INTERNATIONAL TRADE AND SOCIAL CONNECTEDNESS*

MICHAEL BAILEY†
ABHINAV GUPTA‡
SEBASTIAN HILLENBRAND§
THERESA KUCHLER¶
ROBERT RICHMOND||
JOHANNES STROEBEL**

Abstract

We use anonymized data from Facebook to construct a new measure of pairwise social connectedness between 180 countries and between 332 European regions. We show that social connectedness explains a large part of the variation in country-level trade flows. Controlling for social connectedness substantially reduces the estimated effects of geographic distance and country borders on trade volumes. We also find that countries trade more with each other when they share social connections with a similar set of third countries. We show that social connectedness increases trade by reducing information asymmetries and by providing a substitute for both trust and formal mechanisms of contract enforcement; we also present evidence against omitted variables and reverse causality as alternative explanations for the observed relationship between social connectedness and trade flows. We use data on social connectedness at the regional level to show that the social connections that determine international trade in each product are those between the regions where the product is produced in the exporting country and those where it is used in the importing country.

JEL Codes: F1, F5, F6
Keywords: International Trade, Social Connectedness, Contract Enforcement, Information Frictions

*This draft: December 10, 2019. We thank Nick Bloom, Tarek Hassan, Peter Koudijs, Paola Sapienza, Alexi Savov, Tim Schmidt-Eisenlohr, Chenzi Xu, and Moto Yogo for helpful comments. The authors would like to thank Dominic Russel and Steven Zheng for superb research assistance.

†Facebook, Inc: mcbailey@fb.com.
‡Stern School of Business, New York University: agupta4@stern.nyu.edu.
§Stern School of Business, New York Universit: shillenb@stern.nyu.edu.
¶Stern School of Business, New York University, CEPR: tkuchler@stern.nyu.edu.
||Stern School of Business, New York University: rrichmon@stern.nyu.edu.
**Stern School of Business, New York University, NBER, CEPR: johannes.stroebel@nyu.edu.

The propensity for residents of different countries to be connected to one another varies enormously. For example, a U.S.-based Facebook user is 65% more likely to be friends with a given Facebook user living in Germany than with a given Facebook user living in France. Such differences in bilateral social connectedness play an important role in many narratives of economic and political interactions between countries. For example, beginning with Tinbergen (1962), researchers have explored the determinants of international trade using gravity models that relate trade flows between countries to various measures of the relationship between those countries.[1] These models have had substantial empirical success, but their economic underpinnings — and especially the mechanism behind the large estimated negative effect of distance on trade — have remained elusive. One prominent narrative is that geographic proximity proxies for social connections between individuals, which can help facilitate trade. While such a mechanism is intuitively appealing, the absence of comprehensive data on social connections across countries has limited researchers' ability to provide evidence in favor of this interpretation.

In this paper, we introduce a new measure of the pairwise social connectedness between 180 countries and show that a large amount of the variation in global trade flows can indeed be explained by patterns of direct and indirect social connectedness. We also present evidence that the effect of social connectedness on trade comes through at least two channels: by reducing information asymmetries and by providing a substitute for both trust and formal mechanisms of contract enforcement. Our measure of social connectedness is based on an anonymized snapshot of friendship links on Facebook, the world's largest online networking site with more than 2.4 billion active users around the globe. Our *Social Connectedness Index* between countries corresponds to the relative probability of friendship links between Facebook users across country pairs.[2] We argue that social networks as measured by Facebook provide a reasonable representation of real-world friendship networks. This is the result of Facebook's scale, the relative representativeness of its user body, and the fact that people primarily use Facebook to interact with their real-world friends and acquaintances.

We begin by describing the rich patterns of international social connectedness observed in our data. Social connectedness across countries varies with a number of physical and cultural characteristics. About half the variation in social connectedness between countries is explained by geographic distance. Quantitatively, a 10% increase in the geographic distance between two countries is associated with a $10\% - 15\%$ decline in their social connectedness. We find that sharing a common border and being on the same continent are additional determinants of social connectedness. International migration patterns and colonial history further influence the probability of present-day friendship links across country pairs. We also find stronger social connections between countries that share a common

[1]It is impossible to do full justice to the large literature that has studied the determinants of trade across countries. In addition to the papers we reference below, prominent contributions that have explored the role of geographic factors such as distance on trade include Leamer and Levinsohn (1995), Trefler (1995), Obstfeld and Rogoff (2000), Eaton and Kortum (2002), and Hortaçsu et al. (2009). Allen (2014), Chaney (2014), and Steinwender (2018) have studied the role of information and search frictions, while Berkowitz et al. (2006), Nunn (2007), Levchenko (2007), and Ranjan and Lee (2007) have analyzed the role of contract enforcement frictions in determining trade flows. Melitz (2003), Bernard et al. (2003), Chaney (2008, 2018), Melitz and Ottaviano (2008), Helpman et al. (2008), and Schmidt-Eisenlohr (2013) have explored the role of firms in trade. We also contribute to a literature which has proxied for the extent of social connections using measures such as ethnic and business networks (see Rauch, 1999, 2001; Rauch and Trindade, 2002; Combes et al., 2005).

[2]This new and comprehensive measure of international social connectedness can be shared with other researchers, who are invited to submit a 1-page research proposal to sci_data@fb.com. See Bailey et al. (2018b, 2019a) for a description of a related data set measuring the social connectedness between U.S. counties, and between zip codes in the New York metro area.

language, as well as between countries that are similar in terms of economic development, religious beliefs, and the genetic make-up of their populations.

In the second part of the paper, we document that social connectedness is an important determinant of bilateral trade flows. When we introduce social connectedness into a standard gravity model of trade, we find substantial effects. While 83.3% of the variation in bilateral trade is explained by importer and exporter fixed effects that control for factors such as the size of the countries' economies, a model that additionally introduces social connectedness has an R^2 of 91.9%. Indeed, both social connectedness and geographic distance explain similar shares of cross-sectional variation in trade flows. The elasticity of trade flows with respect to social connectedness is 0.33 in specifications that also control for geographic distance. This implies that, all else equal, trade between the U.S. and Germany should be 21% higher than trade between the U.S. and France, since the U.S. is 65% more connected to Germany than it is to France. Controlling for social connectedness reduces the distance elasticity of trade from approximately -1 to about -0.66. This is a substantial decline that does not occur when controlling for other gravity variables such as common language or common colonial origins.[3] The combined evidence suggests that social connectedness is an important determinant of bilateral trade flows, and that the estimates of the relationship between geographic distance and trade in the prior literature partially capture this important role of social connectedness.

We then explore the mechanisms through which social connectedness might help facilitate international trade. In particular, the set of informal trade barriers that are regularly discussed in the literature include contract enforcement frictions and search costs due to information asymmetries (see Chaney, 2016). With this evidence in mind, we show that social connections between countries facilitate trade through at least two channels: by reducing information frictions and by providing a substitute for formal mechanisms of contract enforcement.

We first study how the elasticity of trade with respect to social connectedness varies with measures of the rule of law in the importing and exporting countries. This analysis is motivated by a literature in international trade that finds that weak rule of law reduces trade due to difficulties with the enforcement of contracts (e.g., Anderson and Marcouiller, 2002). We find that social connectedness tends to increase exports particularly to those countries with a weak rule of law, suggesting that one channel through which social connections help to facilitate trade is by alleviating contract enforcement frictions. We also study variation in the effect of social connectedness on trade across different products. In particular, Rauch (1999) and Rauch and Trindade (2002) suggest that information asymmetries are largest for products not traded on organized exchanges. Consistent with the idea that social connections can help to mitigate such information asymmetries, we find that the elasticity of trade with respect to social connectedness is particularly large for these non-exchange-traded goods.

While the direct social connectedness between two countries is an important determinant of their bilateral trade, trade flows might further increase if the countries are also connected to a similar set of other countries. Indeed, being in the same social cluster of countries can increase bilateral trade through similar mechanisms as direct social connections. The first mechanism is by decreasing search costs,

[3]In related work, Portes and Rey (2005) also find a large decrease in the distance effect on equity flows when controlling for proxies for informational barriers using phone call volume for 14 countries.

whereby a common friend in a third country passes information between potential trading partners.[4] The second channel is by improving contract enforcement: contract default is less likely in a shared social community with threat of imposing sanctions (see Greif, 1989, 1993). To test the hypothesis that being in the same social cluster increases bilateral trade, we use a clustering algorithm to group countries into socially connected communities. We find that countries in the same social cluster have substantially higher trade than would be predicted based on only their bilateral social connectedness. For example, Spain is approximately equally well-connected to the United Kingdom and Argentina, but it only shares a cluster with Argentina. Our estimates imply that, all else equal, Spain would thus trade 18% more with Argentina than it would trade with the United Kingdom.

Our evidence suggests that social connectedness has a direct effect on international trade, both by reducing information frictions as well as by facilitating the enforcement of contracts. We next exploit information on social connectedness at the sub-national level to show that other explanations, such as omitted variables or reverse causality, are unlikely to explain the observed aggregate patterns. This section builds on a new data set of social connectedness between 332 NUTS2 regions in Europe (see Bailey et al., 2019c, for a detailed description of this data set). Our findings also highlight that social connections at the firm level are central to influencing trade flows.[5]

Our approach builds on a literature that documents that firms (and individuals working at those firms) are central to facilitating international trade (see Bernard et al., 2007, 2012). Based on this insight, we construct product-specific measures of social connectedness between countries, which overweight the social connectedness between those regions where the products are produced in the exporting country and those regions where the goods are used in the importing country. This measure contrasts with our baseline measure of social connectedness, which corresponds to the population-weighted average connectedness between all regions in the two countries. As an example, more than 80% of Italian exports of non-metallic mineral products to Greece are used as inputs in the Greek construction sector. Our measure of social connectedness relevant for exporting non-metallic mineral products from Italy to Greece thus overweights connectedness between the regions that produce non-metallic mineral products in Italy (primarily the Piedmont region around Torino) and the regions with significant construction employment in Greece (primarily the Attica region around Athens). If social connectedness truly affected trade, we would expect these product-specific measure of social connectedness to matter more for trade in that specific product than the population-weighted social connectedness. On the other hand, if trade between countries was primarily driven by omitted variables such as common preferences that are correlated with social connectedness, we would expect the population-weighted social connectedness to be a more accurate representation of the average similarity in preferences across countries.

We then regress product-level trade between countries on our measures of social connectedness. When controlling for the product-specific measure of social connectedness, our baseline population-

[4] Chaney (2014) presents a model and supporting empirical evidence using French firm-level data that is consistent with social clusters influencing trade. Albornoz et al. (2012) and Morales et al. (2015) also present models where exporters sequentially choose export destinations that are also consistent with our findings.

[5] This follows the direction of an exciting literature that applies research designs built on naturally occurring variation to networks to identify the causal impact of social networks on economic quantities. For example, Burchardi and Hassan (2013) use the fall of the Berlin wall as a natural experiment to show how social connections can lead to growth. Burchardi et al. (2016) study the causal effect of migration flows on foreign direct investment.

weighted measure of social connectedness has no further predictive power for trade at the product-level. This remains true after controlling for product-specific measures of distance, and after including non-linear product specific distance controls. The elasticity of trade to our product-specific measures of social connectedness is also unaffected by the inclusion of country-pair fixed effects (which absorb the population-weighted measures of social connectedness in addition to any other observed or unobserved determinants of trade between countries). This evidence suggests that it is indeed the social connectedness between specific regions where products are produced and used that matters for trade.

Our analysis of the product-specific social connectedness between countries also allows us to rule out the presence of a quantitatively large reverse causality from trade to connectedness. If trade were to cause substantial social connectedness, we would expect the various product-specific measures of social connectedness between countries to be systematically larger than these countries' measures of population-weighted social connectedness. For instance, in the example above, we would expect the Piedmont region in Italy and the Attica region in Greece to be disproportionately more connected than a random pair of regions in the two countries. In contrast with this prediction, we find that the regions that are most important for the trade in a given product are equally likely to be more or less connected than the population-weighted average of regions across a country pair.

In the final part of the paper, we explore the relationship between social connectedness and trade at the sub-national level. We use rail-freight flows between regions in the European Union as our proxy of trade flows. We first explore the determinants of the border effect — the empirical irregularity that, conditional on the distance between two regions, trade is much larger when regions are in same country (see McCallum, 1995; Anderson and Van Wincoop, 2003; Chen, 2004). Consistent with existing estimates, we find that, all else equal, trade within countries is seven times as large as trade across countries. This is true despite the fact that the European Union is a common market with few formal barriers to trade. When we control for the social connectedness between regions, the border effect drops by more than 80% and becomes statistically indistinguishable from zero. This suggests that much of the impact of borders on trade is related to the fact that social connections are reduced at borders.

The region-level data on trade and social connectedness also allows us to understand how social connectedness and trust interact to influence trade patterns. In particular, Guiso et al. (2009) highlight an important role of affinity and trust in facilitating the flow of goods and capital. To test whether trust and social connectedness are substitutes or complements in facilitating trade, we study how the social connectedness elasticity of trade varies across regions that trust one another differentially, as measured by Guiso et al. (2009). We find that the elasticity of trade to social connectedness is higher across regions that have low levels of trust. This suggests that social connectedness and trust are substitutes in their effects on trade, similar to our findings that social connectedness and a strong rule of law are substitutes.

The rest of the paper is organized as follows. Section 1 describes our new measure of international social connectedness. Section 2 explores the determinants of international social connectedness. Section 3 presents results on the relationship between international trade (both aggregate and product-specific) and social connectedness. In Section 4, we present results using our product-specific measures of social connectedness and explore patterns in regional trade within Europe. The final section concludes.

1 Measuring International Social Connectedness

We construct our measure of the social connectedness between countries using anonymized administrative data from Facebook, a global online social networking service. Facebook was created in 2004, and, by the end of the first quarter of 2019, had 2.4 billion monthly active users globally. Of these users, 243 million were based in the United States and Canada, 384 million in Europe, 981 million in the Asia-Pacific region, and 768 million in the rest of the world. With the exception of a few countries where social media services including Facebook are banned — most notably China, Iran, and North Korea — Facebook has a non-trivial footprint in essentially all major countries around the world.

We work with an anonymized snapshot of all active Facebook users from March 2019. For these users, we observe their country of location as well as the set of other Facebook users that they are connected to. Countries of location are assigned based on users' information and activity on Facebook, including the stated city on their Facebook profile and device and connection information. To compare the intensity of social connectedness between countries with varying populations and varying Facebook usage rates, we construct our *Social Connectedness Index*, $SCI_{i,j}$, as the total number of connections between individuals in country i and individuals in country j, divided by the product of the number of Facebook users in those countries, as in Equation 1.

$$SCI_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i \times FB\_Users_j}. \tag{1}$$

We rescale this number to have a minimum value of 1, and a maximum value of 1,000,000. The *Social Connectedness Index* therefore measures the relative probability of a Facebook friendship link between a given user in country i and a given user in country j. We exclude countries that have very low populations, are territories of other countries, or have a ban on Facebook. We also exclude countries for which we do not have other control data available.[6] Overall, this provides us with the pairwise social connectedness between 180 countries, for a total of $180 \times 179 = 32,220$ country-pair combinations.

Interpreting the Social Connectedness Index. Two important issues arise when interpreting $SCI_{i,j}$ as a proxy for the social connectedness between countries: whether Facebook friendships correspond to real-world friendship links of Facebook users, and whether Facebook users are representative of the countries' populations.

On the first issue, we believe that Facebook friendships provide a reasonable proxy for real world friendship networks. For the United States, Duggan et al. (2015) have shown that Facebook friendship patterns correspond quite closely to real-world friendship networks. While similar studies do not exist for most other countries, we believe that there are a number of reasons to think that we are capturing reasonable representations of real-world social networks of Facebook users outside of the United States. For example, establishing a connection on Facebook requires the consent of both individuals,

[6] Due to a ban of Facebook, data is not available for China, Iran, North Korea, Tajikistan, and Turkmenistan. Due to low populations, data are also not available for Andorra, Dominica, Kiribati, St. Kitts and Nevis, San Marino and Tuvalu. Due to a lack of other data (e.g. lack of data on the gravity variables) we exclude Montenegro, Serbia, South Sudan, Kosovo. Finally, we exlude the following territories: Curacao, Guam, Isle of Man, Jersey, Mayotte, Guadeloupe, French Guiana, Martinique, Puerto Rico, Reunion, and Western Sahara.

and there is an upper limit of 5,000 on the number of connections a person can have. As a result, networks formed on Facebook will more closely resemble real-world social networks than those on other online platforms, such as Twitter, where uni-directional links to non-acquaintances, such as celebrities, are common. Consistent with this conclusion, our prior work with micro-data from Facebook has found that many economic decisions, such as whether to buy a house or which phone to purchase, were influenced by related decisions of a person's Facebook friends (Bailey et al., 2018c, 2017, 2019b).

On the second issue, it is likely that the representativeness of Facebook users will differ across countries. While Duggan et al. (2016) have shown that U.S. Facebook users are quite representative of the U.S. population, this is unlikely to be the case everywhere. For example, in countries with relatively low internet penetration, those individuals with access to internet are likely a non-representative subset of the overall population. To the extent that having internet access and having friends abroad are positively correlated, our measure would overstate the international linkages of the average resident in countries with low internet usage. In our analysis, we account for such heterogeneity across countries by including country fixed effects in all specifications. This allows us to explore connectedness between countries i and j, holding fixed the average propensity of having Facebook friends abroad in each country.

While no measure of social connectedness is perfect, we believe that our *Social Connectedness Index*, which is based on hundreds of billion of Facebook friendship links from 2.4 billion Facebook users, presents a substantial improvement over alternative measures.

2 The Determinants of International Social Connectedness

In this section, we examine which factors correlate with the strength of social connectedness between countries. We first explore a number of case studies of the social connectedness of individual counties. We then use a regression framework to formally analyze the role of geography and other factors likely to play a role, including similarity of language, history, religion, and economic outcomes.

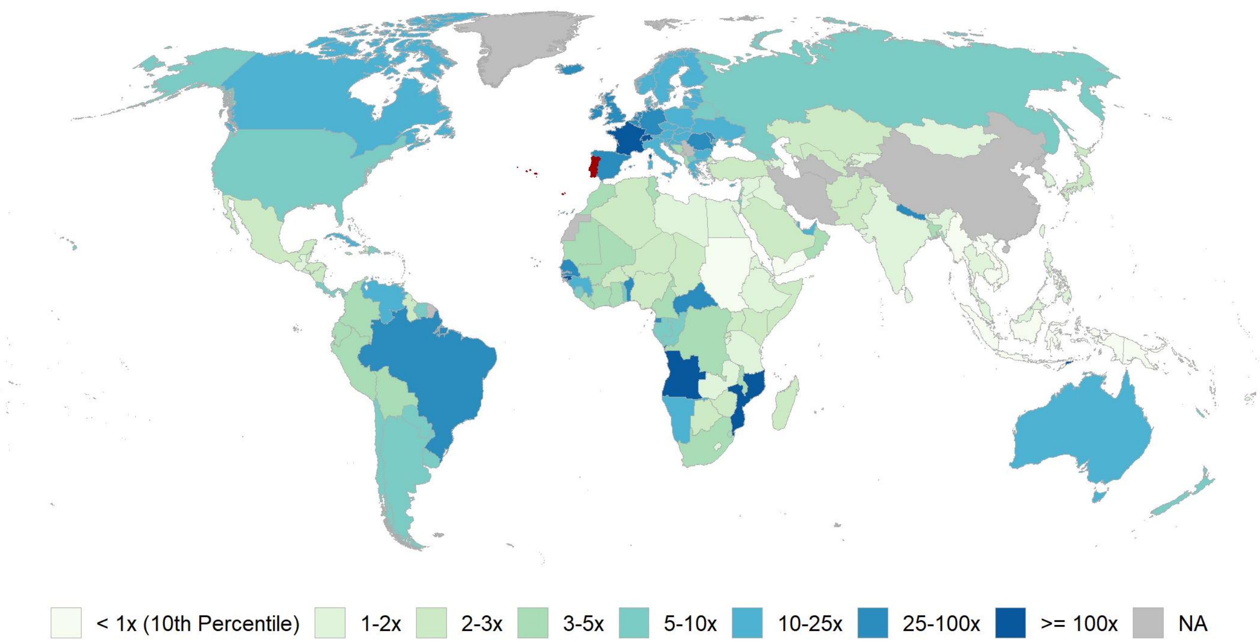
2.1 Case Studies of International Social Connectedness

Figures 1 and 2 show heat maps of the social connections of several countries. Darker colors correspond to closer connections. Each of these case studies reveals a number of interesting and generalizable patterns. Appendix A.2 provides additional examples of country-level social connectedness.

