The Relation Between Global Migration and Trade Networks

Paolo Sgrignoli
Rodolfo Metulini
Stefano Schiavo
Massimo Riccaboni

The Relation Between Global Migration and Trade Networks

Paolo Sgrignoli
IMT Institute for Advanced Studies Lucca

Rodolfo Metulini
IMT Institute for Advanced Studies Lucca

Stefano Schiavo
Department of Economics and Management, University of Trento

Massimo Riccaboni
IMT Institute for Advanced Studies Lucca

The Relation Between Global Migration and Trade Networks

Paolo Sgrignoli, Rodolfo Metulini, Stefano Schiavo, Massimo Riccaboni

Abstract

In this paper we develop a methodology to analyze and compare multiple global networks. We focus our analysis on the relation between human migration and trade. First, we identify the subset of products for which the presence of a community of migrants significantly increases trade intensity. To assure comparability across networks, we apply a hypergeometric filter to identify links for which migration and trade intensity are both significantly higher than expected. Next we develop an econometric methodology, inspired by spatial econometrics, to measure the effect of migration on international trade while controlling for network interdependencies. Overall, we find that migration significantly boosts trade across sectors and we are able to identify product categories for which this effect is particularly strong.

Index Terms

Trade; Migration; Networks; Gravity model; Spatial econometrics

I. Introduction

Since the mid 1990s a growing body of research has investigated the relation between global networks of international trade, information flows and migrations. Whereas traditional trade theory (e.g. the standard Heckscher-Ohlin model) suggests that the movement of goods across borders can provide a substitute for the movement of production factors, the bottom line of this branch of research is that the two actually complement each other. This appears to hold for different countries [for the US, Canada and Spain respectively, see 1, 2, 3] and has recently been confirmed by a meta-analysis covering 48 different studies [4].

Most of the literature referred to above shares a common empirical strategy, based on the estimation of a log-linear gravity model where bilateral trade flows are regressed over standard explanatory variables (economic mass and distance), the stock of immigrants from specific partner countries and other controls aiming at capturing various types of trade costs (common language, colonial relationships and the like). The most influential contributor to the field is probably James Rauch who, in a series of papers with a variety of co-authors, looks at the role of ethnic immigrants networks in facilitating trade. The core of the argument is that formal and informal links among co-ethnic migrants in other countries and at home, the "network" in Rauch's terminology, facilitate trade by providing potential trading partners with easier access to valuable, i.e. qualified, information. The pro-trade effect thus stems from the reduction of the trade barriers and search costs associated with market

transactions. Since these costs are likely to be larger for international trade due to distance, language and cultural differences, legal provisions and the like, "networks" end up being especially relevant in facilitating cross-border transactions. Indeed, the central result in Rauch and Trindade [5] is that the positive effect of migration on trade is larger for "differentiated goods", i.e. those items that are not homogeneous and are not traded in organized exchanges therefore rendering that knowledge about counterpart reputation particularly valuable.

Similar results have been replicated by a number of subsequent authors. Peri and Requena-Silvente [3], for instance, analyze the Spanish case and find that doubling the number of immigrants from a given country increases export to the same destination by 10 percent. This effect is higher for firms selling differentiated products and for more distant countries (geographically or culturally). All of these elements are consistent with the notion that networks (in this case the presence of a large community of expatriates and their connections with co-nationals at home and abroad) lower the hurdle in terms of economic interactions, providing better access to information and trade opportunities and reducing the fixed costs associated with entry into a foreign market.

Easier access to information via co-ethnic migrants does not only increase export from the recipient to the home countries, but also facilitates "chain migration": it is typical for communities with a significant presence of migrants to attract more migrants from the same community. This feature is consistent with the preferential attachment mechanism that accurately describes the evolution of many real-world networks (from airline traffic to the World Wide Web, from social ties to financial networks) and makes migration interesting in terms of complex network analysis. Yet, not much has been written on the subject: Slater [6] studies clustering in US internal migration, whereas Simini et al. [7] presents a stochastic radiation model that can be used to predict international migration patterns.

More recently, Fagiolo and Mastrolillo [8] study the topology of the migration network, and its evolution over the period 1960-2000. They find that the network (where links between two countries A and B are given by the stock of migrants originated in country A and living in country B in a given year) is disassortative and highly clustered, and displays a small-world binary pattern. Furthermore, they show that the structural properties of the network are mainly driven by socioeconomic, geographical, and political factors. In a second paper [9] they look at the interplay between trade and migration, finding that the two networks are strongly correlated, and country-pairs that are more central in the migration network tend to trade more.

In this paper we aim at filling this gap by analyzing the relationship between migration and trade through the lens of complex network analysis and spatial econometrics. To do so, we first make the networks comparable, using a hypergeometric benchmark to judge if the link intensity between two countries is significantly higher than expected. Next, we run a set of econometric regressions, checking for network interdependencies: by means of spatial econometric techniques, we check for the presence of network (auto)correlations between trade and migration, disentangling the analysis in different parts, and using the previously defined network matrices as weights in spatial autoregressive models. Our goal is to identify the subset of products whose export/import is the most highly related to the presence of a community of migrants.

This paper is structured as follows. In Section II we describe migration and trade data, as well as our methodological approach. In Section III we analyze and compare the migration and trade networks by distinguishing

different product types. In Section IV we presents the econometric analysis and discuss the findings. Finally, Section V contains discussions of our main results, further research directions and conclusions.

