UNIVERSITÀ DEGLI STUDI FIRENZE

DISEI
DIPARTIMENTO DI SCIENZE PER L'ECONOMIA E L'IMPRESA

WORKING PAPERS - ECONOMICS

Migrants know better: Migrants' networks and FDI

FILIPPO SANTI, GIORGIA GIOVANNETTI, MARGHERITA VELUCCHI

WORKING PAPER N. 17/2021

DISEI, Università degli Studi di Firenze
Via delle Pandette 9, 50127 Firenze (Italia) www.disei.unifi.it

Migrants know better: Migrants' networks and FDI

Filippo Santi[♠]*, Giorgia Giovannetti*[♦]*, Margherita Velucchi*[♣]

September, 2021

Abstract

We use the instruments of the social network analysis to revisit the relationship between international migration and Foreign Direct Investment (FDI) flows in the period between 2000 and 2015. Applying a multilevel mixed estimator inspired to the gravity literature, we test how and to what extent the structure of the international migrants' network contributes to bilateral FDI flows. We find that the inclusion of network level statistics exposes a much larger degree of complexity in the relationship between international migration and investments. Testing the assumption that migrants networks act as preferential channel for information with their homeland, we find evidence that a more diverse immigrant community in investing countries could "perturb" the flow of information at bilateral level, de facto translating into lower bilateral FDI.

♠ University of Bielefeld and University of Florence – filippo.santi@uni-bielefeld.de (Corresponding Author) –

♦ University of Florence and European University Institute

♣ European University of Rome.

The research as been partially funded with the research grant 11479 from the University of Florence. We are grateful to Gianluca Santoni, Banu Demir, Christine Oughton, Rossana Mastrandrea, Valerio Leone Sciabolazza for their helpful comments and suggestions. We also thank all the participants to the 1st YETI Meeting in Siena, the 21st AEEFI Conference, and the 2021 ITSG meeting in Ancona. All errors are ours.

1. Introduction

Cross-border migration and international capital mobility represent two characterizing features of the economic system of the XXI century. Despite the recent slowdown in international exchanges and mobility dating back to 2015 and exacerbated by the spread of Covid-19 (Antras, 2021;Evenett and Fritz, 2021; UNCTAD, 2021), their importance – both in terms of volumes and relevance in the political discourse - has been growing at accelerating speed for at least three decades.

This pattern is interesting also in light of two considerations. On the one side, migrants flows grew remarkably over the period 1990-2015, in spite of the absence of a clear effort toward reducing the legal barriers to international migration.[1] On the other side, Foreign Direct Investments (FDI) -especially greenfield FDI and unlike international trade - did not grow as much as the substantial removal of most of the legal barriers to capital mobility across countries would have suggested. With most of the physical and legal obstacles removed or highly reduced by technological progress and multilateral cooperation respectively, FDI were expected to increase by an order of magnitude similar to that of international trade. More recently, the attention shifted toward why the response of foreign direct investors to the relaxation such constraints have been mild, compared to other types of flows. Many studies started revisiting the specificities of FDI and other flows with respect to cultural, institutional and social divergence across countries.

In this respect, the role played by migrants' networks has been object of increasing interest in the past decades. Beyond the political debate, it is recognized in the recent literature that migrants establish a social and (in)formal infrastructure across countries, which facilitates communication and economic exchanges (Leblang, 2010). The relationship between bilateral trade flows and migrants networks has been extensively explored in the literature. Gould (1994); Rauch and Trindade, 2002; Peri and Raquena-Silvente (2010); Fensore et al. (2016); Giovannetti and Lanati (2016); Parsons and Vèzina (2018) among others found a strong evidence in favor of the positive impact of migration on trade flows. The existing evidence suggests that the 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'').

The role of migrants' networks on capital flows has been less explored in the existing literature and results are less clear-cut. On the one hand, Gheasi and Nijkamp, (2017). Aroca and Mahoney (2005) and Checchi et al. (2007) among others detected a negative impact of bilateral migration on FDI, contradicting the hypothesis that migrants provide better and more reliable information on economic conditions in the recipient country, a factor to which FDI should be sensitive to. On the other hand, Portes and Rey, (2005) and Kugler and Rapoport, (2007) suggest that the negative sign detected in the earlier studies was mostly related to the fact that the impact of low- and highly educated migrants was not singled out.[2] Controlling for the distribution of migrants across

[1] In this sense, the European Union represents an outlier if we consider internal mobility within the union's borders. Nonetheless, immigration to the EU from outside is still highly regulated and subject to important limitations.

[2] 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).

different educational levels, Kugler and Rapoport (2007); Docquier and Lodigiani (2010); Javorcick et al. (2011); Cuadros et al. (2016, 2019); Burchardi et al. (2019)) detected a positive impact of better educated migrants flows on bilateral investment decisions, suggesting that migration and investments are ultimately complementary.[3]

All the studies reviewed so far focused on migration within a country or as a mostly bilateral phenomenon. In this paper, we instead explore how FDI flow are influenced by the way countries integrate in the International Migrants Network (IMN). Approaching trade and migration as complex, interconnected networks, Fagiolo and Mastrorillo (2014) tested whether and how the migrants' ``business and social network effect'' triggered at bilateral level could also be affected by the overall structure of the IMN. Sgrignoli et al. (2015) and Metulini et al. (2018) found evidence that the efficiency of migrants in providing valuable information depends on the size of bilateral migration as much as on the way both the investing and the recipient economy integrate into the global migration system. Looking at the short term travel network, Fagiolo and Santoni (2015) explored whether better integration might lead to larger productivity gains (mediated by the effect of international mobility on trade flows). Garas et al (2016) detected a positive correlation between a country's integration in the international migration network and its bilateral investment position. Interestingly, despite the evidence seems to suggest that the educational endowment of migrants' network is crucial for triggering investments, no study so far acknowledged the skill and educational heterogeneity of migrants' network within a complex network perspective.

