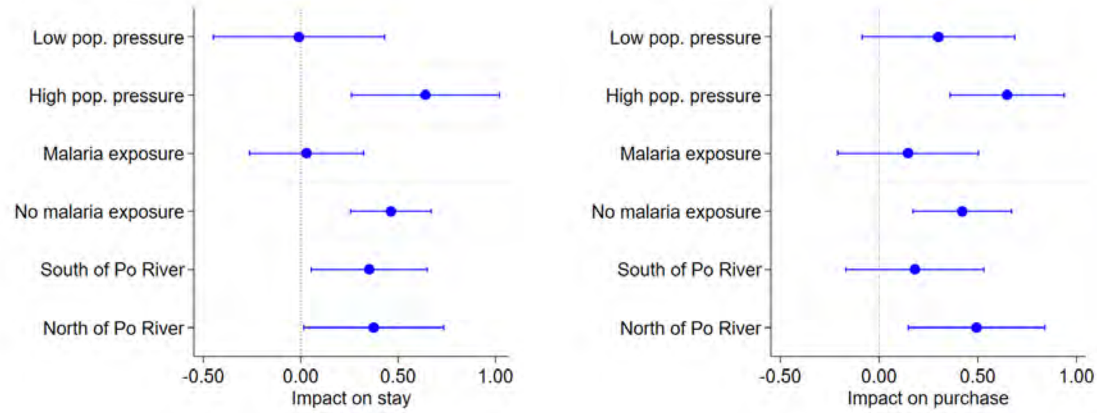


Figure 10: Diversity and solidarity among isolated families

Notes: The figures plot the estimated coefficients and 95% confidence intervals of diversity effects interacted with a dummy for being isolated (zero network) for different sub-samples of families: (i) below/above median family to plot size, (ii) exposure/non-exposure to malaria in origin municipality, and (iii) origin municipality to the south/north of Po River.

To provide evidence on the potential mediating role of social capital among settlers originating from the Po Valley, we investigate whether the effects of network and diversity change progressively across space or rather sharply at that river boundary. In particular, we interact network with dummies for various distances between the settlers' origin municipality and the course of the Po River, thus allowing the effects to vary flexibly by distance. The left panel of Figure 11 shows that the estimates change discretely as one crosses the river. Even settlers originating within 40 kilometers of the Po experienced markedly different effects depending on which bank they came from. The right panel of Figure 11 illustrates a proxy of bonding social capital, the intensity of consanguineous marriages in the 1930s (Zei et al. 2005), reporting the mean values across equal-sized bins on both sides of the Po River. The measure of consanguinity exhibits a similar discontinuity at the river. Interacting this index with network size (or diversity) yields a positive and significant coefficient (Table D12), consistent with stronger externalities in more bonding-oriented origin

[47] We estimate our baseline equation including diversity, an indicator for being isolated, and their interaction. We use the full sample rather than restricting it to isolated families, who represent only 20% of observations.

communities.[48]

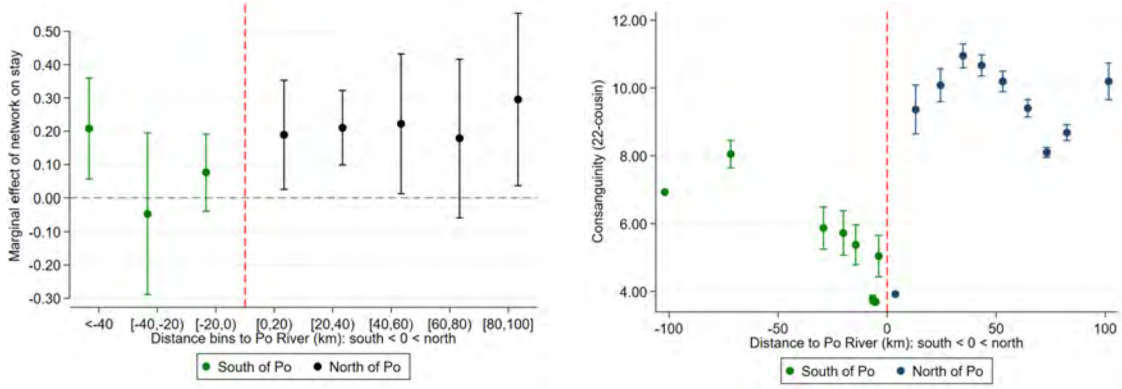


Figure 11: Network effects and social capital along the Po River

Notes: The plot on the left shows the effect of network estimated flexibly in a single regression, interacting network with bin indicators at different distances with respect to the Po River. The plot on the right shows the average of our measure of consanguinity (*i.e.*, cousin marriages per thousand people in 1932-1935) by distance to the Po River, in 10 equal size bins at the south and north.

7 Conclusion

Whether migrants fare better surrounded by people from the same origin or in a more diverse environment has been difficult to establish, as migrants ordinarily choose where they live and can relocate later on. By exploiting a historical quasi-natural experiment (the reclamation and settlement of the Pontine Marshes in 1930s Italy), we provide causal evidence on how neighborhood composition shapes migrants' economic performance in a rural environment. The centralized and quasi-random assignment of families to land plots, together with detailed settler-level data covering the universe of families, allows us to identify the effects of networks and diversity without the selective sorting that typically complicates such analysis.

We find that migrant networks play a fundamental role in improving families' economic performance and resilience in these newly created communities. A larger network of common-origin neighbors within one kilometer significantly increases the probability of remaining in the Pontine Marshes and of purchasing the land. A more diverse neighborhood is associated with worse performance on average, broadly consistent with Unified Growth Theory's prediction that diversity is detrimental to performance in economies with relatively unsophisticated modes of production. This negative effect, however, largely reflects the mechanical crowding out of common-origin ties rather than a distinct cost of heterogeneity. Among the roughly one in five settlers without same-origin neighbors nearby, diversity instead improves outcomes, suggesting that a diverse social environment can partly compensate for the absence of a common-origin network.

Our analysis of the mechanisms consistently points to the importance of cooperation and solidarity rather than knowledge transmission. Settling at the same time as other common-origin

[48] We find similar evidence using a historical proxy of religiosity (based on the prevalence of religious names, as computed by Becker and Voth 2026), although the change in trend at the Po River is less clear. We find no comparable discontinuity for any of our controls: characteristics of the plot (*e.g.*, size, soil fertility, coordinates) and other origin characteristics (*e.g.*, wheat suitability, altitude, malaria exposure, coordinates), which makes it harder to attribute the result to economic or geographic correlates of latitude or other strategic features of the settling process.

families is particularly beneficial, consistent with social bonds being formed through shared experiences during the early stages of settlement. Networks also improve productivity rather than merely providing amenities: settlers embedded in larger networks are less likely to be expelled for failing to meet production quotas. While in rural economies neighbors may transmit relevant knowledge about crops and techniques, in the Pontine Marshes the state agency dictated what was grown and how, leaving little room for this channel to operate. Consistent with this, we find that settlers gain nothing from neighbors with better agricultural knowledge or a better environmental fit. Instead, what neighbors seem to supply is mutual help. Networks matter most where the need for assistance is greatest, including among families facing demographic pressure or less familiar with the environment, and among settlers from origins historically more reliant on bonding social capital. Similar patterns emerge for the beneficial role of diversity among the most isolated settlers, suggesting that a more diverse environment facilitates the formation of ties that provide support similar to that of common-origin networks.

These findings highlight that the economic consequences of neighborhood composition depend on the broader social environment in which migrants settle. In rural settings with limited formal markets and institutions, common-origin networks can provide forms of support that improve economic resilience, while diversity can offer an alternative to migrants who lack such ties. The welfare implications, however, depend on how these gains are distributed, and on the position of migrants excluded from network externalities altogether. This distinction bears on the design of resettlement policies and will become increasingly important as rural migration and displacement continue to rise under climate pressure. The relevant policy choice may not be simply between clustering and dispersion, but between different ways of providing migrants with access to social support.

References

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport. 2016. "Birthplace diversity and economic prosperity." *Journal of Economic Growth* 21 (2): 101–138.

Alfieri, Giulio. 2014. *La terra che non c'era*. Siena, Italy: Betti Editrice.

———. 2018. *Questo piatto di grano*. Rome, Italy: Edizioni di Atlantide.

Algan, Yann, Camille Hémet, and David D Laitin. 2016. "The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation." *Journal of Political Economy* 124 (3): 696–733.