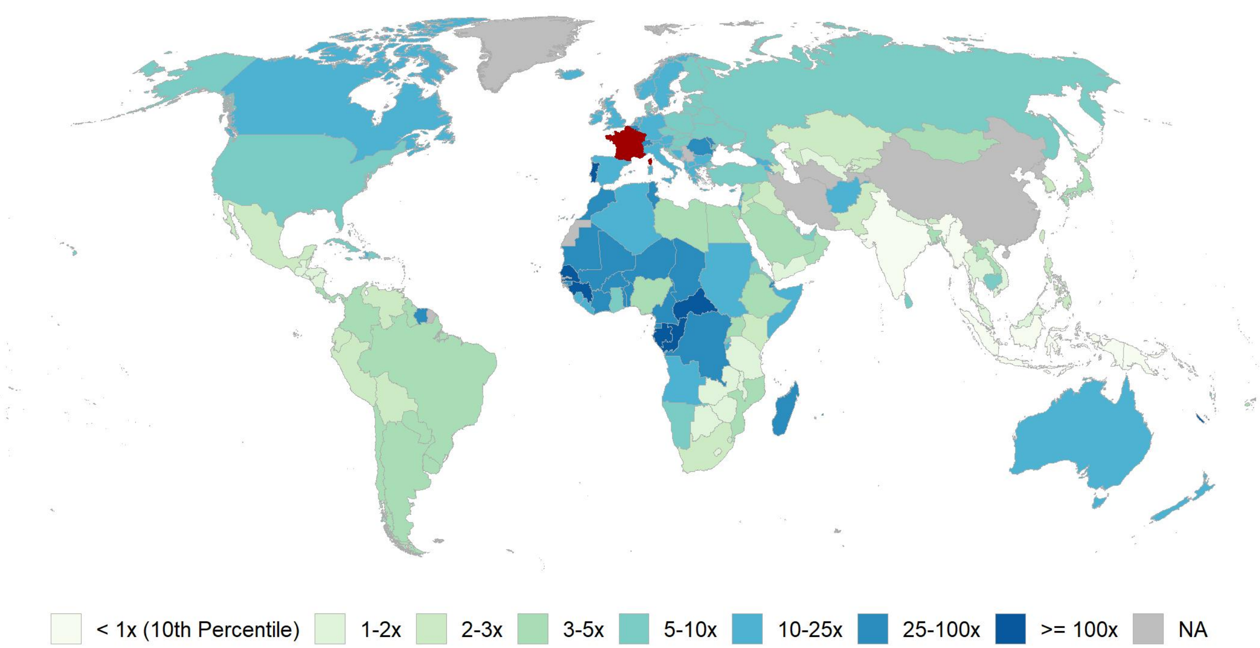
Portugal and France. Figure 1 shows the social connectedness of two Western European countries, Portugal (Panel A) and France (Panel B). Both countries have strong links to other countries in Western Europe and North America. For Portugal in particular, European links are much stronger than links to North African countries despite similar geographic distance. These patterns suggest a role for both cultural similarity and geographical distance in explaining social connectedness. Both countries' connections also highlight an important role played by colonial history and language in shaping present-day social connectedness. Portugal is strongly connected to its former (Portuguese-speaking) colonies Brazil, Angola, Guinea-Bissau, and Mozambique. Similarly, France is strongly connected to its former colonies in Africa (Morocco, Tunisia, Cameroon, Chad, Mali, and Niger), in the Indian Ocean (Madagascar), and in South America (French Guyana). Within Europe, Portugal is most strongly connected to

Figure 1: Social Connectedness of Portugal and France

(A) Social Connectedness to Portugal



(B) Social Connectedness to France



Note: Figure shows a heatmap of the social connectedness to Portugal (Panel A) and France (Panel B). For each country in our data, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Luxembourg. These connections, which are stronger than the connection to Portugal's next-door neighbor Spain, are likely related to the fact that 15% - 20% of Luxembourg's population is of Portuguese origin. Many of these Luxembourg residents are descendants of Portuguese guest workers who moved to Luxembourg as part of a special guest worker agreement in the 1960s. This finding suggests that past migration movements continue to influence social connections today.

India and Malaysia. Figure 2 shows the social connectedness of two Asian countries, India (Panel A) and Malaysia (Panel B). Both countries are strongly connected to the countries on the Arab peninsula, likely a result of migrant workers from India and Malaysia moving to work in these countries in recent years. Similarly, Malaysia is strongly connected to Nepal, likely due to a guest worker program allowing Nepalis to work in Malaysia. Social connections also appear to reflect earlier episodes of migrant and forced labor movements. For instance, India is strongly connected to Guyana in South America. In the 19th century, there was a lack of plantation workers following the abolition of slavery in this former European colony. Indians were selected to fill the gap as they were used to working under tropical condition and willing to accept cheap terms (Davis, 1951), and the resulting social connections to India appear to remain a century later. Finally, the Muslim-majority Malaysia is more strongly connected to the dominantly Muslim countries on the Indian subcontinent (Pakistan, Bangladesh, and the Maldives) than it is connected to India itself, suggesting a role of religion in shaping today's social connections.

2.2 Regression Analysis

These case studies suggest that several factors such as geographic distance, colonial history, and past migration shape today's social connections between countries. We estimate the following regression to analyze the contributions of these factors to social connectedness more systematically:

$$\log(SCI_{i,j}) = \beta + \gamma G_{i,j} + \delta_i + \delta_j + \epsilon_{i,j}. \tag{2}$$

The dependent variable is the logarithm of the *Social Connectedness Index*, $\log(SCI_{i,j})$, between country i and country j. The vector $G_{i,j}$ captures variables that might help us understand the determinants of international social connectedness. We also include fixed effects for each country, δ_i and δ_j, to absorb any country-specific factors that may affect our measure of a county's connections to others, such as patterns of Facebook usage or internet penetration.

2.2.1 Geography and Social Connectedness

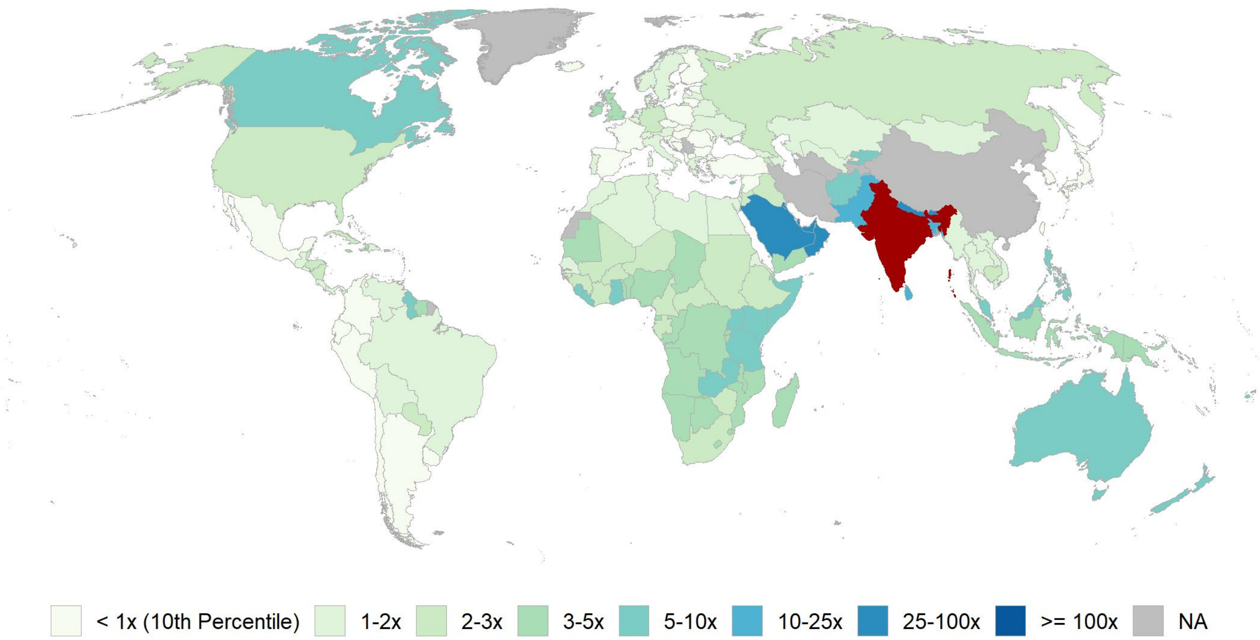
We first explore the role of geographic proximity in shaping international social connectedness. We measure the geographic distance between two countries as the population-weighted distance given by CEPII; summary statistics are presented in Table 1. The binscatter plots in Figure 3 show a broadly log-linear relationship between distance and social connectedness, both with and without country fixed effects. For large distances, the relationship becomes slightly convex, indicating that distance matters somewhat less once countries are already far apart.

Table 2 explores the role of distance using Equation 2. Column 1 only includes fixed effects for country i and country j. The results suggest that about 30% of the pairwise connectedness of countries is explained by the fact that countries differ in their *average* degree of global connectedness. In other

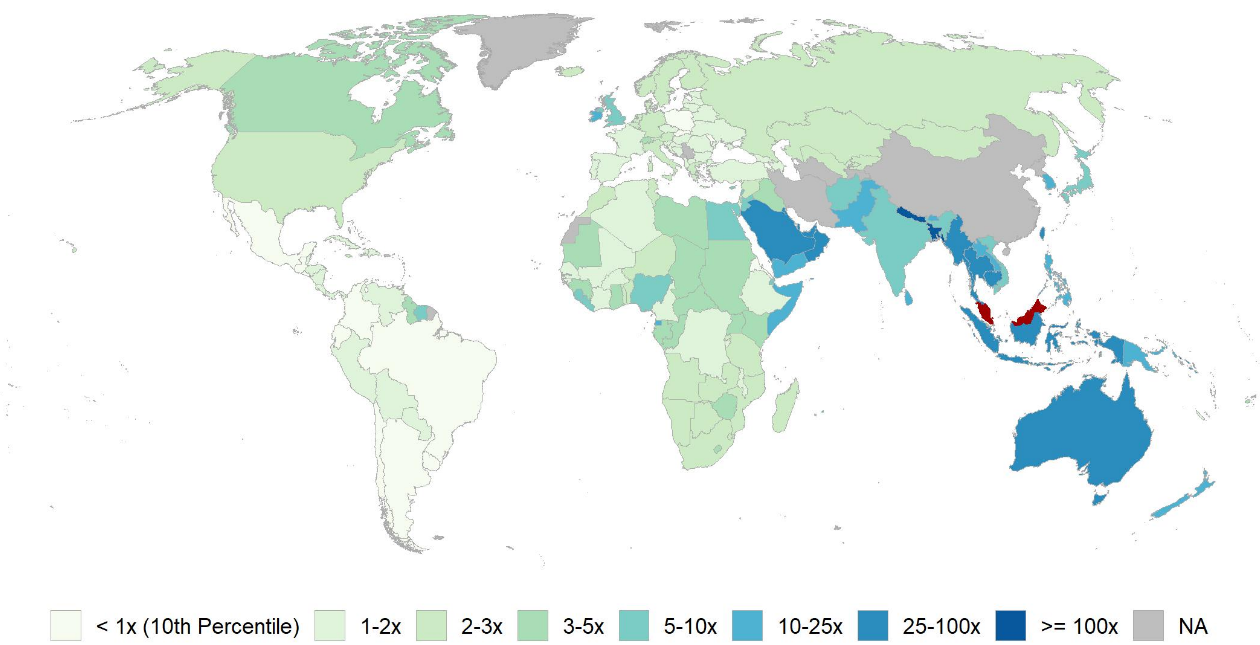
Figure 2: Case Studies of International Social Connectedness

(A) Social Connectedness to India

t



(B) Social Connectedness to Malaysia



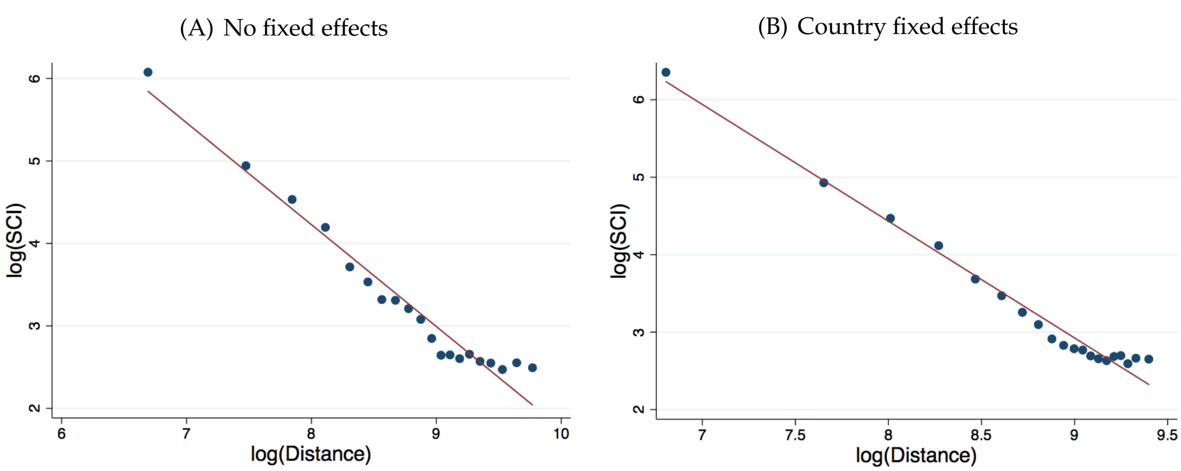
Note: Figure shows a heatmap of the social connectedness to India (Panel A) and Malaysia (Panel B). For each country in our data, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Table 1: Summary Statistics

	Mean	P10	P25	P50	P75	P90	N	Source
log(SCI)	3.30	1.61	2.08	2.94	4.16	5.54	32,220	Facebook
log(Distance)	8.75	7.68	8.39	8.92	9.30	9.58	32,220	CEPII
log(1+Migrant Population)	2.49	0.00	0.00	0.00	4.93	8.13	31,862	United Nations
Common Colonizer	0.11	0.00	0.00	0.00	0.00	1.00	32,220	CEPII
Colonial Relationship	0.01	0.00	0.00	0.00	0.00	0.00	32,220	CEPII
Genetic Distance	0.04	0.01	0.02	0.04	0.05	0.06	27,390	Spolaore et al. (2018)
Common Official Language	0.15	0.00	0.00	0.00	0.00	1.00	32,220	CEPII
Religious Distance	0.76	0.54	0.67	0.80	0.85	0.96	28,730	Spolaore et al. (2016)
Δ GDP per Capita (in '00,000$s)	0.18	0.01	0.03	0.10	0.28	0.49	32,220	CEPII
Common Border	0.02	0.00	0.00	0.00	0.00	0.00	32,220	CEPII
Same Continent	0.23	0.00	0.00	0.00	0.00	1.00	32,220	UNStats
Same Subcontinent	0.13	0.00	0.00	0.00	0.00	1.00	32,220	UNStats

Note: Table presents summary statistics of variables used in Section 2. Variables include the logarithm of SCI, the logarithm of distance, the logarithm of 1 plus the average migrant population, a dummy indicating whether the pair of countries had a common colonizer post 1945, a dummy indicating whether the pair of countries was in a colonial relationship post 1945, genetic distance following Spolaore and Wacziarg (2018), a common official language dummy, religious distance following Spolaore and Wacziarg (2016), differences in GDP per capita (in hundred thousands of dollars), a common border dummy, a same continent dummy, and a same subcontinent dummy.

Figure 3: Social Connectedness vs. Geographic Distance



Note: Figures show binscatter plots of social connectedness and geographic distance based on Equation 2. Panel A regresses log(SCI) on log(Distance) without any fixed effects, while Panel B controls for country fixed effects.

words, most of the cross-sectional variation in social connectedness across country pairs is the result of characteristics that vary at the country-pair level, and not characteristics that vary at the country level. Column 2 adds controls for $\log(Distance_{i,j})$ to the country fixed effects; this specification corresponds to the binscatter plots in Panel B of Figure 3. The estimates imply that a 10% increase in the geographic distance between countries is associated with a 15.1% decline in social connectedness. In terms of magnitude, this elasticity is similar to the elasticity of the connectedness between U.S. counties to geographic distance found by Bailey et al. (2018b). Moreover, geographic distance explains more than one half of the variation in social connectedness across countries after accounting for country fixed effects.

Column 3 explores the importance of any potential non-linearities in the relationship between geographic distance and social connectedness. Specifically, instead of controlling for $\log(Distance_{i,j})$ linearly, we include dummy variables for 500 quantiles of $\log(Distance_{i,j})$. The R^2 of the regression is only slightly higher — 67.2% instead of 65.2% — when we control for these quantiles. This finding confirms that the baseline log-linear specification is reasonable. Nevertheless, in columns 4 and 5 of Table 2, we split the sample into country pairs that are more or less than 6,000km apart. The estimated elasticitiy is -1.8 for countries less than 6,000km apart, and -0.8 for countries more than 6,000km apart. Strikingly, distance explains only 6% of the variation not explained by fixed effects in the sample of countries that are more than 6000km apart, but almost half of the respective variation for countries closer than 6000km.

Column 6 of Table 2 explores whether sharing a border increases social connectedness beyond geographic distance. For instance, a direct border may make it more likely that residents frequently spend time in the other country, either for work or leisure. It might also induce governments to cooperate and establish policies fostering cross-country interactions. Consistent with these hypotheses, the estimates indicate that citizens of two countries that share a border are nearly twice as likely to be friends with each other compared with citizens of two countries that are equally far apart but that do not share a border. However, the incremental R^2 of including this additional control over distance alone is relatively small at 0.6%, since the common border indicator does not allow us to understand the substantial variation in connectedness between the vast majority of pairs of non-neighboring countries.

We also examine how continental borders shape friendships between countries. Following UNStats, we group countries into five continents — Africa, the Americas, Asia, Europe, and Oceania — and into subcontinents such as Northern Africa and Sub-Saharan Africa. The regression results in column 7 of Table 2 show that countries on the same subcontinent are 2.2 times as connected as two countries that are equally far apart but on different continents. Being on the same continent (but not the same subcontinent) is associated with an 61% increase in social connectedness relative to two countries that are equally far apart but on different continents. Column 8 explores whether sharing a continent affects friendships differently on different continents. Countries in Oceania and Africa are substantially more likely to be socially connected to other countries on the same continent than predicted purely by geographic distance. On the other hand, countries in the Americas are not significantly more likely to be connected to each other than to countries on a different continent that are equally far apart.

2.2.2 Social Connectedness and Country Similarity

The previous results highlight that various measures of geographic proximity can explain a little over one half of the variation in social connectedness across country-pairs that is not explained by country

Table 2: The Geographic Determinants of Social Connectedness

	Dependent variable: log(SCI)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Distance)		-1.507***	500	-1.805***	-0.846***	-1.439***	-1.040***	-1.152***
		(0.035)	Quantiles	(0.055)	(0.070)	(0.034)	(0.048)	(0.047)
Common Border						0.925***	0.986***	0.960***
						(0.098)	(0.100)	(0.092)
Same Continent							0.619***	
							(0.074)	
Same Subcontinent							0.385***	
							(0.097)	
Both in Africa								1.327***
								(0.133)
Both in Americas								0.061
								(0.140)
Both in Asia								0.262*
								(0.091)
Both in Europe								0.698***
								(0.159)
Both in Oceania								2.571***
								(0.323)
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y	Y
Sample	All	All	All	Distance < 6000km	Distance > 6000km	All	All	All
R^2	0.291	0.652	0.672	0.647	0.551	0.657	0.678	0.690
Within R^2	0.000	0.510	0.537	0.495	0.060	0.516	0.546	0.563
N	32,220	32,220	32,220	12,478	19,742	32,220	32,220	32,220

Note: Table shows results of regression 2. The dependent variable is the logarithm of social connectedness for a country-pair. Explanatory variables include the logarithm of distance, a dummy indicating a common border, a same continent dummy, a same subcontinent dummy, and dummies indicating whether the pair of countries belongs to Africa, Americas, Asia, Europe or the Oceania region. In column 3, the logarithm of distance is replaced by indicators based on 500 quantiles of distance. All specifications include fixed effects for the origin and destination country. Standard errors are clustered by origin and destination country. We have data on social connectedness for 180 countries, which leads to 32,220 (= 180 x 179) observations. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

fixed effects. We next explore the role of other factors in explaining international friendship linkages. Specifically, we analyze the role of past migration, colonial history, genetic distance, common language, religion, and differences in GDP. Table 1 shows summary statistics on these variables, as well as the data sources; not all variables are available for all country pairs. Table 3 presents the cross-correlation of these gravity variables with the SCI and with each other. Naturally, many of the variables are somewhat correlated with one another: for example, countries that share a colonial history are more likely to have a common official language. Table 4 contains the results from Regression 2 when controlling for these variables. We first explore the conditional relationships between the gravity variables and social connectedness separately for each variable, before analyzing them jointly in a multivariate analysis.

Table 3: Correlation Table: Social Connectedness and Determinants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) log(SCI)	1.00											
(2) log(Distance)	-0.59	1.00										
(3) log(1+Migrant Population)	0.37	-0.41	1.00									
(4) Common Colonizer	0.26	-0.04	-0.07	1.00								
(5) Colonial Relationship	0.10	-0.02	0.17	-0.03	1.00							
(6) Genetic Distance	-0.40	0.56	-0.44	0.01	-0.01	1.00						
(7) Common Official Language	0.44	-0.12	0.12	0.40	0.13	-0.02	1.00					
(8) Religious Distance	-0.27	0.22	-0.18	0.03	0.01	0.11	-0.24	1.00				
(9) Δ GDP per Capita (in '00,000$s)	0.02	0.01	0.31	-0.10	0.05	-0.10	-0.02	-0.01	1.00			
(10) Common Border	0.28	-0.36	0.28	0.05	0.04	-0.17	0.12	-0.12	-0.07	1.00		
(11) Same Continent	0.54	-0.63	0.23	0.10	-0.03	-0.37	0.19	-0.18	-0.15	0.23	1.00	
(12) Same Subcontinent	0.53	-0.51	0.12	0.12	-0.02	-0.24	0.29	-0.20	-0.22	0.27	0.71	1.00

Note: Table presents correlations between variables used in Section 2. Variables include the logarithm of SCI, the logarithm of distance, the logarithm of 1 plus the average migrant population, a dummy indicating whether the two countries had a common colonizer post 1945, a dummy indicating whether the pair of countries was in a colonial relationship post 1945, genetic distance, a common official language dummy, religious distance, differences in GDP per capita (in hundred thousands of dollars), a common border dummy, a same continent dummy and a same subcontinent dummy.

Migration. It is likely that migrants retain many friends and family ties to their country of origin. As a result, we would expect that past migration patterns explain some of the observed cross-sectional variation in social connectedness. To measure migration between country pairs, we use bilateral migration data from the Population Division of the Department of Economic and Social Affairs of the United Nations.[7] The estimates in column 1 of Table 4 show a strong relationship between past migration and current social relationships. Doubling the average migrant population increases social connectedness between countries by over 20%. Including migration in addition to distance and country specific fixed effects increases the R^2 of the regression by 6.7 percentage points.

Common Colonial History. The case studies in Section 2.1 suggested that countries with a common colonial history are likely to maintain closer present-day social ties. To systematically explore this relationship, we use two measures of colonial history. Our first measures is an indicator for having a common colonizer after 1945. The second measure is an indicator for having been in a colonial relationship post 1945. The estimates in column 2 indicate that colonial history correlates strongly with

[7] Most of the data is based on populations censuses from the year 2015. Population registers and surveys are used to supplement the census data. Whenever the number of migrants from a country to another country is missing, we set the number to zero. For each country pair, we then compute the average of the number of migrants from country A living in country B and the number of migrants from country B living in country A. To deal with zero migration between two countries, we add one before taking the logarithm. Dividing this number by the sum of the populations in countries A and B leaves the coefficient almost unchanged, because of the log-log specification (the regression coefficient is not exactly the same because we add one to the number of migrants in order to deal with zeros values).

