NBER WORKING PAPER SERIES

INTERNATIONAL TRADE AND SOCIAL CONNECTEDNESS

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

Working Paper 26960
http://www.nber.org/papers/w26960

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

This research was facilitated through a research consulting agreement between some of the academic authors (Kuchler and Stroebel) and Facebook. Bailey is an employee at Facebook. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

International Trade and Social Connectedness
Michael Bailey, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert J. Richmond, and Johannes Stroebel
NBER Working Paper No. 26960
April 2020
JEL No. F1,F6,G0

ABSTRACT

We use anonymized data from Facebook to construct a new measure of the pairwise social connectedness between 180 countries and 332 European regions. We find that two countries trade more with each other when they are more socially connected and when they share social connections with a similar set of other countries. The social connections that determine 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. Once we control for social connectedness, the estimated effect of geographic distance on trade declines substantially, and the effect of country borders disappears. Our findings suggest 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 relationships between social connectedness and trade flows.

Michael Bailey
Facebook
Economist
mcbailey@fb.com

Abhinav Gupta
New York University
agupta4@stern.nyu.edu

Sebastian Hillenbrand
New York University
shillenb@stern.nyu.edu

Theresa Kuchler
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
tkuchler@stern.nyu.edu

Robert J. Richmond
Leonard N. Stern School of Business
Kaufman Management Center
44 West Fourth Street, 9-81
New York, NY 10012
rrichmon@stern.nyu.edu

Johannes Stroebel
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012
and NBER
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 mechanisms behind the large estimated negative effect of distance on trade — have remained elusive. One prominent explanation is that geographic closeness 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 regions and 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 332 European regions, and show that much of the variation in global trade flows can indeed be explained by patterns of direct and indirect social connectedness. Specifically, it is the social connections between the regions where a product is produced in the exporting country and those where it is used in the importing country that predict trade in that product. Once we control for social connectedness, the effect of geographic distance on trade declines substantially, and the effect of country borders essentially disappears. We also find support for mechanisms proposed by theoretical models in which social connectedness causes trade. In particular, we find evidence consistent with social connections facilitating trade by reducing information asymmetries and by providing a substitute for formal mechanisms of contract enforcement. In addition, we can rule out that our results are primarily driven by omitted country-level variables or by reverse causality from trade to social connectedness. In the end, while the absence of quasi-random variation in social connectedness prevents us from conclusively establishing a causal relationship between social connectedness and trade, our combined evidence is highly consistent with such a relationship, and no alternative explanation can jointly explain our findings.

Our measure of social connectedness is based on an anonymized snapshot of all 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 any pair of regions corresponds to the relative probability of a friendship link between a Facebook user in the first region and a Facebook user in the second region.[2] We construct this pairwise measure of social connectedness for 180 countries and 332 European regions (see also Bailey et al., 2019d, for more details on the European Social Connectedness Index). We

[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 later, 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; Badarinza et al., 2019).

[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, 2019b) 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.

argue that social networks as measured by Facebook provide a good 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 first describe the rich patterns of social connectedness observed in our data. About half the variation in social connectedness between countries is explained by geographic distance. Quantitatively, a 10% increase in the distance between two countries is associated with a 10% − 15% decline in their social connectedness. 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 language, as well as between countries that are similar in terms of economic development, religious beliefs, and the genetic make-up of their populations. Within Europe, common language and common history shape the social connectedness between regions over and above distance and common nationality. Beyond these systematic patterns, our measure of social connectedness is also affected by idiosyncratic factors that are specific to particular country and region pairs.

We then document that patterns of social connectedness explain a substantial part of the variation in international trade flows. We begin by exploring trade across countries and then turn to studying trade and social connectedness between European regions. When we introduce social connectedness into a standard gravity model of country-level trade, we find that social connectedness and geographic distance explain similar shares of the cross-sectional variation in trade flows. The elasticity of trade 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 about −1 to roughly −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 trade flows, and highlights that the relationship between geographic distance and trade in the prior literature might partially capture this role of social connectedness.

