Geographic Nature of Trade Specialization: Economic Distance in the Country Space

Manuel Flores
Departamento de Economía
Facultad de Ciencias Sociales
Universidad de la República
Email: manuel@decon.edu.uy
Address: Constituyente 1502, piso 6
Tel: +(598) 2410.64.49
Fax: +(598) 2410.64.50

Marcel Vaillant
Departamento de Economía
Facultad de Ciencias Sociales
Universidad de la República
Email: marcel@decon.edu.uy
Address: Constituyente 1502, piso 6
Tel: +(598) 2410.64.49
Fax: +(598) 2410.64.50

Abstract

International trade can be represented as a bipartite network connecting products with countries, where a link exists if a country exports a product with Revealed Comparative Advantage (RCA). Projecting this bipartite network on the products partition, Hausmann, Klinger, Barabási and Hidalgo (2007) developed the notion of Product Space. In this paper we present a new version of the Product Space using a highly disaggregated classification of products. In addition, two major extensions of this framework are proposed. In the first place, we consider the alternative projection of the bipartite network on the countries partition, building what we call the "Country Exports Space" (CXS). Second, we replicate the same procedure to the imports bipartite network and present the "Country Imports Space" (CMS). The bases of these two spaces are the notions of "specialization distance" in the CXS and "import pattern distance" in the CMS. Under the hypothesis that geography is associated with some major determinant of specialization, we propose a model where proximities in the trade specialization patterns are explained through geographical variables.

May, 2013

Keywords

Product Space, Trade Specialization, Sophistication, Gravity Models

JEL classification: F14, F63, O33, O57

I. INTRODUCTION

International trade theory focuses in explaining specialization patterns as a static phenomenon, while the dynamics of change from a simple trade pattern towards a more complex structure have received less attention. New growth theory has studied the endogenous determinants, but only in the strand of one-sector macroeconomic models, leaving pattern specialization issues ahead. Since Hirschman (1958), it is conventional knowledge in the Development Theory literature that trade patterns are linked to the basic fundamentals of the endogenous determinants of accumulation capacities (technology, infrastructure, human capital, etc.).

New methodologies permit the analysis of the link between specialization pattern and capacities accumulation. Production structure complexity is then measured through trade specialization pattern, since trade reveals the accumulation of capacities that endogenously determine growth[1]. The information about Revealed Comparative Advantage (RCA) is referred to a product in a particular country[2], and it can be used to define a measure of proximity between products. Proximity relationships permit to analyze structural change from different productive structures as well as the degree of sophistication of a particular product or country.

The aim of this paper is to develop an innovative application of these methodologies, through the definition of new measures of proximity between products and countries. With the proposed magnitudes, new representations of international trade relationships are obtained and applied to a new kind of description of trade networks, in what we call the "country spaces". Based on those spaces, we study the determinants of proximity between countries specialization patterns, using a set of network variables coupled with traditional proximity measures (in geographical, economic, political and cultural terms).

The new trade literature about Global Value Chains put its focus on the relevance and re signification of distance in trade specialization patterns. In particular Johnson and Noguera (2012) studied the determinants of the evolution of the ratio between value added and gross trade on a bilateral basis and use geographical distance as one of the explanatory variables. They show that proximity has a strong role in explaining

[1] Capacities could be interpreted in many different ways: technological knowledge accumulation (classical perspective); factor productivities (neoclassic perspective); market size (new trade models).

[2] We use the conventional Balassa (1965) RCA index.

fragmentation, and that there is a qualitative difference in goods exchanged by countries that are far away (with a higher ratio of value added to exports total value) with respect to those that are close each other (with less value added).

The paper is organized in this introduction and three more sections. The next section presents the methodological framework for the international trade network and describes general results. The third section explores the determinants of trade specialization proximities between pairs of countries. The fourth and final section concludes.

II. INTERNATIONAL TRADE NETWORK

II.1. Methodological Framework

International trade can be represented as a bipartite network connecting products with countries, where a link exists if a country exports a product with RCA. Projecting this bipartite network on the products partition, Hausmann, Klinger, Barabási and Hidalgo (2007, HKBH henceforth) developed the notion of Product Space, based on the notion that two products *(i, j)* are close one another if there is a high probability of a country having RCA in *i* given it has RCA in *j*, or *vice versa*[3]. More precisely, proximities in the product spaces are obtained as shown in Equation 1.

$$\phi_{ij} = min\{P(VCR_i|VCR_j), P(VCR_j|VCR_i)\} \qquad (1)$$

Based on this measure, in this paper we present a new version of the Product Space using a highly disaggregated classification of products[4].

A projection of the bipartite network on the countries partition allows building what we call the “Country Exports Space” (CXS), a graph based on the distances among countries’ export baskets, or “specialization distances”. Hence, two countries (*A*, *B*) are close one another if there is a high probability of a product being exported with RCA by *A* given *B* has RCA in it, or *vice versa.*

[3] For a complete description of the trade network see Appendix A.

[4] Harmonized System (HS-2002), at six-digits level (4956 products).

$$\varphi_{AB} = min\{P(VCR_A|VCR_B), P(VCR_B|VCR_A)\} \quad (2)$$

We replicate the same procedure to the import bipartite network, in order to build what we call the “Country Imports Space” (CMS), based on the notion of “import pattern distance”. The equations are completely analogous to those presented above for the export spaces, replacing exports with imports variables. This means that another bipartite network is now projected on the products and countries partitions. We build this new network turning to a measure analogous to the RCA, obtained as the relation between the weight of each good in a countries’ imports and its weight in world imports (and again, an indicator variable signals the products in which each country has a value greater than unity).

Country and Product spaces can be viewed as the two sides of the same coin; since country space distances are a measure of how close two countries’ projections on the product space are, while product space distances show how two products’ projections resemble on the country space. We focus here on the country spaces, but references to the product spaces are unavoidable.

II.2. Network Structure of Products and Countries Trade Spaces

The considered products and countries spaces of exports and imports can be represented turning to the adequate measures of distance. Their topological descriptions and a cluster-structure analysis for the country spaces serve as an innovative input to the assessment on trading blocs.

Distance calculations were conducted using BACI International Trade Database (CEPII, 2010), with bilateral trade consistency-corrected COMTRADE data for almost 5.000 products (Harmonized System subheadings) and 150 countries[5]. In order to obtain more robust results, we use the mean trade values for the four-year period 2004-2007.

The Minimum Spanning Tree (MST) algorithm is used to define each network structure, in order to keep the strongest proximities and guaranteeing, at the same time, that every node is somehow connected to the graph. Following HKBH, we obtain the main

[5] COMTRADE database is made up by United Nations Statistics Division. CEPII database reconciles data reported by almost 150 countries, and by means of estimation, removes from imports the approximated value of transport and insurance rates.

skeleton of the network using MST and proceed to add the strongest links left aside by this algorithm (in a number equal to the number of nodes in the network).