In this paper, we bring together these two strands of literature, to explore the relationship between human mobility and FDI between 20 OECD countries and the rest of the world. We therefore shed a new light on the relationship between the direct and the indirect effect of migration on FDI. Most of the existing literature mentioned above studies the effect of migration (either from $i \rightarrow j$ or the other way round) on FDI, but falls short from providing a clear picture of the indirect effects. We maintain that such indirect effects depend on the way both the home and the host countries are integrated in the IMN. As an example, let us consider a realistic case in which the resources available for investments in a country are limited. Then, a more diversified immigrant network might also imply a fiercer competition among different ethnic communities to channel the limited resources toward their homeland. Under a conceptual perspective, we make a significant step forward with respect to the existing studies, which consider the IMN as "undirected" (i.e. they do not explicitly account for the direction of the flows). In other words, no previous study makes a distinction between immigration to the host country (the investor in our framework) and the overall emigration from the home country (which we consider as the recipient of FDI). However, the "directed" nature of the network is likely to be significant, as the implications of a highly diversified immigrant community might differ substantially from those of a widely scattered emigrant community.

[3] Other studies include Federici and Giannetti (2010); Flisi and Murat (2011); De Simone and Manchin (2012); D'Agosto et al. (2013); Wang (2017).

We contribute to the existing literature in at three respects. First, we acknowledge the network structure of global migration, and we explore how the position of a country in the International Migration Network (IMN) influences bilateral investments decisions. Despite the substantial agreement on the positive impact of a country's position in the IMN and bilateral trade, the way the position in the IMN affects bilateral investment flows, and the distinction between emigration and immigration flows remain largely under-explored.[4] For instance, the existing evidence is inconclusive on the potential competition that a more diversified immigrant community might foster for directing FDI. To test for this form of "diversion effect", directed (i.e., not symmetric) network statistics - distinguishing the inward connectivity of an investing country in the IMN from to the outward connectivity of the recipient country - are needed.

Furthermore, we link the literature on the heterogeneity of migration flows to recent network approaches. As the scant evidence suggests, the complexity of an investment (as opposed to trade) 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); D'Agosto et al. (2013) among others find evidence that the skill composition of the diaspora affects the flow of investments, both inward and outward.[5] To the best of our knowledge, however, the way the migrants' skill composition affects FDI once the overall structure of the migrants' network is controlled for has not been addressed yet. The inclusion of both bilateral and (directed) network-related measures of international migration also helps disentangling the migration's direct effect from the indirect role played by third parties - as much as the structure of IMN itself - in shaping FDI patterns from the OECD to the rest of the world.[6]

Finally, this paper contributes to the literature also from a methodological point of view. Most of the existing literature adopt a structural gravity approach. We adapt instead a multilevel mixed approach (Goldstein, 2011; Rabe Hesketh and Skrondal, 2012) to dyadic data. This choice enables us to model the hierarchical structure of the data, which derives from the inclusion of country-level network statistics into the bilateral framework.

Our findings are consistent with the existing evidence on the migrants-FDI relationship. All the standard bilateral socioeconomic dimensions are confirmed to be crucial drivers of bilateral greenfield FDI flows, which are also positively influenced by better educated migrants. The results remain robust even after the structure of the overall network is taken into account. Our results suggest that bilateral FDI are both directly and indirectly affected by the degree of integration in the migrants' network. As the presence of an established bilateral migration corridor directly affects the economic exchanges between two countries, bilateral FDI appears to be positively influenced by a more central (in terms of both size and quality of the connections) of

[4] To the best of our knowledge, the study by Garas et al. (2016) represents the only attempt to study the link between migration and FDI within a network perspective. However, their analysis presents some methodological limitations in the way they estimated their gravity equation.

[5] 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 (2016).

[6] 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.

a country in the global migratory network. However, unlike previous studies, we also detect a negative effect of the immigrant community in the investing country at dyadic level. This finding suggests that while both direct and indirect migration positively contributes to bilateral FDI, segregation in the IMN - leading to a greater diversity of the domestic immigrant community in the investing country with respect to the emigrant diaspora from the recipient country - reduces bilateral FDI.

Under a research question perspective, our research resembles the one presented by Garas et al (2016), who studied the migrants' network effect on OECD investment flows toward the rest of the world. This paper takes distance from their work both conceptually and methodologically. First, it controls (consistently with the existing evidence on the heterogeneous effect of migrantss network on FDI) for the bilateral stock of educated migrants at bilateral level, rather than on total migration; second, it preserve the directed structure of the network; third, it embraces a different methodology, which made the two works not directly comparable in terms of outcome. Finally, the results reported by Garas and coauthors are to be taken with caution, due to a misspecification error in their gravity estimates. More precisely, they use PoissonPseudo Maximum Likelihood (PPML, Santos Silva and Tenreyro, 2006;2011) to estimate a gravity model with country and time fixed effect. However, instead of preserving the dependent variable (FDI stocks) in levels, they took the logarithmic transformation, introducing a misspecification in the model. In this paper, we provide to address the conceptual as well as the specification issues raised.[7]

2. Bilateral Greenfield FDI and the Migration Network: Data and Topology

2.1 Data

Data on FDI come from the fDIMarket database (Financial Times), which collects information on all Greenfield transactions who took place in the years 2003-2015. Being based on transactions, the dataset represents a major improvement with respect to standard BoP statistics. On the one hand, it enables us to know the *number* of investments occurred in the period of analysis rather than just their value. As the informative channel fed by migrants' networks is more likely to affect the decision of "making an investment" rather than its size (which depends on factors that might go well beyond the impact of diaspora), the possibility to "count the number of projects" rather than the overall foreign asset position is crucial.[8] On the other hand, transaction-level FDI data allow to distinguish between different type of trans-national investments. This point is relevant, given that the mechanisms linking migrants network to FDI might not be homogeneous, and different types of

[7] The replication of the analysis conducted by Garas and coauthors show that such misspecification error had large effect on the overall estimates: taking LHS variable in levels and running the exact same specification (on the same sample), we found a negative effect of the network level statistics the stock of bilateral FDI. Despite the misspecification has no attenuating explanations, many argument can be used to explain the result after the correct specification is adopted. First, migrants flows are likely to be not as effective at promoting FDI as the presence of long term migrants. Second, FDI stocks (from BoP) require additional modelling, in order to take into account possible disinvestments and devaluations that are unlikely to depend on the presence/diversion of migrants' flows.

[8] Concerning their value, transation-level data only report the initial capital investment, distinguishing between resources raised locally or in foreign markets. The figures are also free from all the subsequent reinvestment (and disinvestment), solving the problems related to the use of BoP data.

investments could respond differently to bilateral migration (Chen and Wang, 2017). A third and final point relates to the characteristics of the global business. In a world dominated by large multinational corporations, investments are often the results of complex strategies conducted on behalf of the Head Quarters (HQ) by controlled affiliates, which might be located abroad. This possibility would undermine the identification of the impact of migration on bilateral FDI. fDIMarket, provides information on the ultimate owner, eliminating this problem, which on the other hand affects other source of FDI data (such as BoP records).

