Journal of Evolutionary Economics
https://doi.org/10.1007/s00191-019-00638-7

REGULAR ARTICLE

Bilateral relatedness: knowledge diffusion and the evolution of bilateral trade

Bogang Jun[1,2] · Aamena Alshamsi[1,3] · Jian Gao[1,4] · César A. Hidalgo[1,5,6,7,8]

Published online: 09 September 2019

Abstract

During the last two decades, two important contributions have reshaped our understanding of international trade. First, countries trade more with those with whom they share history, language, and culture, suggesting that trade is limited by information frictions. Second, countries are more likely to start exporting products that are related to their current exports, suggesting that shared capabilities and knowledge diffusion constrain export diversification. Here, we join both of these streams of literature by developing three measures of bilateral relatedness and using them to ask whether the destinations to which a country will increase its exports of a product are predicted by these forms of relatedness. The first form is product relatedness, and asks whether a country already exports many similar products to a destination. The second is importer relatedness, and asks whether the country exports the same product to the neighbors of the target destination. The third is exporter relatedness, and asks whether a country's neighbors are already exporting the same product to the destination. We use bilateral trade data from 2000 to 2015, and a variety of controls in multiple gravity specifications, to show that countries are more likely to increase their exports of a

✉ Bogang Jun
bogang.jun@inha.ac.kr

[1] Collective Learning Group, MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2] Department of Economics, Inha University, Incheon, South Korea

[3] Masdar Institute, Khalifa University, Abu Dhabi, UAE

[4] CompleX Lab, Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China

[5] ANITI Chair, University of Toulouse, 41 Allée Jules Guesde, 31000 Toulouse, France

[6] Manchester Institute of Innovation Research, University of Manchester, Manchester, M13 9PL, UK

[7] Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[8] Datawheel, Cambridge, MA 02139, USA

product to a destination when they have more product relatedness, importer relatedness, and exporter relatedness. Then, we use several sample splits to explore whether the effects of these forms of relatedness are stronger for products of higher complexity, technological sophistication, and differentiation. We find that, in the case of product relatedness, the effects are stronger for differentiated, complex, and technologically sophisticated products. Also, we find the effects of common language and shared colonial past to increase with differentiation, complexity, and technological sophistication, while the effects of shared borders decrease with these three variables. These results suggest that product relatedness and common language capture dimensions of knowledge relatedness that are more important for the exchange of more sophisticated and differentiated products. These findings extend the ideas of relatedness to bilateral trade and show that the evolution of bilateral trade networks are shaped by relatedness among products, exporters, and importers.

Keywords International trade · Relatedness · Knowledge diffusion · Economic complexity

JEL Classification F1 · O14 · O33

1 Introduction

For more than a century, the literature on international trade explained global commerce as a consequence of differences in factor endowments (Heckscher and Ohlin 1991), product quality, and product differentiation (Krugman 1979; 1991; Anderson 1979; Helpman 1987). More recent streams of literature, however, have shown that there is more to international trade than endowments, costs, and distance, since countries need to learn how to produce and export each product (Hidalgo et al. 2007; Hidalgo 2015), and also need to overcome important information frictions to enter each export destination (Rauch 1999; 2001; Rauch and Trindade 2002; Casella and Rauch 2002; Anderson and Marcouiller 2002; Portes and Rey 2005; Petropoulou 2008; Garmendia et al. 2012; Morales et al. 2015).

During the last two decades scholars have documented that volumes of bilateral trade decrease with the presence of borders (McCallum 1995), and increase with migrants, shared language, and social networks (Rauch 2001; Rauch and Trindade 2002; Combes et al. 2005; Chaney 2014; Morales et al. 2015; Bailey et al. 2017). In fact, using the random re-allocation of the Vietnamese boat people–a population of 1.4 million Vietnamese refugees reallocated in the U.S.–(Parsons and Vézina 2018) showed that states that received a 10% increase in their Vietnamese population experienced a growth in exports to Vietnam of between 4.5% and 14%.

But the evidence in favor of knowledge diffusion is not only expressed in aggregated trade flows, since scholars have also shown the effects of language, social networks, and informal institutions to be larger for differentiated products (Rauch 1999; 2001; Rauch and Trindade 2002). This suggests that factors limiting knowledge and information diffusion (from social networks to language) play a more important role in the diffusion of the knowledge and information needed to exchange more sophisticated goods.

A second stream of literature has focused on the supply side, in particular on the process by which countries learn how to produce the products they export. This literature has shown that the ability of countries and regions to enter new export markets is limited by knowledge diffusion, since countries and regions are more likely to start exporting products when these are related to their current exports (Hidalgo et al. 2007; Hidalgo and Hausmann 2009; Boschma et al. 2013; Hausmann et al. 2014), or when their geographic neighbors are already exporting them (Bahar et al. 2014). The importance of knowledge diffusion in the diversification of economic activities, however, is not limited only to the export of products. It has also been observed in the development of regional industries (Neffke et al. 2011; Gao et al. 2017), research activities (Guevara et al. 2016), and technologies (Kogler et al. 2013; Boschma et al. 2014), suggesting that relatedness between economic activities facilitates knowledge diffusion in general (Hidalgo et al. 2018), not only in the context of international trade.

Together, these findings have given rise to a more nuanced picture of international trade, a picture where factor endowments and transportation costs do not determine trade fully, because information frictions and knowledge diffusion determine the knowledge a country has, and hence, the products it can produce and the partners, with which it can trade.

Here, we contribute to this literature by using more than 15 years of bilateral trade data, disaggregated into more than 1,200 products, to construct several gravity models that validate previous findings and expand them. We use this data to construct three measures of bilateral relatedness. First, following Hidalgo et al. (2007), we construct a measure of product or technological relatedness to explore whether countries increase the exports of a product to a destination when they already export more related products to it. Product relatedness should capture information on the shared knowledge and capabilities needed to make the products. Second, following Chaney (2014) and Morales et al. (2015), we construct a measure of importer relatedness to explore whether countries increase the exports of a product to a destination when they already export the same product to the neighbors of that destination. Importer relatedness should capture information on shared logistics channels, such as knowledge of distribution centers, shipping companies, and customs unions. Third, following Bahar et al. (2014), we construct a measure of exporter relatedness to explore whether countries increase the exports of a product to a destination when the neighbors of this country already export the same product to that destination. This channel may provide information about knowledge diffusion among geographic neighbors, or about an importer's taste for the variety of product produced by countries in a region (e.g. American instead of European cars).

Looking at hundreds of thousands of bilateral trade links confirms that countries are more likely to increase their exports of a product to a destination when they export related products to it, when they export to that destination's neighbors, and when their neighbors export that same product to that destination. Moreover, we find that sharing a colonial past, a language, or a border is also associated with an increase in the volume of trade.

Next, we do several sample splits to explore whether the effects of these forms of bilateral relatedness, as well as shared cultural and geographic factors, increase with the differentiation (Rauch 1999), technological sophistication (Lall 2000), and complexity of products (Hidalgo and Hausmann 2009). We find that the effects of product relatedness, language, and shared colonial past increase with the differentiation, technological sophistication, and complexity of products, suggesting that these channels mediate knowledge flows that are more relevant for the export of complex, sophisticated, and differentiated goods. We also find the effect of borders to decrease with technological sophistication, differentiation, and complexity, suggesting that, once the effects of relatedness and shared cultural factors are taken into account, shared borders are more important for undifferentiated and simple products.

Our empirical results also indicate that the effects of product relatedness in bilateral trade are particularly strong. In fact, a one standard deviation increase in product relatedness is associated with a 21% in a two-year period. This effect is about 46% larger than the effect of exporting that product to a neighbor of the target destination (Chaney 2014; Morales et al. 2015), and more than 170% larger than the effect of having a neighbor exporting the same product to the same destination (Bahar et al. 2014).

Together these findings contribute to our understanding of the role of relatedness in the evolution of international trade by providing a comprehensive extension of the concept of relatedness to bilateral trade data, and by showing that these three forms of relatedness are predictive of increases and decreases in bilateral trade.

2 Data

We use bilateral trade data from 2000 to 2015 from MIT's Observatory of Economic Complexity (Simoes and Hidalgo 2011). The data are disaggregated into the Harmonized System (HS rev 1992, four-digit level) and consist of imports and exports between countries. Because both exporter and importer report their trade information, we clean the data by comparing the data reported by exporters and importers following the work of Feenstra et al. (2005). Also, we exclude countries that have population less than 1.2 million or have a trade volume in 2008 that is below one billion in US dollars. Also, we exclude data from Iraq, Chad and Macau.

Macroeconomic data (GDP at market prices in current US dollar and population) come from the World Bank's World Development Indicators. Data on Product Complexity Index (PCI) are from MIT's Observatory of Economic Complexity (Simoes and Hidalgo 2011), while data on geographical and cultural distance (shared language, physical distance between most populated cities, sharing a border, and shared colonial past) come from GEODIST data from CEPII (Mayer and Zignago 2011). For language proximity, we use one of the global language networks of Ronen et al. (2014): the one considering the number of books translated from one language to another as a proxy for the number of translators, or bilingual speakers, between two languages.

3 Model

Does relatedness among products or geographic neighbors predict increases in bilateral trade flows?