We then explore possible explanations for the observed relationship between trade flows and social connectedness. In particular, we provide evidence consistent with the prominent narrative that social links can facilitate trade by alleviating a number of informal trade barriers that are regularly discussed in the literature, namely contract enforcement frictions and search costs due to information asymmetries (see Chaney, 2016). 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 that shows 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. This finding suggests 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 in different products. In particular, Rauch (1999) and Rauch and Trindade (2002) suggest that information asymmetries are largest for

[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.

products that are 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 to social connectedness is particularly large for these non-exchange-traded goods.

To provide further evidence that social connectedness may increase trade by reducing information asymmetries, we study how unit prices of exports vary with the social connectedness between countries. This test builds on evidence that suggests that search costs inhibit trade by constraining traders in their ability to find the best prices (e.g., Eaton and Kortum, 2002; Jensen, 2007; Aker, 2010; Allen, 2014; Simonovska and Waugh, 2014; Startz, 2016). Social connections are one potential channel through which traders can obtain information on prices and thereby overcome these search costs (Chaney, 2014). Consistent with predictions from this literature, we show that higher social connectedness between countries is associated with lower equilibrium unit prices. We also find that the effect of social connectedness on lowering prices is largest for goods that are not traded on organized exchanges, whereas there is essentially no effect on equilibrium prices for exchange traded goods for which prices are more transparent.

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 can increase bilateral trade through similar channels as direct social connections. The first such mechanism is reducing information asymmetries by decreasing search costs, whereby a common friend in a third country can pass information between potential trading partners.[4] The second channel is by improving contract enforcement, which is easier in a shared social community that can impose sanctions (see Greif, 1989, 1993). To test whether being in the same social cluster increases trade flows, we exploit the broad coverage of our measure of international social connectedness. In particular, since we observe the social connectedness between the vast majority of countries, we can use a clustering algorithm to group countries based on social connectedness. We find that countries in the same social cluster have substantially higher trade than predicted based on their bilateral social connectedness alone. For example, Spain is approximately equally well-connected to the U.K. 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 U.K.

After documenting how social connectedness relates to trade flows at the country level, we use our granular measure of social connectedness across European regions to further understand the mechanism behind the observed relationship. Our results in this section provide new evidence on the interaction of trade, the spatial distribution of production, and the structure of social networks. Our findings also help us to rule out country-level omitted variables or reverse causality as alternative interpretations of the observed aggregate relationship between social connectedness and trade.[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,

[4] Chaney (2014) develops 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 section complements 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.

we construct product-specific measures of social connectedness between countries, which overweight the 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 between two countries, which corresponds to the population-weighted average connectedness between all regions in the 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 proposed measure of social connectedness relevant for exporting non-metallic mineral products from Italy to Greece thus overweights the observed 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).

We then regress product-level trade between countries on both measures of social connectedness. When controlling for the product-specific measure of social connectedness, the baseline population-weighted measure has no further predictive power for trade at the product-level. This remains true after controlling for product-specific measures of distance. This evidence suggests that it really is the social connectedness between the regions where a product is produced and the regions where it is used that matters for trade in that product. We also find that the elasticity of trade to the product-specific measures of social connectedness is unaffected by the inclusion of country pair fixed effects, which absorb the population-weighted measures of social connectedness in addition to any other country-level determinants of trade. This finding dramatically reduces the scope for omitted variables such as common preferences to explain the observed relationships between social connectedness and trade.

Our analysis of the effects on trade of 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 did in fact cause substantial social connections, the various product-specific measures of social connectedness between two countries should 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, as a result of the connections formed from trading non-metallic mineral products. 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 study the relationship between regional social connectedness and subnational trade flows. We use rail-freight flows between regions in the European Union as our proxy of trade flows. This analysis allows us to examine the determinants of the border effect — the empirical regularity that, conditional on the distance between two regions, trade is much larger between regions 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 connections between regions, the estimated border effect drops by more than 80% and becomes statistically indistinguishable from zero. This suggests that much of the effect of borders on trade is related to the fact that social connections fall at borders.