II. DATA & METHODOLOGY

A. Migration and trade data

In this paper we use data regarding migrants from The World Bank's Global Bilateral Migration dataset [10]: it is composed by decennial matrices of bilateral migrant stocks spanning 1960-2000 (5 census rounds), disaggregated by gender and based primarily on the foreign-born definition of migrants. It is the first and only comprehensive picture of bilateral global migration over the second half of the 20th century, taking into account a total of 232 countries. The data reveal that the global migrant stock increased from 92 million in 1960 to 165 million in 2000. Quantitatively, migration between developing countries dominates, constituting half of all international migration in 2000, whereas flows from developing to developed countries represent the fastest growing component of international migration in both absolute and relative terms. As for international trade, we use the dataset that was cleaned and made compatible through a National Bureau of Economic Research (NBER) project led by R. Feenstra [11], disaggregated according to the Standardized International Trade Code at the four-digit level (SITC-4). It provides for each country the value (expressed in thousands of US dollars) exported to all other countries, for 775 product classes. In our analysis, we focus on the years 1970 and 2000, although choosing different years does not alter the results.

In addition, to improve our analysis, we look at the SITC product code traded between each country pair and perform a decomposition of goods following the Rauch classification [5]. We thus distinguish between *homogeneous* and *differentiated* goods: the former are those with a reference price, whether it be a result of organized exchanges or simply of price quotations in a specialized journal; while the latter are without such a reference price and can be thought of as 'branded' commodities. An important distinction between these two categories is that homogeneous product prices can be quoted without mentioning the name of the manufacturer, like many chemical compounds for example, letting specialized traders stay informed regarding their prices around the globe and perform international commodity arbitrage by matching distant buyers and sellers much like traders in an organized exchange. In other words for this category of products it is immediately apparent if the differential between the two countries' markets is large enough to cover customs and transportation costs. On the contrary, differentiated products are such that their commodity categories are usually not well defined (e.g. footwear); they need to be disaggregated into various sub-types, a process that leads to the limit where each category contains only one supplier. These products are, in this respect, 'branded' or differentiated. From this definition one would expect for international trade to be more heavily influenced by migrant networks for this second class of products, as buyers and sellers need to be matched in the product characteristics space. This is indeed the result found by Rauch [5] and one of the aspects we test in this paper.

Our aim is to use the two datasets together. In this respect, we retain only the countries present in both of them to enhance comparability. Doing this we end up with a set of 146 nodes, that populate both the trade and migration networks.

B. Methodology

The different nature of the data, both with regards to type and measurement unit, make a direct comparison of the trade and migration networks unreliable. As such what is needed is a tool to make the two datasets comparable. To address this issue, we use a stochastic benchmark for normalization purposes, as recently used in [12] and previously introduced in [13]. This method has been used to identify statistically significant portions of data in fields ranging from genetics to network theory [14, 15].

For two countries, A and B, let N_A be the value of goods exported by country A and N_B the value of goods imported by country B. The total value of traded goods is N_k and the observed value of goods exported from A to B is N_{AB}. Under the null hypothesis of random co-occurrence, that is to say, customers in country B who are indifferent to the nationality of the exporter, the probability of observing X thousands US dollars of goods traded is given by the hypergeometric distribution

$$H(X|N_k, N_A, N_B) = \frac{\binom{N_A}{X}\binom{N_k-N_A}{N_B-X}}{\binom{N_k}{N_B}} \ , \tag{1}$$

and we can associate a *p*-value with the observed N_{AB} as

$$p(N_{AB}) = 1 - \sum_{X=0}^{N_{AB}-1} H(X|N_k, N_A, N_B) \ . \tag{2}$$

Note that the described null hypothesis directly takes into account the heterogeneity of countries with respect to the total value of goods traded (row and column totals). For each pair of countries, we separately evaluate the *p*-value for each trade relationship and then use a cutoff to select only those links that represent a significant departure from the hypergeometric benchmark ($p < .01$). The resulting matrices are then dichotomized[1]. Thanks to our stochastic approach we can treat a weighted network as a random graph in which any link has a probability of occurrence. We use this method to filter both trade and migration networks, to make them comparable.

In the next Section we study the structural properties of trade and migration networks. We use then networks as weight matrices in a spatial gravity model of trade flows (Section IV).

III. NETWORK ANALYSIS

We first describe and compare some basic topological properties of the two networks separately (Section III-A); we then to analyze the interconnections and correlations among their links (Section III-B).

A. Topological properties

In this section we introduce and analyze some topological indicators of the *migration* and *trade* (both total and separated in differentiated and homogeneous products) networks for the year 2000.

One of the most important topological properties of a network is the degree distribution, $P(k)$. This quantity measures the probability of a randomly chosen vertex to have k neighbors. Figure 1 (main) shows the undirected cumulative distribution of trade and migration networks, defined as $P_c(k) \equiv \sum_{k'>k} P(k')$. In all cases the cumulative distributions show a flat approach from the origin, similar to the Erdös-Rényi (ER) network; but then, at about $k = 10$, the distributions follow a power law decay, with a strong deviation from the exponential

---

[1]The hypergeometric multi-urn benchmark is equivalent to the Monte Carlo degree-preserving network rewiring procedure [16]

tail predicted by the classical random graph theory. This is especially true for the trade networks' cumulative distribution, whereas the migration case has a less evident power law behavior and deviation from the ER case.

Table I presents a selection of well-known statistics for these networks. We can observe that the trade networks share, in general, very similar properties, although they are always slightly different for the migration case. In particular, while the migration network has a lower connectivity, as indicated by the nodes' average degree and density, its links are more reciprocal, meaning that a bilateral bond is more common than in commercial trades.

TABLE I

YEAR 2000. D = DIFFERENTIATED; H = HOMOGENEOUS.