Ashraf, Quamrul, and Oded Galor. 2012. "Cultural diversity, geographical isolation, and the origin of the wealth of nations." *IZA Discussion Paper*, no. 6319.

———. 2013a. "Genetic diversity and the origins of cultural fragmentation." *American Economic Review* 103 (3): 528–533.

———. 2013b. "The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development." *American Economic Review* 103 (1): 1–46.

Baggio, Michele, and Metin Coşgel. 2026. "Racial diversity and team performance: Evidence from the American offshore whaling industry." *Journal of Law, Economics, and Organization*. Forthcoming.

Battisti, Michele, Giovanni Peri, and Agnese Romiti. 2022. "Dynamic effects of co-ethnic networks on immigrants' economic success." *Economic Journal* 132 (641): 58–88.

Battiston, Diego. 2019. "The persistent effects of brief interactions: Evidence from immigrant ships." *Mimeo*.

Bazzi, Samuel, Abel Brodeur, Martin Fiszbein, and Joanne Haddad. 2026. "Frontier history and gender norms in the United States." *Journal of the European Economic Association*. Forthcoming.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. 2020. "Frontier culture: The roots and persistence of "rugged individualism" in the United States." *Econometrica* 88 (6): 2329–2368.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong. 2016. "Skill transferability, migration, and development: Evidence from population resettlement in Indonesia." *American Economic Review* 106 (9): 2658–2698.

———. 2019. "Unity in diversity? How intergroup contact can foster nation building." *American Economic Review* 109 (11): 3978–4025.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak. 2021. "Can network theory-based targeting increase technology adoption?" *American Economic Review* 111 (6): 1918–1943.

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

Becker, Sascha C, and Hans J Voth. 2026. "From the death of God to the rise of Hitler." *Mimeo*.

Belmonte, Alessandro. 2026. "State capacity, schooling, and Fascist values: Evidence from the reclamation of the Pontine Marshes." *Mimeo*.

Blumenstock, Joshua E, Guanghua Chi, and Xu Tan. 2025. "Migration and the value of social networks." *Review of Economic Studies* 92 (1): 97–128.

Boccini, Floriano, and Erminia Ciccozzi. 2007. *Opera Nazionale per i Combattenti: Progetti*. Rome, Italy: Archivio Centrale dello Stato.

Brown, John C, and David Cuberes. 2023. "The effect of first nature and second nature forces on urbanization: the case of Oklahoma (1889-2010)." *Mimeo*.

Buccione, Giulia, and Brian Knight. 2023. "The long-run political effects of the separation of church and state: evidence from the Papal State." *Mimeo*.

Buonanno, Paolo, Matteo Cervellati, Sara Lazzaroni, and Giovanni Prarolo. 2022. "Historical social contracts and their legacy: a disaggregated analysis of the medieval republics." *Journal of Economic Growth* 27 (4): 485–526.

Carillo, Mario F. 2022. "Fascistville: Mussolini's new towns and the persistence of neo-fascism." *Journal of Economic Growth* 27 (4): 527–567.

Clement, Viviane, Kanta Kumari Rigaud, Alex De Sherbinin, Bryan Jones, Susana Adamo, Jacob Schewe, Nian Sadiq, and Elham Shabahat. 2021. *Groundswell part 2: Acting on internal climate migration*. World Bank, Washington, DC.

Comola, Margherita, and Mariapia Mendola. 2015. "Formation of migrant networks." *Scandinavian Journal of Economics* 117 (2): 592–618.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Conley, Timothy G, and Christopher R Udry. 2010. "Learning about a new technology: Pineapple in Ghana." *American Economic Review* 100 (1): 35–69.

Crescenzi, Riccardo, Luisa Gagliardi, and Marco Percoco. 2011. "The bright side of social capital: how 'bridging' makes Italian provinces more Innovative." *SERC Discussion Paper*, no. 96.

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

de Haas, Tymon, and Mans Schepers. 2022. "Wetland reclamation and the development of reclamation landscapes: a comparative framework." *Journal of Wetland Archaeology* 22 (1-2): 75–96.

De la Croix, David, Fabio Mariani, and Marion Mercier. 2023. "Driven by institutions, shaped by culture: human capital and the secularization of marriage in Italy." *International Economic Review* 64 (4): 1777–1818.

Del Bravo, Renzo. 1956. *La colonizzazione dell'Agro Pontino*. Imola, Italy: Cooperative Tipografico-Editrice "Paolo Galeati".

De Nardis, Laura. 2010. *I Coloni del regime. Aspetti della vita rurale nell'Agro Pontino, 1930 – 1939*. Rome, Italy: University of Rome "La Sapienza".

Docquier, Frédéric, Riccardo Turati, Jérôme Valette, and Chrysovalantis Vasilakis. 2020. "Birthplace diversity and economic growth: Evidence from the US states in the Post-World War II period." *Journal of Economic Geography* 20 (2): 321–354.

Dominici, Alice. 2024. "Shadowless theocracies." *Mimeo.*

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment." *Quarterly Journal of Economics* 118 (1): 329–357.

Elsner, Benjamin, Gaia Narciso, and Jacco Thijssen. 2018. "Migrant networks and the spread of information." *Oxford Bulletin of Economics and Statistics* 80 (3): 659–688.

Folchi, Annibale. 2000. *I contadini del duce: Agro Pontino, 1932-1941*. Rome, Italy: Pieraldo.

Foster, Andrew D, and Mark R Rosenzweig. 1995. "Learning by doing and learning from others: Human capital and technical change in agriculture." *Journal of Political Economy* 103 (6): 1176–1209.

Gentilcore, Rocco Louis. 1970. "Reclamation in the Agro Pontino, Italy." *Geographical Review* 60 (3): 301–327.

Groeger, Andre, and Yanos Zylberberg. 2026. "The pick of the crop: agricultural practices and clustered networks in village economies." *Journal of the European Economic Association* 24 (1): 271–310.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-term persistence." *Journal of the European Economic Association* 14 (6): 1401–1436.

Haddad, Joanne. 2026. "Settlers and norms." *Journal of Comparative Economics*, vol. 54.

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

ISTAT. 2019. Confini statistico-amministrativi: analisi storica. https://www.istat.it/notizia/confini-statistico-amministrativi-analisi-storica/.

Jaberg, Karl, and Jakob Jud. 1940. *Sprach-und Sachatlas Italiens und der Südschweiz*. Zofingen, Switzerland: Ringier & Co.

Litina, Anastasia. 2016. "Natural land productivity, cooperation and comparative development." *Journal of Economic Growth* 21 (4): 351–408.

Maertens, Annemie, and Christopher B Barrett. 2013. "Measuring social networks' effects on agricultural technology adoption." *American Journal of Agricultural Economics* 95 (2): 353–359.

Mariani, Fabio, Marion Mercier, and Luca Pensieroso. 2023. "Left-handedness and economic development." *Journal of Economic Growth* 28 (1): 79–123.

Martén, Linna, Jens Hainmueller, and Dominik Hangartner. 2019. "Ethnic networks can foster the economic integration of refugees." *Proceedings of the National Academy of Sciences* 116 (33): 16280–16285.

Massey, Douglas S. 1985. "Ethnic residential segregation: A theoretical synthesis and empirical review." *Sociology and Social Research* 69:315–350.

Maystadt, Jean-Francois, and Philip Verwimp. 2014. "Winners and losers among a refugee-hosting population." *Economic Development and Cultural Change* 62 (4): 769–809.

Mekonnen, Daniel Ayalew, Nicolas Gerber, and Julia Anna Matz. 2018. "Gendered social networks, agricultural innovations, and farm productivity in Ethiopia." *World Development* 105:321–335.

Minale, Luigi, Rudi Rocha, and Bruno Vigna. 2024. "Immigrant diversity and long-run development." *RF Berlin-CReAM Discussion Paper*, no. 08/24.

Montalvo, José G., and Marta Reynal-Querol. 2005. "Ethnic polarization, potential conflict, and civil wars." *American Economic Review* 95 (3): 796–816.

Muñoz-Blanco, Laura. 2026. "Reservoir-induced displacement and social participation: Evidence from the Spanish dictatorship." *European Economic Review* 184:105235.

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

———. 2004. "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution." *Journal of Development Economics* 73 (1): 185–213.

Munshi, Kaivan, and Mark Rosenzweig. 2016. "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." *American Economic Review* 106 (1): 46–98.