The region-level data on trade and social connectedness also allow 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 presents our new measure of international social connectedness and explores its determinants. Section 2 describes the relationship between international trade flows and patterns of social connectedness, focusing on heterogeneities across products and by countries. In Section 3, 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 and European regions 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 U.S. 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 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. For users in Europe, we also observe their region of location at the NUTS2 (Nomenclature of Territorial Units for Statistics level 2 regions) level, similar to Bailey et al. (2019d). These NUTS2 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. They are generally based on existing subnational administrative borders. For example, in Italy the NUTS2 geographies correspond to the 21 "regions", while in the Netherlands they correspond to the 12 "provinces"; smaller countries such as Latvia and Malta are represented by a single NUTS2 region. Location in a country or region is 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 locations 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 location i and individuals in location j, divided by the product of the number of Facebook users in those locations, as in Equation 1:

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

For both countries and regions, 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 Facebook user in location i and a given user in location j.

In our country-level analysis, we exclude countries that have very low populations, that are territories of other countries, or that ban 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. At the NUTS2 level, we observe the social connectedness between $332 \times 331 = 109,892$ pairs of European regions.

Interpreting the Social Connectedness Index. Two important questions arise when interpreting $SCI_{i,j}$ as a proxy for the social connectedness between countries or regions: whether Facebook friendships correspond to real-world friendship links of Facebook users, and whether Facebook users are representative of the countries' or regions' 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 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 also capturing good 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, 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, 2019a,c).

On the second issue, it is likely that the representativeness of Facebook users will differ across locations. 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 penetrations, 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 then overstate the international linkages of the average resident in countries with low internet usage. Differential online access is less of an issue across our European regions. Indeed, statistics from Eurostat show that in 2019, 90% of households in the European Union had access to the internet. However, Facebook usage rates conditional on internet access can still vary in systematic ways across these regions. As a result, one might worry that Facebook users in regions with lower Facebook penetrations are connected to other regions at rates that are not representative of

[6] Due to a ban of Facebook, data are 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.

the full populations in those regions. In our analysis, we account for such heterogeneity in the average connectedness of each country and region by including fixed effects for locations i and j in all specifications. This approach allows us to explore connectedness between locations i and j, holding fixed the average propensity in each location of having Facebook friends in different locations.

In the end, 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, provides a valuable large-scale measure of people's social connections. Indeed, it is hard to imagine an alternative measure that would allow us to measure social connections at this scale and scope. We hope that the easy accessibility of our measures of social connectedness will facilitate more research on the role of social connectedness in economics and across the social sciences.

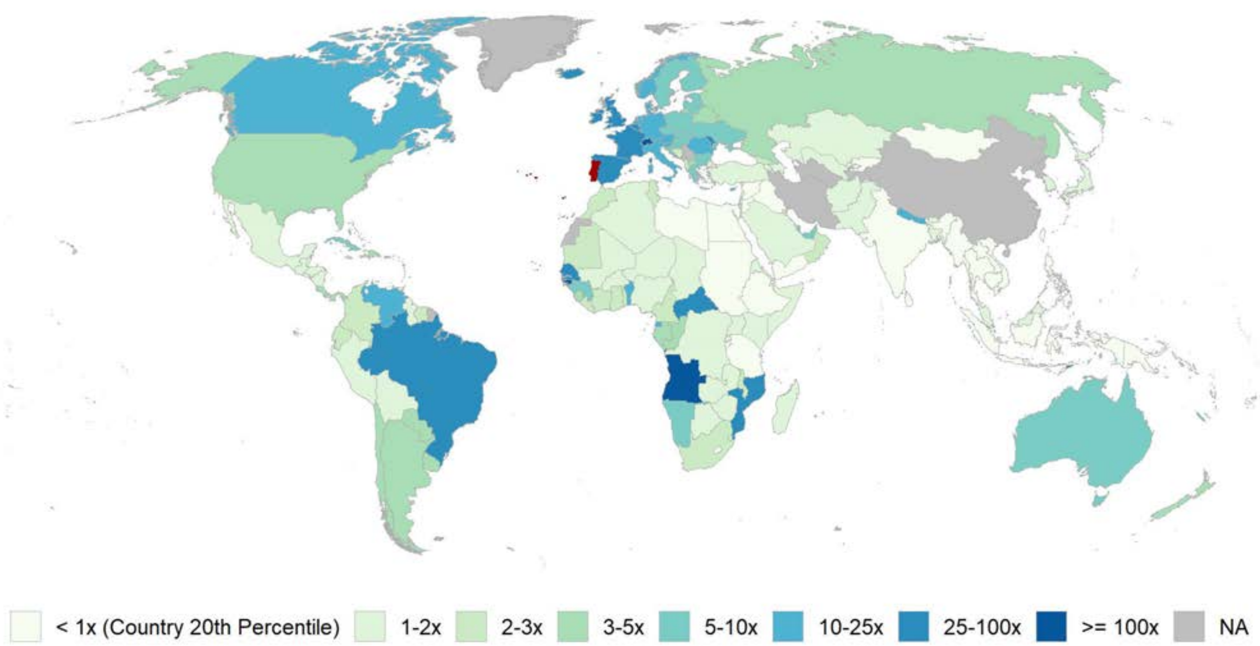
Determinants of Social Connectedness. We next explore a number of factors that help explain the observed patterns of social connectedness across locations. We summarize our central findings in the main body of the paper and present a more extensive analysis in Appendix A.