In order to better identify the relationship between the migrants' network and bilateral FDI, we exclude all the investments in the extractive sector as well as those directed toward countries included in the tax havens list of the European Parliament.[9], since their motivation is hardly linked to migrants (and is rather driven by different reasons).

Bilateral migrants' stock data come from two distinct sources. On the one side, we need global coverage to build the International Migrants Network (IMN). To this end, we use the UNDESA Bilateral Migration Database (hereafter BMD. UNDESA, 2017 revision). As the BMD does not provide any information on the education/skill level of migrants, we integrate it with the information included in the IAB Brain Drain Dataset (Brucker et al., 2013), which collects information on the educational attainment of migrants, but only for 20 OECD destination countries. Data on the presence of highly educated migrants is therefore included to control for the pure bilateral effect. As a consequence, we focus on all channels directed toward the countries the IAB collects data from.[10] With respect to the choice of the data, we take a first step away from few works relating migrants' networks and investment (Garas et al., 2016), where the authors focused on migrants' flows rather than on stocks of foreign citizens at destinations. We maintain that stocks better capture the fact that FDI – more than trade - require collecting information and accumulating social capital at destination, a process that might require time. Furthermore, migrants' flows are more likely to reflect a country's generic openness, a fact that could undermine the identification of the true mechanisms underlying the FDI-Migrants' network relationship. 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, etc) come from the CEPII Gravity database as the information on common linguage (Conte et al., 2021; Melitz and Toubal, 2014) and the World Development Indicators (the World Bank, 2018). We report the main descriptive statistics in Table B-1 in the appendices.

[9] See https://www.europarl.europa.eu/cmsdata/147404/7%20-%2001%20EPRS-Briefing-621872-Listing-tax-havens-by-the-EU-FINAL.PDF for details.

[10] The countries included in the IAB are 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. I 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.

2.2 Topological Description: IMN vs GFDIN

We construct the International Migration Network (IMN) for the years 2005, 2010, and 2015 as N x N directed graph. Formally, the graph can be represented as a (weighed) $220 * 220$ adjacency matrix M^t in which every entry $m_{i,j}^t = MigStock_{i,j}^t, \forall\, i \neq j$ is defined depending on whether we are considering the binary (0-1) or the weighted representation. In the latter case, $m_{i,j}^t$ contains the number of citizens with i citizenship residing in country j in the reference year $t = 2005, 2010, 2015$. Both in the binary and the weighted representation, $m_{i,j}^t$ is set to null if and only if no citizens of country i is resident in country j.

Unlike past related studies (Fagiolo and Mastrorillo, 2014; Fagiolo and Santoni, 2015; Garas et al., 2016), we preserve the directed structure of the migrants' network rather than reducing it to its undirected representation. Indeed, the existing non-network literature reviewed earlier makes a distinction between the effect that emigration have on the origin country from those immigration exert at destination. Failing to account for such distinction in the construction of the network makes it impossible to explore the mechanisms linking human mobility to economic exchanges in general (and FDI in particular). For this reason, we preserve the directed nature of the IMN.[11]

Consistently with the existing literature, we also distinguish between direct and indirect network effects, which might emerge from the extent to which two countries share (or do not share) the same connections in the IMN. In facts, being embedded in shared, rather than segregated networks might change the way information flows among countries. In order to capture the existence of such potential ``third-party'' effects, we also consider the extension to which two countries network overlap.[12] Despite the focus of this paper aims at exploring the migrants' network effects on bilateral greenfield FDI, it is interesting to understand how and to what extent the two flows evolved under the network perspective. This is done in Table 1, where we compare the structure of the IMN and of the global greenfield FDI network (GFDIN). We limit our perspective to the years 2005, 2010 and 2015, years in which we have full data overlapping.[13] Looking at the overall connectivity (Panel A), we notice how the IMN is substantially more dense (higher number of strictly positive connections) than the GFDIN. The fact that more countries exhibit positive migration flows but null (or very

[11] Working with undirected networks generally solves many computational issues, not last, the possibility to estimate in a single step the bilateral network effect on the dependent variable. Nonetheless, it does not allow to understand the channels through which the network dimension of the migrants' network affects bilateral FDI, which constitute an important issue. Suppose to take the aggregate bilateral flow between two countries (denoted i and n). This implies that the value of the flows (stocks) going in either direction between them are summed together ($\boldsymbol{ji = i \rightarrow j + j \rightarrow i}$). Suppose that $\boldsymbol{i}$ receives a modest number of migrants from $\boldsymbol{j}$, despite a large number of investments owing the other direction; on the other side, $\boldsymbol{j}$ could invest very little in $\boldsymbol{i}$, While receiving a huge number of migrants from it. Summing up migrants and FDI flows (stocks) between those two countries would make impossible to distinguish them from another pair (***c,d***) which is characterized by average flows (stocks) of FDI and migrants in either direction.

[12] To the best of our knowledge, this paper represents the first attempt to include measures of indirect network integration of this kind while preserving the direct nature structure of the network.

[13] 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.

low) investment flows is consistent with the low number of firms that are able deal with complex forms of internationalization such as FDI (see for instance Ottaviano and Mayer, 2008). The comparison the two networks over time shows very different dynamics in the extensive margin of the two networks. While the IMN tends to remain stable over time, the GFDIN grew substantially between 2005 and 2010, to stabilize in the following period. Panel B of Table 1, which compares the size of the two networks (and their variation) over time, clarifies this point. Differently from earlier studies (see Fagiolo and Mastrorillo, 2013, 2014; Sciabolazza, 2018) the IMN did not change substantially over the period considered, suggesting that changes in the patterns of human mobility require more time. On the other hand, the GFDIN experienced a noticeable growth, proxied by the decrease in both its average path length (APL) and diameter. However, and in spite of such growth, the GFDIN remains far sparser than the IMN.[14] Additionally, the gap between the two is only partially attenuated if we look at investment stock data (see for instance UNCTAD bilateral stock statistics), despite the number of non-null flows slightly decreases.