Patacchini, Eleonora, and Yves Zenou. 2012. "Ethnic networks and employment outcomes." *Regional Science and Urban Economics* 42 (6): 938–949.

Posch, Max, Jonathan Schulz, and Joseph Henrich. 2026. "How cultural diversity drives innovation: Surnames and patents in US history." *Journal of Political Economy* 134 (4): 1245–1305.

Putnam, Robert D. 1995. "Bowling alone: America's declining social capital." *Journal of Democracy* 6 (1): 65–78.

Putnam, Robert D., Raffaella Y. Nanetti, and Robert Leonardi. 1994. *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Rossetti, Cristina. 1994. *I ferraresi nella colonizzazione dell'Agro Pontino*. Rome, Italy: Bulzoni Editore.

———. 2008. "Un colono ferrarese in Agro Pontino e il riscatto del podere." *Annali del Lazio Meridionale* VIII (1): 7–22.

Ruiz, Isabel, and Carlos Vargas-Silva. 2015. "The labor market impacts of forced migration." *American Economic Review* 105 (5): 581–586.

Sabatini, Fabio. 2009. "Social capital as social networks: A new framework for measurement and an empirical analysis of its determinants and consequences." *Journal of Socio-Economics* 38 (3): 429–442.

Torelli, Luigi. 1882. *La carta della malaria dell'Italia*. Florence, Italy.

Trigilia, Carlo. 1986. "Small-firm development and political subcultures in Italy." *European Sociological Review* 2 (3): 161–175.

UNHCR. 2026. "Global Trends: Forced Displacement in 2025." Technical Report, United Nations High Commissioner for Refugees, Copenhagen, Denmark.

Zei, G, A Lisa, O Fiorani, GA Moroni, and LL Cavalli-Sforza. 2005. "Spatial and temporal variation of consanguinity in Italy." *Human Evolution* 20 (2): 201–215.

Online Appendix

”Networks, Diversity, and the Economic Performance of Migrants: Evidence from the Pontine Marshes (1932-1941)”

Frédéric Docquier, Alice Dominici, Martín Fernández, Fabio Mariani

Table of Contents

A From archival sources to geo-referenced data

Here, we present additional information related to the archival sources and the construction of our dataset.

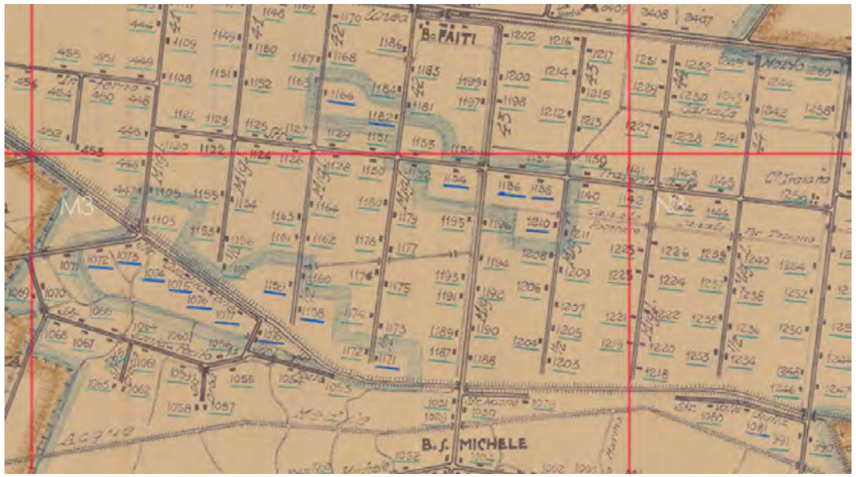
Figure A1 displays an example of the archival records, reporting for each *podere* its number, surface, *azienda*, as well as the settler's name.

Figure A1: Example of archival data

Figure A2 reproduces in panel (a) the original map of the Pontine Marshes allotment. The detail in panel (b) shows how each farmhouse is identified by a sequential number, which allows us to geo-reference the information from the archival registers (see Figure A1).

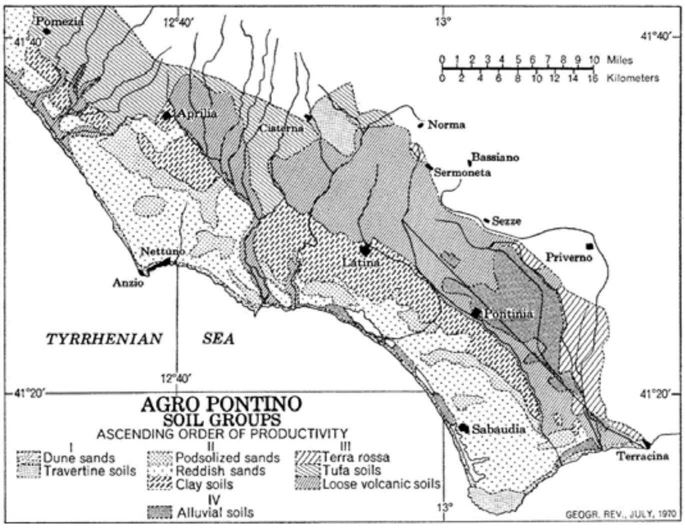
Figure A2: Original map of the Pontine Marshes with numbered plots (ONC, 1939)

(a) Complete view



(b) Detailed view

Figure A3: Maps with soil types



(a) Original map with soil types

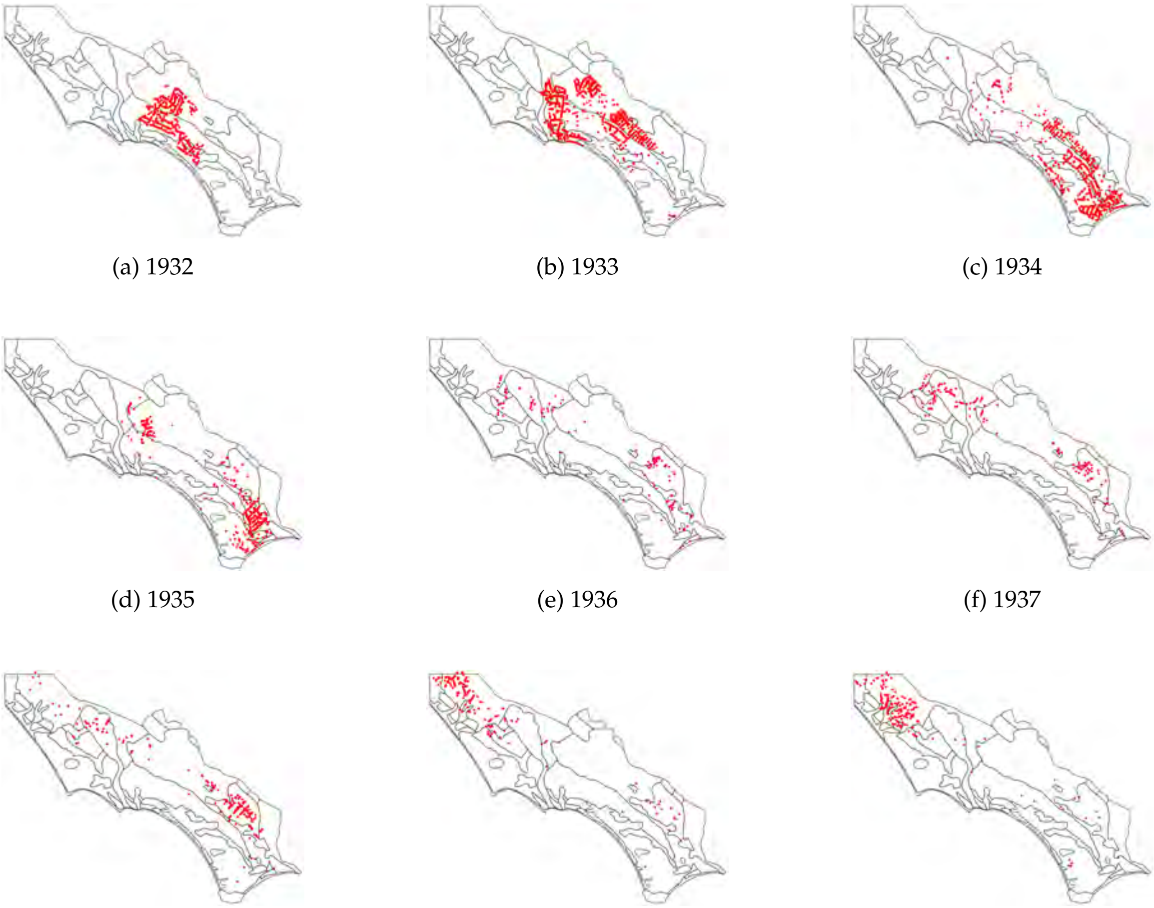
(b) Geo-referenced map with plots and soil types

By using the historical map with the ONC's soil classification from Gentilcore (1970), as in panel (a) of Figure A3, we can further associate to each plot the relevant characteristics of the terrain, thus obtaining the map in panel (b).