Table 4: The Determinants of Social Connectedness

	Dependent variable: log(SCI)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Distance)	-1.080***	-1.467***	-1.161***	-1.381***	-1.415***	-1.477***	-0.579***
	(0.037)	(0.032)	(0.033)	(0.038)	(0.038)	(0.034)	(0.044)
log(1+Migrant Population)	0.207***						0.166***
	(0.01)						(0.01)
Common Colonizer		0.900***					0.325***
		(0.078)					(0.055)
Colonial Relationship		1.911***					1.012***
		(0.196)					(0.115)
Genetic Distance			-30.82***				-20.83***
			(2.103)				(1.939)
Common Official Language				1.100***			0.544***
				(0.071)			(0.058)
Religious Distance					-1.522***		-0.577***
					(0.170)		(0.118)
Δ GDP per Capita (in '00,000$s)						-1.278***	-0.981***
						(0.269)	(0.202)
Common Border							0.193*
							(0.090)
Same Continent							0.502***
							(0.059)
Same Subcontinent							-0.100
							(0.072)
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y
Coeff. on distance w/o regressors	-1.506	-1.507	-1.496	-1.507	-1.494	-1.507	-1.490
R^2	0.714	0.677	0.699	0.691	0.660	0.656	0.793
Within R^2	0.602	0.549	0.578	0.569	0.518	0.520	0.709
Incremental R^2 of regressors	0.067	0.028	0.048	0.042	0.011	0.008	0.142
N	31,862	32,220	27,390	32,220	28,730	32,220	26,082

Note: Table shows results of Regression 2. The dependent variable is the social connectedness of a country-pair. Independent variables include the logarithm of distance, the logarithm of 1 plus the average migrant population, a dummy indicating whether the pair of countries had a common colonizer post 1945, a dummy indicating whether the pair of countries was in a colonial relationship post 1945, genetic distance, a common official language dummy, religious distance, differences in GDP per capita (in hundred thousands of dollars), a common border dummy, a same continent dummy and a same subcontinent dummy. All specifications include fixed effects for the origin and destination country. Standard errors are clustered by origin and destination country. We have data on social connectedness for 180 countries, which leads to 32,220 (= 180 x 179) observations. Not all variables are available for all country pairs; see Section A.1 for more details. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

social connectedness. Countries with a common colonizer have almost twice as many friendship links on Facebook as other countries that are similarly far apart. Having been in a colonial relationship increases the connectedness between countries by a factor close to three. Adding controls for colonial ties to distance and country fixed effects explains additional 2.8% of the cross-sectional variation in social connectedness.

Genetic Distance. Homophily suggests that there are likely to be more connections among countries whose populations are more similar to each other, including along genetic lines. We measure genetic distance following Spolaore and Wacziarg (2018).[8] Table 3 shows that genetic distance has a raw correlation of -40% with social connectedness; this is reflected in the strong negative relationship in the regression analysis. Going from the 10th percentile of genetic distance to the 90th percentile is associated with a decrease in social connectedness of 154%. Including genetic distance leads to an incremental increase in explanatory power, as measured by the R^2, of just under 5%.

Common Language. Language is a natural determinant of social relationships. After all, it is hard to form a personal relationship without speaking the same language. To formally explore the relationship between language and social connectedness, we use an indicator variable for whether two countries share a common official language. The common language indicator is strongly correlated with social connections as evidenced by a raw correlation of 44% presented in Table 3. Column 4 of Table 4 confirms the strong relationships. Having a common language more than doubles the social connectedness between two countries, and increases the R^2 by 4.2%.

Similar Religion. People of the same or similar religion may find it easier to connect to others who share their belief system. Similarly, people of the same religion may also be more likely to meet each other across countries, for instance when traveling for pilgrimage. To explore the role that religion has on social connectedness, we measure religious distance between countries following Spolaore and Wacziarg (2016).[9] The estimates in column 5 of Table 4 suggest that social connections decrease by 64% when moving from the 10th percentile to the 90th percentile of religious distance. The incremental R^2 of religion is about 1%, which is less than the variation explained by the other variables explored so far.

Similarity in GDP. Recent research by Bailey, Cao, Kuchler, Stroebel, and Wong (2018a) has documented that, at the individual level, people are more likely to be friends with others of similar incomes. We next explore whether this is also true for international social connectedness. For each country pair, we compute the absolute difference in GDP per capita. Column 6 of Table 4 shows that differences in GDP correlate with social connectedness. A ten thousand dollar higher absolute difference in GDP per capita

[8]Our measure of genetic distance is based on variation in the human DNA for ethnic groups from Pemberton, DeGiorgio, and Rosenberg (2013) and converted to a country-to-country measure using the shares of each ethnic group in each country's populations from Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003). It is therefore the expected genetic distance between two individuals randomly selected from the two countries.

[9]This measure of religious distance between any two countries is based on two components: the distance between different religions and the share of the population in a country that belongs to a religion. The proximity of two religions is based on how many nodes the two religions share in a tree describing the relationship between different religions. For instance, Roman Catholics are more closely related to Orthodox Christians than to Muslims, but the latter two are more closely related to each other than to religions originating in Asia such as Hinduism. The religious distance between two countries is then obtained by summing across the distances of all religions while weighting each religion by the fractions of people in the two countries that follow the religion.

across two countries corresponds to a 12.8% decline in social connectedness. However, differences in GDP only explain a small fraction of the cross-sectional variation in social connections, with an incremental R^2 of only 0.8%.

Multivariate Regression. The first six columns of Table 4 have shown the relationship between each regressor and social connectedness separately (in addition to distance and country fixed effects). However, many of these variables are correlated with each other, as shown in Table 3, and they often capture related aspects of across-country similarity. Column 7 explores how much of the variation in social connectedness these variables can explain jointly. As expected, the estimated coefficients for most variables decrease somewhat, though all variables retain their economic and statistical significance. The estimated effect of distance drops to about one third of the coefficient in the univariate regressions, suggesting that distance captures some aspects of the other cultural and social similarity in explaining social connectedness. Taken together, the incremental R^2 of all additional regressors (beyond distance and fixed effects) is 14%. Among these additional variables, the largest marginal contribution to explanatory power comes from variables that correspond to an aspect of social connectedness not already captured by any of the other variables. Excluding each variable one-by-one, we find the largest loss in R^2 from excluding controls for migrant population (decline of 3.7%), geographic distance (decline of 2.0%), and genetic distance (decline of 1.9%); see Appendix Table A.1 for these numbers.

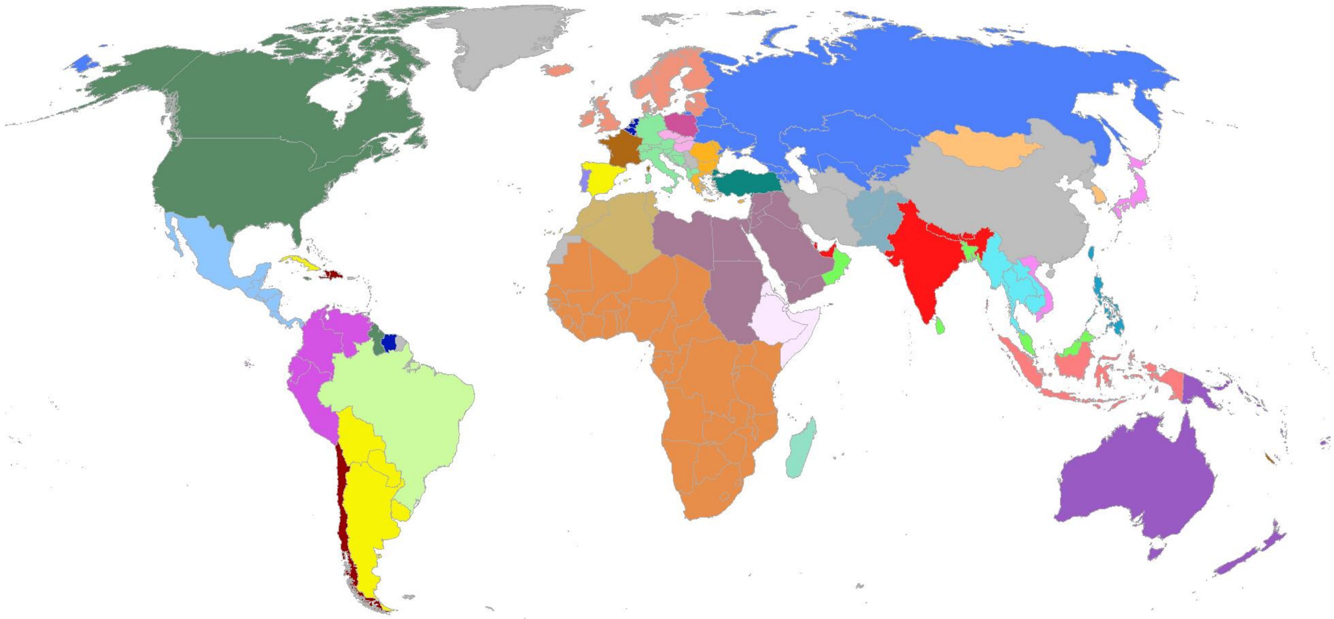
2.3 Groups of Socially Connected Countries

Our analysis above has shown that countries that share certain characteristics are more socially connected to each other. Shared characteristics often lead us to think about groups of countries, such as "Spanish-speaking South America" or "Arab North Africa", where countries in the group are all similar to one another on some salient dimension. In this section, we formalize the idea of groups of countries with strong connections to each other. To do this, we create clusters of countries that have a high within-cluster pairwise social connectedness. There are a number of possible algorithms to construct such clusters. Here, we use hierarchical agglomerative linkage clustering.[10] We use this algorithm to group countries into 30 clusters. Figure 4 shows the 30 different clusters and Table A.2 lists the countries in each cluster. The average number of countries per cluster is 6. There are three clusters that only contain a single country — Brazil, Turkey, and Poland. By far the largest cluster with 37 countries contains all of Southern and Western Africa. All other clusters have substantially fewer countries, with the second and third largest clusters each containing 12 countries.

Different characteristics appear to shape the different clusters. For instance, all countries in Central America and Mexico form one cluster, highlighting the importance of geography. Spain, however, forms a cluster with Argentina, Uruguay, Paraguay, and Bolivia, highlighting that not all clusters are geographically contiguous and pointing at shared language as another important factor. The non-contiguous clus-

[10]Conceptually, the agglomorative clustering algorithm starts by considering each of the N countries as a separate group of size one. In the first step, the two "closest" countries are merged into one larger group, producing N-1 total groups. In each subsequent step, the closest two groups are again merged. This process continues until all the countries are merged into a given number of clusters. We define the "distance" between two countries as the inverse of $SCI_{i,j}$: the lower the probability of a given Facebook user in country i knowing a given Facebook user in country j, the "farther apart" socially the two countries are. We calculate the closeness between clusters with more than one country as the average distance between the countries in the cluster.

Figure 4: Groups of Socially Connected Countries: 30 Clusters



Note: Figure shows countries grouped together when we use hierarchical agglomerative linkage clustering to create 30 distinct groups of connected countries.

ter between Chile, Haiti, and the Dominican Republic is likely explained by recent migrant movements to Chile from Haiti and the Dominican Republic. The cluster comprised of Mongolia and South Korea might reflect both a long common history and common ethnic ancestry, as well as recent migration from Mongolians to South Korea.

3 Trade and Social Connectedness

We begin by studying the relationship between social connectedness and aggregate bilateral trade. We use bilateral goods trade data from CEPII (Gaulier and Zignago, 2010). We analyze data from 2017 to align the trade data most closely with the time at which we measure social connectedness, though the documented patterns are robust to using trade data from other years. The trade data is disaggregated at the 6-digit HS96 code level, and contains information on 4,914 product categories. For this section, we aggregate the product-level trade data into bilateral trade flows between country pairs and then replace missing trade values with zeros.[11] In Section 3.1, we exploit the disaggregated nature of the trade data to understand the mechanism through which social connectedness influences trade patterns. When merged with social connectedness and gravity data, our panel has 30,102 country-pairs.[12] To understand the relationship between social connectedness and trade flows, we follow the literature to estimate the following gravity regression:

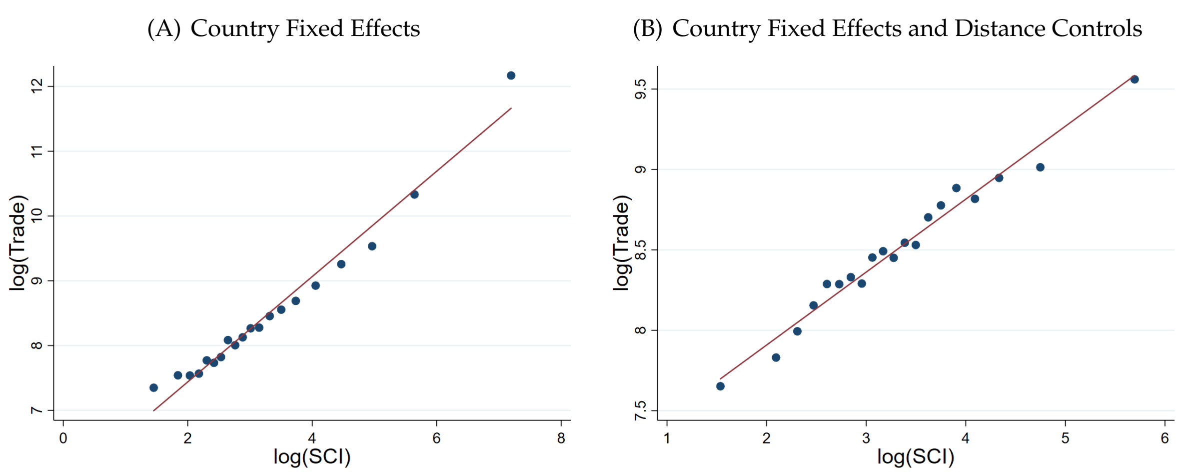
$$X_{i,j} = \exp\left[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(Distance_{i,j}) + \beta_3 G_{i,j} + \delta_i + \delta_j\right] \cdot \epsilon_{i,j} \tag{3}$$

[11] We therefore make the assumption that there was no trade whenever a trade entry is not reported/missing.

[12] This corresponds to pairwise trade data from 174 countries. Relative to the analysis in Section 2, we lose Botswana, Lesotho, Luxembourg, Namibia, Sudan, and Swaziland, for which we have data on social connectedness, but no data on trade.

where $X_{i,j}$ denotes the total value of exports from country i to country j, $G_{i,j}$ represents country-pair characteristics, δ_i and δ_j are exporter and importer fixed effects, and $\epsilon_{i,j}$ is an error term. We include all bilateral trade flows, including zeros, and estimate the regression using Poisson Pseudo Maximum Likelihood (PPML) to account for zero bilateral trade between some countries (see the discussion in Santos Silva and Tenreyro, 2006).[13] We follow a large literature and control for geographic distance between countries by including the log of distance in the PPML specification. The main variable of interest, $SCI_{i,j}$, is also included in logs. This choice of functional form is based on the evidence in Figure 5, which plots the log of social connectedness against the log of trade flows. Panel A controls for importer and exporter fixed effects, while Panel B additionally controls for geographic distance. In both cases, the relationship between aggregate trade and social connectedness is approximately linear in logs. The importer and exporter fixed effects control for country-level characteristics such as population and GDP, which affect the level of trade. These fixed effects also control for average trade barriers at the country-level, as well as country-specific differences in the use of Facebook that might affect our measure of social connectedness.

Figure 5: Trade vs. Social Connectedness



Note: Figures show binscatter plots of aggregate bilateral trade and social connectedness. Panel A regresses log(Trade) on log(SCI) without controlling for distance, while Panel B includes log(Distance) as a control. Both panels control for exporter and importer fixed effects. Here, we focus on the intensive margin of trade, which reduces our sample to 20,054 observations.

The results from estimating Regression 3 are presented in Table 5. Column 1 presents a specification with only importer and exporter fixed effects and shows that 83.3% of the variation in bilateral trade flows is explained by these fixed effects alone. This is perhaps unsurprising, since larger and richer countries will trade substantially more on average. Column 2 introduces controls for the social connectedness between each country pair. The elasticity of trade with respect to social connectedness is an economically significant 0.68, suggesting that a 1% increase in the social connectedness is associated with a 0.68% increase in bilateral trade. Variation in social connectedness accounts for a substantial share

[13]Our estimation uses the algorithm in Correia et al. (2019a) and Correia et al. (2019b). In the Appendix, we present estimates of regression 3 in logs while dropping observations with zero trade flows (i.e., we focus on exploring the effect of social connectedness on the intensive margin of trade). All findings are robust to this deviation from the PPML estimation approach.

of the cross-sectional variation in trade flows: over half of the variation in bilateral trade flows that is not explained by the country fixed effects is explained by social connectedness.

Table 5: Gravity Regressions - Aggregate Trade

	Dependent variable: Aggregate Bilateral Trade							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(SCI)		0.683***			0.325***		0.325***	500
		(0.040)			(0.029)		(0.030)	Quantiles
log(Distance)			-0.996***		-0.660***	-0.863***	-0.563***	-0.626***
			(0.060)		(0.071)	(0.054)	(0.066)	(0.063)
Common border				1.763***		0.439***	0.413***	0.404***
				(0.220)		(0.124)	(0.105)	(0.094)
Common official language				0.169		0.064	-0.131	-0.163**
				(0.145)		(0.102)	(0.086)	(0.072)
Common colonizer				0.958***		0.233	0.040	0.201*
				(0.153)		(0.142)	(0.137)	(0.108)
Colonial relationship				0.445**		0.019	-0.065	0.240
				(0.219)		(0.377)	(0.276)	(0.162)
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.833	0.919	0.929	0.891	0.937	0.932	0.939	0.949
Within R^2		0.086	0.096	0.058	0.104	0.099	0.106	0.116
N	30,102	30,102	30,102	30,102	30,102	30,102	30,102	30,102

Note: Table shows results from Regression 3. The dependent variable is total exports from country i to country j. Controls include the logarithm of SCI, the logarithm of distance, a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, a dummy indicating whether the pair of countries was in a colonial relationship post 1945 and 500 quantiles of SCI. All specifications include fixed effects for the importer and exporter country. Standard errors are clustered by exporter and importer country. The data include 174 countries and 30,102 (= 174 x 173) observations. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

In column 3, we remove the control for social connectedness, and include controls for the geographic distance between countries. The elasticity of trade with respect to distance is -0.996, which is consistent with estimates in prior work (see Head and Mayer, 2014). The increase in the R^2 relative to the specification in column 1 is similar in magnitude to the increase from including social connectedness. Column 4 adds to the specification from column 1 a number of other gravity variables that the literature has focused on (see Anderson and Van Wincoop, 2003; Head and Mayer, 2014). Consistent with prior work, we find that sharing a border, a colonizer, a language, and a colonial relationship all increase bilateral trade. However, these gravity variables jointly explain a smaller share of the cross-sectional variation in trade flows than social connectedness does by itself.

In column 5, we control for both social connectedness and geographic distance. Due to the correlation between these two variables documented in Section 2, the elasticity of trade to social connectedness drops from 0.683 to 0.325, and the distance elasticity drops from -0.996 to -0.660. This additional control for social connectedness increases the R^2 by 0.8% over distance alone. This is a larger magnitude than what is obtained by adding the other gravity variables to geographic distance (see column 6). The decline in distance elasticity from controlling for social connectedness in column 5 is substantial. This finding relates to an important literature that has argued that the estimated effect of distance on trade is too large and time-invariant to primarily capture trading costs (see Disdier and Head, 2008; Head and Mayer, 2014). This literature has proposed that geographic distance might instead be proxying for some other friction such as information asymmetries (e.g., Rauch, 2001; Rauch and Trindade, 2002). Since social connectedness can help overcome many of these frictions, our evidence here and in the rest of this paper is highly consistent with this interpretation of the magnitude of the distance elasticity.

In column 7, we jointly control for social connectedness, geographic distance, and other gravity variables. The results highlight that social connectedness explains variation in bilateral trade flows beyond those explained by distance and other standard gravity variables. Interestingly, the estimated elasticity of trade with respect to social connectedness remains unchanged relative to the estimates from column 5, even though the previous section highlighted that these gravity variables are highly correlated with social connectedness. Quantitatively we find that, even after controlling for a host of control variables that potentially proxy for various aspects of social connectedness, a doubling of social connectedness between two countries is associated with a 33% increase in trade flows.

In the rest of the paper, we explore potential channels through which social connectedness can increase trade. In the next section, we present evidence that social connectedness increases trade through two channels discussed by Chaney (2016): reducing informational frictions and improving contract enforceability. In Sections 3.2 and 4, we discuss and rule out that our findings are the result of reverse causality whereby more trade leads to higher social connections. We also show that our results are unlikely to be driven by omitted variables such as common preferences across trading partners, which could be correlated with both social connectedness and trade.

3.1 Mechanisms Through Which Social Connectedness Affects Trade Patterns

One important channel through which social connectedness might increase trade is by reducing information asymmetries. For example, by allowing importers and exporters to share information about prices and products, social connectedness can be a key channel through which search costs are mitigated and trade is facilitated. In an influential paper, Rauch (1999) argues that these search costs are likely to be lower for goods that are traded on organized exchanges, since those goods are effectively homogeneous and their prices are transparent. In contrast, goods that are not traded on organized exchanges are heterogeneous and prices are more difficult to determine, and therefore more subject to information frictions. We would thus expect a larger role for social connections in reducing information frictions for goods that are not traded on exchanges. To test this hypothesis, this section explores how the elasticity of trade with respect to social connectedness varies with whether goods are traded on exchanges.

A second source of trade barriers that can be mitigated through social connections are contract enforcement frictions. Anderson and Marcouiller (2002) show that weak institutions in the importing country substantially decreases trade (see also Berkowitz et al., 2006; Levchenko, 2007). In the absence of strong institutional enforcement of contracts, Greif (1989, 1993), Rauch and Trindade (2002), and others have shown that ethnic networks can provide reputation-based punishment for contract violation and thereby facilitate trade. Building on these findings, we conjecture that social connections beyond ethnic networks may also help with contract enforcement by deterring contract violations when individuals are trading with personal connections. Since such an enforcement mechanism would be a substitute to contract enforcement through formal institutions, we would expect the elasticity of trade to social connectedness to be larger in places with a weaker rule of law. We also test this hypothesis below.