In this section we present the country spaces (exports and imports). Some basic topographic information of the associated product spaces is reported in the Appendix B. The Product Space of Exports is akin to HKBH's, although new smaller modules start to appear when a higher disaggregation of products is used. The Product Space of Imports is relatively branched out and has a clear cut division between nucleus and periphery, with an extremely dense big core of products that almost every country buys, a basket of international essential products.

By construction both country spaces have the same structure of 167 nodes connected by 333 edges (without self-loops or multi-edge node pairs). They consequently share a density of 0.024 and an average number of neighbors equal to four. The MST Algorithm guarantees that there are no isolated nodes (there is only one connected component). A comparison of the topography descriptive measures of both networks is presented in Table 1.

Table 1

Topographic Measures for the CXS and CMS

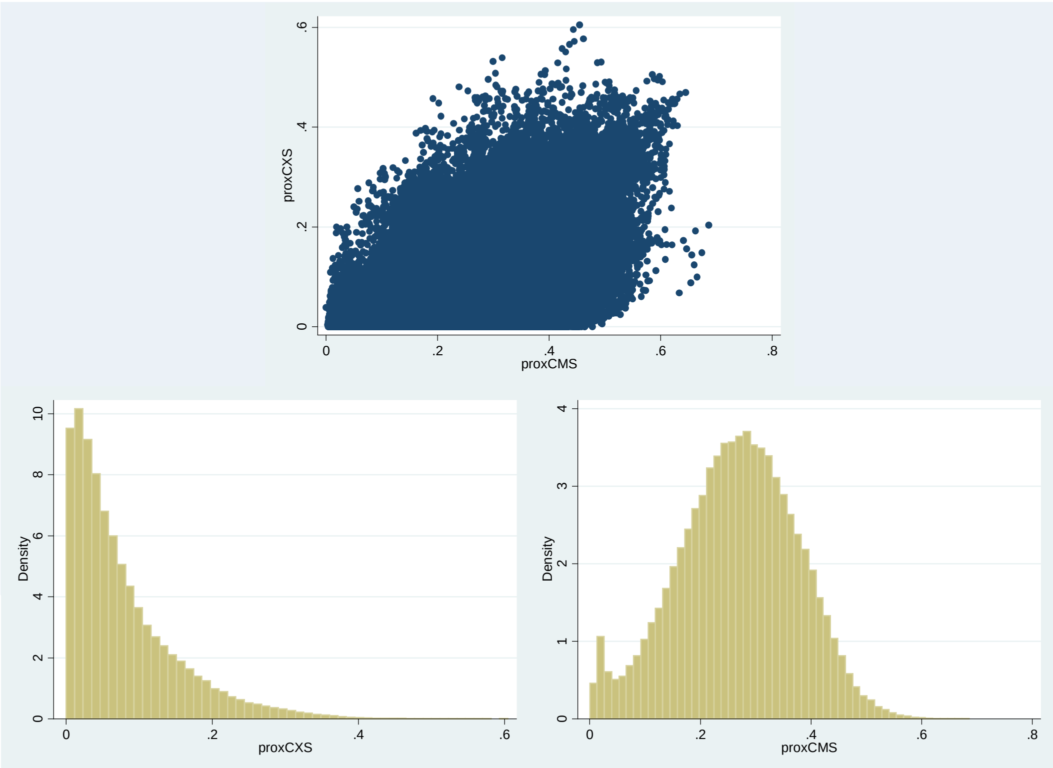
	CXS	CMS
Number of Nodes	167	167
Number of Edges	333	333
Average Number of Neighbors	3.98	3.98
Network Density Measure of cohesion: Number of links as a proportion of the possible connections	0.024	0.024
Network Diameter Largest distance between two nodes	26	17
Network Radius Minimum among the non-zero eccentricities (maximum non-infinite length of a shortest path between nodes)	13	9
Characteristic path length Expected distance between two nodes	8.31	6.176
Network Centralization Measure of neighborhood. A star structure has an index of 1 and decentralized networks have lower values.	0.092	0.085
Clustering Coefficient Average of the clustering coefficient of nodes (ratio of the number of edges between a nodes' neighbors and the number of possible connections among them)	0.175	0.241
Network Heterogeneity Measure of the tendency to contain hub nodes	0.974	0.858

Source: Own elaboration using BACI database and Cytoscape software

The edges of the networks can be treated as weighted using the same proximities used to their construction. The joint and marginal distributions of proximities are illustrated in Figure 1, showing that there is a weak association between them (correlation coefficient of 0.49) and that their marginal distributions are different. Proximities in the CMS have a more symmetric distribution, while proximities in the CXS have a clear right asymmetry[6]. In particular, the high concentration of zero-proximities in the CXS is an issue that will be considered in the next section. These cases represent a 1.2% of the dyads, and mean that those countries have no common product in the baskets of products they export with RCA.

Figure 1

Distribution of proximities in the CXS and the CMS



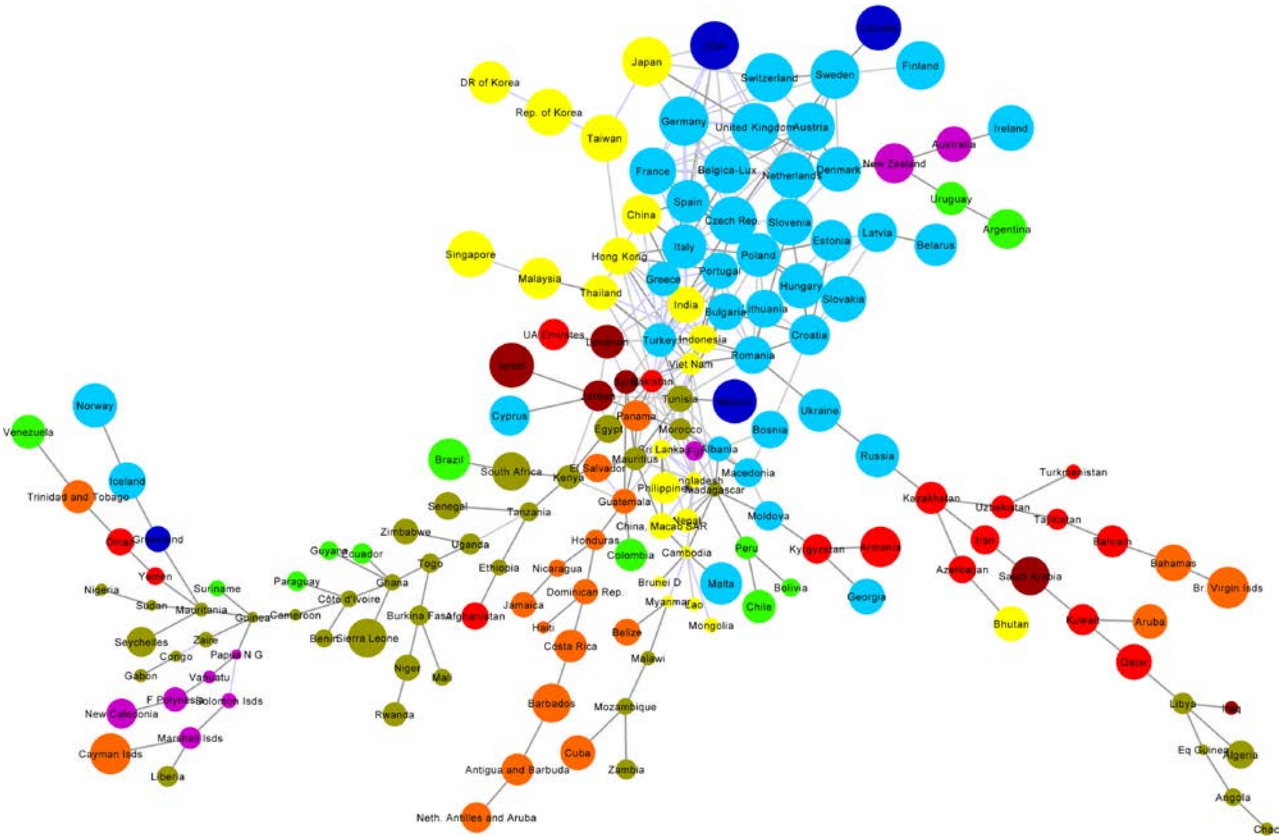
Source: Own elaboration using BACI database.