Table 1: Network Comparison - General Connectivity

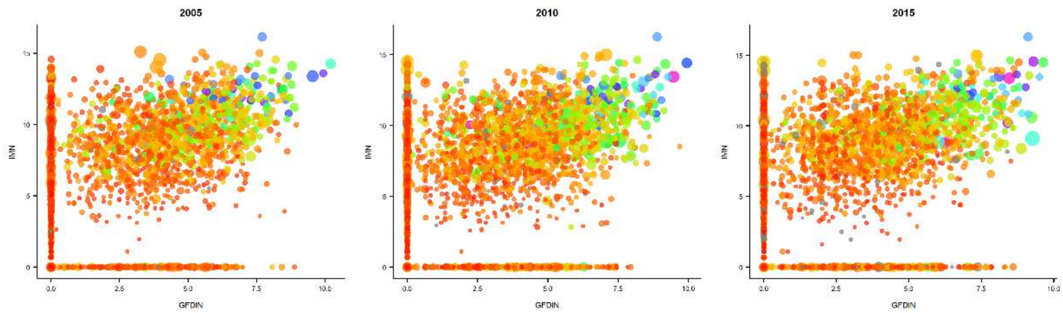
PANEL A: Connectivity						
	2005		**2010**		**2015**	
	IMN	**GFDIN**	**IMN**	**GFDIN**	**IMN**	**GFDIN**
Nodes Count	220	228	220	228	220	228
Edges Count	10534	1802	10685	2399	10688	2347
Average In-degree	47.88	7.9	48.57	10.52	48.58	10.29
Min. In-degree	0	0	2	0	2	0
Max. In-Degree	212	47	204	57	204	64
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
PANEL B: Size						
	2005		**2010**		**2015**	
	IMN	**GFDIN**	**IMN**	**GFDIN**	**IMN**	**GFDIN**
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

Notes: Network Topological Comparison. APL = Average Path Length. ***Source:*** fDIMarket database (FDI) and UNDESA Population Division (Migrant' stock).

[14] There are some discrepancies between the GFDIN and the overall FDI network built by Garas et al. (2016). For instance, they find a much higher (negative) assortative score in both networks. This fact could be due to the different types of data used. To begin with, they focus on FDI stocks rather than flows, and does not distinguish between Greenfields and M&A. Thus, their result might be driven by the other component of aggregate FDI statistics. Conversely, I focus on Greenfield FDI flows only. Concerning the IMN, they build a flow network (Abel and Sanders, 2014). I discussed above the rationale behind choosing to focus on stocks rather than flows: as Flows are more likely to reflect short run synergies that may share the same drivers of FDI, they are also more prone to potential reverse causality issues, which are mitigated by the use of cumulated migrants stock data.

More insight can also be compared by comparing graphically the correlation between the two networks. In Figure 1, we plot the simple correlation between the link weights of every directed pair $i \rightarrow j$ in the IMN and the GFDIN.[15] Not surprisingly, the pattern is similar to the one identified in the literature on Trade and Migration networks: dyads that are characterized by larger migration in either direction ($i \rightarrow j$) are also characterized by larger investment flows going in the opposite direction ($i \leftarrow j$). Furthermore, the economic and demographic size of the countries in the dyad are positively correlated with link weights in either networks, consistently with the gravity literature on FDI and migration.[16] The rough graphical comparison suggests that the IMN and the GFDIN might have co-evolved over time.

Figure 1- Link Weight Trend Over Time



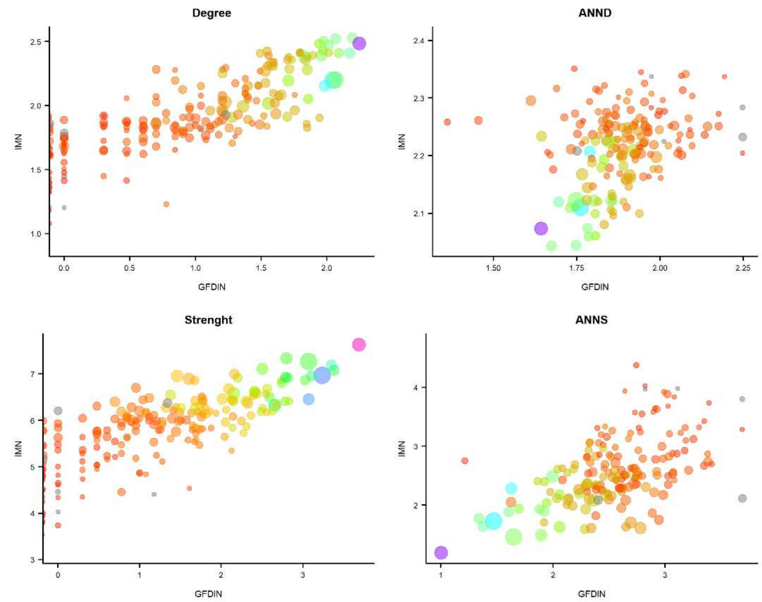
Notes: Log-Log scale. Markers size is proportional to the product of the population of country i and country j. Color (from red to light blue) captures the dyadic product between per capita GDP

Looking at the year 2010, Figure 2 and 3 compare the position of each country and the type of connections that prevails in the two networks. Each point in Figure 2 represents a country as an ordinate couple ($centr_i^{GFDIN}$; $centr_i^{IMN}$), colored and sized according to its per capita GDP and population respectively. Larger and blue nodes are characterized by a larger connectivity, both in binary and in weighted terms. Coherently with the pattern showed at dyadic level (Figure 1), larger economies tend be more central, while smaller and poorer countries (in terms of population and per capita GDP) tend to remain peripheral in both networks. Conversely, the correlation between and the average centrality of a country's directly connected neighbors (right hand side plots of Figure 2) show how wealth and market size are both inversely correlated with the centrality of the neighbors. In other words, it is easier (and more remunerative) for less integrated countries to connect with already central economies (and this trend appears to reinforce as long as the network grows). Moreover, more integrated countries (higher centrality) appears to have a greater probability to connect with peripheral ones. This is shown in Figure 3, in which each point represents an ordinate couple $centr_i^{net}$; $ANNcentr_i^{net}$ (ANNcentr being the average centrality of the direct neighbors of country i in network net. See Barrat et al. 2007).

[15] Conversely from the subsequent econometric exercise, we consider all potential pairs of countries in either networks, irrespectively from the availability of disaggregated migration data.