We begin by exploring a few case studies that highlight how social connectedness varies across specific countries and regions. Panel A of Figure 1 shows the social connectedness of Portugal to other countries around the world. Darker colors correspond to higher connectedness. Portugal has the strongest links to geographically close countries in Western Europe. Portugal's international connections also highlight the role of colonial history and language in shaping present-day social connectedness. The country is strongly connected to its former (Portuguese-speaking) colonies Brazil, Angola, Guinea-Bissau, and Mozambique. Within Europe, Portugal is most strongly connected to 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, following large-scale migration from Portugal to Luxembourg as part of a guest worker program in the 1960s. This finding suggests that past migration movements continue to influence social connections today.

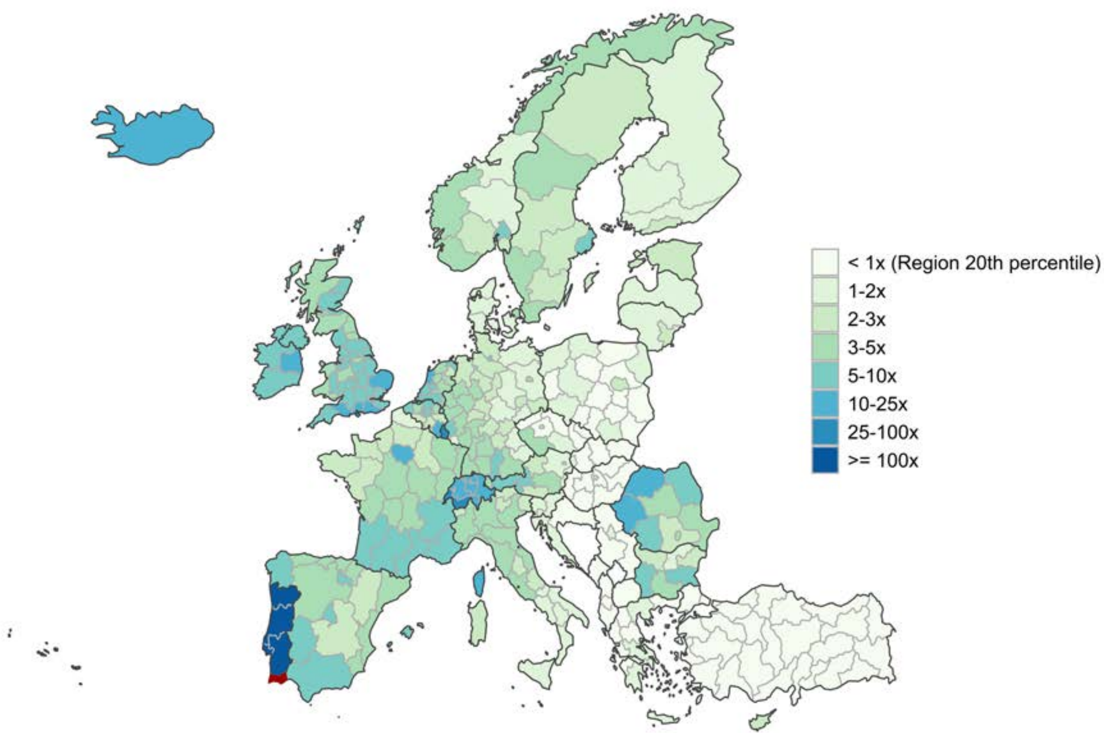
In Appendix A, we explore the determinants of international social connectedness more systematically. We find that a 10% increase in the distance between two countries is associated with a $10\% - 15\%$ decline in their social connectedness. Geographic distance explains about 50% of the variation in social connectedness that remains after accounting for country fixed effects. Consistent with our findings for Portugal, international migration patterns and colonial history strongly influence the probability of present-day friendship links across all country pairs. We also find more social connections between countries sharing a common language, as well as between countries that are similar in terms of economic development, religious beliefs, and the genetic make-up of their populations. However, while about 70% of the variation in social connectedness across country pairs can be explained by distance, language, and other systematic factors, our *Social Connectedness Index* also captures a wide variety of idiosyncratic forces that can determine the social connections between two countries. For example, citizens of Denmark and Australia are 75% more connected than would be predicted by the observable factors described above. These strong social connections between Denmark and Australia are likely the result of the 2004 marriage of the Danish Crown Prince Frederik to Australian-born Mary Donaldson. This marriage led to heightened mutual interest between Danish and Australian citizens, and has sub-