Figure A4 describes the roll-out of the settlement process, showing for each year the plots assigned to incoming settlers. As can be seen in the Figure, the settlement began in the central area of the Pontine Marshes and subsequently extended to peripheral plots, consistent with the historical evidence describing settlement following reclamation Folchi (2000).

Finally, Figure A5 displays the map from Torelli (1882), which has been used to construct our measure of historical malaria exposure in origin municipalities.

Figure A4: Settlement process by year, 1932-1940



(a) 1932

(b) 1933

(c) 1934

(d) 1935

(e) 1936

(f) 1937

(g) 1938

(h) 1939

(i) 1940

Figure A5: Historical malaria exposure: map from Torelli (1882)

B Summary of data sources

Table B1: Summary of data sources

Variable	Source	Level of observation
Municipality of origin Arrival date (month, year) Plot number Family size Stay until 1941 (outcome) Land purchase (outcome) Voluntary departures Expulsions and reasons	Alfieri (2018), originally, ONC archival records (State Archives of Latina)	Settler family
Plot size *Azienda* assignment Land price (subsample)	ONC archival records (State Archives of Latina)	Settler family
Origin coordinates and altitude Altitude (and standard deviation)	Own calculations based on origin municipality from Alfieri (2018) and ISTAT (2019)	Origin municipality
Farmhouse coordinates Distance to nearest rural settlement Distance from closest river	1939 digitized ONC historical map (see Figure A2)	Plot
Soil type	Gentilcore (1970) and archival ONC soil classification (see Figure A3a)	Plot
Historical malaria prevalence at origin	Torelli (1882) (see Figure A5)	Origin municipality
Linguistic similarity between Italian dialects	AIS Dialectometry, Jaberg and Jud (1940)	Settler family
Agroclimatic similarity index components: Average temperature (1900–1930) Average precipitation (1900–1930)	Climatic Research Unit TS4.07, University of East Anglia	Origin municipality - destination plot
Soil nutrient availability Soil rooting conditions Wheat suitability (FAO index) Oxygen availability (drainage) Soil salinity and sodicity	FAO Harmonized World Soil Database (HWSD)	Origin municipality - destination plot

C Descriptive statistics and balance checks

Table C1 reports descriptive statistics for the sample of settlers and associated plots.

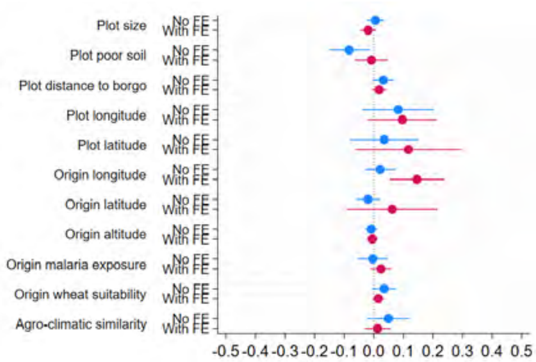
Table C1: Descriptive statistics

	Mean	Std. Dev.	Min.	Max.
Stay	0.72	0.45	0.00	1.00
Purchase	0.53	0.50	0.00	1.00
Network	0.33	0.29	0.00	1.00
Diversity	0.61	0.21	0.00	0.90
Plot size	18.27	7.75	6.00	68.00
Plot poor soil	0.42	0.49	0.00	1.00
Plot distance to borgo	2,574.90	1,347.06	41.51	7,622.96
Plot longitude	12.93	0.17	12.45	13.19
Plot latitude	41.45	0.10	41.26	41.70
Origin longitude	12.53	2.16	7.66	28.98
Origin latitude	44.66	1.59	36.81	46.53
Origin altitude	401.55	499.19	0.56	2,361.12
Origin malaria exposure	0.40	0.49	0.00	1.00
Origin wheat suitability	-0.00	1.00	-4.59	3.96
Agro-climatic similarity	-1.72	0.52	-3.44	-0.26

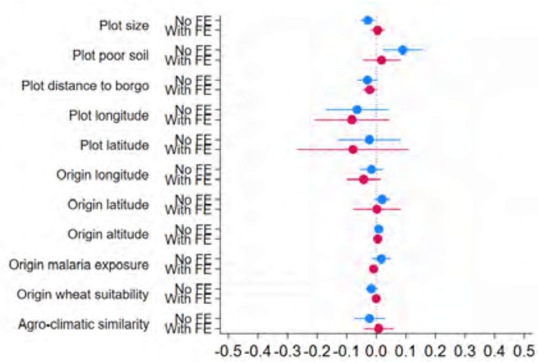
Figure C1 visualizes the results of a balance test, assessing whether the main explanatory variables used in our analysis—namely network, diversity, and isolation—are balanced with respect to observable characteristics of plots and origin municipalities. The balance checks, conducted with and without fixed effects, reveal imbalance only in origin longitude and soil quality, both of which are controlled for in our analysis.

Figure C2 complements the discussion in Section 4 by showing that the settler-level gap between observed and expected common-origin exposure is not systematically related to observable characteristics of plots and origin municipalities. This remains true even when we remove from the sample settlers from Veneto, who were overrepresented in the earliest immigration waves, or when plots that are occupied for the first time or resettled are considered separately.

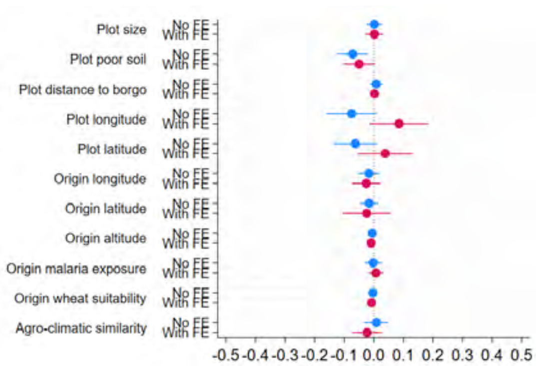
Figure C1: Balance checks with (red) and without (blue) fixed effects



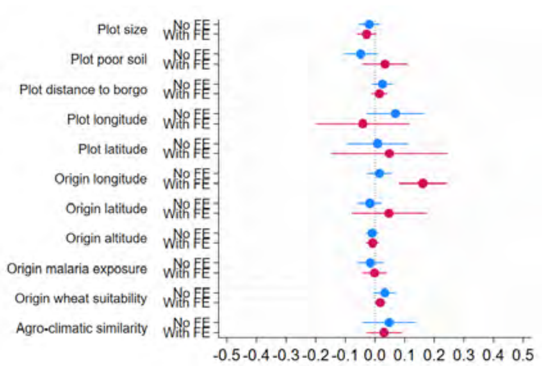
(a) Network



(b) Diversity

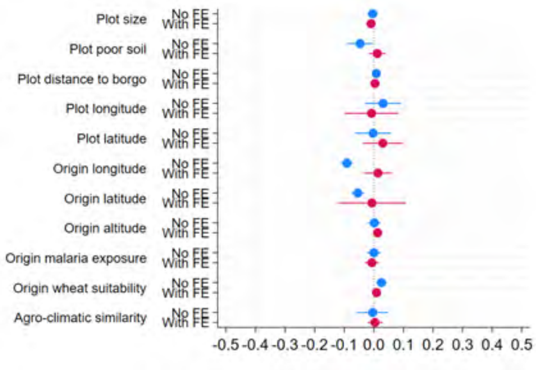


(c) Closest house empty

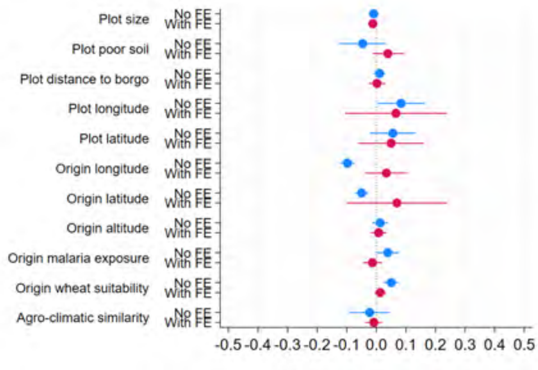


(d) Closest house network

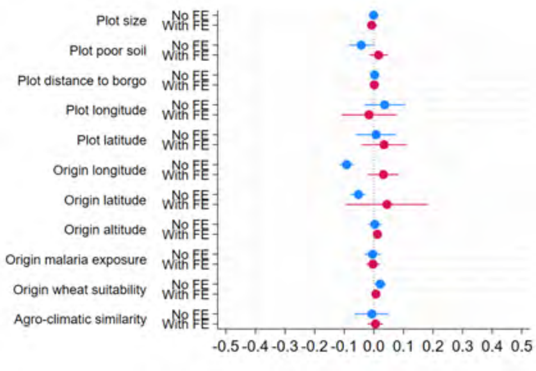
Figure C2: Balance check of the gap between observed-expected common-origin exposure



