



“Migrants know better”: migrants’ networks and FDI

Giorgia Giovannetti² · Filippo Santi¹  · Margherita Velucchi³

Received: 15 February 2023 / Accepted: 16 January 2024
© The Author(s) 2024

Abstract

The interrelations between trade, FDI and migrations have an important impact on the global economy that resulted in the golden age of globalization at first and in a process of slowbalization in the last decade. Against this background, this paper focuses on migrations and FDI and shows that migrant networks encourage cross-border investments. The presence of immigrants is likely to attract new FDI from their country of origin. FDI abroad, furthermore, are positively related to the presence of migrants, thanks to their knowledge of the two markets. We apply a multilevel mixed model to disentangle the hierarchical structure of the data in order to test the relations between FDI and (directional) migration flows. More specifically we test how and to what extent the structure of the international migrants’ network contributes to bilateral FDI flows, besides standard models. Results show that migrants’ networks exploit the information migrants have on both source and destination markets, and that a more diverse migrants’ community in investing countries lowers bilateral FDI.

Keywords International Migration · FDI · Third country effect · Network

JEL Classification F21 · F22

1 Introduction and relevant literature

Over the past decades, international migrations have increased, making the world increasingly interconnected but also more “diverse”. At the same time, also foreign direct investment (FDI) has increased, together with the number of destination countries. In this paper, we show that immigration and FDI are closely related: more

✉ Filippo Santi
filippo.santi6@unibo.it

¹ University of Bologna, Bologna, Italy

² University of Florence and European University Institute, Fiesole, Italy

³ European University of Rome, Rome, Italy

specifically we show that networks of migrants trigger FDI, mainly by facilitating the flow of information (i.e. through the so-called “information channel”).

The interest on the economic role played by migrants’ networks has been growing over time, but so far, the focus has mainly been on the relationship between bilateral trade flows and migrants’ networks. The existing literature has shown that migrants establish a social and (in)formal infrastructure across countries, facilitating communication and economic exchanges (Leblang, 2010). The existing evidence suggests that this relationship is driven by mechanisms affecting both the demand-side (e.g. White, 2007 discusses a “transplanted-home bias”), and the supply-side (e.g. Rauch, 2001’s “Business and Social Network Effect”). Among others, Gould (1994), Rauch and Trindade (2002), Peri and Raquena-Silvente (2010), Fensore (2016), Giovannetti and Lanati (2017) and Parsons and Vèzina (2018) find strong support in favor of a positive effect of migrants’ network on trade flows.

The role of migrants’ networks on capital flows has been less explored and results are less clear-cut, despite the apparent simplicity of the underlying mechanism. Migrants’ network effect, coupled with the privileged information migrants have about the origin-country market, should reduce the transaction costs, and migration would act as a complement for FDI (in the same way as it does for trade). Leblang (2011) suggests that the higher risk typically associated to cross border investments makes any available information on the investment environment extremely valuable. Against this background, the empirical evidence is mixed: Gheasi and Nijkamp (2017), Aroca and Mahoney (2005) and Checchi et al. (2007), among others, detect a negative impact of bilateral migration on FDI, suggesting that human mobility and FDI could substitute each other. Portes and Rey (2005) and Kugler and Rapoport (2007) highlight an heterogeneous relationship, with a negative impact mostly related to the migrants’ education/skill endowment, a dimension often overlooked in the early research.¹ Controlling for the distribution of migrants across different educational levels, Kugler and Rapoport (2007), Docquier and Lodigiani (2010), Javorcick et al. (2011), Cuadros et al., (2016, 2019) and Burchardi et al. (2019) detect a positive impact of better educated migrants’ on bilateral investment flows, suggesting that migration and investments are ultimately complementary.² Considering different types of FDI, Chan and Zheng (2022) find that migrants networks positively affected Mergers and Acquisition in China, and had no effect on Greenfield FDI. All these studies, however, focused on migration within a country or as a mostly bilateral phenomenon. Following a different approach and considering migration as a complex, globally interconnected phenomenon, Fagiolo and Mastroillo (2014) test whether and how the migrants’ “business and social network effect” at bilateral level could also be affected by the overall structure of the International Migration Network (IMN), i.e., by how international migration

¹ FDI constitute a much more complex phenomenon with respect to trade. The evidence suggests that better educated individuals, whose social capital and informative capacity is likely to be larger than their less educated fellows, are more effective in facilitating FDI and capital acquisitions (both back home and in their new destinations).

² Other studies include Federici and Giannetti (2010), Flisi and Murat (2011), De Simone and Manchin (2012), D’Agosto et al. (2013) and Wang (2017).

connects countries globally. The way countries integrate with each other is likely to affect not only the way information on the investment environment in a potential receiving country reaches potential investors, but also how different streams of information from different locations is pooled and elaborated by the investors themselves. Sgrignoli et al. (2015) and Metulini et al. (2018) find evidence that the efficiency of migrants in providing valuable information depends on the size of bilateral migration and on how well both investing and recipient economies integrate in the global migration system. Looking at the business and tourism travel network, Fagiolo and Santoni (2014) explore whether better integration leads to larger productivity gains, through the effect of international mobility on trade flows. Garas et al. (2016) detected a positive correlation between a country's integration in the IMN and its bilateral direct investment position.³ To the best of our knowledge, two main aspects are under-investigated. First, there is no discussion on skill and educational heterogeneity of migrants' network within a complex network perspective, despite evidence suggesting that the educational endowment of migrants' network is crucial for its impact on investments. Second, the relevance of the indirect network effects has been only partially explored, leaving a substantial gap in the understanding of the relationship between migration and FDI flows.

In this paper, we bring together the two strands of literature—gravity and networks—to explore to what extent FDI outflows from 20 OECD countries can be explained by the immigrant network from the rest of the world. We aim at investigating the relationship between the direct and the indirect effect of migration on FDI, considering how countries integrate with respect to the global migratory flows beyond the pure bilateral perspective. We generalize a gravity model to a multilevel mixed approach (Goldstein, 2011; Rabe-Hesketh & Skrondal, 2008) for dyadic data. This choice allows us to model the hierarchical structure of the data, endogenizing the heterogeneity from the country-level network statistics in a bilateral framework. The contribution of this paper is twofold. First, we explore the network structure of global migration, and we study how the position of a country in the International Migration Network (IMN) influences bilateral investments decisions. Despite a substantial agreement on the positive impact of a country's position in the IMN, two aspects remain largely under-explored: (1) how this position affects bilateral investment flows and (2) the distinction between emigration and immigration flows. The existing evidence doesn't clarify if and how a more diversified immigrant community triggers FDI. To test for the “diversion effect”, we use directed (i.e., asymmetric) network statistics—distinguishing the inward connectivity of an investing country in the IMN and the outward connectivity of the recipient country. Second, we link the literature on the heterogeneous economic effect of migrants to network theory. The existing evidence suggests that given that investments are a more complex internationalization mode than trade, the

³ In line with Garas et al. (2016) we study the migrants' network effect on OECD investment flows toward the rest of the world but, instead of focusing on total migration, we control for the bilateral stock of educated migrants (consistently with the existing evidence on the heterogeneous effect of migrants' network on FDI). However, as opposed to Garas et al. (2016) we preserve the direct nature of the IMN and propose an alternative estimation strategy correcting their the dependent variable and error's misspecification.

mechanisms at work may require high skilled human capital. Kugler and Rapoport (2007), Flisi and Murat (2011), Javorcik et al. (2011), Leblang (2011), Foad (2012), Simone and Manchin (2012) and D'Agosto et al. (2013), among others, find evidence that the skill composition of the diaspora affects both inward and outward investments flows.⁴ However, the studies on the way the migrants' skill composition affects FDI do not take into account the role and the characteristics of migrants' network. The inclusion of both bilateral and (directed) network-related measures of international migration also helps disentangling the migration's direct effect and the indirect role played by third parties—as well as the structure of IMN itself—in shaping FDI patterns from the OECD to the rest of the world.⁵

Our findings are consistent with the existing literature and show that the bilateral socioeconomic dimensions are important drivers of bilateral (greenfield) FDI flows. FDI are also positively influenced by better educated (higher skills) migrants. The results are robust to the overall network structure and suggest that bilateral FDI are both directly and indirectly affected by the degree of integration of the migrants' network. As the presence of an established bilateral migration corridor directly affects the economic exchanges between two countries, bilateral FDI appear to be positively influenced by a country central position in the global migratory network, both in terms of connections' size and quality.

Contrary to previous studies, we find that bilateral FDI between any two countries are (i) negatively affected by a more heterogeneous immigrant community in the investing country; (ii) positively related with the number of common connections between the two in the IMN; and (iii) positively affected by either country's network centrality in the IMN, which helps overcoming weaker informative signals coming from relatively smaller direct migratory channels between countries. These findings suggest that, while overall direct and indirect migration positively contribute to bilateral FDI, network effects in the IMN can have a heterogeneous impact on bilateral FDI.⁶

The paper is structured as follows. Section 2 describes the data on bilateral greenfield FDI and the migration network; Sect. 3 introduces the econometric approach and explains the Multilevel Mixed Effect Model. Results are discussed in Sects. 4 and 5 draws conclusions.

⁴ Similarly, the skill composition of the migratory network influences trading patterns between home and host countries. See for instance Peri and Raquena-Silvente (2010) and Giovannetti and Lanati (2017).

⁵ Despite the relatively limited size of the sample considered due to data availability—which accounts for less than 20 percent of global bilateral migration channels) the results are still representative of more than 40 percent of the overall strictly positive greenfield FDI channels worldwide, over the period considered.

⁶ We also carried out an analysis in line with Garas et al. (2016) on our data, and we found that their misspecification error impacted on the overall estimates. We found a negative effect of the network level statistics on the stock of bilateral FDI.

2 Bilateral greenfield FDI and the migration network: data and description

2.1 Data

We reconstruct the bilateral FDI flows for the period 2003–2015 using the *fDIMarket* database (the Financial Times, 2018), to capture individual flows resulting in a new productive activity in the recipient country.⁷ Focusing on individual projects allows us to single out the role of extensive margin, i.e. the number of different transactions, and to analyze how the migrants-induced “information channel” affects the decision of “making an investment” rather than its size. To better identify the relationship between the migrants’ network and bilateral FDI, we exploit the structure of the dataset and focus on the number of bilateral investments project announced (instead of focusing on their value). The value may reflect investment-specific considerations that are independent of the role played by the migrants’ network. Following a similar reasoning, we also exclude investments in the extractive sector as well as those directed toward countries included in the tax havens list of the European Parliament.⁸

We integrate bilateral migrants’ stock data from two distinct sources: the UNDESA Bilateral Migration Database (UNDESA, 2015 revision), which we use to create the International Migrants’ Network (IMN) and the IAB Brain Drain Dataset (Brücker et al., 2013), which provides information on the educational attainment of migrants for 20 OECD destination countries. The presence of highly educated migrants is included to control for the bilateral network effect. Given that information on the skill endowments of the bilateral migrants’ network is not available for countries outside of the IAB sample, we focus on all channels directed toward the IAB countries.⁹ We focus on Migrants’ stocks since flows require collecting information and accumulating social capital at destination, a process that may require time. Furthermore, migrants’ flows are more likely to reflect a country’s generic openness, a possibility that could undermine the identification of the true

⁷ The decision to stop in 2015 is justified by two different motivations. On the one hand, the 5-years span of the migration data would have brought us to 2020, a year which has been ravaged by the outbreak of the Covid-19 pandemic. Second, the 4 years running from 2016 to 2020 have been characterized by a substantial push against international migration and the looming trade war between China and the US (which easily extended to the EU). Controlling for all these factors with a single wave would make the estimates uninformative of the true mechanism at study.