To explore this question, we introduce three measures of relatedness. We use these to estimate: (i) the fraction of the geographic neighbors of a country that import a

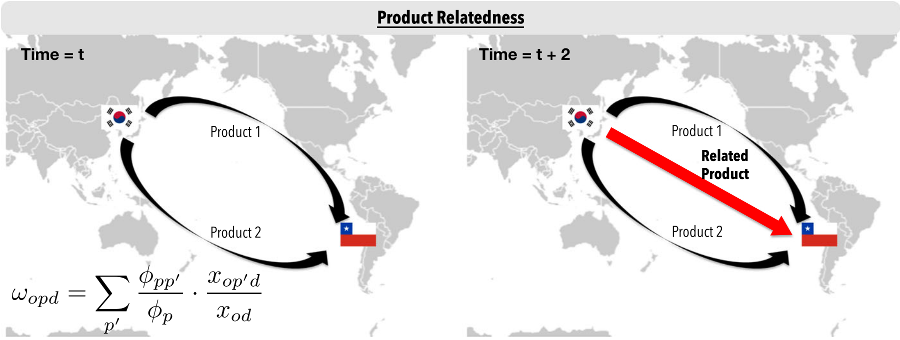

Fig. 1 Relatedness among products. Product Relatedness: the similarity between a product and the other products that a country already exports to a destination

product from the same origin (importer relatedness), (ii) the fraction of neighbors of a country that export a product to the same destination (exporter relatedness), and (iii) the similarity between a product and the other products that a country already exports to a destination (product relatedness). Product relatedness should help us capture information about knowledge flows between products (that range from knowledge flows among industries to knowledge flows among product lines within a firm). Figure 1 illustrates product relatedness in the context of Korea and Chile. In this

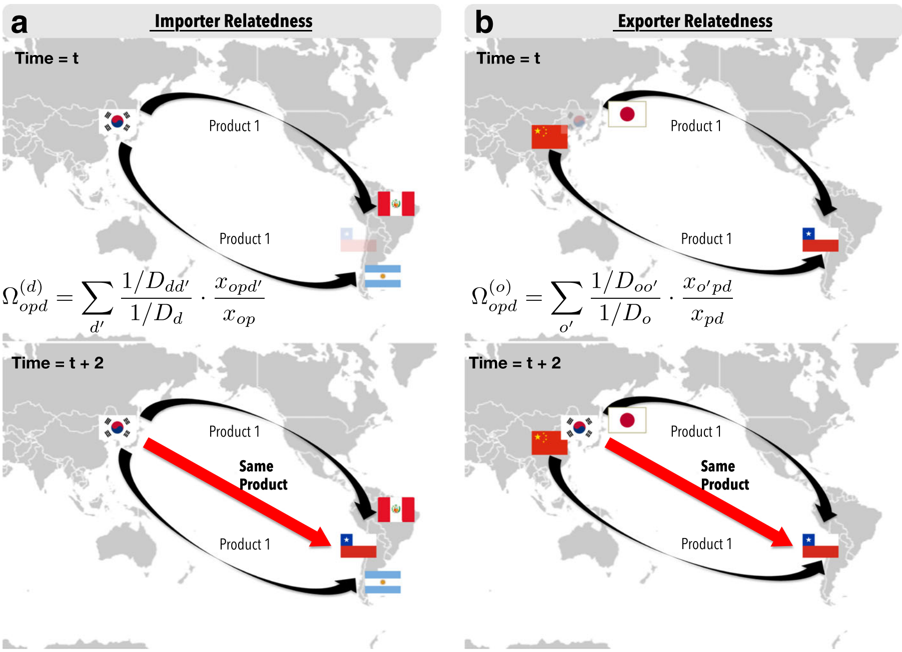

Fig. 2 Relatedness among exporters and importers. **a** Importer Relatedness: the fraction of the geographic neighbors of a country that import a product from the same origin, and **b** Exporter Relatedness: the fraction of neighbors of a country that export a product to the same destination

example, Korea exports Products 1 and 2 to Chile (Shirts and Pants), and this may affect the future exports of a related product (Product 3, Coats) to Chile. Our hypothesis is that knowledge flows on how to export to a destination should be larger among related products, and hence, exports should increase faster when a country exports related products to a destination.

Importer relatedness helps us capture knowledge flows on how to: (i) import a product from the same origin than a neighbor, or (ii) export to a neighbor of a current destination. In the example of Fig. 2a, Korea exports Product 1 (Shirts) to Peru and Argentina and that may affect the future volume of exports of Product 1 (Shirts) to Chile (a geographic neighbor of Peru and Argentina). Here, knowledge on how to import from an origin should be flowing among neighboring importers, or knowledge on how to export to the neighbor's of a country's destinations should be flowing within the exporter. This could be knowledge on how to distribute within a region, knowledge on how to navigate a social network within a certain culture, or knowledge on shipping routes and customs.

Exporter relatedness captures (i) knowledge flows among neighboring exporters on how to export to a destination, or (ii) knowledge flows on how to import from a neighbor of a country from where you currently import. In the example of Fig. 2b, Chile imports Product 1 (Shirts) from China and Japan, and that may affect the future volume of exports of Product 1 (Shirts) from Korea (a neighbor of the places from which Chile is currently importing Product 1). This would be a knowledge flow on how to export to a destination among neighboring exporters, or a knowledge flow within an importer, of how to import from a neighbor of a current origin. Exporter relatedness may also signal information about the taste for a regional variety (e.g. American car versus Korean or Japanese car).

Mathematically, we can construct the three measures of relatedness using a similar formula. The formula is a weighted average of the number of neighbors, or related products, that already have an active trade relationship. In the case of similarity between products, weights are the proximity between products p and p', $\phi_{pp'}$. $\phi_{pp'}$ is the minimum of the conditional probability that two products are co-exported by multiple countries (see Hidalgo et al. (2007) and Appendix A). $\phi_{pp'} = 1$ means products p and p' are always co-exported and $\phi_{pp'} = 0$ means no country exports both products. In the case of geographic neighbors, of both an exporter or an importer, weights are given by the inverse of geographic distance ($1/D_{dd'}$ and $1/D_{oo'}$), where $D_{dd'}$ is the distance in kilometers between the most populated cities in countries d and d'.

Formally, let x_{opd} be a matrix summarizing the trade flow in US dollars of product p from exporter o to destination d. Then, product relatedness is given by:

$$\omega_{opd} = \sum_{p'} \frac{\phi_{pp'}}{\phi_p} \cdot \frac{x_{op'd}}{x_{od}} \tag{1}$$

where x_{od} is the volume of trade between countries o and d ($x_{od} = \sum_p x_{opd}$) and $\phi_p = \sum_{p'} \phi_{pp'}$.

Similarly, importer relatedness is given by:

$$\Omega^{(d)}_{opd} = \sum_{d'} \frac{1/D_{dd'}}{1/D_d} \cdot \frac{x_{opd'}}{x_{op}}, \tag{2}$$

where x_{op} is the volume of trade of product p from origin country o ($x_{op} = \sum_d x_{opd}$), $D_{dd'}$ is the geographic distance between destination country d and its neighbors d', and $1/D_d = \sum_{d'} 1/D_{dd'}$.

Finally, exporter relatedness is given by:

$$\Omega^{(o)}_{opd} = \sum_{o'} \frac{1/D_{oo'}}{1/D_o} \cdot \frac{x_{o'pd}}{x_{pd}}, \tag{3}$$

where x_{pd} is the volume of trade of product p to destination country d ($x_{pd} = \sum_o x_{opd}$), $D_{oo'}$ is the geographic distance between origin country o and its neighbors o', and $1/D_o = \sum_{o'} 1/D_{oo'}$.

Next, we use these three measures of relatedness, together with data on common cultural and geographic factors, to construct an extended gravity model to study the marginal contribution of product, importer, and exporter relatedness, and of shared languages, borders, and colonial past, to the growth of future exports. Our model predicts bilateral trade in a product in two years time while controlling for: (i) initial trade in that product between the same trade partners, (ii) total exports of the product by the exporter, (iii) total imports of the product by the importer, (iv-vii) the GDP per capita and population of exporters and importers, and (viii) their geographic distance. Formally, our model is given by Eq. 4:

$$\begin{aligned} x^{t+2}_{opd} = {} & \beta_0 + \beta_1 \omega^t_{opd} + \beta_2 \Omega^{(d)t}_{opd} + \beta_3 \Omega^{(o)t}_{opd} \\ & + \beta_4 x^t_{opd} + \beta_5 x^t_{op} + \beta_6 x^t_{pd} + \beta_7 D_{od} \\ & + \beta_8 gdp^t_o + \beta_9 gdp^t_d + \beta_{10} Population^t_o + \beta_{11} Population^t_d \\ & + \beta_{12} Border_{od} + \beta_{13} Colony_{od} + \beta_{14} Language_{od} \\ & + \beta_{15} Lang.Proximity_{od} \\ & + \varepsilon^t_{opd} \end{aligned} \tag{4}$$

where the dependent variable, x^{t+2}_{opd}, represents the volume of trade (in US dollar) of product p from exporter o to destination d in year $t+2$. Our main variables of interest are our three measures of relatedness: product relatedness (ω^t_{opd}), importer relatedness ($\Omega^{(d)t}_{opd}$), exporter relatedness ($\Omega^{(o)t}_{opd}$), and shared border ($Border_{od}$), shared language ($Language_{od}$), language proximity (number of bilingual speakers $Lang.Proximity_{od}$), and shared colonial past ($Colony_{od}$). $Border_{od}$, $Language_{od}$, $Colony_{od}$ are binary (dummy) variables (0 or 1). The other factors in the model, GDP per capita, population, and distance (D_{od}), are standard gravity controls (Tinbergen 1962; Pöyhönen 1963). Finally, by incorporating the total volume of exports of a country (x_{op}), the total imports of a destination (x_{pd}), and the present day trade flow for each product between an origin and a destination (x_{opd}), we capture the effects of our variable of interest in the change in trade experience in the subsequent two years. In Eq. 4, we make all variables comparable (except binary variables) by standardizing them by subtracting their means and dividing them by their standard deviations. To prevent the estimates to be dominated by the tails of right skewed distributions, we take the logarithm for x^{t+2}_{opd}, x^t_{opd}, x^t_{op}, x^t_{pd}, D_{od}, gdp^t_o, gdp^t_d, $Population^t_o$, $Population^t_d$ and $Lang.Proximity_{od}$ after checking the distribution of each value.

Table 1 Summary statistics (year 2000-2006)

Statistic	N	Mean	St. Dev.	Min	Max
ω^t_{opd}	10,911,584	0.000	1.000	– 3.366	26.699
$\Omega^{(d)}_{opd}$	10,911,584	0.000	1.000	– 0.959	93.647
$\Omega^{(o)}_{opd}$	10,911,584	0.000	1.000	– 1.172	51.786
$\log x^t_{opd}$	10,911,584	0.000	1.000	– 1.339	4.086
$\log x^t_{op}$	10,911,584	0.000	1.000	– 2.547	3.172
$\log x^t_{pd}$	10,911,584	0.000	1.000	– 2.544	3.758
$\log Distance$	10,911,584	0.000	1.000	– 3.855	1.578
$\log gdp^t_o$	10,911,584	0.000	1.000	– 3.118	1.339
$\log gdp^t_d$	10,911,584	0.000	1.000	– 2.477	1.541
$\log Population_o$	10,911,584	0.000	1.000	– 2.474	2.344
$\log Population_d$	10,911,584	0.000	1.000	– 2.324	2.768
$Border_{od}$	10,911,584	0.089	0.284	0	1
$Colony_{od}$	10,911,584	0.060	0.237	0	1
$Language_{od}$	10,911,584	0.155	0.362	0	1
$\log Lang.Proximity_{od}$	10,911,584	0.000	1.000	– 0.312	5.050

To avoid taking the logarithm of zero, we add a small $\epsilon = 100$ for x^{t+2}_{opd} and x^t_{opd}. We also add $\epsilon \cdot \lambda$ for x^t_{op} and x^t_{pd}, where ϵ is 100 and λ is equal to the number of exporters for x^t_{op} and the number of importers for x^t_{pd}, respectively. Table 1 presents summary statistics of our base data covering the 2000-2006 period.