[6] Proximities in the CXS measure the probability that a country *j* has RCA in a product given that the country *i* has RCA in it. Proximities in the CMS measure the same phenomenon, but instead of considering RCA it takes into account the fact that each country imports a product in a proportion greater than the global average. Proximities in the CXS have a mean value of 0.08 and a standard deviation of 0.07. Proximities in the CMS have a mean value of 0.27 and a standard deviation of 0.11.

Spring-Embedded edge-weighted algorithms are used to set up the basic network layouts. After that, many handmade adjustments have to be done in order to obtain a more eloquent picture of the networks.[7]

Figure 2

The Country Exports Space



Source: Own elaboration using BACI database and Cytoscape software.

The proposed representation of the Country Exports Space (Figure 2) shows that the network has a highly connected nucleus conformed mainly by European countries, very similar as a group in what products they export. The United States are also part to the nucleus, while Canada is directly connected through Sweden. Japan and Hong Kong are the main hubs through which other Asian countries connect to the nucleus. New Zealand is also directly connected, and makes the way for Australia, Uruguay, Argentina and Ireland to approach this pattern of exports. Countries like Turkey, Vietnam, Pakistan, Tunisia, and Morocco form a densely connected region too, placed

[7] The usual criterions of minimizing the number of edge crossings and avoiding overlaps of nodes are applied for handmade adjustments.

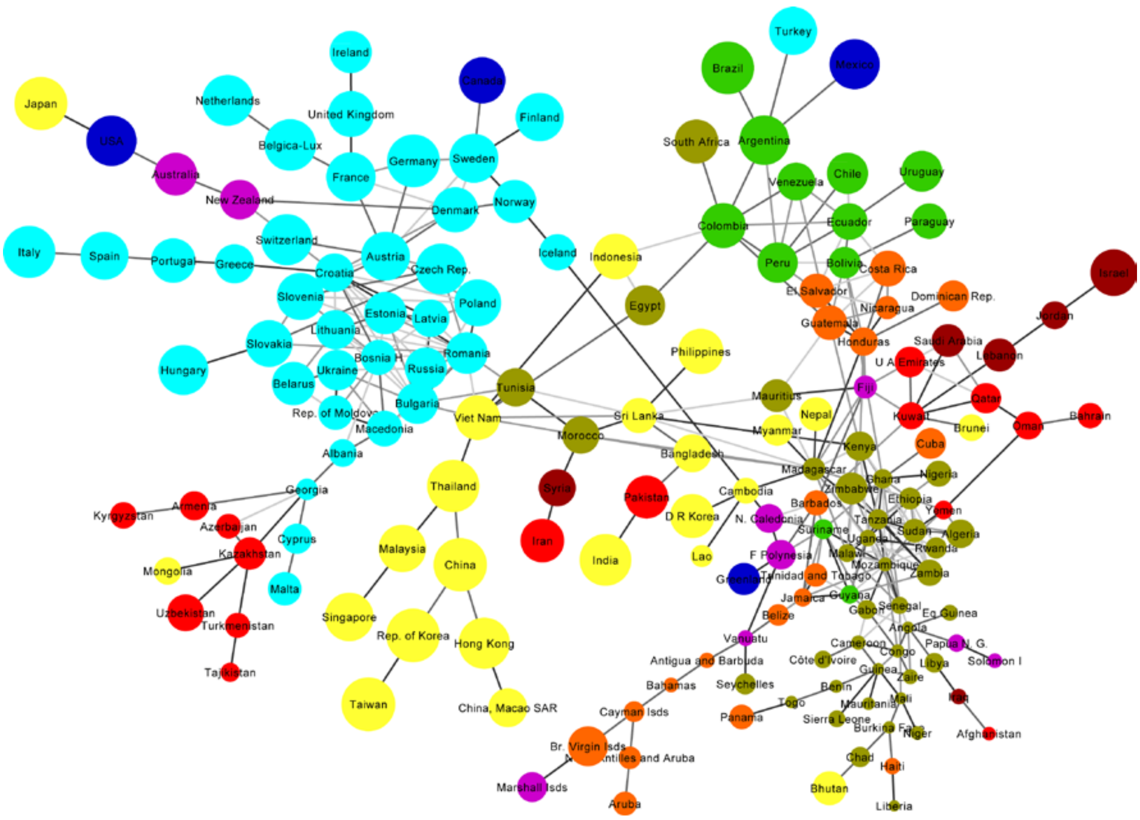
between the nucleus and two of its three branches. The third one is directly linked with the nucleus by Romania, and shows that Russia and Ukraine productive structures are in an intermediate place connecting most of Middle East countries and some North-African and Caribbean countries to the nucleus.

The importance of regional patterns and geographical proximities is the main conclusion that arises from the depicted landscape (note that colors are assigned according to geographical regions), where Europe, North America, and Asia pertain to a connected community, while the rest developing countries are in the periphery of the space.

One can define the “extremely similar” set of countries as those where proximity is greater than 0.40 (49 cases in our sample). The analysis of the results for the “extremely similar” dyads shows that four groups of countries exist. China neighborhood is integrated by many Asian and European countries (Hong Kong, India, Italy, Turkey, Portugal, Spain, Pakistan and Greece). The European group includes a close bilateral proximity between Germany and United States, and also the United States with France and Great Britain (and also includes closeness of Spain, Italy, Czech Republic, Austria, and Polony). The third group has a diverse composition of countries of lower level of development, mainly from the South Asian region (Sri Lanka, Bangladesh, Madagascar, Macao, Cambodia, Myanmar, Laos, Viet Nam, Pakistan, Tunisia, Indonesia, Thailand, and Mauritius). The last group is a selection of bilateral relations among geographical neighbor countries that also have close proximity in trade export specialization patterns (Netherlands-Benelux, Romania-Bulgaria, Latvia-Lithuania-Estonia, North Korea-South Korea, and El Salvador-Guatemala).