⁸ See <https://www.europarl.europa.eu/cmsdata/147404/7%20-%20001%20EPRS-Briefing-621872-Listing-tax-havens-by-the-EU-FINAL.PDF> for details.

⁹ The list of countries included in the IAB includes (in alphabetical order) Australia, Austria, Canada, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States. Whereas UNDESA collects data for 220 countries and autonomous territories around the world every 5 year. Focusing on the countries included in the IAB dataset alone excludes a good number of large investors, such as Italy, Japan, and China. The latter constitutes a particularly relevant loss. We conduct robustness checks to test the sensitivity of the coefficients to the exclusion of those countries using total bilateral migration as a control. The exclusion of these countries does not lead to significant changes in terms of sign, even though the numerical magnitude of the coefficients is reduced.

mechanisms underlying the FDI-Migrants' network relationship.¹⁰ Notice that, since global migration data are only available on a 5-year span basis, we match FDI data by aggregating all transactions at investing-recipient-year level and taking the 5-years cumulate.

Additional controls (such as distance, GDP, common language etc.) are from the CEPII Gravity database (Conte et al., 2021; Melitz & Toubal, 2014) and the World Development Indicators. We report the main descriptive statistics in Table 5 in "Appendix 2".

2.2 Network description

The International Migration Network (IMN) for 2005, 2010, and 2015 can be represented as an $M^t = N_i * N_j$ adjacency matrix, where each row and column refers to a country, considered respectively as origin and destination. We define each entry of the matrix M^t as $m_{i,j}^t = MigStock_{i,j}^t, \forall i \neq j$, which captures the presence of a migrant population from country i to country j . Following this structure, the matrix M^t can be defined both in binary or weighted terms. In the first case, $m_{i,j}^t$ takes value 1 if at least one person from country i is registered in country j ; in the second case $m_{i,j}^t$ represents the number of citizens with j citizenship residing in country i in the reference year. In either case, $m_{i,j}^t$ is set to null if and only if no citizens of country i resides in country j . In less formal terms, the IMN graph represents the whole set of migratory flows, a web of human movements connecting all countries in the world with each other.

This web can be symmetric (that is, a link is considered as strong as the sum of migrants' population in either direction) or directed—when the direction of the migratory flow is kept into account. Contrary to Fagiolo and Mastroiello (2014), Fagiolo and Santoni (2014), and Garas et al. (2016), we preserve the directed structure of the migrants' network, as the mechanisms linking them to FDI might be different depending on whether we consider the perspective of the investing country or that of the recipient one. This in turn requires distinguishing between incoming and outgoing migrants.¹¹

Table 1 compares the structure of the IMN and the global FDI network (GFDIN, built following the same criteria of the IMN), to understand how and to what extent the two flows (co-) evolved under the network perspective.

In Panel A we can see that the IMN is denser (has a higher number of strictly positive connections) than the GFDIN. Nodes and Edges counts refer respectively to

¹⁰ Also in this respect, we take distance from the analytical framework proposed by Garas et al. (2016).

¹¹ Working with undirected networks generally solves many computational issues, e.g. the possibility to estimate in a single step the bilateral network effect. However, it does not allow to understand the channels through which the network dimension of the migrants' network affects bilateral FDI. Consider the aggregate bilateral flow between two countries (denoted i and n). This implies that the value of the flows (or stocks) in both directions are added ($ji = i \rightarrow j + j \rightarrow i$). Suppose that i receives few migrants from j , despite a large number of investments, and that j invests very little in i despite receiving a large number of migrants. This two countries could not be distinguished from a different pair (c,d) , characterized by average flows (stocks) of FDI and migrants in both directions.

the total number of countries in the two networks and to the total number of strictly positive links. The fact that more countries exhibit positive migration flows and low (or zero) investment flows is consistent with the low number of firms with complex modes of internationalization (see Mayer & Ottaviano, 2008). The comparison of the two networks over time shows very different dynamics in the extensive margin, as represented by the trend in the degree statistics. In our case, the degree refers to the total number of strictly positive links a country receives from (IN) or sends to (OUT) other countries. This simple measure of connectivity in both networks shows how the IMN remain stable over time, while the GFDIN expand substantially between 2005 and 2010 and stabilize after.¹²

Panel B of Table 1 compares the size of the two networks (and their variation) over time. Contrary to past studies (Fagiolo & Mastrorillo, 2014; Sciabolazza, 2018), the IMN does not change substantially over the period considered. This may suggest that capturing sharp changes in the patterns of human mobility requires a longer time span than that covered by our data. On the contrary, the GFDIN experiences a substantial growth, proxied by the decrease in the average number of steps to link all countries through the shortest path (the so called average path length, or APL) and diameter (that is, the largest minimal distance in the network). Despite such a growth, the GFDIN remains sparser than the IMN.

The correlation between the two networks provides further information. Figure 1 plots the simple correlation between the migration and FDI flows in the two networks, i.e., the respective link weights—for every directed country pair $i \rightarrow j$.¹³ The pattern is coherent with the literature on Trade and Migration networks: dyads that are characterized by larger emigration ($i \rightarrow j$) are also characterized by larger inflows of investment ($i \leftarrow j$). Furthermore, the economic and demographic size of the countries in the dyad are also positively correlated with link weights in both networks, consistently with the gravity literature on FDI and migration.¹⁴ A simple graphical comparison suggests that the IMN and the GFDIN might have co-evolved over time.

Focusing on 2010, Figs. 2 and 3 compare the position of each country in the two networks and the type of prevailing connections. We focus on different measures of centrality: degree, strength (or, weighted degree), and average nearest neighbor centrality (ANNC, Barrat et al., 2004).¹⁵ The strength statistic represents the weighted counterpart of the degree statistic: they capture the size of a country's local network but weighting each link with the magnitude of the flow

¹² The slowdown was mostly due to the outbreak of the global financial crisis and its aftermath. Yearly data shows a clearer picture of the evolution of the GFDIN in that period.

¹³ Conversely from the subsequent econometric exercise, we extend the description to all potential country pairs, and not just to the reduced OECD 20-to-ROW pairs.

¹⁴ We base the coloring scheme (from red to light blue) on the product of i and j 's populations, while size reflects the product of the GDP of i and j . Coherently with the gravity framework, larger and bluer dots have a higher probability of being in the north-east side of the plot.

¹⁵ A given node's Average Nearest Neighbor Centrality (ANNC) refers to the average (weighted) degree centrality of the other nodes that directly connects to it. "Appendix 1" reports a detailed description of the ANNC and the other measures of centrality used throughout the paper.

Table 1 Network comparison—general connectivity. *Source:* fDIMarket database (FDI) and UNDESA Population Division (Migrant' stock)

	2005		2010		2015	
	IMN	GFDIN	IMN	GFDIN	IMN	GFDIN
<i>Panel A: Connectivity</i>						
Nodes count	220	228	220	228	220	228
Edges count	10,534	1802	10,685	2399	10,688	2347
Average in-degree	47.88	7.9	48.57	10.52	48.58	10.29
Min. in-degree	0	0	2	0	2	0
Max. in-degree	212	47	204	57	204	64
Average out-degree	47.88	7.9	48.57	10.52	48.58	10.29
Min. out-degree	5	0	4	0	3	0
Max. out-degree	153	107	154	120	154	115
<i>Panel B: Size</i>						
Density	0.22	0.03	0.22	0.05	0.22	0.05
APL	1.67	2.17	1.67	2.08	1.67	2.08
Diameter	4	6	4	5	4	5
Assortativity	-0.29	-0.23	-0.28	-0.19	-0.28	-0.25
Transitivity	0.56	0.46	0.57	0.45	0.57	0.45

Network Topological Comparison. APL=Average Path Length. Average in and out degree for both network are equal, as the number of incoming links the same as the number of outgoing links. Figure 4 shows how the distribution of these links however changes depending on whether we look at the outgoing or incoming flows

represented by such a link. The “ANNC” measures refer to the average centrality of the direct connections to a country, binary and weighted respectively.

Each point in Fig. 2 represents a country as an ordinate couple ($centr_i^{GFDIN}; centr_i^{IMN}$), colored from light to dark and sized from small to large, respectively, according to its GDP and population. Focusing on the left panel of Fig. 2, we see a positive correlation between the size of a country—defined considering both its wealth and its market size proxied by demography (the darkest and among the largest dots represents the US)—and its position in both the migrants and the FDI networks. In line with Fig. 1, larger economies tend to be more central while smaller and poorer countries (in terms of population and GDP) are more peripheral in both networks. The correlation and the average centrality of a country's directly connected neighbors (right hand side plots of Fig. 2) show how wealth and market size are inversely correlated with the centrality of the neighbors. In other words, it is easier (and more remunerative) for less integrated countries to connect with more central ones, and vice-versa.

Figures 3 and 4 highlight the dissortative pattern of the two networks (i.e., the prevalence of small-to-big connections over the total), by plotting the correlation between a country's centrality and the average centrality of its neighbors, and the degree distribution in both networks respectively.

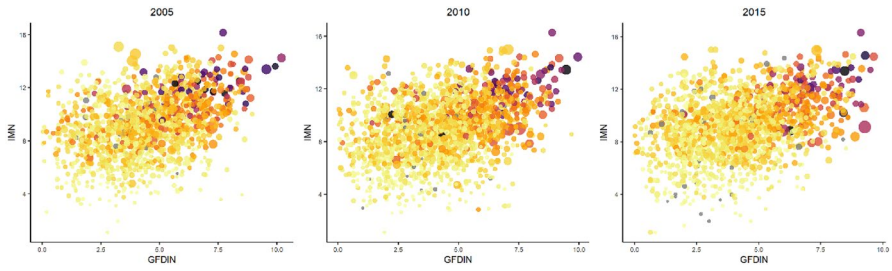


Fig. 1 Correlation between networks: trend over time. *Notes:* Correlation in migrants and FDI networks' link weights. Both y and x axis are in logs. Markers' size is proportional to the product of the population of country i and country j . Color (from light to dark) captures the dyadic product between the two countries' GDP