	Migration	Trade	Trade (D)	Trade (H)
Average degree	29.45	38.47	34.30	36.19
Density (%)	15.1	20.4	18.5	19.3
Corr. coeff.	0.45	0.79	0.70	0.78
Reciprocity (%)	34.5	30.2	28.2	29.1
Average cluster coeff.	0.099	0.088	0.073	0.089
Assortativity	-0.009	-0.428	-0.278	-0.368

Another important aspect is the hierarchical structure of the networks, which is usually analyzed by means of the clustering coefficient and degree-degree correlation. The *clustering coefficient* of vertex i is defined as $c \equiv 2n_i/k_i(k_i - 1)$, where n_i is the number of neighbors of i that are interconnected. The values of the clustering coefficient for our networks are $\simeq 0.1$, much lower than in other analyses of the world trade network [See for example 17], indicating that the hypergeometric filtering we used in our work to dilute it resulted in the loss of the clustered structure of the network. Hierarchy is also reflected in the degree-degree correlation through the conditional probability $P(k|k')$, i.e. the probability that a vertex of degree k' is linked to a vertex of degree k. This function is difficult to measure, due to statistical fluctuations, and it is usually substituted by the *average nearest neighbors degree (ANND)*, defined as $\langle k_{nn}(k) \rangle = \sum_{k'} k' P(k'|k)$ [18]. For independent networks this quantity would result independent of k. Figure 1 (inset) reports the ANND for the migration and trade networks, showing a dependency on the vertex's degree and indicating that in all the networks highly connected vertices tend to connect to poorly connected vertices (which is usually called *dissortative* behavior). It is clear how this phenomenon is much more pronounced in trade than in migration, as can also be seen by the assortativity coefficient[2] in Table I.

B. The Relation between migration and trade

A first approach to analyzing the interplay between international trade and human migration is to observe whether, for a pair of countries, a strong bond in one aspect is related to a strong tie in anther. To do this we make use of the Jaccard index [19] that, given two sets of events, is defined as $J = |A \cap B| / |A \cup B|$, therefore representing the ratio between the number of events shared between the two sets, over the number of events in at least one of them. In our case the events will represent that the link between two countries is significant[3] both in the trade and the migration networks.

---

[2]Also known as *Pearson's degree correlation*

[3]Significant in the sense of Section II-B, so to say with the *p*-value below the threshold we set at 0.01.
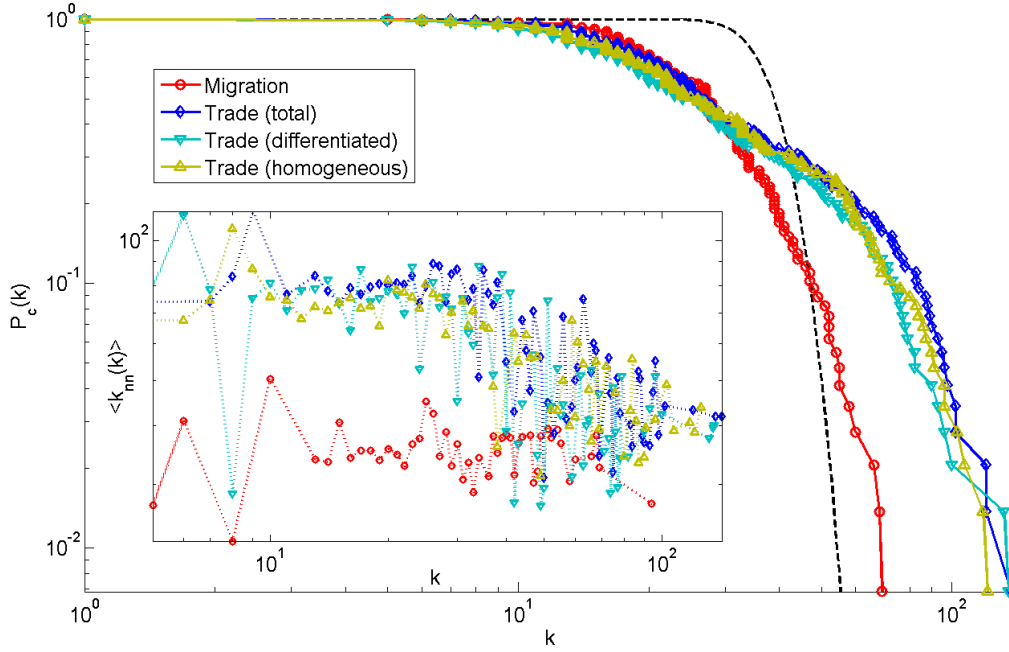
Fig. 1. Main: cumulative degree distribution $P_c(k)$ for the trade and migration networks. The dashed line is the same degree distribution for a random graph with the same average degree. Inset: average nearest neighbors' degree as a function of the total degree of the vertex.

In Figure 2 (main) we can see the distribution of the Jaccard index (J) for all the individual products in the years 1970 and 2000: both present a clear power law behavior and show how for products for the year 2000, the overlap with migration is generally higher. In the inset we plot the total number of overlaps in the networks for individual countries: this too presents a power-law behavior, although the two distributions cross at some point, indicating that countries with a stronger relation between trade and migration had a higher network effects in 1970, while countries with a weaker bond had a higher one in 2000.

In Tables II and III we also list the top and bottom countries and products, respectively, that appear in the distributions.

To verify the validity of the Rauch classification [5] we repeated the same analysis separating *differentiated* and *homogeneous* goods: in the main panel of Fig. 3 we plot the distribution of the Jaccard index for individual products while in the inset we show the Jaccard index for the entire networks in the years 1970-2000. As one can observe, in this first analysis the Rauch classification seems to be verified by our results, with values for the differentiated products generally higher for both individual SITC codes and the whole networks.