Since different countries trade different products, we jointly study heterogeneity across product types and measures of the rule of law across countries.[14] To do so, we use disaggregated data, where the unit of observation is an export of product k from country i to country j. A product represents a 2-digit HS category (there are 96 unique categories). We estimate regressions with the following specification:

$$\begin{aligned} X_{i,j,k} = \exp[&\beta_1 \log(SCI_{i,j}) + \beta_2 \log(SCI_{i,j}) \cdot ET_k + \\ &\beta_3 \log(SCI_{i,j}) \cdot RL_i + \beta_4 \log(SCI_{i,j}) \cdot RL_j + \beta_5 G_{i,j,k}] \cdot \epsilon_{i,j,k} \end{aligned} \tag{4}$$

In this specification, ET_k is the fraction of exchange traded goods for each 2-digit HS product category k.[15] RL_i and RL_j are continuous measures of the rule of law in the exporting and importing countries, as measured by the World Governance Indicators (Kaufmann et al., 2011).[16] As before, $G_{i,j,k}$ are a set of gravity variables and fixed effects. All specifications include origin country $\times$ product and destination country $\times$ product fixed effects, in part to control for differences in country-specific factor endowments. Additionally, product-specific distance controls account for the fact that different products have different shipping costs per unit of distance (see the discussion in Rauch, 1999).

We present the results from estimating Regression 4 in Table 6. Column 1 shows our baseline specification from column 7 of Table 5 for this product-level trade data. The estimated elasticity of trade with respect to social connectedness (as well as the unreported coefficients on the other gravity variables) are quite similar to our baseline specification presented in Table 5. In column 2, we interact $\log(SCI_{i,j})$ with the fraction of exchange traded products in each HS 2-digit category. The coefficient on this interaction is -0.188, which shows that social connectedness matters substantially less for product categories that have more exchange traded goods. Quantitatively, the elasticity is almost 2.2 times as large in a category with no exchange traded goods than it is in a category with primarily exchange traded goods. This provides evidence that one of the channels through which social connectedness increases trade is by decreasing information asymmetries, which are smaller for exchange traded goods. In column 3, rather

[14]For example, if countries with a weak rule of law primarily traded exchange traded products, then studying the heterogeneity in aggregate trade flows by rule of law without distinguishing across product types might incorrectly suggest that social connectedness matters *less* for low-rule-of-law countries.

[15]To construct this measure, we start from trade data at the 6-digit HS level, and use the "conservative" classification scheme by Rauch (1999) to classify goods into "exchange traded" and "not exchange traded"; the results are near-identical using the "liberal" classification scheme in Rauch (1999). We then calculate the fraction of exchange traded goods at the 6-digit HS level using the total global share of trade in those goods within each 2-digit HS category. Across products, ET_k has a range from 0 to 0.91, a mean value of 0.12, and a standard deviation of 0.25. To provide a sense of the variation, within the category HS-19 (preparations of cereals, flour, starch or milk such as pastry products) 0% of goods are exchange traded, within the category HS-27 (mineral fuels, oils, and products of their distillation) 44% are exchange traded, and within category HS-80 (tin and articles thereof) 90% are exchange traded.

[16]This measure captures "perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence." In our analysis, we use values as of 2017. The measure ranges between -2.5 and 2.5. Across countries, it has a mean value of -0.06, and a standard deviation of 0.99. For example, Venezuela has a score of -2.3, Mexico has a score of -0.57, the United States has a score of 1.64, and Finland has a score of 2.0.

Table 6: Gravity Regressions - Heterogeneity

	Dependent variable: Product-Specific Bilateral Trade						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(SCI)	0.319***	0.345***	0.349***	0.339***	0.346***	0.371***	0.375***
	(0.016)	(0.016)	(0.013)	(0.022)	(0.016)	(0.020)	(0.016)
log(SCI) × Share Exchange Traded		-0.188**	-0.193***			-0.206***	-0.209***
		(0.075)	(0.063)			(0.073)	(0.063)
log(SCI) × Rule of Law Destination				-0.032**	-0.035***	-0.033***	-0.035***
				(0.013)	(0.008)	(0.013)	(0.008)
log(SCI) × Rule of Law Origin				-0.013	-0.015	-0.018	-0.019
				(0.017)	(0.014)	(0.016)	(0.013)
Origin Country × Product FE	Y	Y	Y	Y	Y	Y	Y
Destination Country × Product FE	Y	Y	Y	Y	Y	Y	Y
Other Gravity Controls	Y	Y	Y	Y	Y	Y	Y
log(Distance) × Product FE	Y	Y		Y		Y	
Distance Group × Product FE			Y		Y		Y
R^2	0.929	0.929	0.958	0.928	0.957	0.929	0.957
N	2,889,792	2,889,792	2,889,792	2,758,080	2,758,080	2,758,080	2,758,080
N - Explained by FE	381,798	381,798	409,125	345,078	371,692	345,078	371,692

Note: Table shows results from Regression 4. The dependent variable is exports of product category k from country i to country j. Product-level trade data is aggregated up to the first 2 digits of the HS1996 product classification. Other gravity controls include a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, and a dummy indicating whether the pair of countries was in a colonial relationship post 1945. We also separately control for the logarithm of distance interacted with product categories in columns 1, 2, 4 and 6 and for distance groups (dividing distance into 500 quantiles) interacted with product categories in columns 3, 5 and 7. Share Exchange Traded refers to the proportion of exchange traded products (based on the Rauch classification) within a product category. Rule of law is obtained from the World Governance Indicators published by World Bank. All specifications include fixed effects for the importer and exporter country interacted with product categories. Standard errors are clustered by exporter and importer country interacted with product categories. The data includes 174 countries and 96 product categories, which amounts to 2,889,792 observations. In columns 4 to 7, the number of observations reduces to 2,758,080 observations, because we lack the rule of law measure for 4 countries. Observations that are fully explained by fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

than using product specific interactions with log distance, we interact 500 dummies for quantiles of distance with product dummies to allow for a separate non-linear effect of distance on each product's trade. In this specification the coefficient on the interaction of social connectedness with the exchange-traded fraction remains very similar.

Columns 4 and 5 interact our measures of the rule of law in the destination and origin countries with the social connectedness across each pair of countries. Column 4 controls for product-specific effects of $log(Distance_{i,j})$, and column 5 includes non-linear product-specific distance controls. The elasticity of exports to social connectedness is smaller across country pairs in which the destination country has a higher rule of law. The economic magnitude of the estimated effect is sizable: a one-standard-deviation increase in the rule of law in the destination country reduces the elasticity of trade to social connectedness by about 10% of its baseline effect in column 1: all else equal, the elasticity of exporting to Mexico with respect to social connectedness is about 22% larger than the corresponding elasticity of exporting to the United States.[17] The fact that social connectedness is associated with the largest increase of exports to countries with a weaker rule of law suggests that social connections mitigate some of the contract enforcement frictions that have been shown to reduce trade in countries with weak institutions (see Anderson and Marcouiller, 2002). Stronger rule of law in the origin country also reduces the elasticity of trade to social connectedness, though the estimates are not statistically significant. The relatively stronger response of the elasticity to the rule of law in the destination country (compared to the origin country) could be due to social connectedness mitigating importers incentive to delay or default on payments, which would generally be enforced by institutions in the importer's country.

In columns 6 and 7, we jointly explore the heterogeneity of the elasticity of trade with respect to social connectedness across the rule-of-law of the trade partners and the types of products. The coefficients are very similar to the prior specifications. Overall, this evidence suggests that social connectedness increases global trade through two channels. First, by helping to alleviate information asymmetries, especially for non-exchange trade products where asymmetries are likely the highest. Second, by improving contract enforcement when the quality of institutions in importing countries is low.

Groups of Socially Connected Countries. We next explore the effect on trade of two countries being connected to other similar countries over and above the effect of two countries being directly connected with each other. There are several mechanisms through which being in the same social cluster could increase trade. First, being in the same social cluster as other countries may reduce contracting frictions due to community sanctions or other similar mechanisms (see the discussions in Greif, 1989, 1993; Jackson et al., 2017). A second mechanism through which being in the same social cluster can increase trade is by alleviating search frictions. Being in the same social cluster would allow traders to contact direct social connections who can then put them into contact with their social connections. This channel is closely related to the evidence in existing literature that shows that firms make exporting decisions in a sequential manner by searching for new export destinations that are close to their current ones (see Albornoz et al., 2012; Chaney, 2014; Morales et al., 2015). We use the hierarchical clustering introduced in Section 2.3 to build two measures of whether countries are in similar social clusters. Our first measure is simply a dummy variable for whether countries are in the same cluster. For this dummy variable, we use the same 30 clusters from Section 2.3, though our results are robust to variation in the cluster size. Our second measure captures whether countries are placed into the same cluster at an early, middle, or late stage of the agglomerative clustering process.[18] The advantage of this measure is that it does not require us to take a stand on the number of clusters we put countries into.

We present the results in Table 7. For comparison, in column 1 we present our baseline specification

[17]Mexico has a RL coefficient of -0.57, the U.S. has a RL coefficient of 1.64, so the RL difference between the two countries is 1.64 + 0.57 = 2.21. This difference corresponds to a $(0.032 \times 2.21)/0.319 = 22\%$ difference in the elasticity.

[18]Agglomerative clustering is a sequential procedure where at each stage countries are merged into clusters where the merging tends to occur earlier in the process if countries are closer socially. Due to the nature of the hierarchical clustering algorithm, most countries do not join the same cluster until the last stage. Because we want to capture whether countries are in a similar social cluster, we use cutoffs of 95% and 90% to put country-pairs into 3 groups for when they join the same cluster. These cutoffs correspond to the 18th and 42nd clustering step.

for aggregate trade, corresponding to column 7 of Table 5. In column 2 we control for whether countries are in the same cluster when splitting the world into 30 socially connected country clusters. Conditional on the bilateral social connectedness and other gravity variables, being in the same cluster increases trade substantially. As an example, Spain is approximately equally well-connected to the United Kingdom and to Argentina, but it only shares a cluster with Argentina (other countries in that cluster are Paraguay, Cuba, and Bolivia). This suggests that Spain and Argentina share more connected countries compared to Spain and the United Kingdom. Our estimates suggest that, all else equal, Spain would thus trade 19% more with Argentina than with the United Kingdom.

In column 3 we interact the number of countries (in logs) in the cluster with the same cluster dummy variable. We see that the effect of being in the same cluster on trade increases when the cluster includes a larger number of other countries. This is consistent with the interpretation that being socially connected to the same community of countries helps alleviate search costs, with this effect being larger for bigger clusters. In column 4, we add the dummy variables for when countries join the same cluster. We find that, controlling for the direct effect of bilateral social connectedness, as well as the other gravity variables, countries which join the same cluster early rather than late have 24% higher bilateral trade. To rule out that these effects capture non-linear relationship of bilateral social connectedness and distance on trade, columns 5-7 introduce controls for 500 quantiles of social connectedness and geographic distance. Our results remain qualitatively unchanged compared to the estimates in columns 2-4. Overall, we conclude that beyond the effect of direct social connections, being in the same social cluster substantially increases bilateral trade between countries.

3.2 Alternative Interpretations

In the previous sections, we presented evidence consistent with a causal link between social connectedness and trade flows, whereby higher social connectedness facilitates trade through reducing information asymmetries and helping with the enforcement of contracts. There are two potential alternative interpretations for our findings. The first alternative interpretation is reverse causality, whereby trade may cause social connections rather than social connections causing trade. A second alternative interpretation of the aggregate trade patterns is the existence of an omitted variable that correlates with both trade and social connectedness. For example, countries with high social connectedness may share similar tastes. If sharing similar tastes leads countries to trade more, as argued by Linder (1961), this could also explain the observed aggregate trade patterns (even if it might struggle to explain some of the heterogeneities across countries and products that we established above).

Unfortunately, the standard econometric tools to establish causal relationships cannot be credibly applied in this setting. For example, we believe that none of the determinants of social connectedness that we explored in Section 2.2.2 could serve as an instrument that only effects trade flows through its effect on social connectedness. As a result, in the next section, we turn to more disaggregated data on both social connections and trade that allow us to provide evidence against the quantiative importance of both reverse causality and omitted variables. Before doing so, however, it is worth thinking through the quantitative plausibility of the reverse causality story.

Indeed, while we believe that trade likely causes some business relations to form, and a subset of these may lead to Facebook friendship links, such a mechanism is unlikely to be a quantitatively

Table 7: Gravity Regressions - Clustering

	Dependent variable: Aggregate Bilateral Trade						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(SCI)	0.325***	0.313***	0.321***	0.310***			
	(0.030)	(0.030)	(0.029)	(0.031)			
Same Cluster (30 Gr.)		0.188**	-0.472**		0.241***	-0.189	
		(0.084)	(0.193)		(0.082)	(0.235)	
Join Same Cluster - Middle				0.100			0.180***
				(0.082)			(0.069)
Join Same Cluster - Early				0.241**			0.246***
				(0.119)			(0.095)
Same Cluster (30 Gr.) $\times$ log(Cluster Size)			0.326***			0.203*	
			(0.109)			(0.107)	
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y
Other Gravity Controls	Y	Y	Y	Y	Y	Y	Y
SCI Group FE					Y	Y	Y
Distance Group FE					Y	Y	Y
R^2	0.939	0.939	0.940	0.939	0.959	0.959	0.959
N	30,102	30,102	30,102	30,102	30,102	30,102	30,102

Note: Table shows results from Regression 3. The dependent variable is total exports from country i to country j. "Same Cluster (30 Gr.)" is a dummy variable indicating that the 2 countries are in the same cluster when we create 30 SCI-based clusters. "Join Same Cluster" indicates whether countries are placed into the same cluster at an early, middle, or late stage of the agglomerative process. Other gravity controls include the logarithm of distance, a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, and a dummy indicating whether the pair of countries was in a colonial relationship post 1945. We also separately control for distance and SCI groups (dividing distance and SCI into 500 quantiles) in columns 5 to 7. All specifications include fixed effects for the importer and exporter country. Standard errors are clustered by exporter and importer country. We have data on trade and social connectedness for 174 countries and 30,102 (=174 x 173) observations. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

important driver of our findings. In particular, to the extent that reverse causality explains the observed correlation, this effect would have to come through connections formed by individuals that have the opportunity to directly interact with the trading partners. Such individuals are only a small fraction of all Facebook users. As a result, to explain a substantial increase in all Facebook users' connectedness through reverse causality, trade has to explain an implausibly large increase in the Facebook links of the subset of the population that is engaged in trading.[19] In addition, only a subset of connections of social

[19] As a back-of-the-envelope calculation, we start from the observation that U.S. employment by firms engaged in importing and exporting is roughly 40% of total private sector employment (Bernard et al., 2005), or approximately 50 million individuals in the United States. This provides a strong upper bound on the number of people whose international connectedness could possibly be affected by trade: for example, we would not expect that General Motors' exports to Japan increase the international social connections of its factory workers, even if it might affect these links among its sales force. Again, making strong assumptions, let us assume that 20% of the U.S.-based workforce at these large exporting firms are directly involved in the trading of goods such that they would actually meet (and potentially become Facebook friends with) trading partners in the destination countries. This leaves about 10 million U.S.-based individuals whose social network might be causally affected by trade, corresponding to about 5% of U.S.-based Facebook users. It is possible that these users contribute a larger

links formed via trading will actually result in a Facebook friendship link, since Facebook is primarily a platform for personal networks rather than professional networks. In other words, for the observed elasticity to be the result of reverse causality, changes in trade would have to correspond to a very large increase in the Facebook friends among the small subset of Facebook users who are actively engaged in trade, and, if those Facebook links represent only a subset of all social connections formed, an even bigger (and we believe implausibly large) increase in the total number of connections formed.

4 Trade and Sub-National Social Connectedness in Europe

In the previous section, we explored the relationship between social connectedness and trade across countries. Our preferred interpretation of that evidence is that social connectedness can facilitate trade both by reducing information frictions and by helping with the enforcement of contracts. In order to rule out a number of potential alternative interpretations, we next study the relationship between sub-national social connectedness and trade patterns. We focus our analysis on Europe, where we have both measures of social connectedness and some information on trade flows at the regional level. Our data on social connectedness between European NUTS2 regions was introduced by Bailey et al. (2019c). These regions include between 800,000 and 3,000,000 individuals, and are defined for European Union members, European Union candidates, and European Free Trade Association members.

We conduct two exercises. In Section 4.1, we construct product-specific measures of across-country social connectedess. These measures weight the connectedness of sub-national region pairs by the importance that these regions should have for predicting exports of each product. These weights are based on where the product is produced in the exporting country and where it is used as an intermediary input in the importing country. We show that exports for each product vary primarily with these product-specific input-output weighted measures of social connectedness between countries. Our findings allow us to rule out that the observed correlation between social connectedness and trade flows at the country level is driven by either reverse causality or by more similar preferences between individuals in more connected countries. In Section 4.2, we link NUTS2-level social connectedness to rail freight volumes between those regions as a proxy for sub-national trade flows. We find that social connectedness between regions matters for the trade between those regions, even after controlling for country-pair fixed effects. This analysis allows us to control for many potential variables that might have been omitted from the aggregate country-level trade regressions in the previous section. We also find evidence that social connectedness and trust act as substitutes in their effect on trade.

4.1 Input-Output Weighted vs. Population-Weighted Social Connectedness

In Section 3, we related the volume of exports from country i to country j to the probability that a representative Facebook user in country i was friends with a representative Facebook user in country j,

share of international links of U.S.-based Facebook users. Again, making strong assumptions, let us assume that those Facebook users involved in trading hold twice as many foreign links as the average Facebook user. This suggests that the links of people holding 10% of total links could be affected by trading. When switching $\log(Exports_{i,j})$ and $\log(SCI_{i,j})$ in a specification corresponding to column 7 of Table 5, we estimate a coefficient on $\log(Exports_{i,j})$ of about 0.15: doubling exports is associated with a 15% increase in aggregate social connectedness. However, a 15% increase in the aggregate social connectedness corresponds to an implausibly large 150% increase in the links of those people whose links could in principle be affected.

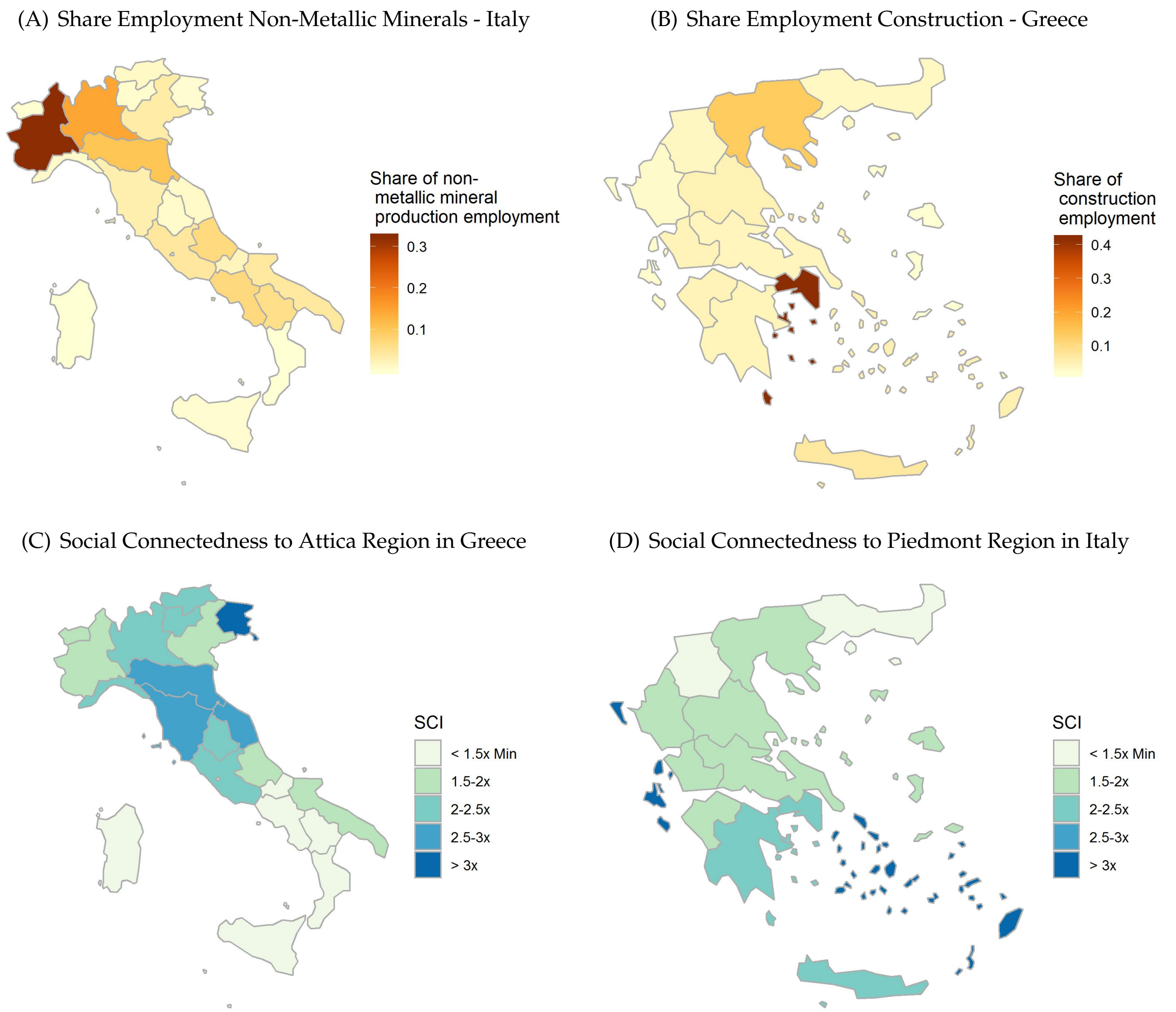
given by $SCI_{i,j}$. This measure of social connectedness is identical to a population-weighted average of social connectedness across the regions in countries i and j. Formally, let us index the regions in each country i by $r_i \in R(i)$, let $Friendships_{r_i,r_j}$ count the number of friendship links between individuals in regions r_i and r_j, let Pop_{r_i} denote the total population in region r_i, and let $PopShare_{r_i}$ denote the share of population in region r_i in country i: $\sum_{r_i \in R(i)} PopShare_{r_i} = 1$. Then:

$$
\begin{aligned}
SCI_{i,j} &= \frac{Friendships_{i,j}}{Pop_i \times Pop_j} = \frac{\sum\limits_{r_i \in R(i)} \sum\limits_{r_j \in R(j)} Friendships_{r_i,r_j}}{\left(\sum\limits_{r_i \in R(i)} Pop_{r_i} \right) \times \left(\sum\limits_{r_j \in R(i)} Pop_{r_j} \right)} && (5) \\
&= \sum_{r_i \in R(i)} \sum_{r_j \in R(j)} \frac{Pop_{r_i}}{\sum\limits_{r_i \in R(i)} Pop_{r_i}} \frac{Pop_{r_j}}{\sum\limits_{r_j \in R(j)} Pop_{r_j}} \frac{Friendships_{r_i,r_j}}{Pop_{r_i} \times Pop_{r_j}} && (6) \\
&= \sum_{r_i \in R(i)} \sum_{r_j \in R(j)} PopShare_{r_i} \times PopShare_{r_j} \times SCI_{r_i,r_j} && (7)
\end{aligned}
$$

In other words, in exploring the role of $SCI_{i,j}$ as a determinant of trade between two countries, we implicitly assume that the relative importance of the connectedness between different regions in explaining country-level trade increased with the population shares of those regions. In this section, we argue that, for each good, the connectedness between the regions in country i where the good is produced and the regions in country j where the good is used should be particularly important for explaining exports of that good from country i to country j. Our prediction that the social connections of individuals at the location of firms should matter disproportionately for predicting trade flows builds on the insights from a large literature that has documented that the vast majority of international trade is being conducted by a small set of firms (see Bernard et al., 2012, for a survey of this literature).