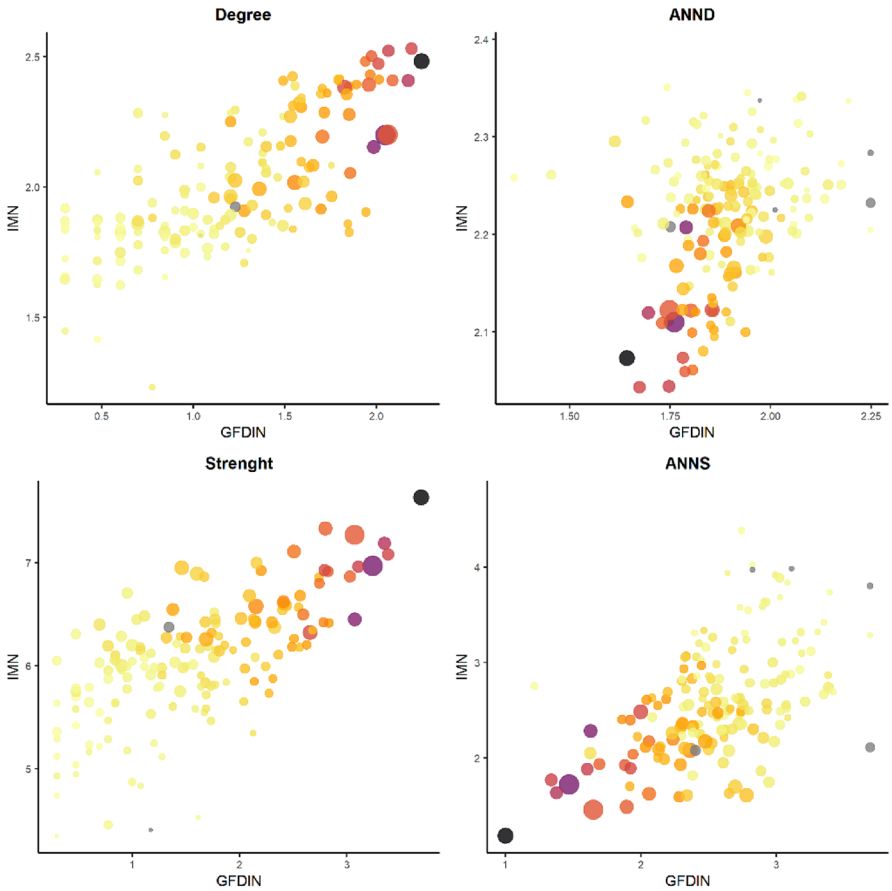


Fig. 2 Country centralities. *Notes:* Correlation between centrality in the migrants and FDI networks. Both y and x axis are in logs. Markets' size is proportional the population of the country, while coloring (from light to dark for higher values) reflects the country's GDP. $ANN(D,S)$ = Average Nearest Neighbor Degree or Strength (Barrat et al., 2004)

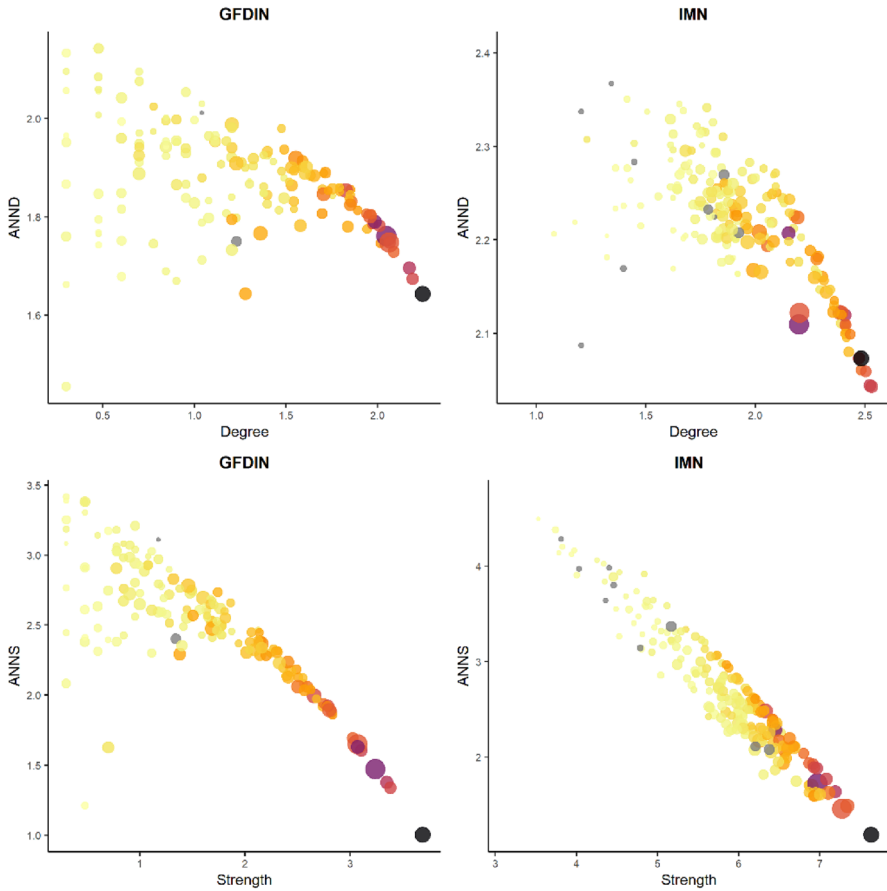
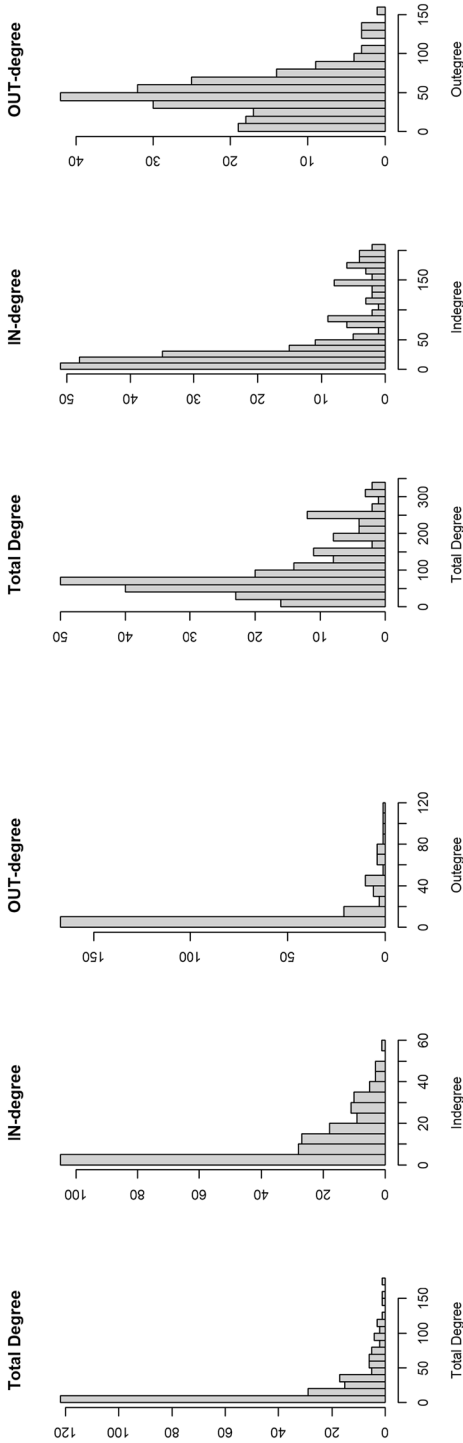


Fig. 3 Dissortative patterns. *Notes:* Dissortative matching between investing and recipient countries (left) and between Immigration and Emigration countries (right). Both y and x axis are in logs. Markers' size is proportional the population of the country, while coloring (from light to dark) reflects the country's GDP. ANNC = Average Nearest Neighbor Centrality (generic for ANND and ANNS)

The comparison of the two networks suggests that the correlation between diaspora and FDI goes beyond the pure bilateral perspective, involving the global structure of the two networks. Figures 1, 2 and 3 also suggest that gravity related dimensions are highly correlated with the link-weight distribution in both the IMN and the GFDIN.



International Migration Network

Greenfield FDI Network

Fig. 4 Degree distributions. *Notes:* Comparison of the degree distributions in the migrants and FDI networks. The reported distribution refer to the year 2010, and is computed on the sample of countries that recorded a non-zero flow in either the Migrants' or the FDI Network

3 Estimation strategy

We now test the relationships between the migrants' network and the bilateral FDI flows, explicitly considering the effect of a country's position within the IMN. We believe that the inclusion of network statistics brings additional complexity to the data, and we consider network-level statistics as belonging to a higher order of information with respect to the dyad level information. The comparison conducted by Herman (2019) suggests that the inclusion of different combinations of fixed effects returns very similar estimates, both in terms of the link formation probability and in terms of trade volume.¹⁶

In this paper, due to the structure of our data and generalizing a gravity approach, we adopt a Multilevel Mixed Effects Regression (MMER) model for (hidden) hierarchical data, where the fixed component of the model controls for the gravity variables as well as for country and time. The additional layers (network and time) are then considered as higher hierarchical levels of information to be modelled specifically to better capture the endogeneity risks.

Multilevel models are used in the international economics and business literature (see for instance, Drzewoszewska, 2014; Giovannetti et al., 2018; Kapás & Czeglédi, 2020), as a generalization of fixed effects exogenous controls, to estimate the multilateral resistance (Anderson & Van Wincoop, 2003). Peterson et al. (2012) points out that the multilateral resistance estimate involves different "levels" of information (hierarchical data) that cannot be fully captured by exogenous fixed effects. MMER represent a useful generalization of gravity models, as they allow both to endogenize the fixed gravity effects when a hierarchical data structure exists, and to disentangles their effect in the error component, allowing the error term correlation to vary across dyads.¹⁷

3.1 The regression model

We model the data hierarchical structure as shown in Eq. 1:

$$y_{ij,t} = \alpha_0 + \sum_{p=1}^P P_{ij,t} + \sum_{net=1}^{nw} net_{ij,t} + u_{ij} + e_n + v_t \quad (1)$$

where $p = (1, \dots, P)$ and $net = (1, \dots, nw)$ collect the variables of interests at a country pair level ($P_{ij,t}$) and the additional controls at a network level ($net_{ij,t}$, which

¹⁶ The author also compares the results of the two types of gravity models (naïve and structural) with a non-structural, simplified gravity model of network formation adding importer and exporter fixed effects and concludes that network formation and growth might depend on complex network patterns that are not accounted for by structural gravity.

¹⁷ Multilevel models control for the heterogeneity of hierarchical data structure, allowing the correlation of the model's stochastic component vary across dyads. Since the between group variability (in this case, the difference across country pairs) is generally larger than the within group variability (that is, within the same group of observations), the adoption of a mixed model allows exploiting and explaining a much larger amount of information leading to a more accurate estimation of data heterogeneity (Bell & Jones, 2015).

also represent our regressors of interest), with u_{ij}, e_n, v_t being the “three levels” of the error term (country pair, network and time). $y_{ij,t}$, the dependent variable, is the cumulative number of FDI over a 5-year span.¹⁸ The inclusion of three stochastic components reflects the assumptions of non-homogeneity and non-constant correlation in the structure of the error term across the different levels considered (as well as across dyads).

In our empirical analysis, Eq. 1 becomes:

$$\ln \text{NumFDI}_{ij,t}^{5y} = \alpha_0 + \ln \text{Mig}_{ji,t-5}^{\text{High}} + \sum_{net=1}^{nw} \gamma net_{i,t-5} + \sum_{net=1}^{nw} \gamma net_{j,t-5} + \sum_{p=1}^P \beta P_{ij,t} \quad (2)$$

where the dependent variable $\ln \text{NumFDI}_{ij,t}^{5y}$ is the (log of the) number of greenfield FDI projects from country i to country n , cumulated over the precedent 5 years period. The set of coefficients of the terms reported in (2) under the term γ represent our coefficients of interest, respectively for the emigration and immigration countries.¹⁹

According to the standard gravity literature, log-linearization would imply the exclusion of all null FDI flow, introducing a bias in our estimates. However, as we confine our analysis to the FDI flows from 20 OECD countries to the rest of the world, we focus on the denser part of the network. This leads to a negligible incidence of null FDI flows, compared to the whole FDI network, limiting the bias that would be introduced by the log transformation. The variable $\ln \text{Mig}_{ji,t-5}^{\text{High}}$ represents the 5-years lagged bilateral stock of tertiary educated migrants from country j to country i . $\sum \gamma net_{i,t-5}$ and $\sum \gamma net_{j,t-5}$ are two vectors of both direct and indirect measures of network centrality (both local and global) as well as measures of third-party network effects’ statistics for country i and country j respectively. In (2) all the network statistics are included with a 5-years lag.²⁰

The decision to introduce all migrants’ related variables (our variables of interest) in lags depends on two main considerations. The first hinges on the necessity of mitigating reverse causality with respect to current FDI flows (Sheperd, 2021). Given the long time that is often required for greenfield FDI to realize, considering lagged migration might not be enough to reduce reverse causality. The time required for a new plant or activity to finalize could create some anticipation in perspective migrants, that might in turn decide to move to a country where they know an investment has been announced but not yet realized. However, FDI Market reports

¹⁸ The indices in (1) refer to the investing countries (i), the recipient country (j), and time (t). The index n refers to the error term related the network effect.