4 Results

Table 2 shows our main results divided into three periods: 2000–2006 (pre-financial crisis), 2007–2012 (crisis period), and 2012–2015 (recovery period). Since our results are qualitatively the same for all of these periods, we will describe them together. Because there is possibility of strong intra-group correlation among errors that inflates the precision of the parameter estimates, we cluster errors by country of origin, country of destination and product, using three-way clustering.[1] Also, since there is potential for omitted variable bias, we add in the Appendices C specifications with Exporter-Importer and Year fixed effects; with Exporter-Product and Year fixed effects; with Importer-Product and Year fixed effects; and with Exporter, Importer, and Product fixed effects (Table 10 of Appendix C. See correlation table and summary statistics in Appendix E).

First, we find that the three relatedness variables correlate positively with future bilateral trade. This means that countries exporting related products to a destination, exporting the same product to the neighbors of a destination (confirming Chaney

[1]Note: We dropped singleton observations, when we apply three-way error clustering (Correia 2015).

Table 2 Bilateral trade volume after two years for periods 2000–2006 (pre-financial crisis), 2007–2012 (crisis period) and 2012–2015 (recovery period)

	Dependent variable: log x_{opd}^{t+2}		
	(1) 2000–2006	(2) 2007–2012	(3) 2012–2015
ω_{opd}^{t}	0.209***	0.180***	0.152***
	(0.017)	(0.017)	(0.015)
$\Omega_{opd}^{(d)}$	0.143***	0.143***	0.151***
	(0.025)	(0.047)	(0.036)
$\Omega_{opd}^{(o)}$	0.077***	0.105***	0.107***
	(0.020)	(0.023)	(0.020)
log x_{opd}^{t}	1.371***	1.600***	1.769***
	(0.055)	(0.057)	(0.057)
log x_{op}^{t}	0.961***	0.971***	0.915***
	(0.032)	(0.035)	(0.032)
log x_{pd}^{t}	0.529***	0.421***	0.406***
	(0.028)	(0.030)	(0.027)
log $Distance$	−0.485***	−0.423***	−0.432***
	(0.036)	(0.042)	(0.038)
log gdp_{o}^{t}	0.165***	0.200***	0.203***
	(0.035)	(0.042)	(0.042)
log gdp_{d}^{t}	0.226***	0.218***	0.281***
	(0.041)	(0.037)	(0.035)
log $Population_{o}$	0.472***	0.457***	0.455***
	(0.029)	(0.047)	(0.048)
log $Population_{d}$	0.344***	0.368***	0.338
	(0.029)	(0.030)	(0.028)
$Border_{od}$	0.712***	0.709***	0.611***
	(0.066)	(0.070)	(0.061)
$Colony_{od}$	0.052	0.135	0.193***
	(0.070)	(0.083)	(0.070)
$Language_{od}$	0.545***	0.436***	0.334***
	(0.061)	(0.070)	(0.071)
log$Lang.Proximity_{od}$	0.032*	0.000*	−0.001
	(0.000)	(0.000)	(0.019)
Constant	9.653***	9.828***	9.793***
	(0.045)	(0.047)	(0.045)
Observations	10,911,584	7,591,489	5,332,257
Adjusted R^2	0.495	0.516	0.558
Root MSE	2.5681	2.637	2.529

Three-way clustering robust standard errors are reported in parentheses and $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

(2014) and Morales et al. (2015)), or with neighbors that are already exporting the same product, are more likely to experience an increase in their exports of a product to a destination. This substantially extends Bahar et al. (2014), who showed that having geographic neighbors increases the probability of exporting a new product, since they did not look at individual export destinations (they aggregate across all destinations). Similarly, this substantially extends Hidalgo et al. (2007), since they also aggregated across all destinations. Also, we extend Chaney (2014) and Morales et al. (2015) by providing evidence with multiple origin countries (not just one). Our findings, therefore, complement and expand Bahar et al. (2014), Chaney (2014) and Morales et al. (2015), and Hidalgo et al. (2007). Interestingly, we find that some of the controls, such as colonial past and language proximity, lose significance with three-way error clustering, while the three relatedness variables hold their significance and sign with three-way error clustering.

When comparing the effects of product and geographical relatedness (variables are standardized), we find that the role of product relatedness (ω^{t}_{opd}) is on average the largest, while that of exporter relatedness (Ω^{o}_{opd}) is the smallest. In addition to these, we find strong and positive effects for the role of shared borders, and shared language either without or with three-way error clustering (for the results without three-way error clustering, see Appendix B).

These findings are evidence in support of the notion that knowledge on how to trade a specific product between a specific pair of countries needs to flow for that trade to be materialized. If this hypothesis is correct, we should also be able to study the varying importance of knowledge flows for products with different levels of complexity (Table 3), technological sophistication (Table 4), and differentiation (Table 5). We should expect the effects of relatedness to increase with the differentiation, technological sophistication, and complexity of products, if relatedness captures information about knowledge flows.

We first separate products into low (first quartile), medium (second and third quartile), and high (fourth quartile) complexity by using the Product Complexity Index (PCI) of products for each respective year (Hidalgo and Hausmann 2009). Table 3 shows that the effect of importer relatedness, $\Omega^{(d)}_{opd}$, increase with product complexity. The effect of product relatedness of medium and high PCI product is larger than that of low PCI product.

We further explore the interaction between our three measures of relatedness and the products' technological sophistication using Lall (2000)'s five technological categories: primary, resource-based manufactures, low-tech, medium-tech, and high-tech products. Since Lall's classification is based on the 3-digit Standard International Trade Classification (SITC-3) rev 2, we match products to our data using the conversion table provided by the UN Trade Statistics site. [2] Following Lall (2000), we also exclude "special transactions" such as electric current, cinema film, printed matter, coins, and pets.

Table 4 shows the gravity model split into the five Lall's categories. To simplify the presentation of these results, we also show the coefficients for our main variables

[2] Available at https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp

Table 3 Bilateral trade volume after two years by different levels of Product Complexity Index (PCI)

	Dependent variable: log x_{opd}^{t+2}		
	(1) Low PCI	(2) Medium PCI	(3) High PCI
ω_{opd}^{t}	0.185***	0.222***	0.218***
	(0.017)	(0.022)	(0.026)
$\Omega_{opd}^{(d)}$	0.133***	0.137***	0.174***
	(0.028)	(0.025)	(0.024)
$\Omega_{opd}^{(o)}$	0.102***	0.065***	0.059**
	(0.020)	(0.020)	(0.027)
log x_{opd}^{t}	1.408***	1.355***	1.349***
	(0.064)	(0.057)	(0.062)
log x_{op}^{t}	0.926***	0.973***	0.989***
	(0.035)	(0.034)	(0.044)
log x_{pd}^{t}	0.458***	0.515***	0.572***
	(0.031)	(0.029)	(0.039)
log $Distance$	−0.436***	−0.501***	−0.512***
	(0.037)	(0.038)	(0.042)
log gdp_{o}^{t}	0.137***	0.140***	0.192***
	(0.030)	(0.038)	(0.050)
log gdp_{d}^{t}	0.276***	0.220***	0.228***
	(0.039)	(0.044)	(0.047)
log $Population_{o}$	0.405***	0.459***	0.532***
	(0.044)	(0.043)	(0.045)
log $Population_{d}$	0.318***	0.338***	0.394***
	(0.025)	(0.031)	(0.037)
$Border_{od}$	0.658***	0.723***	0.734***
	(0.065)	(0.063)	(0.070)
$Colony_{od}$	0.040	0.039	0.098
	(0.054)	(0.074)	(0.083)
$Language_{od}$	0.460***	0.567***	0.580***
	(0.055)	(0.061)	(0.073)
log $Lang.Proximity_{od}$	0.037*	0.030	0.033
	(0.020)	(0.020)	(0.022)
Constant	9.578***	9.643***	9.715***
	(0.045)	(0.046)	(0.050)
Observations	2,531,910	5,113,730	3,238,123
Adjusted R^2	0.459	0.485	0.531
Root MSE	2.616	2.592	2.478

Three-way clustering robust standard errors are reported in parentheses and *p<0.1; **p<0.05; ***p<0.01