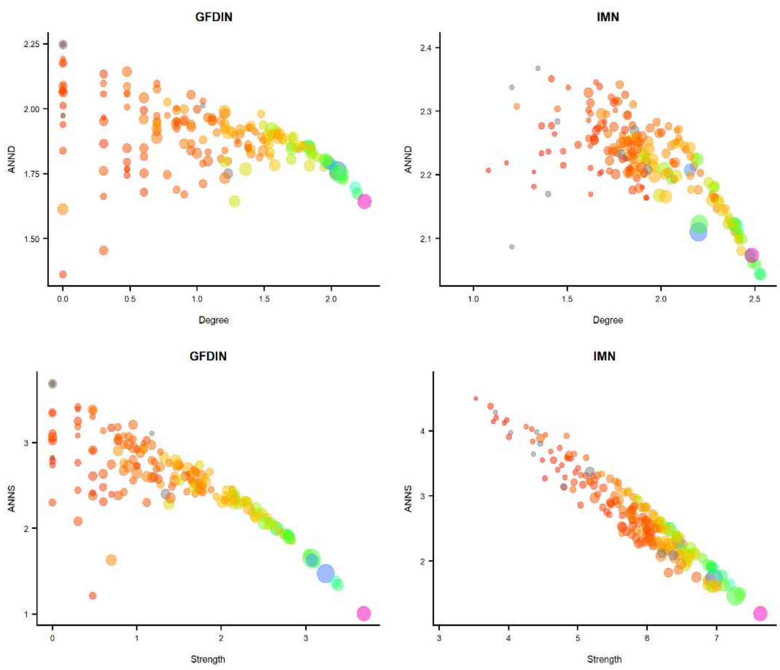
[16] We base the coloring scheme (from red to light blue) on the product of i and j's populations, while size reflects the product of per capita GDP of i and j. Coherently with the gravity framework, larger and bluer dots have a higher probability of being in the north-east side of the plot.

Figure 2- Node Centralities



Notes: Log-Log scale. Markers size is proportional the population of the country, while colors from red to light blue re_ect the per capita GDP. ANN(D,S) = Average Nearest Neighbour Degree or Strength (Barrat et al., 2007)

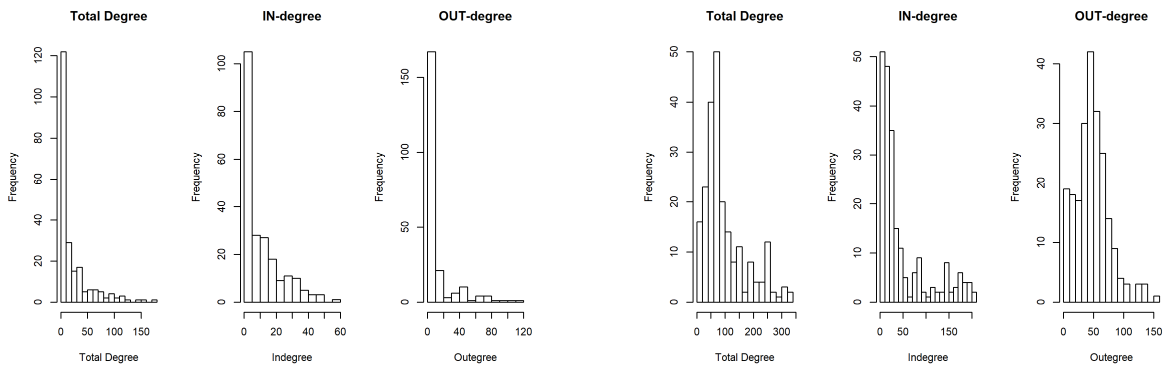
Figure 3 - Dissortative patterns



Notes: Log-Log scale. Markers size is proportional the population of the country, while colors from red to light blue reflect the per capita GDP. ANNC = Average Nearest Neighbor Centrality (generic for ANND and ANNS)

To conclude this descriptive section, we compare the degree distribution of the two networks (Total, IN, and OUT degree respectively) for the GFDIN (on the left) and the IMN (on the right). Figure 4 shows that both networks share a heavily skewed distribution, with the vast majority of nodes having little or no connections, and few large hubs. Together with the evidence provided by Figure 3, we can conclude that both networks are characterized by a dissortative structure. Overall, 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 to 3 also suggest that gravity related dimensions are also highly correlated with the link-weight distribution in both the IMN and the GFDIN.

Figure 4 - Degree Distributions



Greenfield FDI Network *International Migration Network*

Notes: Degree distributions refer to the entire network of countries that recorded a non-null flow in either the IMN or the GFDIN.

3. Econometric Framework and Estimation Strategy

We now move on from the topological comparison of the two networks, to test the relationship between the migrants' network and the flows of FDI at bilateral level, taking explicitly into account the effect of a country's position within the IMN. The inclusion of network statistics has represented a particularly interesting extension to the existing evidence, as they allow to keep the global scale of migration, as well as the way countries integrate in it, into consideration.

We maintain that the inclusion of network statistics bring additional complexity to the structure of the data. On one hand, we consider network-level statistics as belonging to a higher order of information with respect to the dyad level information. On the other hand, the comparison conducted by Herman (2019) for the case of a trade network suggests that the inclusion of different combinations of fixed effects (particularly, $country\ x\ time$ compared to the inclusion of $country\ \&\ time$ fixed effects) returns very similar estimates, both in terms of the probability of link formation and in terms of trade volume. The author also compare the results of the two types of gravity (naïve and structural) with a non-structural, simplified gravity model of network formation, with and without importer and exporter fixed effects, concluding that network formation and growth might depend on complex network patterns that are not accounted for by structural gravity. There is also an additional consideration, that relates to the "structure" of our sample: as data on highly educated migrants are only available for 20 destination countries, we are de facto excluding all FDI flows occurring among most migrants' origin countries (around 200 countries). Thus, we could not draw structural estimates even including the ful set of fixed effects, as most of the interaction occurring at world level are dropped.

In what follows, we adopt a "middle way", by estimating a Multilevel Mixed Effects Regression (MMER), where the "fixed" component controls for the usual gravity controls and for country and time fixed effects. The additional layers (network and time) are then considered as further levels of information to be modelled specifically. Multilevel models are still rarely used in the related literature (see for instance, Drzewoszewska, 2014; and Giovannetti et al., 2018), which usually rely on fixed effects to control for the influence of the multilateral resistance (Anderson and Van Wincoop, 2003).

A multilevel mixed model constitutes a useful alternative approach to the usual fixed effects gravity models, in a setting where the structural interpretation is prevented by the arbitrary exclusion of a large part of the flows. Under an econometric perspective, it also offer an alternative way to control for the heterogeneity of the hierarchical data, allowing the correlation of the model's stochastic component to vary across dyads.[17]

We expect the analysis of this hierarchical dataset via MMER to lead to new and interesting insights complementing the existing literature, despite the departure from structural gravity.