To further investigate this point, in Section IV we compare the results for the classification proposed by Rauch and the that emerges from the overlapping of the trade and migration networks (see Table III). In particular, we take the *top-N* products[4] from our ranking and compare them with Rauch's differentiated ones, and do the

---

[4]N indicates here the number of differentiated products in each year.

same for the bottom ranked with homogeneous commodities. We call these two new categories *overlapping* and *non-overlapping*[5].
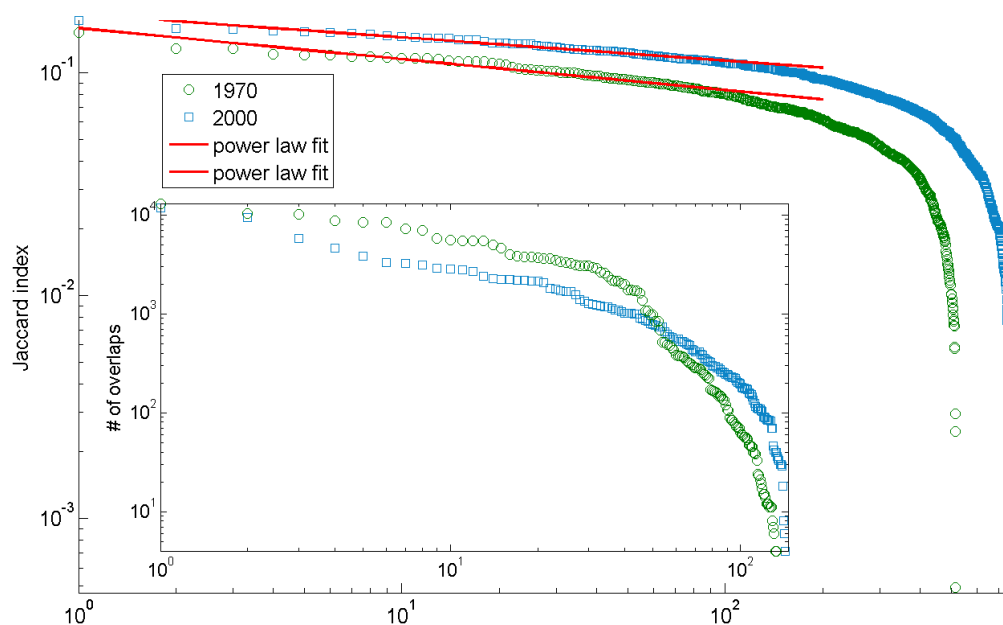

Fig. 2. Main figure: Distribution of the Jaccard index relative to the overlapping of single products with the migration networks for years 1970 and 2000. Red lines are power law fits. Inset: Distribution of the number of overlaps between the trade and migration networks for individual countries.

TABLE II

A LIST OF THE COUNTRIES THAT BEST AND WORST OVERLAP BETWEEN TRADE AND MIGRATION NETWORKS, BASED ON THE JACCARD INDEX, YEAR 2000.

#	Country
1	Germany
2	United Kingdom
3	Italy
4	Spain
5	Netherlands
⋯	⋯
11	China
⋯	⋯
142	Sierra Leone
143	Somalia
144	Saint Pierre and Miquelon
145	Tanzania, United Republic of
146	Uganda

[5]In our new classification, for the year 1970 the overlapping (non-overlapping) products are composed by $54\% - 46\%$ ($27\% - 73\%$) of Rauch's differentiated-homogeneous products; while for the year 2000 the figure changes to $62\% - 38\%$ ($37\% - 63\%$)

TABLE III

A LIST OF THE PRODUCTS (SITC-4) WITH THE MOST AND LEAST OVERLAP WITH THE MIGRATION NETWORK, BASED ON THE JACCARD INDEX, FOR YEAR 2000. D = DIFFERENTIATED; H = HOMOGENEOUS.

#	Product	Type
1	Miscellaneous articles of plastic	D
2	Insulated electric wire. cable. bars. etc	D
3	Plastic packing containers. lids. stoppers and other closures	D
4	Switches. relays. fuses...; switchboards and control panels...	D
5	Edible products and preparations. nes	D
6	Chemical products and preparations. nes	D
7	Other furniture and parts thereof. nes	D
8	Machinery for specialized industries and parts thereof. nes	D
9	Other polymerization and copolymerization products	D
...	...	...
762	Copra	H
763	Palm nuts and kernels	H
764	Manila hemp. raw or processed but not spun. ...	H
765	Uranium depleted in U235. thorium. and alloys. nes; ...	H
766	Ores and concentrates of uranium and thorium	H
767	Castor oil seeds	D
768	Coal gas. water gas and similar gases	H
769	Wood-based panels. nes	D
770	Knitted or crocheted fabrics. elastic or rubberized	D

IV. ECONOMETRIC APPROACH

Gravity models have often been used to explain *Origin-Destination* (OD) flows that arise in fields such as trade, transportation and migration. The term *spatial interaction models* has been used in the literature to label models that focused on flow between origins and destinations [20]. A great deal of literature on theoretical foundations for this model exists [21, 22], as well as literature focused on empirical studies [21, 22, 23, 24]. These models rely on a function of the distance between an origin and destination in addition to explanatory variables pertaining to characteristics of both origin and destination regions. Spatial interaction models assume that using distance as a variable will eradicate the spatial dependence among the OD pairs. However, the notion that using of distance functions in conventional interaction models effectively captures spatial dependence in interregional flows has long been challenged.

It has been recently fully recognized that a spatial interaction effect exists [25, 26, 27, 28], essentially due to the spatial spillover and the third country effect. This motivates the introduction of the spatial autoregressive components in the so-called *spatial gravity model of trade*. The spatial autoregressive component makes use of an $n \times n$ weight matrix in order to define the set of the spatial neighbors. The correct specification of the weight matrix is still debated. It is, however, reputed that the correct specification of this matrix, conditioned to the topic and to the objectives of the analysis, exists. One of the most frequently used weight matrices is based on spatial contiguity: it contains 1 if the pair shares geographical borders, 0 otherwise.