Figure 1: Social Connectedness of Portugal and Algarve Region

(A) Social Connectedness of Portugal to other countries



(B) Social Connectedness of the Algarve Region in Portugal to European regions



Note: Figure shows a heat map of the social connectedness of Portugal to other countries (Panel A) and of the Algarve region in Portugal to other European regions (Panel B). For each location, the colors highlight connections of the focal location, given in red. The lightest color corresponds to the 20th percentile of the connectedness across country pairs that include Portugal in Panel A and region pairs that include Algarve in Panel B; darker colors correspond to closer connections.

stantially increased tourism between the two countries.[7] Examples such as this highlight the power of our approach to measuring social connectedness over and above competing approaches that proxy for social connectedness using a variety of other gravity variables.

Panel B of Figure 1 shows the connectedness of the Algarve region in southern Portugal to other regions within Europe. The strongest social links are to other regions in Portugal. Indeed, the Algarve region is much more strongly connected to the Norte region in the very north of Portugal than it is to the Andalusia region, its neighbor just across the border in Spain. The Algarve's connections to other European regions show some of the same patterns seen for Portugal as a whole, such as the strong connections to Luxembourg. However, additional nuance is visible at the regional level. At the country level, Portugal showed strong connections to France. When exploring regional social connectedness, we find that connections from Algarve are particularly strong to Southern France and Corsica, which has a substantial number of Portugese immigrants. The Algarve also has strong connections to much of Western Europe, from where it attracts substantial numbers of tourists each year.

The forces highlighted in Panel B of Figure 1 also show up in more systematic analyses. Indeed, Bailey et al. (2019d) show that social connectedness within Europe varies with patterns of migration, political borders, geographic distance, language, and other demographics. The elasticity of social connectedness with respect to distance across European regions is similar, at -1.3, as it is across countries. Social connectedness drops off sharply at country borders, even after controlling for distance: depending on the country, the probability of friendship between two individuals living in the same country is five to eighteen times as large as the probability across two individuals living in different countries. In addition, regions that are more similar along demographic measures such as language, religion, education, and age are more socially connected. Interestingly, the relationship between political borders and connectedness can persist many decades after boundaries change. For example, Bailey et al. (2019d) finds higher social connectedness across regions that were originally part of the Austro-Hungarian empire, even after controlling for a host of other determinants of present-day social connectedness.