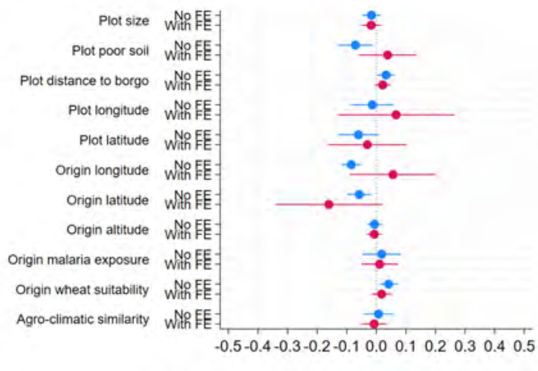
(a) All settlers



(b) Excluding Veneto



(c) Plot first time settled



(d) Plot being resettled

Notes: The figure plots the results of regressing the gap between observed and expected local common-origin exposure on our set of observable characteristics. Observed exposure is the share of settlers whose closest occupied plot is held by a settler from the same province, while expected exposure is implied solely by the province composition of settlers arriving in the same month. A positive (negative) value in the gap implies that the settlers arriving in that wave were disproportionally more (less) likely than expected to settle next to other families from the same origin.

D Robustness and extensions

D.1 Alternative fixed effects, clustering of standard errors

We start by showing how the main results concerning the effects of the network on economic success are robust to alternative sets of fixed effects. In particular, Table D1 reports the results obtained by running the same regression as in column (6) of Table 2, *i.e.*, the most demanding specification for our benchmark network regression with different destination and/or origin fixed effects. Table D2 experiments with alternative clustering approaches, including standard errors that allow for fine-level spatial correlation, which do not appear to alter the significance of the results.

Table D1: Impact of network with alternative fixed effects

	(1) Stay	(2) Stay	(3) Stay	(4) Purchase	(5) Purchase	(6) Purchase
Network	0.182***	0.182***	0.188***	0.179***	0.175***	0.191***
	(0.025)	(0.029)	(0.029)	(0.029)	(0.025)	(0.027)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Municipality	Azienda	Municipality	Municipality	Azienda	Municipality
Origin FE	Region	Province	Province	Region	Province	Province
Mean dep. variable	0.72	0.72	0.72	0.53	0.53	0.53
R^2	0.09	0.10	0.10	0.09	0.07	0.08
Observations	3,784	3,784	3,784	3,784	3,784	3,784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The network variable takes values in the interval [0,1]. Robust standard errors clustered at the *azienda* level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D2: Alternative standard errors

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Stay						
Network	0.181***	0.181***	0.181***	0.181***	0.181***	0.181***
	(0.024)	(0.025)	(0.025)	(0.041)	(0.018)	(0.009)
Mean dep. variable	0.72	0.72	0.72	0.72	0.72	0.72
R^2	0.09	0.09	0.09	0.09	0.09	0.09
Observations	3,784	3,784	3,784	3,784	3,784	3,784
Panel B: Purchase						
Network	0.166***	0.166***	0.166***	0.166***	0.166***	0.166***
	(0.027)	(0.033)	(0.030)	(0.039)	(0.029)	(0.033)
Standard errors	Azienda	Municipality	Origin province	Arrival year	Spatial (5km)	Spatial (10km)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.53	0.53	0.53	0.53	0.53	0.53
R^2	0.08	0.08	0.08	0.08	0.08	0.08
Observations	3,784	3,784	3,784	3,784	3,784	3,784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Standard errors in parentheses clustered at different levels: (1) destination *azienda*, (2) destination municipality, (3) province of origin, (4) arrival year, (5) spatial autocorrelation at 5km, and (6) spatial auto-correlation at 10km following Conley (1999). Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Network effects conditional on staying or leaving

Table D3 shows the effects of network size and a network member in the closest plot on various margins: purchase conditional on staying until 1940 in the full sample, purchase among those who did stay until 1940, and years of stay in the marshes among those who left before 1941. The results suggest that among those who stay until 1941 (columns 2 and 5), a more extensive network increased the probability of redeeming the land. When compared to Table 2, the coefficients are much smaller, and suggest that about 2/3 of the effects on purchase go through a higher likelihood to stay, while about 1/3 through a direct impact beyond staying. We also find evidence that among settlers who left before 1941 (columns 3 and 6), the network size is positively correlated with the duration of stay, with a 0 to 1 expansion of the network associated with 3 additional months of permanence in the Pontine Marshes.

Table D3: Impact of networks conditional on staying/leaving

	(1)	(2)	(3)	(4)	(5)	(6)
	Purchase	Purchase	Duration of stay	Purchase	Purchase	Duration of stay
Network	0.036	0.049	0.291*			
	(0.025)	(0.033)	(0.165)			
Stay	0.719***			0.719***		
	(0.018)			(0.018)		
Closest plot same province				0.036**	0.048**	-0.013
				(0.016)	(0.021)	(0.148)
Controls	Yes	Yes	Yes	Yes		
Full FE	Yes	Yes	Yes	Yes		
Mean dep. variable	0.53	0.74	3.25	0.53	0.74	3.25
R^2	0.46	0.06	0.17	0.46	0.06	0.17
Observations	3,784	2,734	1,050	3,784	2,734	1,050

Notes: Columns 1 and 4 rely on the full sample (arriving between 1932 and 1940), columns 2 and 5 on those who stay until 1940, and columns 3 and 6 on those who leave before 1941. The network variable takes values in the interval [0,1]. Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.3 Alternative diversity measures: Polarization

In Table D4, we run the same regressions as in Table 3, while measuring diversity using polarization rather than fractionalization. As with our preferred fractionalization index, polarization is computed within the 1-km buffer surrounding each settler's farmhouse and based on province of origin. Following Montalvo and Reynal-Querol (2005) and Bazzi et al. (2019), we measure polarization as:

$$P_i = 4 \sum_p q_{i,p}^2 \left(1 - q_{i,p}\right),$$

where $q_{i,p}$ denotes the share of neighbors in the buffer of settler i originating from province p.

The index is maximized when the buffer is split into two equally-sized provincial groups, capturing the potential for segregation within large dominant groups rather than overall origin heterogeneity. Figure D1 illustrates the relationship between fractionalization and polarization in our sample, highlighting a very similar pattern as in other settings Bazzi et al. (2019). We also compute an ”outgroup” version of such measure by restricting the calculation to neighbors who do not originate from settler i's own province.

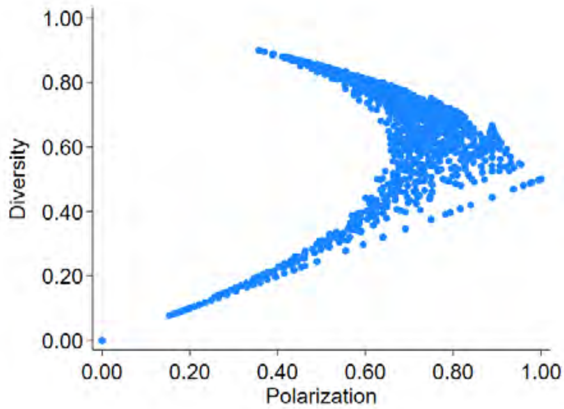


Figure D1: Relationship between fractionalization and polarization

Table D4: Impact of polarization

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Stay			
Polarization (all)	0.016		0.113*		0.270***	0.273***
	(0.049)		(0.055)		(0.074)	(0.078)
Polarization (outgroup)		0.006		0.033		
		(0.030)		(0.022)		
Network (% same province)			0.206***	0.122***		
			(0.024)	(0.027)		
			Panel B: Purchase			
Polarization (all)	0.004		0.091*		0.224*	0.244**
	(0.047)		(0.050)		(0.108)	(0.105)
Polarization (outgroup)		0.023		0.036		
		(0.028)		(0.029)		
Network (% same province)			0.186***	0.139**		
			(0.029)	(0.049)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	Network	Network
					< median	= 0
Observations	3784	3784	3784	3784	1863	734

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The measure of polarization is based on Montalvo and Reynal-Querol (2005). Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.4 Effects of spatial features on the probability of purchasing the land

In Table 4 of Section 5.3, we have seen how some specific spatial features of the network influence the probability to remain in the Pontine Marshes until 1941, even when controlling for network size. Table D5 reports the results of running the same regression, using our purchase variable as the main outcome.