¹⁹ Since the three error terms are not included as variables in the estimation, they are excluded from Eq. 2.

²⁰ We include the previous data point available in the UNDESA migration dataset, which reports information with a five-year time interval between each other.

the date an investment is announced (not realized), so that the anticipation argument is less likely to occur.²¹ The second consideration rests on the potential mechanism driving the relationship between migrants' networks and FDI. Migrants' networks need time to structure and establish the information channel mentioned above, so we consider lagged migration in our models.²²

The term $\sum_{p=1}^P \beta P_{ij,t}$ finally includes a vector of geographical, cultural, and economic bilateral controls such as (a) the geographical distance between the investing (destination) and the recipient (emigration) countries (computed as the geodetic distance between two countries centroid); and (b) the per capita GDP of both the investing and the recipient countries. We also include a set of dummy variables capturing respectively (c) border contiguity; (d) whether the two countries were part of the same colonial empire; (e) if the legal system of the two countries can be associated to the same set of principles; (f) whether a country pair is part of the same free trade agreement.

Tables 2 and 3 focus on the direct relationship between FDI and a country's direct local centrality in the IMN (i.e. how much a country is directly integrated in the global migration network), highlighting degree, strength, and Average Nearest Neighbor Centrality. Table 4 distinguishes common and exclusive connections between country pairs to break down IMN's network centrality. Common connections in the IMN may influence the direct network effect by providing parallel channels for information flows among countries, mediated by national ties, and their effect might be either mitigated or amplified by the (relative) size of non-mutual ties. These mitigating effects can be interpreted as a form of indirect or "third-party" effect. To better understand the migration and FDI connections between countries, Fig. 5 provides a graphical example. Let us consider a simplified network with 4 countries, with different characteristics and connections. Assume Country 1 is mostly an emigration country, with citizens migrating to Country 2 and Country 3. Country 2 is an investing and immigration country, with more limited emigration flows toward country 3 and 4. Country 3 receives migrants from the others, while country 4 only receives migrants from country 2 without sending anyone abroad, while receiving FDI from Country 2. Whether country 3 registers incoming or outgoing FDI is irrelevant to this example. As shown in the graph, Country 1 and 2 share Country 3 as common connection on the IMN. Thus, migrants from country 1 to 3 and from 3 to 2 might facilitate the flow of information between the Country 1 and 2 in the first place. Country 4 is an exclusive connection of Country 2. Thus, it might divert investments from Country 2 to Country 1, provided the flow of information vehiculated by national migrants' networks is strong enough.

Building on Fagiolo and Mastroiello (2014), we create four additional measures of bilateral network integration between any two countries in our dataset:

²¹ This argument is not counteracted by the way we define the dependent variable as the cumulated number of projects over the period $t-4$ to t , given that it is unlikely for migrants to be able to anticipate the announcement of a greenfield FDI happening in the subsequent year.

²² Notice that this argument is reinforced by using the stock of bilateral migrants, which also takes time to build and already mitigates concerns of reverse causality. Notice however that this strategy barely mitigates reverse causality by removing the temporal element that might lead to the co-determination of both migrants network and FDI flows. Additionally, it does not mitigate (let alone solve) other potential endogeneity concerns.

Table 2 Binary and weighted local connectivity

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Multilevel	Multilevel	Multilevel	Multilevel	Multilevel	Multilevel	Multilevel
	$\ln NumFDI_{ijt}^{2y}$	$\ln NumFDI_{ijt}^{5y}$	$\ln NumFDI_{ijt}^{5y}$	$\ln NumFDI_{ijt}^{5y}$	$\ln NumFDI_{ijt}^{5y}$	$\ln NumFDI_{ijt}^{5y}$	$\ln NumFDI_{ijt}^{5y}$
$\ln Mig_{ijt-5}^{High}$	0.263*** (16.97)	0.264*** (15.65)	0.207*** (38.08)	0.246*** (14.95)	0.209*** (32.04)	0.204*** (31.54)	0.121*** (105.72)
$\ln Degree_{ijt-5}^{IN}$		-0.034 (-0.51)				0.082+ (1.87)	
$\ln Degree_{ijt-5}^{OUT}$			0.612*** (5.42)			0.621*** (5.72)	
$\ln Strength_{ijt-5}^{IN}$				0.059*** (19.75)			0.209*** (11.57)
$\ln Strength_{ijt-5}^{OUT}$					0.156*** (5.96)		0.250*** (7.55)
LRest	1545.63*** 9225	1479.55*** 9225	1305.26*** 9225	1386.03*** 9225	1363.26*** 9225	1233.10*** 9225	1219.51*** 9225
Residual DF	9216	9216	9216	9216	9216	9215	9215

t-statistics in parentheses. Robust Standard Errors. Estimates refer to a 3-level regression with random intercepts. Dyads are nested in networks, all nested in time. The table reports the estimates from the Eq. 2 where the dependent variable $\ln NumFDI_{ijt}^{5y}$ represents the cumulative number of FDI projects announced over the past 5 years, and explores the contribution of local network connectivity to bilateral FDI flows. Table 15 in “Appendix 5” compares local network coefficients estimates obtained via Mixed Multilevel Regression with a two-steps FE estimator. Even if their coefficients are not reported, all equations control for distance (in logs); per capita GDP (in logs); three dummies for contiguity, colonial history, and common legal system; FTA; and a constant. Table 7 in “Appendix 3” shows the coefficients of the additional controls. LRest refers to the log-likelihood Ratio test of the multilevel specification against the linear model (Rabe-Hesketh & Skrondal, 2008). The rejection of the null hypothesis implies a systematic difference between MIXED and OLS/FE estimates. Table 8 in “Appendix 3” reports each levels’ group numerosity. Table 9 reports the degree of freedom computed with the ANOVA approach

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 Average nearest neighbor centrality

	(1)	(2)	(3)	(4)
Model	Multilevel	Multilevel	Multilevel	Multilevel
Dependent variable	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$
$\ln \text{Mig}_{ji,t-5}^{\text{High}}$	0.263*** (16.79)	0.259*** (17.22)	0.214*** (26.53)	0.234*** (20.15)
$\ln \text{ANNCentrality}_{i,t-5}^{\text{IN-IN}}$	0.115+ (1.94)			
$\ln \text{ANNCentrality}_{i,t-5}^{\text{IN-OUT}}$		0.177*** (4.63)		
$\ln \text{ANNCentrality}_{j,t-5}^{\text{OUT-IN}}$			-1.782*** (-5.75)	
$\ln \text{ANNCentrality}_{j,t-5}^{\text{OUT-OUT}}$				-2.604*** (-6.53)
LRtest	1524.86***	1540.36***	1218.00***	1382.94***
Obs	9225	9225	9225	9225
Residual DF	9216	9216	9216	9216

t-statistics in parentheses. Robust Standard Errors. Estimates refer to a 3-level regression with random intercepts. Dyads are nested in networks, all nested in time. The table reports the estimates from the Eq. 2 where the dependent variable $\ln \text{NumFDI}_{ij,t}^{5y}$ represents the cumulative number of FDI projects announced over the past 5 years, and explores the contribution of local first-degree connectivity to bilateral FDI flows. The coefficient of the LRtest refers to the log-likelihood Ratio test of the multilevel specification against the linear model. Table 8 in “Appendix 3” reports each levels’ group numerosity. Table 10 reports the degree of freedom computed with the ANOVA approach

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

one measuring the extent of the common (or overlapping) connections between the two, a second measuring the extent of the exclusive (or complementary) network of either one of the two with respect to the other. We consider a third country to be part of the complementary network of a country pair ij if it welcomes migrants from country j but none of its nationals resides in country i or if its people migrated to country i but none of its resident originates from country j .

In this respect, our approach differs from previous work (e.g. Fagiolo & Masrorillo, 2014; Garas et al., 2016), in which a country is considered to be part of the common or of the complementary network of a dyad irrespective of the direction of the migration flows between the three countries. The relevant empirical issue here is to determine which one dominates within a country pair—whether it is the diversion effect which we could expect being at play when the exclusive non-overlapping migrants’ network is large, or the incremental effect a larger common network could generate. This is likely to depend on (a) the total number

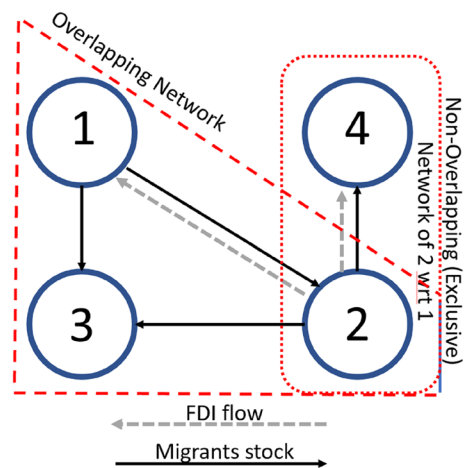
Table 4 Overlapping and complementary networks

	(1)	(2)	(3)	(4)
Model	Multilevel	Multilevel	Multilevel	Multilevel
Dependent variable	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$
$\ln Mig_{ji,t-5}^{High}$	0.243*** (14.27)	0.244*** (14.21)	0.219*** (29.93)	0.208*** (37)
$\ln Overlapping_{8_{i,t-5}}^{IN}$	0.047*** (8.06)	0.051*** (8.38)		
$\ln Complementary_{i,t-5}^{IN}$		-0.010** (-3.19)		
$\ln Overlapping_{j,t-5}^{OUT}$			0.128*** (5.2)	0.082*** (4.14)
$\ln Complementary_{j,t-5}^{OUT}$				0.079*** (9.19)
LRtest	1584.26***	1516.20***	1425.06***	1354.23***
Obs	9225	9225	9225	9225
Residual DF	9197	9196	9216	9179

t-statistics in parentheses. Robust Standard Errors. Estimates refer to a 3-level regression with random intercepts. Dyads are nested in networks, all nested in time. The table reports the estimates from the Eq. 2 where the dependent variable $\ln NumFDI_{ij,t}^{5y}$ represents the cumulative number of FDI projects announced over the past 5 years, and explores the relative contribution of common and exclusive network connectivity to bilateral FDI flows. The coefficient of the LRtest refers to the log-likelihood Ratio test of the multilevel specification against the linear model. Table 8 in “Appendix 3” reports each levels’ group numerosity. Table 11 reports the degree of freedom computed with the ANOVA methodology

+*p*<0.1, **p*<0:05, ***p*<0:01, ****p*<0:001

Fig. 5 Overlapping versus exclusive network



of connections between the two countries and on (b) the extent of the overlapping subnetworks in contrast to the non-overlapping subnetworks.²³

4 Discussion of the main results

We now analyze and discuss the results under three aspects of network integration in the IMN. The estimates reported, refer to 20 OECD countries with detailed migration data by educational level.²⁴ In line with the literature, all equations include a set of standard gravity controls.²⁵ To test for the appropriateness of the multilevel model, for each specification we run a Likelihood Ratio (LR) Test showing the existence of a hierarchical structure in the data. The test statistics strongly reject the null hypothesis of absence of a second/third level in the data, implying a strong bias in the OLS/FE estimates of the baseline model. This suggests that, in our case, the multilevel model is appropriate since a hidden hierarchical structure in the data is present.²⁶

Our base model, with bilateral migration as the only parameter of interest, is reported in Column 1. The estimated coefficient for bilateral migration $\ln Mig_{ij,t-5}^{High}$ is in line with the literature: the positive sign and the significance both confirm the importance of high skilled migration on bilateral investment, reinforcing the idea that migrants and investments might be complementary rather than substitutes. The remaining results extend the benchmark equation, including measures of network centrality. Columns from 2 to 7 in Table 2 explore the effect of different measures of directed local centrality (degree and strength) in the IMN. On average, FDI benefits from a more central position in the IMN of both the recipient and the investing countries. Overall, being more connected implies that more investments are likely to flow from country i to country j ($i \rightarrow j$).²⁷ Quantitatively, our results suggest that the FDI-recipient outward centrality in the IMN outplays the investing side's inward centrality. This result is robust across all specifications and suggests that the information channel is influenced by third party effects.