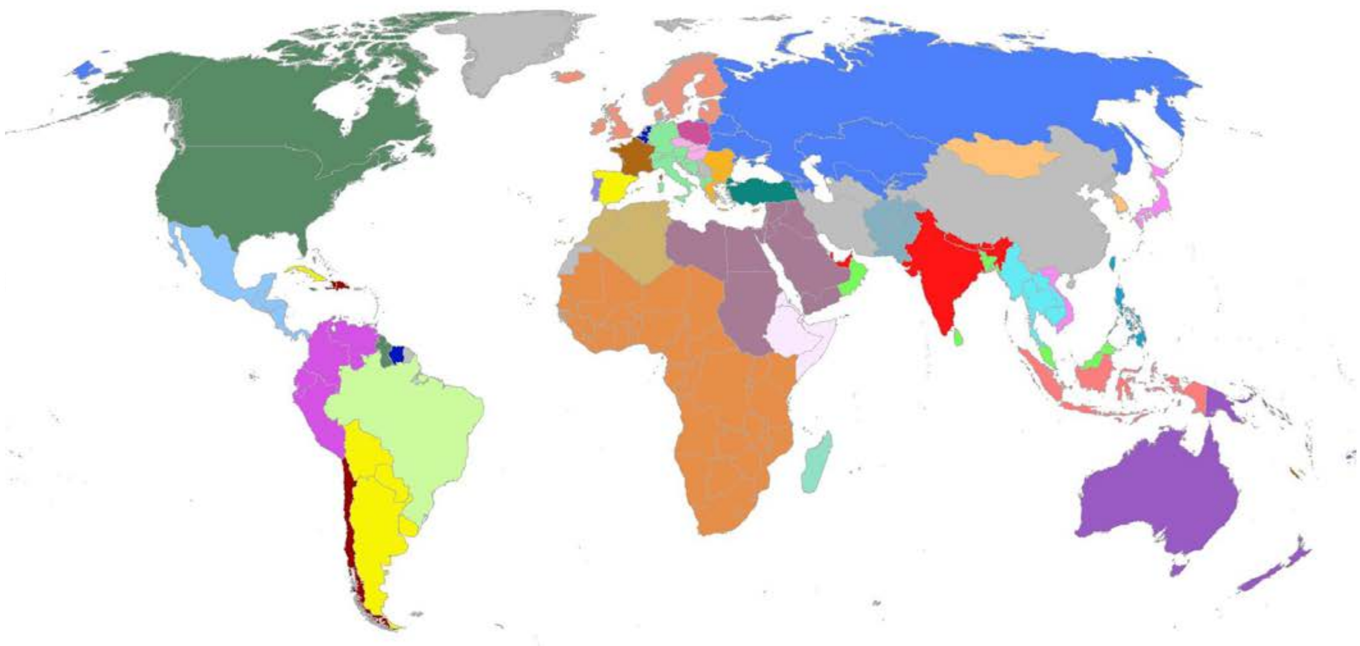
Groups of Socially Connected Countries. One advantage of our measure of social connectedness is that it covers the vast majority of countries, allowing us to understand the structure of global social connectedness beyond just bilateral connections. In particular, we described above that countries that share certain characteristics are more socially connected to each other. Such 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. We next formalize the idea of groups of countries with strong connections amongst each other. In Section 2.2 we use this grouping to show that trade patterns are influenced not only by direct social connections, but also by the social clusters that different countries are part of.

There are a number of possible algorithms to construct clusters of countries that feature a high average within-cluster pairwise social connectedness. Here, we use hierarchical agglomerative linkage

[7] The Australian Department of Foreign Affairs and Trade notes on its country brief on Denmark that "Australia's profile in Denmark, and vice-versa, was boosted by the marriage in May 2004 of Australian-born Mary Donaldson to Denmark's Crown Prince Frederik" and the press reported an increase of 30% in tourism from Australia to Denmark in 2004.

clustering to group countries into 30 clusters.[8] Figure 2 shows the 30 different clusters and Table A.6 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.

Figure 2: 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 socially connected countries.

Different characteristics shape the different clusters. For instance, all countries in Central America and Mexico form one cluster, consistent with 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 cluster between Chile, Haiti, and the Dominican Republic is likely explained by recent migrant movements from Haiti and the Dominican Republic to Chile. The cluster comprised of Mongolia and South Korea reflects both a long common history and common ethnic ancestry, as well as recent migration of Mongolians to South Korea. While standard characteristics can explain some of the variation in clustering, it is important to note that these clusters are constructed from our social connectedness measure that has substantial variation beyond standard gravity characteristics.

[8]Conceptually, the agglomerative 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.

2 Country-level Trade and Social Connectedness

We now turn to understanding the relationship between social connectedness and trade flows. We first analyze how trade patterns vary with social connectedness at the country level, before exploring the channels through which social connectedness can increase trade. In particular, we present evidence that suggests that social connectedness increases trade through two channels discussed by Chaney (2016): by reducing informational frictions and by improving contract enforceability. We also show that two countries that are connected to a similar set of other countries trade more than predicted solely based on their bilateral connectedness. In Section 3, we exploit our granular region-level measures of social connectedness within Europe to document new patterns about the interaction of the spatial structure of social networks, the location of production, and international trade.