Similar to Table 4, distance to network members (both average and to the closest neighbor) emerge as a key determinant of economic success. In addition, the residual effect after controlling for network size appears to be more significant than in the case of the ”stay” outcome.

Table D5: Impact of spatial features of the network on the probability of purchasing the land

	(1)	(2)	(3)	(4)	(5)
Panel A: not controlling for network size					
Closest plot empty	-0.076**				
	(0.027)				
Closest plot network (same province)		0.065***			
		(0.020)			
Distance to the first network family			-0.148***		
			(0.049)		
Avg. distance to network families				-0.149*	
				(0.075)	
Avg. distance among network families					-0.061
					(0.085)
Panel B: controlling for network size					
Closest plot empty	-0.074**				
	(0.026)				
Closest plot network (same province)		0.029			
		(0.023)			
Distance to the first network family			-0.119**		
			(0.056)		
Avg. distance to network families				-0.141*	
				(0.076)	
Avg. distance among network families					-0.034
					(0.089)
Network	0.164***	0.141***	0.078	0.129***	0.130**
	(0.027)	(0.029)	(0.050)	(0.043)	(0.045)
Observations	3,784	3,784	2,795	2,795	2,795
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All distances are expressed in km. The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the *azienda* level in parentheses. The network variable takes values in the interval [0,1].

D.5 Alternative definitions of common origin

In Section 5.1, we presented our main results, based on defining common origin at the province level. Here, we explore the robustness of our results to alternative definitions of common origin.

We first rely on alternative administrative units, namely, regions and municipalities, to measure the extent of common-origin networks. Table D6 shows how a more extensive regional or municipal network has a positive impact on 'stay' and 'purchase', but in a horse race involving all the measures, our baseline province-level indicator is the only one that remains significant and retains a sizable coefficient, thus supporting our choice of the province as a marker of common identity.

Table D6: Impact of networks at the regional and municipality levels

	(1) Stay	(2) Stay	(3) Stay	(4) Purchase	(5) Purchase	(6) Purchase
Network (% same region)	0.142***		0.027	0.161***		0.081
	(0.033)		(0.052)	(0.029)		(0.058)
Network (% same municipality)		0.158**	-0.023		0.201**	0.044
		(0.062)	(0.064)		(0.077)	(0.069)
Network (% same province)			0.169***			0.096
			(0.038)			(0.055)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.72	0.72	0.72	0.53	0.53	0.53
R^2	0.09	0.09	0.09	0.08	0.08	0.08
Observations	3,784	3,784	3,784	3,784	3,784	3,784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We also experiment with "artificial provinces", *i.e.*, buffers of variable radius centered around the settler's municipality of origin. In particular, Table D7 shows that the effect of the network tends to diminish when we define common-origin on the basis of increasingly larger areas, varying the radius from 10 to 150 km. When performing a horse race among multiple "artificial provinces", the largest effect appears to be related to buffers of 25- and 50-km radius (see column (6)), which roughly correspond to the median size of Italian provinces. In column (7), we can also see how the benchmark network measure based on actual provinces outperforms all the "artificial" alternatives.

Table D7: Alternative definitions of common-origin networks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Stay							
Network (families 10km from origin)	0.183***					0.040	0.024
	(0.060)					(0.079)	(0.077)
Network (families 25km from origin)		0.168***				0.024	-0.014
		(0.037)				(0.057)	(0.056)
Network (families 50km from origin)			0.152***			0.165***	0.095
			(0.025)			(0.057)	(0.065)
Network (families 100km from origin)				0.092***		-0.061	-0.086
				(0.026)		(0.059)	(0.055)
Network (families 150km from origin)					0.089***	0.007	0.003
					(0.030)	(0.038)	(0.038)
Network (same province)							0.148***
							(0.029)
Mean dep. variable	0.72	0.72	0.72	0.72	0.72	0.72	0.72
Observations	3,764	3,764	3,764	3,764	3,764	3,764	3,764
Panel B: Purchase							
Network (families 10km from origin)	0.234***					0.095	0.084
	(0.066)					(0.089)	(0.086)
Network (families 25km from origin)		0.197***				0.092	0.064
		(0.036)				(0.067)	(0.066)
Network (families 50km from origin)			0.141***			0.101	0.048
			(0.028)			(0.070)	(0.076)
Network (families 100km from origin)				0.081***		-0.066	-0.085
				(0.027)		(0.076)	(0.076)
Network (families 150km from origin)					0.079*	-0.001	-0.004
					(0.038)	(0.077)	(0.077)
Network (same province)							0.111**
							(0.042)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.53	0.53	0.53	0.53	0.53	0.53	0.53
Observations	3,764	3,764	3,764	3,764	3,764	3,764	3,764

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The network variable takes values in the interval [0,1]. Robust standard errors clustered at the *azienda* level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D8 shows additional results obtained by replacing the benchmark province-based network measures with alternative ones that, as explained in Section 5.4, contain information on geographic and linguistic distance between the origin municipalities of settler *i* and his neighbors.

Table D8: Impact of average geographical and linguistic distance

	(1)	(2)	(3)	(4)	(5)	(6)
	Stay	Stay	Stay	Purchase	Purchase	Purchase
Geographic distance (z-score)	-0.026**		0.011	-0.035***		-0.007
	(0.012)		(0.013)	(0.010)		(0.017)
Linguistic distance (z-score)		-0.001	0.007		-0.001	0.008
		(0.005)	(0.005)		(0.010)	(0.010)
Network			0.180***			0.147***
			(0.025)			(0.033)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.73	0.73	0.73	0.53	0.53	0.53
R^2	0.08	0.08	0.09	0.08	0.08	0.08
Observations	3,686	3,666	3,666	3,686	3,666	3,666

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. The network variable takes values in the interval [0,1]. Robust standard errors clustered at the *azienda* level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.6 Alternative buffer size at destination

Throughout the main text, our neighborhood composition variables are computed by considering all families located within a 1-km buffer centered on family i. Table D9 considers two types of alternative neighborhoods: (i) buffers of different radii (500, 1000, and 1500 meters), and (ii) donut-type areas which exclude the inner space. The results highlight that the benchmark 1-km network measure provides the most statistically meaningful results, systematically outperforming all the alternative indicators.

Table D9: Impact of networks measured in alternative buffers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Stay								
Network (same province, 1000m buffer)	0.207***	0.145**	0.171***	0.160***	0.163***	0.179***	0.165***	
	(0.031)	(0.058)	(0.058)	(0.041)	(0.043)	(0.027)	(0.027)	
Network (same province, 500m buffer)	-0.027		-0.027					0.023
	(0.024)		(0.024)					(0.022)
Network (same province, 1500m buffer)		0.047	0.046					
		(0.065)	(0.065)					
Network (same province, donut 500-750m)				0.033				0.076***
				(0.030)				(0.021)
Network (same province, donut 750-1000m)					0.010			0.054*
					(0.035)			(0.028)
Network (same province, donut 1000-1250m)						0.008		0.002
						(0.019)		(0.023)
Network (same province, donut 1250-1500m)							0.030	0.030
							(0.029)	(0.034)
Panel B: Purchase								
Network (same province, 1000m buffer)	0.164***	0.083	0.080	0.205***	0.131**	0.144***	0.151***	
	(0.045)	(0.071)	(0.081)	(0.059)	(0.051)	(0.037)	(0.028)	
Network (same province, 500m buffer)	0.002		0.003					0.059
	(0.043)		(0.043)					(0.036)
Network (same province, 1500m buffer)		0.108	0.108					
		(0.079)	(0.079)					
Network (same province, donut 500-750m)				-0.027				0.025
				(0.042)				(0.025)
Network (same province, donut 750-1000m)					0.035			0.059
					(0.043)			(0.039)
Network (same province, donut 1000-1250m)						0.038		0.029
						(0.036)		(0.046)
Network (same province, donut 1250-1500m)							0.021	0.012
							(0.030)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,784	3,784	3,784	3,535	3,628	3,685	3,705	3,362

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Network variables take values in the interval [0,1]. Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.7 Absolute *vs* relative network

In Section 5, all our results have been established using the share of neighbors originating from the same province as a measure of network extensiveness. Table D10 moves from the relative to an absolute measure of the network, by showing the results obtained when using the number of same-province neighbors as the main explanatory variable. Among others, we can see that absolute network correlates positively with staying in the Pontine Marshes until 1941 and buying the land. The absolute measure, however, loses the horse race against its relative counterpart, thus supporting our choice of the latter as the benchmark explanatory variable.