The CMS reveals that countries gather around many nucleuses. In addition to the developed countries’ highly connected group, strong similarities appear to exist among many African and Latin American countries. Asian countries are placed now in the middle, connecting the main two differentiated patterns of imports. Middle East countries are now divided in different regions of the network, while Caribbean countries form their own branch. Again, a regional pattern is evident.

Figure 3
The Country Imports Space



Source: Own elaboration using BACI database and Cytoscape software.

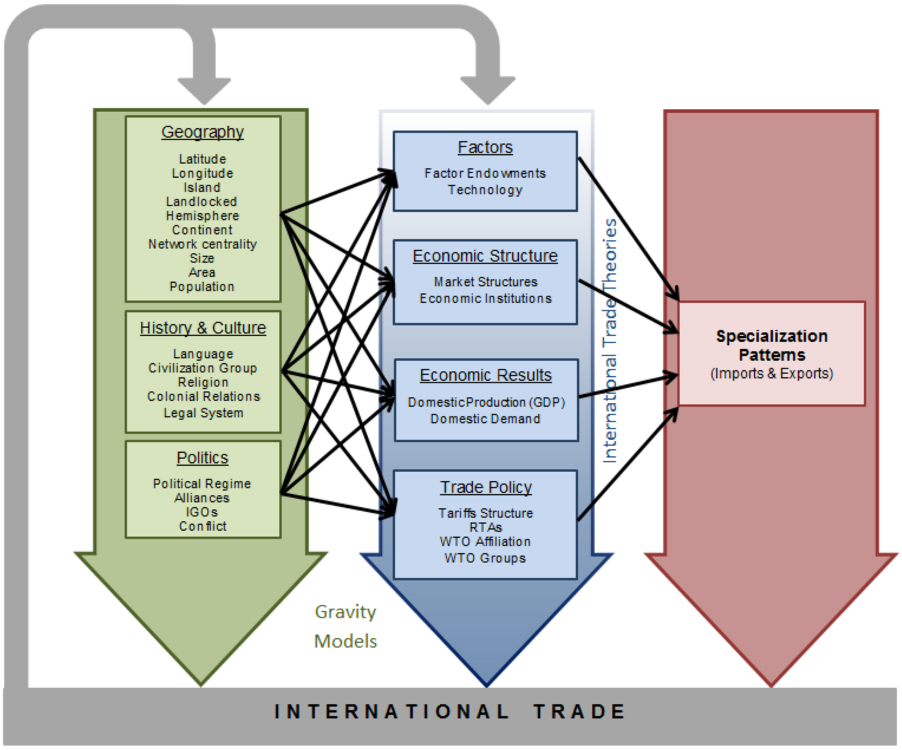
Both country spaces suggest that geography is a key determinant of the baskets of imports and exports of each country. One could argue that geography effects are not direct, but channeled through regional differences in the traditional factors-based determinants of trade. That being true, it suggests an additional reason to focus on the role of distances on specialization and thereby on trade. The next section disentangles the distinct components of geographic distance, and assesses their effects on trade specialization differences (controlling for other historical, cultural, political and economic divides).

III. DETERMINANTS OF TRADE SPECIALIZATION DISTANCES

Figure 3 shows three different levels of analysis in the explanation of international trade, going from the ultimate geographic determinants (green arrow), to the most immediate reason to trade: the pattern of specialization of each country (red arrow). An intermediate level of explanation includes the widely known motives considered by the alternative international trade theories (blue arrow), strongly associated with the exogenous or predetermined geographic, historical, cultural and even political characteristics of countries or regions. Many papers in economic geography, economic history, international relations, political economy, economic growth, development, and international economics could be invoked in a review of the role of these exogenous or predetermined factors in the explanation of the main drivers of a country's performance in the international economy.

Figure 3

Specialization and International Trade Determinants



Source: Own elaboration.

Trade Theories are theories of specialization, and so the effects on trade of primary factors and some economic structure characteristics should be expected to be indirect (shaping the specialization pattern of the economy). On the other hand, the empirical literature has taken advantage of the explicative power of exogenous and predetermined data (including also some variables of economic results and trade policy) to predict bilateral trade by means of gravity models.

In addition, there is an endogeneity problem if medium term causality is to be considered, since trade patterns and trade performance are in turn an explanation for many of the characteristics of the economy.

We propose to analyze the role of the fundamental determinants' differences between countries (in geographic, historical, cultural and political characteristics) on the distance of their specialization patterns and import baskets. As happens with trade, the factors that explain Country Space distances between countries i and j ($CSdist_{ij}$) are somehow related with differences in geographical distances, but the kind of effects that geographical distance gathers is an issue of ongoing debate (Johnson and Noguera, 2012). Traditionally interpreted as a proxy of transport costs, geographic distance can also be a proxy of differences in climate and other primary resources. In order to get a more precise idea of its effects on the distance in specialization patterns, we consider the usual distance between most populated cities of each country ($dist_{ij}$) and we add two geometry inspired regressors: the difference in time zones ($tdif_{ij}$, a proxy of distances in longitude) and the difference in absolute latitudes ($diflat_{ij}$, a proxy of climate differences).[8]

Two other geographic regressors complement the notion of geographic proximity: a contiguity variable reflecting the existence of a common border between each pair of countries ($border_{ij}$), and a common continent dummy variable reflecting the availability of earth transportation ($comcont_{ij}$). The inclusion of a common hemisphere dummy variable ($comhemis_{ij}$) also adds information on proximity (given the climatic bias given to our variable of difference in latitudes), and encloses the differences in economic structures between North and South (including the greater weight of trade in North-North relationships). Three last geographic variables control for characteristics that could be relevant in explaining geographic-based differences in the countries'

[8] Distances in latitudes are calculated as the absolute value of the difference between latitudes of capitals, without taking into account if the capital is located in the Northern or Southern hemisphere. Doing so, we obtain a variable that takes a zero value when two countries share the same parallel (with a parallel quarter precision), with disregard of the fact that they can be located in different hemispheres. Together, these two variables explain only a twenty percent of the variation in the linear distance, since the time zone difference is a loose indicator of the distance in meridians, the difference in absolute latitudes is nonlinear (linear over intervals), and the width of a meridian depends on the parallel (tending to zero at the poles).

production structures: a dummy variable that signals if only one of the two countries is an island (*o1island*$_{ij}$), another taking the value one if only one country is a landlocked territory (*o1landlock*$_{ij}$), and a variable that measures the difference in countries' areas (*difarea*$_{ij}$).

Consistent with the symmetric notions of proximity proposed, we estimate a symmetric model where all the variables are taken in a relational form, i.e. they are defined on the support of dyads.