Table 4 Bilateral trade volume after two years for five technological categories

	Dependent variable: $\log x_{opd}^{t+2}$				
	Primary product	Resource-based manufactures	Low-tech manufactures	Medium-tech manufactures	High-tech manufactures
ω_{opd}^{t}	0.183***	0.164***	0.204***	0.203***	0.229***
	(0.019)	(0.018)	(0.021)	(0.022)	(0.028)
$\Omega_{opd}^{(d)}$	0.152***	0.144***	0.131***	0.154***	0.128***
	(0.028)	(0.029)	(0.027)	(0.032)	(0.020)
$\Omega_{opd}^{(o)}$	0.098***	0.076***	0.072***	0.084***	0.078***
	(0.017)	(0.019)	(0.022)	(0.031)	(0.027)
$\log x_{opd}^{t}$	1.260***	1.341***	1.425***	1.348***	1.476***
	(0.068)	(0.063)	(0.061)	(0.075)	(0.079)
$\log x_{op}^{t}$	0.996***	0.923***	0.866***	1.006***	0.995***
	(0.035)	(0.036)	(0.040)	(0.042)	(0.060)
$\log x_{pd}^{t}$	0.565***	0.490***	0.428***	0.509***	0.615***
	(0.038)	(0.034)	(0.033)	(0.036)	(0.044)
$\log Distance$	−0.421***	−0.439***	−0.495***	−0.544***	−0.477***
	(0.040)	(0.034)	(0.042)	(0.044)	(0.039)
$\log gdp_{o}^{t}$	0.103***	0.082**	0.166***	0.132***	0.296***
	(0.033)	(0.038)	(0.039)	(0.042)	(0.054)
$\log gdp_{d}^{t}$	0.151***	0.184***	0.292***	0.164***	0.271***
	(0.038)	(0.042)	(0.045)	(0.053)	(0.056)
$\log Population_{o}$	0.326***	0.419***	0.482***	0.511***	1.517***
	(0.035)	(0.040)	(0.059)	(0.050)	(0.047)
$\log Population_{d}$	0.298***	0.323***	0.339***	1.349***	0.379***
	(0.030)	(0.032)	(0.030)	(0.037)	(0.043)
$Border_{od}$	0.741***	0.754***	0.668***	0.733***	0.664***
	(0.074)	(0.064)	(0.064)	(0.070)	(0.080)
$Colony_{od}$	−0.112*	− 0.016	0.077	0.037	0.182*
	(0.067)	(0.067)	(0.068)	(0.079)	(0.094)
$Language_{od}$	0.395***	0.479***	0.553***	0.556***	0.619***
	(0.056)	(0.053)	(0.066)	(0.071)	(0.080)
$\log Lang.Proximity_{od}$	0.043**	0.038*	0.031	0.036*	0.021
	(0.019)	(0.019)	(0.023)	(0.021)	(0.023)
Constant	9.445***	9.513***	9.358***	10.024***	10.140***
	(0.044)	(0.046)	(0.050)	(0.054)	(0.053)
Observations	1,127,670	2,241,432	3,314,246	1,110,342	1,049,765
Adjusted R^2	0.399	0.446	0.527	0.471	0.565
Root MSE	2.874	2.658	2.371	2.689	2.478

Three-way clustering robust standard errors are reported in parentheses and *p<0.1; **p<0.05; ***p<0.01

Table 5 Bilateral trade volume after two years by the differentiation of products (Rauch classification)

	Dependent variable: log x_{opd}^{t+2}		
	(1) Organized exchange	(2) Reference priced	(3) Differentiated
ω_{opd}^{t}	0.212***	0.140***	0.214***
	(0.050)	(0.024)	(0.022)
$\Omega_{opd}^{(d)}$	0.140***	0.122***	0.151***
	(0.035)	(0.031)	(0.033)
$\Omega_{opd}^{(o)}$	0.111***	0.075***	0.091***
	(0.025)	(0.020)	(0.030)
log x_{opd}^{t}	0.986***	1.395***	1.513***
	(0.113)	(0.079)	(0.075)
log x_{op}^{t}	1.111***	0.976***	0.938***
	(0.051)	(0.048)	(0.048)
log x_{pd}^{t}	0.594***	0.599***	0.563***
	(0.070)	(0.060)	(0.043)
log $Distance$	−0.429***	−0.446***	−0.495***
	(0.061)	(0.051)	(0.042)
log gdp_{o}^{t}	0.108**	0.086*	0.208***
	(0.049)	(0.043)	(0.044)
log gdp_{d}^{t}	0.139**	0.132**	−0.223***
	(0.061)	(0.055)	(0.047)
log $Population_{o}$	0.301***	0.338***	0.501***
	(0.038)	(0.044)	(0.055)
log $Population_{d}$	0.302***	0.243***	0.326***
	(0.048)	(0.037)	(0.035)
$Border_{od}$	0.920***	0.851***	0.741***
	(0.100)	(0.077)	(0.067)
$Colony_{od}$	−0.123	−0.036	0.156*
	(0.110)	(0.082)	(0.088)
$Language_{od}$	0.426***	0.472***	0.581***
	(0.072)	(0.056)	(0.069)
log$Lang.Proximity_{od}$	0.041**	0.025	0.040*
	(0.019)	(0.018)	(0.024)
Constant	9.482***	9.551***	9.624***
	(0.089)	(0.059)	(0.051)
Observations	134,015	392,533	1,084,194
Adjusted R^2	0.340	0.420	0.547
Root MSE	3.401	2.852	2.496

Three-way clustering robust standard errors are reported in parentheses and $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

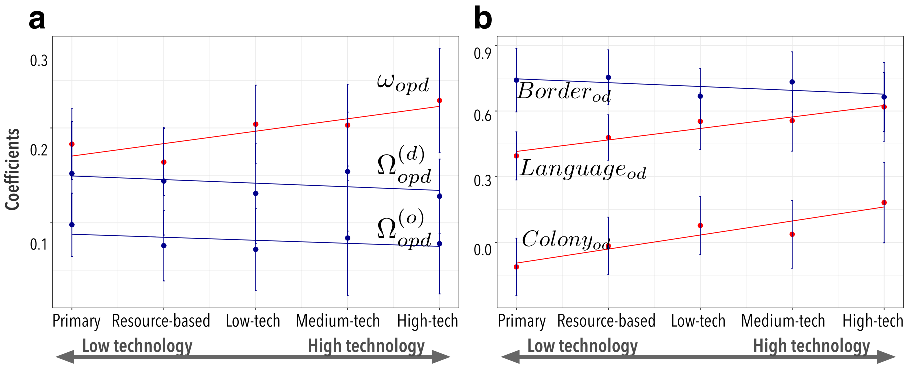

Fig. 3 Coefficients of variables by the technological sophistication of products; **a** Coefficients of ω_{opd}, $\Omega^{(d)}_{opd}$, and $\Omega^{(o)}_{opd}$, and **b** Coefficients of $Border_{od}$, $Language_{od}$, and $Colony_{od}$. The fitted lines in red are statistically significant, while the lines in blue are not statistically significant. The R^2s of line for ω_{opd}, $\Omega^{(d)}_{opd}$, and $\Omega^{(o)}_{opd}$ are 0.62, 0.01, and 0.001, respectively, while those for $Border_{od}$, $Language_{od}$, and $Colony_{od}$ are 0.23, 0.82, and 0.91, respectively. *p<0.1; **p<0.05; ***p<0.01

of interest in Fig. 3. Trends that increase significantly with technological sophistication ($p < 0.1$) are presented in red, whereas non-significant trends are shown in blue. Figure 3a confirms the results found for product complexity by showing that the effect of product relatedness, but not that of importer or exporter relatedness, increases with technological sophistication. These two results support our idea that product relatedness captures channels of knowledge and information flow that are more relevant for the export of sophisticated products.

Also, Fig. 3b shows that the effect of sharing a language and a colonial past, but not those of sharing a border, are larger for more technologically sophisticated products (although the individual coefficients for colony are not significant for products with intermediate sophistication, when the coefficient is close to zero). Once again, this reiterates the idea that borders and geographic distance affect knowledge flows by limiting social interactions (Singh 2005; Breschi and Lissoni 2009), so we do not see much of a geographic effect once we take cultural and linguistic similarity into account. In fact, the effect of borders decreases with sophistication once we have taken into account the effects of culture and relatedness and we have included error clustering. Together, these findings support the idea that trade is driven partly by the diffusion of knowledge and information on how to export each product to each destination.

Next, we separate products by level of differentiation following the work of Rauch (1999).[3] According to Rauch (1999), homogeneous products that are traded with a reference price can be distinguished from differentiated products, and "further divided into those products whose reference prices are quoted on organized exchanges and those products whose reference prices are quoted only in trade publications." He showed that factors relating to decreasing information friction in trade,

[3] Available at http://www.freit.org/TradeResources/TradeData.php

such as sharing colonial past and same language, make the greatest effect on the trade of differentiated products and the weakest effect on that of products traded on an organized exchange (but have unclear effects on the other referenced priced homogeneous products), since traders need to engage in a consecutive search for buyer/seller to trade the differentiated products with searching cost and "this search is facilitated by proximity and common language, and by any contacts who know the market."

Table 5 shows the results for differentiated products, reference-priced homogeneous products, and homogeneous products traded on an organized exchange. Since Rauch (1999) provided the classification in two ways, conservative and liberal, we apply the conservative classification in the main text and the liberal classification in the Appendix (Table 10 of Appendix D). First, our results confirm Rauch (1999)'s claim that proximity factors that decrease information friction in trade, such as sharing colonial past and common language, have the greatest impact on bilateral trade for differentiated products. In addition to those results, our main three relatedness channels, product and importer relatedness, but not exporter relatedness, have the greatest impact on trade increases for differentiated products, showing once again that relatedness is more important, in this case, for differentiated products.

5 Discussion

During the past few decades, two ideas have re-framed our understanding of international trade. The first idea is that information and knowledge frictions, not just differences in transportation costs, factor endowments, and differences in productivity, shape global trade (Rauch 1999; 2001; Rauch and Trindade 2002; Casella and Rauch 2002; Anderson and Marcouiller 2002; Portes and Rey 2005; Petropoulou 2008; Garmendia et al. 2012). The second idea is that countries need to learn how to produce the products they export, and hence, evolve their productive structures in a path-dependent manner that is constrained by knowledge flows (Hidalgo et al. 2007; Hidalgo and Hausmann 2009; Hidalgo et al. 2018; Boschma et al. 2013; Bahar et al. 2014; Chaney 2014; Morales et al. 2015). Here, we use bilateral trade data, together with various measures of economic size, culture, and geographic proximity, to put these two ideas together. Our findings confirm much of the existing work involving the role of language, and culture, but also add to the body of knowledge by showing that relatedness among products and countries shapes future trade volumes. In particular, we showed that relatedness among products, and among geographic neighbors, explains a substantial fraction of future bilateral trade: trade volumes increase when they share neighbors who export to that destination, or when they are already exporting to a destination's neighbors, but also when countries export related products to a destination. When comparing these three forms of relatedness, we found that relatedness among products is by far the strongest, suggesting that there may be product or industry specific learning channels that play an important role in the diffusion of the knowledge needed to establish or increase trade relationships. Moreover, we found that the effects of product relatedness is likely to be stronger for more complex, technologically sophisticated, and differentiated products. These additional considerations support the idea that the presence of related activities facilitates the knowledge flows that countries need to learn how to produce and export products to specific destinations.