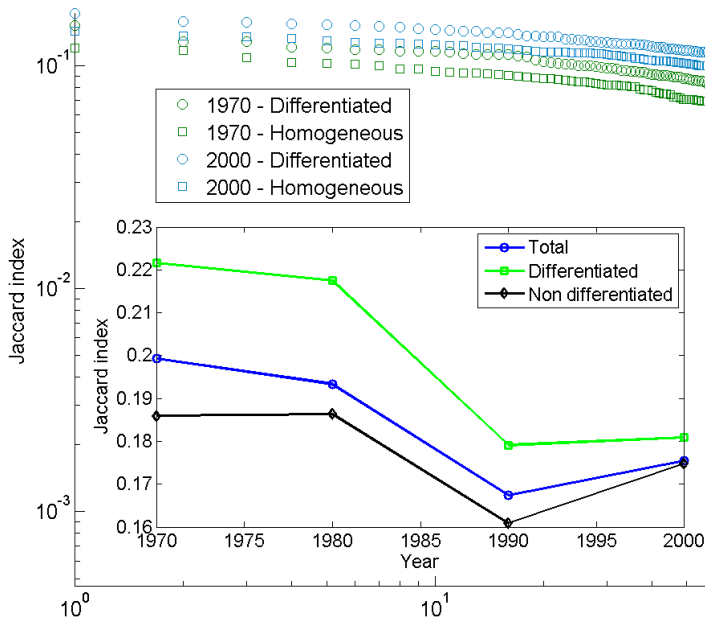W: $w_{i,j} = 1$ if i is neighbor of j, $w_{i,j} = 0$ otherwise.

Fig. 3. Main figure: Distribution of the Jaccard index relative to the overlapping of single products with the migration networks, shown separately for *differentiated* and *homogeneous* products, for years 1970 and 2000. Inset: Jaccard indexes relative to the overlap of the entire trade and migration networks for different years. A distinction is made between all the products and just differentiated or homogeneous ones.

Other formulations are also proposed based on an inverse distance approach, as well as on the similarity of technologies in the pair or regions. It is important to underline that the neighboring regions of OD flows include neighbors of the origin location, neighbors of the destination location, and perhaps a link between neighbors of the origin and neighbors of the destination regions. To account for this, the matrix dimension that suits in this case is $n^2 \times n^2$, and it is generally constructed (as LeSage and Pace proposed [29]) as the Kronecker product of the W matrix with itself.

$$W_K = W \otimes W$$

Network theory and spatial econometrics are intimately connected. Leenders [30] proposed using SAR models employing an ad-hoc W matrix based on the network relations (in terms of social influences and communication); Farber et al. [31] analyzed the relationship between the topology property of networks and the properties of spatial models, performing some simulation tests. Mansky's paper [32] is seminal, as it laid the foundation for analyzing the exogenous, endogenous and correlated effects that researchers encounter both in network and econometric theory. Lee, Liu and Lin [33], following Mansky's work, proposed a specification and estimation of network econometric models, in the presence of exogenous, endogenous and correlated effects. Furthermore, the discussion on the specification and estimation of the network econometric models has become quite widespread as of late [34, 35]. To control for network effects, here we use a similar approach and replace the spatial weight the network of trade weight matrix.

The specification of the W matrix becomes the following:

$$W : w_{i,j} = 1 \text{ if } i \text{ has a significant trade relationship}$$

$$\text{with } j \ (p < 0.1), w_{i,j} = 0 \text{ otherwise.}$$

A. Model specification

Le Sage and Pace [29] proposed to revise spatial specification for OD relationships based on two classes of models: *Spatial autoregressive models* (SAR) and *Spatial Durbin / Spatial error models* (SDM/SEM). The former is motivated by a time relationship describing a diffusion process over space; the latter by the omitted variables argument. We can think of the omitted variable as an unobserved tendency of trade due to network interdependencies.

The econometric representation of the models can be illustrated as:

SAR:

$$y = \rho \mathbf{W} y + x \beta_1 + \epsilon \tag{3}$$

SDM:

$$y = \rho \mathbf{W} y + x \beta_1 + \mathbf{W} x \beta_2 + \epsilon \tag{4}$$

which becomes the SEM model in the event that included and excluded variables are not correlated (Common factor test can be performed [29]).

SEM:

$$y = x \beta_1 + \lambda \mathbf{W} \epsilon + \epsilon \tag{5}$$

B. Estimation methods

The concentrated maximum likelihood estimated proposed by Anselin [36] and revised by LeSage and Pace [29] generally fits the spatial gravity model of trade. This methods is based on the computation of the log-determinant, which is quite time-consuming. Fortunately, new technical results regarding the calculation of the log determinant have recently come to light, which facilitates estimations [37, 38, 39]. One caveat for this method is that for cases in which a large number of zero flows exists, the method is not appropriate. Maximum likelihood (ML) estimation requires that the dependent variable vector follow a normal distribution, or that it is able to be suitably transformed to achieve normality. A partial solution is generally the logarithmic transformation. Another solution is to consider trade as count data using the Poisson model [40].

C. Data, application and econometric results

We conduct a cross-section on two different years, 1970 and 2000, on a sample of 146 world countries. For each year we test a cross section using three different dependent variables: (i) total trade; (ii) differentiated goods; (iii) non-differentiated goods. This allows us to test the Rauch theory, in addition to commenting on the relations between trade of differentiated goods and migration.

The gravity model is a benchmark in international trade and seems to be the best choice for our aims. For the gravity model the trade flow depends on the economic dimension of both the origin and the destination country,

and on the distance between the couple. Following this approach, we use the gross domestic product (per capita) and the population as control variables. Furthermore, we control for contiguity, for free trade agreements and for common language and currency, much like what is done in many empirical applications regarding the gravity model of trade. We found all of these variables within the CEPII dataset [41, 42]. However, the control variable we are most interested in are the migrants stocks: we collected this data from the World Bank dataset [10].