²³ Differently from the existing literature, we preserve the directed structure of the network also in constructing the two overlapping and the exclusive subnetworks.

²⁴ For the sake of space, we do not report the coefficients for the non-migrant-related controls in the main text. Such controls include geographical distance, colonial relationship, contiguity, the presence of similar legal system; and measures of country specific economic, such as per capita GDP. The estimates of the controls are consistent with the existing literature on the determinants of bilateral FDI, in terms of both sign and magnitude, and across all specifications. For reference, Table 7 in "Appendix 3" reports the full specification of the model discussed in Table 2.

²⁵ As a robustness check, we compare the results from the multilevel model with the estimates we obtain from a gravity model estimated with high dimensional fixed effects. The estimates are reported in Table 15 in "Appendix 5": all the standard gravity controls estimated effect are coherent with the literature (sign and magnitude).

²⁶ We also compare the estimates from the Multilevel Model to both FE (two step gravity and dyad FE) and Bonus Vetus OLS. The estimates are consistent across different models. Results are available in the Online Appendix.

²⁷ The only exception in this sense is represented by the inward binary degree centrality of the investing country, which remains not statistically significant when included as the only additional regressor in the equation.

The measures of centrality in Table 2, however, do not consider the quality of each country's position in the network, and do not allow to explore the sources of third party's effects described in Fig. 5. For instance, it may happen that a country's local centrality on the IMN captures the economy's openness, with little correlation with the information channel triggered by bilateral migration. We believe that a better way to explore the presence and impact of third-party's effects consists in splitting local centrality between overlapping and complementary bilateral networks. A positive and significant coefficient on the overlapping network can be interpreted as a bridge operating through countries (and the migrants' community within them), mutually connecting i and j in the IMN.²⁸

Table 3 examines the relative importance of the first-order connections (i.e., countries that are directly connected to either i or j —or to both) rather than the direct centrality of each country pair. The first two columns show that the inward and outward centrality of the investing country i 's first-order connections ($\ln ANN Centrality_{i,t-5}^{IN-IN}$ and $\ln ANN Centrality_{i,t-5}^{IN-OUT}$) have a positive impact on bilateral FDI toward j . This suggests that investing countries with more central connections on the IMN also have a higher probability of investing in country j . Country j can also convey information through its indirect connections with i . However, the relative importance of country j 's outward connectivity works against the positive effect of the direct bilateral channel. When considering the effect of direct bilateral migration, the more important the neighbors of country j , the stronger the ties. The results reported in Table 3 are in line with the existing literature on trade and migrants' networks.

Table 4 compares the indirect network effect by disentangling the relative importance of the two countries common and exclusive connections. We consider investing (immigration) country i 's inward connectivity against a recipient (emigration) country j 's outward connectivity. An interesting pattern emerges. The coefficient for both the investing country's inward and recipient countries outward overlapping networks (respectively, $\ln Overlapping_{i,t-5}^{IN}$ and $\ln Overlapping_{j,t-5}^{OUT}$) show that a larger number of shared connections between i and j translates into a higher number of bilateral FDI going from $i \rightarrow j$.²⁹ Focusing on the size of the non-shared ties, we notice that, while the size and the volume of the complementary network for country j has a positive impact on bilateral FDI from $i \rightarrow j$, a more diversified immigrant community in the investing country i (with respect to j 's network, $\ln Complementary_{i,t-5}^{IN}$) significantly reduces bilateral FDI between the two countries, even if the numerical magnitude is low, on average. This suggests that, if shared and exclusive connections are considered separately, the positive effect of bilateral migration is mitigated, potentially due to the higher noise caused by the larger set of distinct streams of information generated by the different national communities. The bias introduced by the

²⁸ In network terms, belonging to different relational structures is defined in terms of communities. See Danchev and Porter (2018) for an extensive analysis of the evolution of the community structure within the IMN.

²⁹ This refers to the number and the volume of countries to which the recipient j sends migrants to, and that in turn send migrants to i .

information from exclusive migratory channels, therefore, reduces on average the bilateral migration’ positive effect. This effect is coherent with the idea that the capacity and the resources required for international investments are lower, on average, than those required to foster bilateral trade.

4.1 Network scale effect

Our results show that being integrated in the IMN positively affects both the ability of a country to attract FDIs and to promote domestic investors abroad. Being more central in the IMN leads to larger in-/out-flows of FDI and makes information flows more efficient in promoting bilateral investments. This may result from the availability of information that pushes investors towards countries characterized by a favorable investment environment (Fiorini et al., 2021; Lücke & Stöhr, 2018). Notice that in our paper, we assumed that being a central hub in the IMN has the same relevance as a driver of bilateral FDI at any level of total migration. Relatively low/high migration levels can impact the effectiveness of networks ties, everything else equal, as the size of a nationals’ community in a potential investing economy (and vice versa) is likely to determine how information reaches a potential investor. In this case, being central in the IMN helps reaching the critical mass of information that is necessary to trigger investments, even at low levels of bilateral migration. To test this hypothesis, we modify Eq. (2) as

$$\begin{aligned} \ln NumFDI_{ij,t}^{5y} = & \alpha_0 + \ln Mig_{ji,t-5}^{High} + \sum_{net=1}^{nw} \gamma_{net} X_{i,t-5}^{net} + \sum_{net=1}^{nw} \gamma_{net} X_{j,t-5}^{net} + \ln Mig_{ji,t-5}^{High} \times HighMig_{i,t-5} \\ & + \sum_{net=1}^{nw} (\gamma_{net} X_{i,t-5}^{net} \times HighMig_{i,t-5}) + \sum_{net=1}^{nw} (\gamma_{net} X_{j,t-5}^{net} \times HighMig_{i,t-5}) + \sum_{r=1}^R \beta_r X_{jt}^r \end{aligned} \tag{3}$$

where $HighMig_{i,t}$ is a dummy variable signaling if the migrant’ community from county j in country i exceeds the average size of all other immigrant community in country i. Using this approach, we divide the sample based on the relative importance of each potential investing country’ immigrant community. Coefficients plotted in Fig. 6 summarize the main results, while Table 12, 13 and 14 in “Appendix 4” report the coefficients for each specification.

Figure 6 shows that both the bilateral stock of highly skilled migrants’ and the different measures of network centrality are significant. The bilateral migrants’ estimate is larger for pairs in the above-average stock sample, while those related to network centrality are lower. These results are consistent with the existence of a critical mass in the information flows between two countries. A high degree of attractiveness may not be enough to create a national community, since the community needs to be big enough to provide useful information to local investors. Thanks to its indirect effects, network centrality (of both the potential investing country and the recipient) becomes extremely relevant for FDI when the bilateral stock of migrants from a given emigration country j living in destination i is relatively small, compared to other migrants’ communities in the same country.

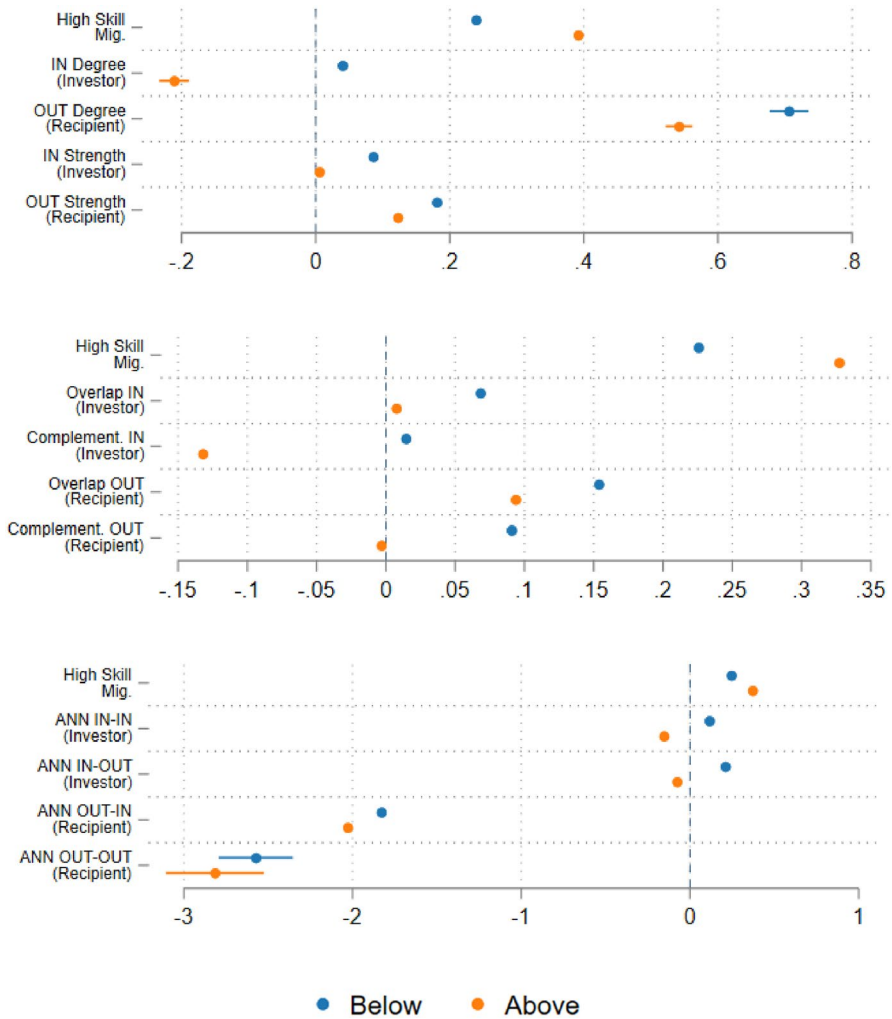


Fig. 6 Heterogeneous migrants’ network effects. *Notes:* Estimated bilateral and migrants’ network coefficients from Eq. 3. Here we assess to what extent the network effect identified above are driven by the relative size of the total bilateral migrant population from country j to country i . The x-axis refers to the size of the coefficient for each sub-group. The labels “Below” and “Above” refer to the subsamples of country pairs (i, j) for whom the stock of migrants from country j to country i ranks below (above) the average size of the immigrants population in country i in a given year

5 Conclusions

In this paper, we generalize the existing evidence on migration and cross border investments’ relationships. Using social network analysis, we offer a different perspective on the possible impact of migrants’ networks on bilateral FDI. We show that the countries’ integration in the International Migration Network (IMN) may trigger channels that, in turn, can affect Greenfield FDI.