We find that the weights of regions in the production of goods often deviates substantially from their population weights, in particular for goods that are used as intermediate inputs in geographically clustered industries. Let us give a concrete example. More than 80% of Italian exports of non-metallic mineral products (e.g., cement) to Greece are used as inputs in the Greek construction sector. Panel A of Figure 6 shows the share of Italian employment in the sector that manufactures non-metallic mineral products in each of the country's NUTS2 regions. The largest shares are in the northwestern Piedmont region, which includes the city of Torino and a number of major industrial sites (note that these employment shares do not line up closely with population shares). Similarly, Panel B of Figure 6 shows the share of Greek construction employment that is in each of the country's NUTS2 regions. The largest employment shares are in the Attica region covering metropolitan Athens. Based on this information, we propose that for exporting non-metallic mineral products from Italy to Greece, connectedness between the Piedmont region and the Attica region should be particularly important, since firms located in those regions are most likely to be involved in any trade in this product. The bottom row of Figure 6 shows that there is substantial variation in which regions in Italy are connected to the Attica region in Greece (Panel C), and which region in Greece are connected to the Piedmont region in Italy (Panel D). These figures highlight that the strongest connections are not necessarily between the regions with firms that should matter most for the trading of non-metallic mineral products. We next test whether it is indeed the connections of those regions with firms most likely involved in trading a particular product that matter the most for explaining country-level trade of that product.

Figure 6: Regional Employment Shares And Social Connectedness



Note: Panel A shows the regional distribution of employment in the non-metallic minerals industry across NUTS2 regions in Italy. Panel B plots the regional shares of employment in the construction sector across NUTS2 regions in Greece. Panels C and D, respectively, show heatmaps of social connectedness from the Attica Region in Greece to Italian NUTS2 regions, and from the Piedmont Region in Italy to Greek NUTS2 regions.

To conduct this exercise, we construct, for each country $i \times$ country $j \times$ product p triplet, the input-output weighted social connectedness of regions in countries i and j that should be most important for predicting trade of product p. This construction involves a number of important steps. First, since the trade data is at the product level, while the employment and input-output data is at the industry level, we match products in the trade data to industries (see Appendix C.1 for details). Accordingly, for the description of the methodology that follows, we will interchangeably refer to p as representing a product or an industry. For each product p produced in country i, we then use the World Input-Output Tables (WIOT) described in Timmer et al. (2015) to measure the share of that product that is used as an intermediate input in each industry p' in country j, $IO_{i,j}^{p,p'}$. We focus on uses of products as intermediate inputs, such that $\sum_{p'} IO_{i,j}^{p,p'} = 1$, and only consider products where at least 50% of the exports across countries in our sample are used as intermediate inputs (rather than in final consumption).[20] This leaves us with a set P that includes 20 products, which we list in Appendix C.1. For each product $p \in P$, we then construct a measure of the social connectedness between countries i and j, $SCI_{i,j}^{p}$, that corresponds to the weighted average of the social connectedness between the NUTS2 regions in these countries that are most relevant for exporting product p from i to j:

$$SCI_{i,j}^{p} = \sum_{r_i \in R(i)} EmpShare_{p,r_i} \times \left[\sum_{p' \in P} IO_{i,j}^{p,p'} \times \left(\sum_{r_j \in R(j)} EmpShare_{p',r_j} \times SCI_{r_i,r_j} \right) \right]. \tag{8}$$

The variable $EmpShare_{p,r_i}$ represents the share of employment in industry p in country i that is in region r_i: $\sum_{r_i \in R(i)} EmpShare_{p,r_i} = 1$. These regional employment shares are constructed using data from Eurostat. We limit to 28 countries for which we have trade data, WIOT data, and regional employment data.[21] Similarly, we construct a product-specific measure of the input-output-weighed geographic distance between each country – again, under the maintained hypothesis that the geographic distance that should matter the most for exports in each country-pair-product is the distance between those regions where the product would be produced and used:

$$Distance_{i,j}^{p} = \sum_{r_i \in R(i)} EmpShare_{p,r_i} \times \left[\sum_{p' \in P} IO_{i,j}^{p,p'} \times \left(\sum_{r_j \in R(j)} EmpShare_{p',r_j} \times Distance_{r_i,r_j} \right) \right]. \tag{9}$$

Quantitatively, most of the cross-sectional variation in $SCI_{i,j}^{p}$ and $Distance_{i,j}^{p}$ comes from a common component that drives the social connectedness and geographic distance between all regions in a given country-pair. For example, all regions of Germany are more connected to regions in Austria than they are to regions in Finland. Indeed, regressions of $SCI_{i,j}^{p}$ and $Distance_{i,j}^{p}$ on country i × country j fixed effects have R^2s above 90%. The remaining variation comes from the fact that, for some products, the producing or using industries are geographically concentrated in regions that might be differentially connected (or differentially distant) than the average region in a country pair. For instance, in the example given above, the input-output-weighted social connectedness for non-metallic mineral products between Greece and Italy is about 7% higher than the population-weighted social connectedness be-

[20]One possible concern with this construction is that the actual measure of $IO_{i,j}^{p,p'}$ is based on observed trade flows, which are the eventual object of interest. Here it is important to note that $IO_{i,j}^{p,p'}$ does not depend on the overall level of exports of good p from country i to country j, but only on the relative shares of exports of product p from country i to country j that are used in each industry p' in country j. Nevertheless, we have also constructed a version of $SCI_{i,j}^{p}$ that uses a predicted value of $IO_{i,j}^{p,p'}$, based on the share of product p that is used in each industry p' when p is traded between all countries other than i and j. The results are unchanged using this procedure.

[21]These countries are Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Slovenia, Slovakia and the United Kingdom.

tween these countries. These differences provide the identifying variation in the following regressions.

Next, we explore how trade of different products correlates with the product-specific measures of social connectedness, $SCI^p_{i,j}$, and geographic distance, $Distance^p_{i,j}$, by estimating the following regression:

$$X_{i,j,p} = \exp[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(Distance_{i,j}) + \beta_3 \log(SCI^p_{i,j}) + \beta_4 \log(Distance^p_{i,j}) + \delta_{i,j,p}] \cdot \epsilon_{i,j,p} \quad (10)$$

Here, $X_{i,j,p}$ denotes the total value of exports from country i to country j of product p. We also include the logarithm of population-weighted measures of social connectedness, $\log(SCI_{i,j})$, and distance, $\log(Distance_{i,j})$, as controls; these are the same covariates that we used in Section 3. The vector $\delta_{i,j,p}$ represents various fixed effects. In all specifications we add country $i \times$ product p fixed effects as well as country $j \times$ product p fixed effects, which controls for the average propensity of each country to export and import each good.

Table 8 shows results from Regression 11. In column 1, we control only for the population-weighted social connectedness and distance. The estimated elasticities of trade to social connectedness is similar to that estimated in Section 3. This suggests that the set of countries and products for which we can construct input-output-weighted social connectedness has similar trade elasticities to the full sample of countries. In column 2, we instead control for the product-specific input-output-weighted social connectedness between countries i and j. The magnitudes of the elasticities of trade to social connectedness and geographic distance are similar to those estimated in column 1. As discussed above, this is consistent with the fact that much of the cross-sectional variation is captured by the component of social connectedness between i and j that is common across countries.

In column 3, we control for both the population-weighted and input-output weighted measures of social connectedness. While these two objects have a correlation of 95%, the regression loads strongly on the input-output weighted measure of social connectedness – once this is controlled for, the population-weighted social connectedness has no additional predictive power. In column 4, we include fixed effects for each country pair. These fully absorb the population-weighted social connectedness and geographic distance between country pairs. Importantly, the inclusion of country-pair fixed effects also controls for any observable or unobservable differences between countries that might have been correlated with both social connectedness and trade flows at the country level, leading to an omitted variables bias in the previous regressions. For example, including country-pair fixed effects controls for whether country pairs share a common language, a common religion, or a common historical origin, all of which might be correlated both with trade flows and with social connectedness. The estimated elasticity of trade flows to the product-specific social connectedness even increases somewhat in this specification, though the fact that we are now identifying our effect from less than 10% of the variation in $SCI^p_{i,j}$ and $Distance^p_{i,j}$ has also increased standard errors.

One specific concern alleviated by the specifications in columns 3 and 4 of Table 8 is that the correlation between country-level social connectedness and trade documented in Section 3 might pick up an effect of unobserved common preferences in consumption. Under this alternative theory, higher social connectedness between the populations of two countries coincides with more similar consumption preferences of the populations, for example because social connectedness is partially driven by migration, and migrants have similar preferences to individuals in their countries of origin. This similiarity of

Table 8: Input-Output Weighted Social Connectedness and Trade

	Dependent variable: Product-Specific Bilateral Trade					
	(1)	(2)	(3)	(4)	(5)	(6)
log(SCI)	0.272***		-0.071			
	(0.023)		(0.131)			
log(Distance)	-1.002***		-0.372*			
	(0.053)		(0.207)			
log(SCI^p)		0.248***	0.321**	0.382**	0.297**	0.227*
		(0.022)	(0.132)	(0.152)	(0.123)	(0.132)
log($Distance^p$)		-0.973***	-0.610***	-0.985***		
		(0.052)	(0.200)	(0.195)		
Origin Country × Product FE	Y	Y	Y	Y	Y	Y
Destination Country × Product FE	Y	Y	Y	Y	Y	Y
Undir. Country-Pair FE				Y	Y	
log($Distance^p$) Group FE					Y	Y
Undir. Country-Pair × Product FE						Y
R^2	0.953	0.955	0.955	0.970	0.976	0.993
N	15,120	15,120	15,120	15,120	15,120	15,120
N - Explained by FE	262	591	591	591	591	2,250

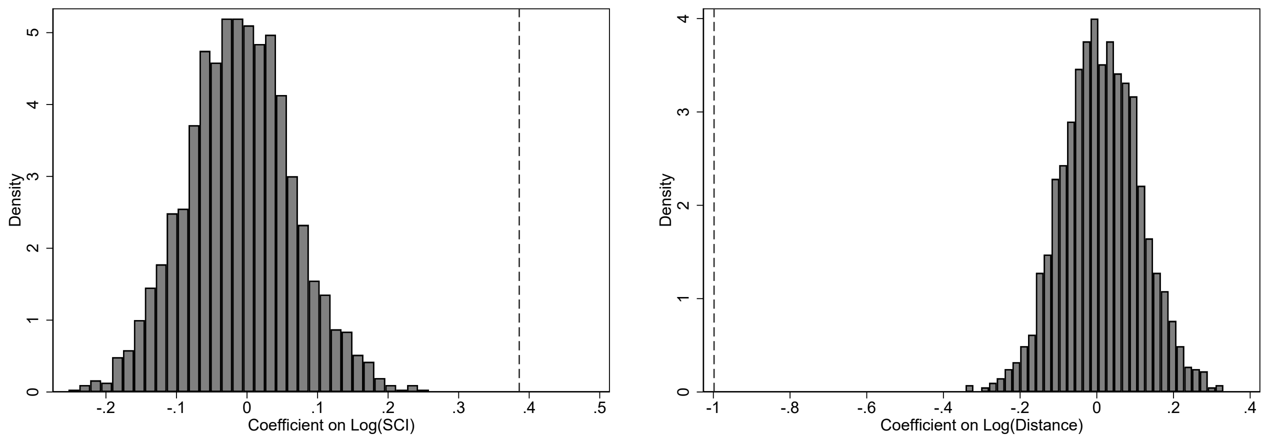
Note: Table shows the results from regression 11. The dependent variable is exports of product category k from country i to country j. The variable $SCI_{i,j}$ is the population-weighted average of NUTS2 region-level social connectedness. The variable $SCI_{i,j}^p$ is an employment share weighted measure that uses information on input-output trade, as defined in equation 9. The measures $Distance_{i,j}$ and $Distance_{i,j}^p$ are constructed in the same way as the corresponding social connectedness measures. All specifications include exporter-product and importer-product fixed effects. Column 4 and 5 add country-pair fixed effects that do not distinguish the direction of trade (undirected). In column 5, we replace the control for log($Distance_{i,j}^p$) with 500 dummy variables representing distance quantiles. Column 6 includes fixed effects that interact each industry with the undirected country-pair fixed effects. Standard errors are clustered by country pairs. The data includes 28 countries and 20 products leading to 15,120 (= 28 × 27 × 20) observations. Observations that are fully explained by fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

preferences might then be the source of trade in both final consumption goods and intermediate goods used in the production of the final consumption goods (see Linder, 1961). For intuition, we provide a concrete example: there are lots of Italian migrants in Germany, which increases the social connectedness between Italy and Germany. Due to Italians' desire for high-quality pizza, they demand substantial imports of high-quality pizza ovens from Italy, thereby increasing the trade between Italy and Germany, independent of any direct effect of social connectedness. In this story, the social connectedness between countries would primarily be a proxy for the similarity in preferences. However, if such an omitted variable explained the patterns in Section 3, the population-weighted social connectedness across regions would be the most appropriate measure of the similarity of preferences between the populations of two

Figure 7: Randomization Inference

(A) Coefficient on $\log(SCI^p_{i,j})$

(B) Coefficient on $\log(Distance^p_{i,j})$



Note: Figures show the distribution of regression coefficients for randomly selected values of social connectedness and distance; Panel A shows the coefficients for social connectedness, Panel B the coefficients for distance. The regression specification is equal to column 4 in Table 8; namely it is a regression of industry-level trade between countries on industry-specific measures of social connectedness and distance. The coefficients obtained in the original regression are shown as the dashed lines. We contrast the actual estimates with regression coefficients that are obtained when choosing "random" values for social connectedness and distance. To be more precise, for each country $i \times$ country $j \times$ product p triplet, we assign a value of $\log(SCI^p_{i,j})$ and $\log(Distance^p_{i,j})$ from a randomly chosen product in the same country pair. We then estimate a regression based on these "random" values and repeat this exercise 2,000 times. The distribution of estimated coefficients is then plotted in a histogram.

countries. The fact that it is, instead, the social connectedness between the locations of output and input industries for each product that determines the amount of trade suggests that similarities in preference between more connected countries does not constitute a quantitatively important determinant of trade.

Another way of exploring the statistical significance of the estimates in column 4 of Table 8 is the following. For each country $i \times$ country $j \times$ product p triplet, we assign a value of $\log(SCI^p_{i,j})$ and $\log(Distance^p_{i,j})$ from a randomly chosen product in the same country pair, and then re-run the regression. We repeat this exercise 2,000 times. The histograms in Figure 7 show the distribution of the coefficients on the re-shuffled values of input-output weighted social connectedness and distance; the dashed lines shows the estimated effect corresponding to column 4 in Table 8. The randomized coefficients are centered around zero: conditional on country-pair fixed effects, there is no additional explanatory power for trade in a given product coming from variation in input-output-weighted social connectedness of a random product. Said differently, what matters is not the social connectedness of regions involved in trade generally; instead, what matters for the trade in a specific product is the social connectedness across regions that produce and use that specific good.

In column 5, we replace the control for $\log(Distance^p_{i,j})$ with 500 dummy variables representing percentiles of the distribution of that variable. The fact that the coefficient on $\log(SCI^p_{i,j})$ remains unaffected rules out concerns that the loading on social connectedness may be picking up non-linearities in the relationship between the product-specific geographic distance and trade flows. Finally, in column 6 we include fixed effects that interact each product type with undirected country $i \times$ country j pair fixed effects: in other words, we are comparing exports of a specific good from country i to country j to the exports of the same good from country j to country i. The remaining variation in $SCI^p_{i,j}$ comes from the fact that the industries that produce the product in each country are not located in the same regions as the industries that use these products as an input. The inclusion of these fixed effects decreases the coefficient on $SCI^p_{i,j}$ only slightly, providing further evidence that common preferences across countries (which should affect the trade of a given good in both directions) are not a quantitatively large driver of the findings in Section 3.

Reverse Causality. Another benefit of exploring input-output-weighted social connectedness is that it allows us to further address a concern outlined above regarding reverse causality as an explanation for the observed relationship between trade and social connectedness. While we have argued above that such reverse causality cannot explain the observed correlation from a quantitative perspective, we next provide additional evidence against reverse causality as the mechanism behind our findings.

Our approach starts from the observation that, under the reverse causality story, the social connectedness between input-output-weighted regions should be systematically larger in magnitude than the social connectedness between population-weighted regions, since reverse causality would increase the connectedness between those regions actually engaged in trade relative to the connectedness of other regions not engaged in trade. To test whether this is indeed the case, we construct for each country i $\times$ country j $\times$ product p triplet the difference between the input-output-weighted social connectedness and the population-weighted social connectedness across countries i and j. To interpret the magnitude of the differences, we express them as as a fraction of the cross-sectional standard deviation of $SCI^p_{i,j}$:

$$SCI\_Divergence^p_{i,j} = \frac{SCI^p_{i,j} - SCI_{i,j}}{SD(SCI^p_{i,j})}. \tag{11}$$

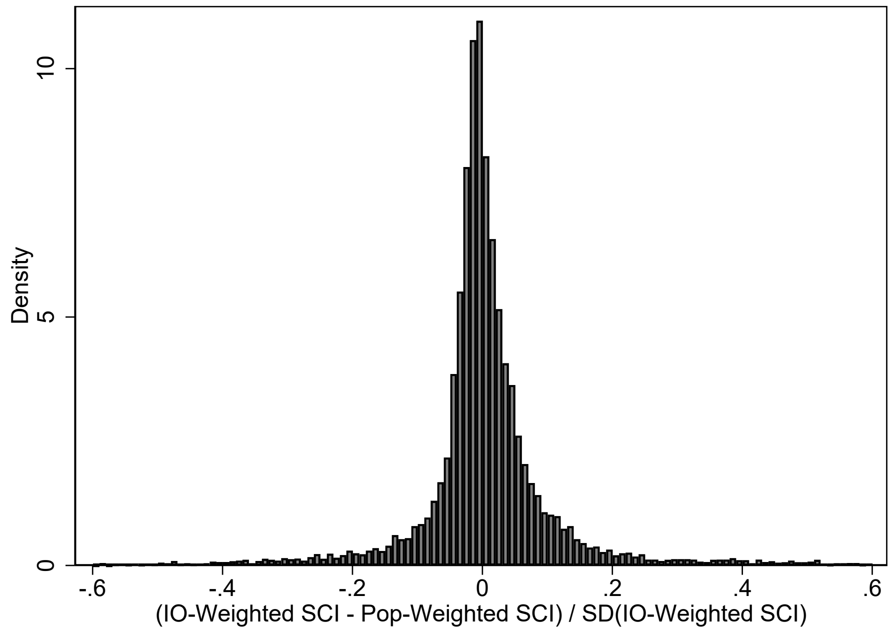
Figure 8 shows a histogram of $SCI\_Divergence^p_{i,j}$ across all country i $\times$ country j $\times$ product p triplets. The distribution has a mean of 0.008, and a median of -0.003. In other words, the regions that were shown to be most important for the trade in a given product are equally likely to be more connected and less connected than the population-weighted average of regions across a country-pair. This provides strong evidence against a quantitatively large reverse causality story in which the fact that two regions trade more with each other causes them to be more connected.[22]

4.2 Subregional Social Connectedness and Rail Freight Flows

The challenge for exploring sub-national trade patterns is the absence of trade data at sub-national level. However, within Europe, we observe data on rail freight tonnage shipped between pairs of NUTS2

[22]It is extremely unlikely that the absence of an average difference between input-output-weighted social connectedness and population-weighted social connectedness is driven by two offsetting forces, whereby reverse causality would push trading regions to be more connected, which is offset by a second force that would cause them to be less connected. Indeed, most plausible additional mechanisms would also lead regions with industries that would trade with each other to be more connected. For example, an endogenous location of industries in a given exporting country into regions that are more connected to regions in the importer country that use the products as an intermediary input would bias $SCI\_Divergence^p_{i,j}$ towards being larger than zero.

Figure 8: Ruling Out Reverse Causality



Note: Figure shows the distribution of the difference between the input-output-weighted social connectedness and the population-weighted social connectedness. More specifically, for each country i × country j × product p triplet, we construct the input-output weighted social connectedness $SCI^p_{i,j}$, then subtract the population-weighted social connectedness $SCI^p_{i,j}$ and divide by the cross-sectional standard deviation of $SCI^p_{i,j}$. The exact formula is given in equation 12.

regions for a number of countries.[23] While this will not capture all trade flows between these regions, rail freight transport accounted for 12.2% of all intra-EU freight transport in 2015 (Eurostat, 2015).