Our three relatedness measures tell us about three different channels shaping bilateral trade flows (the three measures are not strongly correlated with $R^2 < 1\%$). Product relatedness captures information on the technological similarity among products (the knowledge needed to produce and trade it). The fact that this relatedness channel was strong in all specifications suggests that the knowledge needed to trade products is quite specific and flows more effectively among related products. The second relatedness channel (importer relatedness) is more directly connected to sales, and may capture information about relationships with established regional distribution centers, logistic companies, and customs unions. This channel may also reflect knowledge of how to do business or cater to the tastes of customers in a region. The third channel, exporter relatedness, may signal knowledge diffusion among geographic neighbors, or may signal a taste for the variety of the product that is common in the exporter's region.

By splitting the data by complexity, Lall's technological sophistication, and Rauch's differentiation classification, we support the idea that knowledge on how to export to specific destinations flows among product lines or related industries. Yet, these classifications are likely to be highly correlated, since products that are more complex, should also be more technologically sophisticated and more likely to be differentiated. Also, we don't provide micro level evidence of the mechanisms, nor are we able to pin-point each of the mechanisms narrowly.

Despite our limitations, our results do provide some light in the long quest to understand how social networks, culture, and knowledge flows shape international trade. They tell us that product relatedness plays an important role since the size of its effect is larger than the one observed among geographic neighbors. This suggests that looking at knowledge flows among product lines and among industries with micro level data should be an avenue of inquiry for improving our understanding of the social and economic forces that govern global trade.

Acknowledgments We thank Mauricio (Pacha) Vargas and Alex Simoes for help with the data. We also thank Cristian-Jara Figueroa, Flávio Pinheiro, Tarik Roukny, Dogyoon Song, and Robert A Irwin for helpful comments. This project is supported by the MIT Skoltech Program, by the National Research Foundation of Korea (NRF) (No. 2019R1G1A1100322), and by the Cooperative Agreement between the Masdar Institute of Science and Technology (Masdar Institute), Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA—Reference 02/MI/MIT/CP/11/07633/GEN/G/00. CAH also acknowledges the support of the ANITI Chair from the University of Toulouse. This paper was circulated as a working paper: "Relatedness, knowledge diffusion, and the evolution of bilateral trade." arXiv preprint arXiv:1709.05392 (Sep 2017).

Appendix A: Building a product space for 2000–2015

To calculate the ω_{opd}, we need first build a product space. We define the product space by looking at all proximity measures between products (Hidalgo et al. 2007) after aggregating all the data that covers from 2000 to 2015. To capture the significant trade flow, we calculate the revealed comparative advantage (RCA) following Balassa (1965):

$$RCA_{o,i} = \frac{x_{o,i}}{\sum_i x_{o,i}} \Bigg/ \frac{\sum_o x_{o,i}}{\sum_{o,i} x_{o,i}} \tag{A1}$$

Based on the result of RCA, we measure the proximity between product by calculating $\phi_{i,j}$ between product i and j (Hidalgo et al. 2007).

$$\phi_{i,j} = \min\left\{P(RCA_i|RCA_j, P(RCA_j|RCA_i))\right\} \tag{A2}$$

Using this significant trade flow over 2000–2015, we can create 1242 × 1242 matrix, which entities the proximity between products. Figure 4 shows the product space of world market in the period from 2000 to 2015.

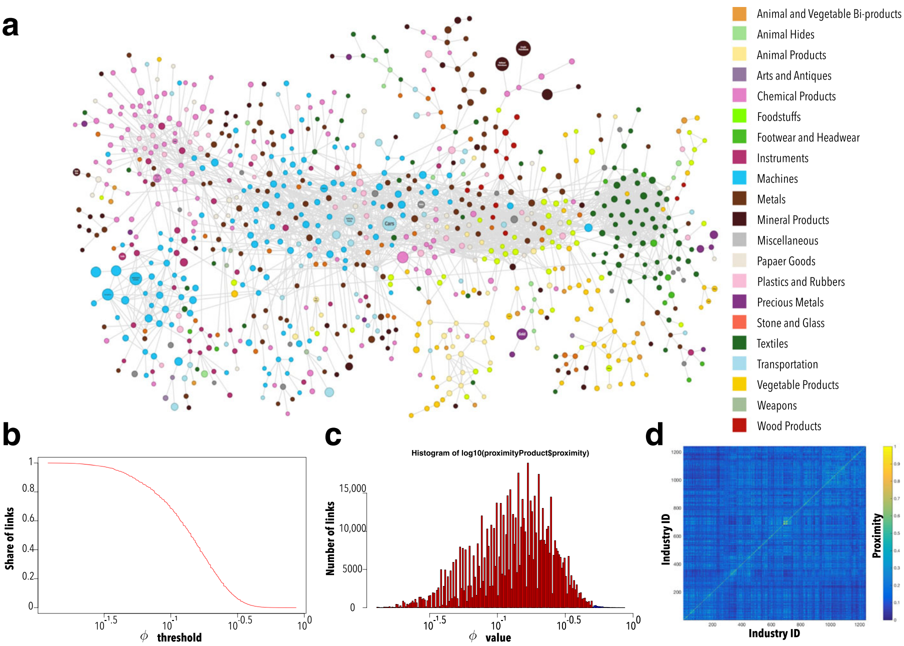

Fig. 4 Product space over 2000 to 2015: **a** Network representation of product space, **b** Cumulative distribution of proximity values, **c** Density distribution of proximity values, and **d** the product space matrix sorted in increasing order of the is numerical code.

Appendix B: Regression results without three-way error clustering

Table 6 Bilateral trade volume after two years for periods 2000-2006 without three-way error clustering

	Dependent variable: log x_{opd}^{t+2}						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ω_{opd}^{t}		0.122***			0.116***	0.202***	0.209***
		(0.001)			(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(d)}$			0.118***		0.109***	0.139***	0.143***
			(0.001)		(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(o)}$				0.041***	0.027***	0.042***	0.077***
				(0.001)	(0.001)	(0.001)	(0.001)
log x_{opd}^{t}	1.601***	1.595***	1, 600***	1.603***	1.597***	1.438***	1.371***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log x_{op}^{t}	0.937***	0.910***	0.945***	0.931***	0.915***	0.888***	0.961***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log x_{pd}^{t}	0.569***	0.578***	0.565***	0.569***	0.573***	0.514***	0.529***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log $Distance$	−0.498***	−0.490***	−0.437***	−0.478***	−0.419***	−0.596***	−0.485***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log gdp_{o}^{t}						0.150***	0.165***
						(0.001)	(0.001)
log gdp_{d}^{t}						0.193***	0.226***
						(0.001)	(0.001)
log $Population_{o}$						0.479***	0.472***
						(0.001)	(0.001)
log $Population_{d}$						0.342***	0.344***
						(0.001)	(0.001)
$Border_{od}$							0.712***
							(0.003)
$Colony_{od}$							0.052***
							(0.003)
$Language_{od}$							0.545***
							(0.002)
log$Lang.Proximity_{od}$							0.032***
							(0.001)
Constant	0.407***	9.803***	9.803***	9.803***	9.803***	9.803***	9.653***
	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584	10,911,584
Adjusted R^2	0.469	0.470	0.470	0.469	0.471	0.489	0.494
Root MSE	2.632	2.629	2.630	2.632	2.628	2.583	2.568

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7 Bilateral trade volume after two years for periods 2000–2006 (pre-financial crisis), 2007–2012 (crisis period) and 2012–2015 (recovery period) without three-way error clustering

	Dependent variable: log x_{opd}^{t+2}					
	2000–2006		2007–2012		2012–2015	
	(1)	(2)	(3)	(4)	(5)	(6)
ω_{opd}^{t}	0.116***	0.209***	0.082***	0.180***	0.054***	0.152***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(d)}$	0.109***	0.143***	0.105***	0.143***	0.123***	0.151***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(o)}$	0.027***	0.077***	0.073***	0.105***	0.087***	0.107***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log x_{opd}^{t}	1.597***	1.371***	1.835***	1.600***	2.011***	1.769***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log x_{op}^{t}	0.915***	0.961***	0.932***	0.972***	0.872***	0.915***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
log x_{pd}^{t}	0.573***	0.529***	0.454***	0.421***	0.450***	0.406***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
log $Distance$	−0.419***	−0.485***	−0.345***	−0.419***	−0.329***	−0.432***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
log gdp_{o}^{t}		0.165***		0.198***		0.203***
		(0.001)		(0.001)		(0.001)
log gdp_{d}^{t}		0.226***		0.216***		0.281***
		(0.002)		(0.001)		(0.001)
log $Population_{o}$		0.472***		0.456***		0.455***
		(0.001)		(0.001)		(0.002)
log $Population_{d}$		0.344***		0.367***		0.338***
		(0.001)		(0.001)		(0.001)
$Border_{od}$		0.712***		0.717***		0.611***
		(0.003)		(0.004)		(0.005)
$Colony_{od}$		0.052***		0.127***		0.193***
		(0.003)		(0.004)		(0.005)
$Language_{od}$		0.545***		0.441***		0.334***
		(0.002)		(0.003)		(0.003)
log$Lang.Proximity_{od}$		0.032***		0.027***		−0.001
		(0.001)		(0.001)		(0.001)
Constant	9.803***	9.653***	9.963***	9.830***	9.898***	9.7935***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	10,911,584	10,911,584	7,591,489	7,591,489	5,332,257	5,332,257
Adjusted R^2	0.471	0.494	0.496	0.516	0.539	0.558
Root MSE	2.628	2.568	2.691	2.637	2.582	2.529

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 8 Bilateral trade volume after two years for five technological categories without three-way error clustering