2.1 Aggregate Bilateral Trade

We measure country-level trade flows using 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 our results are robust to using trade data from earlier years. The trade data is disaggregated at the 6-digit HS96 code level, and contains information on 4,914 product categories. For our first analysis, we aggregate the product-level trade data into bilateral trade flows between country pairs. When merged with social connectedness and gravity data, we have information on 30,102 country pairs.[9] To understand the relationship between social connectedness and trade flows, we follow the literature to estimate the following gravity regression:

$$X_{i,j} = \exp\left[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(Distance_{i,j}) + \beta_3 G_{i,j} + \delta_i + \delta_j\right] \cdot \epsilon_{i,j}, \tag{2}$$

where $X_{i,j}$ denotes the total value of exports from country i to country j, $G_{i,j}$ captures country pair characteristics, δ_i and δ_j are exporter and importer fixed effects, and $\epsilon_{i,j}$ is an error term. We estimate regression 2 using Poisson Pseudo Maximum Likelihood (PPML) to account for zero bilateral trade between some country pairs (see the discussion in Santos Silva and Tenreyro, 2006).[10] 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 from the binscatter plots in Figure 3, which show that the relationship between exports and social connectedness is approximately log-linear. The importer and exporter fixed effects in regression 2 control for country-level characteristics such as population, GDP, and average tariffs, all of which affect the overall level of trade. These fixed effects also control for country-specific differences in the use of Facebook that might affect our measure of social connectedness.

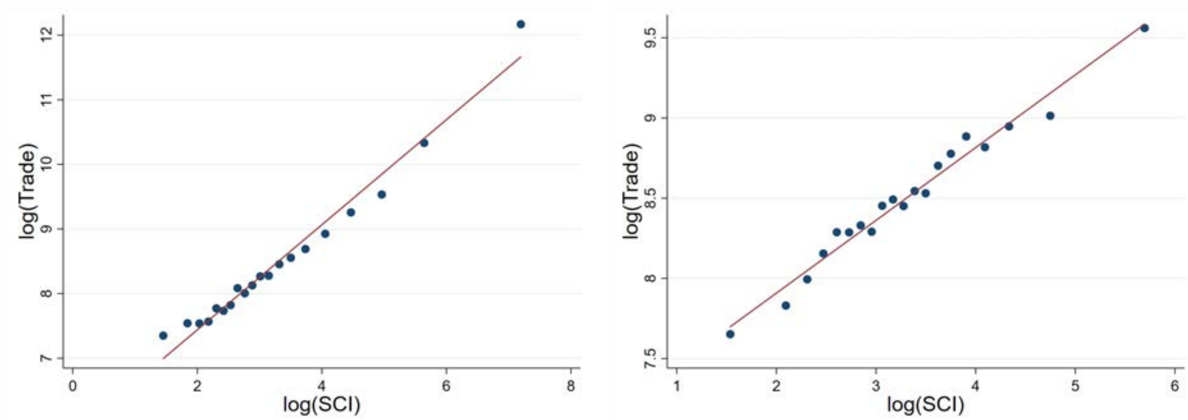
The results from estimating regression 2 are presented in Table 1. Column 1 shows results from a specification with only importer and exporter fixed effects. The R^2 highlights that 83.3% of the varia-

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

[10] Our estimation uses the algorithm in Correia et al. (2019a) and Correia et al. (2019b). In the Appendix, we present estimates of regression 2 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.

Figure 3: Trade vs. Social Connectedness

(A) Country Fixed Effects (B) Country Fixed Effects and Distance Controls



Note: Figures show binscatter plots of aggregate bilateral trade and social connectedness. Panel A regresses log(Exports) 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.

tion in bilateral trade flows is explained by these fixed effects alone. This finding is 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 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.