3.1 The model

The hierarchical structure of our data can be represented as

$$y_{ni,t} = \alpha_0 + \sum_{s=1}^{P} Z_{sit} + \sum_{r=1}^{R} X_i^r + u_{ni} + e_{ni} + v_t$$

Where $r = (1, \dots, R)$ and $s = (1, \dots, P)$ collect the variables of interests at a country pair level and the additional controls at a network level with u_{ni}, e_{ni}, v_t being the "three levels" of the error term (country pair, network and time) . 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).

The associated empirical equation is:

$$\ln NumFDI_{ij,t}^{5y} = \alpha_0 + \ln Mig_{ji,t}^{High,5y} + \sum_{net=1}^{nw} \gamma_{net} X_{i,t}^{net,5y} + \sum_{net=1}^{nw} \gamma_{net} X_{j,t}^{net,5y} + \sum_{r=1}^{R} \beta_r x_{ji,t}^r$$

The dependent variable $\ln NumFDI_{ij,t}^{5y}$ represents the log of number of greenfield FDI projects from country i to country n, cumulated over the precedent 5 years period.. Log-linearization implies that we are excluding null FDI flows. The low share of null flows in our sample however suggests that distortion to the

[17] 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 and Jones, 2015).

estimates, if any, are negligible. The variable $\ln Mig_{ji,t-5}^{High,5y}$ represents the 5-years lagged bilateral stock of tertiary educated migrants from country n to country i. $\sum_{net=1^R} \gamma_{net} X_{i,t-5}^{net,5y}$ and $\sum_{net=1^R} \gamma_{net} X_{j,t-5}^{net,5y}$ are two vectors of both direct and indirect measures of centrality (local and global) as well as measures of third-party network statistics. Similarly to the bilateral term, also the network statistics are included as lags. There are two order of considerations for using lagged migration-related variables. On the one hand, lags reduce reverse causality, which might affect the relationship between migration and FDI; on the other hand, they acknowledge that migrants' networks need time to structure themselves and become able to establish those channels for information to flow. A description of the network statistics included in the econometric exercise is reported in Appendix A.

Finally, the term $\sum_{r=1}^{R} \beta_r x_{ji,t}^r$ includes the set of geographical, cultural and economic bilateral controls.

4. Results

Table 2: Binary and Weighted Local Connectivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	*Multilevel*	*Multilevel*	*Multilevel*	*Multilevel*	*Multilevel*	*Multilevel*	*Multilevel*
Dependent Variable	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$	$\ln NumFDI_{ij,t}^{5y}$
$\ln \boldsymbol{Mig_{ji,t}^{High,5y}}$	0.263***	0.264***	0.207***	0.246***	0.209***	0.204***	0.121***
	(16.97)	(15.65)	(38.08)	(14.95)	(32.04)	(31.54)	(105.72)
$\ln \boldsymbol{Degree_{i,t-5}^{IN}}$		-0.034				0.082+	
		(-0.51)				(1.87)	
$\ln \boldsymbol{Degree_{j,t-5}^{OUT}}$			0.612***			0.621***	
			(5.42)			(5.72)	
$\ln \boldsymbol{Strength_{i,t-5}^{IN}}$				0.059***			0.209***
				(19.75)			(11.57)
$\ln \boldsymbol{Strength_{i,t-5}^{OUT}}$					0.156***		0.250***
					(5.96)		(7.55)
LRtest	1545.63***	1479.55***	1305.26***	1386.03***	1363.26***	1233.10***	1219.51***
Obs	9225	9225	9225	9225	9225	9225	9225

Notes: * p < 0:05, ** p < 0:01, *** p < 0:001. 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. Table C-1 in appendix C 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; and a constant. ***LRtest*** refers to the log-likelihood Ratio test of the multilevel specification against the linear model (Rabe-Hesketh and Skrondal, 2012). The rejection of the null hypothesis implies a systematic difference between MIXED and OLS/FE estimates.

We analyze and discuss the results for three distinct aspects of network integration in the IMN. All the estimates are drawn from our sample of 20 OECD countries investing in - and receiving migrants from - all other countries in the world.[18] All equations include a set of standard gravity controls (not reported in the tables

[18] For reason of space, we do not report the coefficients for the non migrant-related controls. 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. Complete tables are available upon request.

but available on request), which all maintain the expected sign and magnitude across different specification. Additionally, we test each model against its OLS/FE counterparts, and we report the related likelihood ratio statistics at the bottom of each column. The test suggest that our multilevel specification dominates OLS/FE alternatives. Overall, our results are in line (as far as it concerns magnitude and significance) with the literature on the effect of migration on FDI: the coefficient for $\ln Mig_{ni,t}^{High,5y}$ confirms the importance of the highly skilled migration on bilateral investment, reinforcing the idea that migrants and investments might be complementary rather than substitutes. What happens when we include network centrality into our model? The remaining columns of Table 2 and the subsequent sets of results explore different measures of directed local centrality (degree and strength) in the IMN as additional explanatory variables.

On average, FDI appears to benefit from a more central position in the IMN of either the recipient and the investing countries. This means that the more connections a country forms (columns 2,3, and 6 respectively), the more investments are likely to flow from $i \rightarrow j$. In line with the empirical gravity literature on FDI and migration, also the volume of such connections is a relevant determinant of outward FDI.[19] However, the recipient outward centrality seems to outplay the investing side's inward centrality. This finding is consistent across all specifications, and suggests that the informative channel fed by migrants might be influenced by third party effects.

The measures of centrality considered in Table 2 however do not take into account the quality of each country's position in the network, nor allow us to explore the sources of third party effects. For instance, it could be that a country's local centrality on the IMN captures a generic openness of an economy, with little correlation with the informative channels triggered by bilateral migration.