The same basic specification of the model is used for CXS and CMS distances (Equation 3), and the mentioned control variables are always included. In order to ease coefficients interpretation, and to reduce some heteroskedasticity problems, all the models are log-linearly transformed.

$$CSdist_{ij} = a_0 + a_1 ldist_{ij} + a_2 ltdif_{ij} + a_3 ldiflat_{ij} + a_4 ldifarea_{ij} + a_5 o1island_{ij} + \\ + a_6 o1landlock_{ij} + a_7 border_{ij} + a_8 comcont_{ij} + a_9 comhemis_{ij} + \varepsilon_{ij} \qquad (3)$$

We expect a_1 to a_6 to be positive, while the remaining coefficients should show a negative effect of the respective proximity variables on the trade pattern distance.

Bilateral trade would be an interesting regresor, because geographical and cultural proximity could imply less trade costs and so more trade between countries. Increased trade would have a dynamic effect through the diffusion of technical progress; convergence in domestic capabilities and an associated increment in intra-industry trade when countries are geographically close to one another. Nevertheless, the whole international trade theory shows that the inclusion of bilateral trade value (*trd*$_{ij}$) in this model would lead to an inconsistent estimator due to endogeneity. An appropriate estimation would be possible if an exogenous source of variation were to be found, i.e. a variable that is empirically correlated with trade (once the influence of the exogenous regressors has been taken into account), and theoretically independent of the error term of the model (in our model the error term should be gathering the influence of primary factors and economic structure not explained by the included exogenous variables).

Finally, we also try the inclusion of other control variables in the models, like historical, cultural, institutional and economic relational characteristics. These variables are the difference in GDPs (*difgdp*$_{ij}$) and *per capita* GDP (*difgdppc*$_{ij}$), and a set of dummy variables signaling: if colonial relationship between *i* and *j* ever existed (*colony*$_{ij}$); if both have had a common colonizer after 1945 (*comcol*$_{ij}$); if they share a common official language (*comlang*$_{ij}$), a common majority religion (*comrel*$_{ij}$), a common legal system (*comleg*$_{ij}$), or a common currency (*comcur*$_{ij}$); if both or none are affiliated to WTO (*wtoboth*$_{ij}$, and *wtonone*$_{ij}$); if there is an ongoing conflict between them (*conflict*$_{ij}$); and the existence of a Regional Trade Agreement (RTA) where both are participants (*rta*$_{ij}$).

Needed information is obtained from the CEPII Gravity Database (Head, Mayer and Ries, 2010), together with average latitude and longitude for the 166 countries for which trade pattern distances are available.

The empirical assessment on the determinants of trade networks distances is relatively more complex in the case of CXS distances, since as shown before, its distribution is right censored[9]. This could lead to inconsistent estimations if the probability of being censored is correlated with the covariates included in the model.

A proper way to face this problem is to use a Heckman (1979) sample selection model, based on a selection equation that should incorporate at least one regresor not included in the proximity equation. Then, the “exclusion restriction” requires at least one variable explaining the fact that the countries of a dyad have any export product in common, and at the same time being not relevant in the explanation of their exports’ baskets differences. We propose that such a variable could be one that signals the dyads with zero trade, considering both theoretical and empirical arguments. While Trade Theories focus on how differences in specialization are associated with more trade, it is also true that more trade could have a convergence effect on industrial structures. The proposed selection variable could reflect the extreme cases in which this process has never started. In addition, a very strong association between null proximity in the CXS and zero trade is clearly observed in the data[10]. Since the use of one dummy variable for the exclusion restriction makes identification problematic, we add two other binary variables taken from our set of controls (*comlang*$_{ij}$ and *comcol*$_{ij}$). Being non-significant in the model, these variables proved to be significant in the selection equation.

---

[9] Distance is obtained as –ln(proximity$_{ij}$). With proximity having a discrete accumulation point in zero (meaning that in many cases the two countries of the dyad don’t have any common product in their export baskets), distance is infinite in these cases, that represent 1.25% of the sample.

[10] The percentage of the dyads that have zero trade is 24.1% for the whole sample and increases to 72.7% for the zero-proximity cases (period 1995-2007).

Four other strategies are widely followed in the gravity models literature to deal with zero-trade observations: OLS estimation with the whole sample once zero proximities are replaced with an arbitrarily small constant (only approximate and with no theoretical support), OLS estimation with non-censored observations (possibly biased estimators due to information loss), Tobit model estimation (also replacing zero proximity with a constant, but adequately considering the particular distribution of the dependent variable), or Poisson counting model Maximum Likelihood estimator (equivalent to a weighted non-linear least squares estimator). Although none of these methods deals adequately with sample selection, in order to assess the robustness of the results all the alternatives were conducted and are reported in Appendix C.[11]

Estimations of the models for distances in both spaces are presented in Table 2. The CXS distance equation is estimated with the two-steps Heckman Selection Model, and CMS distance equation is estimated by OLS. In both cases the sample includes 166 countries (13695 dyads) and covers the period 1995-2007. For each equation two alternative estimations are reported, using Country Fixed Effects (CFE) and using Time Varying Country Fixed Effects (TVCFE). Panel Data techniques would require the inclusion of Country-Pair Fixed Effects, but this alternative leads to the loss of almost all the covariates, since they are time-invariant dyad-supported.

The results show that estimations are also robust to the different specifications of the fixed effects, at least for the relevant variables. All the significant coefficients have the expected signs, except for negative effect of time zone differences in both equations, and their magnitudes are larger in the CMS equation. The goodness of fit is remarkable good in the case of the CMS model, reaching a 0.64 determination coefficient, and is considerably high for the CXS model (about 0.45, according to OLS estimation, see Appendix C).

[11] The OLS estimation with the whole sample requires imputing a value to the infinite distance obtained in zero proximity cases. A value of 0.00001 was assigned, then the distribution of distance is censored with an accumulation point in –ln(0.00001). Tobit and Poisson results are the estimated partial effects at the mean of variables. Presented standard errors are robust, clustered by dyad.