Appending the spatial autoregressive components in the econometric formula of the gravity model seems to be fundamental in order to grasp the potential contributions of the networks of trade and migration. The concentrated maximum likelihood estimation is the most frequently used estimation method in spatial econometrics because it overcomes issues related to the intrinsic endogeneity in the spatial models. However, the ML estimates require normal distribution, which rarely happens when zero flows are present. We resolve this issue by performing a logarithmic transformation of the dependent and the control variables.

We did a likelihood ratio (LR) and common factor test to choose from different specifications of spatial econometric models, and the results of these tests point in favor of the Durbin model with both the autoregressive component in the dependent variable (trade) and in the control variables (including stock of migration).

In the previous section, we computed four 146×146 matrices for two distinct time periods (the years 1970 and 2000), representing the (i) network of country to country migration; (ii) network of country to country total trades; (iii) network of country to country differentiated goods; (iv) network of country to country non-differentiated goods. We use these matrices as weight matrices in the spatial econometric model, which can be basically represented in this fashion:

$$t_{00} = \rho \boldsymbol{W_{t00}} t_{00} + x_{00}\beta_1 + \beta_2 m_{00} + \gamma \boldsymbol{W_{t00}} m_{00} + \epsilon$$

$$dt_{00} = \rho \boldsymbol{W_{tf00}} dt_{00} + x_{00}\beta_1 + \beta_2 m_{00} + \gamma \boldsymbol{W_{tf00}} m_{00} + \epsilon$$

$$ndt_{00} = \rho \boldsymbol{W_{tn00}} ndt_{00} + x_{00}\beta_1 + \beta_2 m_{00} + \gamma \boldsymbol{W_{tn00}} m_{00} + \epsilon$$

$$t_{70} = \rho \boldsymbol{W_{t70}} t_{70} + x_{70}\beta_1 + \beta_2 m_{70} + \gamma \boldsymbol{W_{t70}} m_{70} + \epsilon$$

$$dt_{70} = \rho \boldsymbol{W_{tf70}} dt_{70} + x_{70}\beta_1 + \beta_2 m_{70} + \gamma \boldsymbol{W_{tf70}} m_{70} + \epsilon$$

$$ndt_{70} = \rho \boldsymbol{W_{tn70}} ndt_{70} + x_{70}\beta_1 + \beta_2 m_{70} + \gamma \boldsymbol{W_{tn70}} m_{70} + \epsilon$$

where with tr we represent the total trade between pair of countries, dtr the amount for the differentiated goods and $ndtr$ the non-differentiated goods. 1970 and 2000 remain the time periods. ρ is the coefficient of the lagged trade term, β_1 the coefficients of the benchmark control variables like *GDP* and population, β_2 is the migration coefficient and γ is the lagged migration coefficient. Moreover, m stay for stock of migration. To conclude, W_t is the $n^2 \times n^2$ network of total trade Kronecker matrix, W_{tf} is the $n^2 \times n^2$ network of differentiated trade, W_{tn} is the $n^2 \times n^2$ network of non-differentiated trade and W_m is the $n^2 \times n^2$ network of migration.

We also perform the baseline gravity model using a new dependent variable representing a novel version of differentiated products that we call *overlapping* (see Section III-B).

We computed these models in R with the *spdep* package. Results are shown in Tables IV, VI and V. The most important results can be summarized as follows:

- in all models the positive effects on trade of the benchmark variables is confirmed (positive effect for *gdpcap* and *pop*, negative for distance);

TABLE IV

BASELINE GRAVITY MODEL FOR TRADE - YEAR: 2000, STANDARD ERRORS IN PARENTHESES

	Total	Diff.	Homog.	Overlap.	Non-Overlap
intercept	-7.037***	-8.869***	-7.903***	5.090***	5.343***
	(0.493)	(0.547)	(0.551)	(0.394)	(0.414)
distw	-0.673***	-0.575***	-0.570***	-0.333***	-0.272***
	(0.037)	(0.041)	(0.040)	(0.035)	(0.037)
gdpcap_o	0.343***	0.309***	0.353***	0.144***	0.009
	(0.015)	(0.027)	(0.017)	(0.016)	(0.015)
gdpcap_d	0.538***	0.606***	0.453***	0.306***	0.253***
	(0.014)	(0.016)	(0.016)	(0.012)	(0.014)
pop_o	0.714***	0.610***	0.783***	0.283***	0.133***
	(0.022)	(0.024)	(0.024)	(0.013)	(0.018)
pop_d	1.015***	1.146***	0.885***	0.268***	0.260***
	(0.023)	(0.026)	(0.026)	(0.012)	(0.015)
migrat	0.264***	0.238***	0.224***	0.396***	0.322***
	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)
Adj. R^2	0.499	0.501	0.464	0.352	0.372

TABLE V

DURBIN MODEL FOR TRADE - YEAR: 2000, STANDARD ERRORS IN PARENTHESES

	Total	Diff.	Homog.	Overlap.	Non-Overlap
intercept	-14.793***	-10.904***	-14.275***	- 10.260***	-8.672***
	(0.792)	(0.808)	(0.845)	(0.745)	(0.683)
distw	-0.364***	-0.375***	-0.373***	-0.098	-0.220***
	(0.061)	(0.071)	(0.074)	(0.087)	(0.080)
gdpcap_o	0.170***	0.071**	0.169***	0.315***	0.126***
	(0.030)	(0.027)	(0.027)	(0.027)	(0.024)
gdpcap_d	0.294***	0.278***	0.256***	0.531***	0.514***
	(0.026)	(0.023)	(0.024)	(0.026)	(0.025)
pop_o	0.713***	0.402***	0.675***	0.367***	0.208***
	(0.055)	(0.049)	(0.050)	(0.026)	(0.043)
pop_d	0.369***	0.326***	0.319***	0.795***	0.892***
	(0.025)	(0.023)	(0.023)	(0.047)	(0.024)
migrat	0.406***	0.391***	0.371***	0.690***	0.593***
	(0.023)	(0.021)	(0.022)	(0.025)	(0.024)
γ	0.009***	0.005*	0.002	0.003	0.008**
	(0.003)	(0.002)	(0.002)	(0.005)	(0.004)
ρ	-0.005	-0.006	-0.005	-0.003	-0.0001
LR-test	17.377	12.666	15.433	2.035	0.003
p-value	3E-05	0.0003	0.855	0.153	0.959