Table D10: Impact of total network families

	(1) Stay	(2) Stay	(3) Stay	(4) Stay	(5) Purchase	(6) Purchase	(7) Purchase	(8) Purchase
Total network families (1000m buffer)	0.012***	0.011***	0.012***	0.006	0.010***	0.011***	0.011***	0.005
	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)
Total number of plots (1000m buffer)		0.001				-0.001		
		(0.002)				(0.002)		
Total non-network families (1000m buffer)			0.001				0.001	
			(0.001)				(0.002)	
Network				0.111**				0.112**
				(0.039)				(0.052)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3784	3784	3784	3784	3784	3784	3784	3784

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Network refers to families from the same province of origin. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.8 Sensitivity to excluding specific groups of settlers

Our main analysis uses the full sample of settlers arriving between 1932 and 1940. Here we assess whether our results may be influenced by settlers who faced distinctive circumstances. In columns 1 and 4, we exclude individuals who settled a plot formerly occupied by another family. These individuals may have had an advantage, as the plot was fully ready to exploit and had been farmed before; yet they settled in an area with already-established connections, the effect of which is ambiguous. In columns 2 and 5, we exclude individuals who came from nearby areas, namely the Lazio province. These settlers could have been more familiar with the area, had different outside options, or been able to preserve social ties outside the marshes. In columns 3 and 6, we exclude individuals who came from outside of Italy. These settlers were recruited under particular circumstances and may have been subject to differential treatment. Across all specifications, the estimated coefficients remain very similar in magnitude and statistical significance to those in our baseline, suggesting that our findings are not driven by any of these particular subgroups of settlers.

Table D11: Impact of network excluding specific groups of settlers

	Stay			Purchase		
	(1)	(2)	(3)	(4)	(5)	(6)
Network	0.202***	0.191***	0.172***	0.184***	0.187***	0.148***
	(0.031)	(0.026)	(0.025)	(0.034)	(0.035)	(0.024)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.72	0.73	0.72	0.54	0.54	0.52
R^2	0.11	0.08	0.09	0.09	0.08	0.08
Observations	2,917	3,250	3,617	2,917	3,250	3,617

Notes: Columns 1 and 4 exclude settlers who occupy a plot that had a settler family in the past. Columns 2 and 5 exclude settlers who come from nearby areas (Lazio province). Columns 3 and 6 exclude settlers who come from outside of Italy. Robust standard errors clustered at the *azienda* level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.9 Networks and bonding social capital

This section reports the regression results underlying the discussion of bonding social capital in Section 6.4. We proxy for bonding social capital using the prevalence of consanguineous marriage in the province of origin, measured as cousin marriages per thousand people in 1932–1935 (Zei et al. 2005) and standardized. Higher values characterize origin communities in which cooperation was historically mediated more by kinship than by generalized trust. Because consanguinity varies little within regions, we omit region-of-origin fixed effects, which would absorb most of its variation; all other controls and fixed effects are as in our baseline specification.

Table D12 reports three exercises. Columns (1)–(2) split the sample at the median of consanguinity: network effects on both outcomes are roughly twice as large among settlers from high-consanguinity origins. Columns (3)–(4) interact network size with the standardized index in the full sample, yielding positive and significant interactions for both staying and purchasing. Columns (5)–(6) estimate the same interaction separately for settlers from the Po Valley and elsewhere. The interaction remains positive and significant within the Po Valley, where Figure 11 documents a discontinuity in consanguinity at the river, but is small and imprecise elsewhere. This pattern suggests that the results are driven by variation along the Po boundary rather than by broader North–South differences.

Table D12: Consanguinity and Network Effects

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Stay			
Network	0.084**	0.191***	0.143***	0.175***	0.201***	0.233*
	(0.035)	(0.031)	(0.024)	(0.025)	(0.031)	(0.113)
Consanguinity			0.027*	-0.007	-0.003	-0.005
			(0.014)	(0.015)	(0.019)	(0.027)
Network x Consanguinity				0.119***	0.135***	-0.061
				(0.020)	(0.036)	(0.136)
Mean dep. variable	0.72	0.72	0.72	0.72	0.73	0.69
R^2	0.05	0.05	0.04	0.04	0.04	0.07
Observations	1,533	2,249	3,497	3,497	2,607	890
			Panel B: Purchase			
Network	0.124*	0.186***	0.168***	0.191***	0.243***	0.181
	(0.066)	(0.036)	(0.040)	(0.037)	(0.048)	(0.120)
Consanguinity			0.048***	0.023	0.041	0.033
			(0.014)	(0.019)	(0.024)	(0.028)
Network x Consanguinity				0.089**	0.147***	-0.027
				(0.034)	(0.050)	(0.148)
Mean dep. variable	0.52	0.54	0.53	0.53	0.53	0.50
R^2	0.04	0.04	0.03	0.03	0.04	0.05
Observations	1,533	2,249	3,497	3,497	2,607	890
Sample	Low consang.	High consang.	Full	Full	Po Valley	Outside Po Valley
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Columns (1)-(2) split the sample at the median of consanguinity. Columns (5)-(6) split by Po Valley origin. The consanguinity variable measures the share of cousin marriages in the province and is further standardized. The specification does not include region fixed effects because the consanguinity measure has limited variation within a region. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E Measures of agroclimatic similarity

E.1 Overall agroclimatic similarity

This section details how we constructed the agroclimatic similarity index used as a proxy for skill transferability.

The index is computed for each settler by comparing nine elements of agroclimatic conditions between the municipality of origin and the assigned plot in the Pontine Marshes. Primary data is collected from two high-resolution raster datasets: the FAO Harmonized World Soil Database (HWSD) and the Climatic Research Unit TS4.07 database at the University of East Anglia (CRU):

- Average altitude (HWSD);
- Altitude's standard deviation (HWSD);
- Distance from the closest river (HWSD/own Pontine Marshes map);
- Soil nutrient availability (HWSD). Index based on soil texture, organic carbon, and pH;
- Soil rooting conditions (HWSD). Index based on soil textures, coarse fragments, and vertic soil properties;
- Oxygen availability (soil drainage, HWSD);
- Soil salinity and sodificity (HWSD);
- Average precipitation (CRU);
- Average temperature (CRU);

For climate variables (precipitation and temperature), values are extracted at the centroid of each origin municipality, defined according to the 1991 ISTAT administrative boundaries, and at the coordinates of each plot in the Pontine Marshes. Monthly grids are first aggregated to annual values by averaging across the 12 months of the year; the resulting annual series is then averaged over the period 1900–1930.