Table 2

Geographic Determinants of Countries' Specialization Patterns

	Distances in CXS		Distances in CMS	
	Heckman CFE	**Heckman TVCFE**	**OLS CFE**	**OLS TVCFE**
ldist	0.20***	0.21***	0.10***	0.05***
	[0.012]	[0.019]	[0.006]	[0.004]
ldiflat	0.04***	0.05***	0.00***	0.01***
	[0.003]	[0.005]	[0.001]	[0.001]
ltdif	-0.02***	-0.01***	-0.00**	-0.00**
	[0.002]	[0.004]	[0.001]	[0.001]
ldifarea	0.03***	0.01**	0.03***	0.01***
	[0.003]	[0.006]	[0.002]	[0.002]
o1island	0.08***	0.09***	0.04***	0.02***
	[0.014]	[0.025]	[0.008]	[0.006]
o1landlock	0.11***	0.10***	0.00	0.00
	[0.015]	[0.027]	[0.007]	[0.004]
border	-0.22***	-0.20***	-0.02	-0.03**
	[0.043]	[0.066]	[0.019]	[0.014]
comcont	-0.09***	-0.12***	0.03***	-0.04***
	[0.020]	[0.031]	[0.010]	[0.006]
comhemis	-0.07***	-0.06**	0.09***	0.01*
	[0.016]	[0.028]	[0.007]	[0.005]
ldifgdp	0.15***	0.16***	0.12***	0.09***
	[0.004]	[0.006]	[0.002]	[0.002]
ldifgdppc	0.07***	0.07***	0.05***	0.04***
	[0.004]	[0.007]	[0.002]	[0.001]
colony	-0.25***	-0.04	-0.11***	-0.05***
	[0.050]	[0.079]	[0.023]	[0.016]
comcol			-0.03***	-0.02***
			[0.011]	[0.007]
comlang			0.05***	-0.01
			[0.012]	[0.007]
comrel	0.01	-0.03	0.00	-0.01***
	[0.013]	[0.021]	[0.007]	[0.004]
comleg	0.00	-0.03*	-0.04***	-0.01***
	[0.012]	[0.019]	[0.007]	[0.004]
comcur	-0.04	-0.27***	0.03	-0.09***
	[0.063]	[0.098]	[0.022]	[0.022]
wto_both	-0.42***	-2.37	-0.20***	0.34
	[0.016]	[2.162]	[0.008]	[492.180]
wto_none	0.27***	2.30	0.11***	-0.45
	[0.023]	[2.163]	[0.014]	[.]
rta	-0.31***	-0.14***	-0.05***	-0.08***
	[0.027]	[0.043]	[0.011]	[0.007]
Constant	3.17***	2.17***	0.67***	1.84
	[0.148]	[0.408]	[0.077]	[224.446]
Observations	178,022	178,022	178,022	178,022
R-squared			0.639	0.899

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Differences in specialization patterns are greater when geographic distance is large, when climate difference is more pronounced, when countries' sizes asymmetry is marked, when only one of them is an island, or when GDPs or *per capita* GDPs are very distinct. The fact that countries are landlocked makes they export more similar products, but it has no effect on the dissimilarity of their imports. Contiguity is an important predictor of similarities in exports, but presents mixed results in the CMS equation, and the same conclusion is valid for the variable that captures when countries belong to the same continent. On the other hand, being in the same hemisphere makes the export basket resemble, while it is associated with increased differences in imports. Finally, colonial relations have a negative coefficient on import distances, but besides a significant negative point estimate in one of the models they have no effects on export differences.

In sum, these results are in concordance with common sense expectations, giving support to the use of export and import pattern distances in other lines of research. In addition, they illustrate how a set of different geographic factors contributes to explain most of the variation in specialization patterns.

V. CONCLUSIONS

Further developing the idea of product space, presented by HKBH (2007), we have shown that the main characteristics of the product space remain unchanged when a higher level of detail on products is considered, while less aggregation permits to find new densely connected groups of products. We also extended their proposal in two directions. First, the notion of Country Space leads to a suggestive representation of nation's specialization patterns, where the core of the space includes almost all the European and some other highly developed countries (some other clubs were also found and detailed). Second, the adaptation to import trade flows of both product and country spaces complements the picture of how countries resemble in the kind of goods they sell and they buy abroad. Again, the Country Imports Space suggests that countries gather around definite patterns of imports.

Both country spaces posit that trade patterns are strongly associated with geographic variables. Estimating two separate models, one for distances in export and the other for distances in imports, we have shown that the structure of production (proximity in exports and imports) is heavily related with geographical factors. Even after controlling for many different kinds of geographic distance, and besides a large number of control variables taking into account other sorts of distance, like historical, cultural, political, institutional or economic differences between countries, linear geographic distance stills being a very important factor.

Our results could be useful to understand the role of geographical variables in gravity models, being useful to reanalyze the role of distances as a determinant of bilateral trade. The constructed variables of distance in the CXS and CMS could also be useful in several lines of research that analyze dyad data

REFERENCES

Anderson, J. E. and van Wincoop, E., 2003. "Gravity with Gravitas: A Solution to the Border Puzzle". *American Economic Review*, American Economic Association, vol. 93(1), pages 170-192, March.

Anderson, K. and Norheim, H., 1993. “From imperial to regional trade preferences: Its effect on Europe’s intra- and extra-regional trade”, *Weltwirtschaftliches Archiv*, vol. 129, n. 1, pp. 78- 102.

Balassa, B., 1965. “Trade Liberalization and ‘Revealed’ Comparative Advantage”, The Manchester School of Economic and Social Studies, vol. 33, n. 2, pp. 99-123.

Hausmann, R., Hwang J., and Rodrik, D., 2007. “What you exports matters”. *Journal of Economic Growth*, 12(1), pp, 1-25.

Hausmann, R. and Hidalgo, C., 2009. “The Building Blocks of Economic Complexity”. PNAS, 106(26), pp 10575-10575.

Hidalgo, C.A., Klinger, B., Barabási, A.L., and Hausmann, R., 2007. “The product space conditions the development of nations”. *Science*, 317(5837), pp, 482-487.

Hirschman, A., 1958. *The strategy of economic development*. New Haven, Conn.: Yale Press.

Krugman, P., 1988. “La nueva teoría del comercio internacional y los países menos desarrollados”. *El trimestre económico*, enero-marzo, LV (217).

Johnson R. and Noguera G., 2012. “Proximity and production fragmentation”. American Economic Review, Paper&Proceedings, 102 (3):407-411.

Leamer, E. 1984. *Sources of Comparative Advantage: Theory and Evidence*. Cambridge MA: The MIT Press.

Pavitt, K., 1984. “Sectoral patterns of technical change: towards a taxonomy and a theory”. *Research Policy* 13, 343-73, North Holland.

Van Dongen, S., 2000. *Graph Clustering by Flow Simulation*. PhD Thesis, University of Utrecht, The Netherlands.

APPENDIX A

International Trade Networks

A.1 Trade bipartite network

The trade bipartite network ($\mathcal{T}$) is defined by two groups of nodes: products ($\mathcal{P}$) and countries ($\mathcal{C}$). There are edges ($\mathcal{E}$) that connect countries with products. The edges are defined taking into account a trade specialization index[12].

$$\mathcal{T} = (\mathcal{P}, \mathcal{C}, \mathcal{E}) \tag{A.1}$$

The network is summarized in a matrix (***T***), with products as rows and countries as columns. A binary entry signals with 1 if there is an edge and 0 if it is not.

Figure A.1

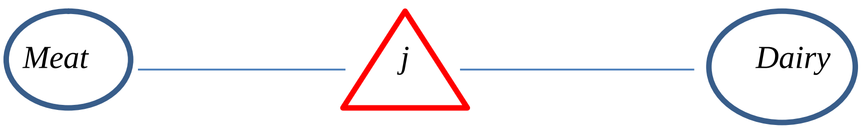
Matrix *T*

$$\begin{array}{c} \\ 1 \\ \vdots \\ i \\ \vdots \\ p \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & \dots & j & \dots & c \end{array} \\ \left[\begin{array}{ccccc} 0 & & & & 1 \\ & & & & \\ & & 1 & & \\ & & & & \\ 0 & & & & 0 \end{array}\right] \end{array}$$

The matrix ***T*** summarizes the information of exports (X) or imports (M). Taking the export space as an example, to find the proximity between any two pair of products it is necessary to answer the following question: What's the probability that a country exports product i with RCA given that he exports with RCA product j?