- the stock of migrants is highly significant both in 1970 and in 2000, for total trade as well as differentiated and homogeneous goods; however, using our classification of differentiated and homogeneous goods, we can see a more substantial effect of migration on the differentiated ones (see Table IV);

TABLE VI

DURBIN MODEL FOR TRADE - YEAR: 1970, STANDARD ERRORS IN PARENTHESES

	Total	Diff.	Homog.	Overlap.	Non-Overlap
intercept	-5.100***	-2.671***	-5.168***	0.711	-0.484
	(0.560)	(0.437)	(0.532)	(0.433)	(0.386)
distw	-0.129**	-0.105*	-0.135*	-0.159**	-0.190***
	(0.057)	(0.044)	(0.054)	(0.053)	(0.048)
gdpcap_o	0.113***	0.056***	0.095***	0.154***	0.113***
	(0.015)	(0.012)	(0.014)	(0.019)	(0.017)
gdpcap_d	0.171***	0.136***	0.155***	0.110***	0.115***
	(0.016)	(0.013)	(0.016)	(0.016)	(0.014)
pop_o	0.301***	0.111***	0.293***	0.232***	0.176***
	(0.024)	(0.019)	(0.024)	(0.019)	(0.018)
pop_d	0.217***	0.158***	0.195***	0.078**	0.180***
	(0.016)	(0.013)	(0.017)	(0.027)	(0.024)
migrat	0.389***	0.346***	0.352***	0.624***	0.520***
	(0.017)	(0.013)	(0.016)	(0.021)	(0.019)
lag.migrat	0.011***	0.003	0.010***	0.011	0.014***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
ρ	-0.011	-0.005	-0.009	-0.002	- 0.004
LR-test	54.684	7.253	33.671	1.387	9.574
p-value	1.4E-09	0.007	7E-05	0.239	0.002

- the lagged stock of migration (γ) also has a positive impact of trade. This means that, if country i trades with country j, and the country j has an high stock of migrants, the trade from i to a third country k will also increase;

- using the Rauch classification, this phenomenon is stronger for non-differentiated goods in 1970, while for differentiated goods in 2000. Using our classification, this phenomenon is still not strongly significative for differentiated goods (see Table V);

- ρ has negative sign for all the models used in Rauch classification. It means that, if a country i imports from country j, and the country j has a high level of trade, imports of i from a third country k will decrease (substitution effect). Using the Rauch classification, these coefficients are still negative, altough they are not significative;

- the impact of the stock of migrants increases (the coefficient goes from 0.322 to 0.396 in 2000) if we use the new classification. The results confirm our improvements on defining the set of differentiated goods.

All in all, the control for network interdependencies is always significant in our analysis. This means that the baseline OD model does not account for network effects, which play a relevant role in shaping the world trade web.

V. CONCLUSIONS

Sets of interdependent phenomena on a global scale are increasingly analyzed as complex networks. International trade, human mobility, communication and transportation infrastructures are just a few examples. Only

recently have new methodologies been developed to analyze the dynamics of intertwined networks, including cascading failures and the transmission of shocks across multiple and heterogeneous network structures. In this paper we contribute to this emerging field of research by analyzing the relationship between the network of human migration and the world trade web. To make different networks comparable, we apply a hypergeometric benchmark to identify significant cross-country relationships. We find that the presence of a stock of migrants in a foreign country has a positive impact on trade. This is particularly true for a subset of products that we identify in Section III. However, when we move to test a gravity model of trade controlling for network interdependencies, we do not find any significant difference between the trade of differentiated and homogeneous goods. Since multiple communities of migrants are simultaneously present in developed countries, when we move to consider competition among them and network substitution effects we do not find any particular information advantage for differentiated goods. However, our results in this paper must be considered as preliminary. Spatial econometrics techniques have a huge potential to include network effects in traditional gravity models. In future work we plan to accommodate multiple and different network effects in this setting. Moreover, since the traditional distinction between differentiated and homogeneous goods does not seem to be particularly informative, we plan to investigate the changing role that migrants play in favoring the trade of specific commodities.

REFERENCES

[1] D. M. Gould, "Immigrant links to the home country: empirical implications for us bilateral trade flows," *The Review of Economics and Statistics*, pp. 302–316, 1994.

[2] K. Head and J. Ries, "Immigration and trade creation: econometric evidence from Canada," *Canadian journal of economics*, pp. 47–62, 1998.

[3] G. Peri and F. Requena-Silvente, "The trade creation effect of immigrants: evidence from the remarkable case of Spain," *Canadian Journal of Economics/Revue canadienne d'économique*, vol. 43, no. 4, pp. 1433–1459, 2010.

[4] M. Gen, M. Gheasi, P. Nijkamp, and J. Poot, "The impact of immigration on international trade: a meta-analysis," 2011.

[5] J. E. Rauch and V. Trindade, "Ethnic Chinese networks in international trade," *Review of Economics and Statistics*, vol. 84, no. 1, pp. 116–130, 2002.

[6] P. B. Slater, "Hubs and clusters in the evolving us internal migration network," *arXiv preprint arXiv:0809.2768*, 2008.

[7] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012.

[8] G. Fagiolo and M. Mastrorillo, "International migration network: Topology and modeling," *Phys. Rev. E*, vol. 88, p. 012812, Jul 2013.