We introduce a series of network statistics in a Multilevel Mixed Effects Model to capture the hierarchical structure of our data. Controlling for a range of gravity-related factors, our results are coherent with the existing literature and show how high-skilled migration and FDI are complementary. Furthermore, we show that the degree of integration of both emigration and immigration countries in the IMN reinforces this complementarity. In detail, we show that (i) it is necessary to distinguish the emigration from the immigration perspective in order to understand how the position of a country in the migrants' network affects bilateral (FDI) flows; (ii) differentiating between common and exclusive connections on the International Migrants Network results in heterogeneous network effects; and (iii) the composition of the immigrants community in the investing country shapes bilateral FDI flows, independently of the bilateral migratory history between two countries. We highlight that the immigration network of the investing economy has a differentiated impact on bilateral FDI flows, as the size of the exclusive network negatively affects bilateral investments. The migrants' information channel at bilateral level becomes noisy, as the network becomes more heterogeneous. This is consistent with the idea of considering FDI as rival goods, with potential destination countries competing for them.

In summary, our results highlight the complex yet largely positive relationship between international migration and FDI. From the potential recipient economy perspective, leveraging on the presence of fellow national communities abroad and on foreign national communities at home, constitutes an additional tool for pursuing strategic investments (from) abroad.

Appendix 1: Definitions of network centralities

(Weighted) Degree centrality

To intuitively understand how much integrated an actor is in the network, measures of local connectivity represent the most intuitive and straightforward indicators to deal with. as they do not take into consideration anything but the actor itself and the number of connections centered on her. Degree centrality In its binary formulation merely counts the number of connections of a node

$$k_{ij} = \sum_{j=1}^n a_{ij} \quad \forall i \in G$$

where a_{ij} represents the ij th entry of A , the binary adjacency matrix. Degree centrality is associated to the extensive margin of network, i.e. to the number of (new) connections that an actor establishes. The intensive margin of a network can be capture by the Weighted degree centrality (or, Strength Centrality). Similarly to degree centrality, sums all entries of the weighted adjacency matrix W , while considering the size of the link rather than just its existence.

$$s_{ij} = \sum_{j=1}^n v_{ij} \quad \forall i \in G$$

Both Degree and Strength centrality can be normalized in order to obtain a measure ranging between 0 and 1. Interestingly, despite both measures are based on a very similar type of information, they may not perfectly coincide in complex networks (even if their correlation is usually very large).³⁰

Average nearest neighbor centrality (Barrat et al., 2004)

Measures of degree centrality taken alone offer a very limited perspective about the context an actor operates in. For instance, while they can provide information on the size/intensity of an actor own network, they do not take into consideration any characteristic of the partners it is connected with. Conversely, ANN measures are based on the average size and intensity of an actor's neighbors connectivity, and represent a form of indirect measure of centrality. Similarly to its actor centered counterparts, ANN Degree (ANND) averages the number of connections centered on the partners of an actors (analogously, ANN Strength measures the average intensity of those connections). ANN centrality (Barrat et al., 2004) is defined as

$$ANNCentrality = \frac{1}{s_i} \sum_{j=1}^n a_{ij} w_{ij} k_j$$

where s_i represents the strength of node i and constitutes the normalizing factor (in the binary version, s_i is replaced by k_i). a_{ij} represents the ij th entry of the adjacency matrix A . w_{ij} is the weight of the link (that is conveniently set to 1 in the binary case); k_j indicates node j 's degree centrality. Comparing Degree (strength) and ANNC measures is particularly interesting under a topological perspective (see Fig. 3), as it allows to identify a network assortativity mixing patterns. In the econometric exercise, we consider the direction of each connection in the IMN to build different definitions of ANNC, depending on whether we are considering the inward or outward centrality of each neighbor.

Overlapping and complementary networks

In our bilateral framework, the way the IMN affects economic exchanges might also depend on the type of connections they share (or, alternatively, they do not share). The overlapping (non-overlapping, or complementary) network is defined as size of the common neighborhood shared (not shared) by two actors

³⁰ For instance, actors with a lot of small connections may coexist with nodes with a small number of extremely large links. Measures of purely local connectivity fail to capture this possibility, since they tend to ignore the structural characteristics of the neighbors, or of the network itself.

$$\text{Overlap}_{ij} = I = (1, \dots, n), J = (1, \dots, m) \quad I \cap J = 0$$

$$\text{Complementary}_{ij} = I = (1, \dots, n), J = (1, \dots, m) \quad I \cap J \neq 0$$

where $I = (1, \dots, n)$ and $J = (1, \dots, m)$ represent the set of channels flowing in/to country i and j respectively.³¹ According to the literature, international migrants favors economic exchanges by widening and forming a network through which information can flow. Migrants networks might be therefore more effective in channeling information between countries in the same community rather than between countries which are not much integrated with each other. The inclusion of both measures allow to control for this occurrence.

Appendix 2: Summary statistics

See Table 5 and 6.

Appendix 3: Additional information to the main tables

See Table 7, 8, 9, 10 and 11.

Appendix 4: Additional results: network scale effect

See Table 12, 13 and 14.

Appendix 5: A multilevel mixed effect versus gravity approach

Economic literature analyzes bilateral data using gravity equations, generally estimated using fixed effects (FE) (see for instance Anderson & Van Wincoop, 2003; Bertoli & Huertas-Moraga, 2013). In this paper, we depart from the usual gravity literature in terms of estimation approach, due to the hidden hierarchical structure of our dataset.

In our framework, network variables represent second-level data, difficult to be captured in gravity analysis. We decided to endogenize this hierarchy using a multilevel approach, rather than ruling it out by fixed effects estimation that might lead to specification errors (Baldwin & Taglioni, 2006; Eaton & Kortum, 2001; Head & Ries, 2008).³² We adopt a Multilevel Mixed Effects Model (Rabe-Hesketh &

³¹ The reverse holds to compute the overlapping and complementary outward network.

³² FE gravity estimation presents some weaknesses that might highly affect the analysis. FE imply a strong analytical assumption concerning the structure of the error term and the degree of interdependence between the observations: by considering the correlation in the error term to be constant across observations (country pairs), FE models often ignore the specificity of each bilateral relationship. This

Table 5 Summary statistics: network characteristics

	Obs	Mean	SD	Min	Max
$\ln \text{NumFDI}_{ij,t-5,t}$	9225	0.999	1.396	0	7.659
$\ln \text{Mig}_{ji,t-5}^{\text{High}}$	9225	5.944	2.639	0	14.090
<i>Measures of Local Connectivity (centrality)</i>					
$\ln \text{Degree}_{i,t-5}^{\text{IN}}$	9225	5.151	0.226	4.111	5.357
$\ln \text{Degree}_{j,t-5}^{\text{OUT}}$	9145	3.954	0.447	1.792	5.037
$\ln \text{Strength}_{i,t-5}^{\text{IN}}$	9225	14.428	1.352	11.921	17.590
$\ln \text{Strength}_{j,t-5}^{\text{OUT}}$	9145	13.024	1.591	7.006	16.561
<i>(NON)-overlapping network</i>					
$\ln \text{Overlapping}_{i,t-5}^{\text{IN}}$	9225	3.406	1.113	0	5.313
$\ln \text{Complementary}_{i,t-5}^{\text{IN}}$	9225	3.699	0.522	1.386	5.142
$\ln \text{Overlapping}_{j,t-5}^{\text{OUT}}$	9225	3.702	0.435	1.099	4.875
$\ln \text{Complementary}_{j,t-5}^{\text{OUT}}$	9225	4.811	0.259	3.89	5.236
<i>Barrat et al.'s ANNCentrality measures</i>					
$\ln \text{ANNCentrality}_{i,t-5}^{\text{IN-IN}}$	9225	3.862	0.222	3.586	4.849
$\ln \text{ANNCentrality}_{i,t-5}^{\text{IN-OUT}}$	9225	3.406	0.277	3.021	4.327
$\ln \text{ANNCentrality}_{j,t-5}^{\text{OUT-IN}}$	9225	4.762	0.176	4.157	5.204
$\ln \text{ANNCentrality}_{j,t-5}^{\text{OUT-OUT}}$	9225	4.215	0.088	3.929	4.505

Summary Statistics based on the estimation sample, computed on the observations for which data on highly educated migration is available. The estimated sample includes all dyads involving the IAB-20 group to the rest of the world (excluding dyads that have been singled out by the statistical routine)

Table 6 Summary statistics: covariates

	Obs	Mean	SD	min	MAX
$\ln \text{Distance}_{ij,t}$	9225	8.581	0.838	4.708	9.880
Contiguity _{ij}	9225	0.019	0.137	0	1
$\ln \text{GDPcap}_{i,t}$	9225	10.593	0.478	8.938	11.541
$\ln \text{GDPcap}_{j,t}$	9225	8.436	1.568	4.968	11.541
Colony _{ij}	9225	0.051	0.219	0	1
ComLegal _{ij}	9225	0.304	0.460	0	1
FTA _{ij}	9225	0.297	0.457	0	1

Summary Statistics based on the estimation sample, computed on the observations for which data on highly educated migration is available. The estimated sample includes all dyads involving the IAB-20 group to the rest of the world (excluding dyads that have been singled out by the statistical routine)

Footnote 32 (continued)

might be particularly relevant in the case of historical, geographical, as well as relational features (Egger, 2000). Back to our case, the homogeneity across units assumption is poorly realistic due to the hidden hierarchical structure of the data.