Figure A.2

Edges of order two between two products



In this case two products (e.g. meat and dairy) are connected through a countries' partition node (see Figure A.2). We obtain the edges of order two by summing over a country's index all the countries that have RCA in meat and dairy products

[12] We use a traditional Revealed Comparative Advantage Index a la Balassa.

simultaneously. We repeat the operation for every pair of products. It is easy to show that all edges of degree two are summarized in the following matrix $\boldsymbol{P}$.

$$\boldsymbol{P}_{p\times p} = \boldsymbol{TT}' \quad \text{(A.2)}$$

Proximity, as defined in Equation 1, comes from a normalization of the entries of the non-symmetric matrix $\boldsymbol{P}$ by the total number of countries that export each product with RCA ($diag(\boldsymbol{X}\boldsymbol{i}_c)^{-1}$, where $\boldsymbol{i}_c$ is a vector of ones with dimension equal to the number of countries). Then, a non-symmetric square matrix $\tilde{\phi}$ is obtained, and a symmetric matrix of proximities ($\boldsymbol{\phi}$) results from taking the minimum value found in its symmetric positions.

$$\widetilde{\boldsymbol{\phi}} = diag(\boldsymbol{X}\boldsymbol{i}_c)^{-1}\boldsymbol{P} \quad \text{(A.3)}$$

$$\boldsymbol{\phi} = min\{\widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{\phi}}'\} \quad \text{(A.4)}$$

Hence, each element of the proximity matrix is of the form presented in Equation 1.

The distance matrix between any pair of products is defined as a log transformation of the proximity matrix:

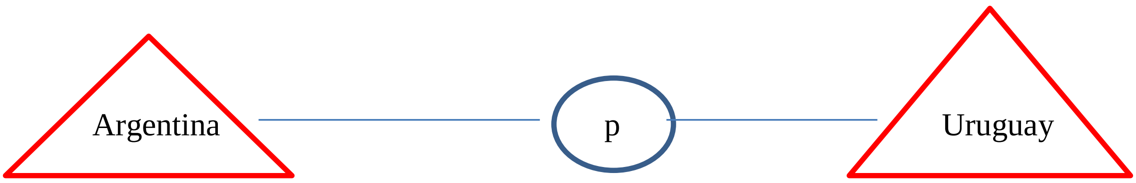
$$distPS_{ij} = -Ln(\phi_{ij}) \quad \text{(A.5)}$$

With this information it is possible to make a graph representation of the product projection of the bipartite network[13].

Another projection of the trade bipartite network could be made in the country space. With this purpose the proximity between countries must be measured answering the following question: what is the probability of a product being exported by country A with RCA given that it is exported with RCA by the country B?

Figure A.3

Edges of order two between two countries



[13] Minimum Spanning Tree (MST) algorithm, and network representation algorithms. MST is an algorithm that chooses C-1 connections (C numbers of nodes) which, while minimizing distance, guarantees the connection of every node to the net. Then the more strong C-1 edges are added. So we obtain a reduced net of 334 edges. Graph visualization Force Direct algorithm (Kamada-Kawai). Cytoscape software.

In the case of two countries (e.g. Argentina and Uruguay) they are connected through a products' partition. We obtain the edges of order two summing over product's index all the products for which both countries have RCA simultaneously. We repeat the operation for every pair of countries. It is easy to show that all edges of degree two are summarized in the following matrix.

$$\boldsymbol{C}_{\boldsymbol{c}\times\boldsymbol{c}} = \boldsymbol{T}'\boldsymbol{T} \tag{A.6}$$

The measure of proximity is obtained normalizing by the number of products that each country exports with RCA ($diag(\boldsymbol{X}'\boldsymbol{i_p})^{-1}$, where $\boldsymbol{i_p}$ is a vector of ones with dimension equal to the number of products). Then:

$$\widetilde{\boldsymbol{\varphi}} = diag(\boldsymbol{X}'\boldsymbol{i_p})^{-1}\boldsymbol{C} \tag{A.7}$$

And a symmetric proximity matrix is defined:

$$\boldsymbol{\varphi} = min\{\widetilde{\boldsymbol{\varphi}}, \widetilde{\boldsymbol{\varphi}}'\} \tag{A.8}$$

Each element of the proximity matrix is of the form defined in Equation 2.

The distance matrix between any pair of countries is defined as a log transformation of the proximity matrix:

$$distCS_{ij} = -Ln(\varphi_{ij}) \tag{A.10}$$

APPENDIX B

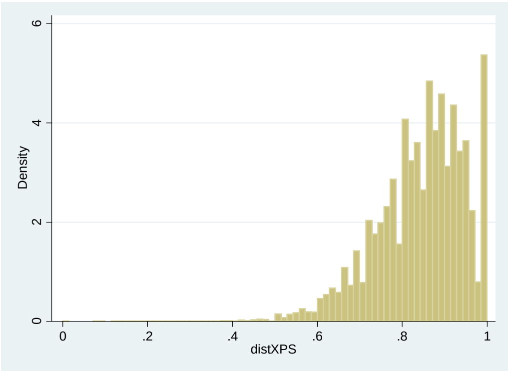
Product space projections

Both import and export product networks have 4955 nodes and 9910 edges, with a density of 0.001, and an average number of neighbors of 4

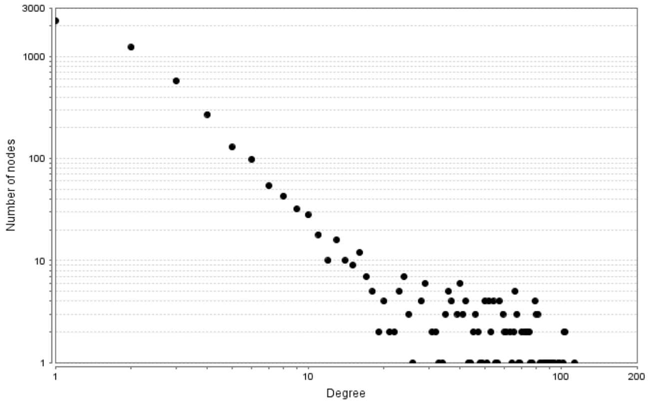
B.1 Exports

This network has a diameter of 57 nodes and a clustering coefficient of 0.058.