	Dependent variable: $\log x_{opd}^{t+2}$				
	Primary product	Resource-based manufactures	Low-tech manufactures	Medium-tech manufactures	High-tech manufactures
ω_{opd}^{t}	0.183***	0.164***	0.204***	0.203***	0.229***
	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)
$\Omega_{opd}^{(d)}$	0.152***	0.144***	0.131***	0.154***	0.128***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
$\Omega_{opd}^{(o)}$	0.098***	0.076***	0.072***	0.084***	0.078***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
$\log x_{opd}^{t}$	1.260***	1.341***	1.425***	1.348***	1.476***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
$\log x_{op}^{t}$	0.996***	0.923***	0.866***	1.006***	0.995***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)
$\log x_{pd}^{t}$	0.565***	0.490***	0.428***	0.509***	0.615***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)
$\log Distance$	−0.421***	−0.439***	−0.495***	−0.544***	−0.477***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)
$\log gdp_{o}^{t}$	0.103***	0.082***	0.166***	0.132***	0.296***
	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)
$\log gdp_{d}^{t}$	0.151***	0.184***	0.292***	0.164***	0.271***
	(0.004)	(0.002)	(0.002)	(0.004)	(0.004)
$\log Population_{o}$	0.326***	0.419***	0.482***	0.511***	0.517***
	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)
$\log Population_{d}$	0.298***	0.323***	0.339***	0.349***	0.379***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
$Border_{od}$	0.741***	0.754***	0.668***	0.733***	0.664***
	(0.010)	(0.007)	(0.006)	(0.011)	(0.011)
$Colony_{od}$	−0.112***	−0.016**	0.077***	0.037***	0.182***
	(0.011)	(0.007)	(0.006)	(0.011)	(0.012)
$Language_{od}$	0.395***	0.479***	0.553***	0.556***	0.619***
	(0.008)	(0.005)	(0.004)	(0.008)	(0.008)
$\log Lang.Proximity_{od}$	0.043***	0.038***	0.031***	0.036***	0.021***
	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)
Constant	9.445***	9.513***	9.358***	10.024***	10.140***
	(0.003)	(0.002)	(0.001)	(0.003)	(0.003)
Observations	1,127,670	2,241,432	3,314,246	1,110,342	1,049,765
Adjusted R^2	0.399	0.446	0.527	0.471	0.565
Root MSE	2.874	2.658	2.371	2.689	2.478

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix C: Fixed effects model

Table 9 Bilateral trade volume after two years with various fixed effects

	Dependent variable: $\log x_{opd}^{t+2}$			
	(1)	(2)	(3)	(4)
ω_{opd}^{t}	0.232***	0.229***	0.215***	0.235***
	(0.001)	(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(d)}$	0.158***	0.156***	0.162***	0.169***
	(0.001)	(0.001)	(0.001)	(0.001)
$\Omega_{opd}^{(o)}$	0.090***	0.103***	0.066***	0.090***
	(0.001)	(0.001)	(0.001)	(0.001)
$\log x_{opd}^{t}$	1.088***	1.186***	1.217***	1.242***
	(0.001)	(0.001)	(0.001)	(0.001)
$\log x_{op}^{t}$	1.157***	−0.876***	1.174***	1.107***
	(0.001)	(0.005)	(0.001)	(0.002)
$\log x_{pd}^{t}$	0.652***	0.744***	−0.791***	0.700***
	(0.001)	(0.002)	(0.005)	(0.001)
$\log Distance$		−0.647***	−0.687***	−0.751***
		(0.001)	(0.001)	(0.002)
$\log gdp_{o}^{t}$	−0.186***	0.070***	0.142***	0.148***
	(0.011)	(0.011)	(0.001)	(0.012)
$\log gdp_{d}^{t}$	−0.969***	0.215***	−0.334***	−0.697***
	(0.011)	(0.001)	(0.012)	(0.010)
$\log Population_{o}$	−2.134***	−1.310***	0.475***	1.028***
	(0.085)	(0.084)	(0.001)	(0.076)
$\log Population_{d}$	−1.687***	0.341***	−0.749***	1.173***
	(0.068)	(0.001)	(0.068)	(0.060)
$Border_{od}$		0.703***	0.687***	0.548***
		(0.003)	(0.003)	(0.003)
$Colony_{od}$		0.260***	0.130***	0.362***
		(0.004)	(0.004)	(0.004)
$Language_{od}$		0.606***	0.554***	0.416***
		(0.003)	(0.003)	(0.003)
$\log Lang.Proximity_{od}$		0.028***	0.035***	0.041***
		(0.001)	(0.001)	(0.001)
Fixed effects	Exporter-Importer, Year	Exporter-Product, Year	Importer-Product, Year	Exporter, Importer, Product
Observations	10,911,584	10,911,584	10,911,584	10,911,584
Adjusted R^2	0.528	0.537	0.517	0.512
Root MSE	2.480	2.458	2.509	2.524

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Appendix D: Rauch classification

Table 10 Bilateral trade volume after two years for homogeneous goods, reference priced, and differentiated products

	Dependent variable: log x_{opd}^{t+2}					
	Conservative classification			Liberal classification		
	Organized exchange	Reference priced	Differentiated	Organized exchange	Reference priced	Differentiated
ω_{opd}^{t}	0.212***	0.140***	0.214***	0.222***	0.127***	0.216***
	(0.050)	(0.024)	(0.022)	(0.042)	(0.021)	(0.022)
$\Omega_{opd}^{(d)}$	0.140***	0.122***	0.151***	0.147***	0.122***	0.148***
	(0.035)	(0.031)	(0.033)	(0.034)	(0.030)	(0.033)
$\Omega_{opd}^{(o)}$	0.111***	0.075***	0.091***	0.107***	0.080***	0.088***
	(0.025)	(0.020)	(0.030)	(0.022)	(0.021)	(0.932)
log x_{opd}^{t}	0.986***	1.395***	1.513***	1.038***	1.418***	1.526***
	(0.113)	(0.079)	(0.075)	(0.104)	(0.082)	(0.076)
log x_{op}^{t}	1.111***	0.976***	0.938***	1.117***	0.963***	0.927***
	(0.051)	(0.048)	(0.048)	(0.045)	(0.051)	(0.049)
log x_{pd}^{t}	0.594***	0.599***	0.563***	0.610***	0.582***	0.565***
	(0.070)	(0.060)	(0.043)	(0.063)	(0.057)	(0.043)
log $Distance$	−0.429***	−0.446***	−0.495***	−0.410***	−0.463***	−0.495***
	(0.061)	(0.051)	(0.042)	(0.059)	(0.047)	(0.042)
log gdp_{o}^{t}	0.108**	0.086*	0.208***	0.107**	0.089**	0.213***
	(0.049)	(0.043)	(0.044)	(0.047)	(0.043)	(0.045)
log gdp_{d}^{t}	0.139**	0.132**	−0.223***	0.137**	0.154***	0.219***
	(0.061)	(0.055)	(0.047)	(0.057)	(0.053)	(0.048)
log $Population_{o}$	0.301***	0.338***	0.501***	0.290***	0.363***	0.504***
	(0.038)	(0.044)	(0.055)	(0.044)	(0.044)	(0.380)
log $Population_{d}$	0.302***	0.243***	0.326***	0.260***	0.260***	0.323***
	(0.048)	(0.037)	(0.035)	(0.042)	(0.044)	(0.035)
$Border_{od}$	0.920***	0.851***	0.741***	0.924***	0.839***	0.732***
	(0.100)	(0.077)	(0.067)	(0.086)	(0.077)	(0.067)
$Colony_{od}$	−0.123	−0.036	0.156*	−0.142	0.012	0.152*
	(0.110)	(0.082)	(0.088)	(0.094)	(0.080)	(0.089)
$Language_{od}$	0.426***	0.472***	0.581***	0.451***	0.471***	0.582***
	(0.072)	(0.056)	(0.069)	(0.058)	(0.058)	(0.070)
log $Lang.Proximity_{od}$	0.041**	0.025	0.040*	0.040*	0.025	0.040
	(0.018)	(0.018)	(0.024)	(0.020)	(0.017)	(0.024)
Constant	9.482***	9.551***	9.624***	9.520***	9.553***	9.624***
	(0.089)	(0.059)	(0.051)	(0.073)	(0.059)	(0.052)
Observations	134,015	392,533	1,084,194	191,666	384,589	1,034,487
Adjusted R^2	0.340	0.420	0.547	0.350	0.431	0.552
Root MSE	3.401	2.852	2.496	3.295	2.828	2.475

Three-way clustering robust standard errors are reported in parentheses and *p<0.1; **p<0.05; ***p<0.01

Appendix E: Summary statistics and correlation table

Table 11 Summary statistics: 2000–2006

Statistic	N	Mean	St. Dev.	Min	Max
ω^t_{opd}	10,911,584	0.000	1.000	−3.366	26.699
$\Omega^{(d)}_{opd}$	10,911,584	0.000	1.000	−0.959	93.647
$\Omega^{(o)}_{opd}$	10,911,584	0.000	1.000	−1.172	51.786
$\log x^t_{opd}$	10,911,584	0.000	1.000	−1.339	4.086
$\log x^t_{op}$	10,911,584	0.000	1.000	−2.547	3.172
$\log x^t_{pd}$	10,911,584	0.000	1.000	−2.544	3.758
$\log Distance$	10,911,584	0.000	1.000	−3.855	1.578
$\log gdp^t_o$	10,911,584	0.000	1.000	−3.118	1.339
$\log gdp^t_d$	10,911,584	0.000	1.000	−2.477	1.541
$\log Population_o$	10,911,584	0.000	1.000	−2.474	2.344
$\log Population_d$	10,911,584	0.000	1.000	−2.324	2.768
$Border_{od}$	10,911,584	0.089	0.284	0	1
$Colony_{od}$	10,911,584	0.060	0.237	0	1
$Language_{od}$	10,911,584	0.155	0.362	0	1
$\log Lang.Proximity_{od}$	10,911,584	0.000	1.000	−0.312	5.050

Table 12 Summary statistics: 2007–2012

Statistic	N	Mean	St. Dev.	Min	Max
ω^t_{opd}	7,591,489	0.000	1.000	−3.347	26.234
$\Omega^{(d)}_{opd}$	7,591,489	0.000	1.000	−0.915	91.287
$\Omega^{(o)}_{opd}$	7,591,489	0.000	1.000	−1.171	25.338
$\log x^t_{opd}$	7,591,489	0.000	1.000	−1.374	3.977
$\log x^t_{op}$	7,591,489	0.000	1.000	−2.673	3.094
$\log x^t_{pd}$	7,591,489	0.000	1.000	−2.830	3.745
$\log Distance$	7,591,489	0.000	1.000	−3.951	1.580
$\log gdp^t_o$	7,591,489	0.000	1.000	−3.285	1.479
$\log gdp^t_d$	7,591,489	0.000	1.000	−2.633	1.637
$\log Population_o$	7,591,489	0.000	1.000	−1.990	2.328
$\log Population_d$	7,591,489	0.000	1.000	−1.916	2.749
$Border_{od}$	7,591,489	0.082	0.274	0	1
$Colony_{od}$	7,591,489	0.053	0.224	0	1
$Language_{od}$	7,591,489	0.153	0.360	0	1
$\log Lang.Proximity_{od}$	7,591,489	−0.000	1.000	−0.298	5.349