A better way explore the presence and impacts of third party 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 effect operating through those countries (and the migrants' community within them) who mutually connect in the IMN.to both i and j.[20] Table 3 reports the estimates of the coefficients related to the weighted IMN subnetwork. Consistently with our framework, we consider investing (immigration) country i's inward connectivity as opposed to 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 $\ln Overlapping_{j,t-5}^{OUT}$) indicate that a larger number of the shared connections between i and j (i.e. higher the number and volume of countries to which the recipient j sends migrants to, and that in turn send migrants to i) translates into a higher number of bilateral FDI going from $i \rightarrow j$. Looking at the size of the non-shared ties however allows us to detect an interesting distinction. In fact, while the size and the volume of the complementary network for country j still has a positive impact on bilateral FDI from $i \rightarrow j$, a more diversified immigrant community in

.
[19] 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.
[20] 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.

the investing country i (with respect to j's network, captured by the coefficient $\ln Complementary_{i,t-5}^{IN}$) appears to reduce bilateral FDI between the two countries, even if the magnitude of such effects remain low. This finding suggests that, once shared and exclusive connections are taken into account separately, the overall positive effect of bilateral migration is ultimately mitigated, due to competition among different ethnic communities to attract FDI. The "noise" introduced by information from exclusive migratory channels might therefore dilute the direct positive effect of bilateral migration. This effect is coherent with the idea that the capacity and the resources required for international investments are much more limited than those required to foster bilateral trade. Interestingly, the empirical literature has not highlighted this result yet, even though it could partially explain the less intense response to migrants' network of bilateral FDI flows with respect to international trade.

Table 3 - Overlapping and Complementary Networks

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

Notes: * p < 0:05, ** p < 0:01, *** p < 0:001. t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln NumFDI_{ni,t}^{5y}$ represents the log of the total number of FDI projects over the past 5 year. The coefficient ***LRtest*** refers to the log-likelihood Ratio test of the multilevel specification against the linear model.

4.1 Additional results

In the previous section, we have considered the effect of local connectivity in the IMN on bilateral FDI. This means that we only explored the effect of the number (and volume) of connections countries i and j have in the IMN, leaving apart the structure of the network centered on each country as well as the relative importance of the countries to which i and j are connected with in the IMN. In order to further investigate the mechanisms linking the IMN to bilateral FDI, we can consider different indicators.

In Table 4, we expand further on the result that emerged from Table 3 by looking at the relative importance of a dyad first-order connections (i.e. those countries that are directly connected to either i or j – or both).

Looking at the first two columns, it appears that both the inward and the outward centrality of the first-order connections of the investing country i (respectively $\ln ANNCentrality_{i,t-5}^{IN-IN}$ and $\ln ANNCentrality_{i,t-5}^{IN-OUT}$) always exert a positive impact on bilateral FDI toward j. This might be due to the fact that countries with more central connections have higher probability of receiving migrants from country j, that in turn might be able to convey information also through their indirect connections with i. Conversely, the relative importance of the outward connectivity of country j seems to play against the positive effect of the direct bilateral channel. This result is not at odd with the previous ones: if we consider that the direct bilateral migration has an effect on its own, the more important are the neighbors of country j, the stronger are likely to be the ties with the neighbors as opposed to a generic investing country i. The results from Table 4 are partially in line with the existing literature on trade and migrants' network.

Table 4 – Average Nearest Neighbor Centrality

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

Notes: * p < 0:05, ** p < 0:01, *** p < 0:001. t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln NumFDI_{ni,t}^{5y}$ represents the log of the total number of FDI projects over the past 5 year. The coefficient of the ***LRtest*** refers to the log-likelihood Ratio test of the multilevel specification against the linear model.

5. Conclusions

Most of the literature studying the effect of migrants' network on international economic exchanges focuses on the bilateral direct effect of migration, relying on different specification of the gravity model to account for the dyadic nature of such flows. Despite a substantial agreement about the positive impact of migrants on economic exchanges– as consumers, investors, and valuable source of information across countries, there are still many unstudied aspects of the role they may play. Economic relationships between countries tend in facts to be defined on a much larger and global scale, where each movement might have non-trivial repercussions the entire economic system. Only recently the attention has turned to the role played by interconnected interactions. The inclusion of network statistics into more traditional econometric methodologies is an attempt to account for such complexity.

This paper contributes to this evolving strand of literature, by reconciling the existing evidence on the heterogeneous effects of migrants' networks on bilateral FDI with the insights offered by the social network analysis tools. First, we use network related statistics to describe the structure of the International Migration Network and of the Greenfield FDI Network, and we show why it is relevant to consider bilateral investments and migration as just two narrow segments of wider phenomena. Differently from previous studies, we preserve the directed nature of the IMN, to explore the specific differences between emigration and immigration connections across countries. Next, we use different measure of network centrality to shed a new light on the way migration affects greenfield FDI.

Finally, due to the hierarchical structure of our dataset, we leave the widely adopted structural gravity framework, adopting a multilevel mixed approach that allows a better fit of the data and an improved treatment of the error term. Running a Longitudinal Multilevel Model, we properly treat three-level data (country pair, network and time) and, controlling for the bilateral stock of highly educated migrants as well as for an extensive set of gravity related controls, we found a positive and statistically significant effect of a more central position of both the investing and the recipient countries in the IMN. Our results point in the direction of a much more complex relationship between migration and bilateral FDI. In particular, we find evidence of a competition between migrants' communities from different countries. If FDI can be thought of in the same terms of rival goods, our results suggest that the real effect of international migration on FDI could be far more complex than previously thought.

REFERENCES

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science*, *343*(6178), 1520-1522.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American economic review*, *93*(1), 170-192.

Anderson, J. E., Larch, M., Yotov, Y. V. (2019). Trade and Investment in the Global Economy: a Multi-country dynamic analysis. *European Economic Review*, °120, 103311

Antràs, Pol. 2021. "De-Globalisation? Global Value Chains in the Post-COVID-19 Age." 2021 ECB Forum: "Central Banks in a Shifting World" Conference Proceedings.

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*(19,198,268), 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. (2017). *FDI on the Move: the Facilitation of Greenfield FDI and Cross-border M&A by Migrant Networks*. Working Paper, Chinese University of Hong Kong.

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 & 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.

Evenett, Simon and Fritz, johannes (2021) THE 27TH GLOBAL TRADE ALERT REPORT Advancing Sustainable Development With FDI: Why Policy Must Be Reset

Fagiolo, G., & Mastrorillo, M. (2013). International migration network: Topology and modeling. *Physical Review E*, *88*(1), 012812.

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.

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., & Poulios, 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). John Wiley & Sons.

Gould, D. M. (1994). Immigrant links to the home country: empirical implications for US bilateral trade flows. *The Review of Economics and Statistics*, 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.

Kugler, M., & Rapoport, H. (2007). International labor and capital flows: Complements or substitutes?. *Economics Letters*, *94*(2), 155-162.

Leblang, D. (2010). Familiarity breeds investment: Diaspora networks and international investment. *American political science review*, 584-600.