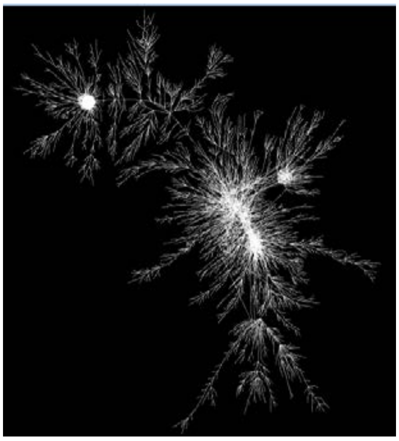
Distance in XPS Distribution



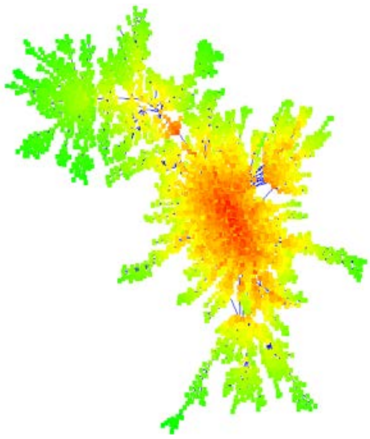
Degree Distribution



Links



Closeness Centrality

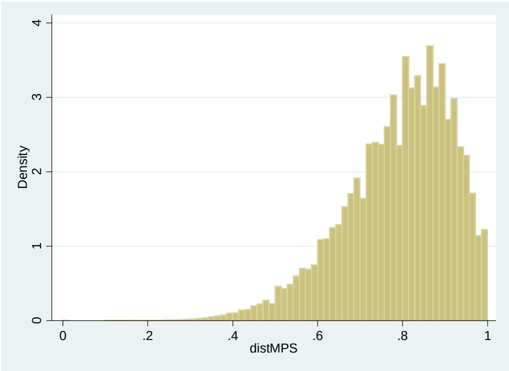


Sophistication

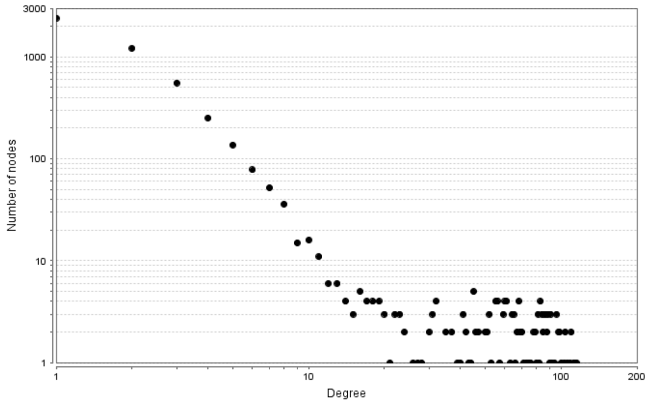
B.2 Imports

This network has a diameter of 72 nodes and a clustering coefficient of 0.032.

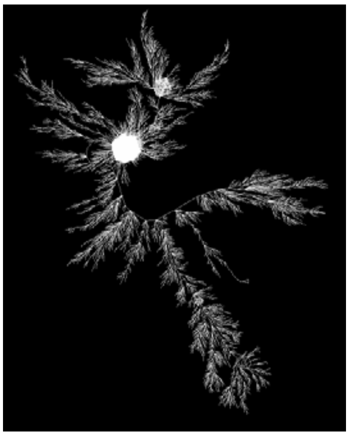
Distance in MPS Distribution



Degree Distribution

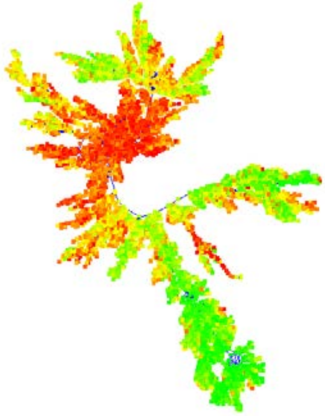


Links



Closeness Centrality

Sophistication



APPENDIX C

Geographic Determinants of Countries' Specialization, Alternative Methods Results

	Heckman Selection Model	OLS Whole sample, inputed values	OLS Conditional to non-censored	Tobit Model	Poisson Counting Model
ldist	0.20***	0.22***	0.18***	0.22***	0.21***
	[0.012]	[0.016]	[0.014]	[0.016]	[0.016]
ldiflat	0.04***	0.04***	0.05***	0.04***	0.05***
	[0.003]	[0.005]	[0.005]	[0.005]	[0.006]
ltdif	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
	[0.002]	[0.003]	[0.003]	[0.004]	[0.003]
ldifarea	0.03***	0.03***	0.02***	0.03***	0.03***
	[0.003]	[0.005]	[0.004]	[0.005]	[0.004]
o1island	0.08***	0.10***	0.07***	0.10***	0.08***
	[0.014]	[0.020]	[0.016]	[0.020]	[0.019]
o1landlock	0.11***	0.13***	0.09***	0.13***	0.14***
	[0.015]	[0.018]	[0.015]	[0.018]	[0.019]
border	-0.22***	-0.22***	-0.22***	-0.22***	-0.38***
	[0.043]	[0.054]	[0.047]	[0.054]	[0.060]
comcont	-0.09***	-0.09***	-0.10***	-0.09***	-0.09***
	[0.020]	[0.029]	[0.024]	[0.029]	[0.030]
comhemis	-0.07***	-0.08***	-0.07***	-0.08***	-0.08***
	[0.016]	[0.022]	[0.018]	[0.022]	[0.022]
ldifgdp	0.15***	0.15***	0.16***	0.14***	0.14***
	[0.004]	[0.005]	[0.004]	[0.005]	[0.004]
ldifgdppc	0.07***	0.07***	0.07***	0.07***	0.07***
	[0.004]	[0.006]	[0.005]	[0.006]	[0.006]
colony	-0.25***	-0.26***	-0.25***	-0.26***	-0.27***
	[0.050]	[0.063]	[0.062]	[0.063]	[0.063]
comcol		0.03	0.03	0.03	0.04
		[0.029]	[0.026]	[0.030]	[0.030]
comlang		0.02	0.02	0.02	0.01
		[0.027]	[0.023]	[0.027]	[0.027]
comrel	0.01	0.01	-0.00	0.01	0.00
	[0.013]	[0.019]	[0.016]	[0.019]	[0.019]
comleg	0.00	-0.01	-0.01	-0.01	-0.00
	[0.012]	[0.018]	[0.015]	[0.018]	[0.018]
comcur	-0.04	-0.06	-0.05	-0.06	-0.07
	[0.063]	[0.077]	[0.066]	[0.078]	[0.084]
rta	-0.31***	-0.31***	-0.32***	-0.31***	-0.44***
	[0.027]	[0.029]	[0.026]	[0.029]	[0.031]
wto_both	-0.42***	-0.46***	-0.39***	-0.46***	-0.45***
	[0.016]	[0.021]	[0.017]	[0.021]	[0.020]
wto_none	0.27***	0.37***	0.20***	0.38***	0.33***
	[0.023]	[0.037]	[0.025]	[0.037]	[0.034]
Constant	3.17***	3.26***	2.94***		
	[0.148]	[0.241]	[0.178]		
Observations	178,022	178,022	175,768	178,022	178,022
R-squared		0.406	0.465		

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1