Table 13 Summary statistics: 2012–2015

Statistic	N	Mean	St. Dev.	Min	Max
ω^t_{opd}	5,332,257	0.000	1.000	−3.373	24.428
$\Omega^{(d)}_{opd}$	5,332,257	0.000	1.000	−0.965	98.891
$\Omega^{(o)}_{opd}$	5,332,257	0.000	1.000	−1.185	34.256
$\log x^t_{opd}$	5,332,257	0.000	1.000	−1.377	3.976
$\log x^t_{op}$	5,332,257	0.000	1.000	−2.787	3.116
$\log x^t_{pd}$	5,332,257	0.000	1.000	−2.900	3.591
$\log Distance$	5,332,257	0.000	1.000	−3.993	1.581
$\log gdp^t_o$	5,332,257	0.000	1.000	−3.193	1.527
$\log gdp^t_d$	5,332,257	0.000	1.000	−2.393	1.652
$\log Population_o$	5,332,257	0.000	1.000	−1.994	2.340
$\log Population_d$	5,332,257	0.000	1.000	−1.861	2.754
$Border_{od}$	5,332,257	0.078	0.268	0	1
$Colony_{od}$	5,332,257	0.052	0.222	0	1
$Language_{od}$	5,332,257	0.141	0.348	0	1
$\log Lang.Proximity_{od}$	5,332,257	0.000	1.000	−0.298	5.383

Table 14 Correlation Matrix: 2000–2006

	ω^t_{opd}	$\Omega^{(d)}_{opd}$	$\Omega^{(o)}_{opd}$	$\log x^t_{opd}$	$\log x^t_{op}$	$\log x^t_{pd}$	$\log Distance$	$\log gdp^t_o$	$\log gdp^t_d$	$\log Population_o$	$\log Population_d$	$Border_{od}$	$Colony_{od}$	$Language_{od}$	$\log Lang.Proximity_{od}$
ω^t_{opd}	1	0.038	0.080	0.132	0.215	−0.020	−0.026	−0.006	−0.055	−0.060	−0.081	−0.017	−0.025	−0.035	−0.018
$\Omega^{(d)}_{opd}$	0.038	1	0.287	0.049	−0.182	−0.009	−0.536	−0.037	0.091	−0.188	−0.150	0.229	0.013	0.044	0.095
$\Omega^{(o)}_{opd}$	0.080	0.287	1	0.064	−0.003	−0.035	−0.457	0.123	0.025	−0.176	−0.123	0.126	0.005	−0.068	0.100
$\log x^t_{opd}$	0.132	0.049	0.064	1	0.459	0.344	−0.117	0.170	0.159	0.135	0.104	0.110	0.063	0.019	0.132
$\log x^t_{op}$	0.215	−0.182	−0.003	0.459	1	0.171	0.239	0.345	−0.120	0.288	−0.083	−0.190	−0.028	−0.163	0.015
$\log x^t_{pd}$	−0.020	−0.009	−0.035	0.344	0.171	1	0.078	−0.038	0.486	−0.140	0.331	−0.072	−0.026	−0.083	0.085
$\log Distance$	−0.026	−0.536	−0.457	−0.117	0.239	0.078	1	−0.015	−0.066	0.285	0.211	−0.486	−0.042	−0.044	−0.217
$\log gdp^t_o$	−0.006	−0.037	0.123	0.170	0.345	−0.038	−0.015	1	0.016	−0.338	−0.041	−0.115	0.029	−0.072	0.110
$\log gdp^t_d$	−0.055	0.091	0.025	0.159	−0.120	0.486	−0.066	0.016	1	−0.084	−0.148	−0.032	−0.001	−0.065	0.138
$\log Population_o$	−0.060	−0.188	−0.176	0.135	0.288	−0.140	0.285	−0.338	−0.084	1	−0.013	−0.052	0.027	−0.044	0.004
$\log Population_d$	−0.081	−0.150	−0.123	0.104	−0.083	0.331	0.211	−0.041	−0.148	−0.013	1	−0.022	0.019	−0.022	0.058
$Border_{od}$	−0.017	0.229	0.126	0.110	− 0.190	−0.072	−0.486	−0.115	−0.032	−0.052	−0.022	1	0.130	0.170	0.114
$Colony_{od}$	−0.025	0.013	0.005	0.063	−0.028	−0.026	−0.042	0.029	−0.001	0.027	0.019	0.130	1	0.264	0.071
$Language_{od}$	−0.035	0.044	−0.068	0.019	−0.163	−0.083	−0.044	−0.072	−0.065	−0.044	−0.022	0.170	0.264	1	−0.103
$\log Lang.Proximity_{od}$	−0.018	0.095	0.100	0.132	0.015	0.085	−0.217	0.110	0.138	0.004	0.058	0.114	0.071	−0.103	1

Table 15 Correlation matrix: 2007–2012

	ω^t_{opd}	$\Omega^{(d)}_{opd}$	$\Omega^{(o)}_{opd}$	$\log x^t_{opd}$	$\log x^t_{op}$	$\log x^t_{pd}$	$\log Distance$	$\log gdp^t_o$	$\log gdp^t_d$	$\log Population_o$	$\log Population_d$	$Border_{od}$	$Colony_{od}$	$Language_{od}$	$\log Lang.Proximity_{od}$
ω^t_{opd}	1	0.036	0.083	0.139	0.233	−0.012	−0.030	−0.001	−0.049	−0.066	−0.086	−0.018	−0.024	−0.027	−0.015
$\Omega^{(d)}_{opd}$	0.036	1	0.273	0.059	−0.161	−0.022	−0.512	−0.024	0.093	−0.179	−0.153	0.217	0.013	0.053	0.087
$\Omega^{(o)}_{opd}$	0.083	0.273	1	0.086	−0.004	−0.036	−0.452	0.139	0.061	−0.167	−0.135	0.129	0.004	−0.050	0.101
$\log x^t_{opd}$	0.139	0.059	0.086	1	0.481	0.343	−0.128	0.137	0.138	0.166	0.100	0.126	0.057	0.004	0.132
$\log x^t_{op}$	0.233	−0.161	−0.004	0.481	1	0.206	0.219	0.306	−0.107	0.298	−0.088	−0.168	−0.028	−0.182	0.021
$\log x^t_{pd}$	−0.012	−0.022	−0.036	0.343	0.206	1	0.078	−0.036	0.414	−0.149	0.342	−0.063	−0.017	−0.091	0.083
$\log Distance$	−0.030	−0.512	−0.452	−0.128	0.219	0.078	1	−0.043	−0.104	0.290	0.224	−0.477	−0.046	−0.053	−0.213
$\log gdp^t_o$	−0.001	−0.024	0.139	0.137	0.306	−0.036	−0.043	1	0.034	−0.349	−0.053	−0.090	0.038	−0.117	0.131
$\log gdp^t_d$	−0.049	0.093	0.061	0.138	−0.107	0.414	−0.104	0.034	1	−0.097	−0.177	−0.014	0.012	−0.096	0.160
$\log Population_o$	−0.066	−0.179	−0.167	0.166	0.298	−0.149	0.290	−0.349	−0.097	1	−0.012	−0.046	0.025	−0.034	0.001
$\log Population_d$	−0.086	−0.153	−0.135	0.100	−0.088	0.342	0.224	−0.053	−0.177	−0.012	1	−0.020	0.022	−0.007	0.045
$Border_{od}$	−0.018	0.217	0.129	0.126	−0.168	−0.063	−0.477	−0.090	−0.014	−0.046	−0.020	1	0.126	0.169	0.111
$Colony_{od}$	−0.024	0.013	0.004	0.057	−0.028	−0.017	−0.046	0.038	0.012	0.025	0.022	0.126	1	0.252	0.070
$Language_{od}$	−0.027	0.053	−0.050	0.004	−0.182	−0.091	−0.053	−0.117	−0.096	−0.034	−0.007	0.169	0.252	1	−0.099
$\log Lang.Proximity_{od}$	−0.015	0.087	0.101	0.132	0.021	0.083	−0.213	0.131	0.160	0.001	0.045	0.111	0.070	−0.099	1

Table 16 Correlation matrix: 2012–2015

	ω^t_{opd}	$\Omega^{(d)}_{opd}$	$\Omega^{(o)}_{opd}$	$\log x^t_{opd}$	$\log x^t_{op}$	$\log x^t_{pd}$	$\log Distance$	$\log gdp^t_o$	$\log gdp^t_d$	$\log Population_o$	$\log Population_d$	$Border_{od}$	$Colony_{od}$	$Language_{od}$	$\log Lang.Proximity_{od}$
ω^t_{opd}	1	0.033	0.084	0.143	0.235	−0.006	−0.025	−0.012	−0.044	−0.064	−0.084	−0.021	−0.023	−0.026	−0.014
$\Omega^{(d)}_{opd}$	0.033	1	0.283	0.069	−0.167	−0.023	−0.542	−0.036	0.095	−0.180	−0.154	0.236	0.017	0.044	0.095
$\Omega^{(o)}_{opd}$	0.084	0.283	1	0.082	−0.015	−0.042	−0.440	0.118	0.059	−0.161	−0.132	0.125	0.002	−0.053	0.101
$\log x^t_{opd}$	0.143	0.069	0.082	1	0.480	0.350	−0.125	0.109	0.128	0.185	0.108	0.138	0.061	0.022	0.124
$\log x^t_{op}$	0.235	−0.167	−0.015	0.480	1	0.213	0.222	0.271	−0.124	0.309	−0.083	−0.157	−0.031	−0.162	0.009
$\log x^t_{pd}$	−0.006	−0.023	−0.042	0.350	0.213	1	0.085	−0.055	0.387	−0.148	0.349	−0.054	−0.021	−0.073	0.076
$\log Distance$	−0.025	−0.542	−0.440	−0.125	0.222	0.085	1	−0.010	−0.093	0.283	0.227	−0.472	−0.051	−0.042	−0.215
$\log gdp^t_o$	−0.012	−0.036	0.118	0.109	0.271	−0.055	−0.010	1	0.019	−0.332	−0.048	−0.089	0.030	−0.104	0.116
$\log gdp^t_d$	−0.044	0.095	0.059	0.128	−0.124	0.387	−0.093	0.019	1	−0.101	−0.169	−0.006	0.008	−0.075	0.158
$\log Population_o$	−0.064	−0.180	−0.161	0.185	0.309	−0.148	0.283	−0.332	−0.101	1	−0.007	−0.035	0.028	−0.005	−0.001
$\log Population_d$	−0.084	−0.154	−0.132	0.108	−0.083	0.349	0.227	−0.048	−0.169	−0.007	1	−0.020	0.021	0.003	0.042
$Border_{od}$	−0.021	0.236	0.125	0.138	−0.157	−0.054	−0.472	−0.089	−0.006	−0.035	−0.020	1	0.136	0.165	0.114
$Colony_{od}$	−0.023	0.017	0.002	0.061	−0.031	−0.021	−0.051	0.030	0.008	0.028	0.021	0.136	1	0.260	0.070
$Language_{od}$	−0.026	0.044	−0.053	0.022	−0.162	−0.073	−0.042	−0.104	−0.075	−0.005	0.003	0.165	0.260	1	−0.092
$\log Lang.Proximity_{od}$	−0.014	0.095	0.101	0.124	0.009	0.076	−0.215	0.116	0.158	−0.001	0.042	0.114	0.070	−0.092	1