We next explore the relationship between rail freight and social connectedness and geographic distance across European NUTS2 regions. We include the trade for regions within and across countries:

$$RailFreight_{r_i,r_j} = \exp[\beta_1 \log(SCI_{r_i,r_j}) + \beta_2 \log(Distance_{r_i,r_j}) + \delta_{r_i,r_j}] \cdot \epsilon_{r_i,r_j} \tag{12}$$

The dependent variable, $RailFreight_{r_i,r_j}$, is the amount of goods (in tons) shipped by rail from region i to region j. The variables $\log(SCI_{r_i,r_j})$ and $\log(Distance_{r_i,r_j})$ are the logarithm of the social connectedness and distance between NUTS2 regions, and δ_{r_i,r_j} represents various fixed effects.

Table 9 shows the results from Regression 13. Column 1 shows that the elasticity of rail freight to social connectedness is larger than the elasticity of all trade to social connectedness estimated in previous sections. This higher elasticity is not a feature of the set of countries included in our analysis, since running Regression 3 only on countries included in the rail freight data yields elasticities similar to those in the baseline regression in Table 5. As a result, the higher elasticity observed here could be the result of social connectedness being more important for the type of products shipped by rail, or it could

[23] We use data on region-to-region rail goods transport made available by Eurostat in the series *tran_rt_rago*. The data are built from individual country reports to the European Union on national and international rail transport in 2015. For each pair of NUTS2 regions r_i and r_j, the data include the tons of goods that were loaded on a railway vehicle in region r_i and unloaded in region r_j. We take a number of steps to standardize and clean the data, as described in Appendix C.2.

be that the importance of social connectedness varies with the means of transportation.

Table 9: Sub-national Social Connectedness and Rail Goods Transport

	Dependent variable: Regional Bilateral Rail Freight								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(SCI)	0.630***	0.622***		0.548***	0.419***	0.359**	0.504**		
	(0.055)	(0.040)		(0.054)	(0.096)	(0.162)	(0.236)		
log(Distance)	-0.835***								
	(0.130)								
Same Country			1.941***	0.319					
			(0.196)	(0.261)					
log(SCI) $\times$ Low Trust								1.064**	1.298**
								(0.440)	(0.508)
log(SCI) $\times$ High Trust								0.486**	0.563**
								(0.239)	(0.241)
Orig. and Dest. Region FE	Y	Y	Y	Y	Y	Y	Y	Y	Y $\times$ Trust
Distance Group FE		Y	Y	Y	Y	Y	Y	Y	Y $\times$ Trust
Undir. Country-Pair FE					Y				
Orig. Reg. x Dest. Ctry FE						Y	Y	Y	Y
Dest. Reg. x Orig. Ctry FE						Y	Y	Y	Y
Sample: Has Trust Data							Y	Y	Y
R^2	0.761	0.805	0.794	0.805	0.824	0.859	0.839	0.840	0.861
N	74,862	74,862	74,862	74,862	74,862	74,862	34,572	34,572	34,572
N - Explained by FE	27,442	37,233	37,233	37,233	48,495	59,465	27,200	27,200	27,479

Note: Table shows the results from Regression 13. The dependent variable is the rail freight shipped from a NUTS2 region to another NUTS2 region. "Same Country" is a dummy variable indicating whether the rail shipment is between two NUTS2 regions within the same country (domestic shipment). In columns 8 and 9, we divide the sample into Low Trust and High Trust observations based on the country-level trust measure of Guiso et al. (2009). In columns 1 to 6, we use the full panel with 332 NUTS2 regions and 74,862 non-missing trade observations. In columns 7 to 9 the sample is reduced to 34,572 observations, because we have trust data on only 15 countries (which contain 225 NUTS2 regions). All specifications include origin region and destination region fixed effects. Columns 2 to 9 introduce 500 quantiles of distance as a control. Column 5 adds country-pair fixed effects that do no distinguish the direction of trade (undirected). Columns 6 to 9 add origin region x destination country and destination region x origin country fixed effects. The standard errors are clustered by NUTS2 origin region and NUTS2 destination region. Observations that are fully explained by fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

In column 2 and all following columns of Table 9, we replace our controls for geographic distance with dummy variables for each percentile of the distance distribution. This ensures that the estimated elasticity between social connectedness and trade flows is not in part determined by any non-linearities in the relationship between $\log(Distance_{r_i,r_j})$ and trade.

In column 3 of Table 9 we remove our controls for the measure of social connectedness between regions, and instead include a "same country" dummy variable. Conditional on the geographic distance between two regions, trade is about seven times as large between regions in the same country than between regions in different countries. This estimate is similar in magnitude to that of Chen (2004), who finds that EU countries trade about 6 times more with themselves than with other countries. Our estimates are also similar those in Tan (2016), who looks at truck freight shipments in Europe and finds that trade flows are 5.75 times higher for shipments within the same country (see McCallum, 1995; Anderson and Van Wincoop, 2003, for other contributions to the literature estimating border effects in trade). These estimated border effects are large in light of the fact that most of the countries in the sample are part of the European Common Market, and therefore face no formal barriers to trade such as tariffs; indeed, all results in this section are the same when we restrict our sample to exclusively focus on NUTS2 regions from countries within the single market. In column 4, we bring back the control for social connectedness. The estimate of trade declines at the border drop dramatically, from a border effect of 597% to a statistically insignificant border effect of about 38%. This finding suggests that much of the reason we see border effects is the fact that social connectedness is much stronger across regions within countries than it is across equidistant region pairs in different countries.

In column 5, we include country-pair fixed effects. As before, this controls for any differences across country-pairs that might affect trade between regions of these countries, and that might be correlated with social connectedness (e.g., common language, common history, or common tastes). The estimated elasticity of trade flows to social connectedness barely changes, suggesting that country-pair-level omitted variables that might correlate with social connectedness are not a key driver of our results. However, within Europe, some of these omitted variables do not just vary at the country-pair level, but can also vary at the region-country level. For example, the Alsace region in France has common historical heritage with regions in Germany (for example, during the Franco-Prussian war, France ceded Alsace to the German Empire, while the Treaty of Versailles ceded it back to France). Similarly, the Zentralschweiz region of Switzerland has a common language with Germany, while the Lake Geneva region shares a common language with France. To control for such determinants of trade at the region-country level, column 6 includes origin region $\times$ destination country and destination region $\times$ origin country fixed effects. The estimated elasticity between social connectedness and trade is unaffected, though standard errors increase as our fixed effects remove more and more of the cross-sectional variation in SCI_{r_i,r_j}. Again, these estimates suggest that our central findings are not confounded by omitted variables bias.

Trust and Social Connectedness. As discussed above, a central mechanism through which we believe that social connections affect the level of trade flows is through helping with the enforcement of contracts. Guiso et al. (2009) have shown that, within Europe, trade increases in the amount of trust within countries, in part because trust can also help alleviate concerns about adherence to contracts. This suggests that trust and social connectedness might act as substitutes in fostering trade. We next test that prediction in the data, using the same country-pair-specific trust data studied in Guiso et al. (2009)[24]

[24] Guiso et al. (2009) construct data on trust between countries by using a set of surveys that was conducted by Eurobarometer in 1995. Eurobarometer conducted the survey in 15 European countries and surveyed around 1,000 individuals per country. The survey participants were asked how much they trust their fellow citizens and how much they trust citizens from other countries. The exact question was *"I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust or no trust at all"*. The answers are then coded into numerical scores that are increasing in the amount of trust; the scoring scheme was 1 (*"no trust at all"*), 2 (*"not very much trust"*), 3 (*"some trust"*), and 4 (*"a lot of trust"*). The trust of a country to another country is then obtained by computing the average score across all participants surveyed in a given country. The countries that were surveyed are Austria, Belgium, Britain, Denmark, Finland, France, Germany, Greece, Italy, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden, and

The correlation between trust and social connectedness across country pairs is 0.35.

In column 7 of Table 9, we replicate the estimates from column 6, now only for country-pairs for which we observe trust. The estimates are of broadly similar magnitude as those in the full sample. We then split country-pairs into those with above-median and below-median trust, and interact $\log(SCI)$ between regions with this dummy. This specification allows us to obtain separate estimates of the elasticity of trade with respect to social connectedness across regions in country pairs with high and low levels of trust. Column 8 mirrors the specification in column 7, but with seperate SCI coefficients for region-pairs in low and high trust country-pairs. We can observe that trade varies substantially more with social connectedness across regions in country pairs where trust is low than it does across regions in country pairs where trust is high. Column 9 highlights that this result survives interacting all control variables and fixed effects with the high-trust dummies, allowing, for example, the effect of geographic distance on trade to differ with the trust across countries. Overall, it appears that trust and social connectedness are substitutes in their effects on trade.

5 Conclusion

In this paper we use data from Facebook to construct the most comprehensive measure to date of social connectedness between pairs of countries. Using this measure, we show that social connectedness has a substantial impact on the patterns of bilateral trade between countries and regions through multiple economic channels. The richness of our findings of the impact of social connectedness suggests that our measure can be used in a broad range of settings beyond international trade. For example, within the field of economics, existing theoretical work suggests that the diversity of social networks is an important determinant of economic development; conversely, tightly clustered social ties can limit access to a broad range of social and economic opportunities (for example Granovetter, 1977, 2005). Exploring the literature beyond economics, researchers in international relations have argued that it is relationships between citizens of countries, rather than between political leaders, that are an important force for the maintenance of peaceful relationships between countries. Our measure could be used in these settings, as well as many others, to understand the impact of social connectedness globally. Indeed, we hope that future research will further broaden our understanding of the importance of social connectedness across a wide range of questions across the social sciences.

In this light, our research emphasizes the increasingly important role of data from online services—such as Facebook, LinkedIn, Twitter, eBay, Mint, Trulia, and Zillow—in overcoming important measurement challenges across the social sciences (see, for example, Baker, 2018; Giglio et al., 2015; Einav et al., 2015; Piazzesi et al., 2015). Specifically, we hope that the increasing availability of social network data, such as the *Social Connectedness Index* described in this paper, will help to improve our understanding of the effects of social connectedness on social, political, financial, and economic outcomes.

the United Kingdom.

References

Jenny C Aker. Information from markets near and far: Mobile phones and agricultural markets in niger. *American Economic Journal: Applied Economics*, 2(3):46–59, 2010.

Facundo Albornoz, Héctor F Calvo Pardo, Gregory Corcos, and Emanuel Ornelas. Sequential exporting. *Journal of International Economics*, 88(1):17–31, 2012.

Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic growth*, 8(2):155–194, 2003.

Treb Allen. Information frictions in trade. *Econometrica*, 82(6):2041–2083, 2014.

James E Anderson and Douglas Marcouiller. Insecurity and the pattern of trade: An empirical investigation. *Review of Economics and statistics*, 84(2):342–352, 2002.

James E Anderson and Eric Van Wincoop. Gravity with gravitas: a solution to the border puzzle. *American economic review*, 93(1):170–192, 2003.

Michael Bailey, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. House price beliefs and mortgage leverage choice. Technical report, National Bureau of Economic Research, 2017.

Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018a.

Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80, 2018b.

Michael Bailey, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, 126(6):2224–2276, 2018c.

Michael Bailey, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. Social connectedness in urban areas. *NBER Working Paper*, (w26029), 2019a.

Michael Bailey, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. Peer effects in product adoption. Technical report, National Bureau of Economic Research, 2019b.

Michael Bailey, Theresa Kuchler, Domninic Russel, and Johannes Stroebel. Social connectedness in europe. Technical report, Working Paper, 2019c.

Scott R Baker. Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, 126(4):1504–1557, 2018.

Richard Baldwin and James Harrigan. Zeros, quality, and space: Trade theory and trade evidence. *American Economic Journal: Microeconomics*, 3(2):60–88, 2011.

Daniel Berkowitz, Johannes Moenius, and Katharina Pistor. Trade, law, and product complexity. *the Review of Economics and Statistics*, 88(2):363–373, 2006.

Andrew B Bernard, Jonathan Eaton, J Bradford Jensen, and Samuel Kortum. Plants and productivity in international trade. *American economic review*, 93(4):1268–1290, 2003.

Andrew B Bernard, J Bradford Jensen, and Peter K Schott. Importers, exporters, and multinationals: A portrait of firms in the us that trade goods. Technical report, National Bureau of Economic Research, 2005.

Andrew B Bernard, J Bradford Jensen, Stephen J Redding, and Peter K Schott. Firms in international trade. *Journal of Economic perspectives*, 21(3):105–130, 2007.

Andrew B Bernard, J Bradford Jensen, Stephen J Redding, and Peter K Schott. The empirics of firm heterogeneity and international trade. *Annu. Rev. Econ.*, 4(1):283–313, 2012.

Antoine Berthou and Charlotte Emlinger. The trade unit values database. *Economie internationale*, (4): 97–117, 2011.

Konrad B Burchardi and Tarek A Hassan. The economic impact of social ties: Evidence from german reunification. *The Quarterly Journal of Economics*, 128(3):1219–1271, 2013.

Konrad B Burchardi, Thomas Chaney, and Tarek A Hassan. Migrants, ancestors, and investments. Technical report, National Bureau of Economic Research, 2016.

Thomas Chaney. Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–21, 2008.

Thomas Chaney. The network structure of international trade. *American Economic Review*, 104(11):3600–3634, 2014.

Thomas Chaney. Networks in international trade. In *The Oxford handbook of the economics of networks*. 2016.

Thomas Chaney. The gravity equation in international trade: An explanation. *Journal of Political Economy*, 126(1):150–177, 2018.

Natalie Chen. Intra-national versus international trade in the european union: why do national borders matter? *Journal of international Economics*, 63(1):93–118, 2004.

Pierre-Philippe Combes, Miren Lafourcade, and Thierry Mayer. The trade-creating effects of business and social networks: evidence from france. *Journal of international Economics*, 66(1):1–29, 2005.

Sergio Correia, Paulo Guimarães, and Thomas Zylkin. Verifying the existence of maximum likelihood estimates for generalized linear models. 2019a.

Sergio Correia, Paulo Guimarães, and Thomas Zylkin. ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects. 2019b.

Kingsley Davis. The population of india and pakistan. 1951.

Anne-Célia Disdier and Keith Head. The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics*, 90(1):37–48, 2008.

Maeve Duggan, Nicole B Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. Social media update 2014. pew research center., January 2015.

Maeve Duggan, Shannon Greenwood, and Andrew Perrin. Social media update 2016. Pew Research Center., November 2016.

Jonathan Eaton and Samuel Kortum. Technology, geography, and trade. *Econometrica*, 70(5):1741–1779, 2002.

Liran Einav, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, 7(2):215–47, 2015.

Eurostat. Freight transport statistics - model split. Technical report, 2015. URL https://ec.europa.eu/eurostat/statistics-explained/index.php/Freight_transport_statistics_-_modal_split.

Robert C Feenstra and John Romalis. International prices and endogenous quality. *The Quarterly Journal of Economics*, 129(2):477–527, 2014.

Guillaume Gaulier and Soledad Zignago. Baci: international trade database at the product-level (the 1994-2007 version). 2010.

Stefano Giglio, Matteo Maggiori, Johannes Stroebel, and Andreas Weber. Climate change and long-run discount rates: Evidence from real estate. Technical report, National Bureau of Economic Research, 2015.

Mark Granovetter. The impact of social structure on economic outcomes. *The Journal of Economic Perspectives*, 19(1):33–50, 2005.

Mark S Granovetter. The strength of weak ties. In *Social networks*, pages 347–367. Elsevier, 1977.

Avner Greif. Reputation and coalitions in medieval trade: evidence on the maghribi traders. *The journal of economic history*, 49(4):857–882, 1989.

Avner Greif. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American economic review*, pages 525–548, 1993.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Cultural biases in economic exchange? *The Quarterly Journal of Economics*, 124(3):1095–1131, 2009.

Juan Carlos Hallak. Product quality and the direction of trade. *Journal of international Economics*, 68(1): 238–265, 2006.

Juan Carlos Hallak and Peter K Schott. Estimating cross-country differences in product quality. *The Quarterly journal of economics*, 126(1):417–474, 2011.

Keith Head and Thierry Mayer. Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, volume 4, pages 131–195. Elsevier, 2014.

Elhanan Helpman, Marc Melitz, and Yona Rubinstein. Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, 123(2):441–487, 2008.

Ali Hortaçsu, F Martínez-Jerez, and Jason Douglas. The geography of trade in online transactions: Evidence from ebay and mercadolibre. *American Economic Journal: Microeconomics*, 1(1):53–74, 2009.

Matthew O Jackson, Brian W Rogers, and Yves Zenou. The economic consequences of social-network structure. *Journal of Economic Literature*, 55(1):49–95, 2017.

Robert Jensen. The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The quarterly journal of economics*, 122(3):879–924, 2007.

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi. The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2):220–246, 2011.

Amit Khandelwal. The long and short (of) quality ladders. *The Review of Economic Studies*, 77(4):1450–1476, 2010.

Edward E Leamer and James Levinsohn. International trade theory: the evidence. *Handbook of international economics*, 3:1339–1394, 1995.

Andrei A Levchenko. Institutional quality and international trade. *The Review of Economic Studies*, 74(3): 791–819, 2007.

Staffan Burenstam Linder. *An essay on trade and transformation*. Almqvist & Wiksell Stockholm, 1961.

John McCallum. National borders matter: Canada-us regional trade patterns. *The American Economic Review*, 85(3):615–623, 1995.

Marc J Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, 71(6):1695–1725, 2003.

Marc J Melitz and Gianmarco IP Ottaviano. Market size, trade, and productivity. *The review of economic studies*, 75(1):295–316, 2008.

Eduardo Morales, Gloria Sheu, and Andrés Zahler. Extended gravity. *The Review of Economic Studies*, 2015.

Nathan Nunn. Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600, 2007.

Maurice Obstfeld and Kenneth Rogoff. The six major puzzles in international macroeconomics: is there a common cause? *NBER macroeconomics annual*, 15:339–390, 2000.

Trevor J Pemberton, Michael DeGiorgio, and Noah A Rosenberg. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3: Genes, Genomes, Genetics*, 3(5):891–907, 2013.

Monika Piazzesi, Martin Schneider, and Johannes Stroebel. Segmented housing search. Technical report, National Bureau of Economic Research, 2015.

Richard Portes and Helene Rey. The determinants of cross-border equity flows. *Journal of international Economics*, 65(2):269–296, 2005.

Priya Ranjan and Jae Young Lee. Contract enforcement and international trade. *Economics & Politics*, 19 (2):191–218, 2007.

James E Rauch. Networks versus markets in international trade. *Journal of international Economics*, 48(1): 7–35, 1999.

James E Rauch. Business and social networks in international trade. *Journal of economic literature*, 39(4): 1177–1203, 2001.

James E Rauch and Vitor Trindade. Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130, 2002.

J. M. C. Santos Silva and Silvana Tenreyro. The log of gravity. *Review of economics and statistics*, 88(4): 641–658, 2006.

Tim Schmidt-Eisenlohr. Towards a theory of trade finance. *Journal of International Economics*, 91(1):96–112, 2013.

Peter K Schott. Across-product versus within-product specialization in international trade. *The Quarterly Journal of Economics*, 119(2):647–678, 2004.

Ina Simonovska and Michael E Waugh. The elasticity of trade: Estimates and evidence. *Journal of international Economics*, 92(1):34–50, 2014.

Enrico Spolaore and Romain Wacziarg. Ancestry, language and culture. In *The palgrave handbook of economics and language*, pages 174–211. Springer, 2016.

Enrico Spolaore and Romain Wacziarg. Ancestry and development: New evidence. *Journal of Applied Econometrics*, 33(5):748–762, 2018.

Meredith Startz. The value of face-to-face: Search and contracting problems in nigerian trade. *Available at SSRN 3096685*, 2016.

Claudia Steinwender. Real effects of information frictions: When the states and the kingdom became united. *American Economic Review*, 108(3):657–96, 2018.

Shawn W Tan. The geography of trade in the european single market. 2016.

Marcel P Timmer, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J De Vries. An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics*, 23(3):575–605, 2015.

Jan J Tinbergen. Shaping the world economy; suggestions for an international economic policy. 1962.

Daniel Trefler. The case of the missing trade and other mysteries. *The American Economic Review*, pages 1029–1046, 1995.

Appendix for "International Trade and Social Connectedness"

Michael Bailey
Abhinav Gupta
Sebastian Hillenbrand
Theresa Kuchler
Robert Richmond
Johannes Stroebel

A The Determinants of International Social Connectedness

A.1 Construction of SCI Data

In Section 2, we explored the determinants of social connectedness across countries. Data on social connectedness is not available for countries that have banned the use of Facebook (and other social media companies). These countries include China, Iran, North Korea, Tajikistan and Turkmenistan. Data is also not available for countries that have low population numbers such as Andorra, Dominica, Kiribati, St. Kitts and Nevis, San Marino and Tuvalu. We do not use data on territories and therefore exclude Curacao, Guam, Isle of Man, Jersey, Mayotte, Guadeloupe, French Guiana, Martinique, Puerto Rico, Reunion, and Western Sahara. Finally, we drop countries for which CEPII does not provide gravity variables, namely Kosovo, Montenegro, Serbia and South Sudan. We end up with data on social connectedness for 180 countries and 32,220 (= 180 x 179) observations.