For each settler i and element e of the list above, we compute the standardized distance between origin municipality o and destination plot p as:

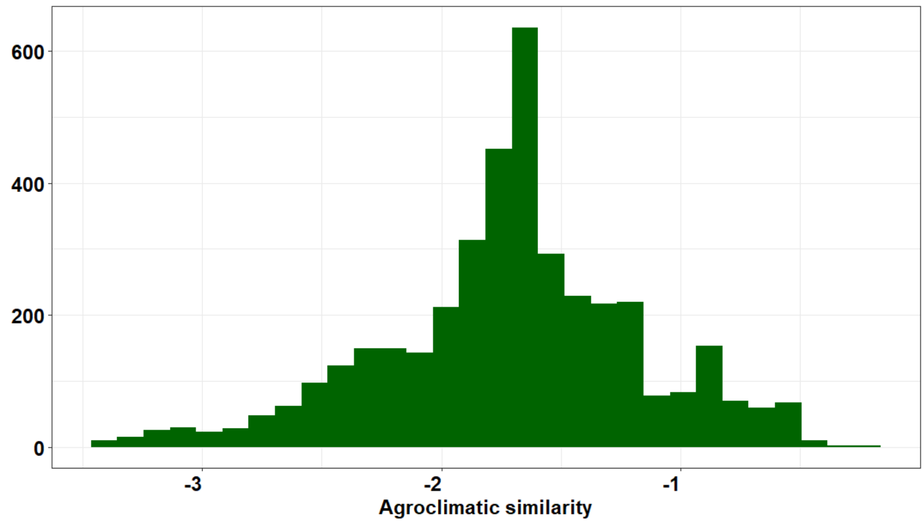
$$d_i^e = \left|\frac{\bar{e}_i^o - \bar{e}_i^p}{\sigma_e^o}\right|$$

The overall index is then computed as:

$$A_i = -\frac{\sum_e (d_i^e)}{9}$$

where nine is the total number of elements included. The minus sign is included so that the distribution is entirely below 0 and lower values correspond to higher agroclimatic distance (0 being agroclimatic equivalence). Figure E1 plots the distribution of A_i. The mode of the distribution is around -1.5, meaning the typical settler faced an environment differing from that of their origin municipality by about 1.5 standard deviations on average across dimensions.

Figure E1: Agroclimatic similarity: distribution



In our analyses, we use both the settler-level agroclimatic similarity and the averages computed over neighbors within the 1-km radius buffer.

E.2 Wheat-specific agroclimatic similarity

In addition to the general agroclimatic similarity index described above, we construct a measure that focuses specifically on land suitability for wheat cultivation. This is motivated by the central role of wheat in the colonization of the Pontine Marshes, as it was the main crop imposed onto settlers by the ONC in the framework of Mussolini's *"Battle for Grain"*.

We rely on the FAO Global Agro-Ecological Zones (GAEZ) database, which provides high-resolution, crop-specific measures of land suitability. The GAEZ wheat suitability index combines several determinants of wheat-growing potential, including climate variables (temperature and precipitation, based on the 1960s-1990s reference period), terrain features and soil characteristics from the HWSD FAO database, water availability, and the possibility of relying on irrigation and groundwater. The result is a single score, with higher values reflecting more favorable conditions for wheat cultivation.

For each origin municipality, we compute the average of the wheat suitability index across all the raster pixels falling within the municipality's administrative boundaries. For each Pontine Marshes plot, we instead take the value of the index in the pixel where the farmhouse is located.

To capture the distance between origin and destination in terms of wheat-growing potential, we then compute, for each settler i, the absolute difference between the wheat suitability of the origin municipality o and that of the assigned plot p:

$$w_i = -\frac{\left|W_i^o - W_i^p\right|}{\sigma_w^o},$$

where W_i^o and W_i^p denote the wheat suitability indices at origin and destination, respectively. σ_w^o is the standard deviation of the wheat suitability distribution at origin, thus obtaining a standardized measure of distance.

F Heterogeneity

Here, we present a series of tables that assess the heterogeneity of our results across different dimensions.

Tables F1 and F2 show how the effects of networks and diversity vary by settlers' region of origin. Table F1, in particular, complements Figure 9 in Section 6.4 by highlighting a stark difference in the importance of networks for settlers from Veneto compared with those from Emilia-Romagna–the two main regions of origin in our sample. The results on diversity are consistently negative but imprecisely estimated once splitting the sample.

Table F1: Impact of network

	(1)	(2)	(3)	(4)	(5)
Panel A: Stay					
Network (% same province)	0.217***	0.107	0.143	0.272*	0.253***
	(0.041)	(0.086)	(0.092)	(0.140)	(0.034)
Mean dep. variable	0.76	0.65	0.67	0.72	0.76
R^2	0.06	0.10	0.21	0.13	0.39
Observations	1,855	761	534	350	284
Panel B: Purchase					
Network (% same province)	0.211***	0.099	0.165*	0.198	0.488***
	(0.047)	(0.072)	(0.095)	(0.152)	(0.082)
Subsample	Veneto	Emilia Romagna	Lazio	Friuli Venezia Giulia	Other regions
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.57	0.45	0.48	0.54	0.58
R^2	0.06	0.13	0.13	0.15	0.34
Observations	1,855	761	534	350	284

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Tables F3, F4 and F5 explore heterogeneity by year of arrival and soil quality, according to the ONC classification. We find that while the role of networks and diversity does not vary substantially by arrival year, settlers with poor-quality soil are particularly vulnerable to the adverse effects of diversity.

Table F2: Impact of diversity

	(1)	(2)	(3)	(4)	(5)
			Panel A: Stay		
Diversity (all)	-0.134*	-0.131	0.066	-0.027	-0.108
	(0.065)	(0.160)	(0.088)	(0.218)	(0.102)
Mean dep. variable	0.76	0.65	0.67	0.72	0.76
R^2	0.05	0.10	0.21	0.12	0.38
Observations	1,855	761	534	350	284
			Panel B: Purchase		
Diversity (all)	-0.105	-0.150	0.079	0.034	-0.227*
	(0.094)	(0.165)	(0.107)	(0.198)	(0.126)
Subsample	Veneto	Emilia Romagna	Lazio	Friuli Venezia Giulia	Other regions
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.57	0.45	0.48	0.54	0.58
R^2	0.05	0.13	0.13	0.15	0.30
Observations	1,855	761	534	350	284

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table F3: Heterogeneous effects of network

	(1)	(2)	(3)	(4)	(5)
			Panel A: Stay		
Network	0.219***	0.179***	0.113**	0.161***	0.173***
	(0.062)	(0.028)	(0.041)	(0.043)	(0.033)
Mean dep. variable	0.72	0.64	0.79	0.67	0.76
R^2	0.09	0.10	0.09	0.12	0.08
Observations	1,215	1,139	1,430	1,589	2,195
			Panel B: Purchase		
Network	0.304***	0.137***	0.126**	0.201***	0.145***
	(0.076)	(0.039)	(0.053)	(0.050)	(0.031)
Subsample	1932-33	1934-35	1936-40	Poor soil	Better soil
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.55	0.46	0.57	0.50	0.56
R^2	0.10	0.10	0.09	0.13	0.07
Observations	1,215	1,139	1,430	1,589	2,195

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. The network variable takes values in the interval [0,1].

Table F4: Heterogeneous effects of diversity (all)

	(1)	(2)	(3)	(4)	(5)
			Panel A: Stay		
Diversity (all)	-0.099	-0.095	-0.070	-0.244***	-0.021
	(0.055)	(0.112)	(0.047)	(0.069)	(0.048)
Mean dep. variable	0.72	0.64	0.79	0.67	0.76
R^2	0.08	0.09	0.09	0.12	0.07
Observations	1,215	1,139	1,430	1,589	2,195
			Panel B: Purchase		
Diversity (all)	-0.070	-0.121	-0.061	-0.259***	-0.018
	(0.085)	(0.116)	(0.064)	(0.087)	(0.058)
Subsample	1932-33	1934-35	1936-40	Poor soil	Better soil
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.55	0.46	0.57	0.50	0.56
R^2	0.09	0.10	0.08	0.13	0.06
Observations	1,215	1,139	1,430	1,589	2,195

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table F5: Heterogeneous effects of diversity (out-group)

	(1)	(2)	(3)	(4)	(5)
			Panel A: Stay		
Diversity (out-group)	-0.018	-0.020	0.024	-0.150**	0.059
	(0.037)	(0.084)	(0.039)	(0.068)	(0.039)
Mean dep. variable	0.72	0.64	0.79	0.67	0.76
R^2	0.08	0.09	0.09	0.12	0.07
Observations	1,215	1,139	1,430	1,589	2,195
			Panel B: Purchase		
Diversity (out-group)	0.043	-0.039	0.028	-0.145*	0.054
	(0.053)	(0.090)	(0.048)	(0.069)	(0.050)
Subsample	1932-33	1934-35	1936-40	Poor soil	Better soil
Controls	Yes	Yes	Yes	Yes	Yes
Full fixed effects	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.55	0.46	0.57	0.50	0.56
R^2	0.09	0.10	0.08	0.13	0.06
Observations	1,215	1,139	1,430	1,589	2,195

Notes: The sample comprises all the settlers who arrived between 1932 and 1940. Robust standard errors clustered at the azienda level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.