Appendix F: Relationship between bilateral trade volume after two years and the three learning channels by exporters' comparative advantage

We also test the effects of the exporters' levels of competitiveness on the diffusion of the information needed to trade by dividing exporters of each product into small, medium, and large exporters (Table 17). We do this by calculating the revealed comparative advantage (RCA) of each exporter in each product. RCA is the ratio between the exports of a country in a product, and the exports that are expected based on a country's total export market and the size of the global market for that product. We classify as small exporters all countries with an RCA below 0.2 in a product (countries that export less than 20% of what they are expected to export by chance). We classify as medium exporters all countries with an RCA between 0.2 and 1. We classify as the large exporters of a product all countries that have revealed comparative advantage in it (RCA $>$ 1). To rule out temporary changes of exporters' comparative advantage, we restrict the condition for being a small, medium, and large exporter: small, medium, and large exporters need to keep RCA below 0.2, between 0.2 and 1, and above 1 for three years before the beginning of the period.

Table 17 divide country-product pairs into small, medium, and large exporters. The results are consistent with those presented in Table 2, but also reveal two important distinctions. First, the effects of product and geographic relatedness, especially exporter relatedness, are stronger for small exporters, suggesting that knowledge and information frictions impose larger constraints for countries that are not exporting a product on a large scale. Second, the overall explanatory power of the model is considerably larger for large exporters ($R^2 \approx 53\%$ vs $R^2 \approx 46\%$ for medium exporters and $R^2 \approx 28\%$ for small exporters; these are large differences, even considering that the sample sizes are not the same). This suggests that smaller exporters face more uncertainty (less predictable because of lower R-squared), and, hence, other factors are needed to predict their bilateral trade volume.

Table 17 Bilateral trade volume after two years by different levels of exporters' comparative advantage

	Dependent variable: $\log x_{opd}^{t+2}$		
	(1) Small	(2) Medium	(3) Large
ω_{opd}^{t}	0.165***	0.135***	0.201***
	(0.015)	(0.021)	(0.021)
$\Omega_{opd}^{(d)}$	0.089***	0.122***	0.150***
	(0.027)	(0.024)	(0.023)
$\Omega_{opd}^{(o)}$	0.126***	0.086**	0.068***
	(0.038)	(0.034)	(0.018)
$\log x_{opd}^{t}$	0.575***	1.481***	1.482***
	(0.103)	(0.070)	(0.057)
$\log x_{op}^{t}$	0.990***	0.831***	0.961***
	(0.047)	(0.051)	(0.035)

Table 17 (continued)

	Dependent variable: log x_{opd}^{t+2}		
	(1) Small	(2) Medium	(3) Large
log x_{pd}^{t}	0.484***	0.558***	0.586***
	(0.036)	(0.054)	(0.030)
log $Distance$	−0.499***	−0.486***	−0.471***
	(0.038)	(0.046)	(0.037)
log gdp_o^t	0.251***	0.143***	0.109***
	(0.057)	(0.045)	(0.039)
log gdp_d^t	0.010	0.045	0.265***
	(0.052)	(0.048)	(0.042)
log $Population_o$	0.396***	0.408***	0.456***
	(0.052)	(0.058)	(0.044)
log $Population_d$	0.189***	0.215***	0.364***
	(0.059)	(0.040)	(0.030)
$Border_{od}$	0.651***	0.817***	0.713***
	(0.076)	(0.081)	(0.074)
$Colony_{od}$	0.218*	0.259**	0.097
	(0.129)	(0.127)	(0.076)
$Language_{od}$	0.557***	0.782***	0.481***
	(0.071)	(0.073)	(0.070)
log$Lang.Proximity_{od}$	0.012	0.018	0.027
	(0.031)	(0.031)	(0.019)
Constant	9.272***	9.414***	9.690***
	(0.096)	(0.070)	(0.047)
Observations	922,092	463,388	8,045,262
Adjusted R^2	0.281	0.456	0.526
Root MSE	2.641	2.5349	2.487

Three-way clustering robust standard errors are reported in parentheses and $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

References

Anderson JE (1979) A theoretical foundation for the gravity equation. Am Econ Rev 69(1):106–116

Anderson JE, Marcouiller D (2002) Insecurity and the pattern of trade: an empirical investigation. Rev Econ Stat 84(2):342–352

Bahar D, Hausmann R, Hidalgo CA (2014) Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? J Int Econ 92(1):111–123

Bailey M, Cao RR, Kuchler T, Stroebel J, Wong A (2017) Measuring social connectedness. NBERWorking Paper No. 23608

Balassa B (1965) Trade liberalisation and "revealed" comparative advantage. Manchester School 33(2):99–123

Boschma R, Balland P-A, Kogler DF (2014) Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. Ind Corp Chang 24(1):223–250

Boschma R, Minondo A, Navarro M (2013) The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. Econ Geogr 89(1):29–51
Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J Econ Geogr 9(4):439–468
Casella A, Rauch JE (2002) Anonymous market and group ties in international trade. J Int Econ 58(1):19–47
Chaney T (2014) The network structure of international trade. Am Econ Rev 104(11):3600–3634
Combes P-P, Lafourcade M, Mayer T (2005) The trade-creating effects of business and social networks: Evidence from France. J Int Econ 66(1):1–29
Correia S (2015) Singletons, cluster-robust standard errors and fixed effects: a bad mix. Technical Note, Duke University
Feenstra RC, Lipsey RE, Deng H, Ma AC, Mo H (2005) World trade flows: 1962-2000. NBER Working Paper No. 11040
Gao J, Jun B, Pentland A, Zhou T, Hidalgo CA (2017) Collective learning in China's regional economic development. arXiv:1703.01369
Garmendia A, Llano C, Minondo A, Requena F (2012) Networks and the disappearance of the intranational home bias. Econ Lett 116(2):178–182
Guevara MR, Hartmann D, Aristarán M, Mendoza M, Hidalgo CA (2016) The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics 109(3):1695–1709
Hausmann R, Hidalgo CA, Bustos S, Coscia M, Simoes A, Yildirim MA (2014) The atlas of economic complexity: Mapping paths to prosperity. MIT Press, Cambridge
Heckscher EF, Ohlin BG (1991) Heckscher-Ohlin trade theory. MIT Press, Cambridge
Helpman E (1987) Imperfect competition and international trade: Evidence from fourteen industrial countries. Journal of the Japanese and International Economies 1(1):62–81
Hidalgo CA (2015) Why information grows: the evolution of order, from atoms to economies. Basic Books, New York
Hidalgo CA, Balland P-A, Boschma R, Delgado M, Feldman M, Frenken K, Glaeser E, He C, Kogler DF, Morrison A, et al. (2018) The principle of relatedness. In: International conference on complex systems. Springer, pp 451–457
Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. Proc Natl Acad Sci USA 106(26):10570–10575
Hidalgo CA, Klinger B, Barabási A-L, Hausmann R (2007) The product space conditions the development of nations. Science 317(5837):482–487
Kogler DF, Rigby DL, Tucker I (2013) Mapping knowledge space and technological relatedness in US cities. Eur Plan Stud 21(9):1374–1391
Krugman P (1979) Increasing returns, monopolistic competition, and international trade. J Int Econ 9(4):469–479
Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99(3):483–499
Lall S (2000) The technological structure and performance of developing country manufactured exports, 1985-98. Oxf Dev Stud 28(3):337–369
Mayer T, Zignago S (2011) Notes on CEPII's distances measures: the geodist database. CEPII Working Paper No. 2011-25
McCallum J (1995) National borders matter: Canada-US regional trade patterns. Am Econ Rev 85(3):615–623
Morales E, Sheu G, Zahler A (2015) Extended gravity. NBER Working Paper 21351
Neffke F, Henning M, Boschma R (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. Econ Geogr 87(3):237–265
Parsons C, Vézina PL (2018) Migrant networks and trade: the Vietnamese boat people as a natural experiment[J]. Econ J 128(612):F210–F234
Petropoulou D (2008) Information costs, networks and intermediation in international trade. CEP Discussion Paper No. 848
Portes R, Rey H (2005) The determinants of cross-border equity flows. J Int Econ 65(2):269–296
Pöyhönen P (1963) A tentative model for the volume of trade between countries. Weltwirtschaftliches Archiv, 93–100
Rauch JE (1999) Networks versus markets in international trade. J Int Econ 48(1):7–35
Rauch JE (2001) Business and social networks in international trade. J Econ Lit 39(4):1177–1203

Rauch JE, Trindade V (2002) Ethnic Chinese networks in international trade. Rev Econ Stat 84(1):116–130

Ronen S, Gonçalves B, Hu KZ, Vespignani A, Pinker S, Hidalgo CA (2014) Links that speak: the global language network and its association with global fame. Proc Natl Acad Sci USA 111(52):E5616–E5622

Simoes AJG, Hidalgo CA (2011) The economic complexity observatory: An analytical tool for understanding the dynamics of economic development. In: Proceedings of the 17th AAAI conference on scalable integration of analytics and visualization. AAAIWS'11-17. AAAI Press, pp 39–42

Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. Manag Sci 51(5):756–770

Tinbergen J (1962) Shaping the world economy. Twentieth Century Fund, New York