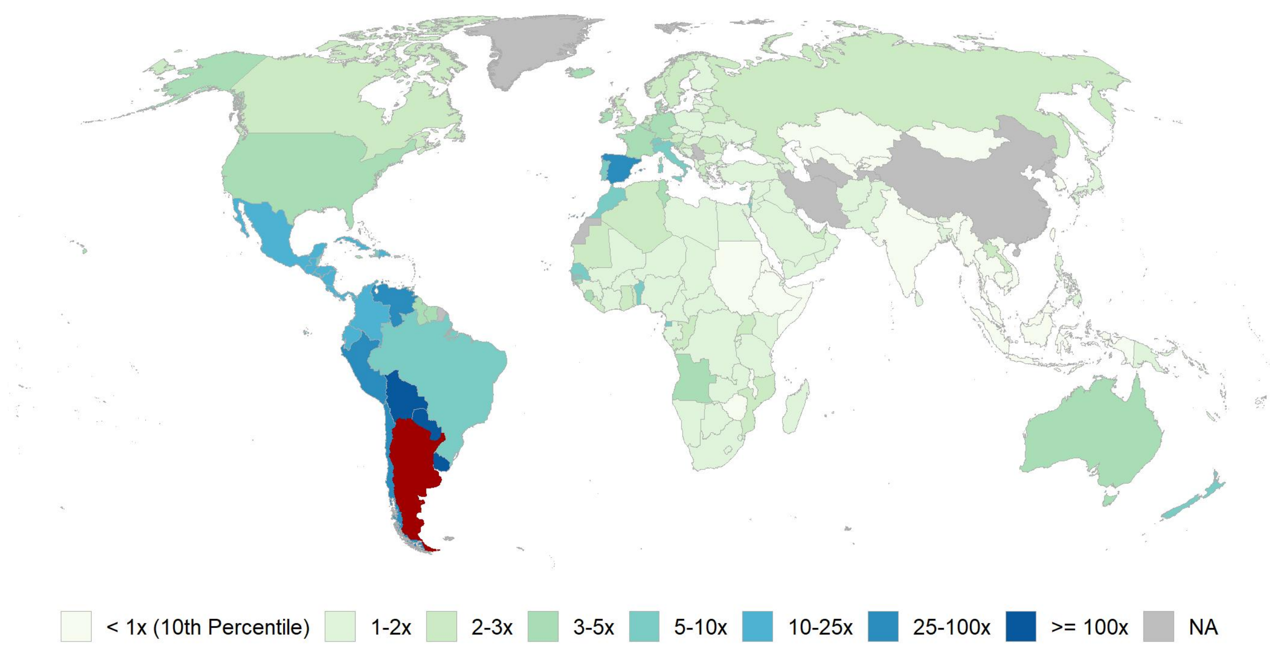
We additionally use data on migration, genetic and religious distance obtained from sources other than CEPII; more information on the sources can be found in the main text. The set of countries in these data can be different from the one for which we have international SCI data. More concretely, we do not observe migration data on Taiwan. In the genetic distance data, we lack information on 14 countries: Bosnia & Herzegovina, Cook Islands, Cayman Islands, Democratic Republic of the Congo, French Polynesia, New Caledonia, Macao, Maldives, Palestine, Sao Tome and Principe, Tanzania, Timor Leste, Togo, and Yemen. We also do not have the religious distance measure for 10 countries: Bosnia & Herzegovina, Belarus, Cook Islands, Cayman Islands, Democratic Republic of the Congo, Hong Kong, Macao, Palestine, the Republic of North Macedonia, and Timor Leste.

A.2 Additional Case Studies of International Social Connectedness

Argentina. Figure A.1 shows the social connectedness of Argentina. Argentina is strongly connected to all Spanish-speaking countries in Latin America. Connections to Portuguese-speaking Brazil are substantially lower, even though the two countries are geographically close and share a common border. Similarly, Argentina's strongest connection in Europe is to Spain. It is much less connected to Italy, despite the fact that Italians were the largest group of post-colonial immigrants (more than from Spain) and 60% of Argentinians have some Italian ancestry. These findings suggest an important role of shared language for today's connections.

Figure A.1: Social Connectedness to Argentina

(A) Social Connectedness to Argentina



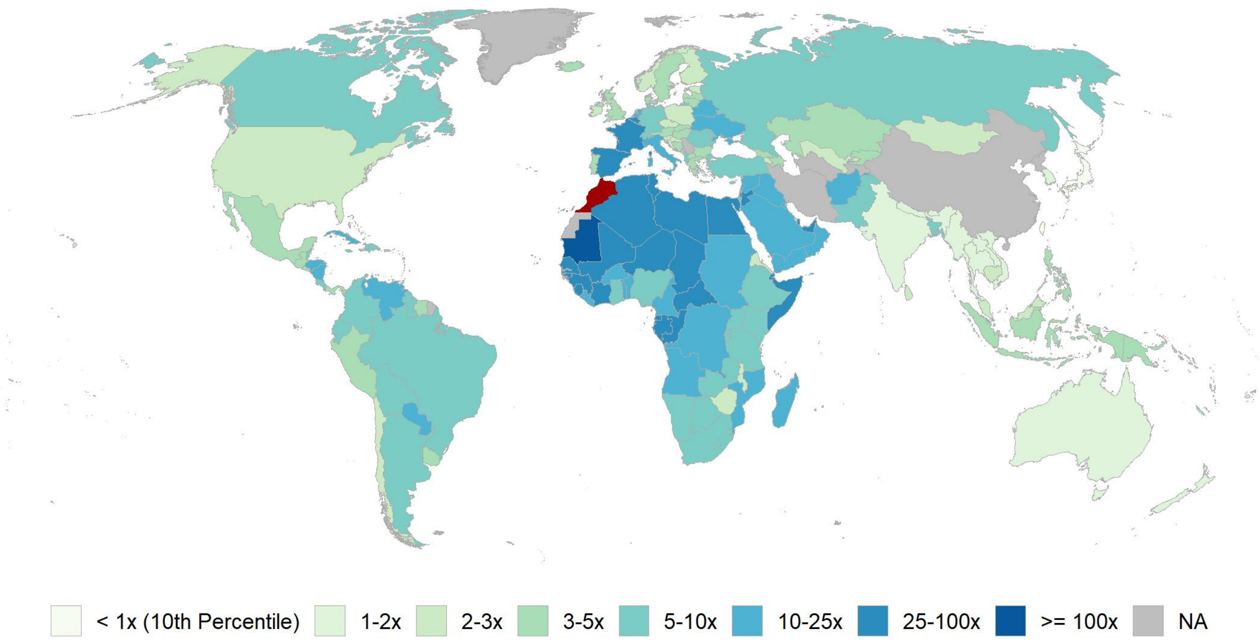
Note: Figure shows a heatmap of the social connectedness to Argentina. For each country in the world, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Morocco and Mauritania. Figure A.2 shows the social connectedness of two neighboring countries in Northwest Africa, Morocco (Panel A) and Mauritania (Panel B). Both Morocco and Mauritania are former French colonies and, as such, still have strong social ties to France. The populations of both countries are predominantly Muslim, which helps to explain their strong ties to other Muslim countries in Northern Africa and the Middle East. On the other hand, Mauritania has much stronger ties to sub-Saharan Africa than Morocco does. This is likely related to the fact that Morocco's population is almost entirely Arab-Berber, while Mauritania has a substantial population of individuals of Haratin and West African ethnicity. These patterns suggest that ethnic ties might be important in shaping friendships across countries.

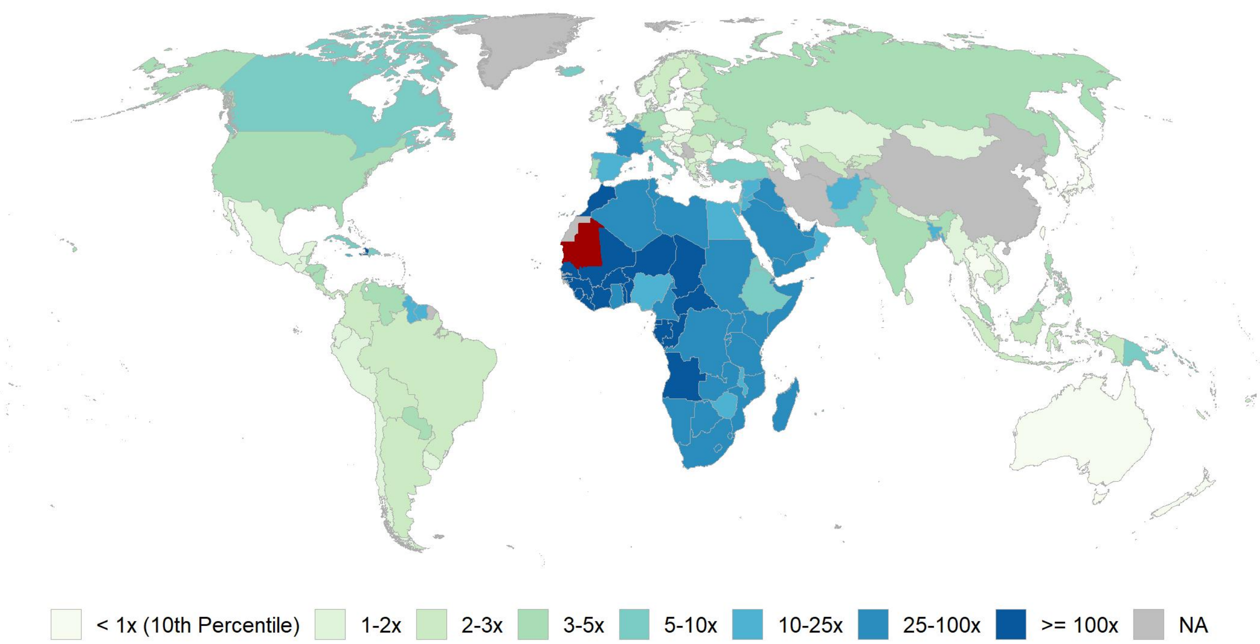
Azerbaijan and Turkey. Figure A.3 shows the social connectedness of Azerbaijan (Panel A) and Turkey (Panel B), two countries in Northwest Asia that share a short border. Both are strongly connected to each other, the nearby Caucasus countries of Armenia and Georgia, and Central European countries which have welcomed migrants from the two nations. However, Azerbaijan, a former Soviet Republic, is much more connected with countries in Europe and Central Asia that were also part of the Soviet Union, including Russia, Kazakhstan, Uzbekistan, Ukraine, Belarus, Lithuania, Latvia, and Estonia. Turkey, whose residents are predominately Muslim, is more connected to other predominately Muslim countries including Afghanistan, Syria, Lebanon, Iraq, Saudi Arabia, Yemen, and Libya. These patterns emphasize that historical ties and religion play important roles in shaping today's social connections.

Figure A.2: Social Connectedness of Morocco and Mauritania

(A) Social Connectedness to Morocco



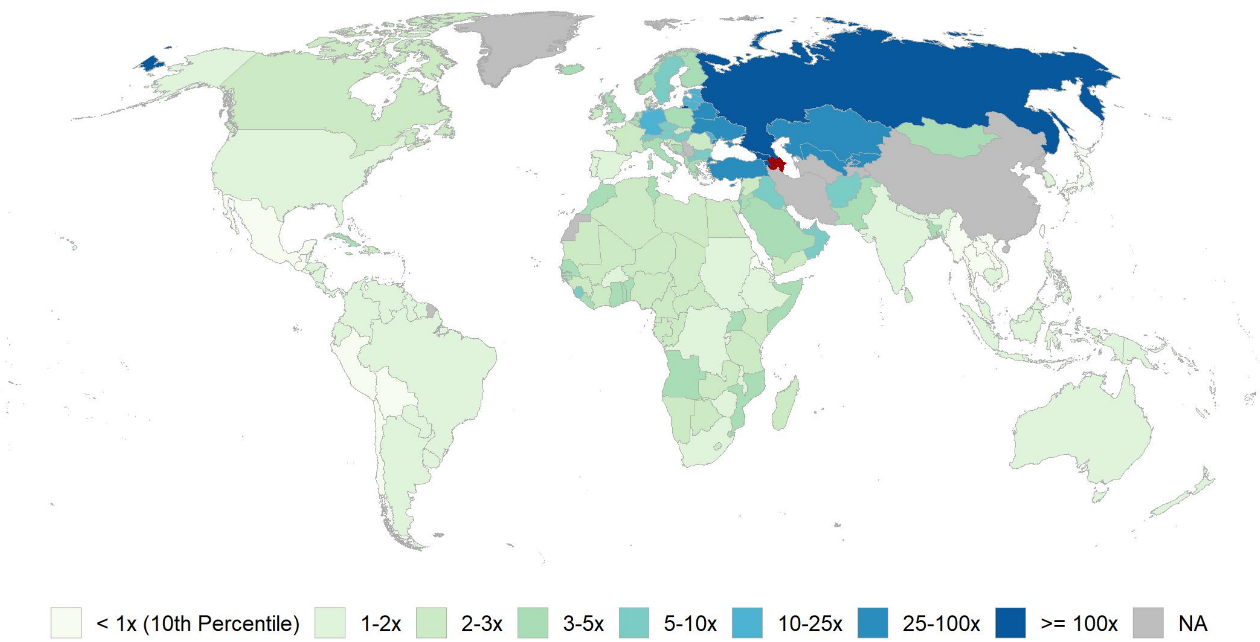
(B) Social Connectedness to Mauritania



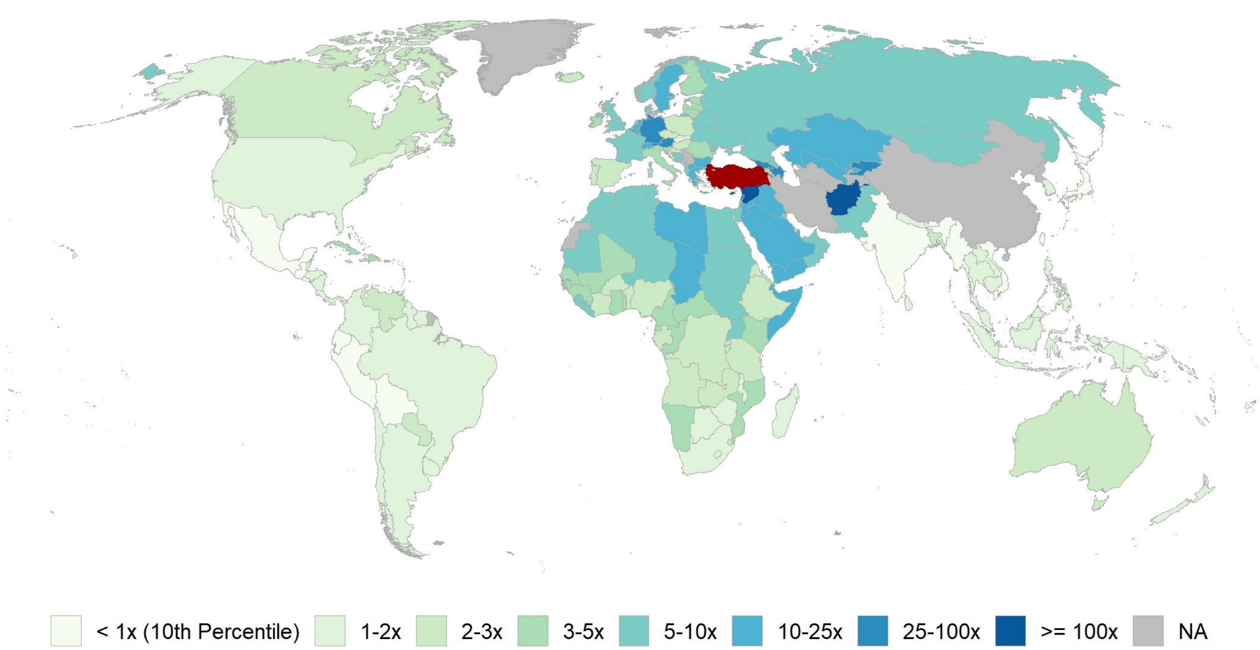
Note: Figure shows a heatmap of the social connectedness to Morocco (Panel A) and Mauritania (Panel B). For each country in the world, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Figure A.3: Social Connectedness to Azerbaijan and Turkey

(A) Social Connectedness to Azerbaijan



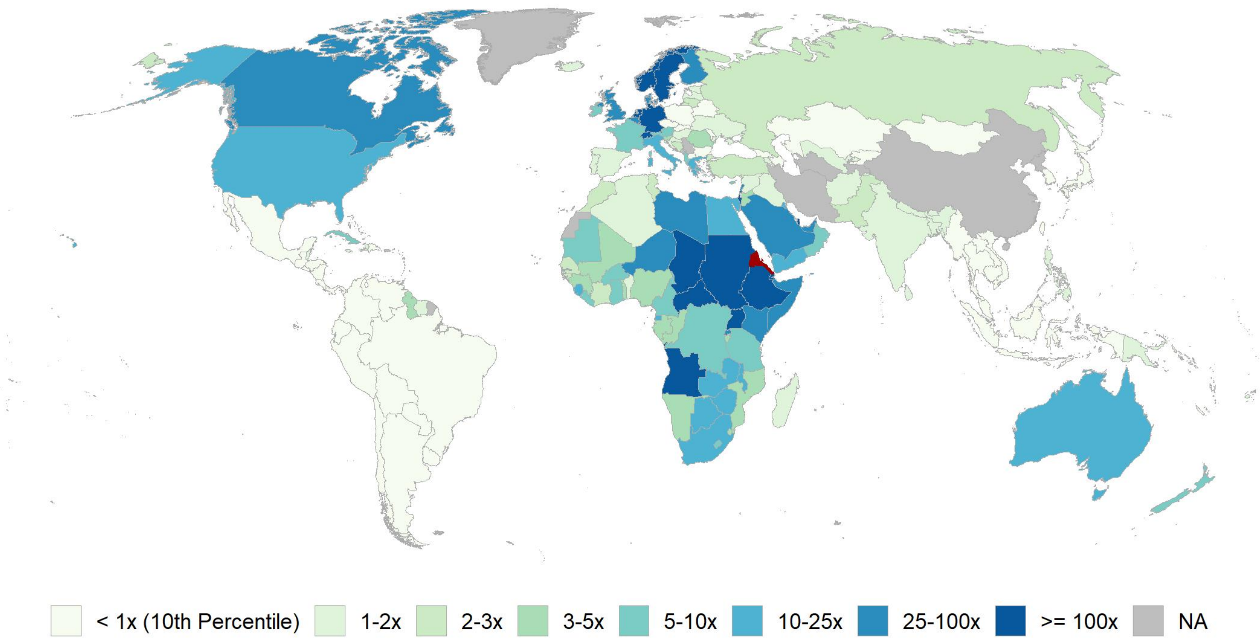
(B) Social Connectedness to Turkey



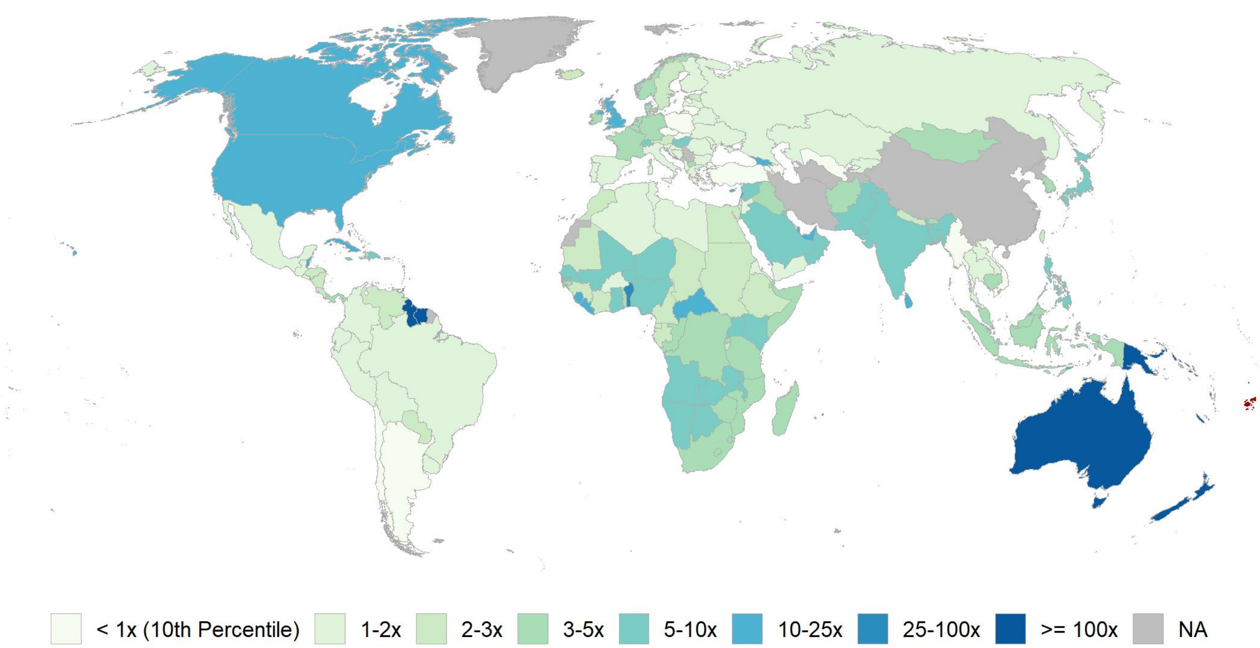
Note: Figure shows a heatmap of the social connectedness to Azerbaijan (Panel A) and Turkey (Panel B). For each country in the world, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Figure A.4: Social Connectedness to Eritrea and Fiji

(A) Social Connectedness to Eritrea



(B) Social Connectedness to Fiji



Note: Figure shows a heatmap of the social connectedness to Eritrea (Panel A) and Fiji (Panel B). For each country in the world, the colors highlight connections to the focal country, given in red. The lightest color corresponds to the 10th percentile of the connectedness across country-pairs globally; darker colors correspond to closer connections.

Eritrea and Fiji. Panel A of Figure A.4 shows the social connectedness of Eritrea. Eritrea is strongly connected to nearby countries in Africa including Ethiopia and Sudan. In addition, it is strongly connected to Israel and a number of countries in northern Europe, including Germany, Switzerland, Sweden, Norway, and the Netherlands. Eritrea had the ninth most exiting refugees in the world in 2018. While more than half of these refugees went to Ethiopia and Sudan, the other top countries of destination were Germany, Switzerland, Sweden, Norway, the Netherlands, and Israel.

Panel B of Figure A.4 shows the social connectedness of Fiji. Fiji, a former British colony, is strongly connected to other countries that were part of the British empire in Oceania (Australia, New Zealand, Papa New Guinea, and the Solomon Islands), as well as other English speaking nations, including Canada, the United States, the United Kingdom, and Guyana. The country's connections to Guyana and Suriname are particularly strong compared to its connections with countries throughout the rest of South America. A potential explanation for this lies in the Indian worker program described in Section 2.1. Between 1830 and 1930, over a million indentured laborers from India were relocated to European colonies, including Dutch Suriname and British Fiji and Guyana. These patterns suggest that ties from migratory movements significantly impact international connectedness.