Leblang, D. (2011). Another link in the chain: Migrant networks and international investment. *DIASPORA for*, 79.

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., Patuelli, R., & Griffith, D. A. (2018). A spatial-filtering zero-inflated approach to the estimation of the gravity model of trade. *Econometrics*, *6*(1), 9.

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.

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.

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.

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*, 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.

The Financial Times (2018). fDI Markets database.

UNCTAD 2021. The World Investment Report

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.

APPENDICES

A. Definitions of Network Centrality

(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

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

where aij 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.

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

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).[21]

Average Nearest Neighbor Centrality (Barrat et al., 2007). 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., 2007) is defined as

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

[21] 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.

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 figure 3), as it allows to identify a network assortativity mixing patterns. In the econometric exercise, we consider th 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

$$Overlap_{ij} = I = (1, \dots, n), J = (1, \dots, m) \qquad I \cap J = 0$$
$$Complementary_{ij} = I = (1, \dots, n), J = (1, \dots, m) \qquad I \cap J \neq 0$$

Where $I = (1, \dots, n)$ and J= $(1, \dots, m)$ represent the set of channels flowing in/to country i and j respectively.[22] 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.

[22] The reverse holds to compute the overlapping and complementary outward network

B. Summary Statistics

Table B1: Summary statistics

	Obs	Mean	SD	min	MAX
$\ln NumFDI_{ij,t;t}^{5y}$	9,225	0.999	1.396	0	7.659
$\ln Mig_{ji,t-5}^{High,5y}$	9,225	5.944	2.639	0	14.090
Measures of Local Connectivity (centrality)					
$\ln Degree_{i,t-5}^{IN}$	9,225	5.151	0.226	4.111	5.357
$\ln Degree_{j,t-5}^{OUT}$	9,145	3.954	0.447	1.792	5.037
$\ln Strength_{i,t-5}^{IN}$	9,225	14.428	1.352	11.921	17.590
$\ln Strength_{j,t-5}^{OUT}$	9,145	13.024	1.591	7.006	16.561
(NON)-Overlapping Network					
$\ln Overlapping_{i,t-5}^{IN}$	9,225	3.406	1.113	0	5.313
$\ln Complementary_{i,t-5}^{IN}$	9,225	3.699	0.522	1.386	5.142
$\ln Overlapping_{j,t-5}^{OUT}$	9,225	3.702	0.435	1.099	4.875
$\ln Complementary_{j,t-5}^{OUT}$	9,225	4.811	0.259	3.89	5.236
Barrat et al.'s ANNCentrality Measures					
$\ln ANNCentrality_{i,t-5}^{IN-IN}$	9,225	3.862	0.222	3.586	4.849
$\ln ANNCentrality_{i,t-5}^{IN-OUT}$	9,225	3.406	0.277	3.021	4.327
$\ln ANNCentrality_{j,t-5}^{OUT-IN}$	9,225	4.762	0.176	4.157	5.204
$\ln ANNCentrality_{j,t-5}^{OUT-OUT}$	9,225	4.215	0.088	3.929	4.505

Notes: Summary Statistics based on the estimation sample, computed on the observations for which data on highly educated migration is available. The sample includes flows from the IAB-20 to the rest of the world.

C. Multilevel vs Fixed Effects Estimator

Economic literature usually analyzes bilateral data with resorting to variations of a gravity equation, which are generally estimated by means of fixed effects (FE) estimation. However, while the gravity of trade and the gravity of migration have a convincing theoretical justification to prefer FE estimators (see for instance Anderson and Van Wincoop, 2003; and Bertoli and Huertas-Moraga, 2013). In this paper, instead, 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 , higher than the bilateral perspective included in gravity analysis. This hierarchy can be endogenized using a multilevel approach, rather than ruled out by fixed effects estimation that might lead to specification errors.[23](Eaton and Kortum, 2001; and Head and Ries, 2008, Baldwin and Taglioni, 2006).

[23] 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 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.

In this paper, we directly model the hierarchical structure of data, leaving the correlation of the error term to vary across dyads, by adopting a Multilevel Mixed Model (Rabe/Hesketh and Skrondal, 2012). This approach merges the fixed 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 (Jones and Bell, 2015).

Table C-1 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 $country \times year$ FE on country specific features, it is not possible to estimate a second stage including partner's characteristics. For this reason, table c-1 does not reports the replication of columns (6) and (7) , where both in-degree(-strength) of country i and out-degree(-strength) of country n were included. Beyond the greater flexibility, this last issue reinforces our argument in favor of a multilevel mixed modelling.

Table C-1: Binary and Weighted Local Connectivity

Panel A: FE second stage FE estimates				
	(1)	(2)	(3)	(4)
Model	*Fixed Effects*	*Fixed Effects*	*Fixed Effects*	*Fixed Effects*
Dependent Variable	$i \times year\ FE$	$j \times year\ FE$	$j \times year\ FE$	$j \times year\ FE$
$\ln \boldsymbol{Degree^{IN}_{i,t-5}}$	0.298***			
	(47.46)			
$\ln \boldsymbol{Degree^{OUT}_{j,t-5}}$		0.804***		
		(82.76)		
$\ln \boldsymbol{Strength^{IN}_{i,t-5}}$			0.172***	
			(117.69)	
$\ln \boldsymbol{Strength^{OUT}_{i,t-5}}$				0.171***
				(60.01)

Panel B: Multilevel Regression				
	(1)	(2)	(3)	(4)
Model	*Multilevel*	*Multilevel*	*Multilevel*	*Multilevel*
Dependent Variable	$\ln NumFDI^{5y}_{ij,t}$	$\ln NumFDI^{5y}_{ij,t}$	$\ln NumFDI^{5y}_{ij,t}$	$\ln NumFDI^{5y}_{ij,t}$
$\ln \boldsymbol{Degree^{IN}_{i,t-5}}$	-0.0343			
	(-0.51)			
$\ln \boldsymbol{Degree^{OUT}_{j,t-5}}$		0.612***		
		(5.42)		
$\ln \boldsymbol{Strength^{IN}_{i,t-5}}$			0.059***	
			(19.75)	
$\ln \boldsymbol{Strength^{OUT}_{i,t-5}}$				0.156***
				(5.96)

Notes: * p < 0:05, ** p < 0:01, *** p < 0:001. t-statistics in parentheses. Robust Standard Errors. The dependent variable $\ln NumFDI^{5y}_{ni,t}$ 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.