Table 7 Binary and weighted local connectivity—all control variables

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel	In $NumFDI_{ijt}^{5y}$ Multilevel
In $Mig_{ijt}^{High,5y}$	0.263*** (16.97)	0.264*** (15.65)	0.207*** (38.08)	0.246*** (14.95)	0.209*** (32.04)	0.204*** (31.54)	0.121*** (105.72)
In $Degree_{ijt-5}^{IN}$		-0.034 (-0.51)				0.082+ (1.87)	
In $Degree_{ijt-5}^{OUT}$			0.612*** (5.42)			0.621*** (5.72)	
In $Strength_{ijt-5}^{IN}$				0.059*** (19.75)			0.209*** (11.57)
In $Strength_{ijt-5}^{OUT}$					0.156*** (5.96)		0.250*** (7.55)
In $Distance_{ij}$	-0.121*** (-4.39)	-0.122*** (-4.62)	-1.08*** (-3.95)	-0.140*** (-5.03)	-0.106*** (-3.88)	-0.104*** (-4.00)	-0.162*** (-5.61)
In $GDPcap_i$	0.350*** (5.25)	0.346*** (5.79)	0.364*** (4.99)	0.342*** (5.06)	0.359*** (5.31)	0.373*** (5.45)	0.341*** (5.02)
In $GDPcap_j$	0.197*** (5.56)	0.196*** (5.67)	0.187*** (6.00)	0.201*** (5.67)	0.244*** (5.81)	0.188*** (6.11)	0.285*** (6.32)
Contiguity _{ij}	0.846*** (18.65)	0.843*** (16.04)	1.033*** (24.67)	0.843*** (17.85)	1.060*** (26.70)	1.042*** (22.14)	1.108*** (23.82)
Colony _{ij}	0.159+ (1.71)	0.163+ (1.66)	0.375** (2.79)	0.141 (1.52)	0.364** (2.81)	0.369** (2.70)	0.418** (3.20)
LegalSystem _{ij}	-0.0585*** (-3.81)	-0.0618*** (-6.04)	-0.0103 (-0.45)	-0.0538*** (-3.65)	-0.0148 (-0.72)	-0.00177 (-0.09)	0.0261 (1.08)
FTA _{ij}	0.0010	0.0005	-0.060	0.024	-0.060	-0.060	-0.038

Table 7 (continued)

Model	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$	Multilevel	In $NumFDI_{ijt}^{5y}$
Dependent variable	In $NumFDI_{ijt}^{5y}$	(0.03)	In $NumFDI_{ijt}^{5y}$	(0.03)	In $NumFDI_{ijt}^{5y}$	(0.03)	In $NumFDI_{ijt}^{5y}$	(0.04)	In $NumFDI_{ijt}^{5y}$	(0.04)	In $NumFDI_{ijt}^{5y}$	(0.03)	In $NumFDI_{ijt}^{5y}$	(0.05)
LRtest	1545.63***		1479.55***		1305.26***		1386.03***		1363.26***		1233.10***		1219.51***	
Obs	9225		9225		9225		9225		9225		9225		9225	
Residual DF	9216		9216		9216		9216		9216		9215		9215	

t-statistics in parentheses. Robust Standard Errors. The table reports the coefficients from Eq. 2 reported in Table 2, also displaying the coefficients of the additional sets of controls included in all regression

† $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8 Group level information

	# of Groups	Observations per group		
		Min	Avg	Max
Time	3	3026	3075	3112
ID	8660	1	1.1	2

Group information from the Multilevel mixed regression. The table reports group numerosity across all models discussed in Table 2, 3 and 4 and in Fig. 6 (related tables in “Appendix 3”)

Table 9 Degree of freedom (from ANOVA)—cfr Table 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\ln Mig_{ji,t}^{High,5y}$	562	561	561	561	561	560	560
$\ln Distance_{ij}$	562	561	561	561	561	560	560
$\ln GDPcap_i$	562	561	561	561	561	560	560
$\ln GDPcap_j$	562	561	561	561	561	560	560
Contiguity _{ij}	562	561	561	561	561	560	560
Colony _{ij}	562	561	561	561	561	560	560
LegalSystem _{ij}	562	561	561	561	561	560	560
FTA _{ij}	562	561	561	561	561	560	560
$\ln Degree_{i,t-5}^{IN}$		561				560	
$\ln Degree_{j,t-5}^{OUT}$			561			560	
$\ln Strength_{i,t-5}^{IN}$				561			560
$\ln Strength_{j,t-5}^{OUT}$					561		560
Constant	2	2	2	2	2	2	2

Degree of freedom for each coefficient included in all the specifications reported in Table 2

DF have been computed using ANOVA

Skrondal, 2008) allowing us to run the fixed effect model with one or more random components, representing the levels of the data hierarchy. In doing this, this approach relaxes the strict homogeneity assumption on the error term (by specifying the structure of the dataset instead of controlling out the heterogeneity across the observed dyads) limiting the risk of misspecification or overfitting, both frequent in FE estimation, while providing an efficient estimator (Bell & Jones, 2015).

Table 15 below provides a consistent check for the robustness of our approach. The table replicates the results from Table 2, comparing the results of a second stage FE estimation (Panel A) against the baseline mixed estimates (Panel B). Since the second stage regresses the *country* × *year* FE on country specific features, it is not possible to estimate a second stage including partner’s characteristics. For this reason, Table 15 does not report the replication of columns (6) and (7), where both in-degree (strength) of country *i* and out-degree (strength) of country *n* were included.

Table 10 Degree of freedom (from ANOVA)—cfr Table 3

	(1)	(2)	(3)	(4)
$\ln Mig_{j,t}^{High,5y}$	561	561	561	561
$\ln Distance_{ij}$	561	561	561	561
$\ln GDPcap_i$	561	561	561	561
$\ln GDPcap_j$	561	561	561	561
$Contiguity_{ij}$	561	561	561	561
$Colony_{ij}$	561	561	561	561
$LegalSystem_{ij}$	561	561	561	561
FTA_{ij}	561	561	561	561
$\ln ANNCentrality_{i,t-5}^{IN-IN}$	561			
$\ln ANNCentrality_{i,t-5}^{IN-OUT}$		561		
$\ln ANNCentrality_{j,t-5}^{OUT-IN}$			561	
$\ln ANNCentrality_{j,t-5}^{OUT-OUT}$				561
Constant	2	2	2	2

Degree of freedom for each coefficient included in all the specifications reported in Table 3

DF have been computed using ANOVA

Table 11 Degree of freedom (from ANOVA)—cfr Table 4

	(1)	(2)	(3)	(4)
$\ln Mig_{j,t}^{High,5y}$	561	560	561	560
$\ln Distance_{ij}$	561	560	561	560
$\ln GDPcap_i$	561	560	561	560
$\ln GDPcap_j$	561	560	561	560
$Contiguity_{ij}$	561	560	561	560
$Colony_{ij}$	561	560	561	560
$LegalSystem_{ij}$	561	560	561	560
FTA_{ij}	561	560	561	560
$\ln Overlapping_{i,t-5}^{IN}$	561			
$\ln Complementary_{i,t-5}^{IN}$		560		
$\ln Overlapping_{j,t-5}^{OUT}$			561	
$\ln Complementary_{j,t-5}^{OUT}$				561
Constant	2	2	2	2

Degree of freedom for each coefficient included in all the specifications reported in Table 4

DF have been computed using ANOVA

Table 12 Interacting local connectivity

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}	Multilevel In NumFD _{ijt} ^{5y}
In Mig _{ijt} ^{High5y}	↓ 0.260*** (13.849)	0.239*** (12.562)	0.165*** (61.462)	0.228*** (12.481)	0.169*** (31.356)	0.176*** (26.270)	0.093*** (26.578)
In Degree _{ijt-5} ^{IN}	↓ (18.465)	0.392*** (8.712)	0.276*** (22.222)	0.368*** (9.315)	0.290*** (15.602)	0.226*** (36.885)	0.219*** (15.984)
In Degree _{ijt-5} ^{OUT}	↑ (-1.982)	0.040 (0.650)				0.173*** (3.853)	
In Mig _{ijt-5} ^{IN}	↓	-0.211* (-1.982)	0.706*** (5.823)			-0.505*** (-6.216)	
In Mig _{ijt-5} ^{OUT}	↑		0.542*** (5.359)			0.511*** (5.491)	
In Strength _{ijt-5} ^{IN}	↓			0.086*** (13.587)		1.328*** (8.226)	0.213*** (11.692)
In Strength _{ijt-5} ^{OUT}	↑			0.006 (0.524)	0.181***		0.158*** (9.927)
Obs	9225	9225	9225	9225	9225	9225	9225
Residual DF	9216	9214	9214	9214	9214	9212	9212

Table 12 (continued)

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}	In NumFDI ^{sy} _{ijt}
Wald test: High Skill Migrants	Chi2 0.61	29.32	102.99	35.64	63.54	5856.1	74.94
	Pval	0	0	0	0	0	0
Wald test: Investor	Chi2 0	17.06	22.49	0	120.12	154.2	0
	Pval	0	0	0	0	0	0
Wald test: Recipient	Chi2 0	58.00	34.23	103.19	0	0	.21
	Pval	0	0	0	0	0	.64

t-statistics in parentheses. Robust Standard Errors. Estimates refer to a 3-level regression with random intercepts. Dyads are nested in networks, all nested in time. All equations control for distance (in logs); per capita GDP (in logs); three dummies for contiguity, colonial history, and common legal system; and a constant. All regressors have been interacted with a dummy variable taking value 1 if total bilateral migrants' stock from country *j* to country *i* is below (↓) or above (↑) the average number of migrants received by country *i* from all countries in our sample in the current year. The Wald tests for equality of the interaction coefficients (Below vs. Above average of migrants at destination) of the migrants' network related statistics are reported at the end of the table

[†] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13 Interacting overlapping and complementary networks

Model		(1)	(2)	(3)	(4)
Dependent variable		Multilevel	Multilevel	Multilevel	Multilevel
		$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$
$\ln Mig_{ji,t-5}^{High,5y}$	↓	0.226*** (11.501)	0.237*** (11.431)	0.179*** (30.934)	0.174*** (38.512)
	↑	0.327*** (11.227)	0.329*** (9.574)	0.298*** (15.289)	0.286*** (16.324)
$\ln Overlapping_{i,t-5}^{IN}$	↓	0.068*** (11.685)	0.044*** (7.178)		
	↑	0.008 (0.773)	0.133*** (17.710)		
$\ln Complementary_{i,t-5}^{IN}$	↓		0.014*** (5.401)		
	↑		-0.132*** (-8.744)		
$\ln Overlapping_{j,t-5}^{OUT}$	↓			0.154*** (5.905)	0.096*** (4.679)
	↑			0.094*** (5.829)	0.123*** (5.674)
$\ln Complementary_{j,t-5}^{OUT}$	↓				0.091*** (9.017)
	↑				-0.003 (-0.334)
Obs		9225	9225	9225	9225
Residual DF		9195	9193	9214	9176
Wald test: high skill migrants:	Chi2	58.92	31.06	64.56	60.45
	Pval	0	0	0	0
Wald test: overlapping	Chi2	32.22	1154.89	33.88	9.61
	Pval	0	0	0	0
Wald test: complementary	Chi2		82.40		31.80
	Pval		0		0

t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln NumFDI_{ij,t}^{5y}$ represents the log of the total number of FDI projects over the past 5 year. All regressors have been interacted with a dummy variable taking value 1 if total bilateral migrants' stock from country *j* to country *i* is below (↓) or above (↑) the average number of migrants received by country *i* from all countries in our sample in the current year. The Wald tests for equality of the interaction coefficients (Below vs. Above average of migrants at destination) of the migrants' network related statistics are reported at the end of the table

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14 Interacting ANN

Model		(1)	(2)	(3)	(4)
		Multilevel	Multilevel	Multilevel	Multilevel
Dependent variable		$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$
$\ln Mig_{ji,t-5}^{High,5y}$	↓	0.246*** (13.016)	0.245*** (13.196)	0.171*** (24.973)	0.202*** (16.501)
	↑	0.372*** (9.500)	0.365*** (10.230)	0.314*** (12.436)	0.342*** (10.296)
$\ln ANNCentrality_{i,t-5}^{IN-IN}$	↓	0.115 + (1.725)			
	↑	-0.154** (-2.679)			
$\ln ANNCentrality_{i,t-5}^{IN-OUT}$	↓		0.210*** (3.787)		
	↑		-0.076 (-1.566)		
$\ln ANNCentrality_{j,t-5}^{OUT-IN}$	↓			-1.828*** (-6.275)	
	↑			-2.029*** (-6.164)	
$\ln ANNCentrality_{j,t-5}^{OUT-OUT}$	↓				-2.573*** (-7.692)
	↑				-2.815*** (-7.305)
Obs		9225	9225	9225	9225
Residual DF		9214	9214	9214	9214
Wald test: High Skill Migrants:	Chi2	30.59	37.2	49.84	37.95
	Pval	0	0	0	0
Wald test: ANN coefficients	Chi2	17.64	21.11	25.79	19.88
	Pval	0	0	0	0

t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln NumFDI_{ij,t}^{5y}$ represents the log of the total number of FDI projects over the past 5 year. All regressors have been interacted with a dummy variable taking value 1 if total bilateral migrants' stock from country *j* to country *i* is below (↓) or above (↑) the average number of migrants received by country *i* from all countries in our sample in the current year. The Wald tests for equality of the interaction coefficients (Below vs. Above average of migrants at destination) of the migrants' network related statistics are reported at the end of the table