A.3 Incremental Explanatory Power of Regressors for SCI

This section explores the explanatory power of the determinants of SCI when we remove each regressor one-by-one. Column 7 in Table 4 shows a multivariate regression of SCI on all explanatory variables. A number of these variables are highly correlated; for example, the correlation between genetic distance and our measure of migration is -0.44. Therefore, another metric of interest might be how much explanatory power each variable contributes when controlling for all other variables.

Table A.1: Incremental Explanatory Power of Regressors for SCI

	Dependent variable: log(SCI)
	Incremental R^2
log(Distance)	1.95
log(1+Migrant Population)	3.72
Common Colonizer	0.22
Colonial Relationship	0.23
Genetic Distance	1.93
Common Official Language	0.84
Religious Distance	0.15
Δ GDP per Capita (in '00,000$s)	0.41
Common Border	0.02
Same Continent	0.61
Same Subcontinent	0.01

Note: The table reports the incremental R^2 of each regressor when included/excluded in a regression of the logarithm of SCI on the full panel of regressors. This is alike the specification in column 7 of Table 4. To get the incremental R^2, we compute the difference between the R^2 of a regression where we exclude the variable from the set of explanatory variables and the R^2 when we include the variable in the set of explanatory variables. For each variable considered, we run both regressions on the same set of observations.

For each variable, we run two regressions with the specification equal to column 7 of Table 4; one where we excludes the variable of interest and one where we include it. We then compare the R-squareds of the two regressions, giving us an estimate of the incremental explanatory power of the variable beyond all other explanatory variables. The results for this exercise are reported in Table A.1. We find that migration has the largest incremental R-squared with 3.74% followed by distance with 1.97%. Genetic distance (1.93%) and the common official language dummy (0.80%) also add sizeable explanatory power. Colonial heritage, religious distance, and the common border and same subcontinent dummies have very little explanatory power once we control for other factors.

A.4 Groups of Socially Connected Countries

In Table A.2 below, we list the countries in the 30 clusters described in Section 2.3.

Table A.2: 30 Clusters of Socially Connected Countries

Cluster	Countries
1	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo, Cote d'Ivoire, Democratic Republic of the Congo, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Swaziland, Togo, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
2	Antigua and Barbuda, Bahamas, Barbados, Canada, Cayman Islands, Grenada, Guyana, Jamaica, Saint Lucia, Saint Vincent and the Grenadines, Trinidad and Tobago, United States
3	Egypt, Iraq, Israel, Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Palestine, Saudi Arabia, Sudan, Syrian Arab Republic, Yemen
4	Albania, Austria, Bosnia and Herzegovina, Croatia, Germany, Italy, Malta, Slovenia, Switzerland, the Republic of North Macedonia
5	Denmark, Estonia, Finland, Iceland, Ireland, Latvia, Lithuania, Norway, Sweden, United Kingdom
6	Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Russia, Ukraine, Uzbekistan
7	Australia, Cook Islands, Fiji, New Zealand, Papua New Guinea, Solomon Islands, Tonga, Vanuatu
8	Bahrain, Bangladesh, Brunei Darussalam, Malaysia, Maldives, Oman, Singapore, Sri Lanka
9	Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama
10	Argentina, Bolivia, Cuba, Paraguay, Spain, Uruguay
11	Bhutan, India, Nepal, Qatar, United Arab Emirates
12	Bulgaria, Cyprus, Greece, Republic of Moldova, Romania
13	Colombia, Ecuador, Peru, Venezuela
14	Comoros, Madagascar, Mauritius, Seychelles
15	Cape Verde, Luxembourg, Portugal, Sao Tome and Principe
16	Djibouti, Eritrea, Ethiopia, Somalia
17	Hong Kong, Macau, Philippines, Taiwan
18	Burma, Cambodia, Lao People's Democratic Republic, Thailand
19	Belgium, Netherlands, Suriname
20	Chile, Dominican Republic, Haiti
21	Czech Republic, Hungary, Slovakia
22	Algeria, Morocco, Tunisia
23	France, French Polynesia, New Caledonia
24	Afghanistan, Pakistan
25	Indonesia, Timor-Leste
26	Japan, Viet Nam
27	Republic of Korea, Mongolia
28	Brazil
29	Poland
30	Turkey

Note: The table reports 30 groups generated by hierarchical agglomerative linkage clustering.

B Trade and Social Connectedness

B.1 Contruction of Trade and SCI Data

For the analyses in Section 3, we merge data on the international social connectedness with product-level trade data from CEPII. Of the 180 countries for which we have data on international social connectedness, six countries are not contained in the trade data from CEPII. These countries are Botswana, Lesotho, Luxembourg, Namibia, Sudan, and Swaziland. In the original data, products are classified according to the 6-digit HS1996 classification into 4,914 product categories. In order to construct total bilateral trade, we aggregate trade across all 96 products for each exporter-import pair. We replace missing trade values with zero trade values. Our final data that contains both social connectedness and aggregate bilateral trade includes 174 countries and 30,102 observations.

In Section 3.1, we use data on product-level trade instead of aggregate-level trade. For computational reasons, we aggregate trade up to the first 2 digits of the HS1996 product code for each exporter-importer pair.[1] This procedure results in 96 product categories. We then replace missing trade values with zero values. The panel on product-level trade and social connectedness contains 2,889,792 (=174 x 173 x 96) observations.

Additionally, we use rule of law measures from the Worldwide Governance Indicators provided by the World Bank. This data lacks four countries: Cook Islands, New Caledonia, Palestine, and French Polynesia. The data that contains product-level trade, international SCI and the rule of law measure includes 170 countries and 2,758,080 (=170 x 169 x 96) observations.

B.2 Gravity Regressions (OLS) - Intensive Margin of Trade

Table 3 shows results from regressing aggregate bilateral trade on social connectedness and other gravity variables. The regression is estimated using Poisson Pseudo Maximum Likelihood (PPML) to account for zero bilateral trade between countries. Here, we show that the results are robust to estimating the relationship using OLS. We focus on the intensive margin of trade in order to avoid problems with zero-trade observations. We estimate the following regression:

$$log(X_{i,j}) = \beta + \gamma G_{i,j} + \delta_i + \delta_j + \epsilon_{i,j} \tag{A.1}$$

The dependent variable $log(X_{i,j})$ denotes the logarithm of the total value of exports from country i to country j. As before, social connectedness is successful in explaining a substantial part of the variation in bilateral trade at the intensive margin. Column 2 shows that social connectedness explains 28.1% of the within variation of trade after controlling for exporter and importer fixed effects, nearly as high as the within-R^2 of 28.9% for distance (see column 3). Similarly to Table 3, column 4 shows that gravity variables together explain a much smaller share of variation—here, less than half as much—than social connectedness. Column 7 shows that, after controlling for distance and other gravity variables, the coefficient on social connectedness is similar to that obtained using PPML (0.387 vs. 0.325).

[1] The first 2 digits of the product code are referred to as the "HS chapter". One example is for example this chapter 09, which includes " Coffee, Tea, Maté and Spices".

Table A.3: Gravity Regressions (OLS) - Intensive Margin of Trade

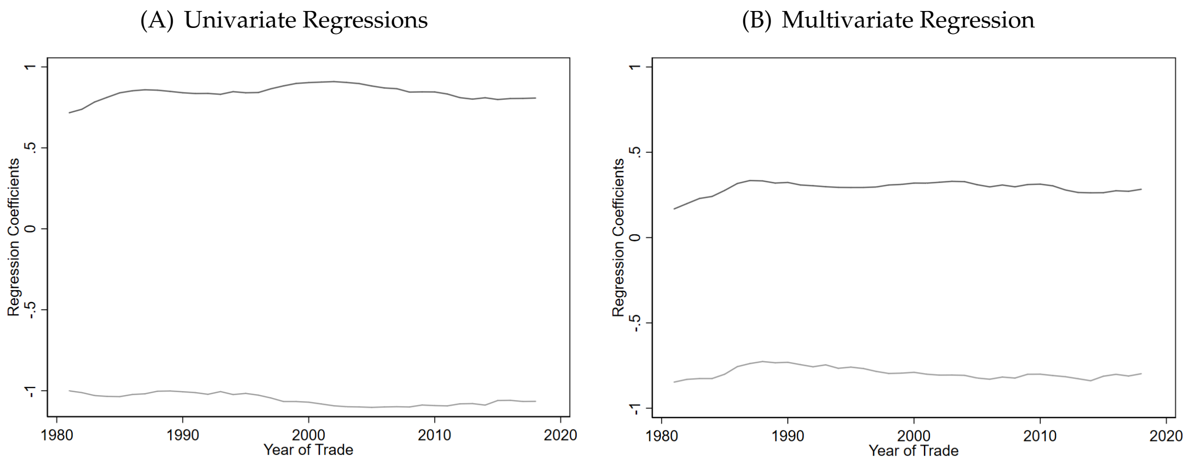
	Dependent variable: Aggregate Bilateral Trade							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(SCI)		0.813***			0.454***		0.387***	500
		(0.030)			(0.033)		(0.034)	Quantiles
log(Distance)			-1.706***		-1.027***	-1.534***	-1.026***	-1.019***
			(0.063)		(0.072)	(0.064)	(0.073)	(0.072)
Common border				3.573***		0.866***	0.660***	0.467***
				(0.222)		(0.182)	(0.169)	(0.155)
Common official language				1.289***		0.586***	0.241***	0.266***
				(0.151)		(0.088)	(0.081)	(0.083)
Common colonizer				0.816***		0.628***	0.350***	0.276**
				(0.184)		(0.140)	(0.118)	(0.120)
Colonial relationship				0.630**		1.168***	0.629***	0.591***
				(0.249)		(0.176)	(0.134)	(0.132)
Orig. and Dest. Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	20,054	20,054	20,054	20,054	20,054	20,054	20,054	20,054
Adjusted R^2	0.679	0.769	0.772	0.719	0.785	0.778	0.787	0.788
Adjusted within R^2		0.281	0.289	0.125	0.330	0.310	0.336	0.337

Note: Table shows results from regression A.1. We estimate the regression on non-zero trade observations (intensive margin) using OLS. The dependent variable is total exports from country i to country j. Controls include the logarithm of SCI, the logarithm of distance, a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, a dummy indicating whether the pair of countries was in a colonial relationship post 1945 and 500 quantiles of SCI. All specifications include exporter and importer country fixed effects. Standard errors are clustered by exporter and importer country. We have data on trade and social connectedness for 174 countries, which leads to 20,054 non-zero trade observations. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

B.3 Time-variation in Trade Elasticities

In this section we examine how the elasticity of trade with respect to social connectedness and distance varies over time. For this purpose, we use *The Direction of Trade Statistics* from the IMF that reports aggregate bilateral trade data from 1948 to 2018. Because the earlier years contains a lot of missing trade entries and we want to keep the set of countries constant across time, we focus on the time period from 1981 to 2018. For each year, we regress the bilateral trade on our distance measure and our social connectedness measure from March 2019. This allows to year-specific regression coefficients that are shown in Figure A.5.

Figure A.5: Time-variation in Elasticity of trade to Social Connectedness



Note: Figures show year-specific regression coefficients from regressing trade on $\log(SCI)$ and $\log(Distance)$ as specified in Equation 3. Panel A shows the regression coefficients obtained in univariate regressions, Panel B shows the regression coefficients obtained in a multivariate regression. The dark grey line shows the coefficient on social connectedness, while the light grey line shows the coefficient on distance. Both regression specifications include exporter and importer country fixed effects.

B.4 Gravity Regressions - A Horse Race of Predictors

In Table 3 we regress aggregate bilateral trade on the logarithm of social connectedness and all other gravity variables. However, it might be instructive to look at how trade interacts with each of these variables individually, exploring the relative economic and statistical significance of each variable. In this section, we conduct this analysis using both PPML regression and an OLS specification on the intensive margin of trade. Table A.4 shows this horse race among variables on the aggregate data using a PPML specification, as formalized in equation 3. Table A.5 shows the results of a similar horse race regression for the intensive margin of trade following regression A.1.

Consistent with our other results, we find that distance and social connectedness are the most successful in explaining variation in bilateral trade. In the PPML regression, distance and social connectedness are the only two variables that individually explain more than 90% of bilateral trade after controlling for exporter and importer country fixed effect. In the OLS intensive margin regression, distance and social connectedness explain 29% and 28% of the within variation in trade, compared to between 0.4% and 7.5% for other gravity variables.

Table A.4: Gravity Regressions Horse Race - PPML

	Dependent variable: Aggregate Bilateral Trade					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Distance)	-0.996***					
	(0.06)					
log(SCI)		0.683***				
		(0.04)				
Common Border			1.900***			
			(0.204)			
Common Official Language				0.936***		
				(0.146)		
Common Colonizer					1.439***	
					(0.156)	
Colonial Relationship						0.607*
						(0.325)
Orig. and Dest. Country FE	Yes	Yes	Yes	Yes	Yes	Yes
N	30,102	30,102	30,102	30,102	30,102	30,102
Adjusted R^2	0.929	0.919	0.887	0.845	0.839	0.835

Note: Table shows results from regression 3. We estimate the regression using PPML. The dependent variable is total exports from country i to country j. Controls include the logarithm of SCI, the logarithm of distance, a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, and a dummy indicating whether the pair of countries was in a colonial relationship post 1945. All specifications include exporter and importer country fixed effects. Standard errors are clustered by exporter and importer country. We have data on trade and social connectedness for 174 countries, which leads to 30,102 (= 174 x 173) observations. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table A.5: Gravity Regressions Horse Race - Intensive Margin of Trade (OLS)

	Dependent variable: Aggregate Bilateral Trade					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Distance)	-1.706***					
	(0.063)					
log(SCI)		0.813***				
		(0.03)				
Common Border			4.205***			
			(0.230)			
Common Official Language				1.809***		
				(0.163)		
Common Colonizer					1.564***	
					(0.199)	
Colonial Relationship						1.544***
						(0.404)
Orig. and Dest. Country FE	Yes	Yes	Yes	Yes	Yes	Yes
N	20,054	20,054	20,054	20,054	20,054	20,054
Adjusted R^2	0.772	0.769	0.703	0.699	0.688	0.680
Adjusted within R^2	0.289	0.281	0.0741	0.0628	0.0291	0.00464

Note: Table shows results from regression A.1. We estimate the regression on non-zero trade observations (intensive margin) using OLS. The dependent variable is total exports from country i to country j. Controls include the logarithm of SCI, the logarithm of distance, a common border dummy, a common official language dummy, a dummy indicating whether the two countries had a common colonizer post 1945, and a dummy indicating whether the pair of countries was in a colonial relationship post 1945. All specifications include exporter and importer country fixed effects. Standard errors are clustered by exporter and importer country. We have data on trade and social connectedness for 174 countries, which leads to 20,054 non-zero trade observations. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

C Trade and Sub-National Social Connectedness in Europe

C.1 Construction of Data for Input-Output Weighted Connectedness

Our analyses in Section 4.1 use information on employment at the industry/NUTS2 region level, mapped to input-output data at the industry/country level and trade data at the product/country level. (As described in the text, we use product and industry interchangeably.) The final analyses include 28 countries for which all three sets of data were available.[2] The industry employment data come from the Eurostat Structural Business Statistics series, which includes employment in NUTS2 regions for NACE Rev. 2 industry classifications at the division level.[3] In each region we use the most recent year the data were

[2] These countries are Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Slovenia, Slovakia and the United Kingdom.

[3] Notably, the SBS series does not cover agriculture, forestry, and fishing. For more information on the series see: https://ec.europa.eu/eurostat/web/structural-business-statistics. For more information on NACE Rev. 2 classifications see: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-07-015.

available starting with 2017. 64% of the data comes from 2017, 26% from 2016, 1.6% from 2015, 1.3% from 2014, and 3.22% from between 2008 and 2013. Observations prior to 2016 may be categorized using different NUTS2 regions, as these boundaries periodically change. In instances when we use an observation prior to 2016 in a region that changed, we use a crosswalk described in Section C.2, below.[4] For each industry and country, we calculate the share of employment in each region (e.g. the share of Greek construction industry workers that are in the Attica Region).

We then match this data to the World Input-Output data by country and industry of origin and destination. Our mapping from NACE Rev. 2 industry classifications to the World Input-Output classifications comes from correspondence tables provided by Eurostat. We then add product-level trade data from CEPII by mapping the Harmonized Commodity Description and Coding System (HS) 1996 product classifications to the World Input-Output industry classifications. This mapping comes from correspondence tables provided by the World Bank, UN Statistics Division, and Eurostat. For the purpose of our analysis, our focus is on goods that are used as an intermediate input to another production process in the country of destination. Accordingly, we drop industries for which more than half of the exports are used for final consumption. Our final analysis includes the 20 industries listed in Table A.6.

C.2 Construction of Rail Freight Data

Our analyses in Section 4.2 use information on European region-to-region rail goods transportation from Eurostat. The data are based on individual reports from European Union members, European Free Trade Association members, and European Union candidates. For each observation in the data, the reporting country is identified. The data are reported for 2005, 2010, and 2015, in accordance with Directive 80/1177/EC of the European Commission and subsequent regulatory updates. We use the 2015 data and take a number of steps to clean it for our analyses. This process was informed by both the "Reference Manual on Rail transport statistics" and correspondence with the Eurostat data providers.

We first restrict the data to observations that are at the NUTS2 regional level, including country-level data for countries which consist of only a single NUTS2 region. We also exclude all pairs that include a region with the unknown indicator "XX" or the extra-regio territory indicator "ZZ."[5] From here, we are faced with four challenges: 1) As confirmed by the authors' correspondence with Eurostat, when the data appear as "non-available" in a particular row, this could mean either that there was no rail traffic or that the relevant country did not provide the data.[6] 2) There are a number of hypothetical regional pairs missing, even between countries that did report data elsewhere. 3) For some international regional pairs, there are data reported from both countries on the same train flows, and often the tons of goods transported does not match. 4) The 2015 data are reported by 2013 NUTS2 region, while our social connectedness and distance data are by 2016 NUTS2 region.

With respect to challenges 1 and 2, we use the fact that each country reports data to Eurostat in two intermediate data sets: one for domestic transport of goods and another for international transport of goods. To identify countries that submitted a particular set of data in a particular year, we group the

[4]Here, the necessary assumption we make for regions that split is that employment by industry in each new region is proportional to 2015 populations.

[5]Observations with these two codes makeup 4.9% of tonnage transported in the data.

[6]In some instances, countries report the data to Eurostat, but flag them as confidential so that they are not included in the public release. We always treat these data as missing in our final analysis.

data by the reporting country, year, and whether the region pair is international or domestic. We then generate a list of countries that had at least one non-missing entry in each year/domestic-international group. These lists are provided in Table A.7. When "non-available" values are reported by a country that *did not* report data elsewhere in the year/domestic-international group, we treat the observation as missing and exclude it. When "non-available" values are reported by a country that *did* report data elsewhere in the group, we treat this value as a zero (no traffic). Additionally, for countries that reported data in a particular group, we fill any missing regional pairs (i.e. pairs that are not in the data) in the group with zeros. Together, these assumptions handle challenges 1 and 2.

For each international regional pair, there still remains two possible reports: one from each of the regions' home countries in the pair. In instances when only one country reports the data, we take the non-missing value from the reporting country. However, there are a number of instances when each country reports data for the same international regional pair (challenge 3). In these instances, we take the average of the two reports.[7] Finally, to update the data to the 2016 NUTS2 regions (challenge 4) we build a crosswalk using the history of NUTS information provided by Eurostat.[8] In instances when a 2013 NUTS2 region split into multiple regions, we set the tons of goods transported in each of the 2016 NUTS2 observations equal to the corresponding 2013 NUTS2 region tons of goods transported, multiplied by the region's 2015 population share of the 2013 NUTS2 region's 2015 population (i.e. we assume that tons of goods transported in each of these regions is proportional to the 2015 population).

[7] In a couple instances, a "third-party" country will report transport between regions in two other countries. We exclude these observations for our analysis.

[8] Available at: `https://ec.europa.eu/eurostat/web/nuts/history`

Table A.6: Products Used in Input-Output Weighted Regressions

Industry
Architectural and engineering activities; technical testing and analysis
Electricity, gas, steam and air conditioning supply
Manufacture of basic metals
Manufacture of chemicals and chemical products
Manufacture of coke and refined petroleum products
Manufacture of computer, electronic and optical products
Manufacture of electrical equipment
Manufacture of fabricated metal products, except machinery and equipment
Manufacture of machinery and equipment n.e.c.
Manufacture of motor vehicles, trailers and semi-trailers
Manufacture of other non-metallic mineral products
Manufacture of other transport equipment
Manufacture of paper and paper products
Manufacture of rubber and plastic products
Manufacture of wood/products of wood and cork, except furniture; manufacture of straw articles and plaiting materials
Mining and quarrying
Other professional, scientific and technical activities; veterinary activities
Other service activities
Printing and reproduction of recorded media
Publishing activities

Note: Table shows the 20 industries that are included in the input-output weighted analyses in Section 4.1. Industry descriptions come from the NACE Rev. 2 European statistical classifications of economic activities.

Table A.7: Rail Freight Data Availability By Reporting Country

Reporting Country	Domestic Data	International Data
Albania	N	N
Austria	N	N
Belgium	N	N
Bulgaria	Y	Y
Croatia	Y	Y
Cyprus	N	N
Czech Republic	Y	Y
Denmark	Y	Y
Estonia	Y	Y
Finland	Y	Y
France	N	N
Germany	Y	Y
Greece	N	N
Hungary	N	N
Iceland	N	N
Ireland	Y	Y
Italy	Y	Y
Latvia	Y	Y
Liechtenstein	N	N
Lithuania	Y	Y
Luxembourg	Y	Y
Malta	N	N
Montenegro	N	N
Netherlands	Y	N
Norway	Y	Y
Poland	Y	Y
Portugal	Y	N
Romania	Y	N
Serbia	N	N
Slovakia	Y	Y
Slovenia	Y	Y
Spain	Y	Y
Sweden	N	N
Switzerland	N	N
The Republic of North Macedonia	N	N
Turkey	Y	N
United Kingdom	N	N

Note: Table shows the rail goods transportation data availability by reporting country and by type of trade (domestic or international). Y (N) indicates the data is (not) available. The table only shows availability at the reporter level, not whether any regions from this country are included in the final analysis. For example, although Austria did not report international data in 2015, pairs that include an Austrian region and a region in a country that did report international data in 2015 are included.