[9] ——, "Migration and trade: A complex-network approach," 2013, arXiv preprint arXiv:1309.5859.

[10] Ç. Özden, C. R. Parsons, M. Schiff, and T. L. Walmsley, "Where on earth is everybody? The evolution of global bilateral migration 1960–2000," *The World Bank Economic Review*, vol. 25, no. 1, pp. 12–56, 2011.

[11] R. C. Feenstra, R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo, "World trade flows: 1962-2000," National Bureau of Economic Research, Working Paper 11040, January 2005. [Online]. Available: http://www.nber.org/papers/w11040

[12] M. Riccaboni, A. Rossi, and S. Schiavo, "Global networks of trade and bits," *Journal of Economic Interaction and Coordination*, pp. 1–24, 2013.

[13] M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, and R. N. Mantegna, "Statistically validated networks in bipartite complex systems," *PloS one*, vol. 6, no. 3, p. e17994, 2011.

[14] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture," *Nature genetics*, vol. 22, no. 3, pp. 281–285, 1999.

[15] S. Wuchty, A.-L. Barabási, and M. Ferdig, "Stable evolutionary signal in a yeast protein interaction network," *BMC evolutionary biology*, vol. 6, no. 1, p. 8, 2006.

[16] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.

[17] M. A. Serrano and M. Boguná, "Topology of the world trade web," *Physical Review E*, vol. 68, no. 1, p. 015101, 2003.

[18] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Physical review letters*, vol. 87, no. 25, p. 258701, 2001.

[19] P. Jaccard, "Distribution de la flore alpine dans le bassin des drouces et dans quelques regions voisines," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, no. 140, pp. 241–272, 1901.

[20] M.-L. Lee and R. K. Pace, "Spatial distribution of retail sales," *The Journal of Real Estate Finance and Economics*, vol. 31, no. 1, pp. 53–69, 2005.

[21] J. E. Anderson, "A theoretical foundation for the gravity equation," *The American Economic Review*, vol. 69, no. 1, pp. 106–116, 1979.

[22] E. Van Wincoop and J. E. Anderson, "Gravity with gravitas: a solution to the border puzzle," *American Economic Review*, vol. 93, no. 1, pp. 170–192, 2003.

[23] J. Frankel and A. Rose, "An estimate of the effect of currency unions on trade and growth," *NBER Working Paper*, vol. 7857, 2000.

[24] P. Egger, "A note on the proper econometric specification of the gravity equation," *Economics Letters*, vol. 66, no. 1, pp. 25–31, 2000.

[25] J. T. Bang, "Regional integration, spatial effects, and the gravity equation," 2006.

[26] H. Kelejian, G. S. Tavlas, and P. Petroulas, "In the neighborhood: The trade effects of the euro in a spatial framework," *Regional Science and Urban Economics*, vol. 42, no. 1, pp. 314–322, 2012.

[27] B. H. Baltagi, P. Egger, and M. Pfaffermayr, "Estimating models of complex fdi: Are there third-country effects?" *Journal of Econometrics*, vol. 140, no. 1, pp. 260–281, 2007.

[28] S. G. Hall and P. Petroulas, *Spatial interdependencies of FDI locations: A lessening of the tyranny of distance?* Bank of Greece, 2008, vol. 67.

[29] J. P. LeSage and R. K. Pace, "Spatial econometric modeling of origin-destination flows," *Journal of Regional Science*, vol. 48, no. 5, pp. 941–967, 2008.

[30] R. T. A. Leenders, "Modeling social influence through network autocorrelation: constructing the weight matrix," *Social Networks*, vol. 24, no. 1, pp. 21–47, 2002.

[31] S. Farber, A. Páez, and E. Volz, "Topology, dependency tests and estimation bias in network autoregressive models," in *Progress in Spatial Analysis*. Springer, 2010, pp. 29–57.

[32] C. F. Manski, "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, vol. 60, no. 3, pp. 531–542, 1993.

[33] L.-f. Lee, X. Liu, and X. Lin, "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, vol. 13, no. 2, pp. 145–176, 2010.

[34] Y. Bramoullé, H. Djebbari, and B. Fortin, "Identification of peer effects through social networks," *Journal of econometrics*, vol. 150, no. 1, pp. 41–55, 2009.

[35] A. G. Chandrasekhar and R. Lewis, "Econometrics of sampled networks," *Unpublished manuscript, MIT.[422]*, 2011.

[36] L. Anselin, *Spatial econometrics: methods and models*. Springer, 1988, vol. 4.

[37] R. K. Pace and J. P. LeSage, "Chebyshev approximation of log-determinants of spatial weight matrices," *Computational Statistics & Data Analysis*, vol. 45, no. 2, pp. 179–196, 2004.

[38] R. P. Barry and R. Kelley Pace, "Monte Carlo estimates of the log determinant of large sparse matrices," *Linear Algebra and its Applications*, vol. 289, no. 1, pp. 41–54, 1999.

[39] O. Smirnov and L. Anselin, "Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach," *Computational Statistics & Data Analysis*, vol. 35, no. 3, pp. 301–319, 2001.

[40] J. P. LeSage, M. M. Fischer, and T. Scherngell, "Knowledge spillovers across Europe: Evidence from a poisson spatial interaction model with spatial effects," *Papers in Regional Science*, vol. 86, no. 3, pp. 393–421, 2007.

[41] T. Mayer and S. Zignago, "Notes on cepiis distances measures: The geodist database," 2011.

[42] K. Head, T. Mayer, and J. Ries, "The erosion of colonial trade linkages after independence," *Journal of International Economics*, vol. 81, no. 1, pp. 1–14, 2010.

Institutions
Markets
Technologies

IMT

INSTITUTE FOR ADVANCED STUDIES LUCCA