+*p* < 0.1, **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Table 15 Binary and weighted local connectivity

	(1)	(2)	(3)	(4)
Model	Fixed effects	Fixed effects	Fixed effects	Fixed effects
Dependent variable	$i \times \text{yearFE}$	$j \times \text{yearFE}$	$j \times \text{yearFE}$	$j \times \text{yearFE}$
<i>Panel A: FE second stage FE estimates</i>				
$\ln \text{Degree}_{i,t-5}^{\text{IN}}$	0.298*** (47.46)			
$\ln \text{Degree}_{j,t-5}^{\text{OUT}}$		0.804*** (82.76)		
$\ln \text{Strength}_{i,t-5}^{\text{IN}}$			0.172*** (117.69)	
$\ln \text{Strength}_{i,t-5}^{\text{OUT}}$				0.171*** (60.01)
Model	Multilevel	Multilevel	Multilevel	Multilevel
Dependent variable	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$	$\ln \text{NumFDI}_{ij,t}^{5y}$
<i>Panel B: Multilevel regression</i>				
$\ln \text{Degree}_{i,t-5}^{\text{IN}}$	-0.0343 (-0.51)			
$\ln \text{Degree}_{j,t-5}^{\text{OUT}}$		0.612*** (5.42)		
$\ln \text{Strength}_{i,t-5}^{\text{IN}}$			0.059*** (19.75)	
$\ln \text{Strength}_{i,t-5}^{\text{OUT}}$				0.156*** (5.96)

t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln \text{NumFDI}_{ij,t}^{5y}$ represents the log of the total number of FDI projects over the past 5 year. First stage specifications in Panel A include country-time fixed effects for both origin and destination

$^+p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40888-024-00324-3>.

Acknowledgements We are grateful the two anonymous referees and to the editors for their suggestions and comments. We also thank Gianluca Santoni, Banu Demir, Christine Oughton, Rossana Mastrandrea, Valerio Leone Sciabolazza, and all the participants to the 1st YETI Meeting in Siena, the 21st AEEFI Conference, and the 2021 ITSG meeting in Ancona for their helpful comments and suggestions on earlier version of the paper. An earlier version of the paper with the same title circulated as a working paper. All errors are ours.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Data availability The dataset used in this paper relies on proprietary data and cannot be shared. All sources have been correctly listed in the text. The corresponding author will gladly share the code used to create the dataset and will make it available on his personal webpage.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1), 170–192.
- Aroca, P., & Maloney, W. F. (2005). Migration, trade, and foreign direct investment in Mexico. *The World Bank Economic Review*, 19(3), 449–472.
- Baldwin, R., & Taglioni, D. (2006). *Gravity for dummies and dummies for gravity equations* (No. w12516). National Bureau of Economic Research.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11), 3747–3752.
- Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1), 133–153.
- Bertoli, S., & Moraga, J. F. H. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102, 79–100.
- Brücker, H., Capuano, S., & Marfouk, A. (2013). Measuring migration flows. *Migration: New Developments*, 46(19198268), 31.
- Burchardi, K. B., Chaney, T., & Hassan, T. A. (2019). Migrants, ancestors, and foreign investments. *The Review of Economic Studies*, 86(4), 1448–1486.
- Chan, J. M., & Zheng, H. (2022). FDI on the move: Cross-border M&A and migrant networks. *Review of World Economics*, 158(3), 947–985.
- Checchi, D., De Simone, G., & Faini, R. (2007). *Skilled migration, FDI and human capital investment* (No. 2795). IZA Discussion Papers.
- Conte, M., Cotterlaz, P., & Mayer, T. (2021). *The CEPII gravity database*.
- Cuadros, A., Martín-Montaner, J., & Paniagua, J. (2016). Homeward bound FDI: Are migrants a bridge over troubled finance? *Economic Modelling*, 58, 454–465.
- Cuadros, A., Martín-Montaner, J., & Paniagua, J. (2019). Migration and FDI: The role of job skills. *International Review of Economics and Finance*, 59, 318–332.
- D'Agosto, E., Solferino, N., & Tria, G. (2013). Does FDI affect migration flows? The role of human capital. *Migration Letters*, 10(3), 359.
- Danchev, V., & Porter, M. A. (2018). Neither global nor local: Heterogeneous connectivity in spatial network structures of world migration. *Social Networks*, 53, 4–19.
- De Simone, G., & Manchin, M. (2012). Outward migration and inward FDI: Factor mobility between eastern and western Europe. *Review of International Economics*, 20(3), 600–615.
- Docquier, F., & Lodigiani, E. (2010). Skilled migration and business networks. *Open Economies Review*, 21(4), 565–588.
- Drzewoszewska, N. (2014). Multilevel modelling of bilateral trade flows between European Union countries. *Przegląd Statystyczny*, 61(3), 283–306.
- Eaton, J., & Kortum, S. (2001). Technology, trade, and growth: A unified framework. *European Economic Review*, 45(4–6), 742–755.
- Egger, P. (2000). A note on the proper econometric specification of the gravity equation. *Economics Letters*, 66(1), 25–31.

- Fagiolo, G., & Mastrorillo, M. (2014). Does human migration affect international trade? A Complex-Network Perspective. *PLoS ONE*, *9*(5), e97331.
- Fagiolo, G., & Santoni, G. (2016). Revisiting the role of migrant social networks as determinants of international migration flows. *Applied Economics Letters*, *23*(3), 188–193.
- Federici, D., & Giannetti, M. (2010). Temporary migration and foreign direct investment. *Open Economies Review*, *21*(2), 293–308.
- Fensore, I. (2016). *The impact of migration on foreign direct investments*. Mimeo.
- Fiorini, M., Giovannetti, G., Lanati, M., & Santi, F. (2021). Asymmetric cultural proximity and green-field foreign direct investment. *The World Economy*, *44*(9), 2572–2603.
- Flisi, S., & Murat, M. (2011). The hub continent. Immigrant networks, emigrant diasporas and FDI. *The Journal of Socio-Economics*, *40*(6), 796–805.
- Foad, H. (2012). FDI and immigration: A regional analysis. *The Annals of Regional Science*, *49*(1), 237–259.
- Garas, A., Lapatinas, A., & Poullos, K. (2016). The relation between migration and FDI in the OECD from a complex network perspective. *Advances in Complex Systems*, *19*(06n07), 1650009.
- Gheasi, M., & Nijkamp, P. (2017). A brief overview of international migration motives and impacts, with specific reference to FDI. *Economies*, *5*(3), 31.
- Giovannetti, G., & Lanati, M. (2017). Do high-skill immigrants trigger high-quality trade? *The World Economy*, *40*(7), 1345–1380.
- Giovannetti, G., Sanfilippo, M., & Velucchi, M. (2018). Diverse twins: Analysing China's impact on Italian and German exports using a multilevel quantile regressions approach. *Applied Economics*, *50*(28), 3051–3065.
- Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). Wiley.
- Gould, D. M. (1994). Immigrant links to the home country: empirical implications for US bilateral trade flows. *The Review of Economics and Statistics*, *1994*, 302–316.
- Head, K., & Ries, J. (2008). FDI as an outcome of the market for corporate control: Theory and evidence. *Journal of International Economics*, *74*(1), 2–20.
- Herman, P. R. (2019). *Modeling complex network patterns in international trade (2019 update)*. USITC Economics Working Paper Series No. 2017-04-A.
- Javorcik, B. S., Özden, Ç., Spatareanu, M., & Neagu, C. (2011). Migrant networks and foreign direct investment. *Journal of Development Economics*, *94*(2), 231–241.
- Kapás, J., & Czeglédi, P. (2020). The impact of culture on FDI disentangled: Separating the “level” and the “distance” effects. *Economia Politica*, *37*(1), 223–250.
- Kugler, M., & Rapoport, H. (2007). International labor and capital flows: Complements or substitutes? *Economics Letters*, *94*(2), 155–162.
- Leblang, D. (2011). *Another link in the chain: Migrant networks and international investment*. DIASPORA for, 79.
- Leblang, D. (2010). Familiarity breeds investment: Diaspora networks and international investment. *American Political Science Review*, *2010*, 584–600.
- Lücke, M., & Stöhr, T. (2018). Heterogeneous immigrants, exports and foreign direct investment: The role of language skills. *The World Economy*, *41*(6), 1529–1548.
- Mayer, T., & Ottaviano, G. I. (2008). The happy few: The internationalisation of European firms. *Intereconomics*, *43*(3), 135–148.
- Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, *93*(2), 351–363.
- Metulini, R., Sgrignoli, P., Schiavo, S., & Riccaboni, M. (2018). The network of migrants and international trade. *Economia Politica*, *35*, 763–787.
- Parsons, C., & Vézina, P. L. (2018). Migrant networks and trade: The Vietnamese boat people as a natural experiment. *The Economic Journal*, *128*(612), F210–F234.
- Peri, G., & Requena-Silvente, F. (2010). The trade creation effect of immigrants: Evidence from the remarkable case of Spain. *Canadian Journal of Economics/revue Canadienne D'économique*, *43*(4), 1433–1459.
- Peterson, M. F., Arregle, J. L., & Martin, X. (2012). Multilevel models in international business research. *Journal of International Business Studies*, *43*(5), 451–457.
- Portes, R., & Rey, H. (2005). The determinants of cross-border equity flows. *Journal of International Economics*, *65*(2), 269–296.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA Press.

- Rauch, J. E. (2001). Business and social networks in international trade. *Journal of Economic Literature*, 39(4), 1177–1203.
- Rauch, J. E., & Trindade, V. (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1), 116–130.
- Santi, F., Giovannetti, G., & Velucchi, M. (2021). *Migrants know better: Migrants' networks and FDI*. In DISEI Working Paper—Economics n.17/2021. Università degli Studi di Firenze.
- Sciabolazza, V. L. (2018). *Moving people: network analysis of international migration*. Networks of International Trade and Investment: Understanding globalisation through the lens of network analysis (p. 249).
- Sgrignoli, P., Metulini, R., Schiavo, S., & Riccaboni, M. (2015). The relation between global migration and trade networks. *Physica a: Statistical Mechanics and Its Applications*, 417, 245–260.
- Sheperd, B. (2021). *Gravity modeling tips and traps: A crash course for practitioners*. Trade Policy Research Forum.
- The Financial Times. (2018). *fDI markets database*.
- UNDESA. (2015). *Trends in international migrant stock: Migrants by destination and origin*. United Nations database, POP/DB/MIG/Stock/Rev.2015
- Wang, L. H. (2017). Accounting quality and information asymmetry of foreign direct investment firms. *Research in International Business and Finance*, 42, 950–958.
- White, R. (2007). Immigrant-trade links, transplanted home bias and network effects. *Applied Economics*, 39(7), 839–852.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.