In column 3, we remove the control for social connectedness, and instead include controls for the geographic distance between countries. The elasticity of trade with respect to distance is -1.00. This magnitude 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 A.2, the elasticity of trade to social connectedness drops from 0.68 to 0.33, and the distance elasticity drops from -1.00 to -0.66. The additional control for social connectedness increases the R^2 by 0.8% relative to that in column 3, which only included controls for distance in addition to importer and exporter fixed effects. This increase in the R^2 corresponds to about 10% of the variation not already explained by distance and fixed effects, and is larger than the incremental variation explained by the other gravity variables (see column 6). The decline in the distance elasticity from adding controls for social connectedness between columns 3 and 5

Table 1: Gravity Regressions - Aggregate Trade

	Dependent variable: Aggregate Exports						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(SCI)		0.683***			0.325***		0.325***
		(0.040)			(0.029)		(0.030)
log(Distance)			-0.996***		-0.660***	-0.863***	-0.563***
			(0.060)		(0.071)	(0.054)	(0.066)
Common Border				1.763***		0.439***	0.413***
				(0.220)		(0.124)	(0.105)
Common Official Language				0.169		0.064	-0.131
				(0.145)		(0.102)	(0.086)
Common Colonizer				0.958***		0.233	0.040
				(0.153)		(0.142)	(0.137)
Colonial Relationship				0.445**		0.019	-0.065
				(0.219)		(0.377)	(0.276)
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y
R^2	0.833	0.919	0.929	0.891	0.937	0.932	0.939
N	30,102	30,102	30,102	30,102	30,102	30,102	30,102

Note: Table shows results from regression 2. 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).

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 frictions 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 baseline 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 other predictors. Interestingly, the estimated elasticity of trade with respect to social connectedness remains unchanged relative to the estimates from column 5, even though the newly added gravity variables are 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.

2.2 Mechanisms Through Which Social Connectedness Can Affect Trade Patterns

We now turn to understanding the mechanisms driving the positive relationship between social connectedness and trade. As discussed above, while we cannot conclusively establish that higher social connectedness causes increases in trade, we find evidence that is highly consistent with a number of plausible causal channels behind the observed positive relationship.

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

Contract enforcement frictions are a second source of trade barriers that might be mitigated through social connections. 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 argued that ethnic networks can provide reputation-based punishment for contract violations 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.[11] To do so, we use trade disaggregated data, where the unit of observation is an export of product k from country i to country j. A product corresponds to one of 96 unique 2-digit HS96 categories. We estimate the following regression specification:

$$X_{i,j,k} = \exp[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(SCI_{i,j}) \cdot ET_k + \\ \beta_3 \log(SCI_{i,j}) \cdot RL_i + \beta_4 \log(SCI_{i,j}) \cdot RL_j + \beta_5 G_{i,j,k}] \cdot \epsilon_{i,j,k}. \tag{3}$$

In this specification, ET_k is the fraction of exchange traded goods for each 2-digit HS96 product category k.[12] RL_i and RL_j are continuous measures of the rule of law in the exporting and importing countries,

[11] 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.

[12] To construct this measure, we start from trade data at the 6-digit HS96 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 2-digit HS96 level using the total global share of trade in those goods within each 2-digit HS96 category. Across products, ET_k ranges from 0 to 0.91. It has a mean value of 0.12, and a standard deviation of 0.25. To provide a sense of the variation, within

as measured by the World Governance Indicators (Kaufmann et al., 2011).[13] As before, $G_{i,j,k}$ are a set of gravity variables and fixed effects. All specifications include origin country × product and destination country × 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).

Table 2: Gravity Regressions - Heterogeneity

	Dependent variable: Product-Specific Exports						
	(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 3. The dependent variable is exports of product category *k* from country *i* to country *j*. Product-level trade data are aggregated up to the first 2 digits of the HS96 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 include 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 the fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

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% of value-weighted goods are exchange traded, and within category HS-80 (tin and articles thereof) 90% are exchange traded.

[13] 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.

We present the results from estimating regression 3 in Table 2. Column 1 shows our baseline specification from column 7 of Table 1 for 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 very similar to our baseline specification presented in Table 1. In column 2, we interact $\log(SCI_{i,j})$ with the fraction of exchange traded products in each product category. The coefficient on this interaction is -0.188, suggesting that social connectedness matters substantially less for trade in product categories that have more exchange traded goods. Quantitatively, the elasticity is more than twice as large in a category with no exchange traded goods than it is in a category with primarily exchange traded goods. This finding provides suggestive 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 than using product-specific interactions with $log(Distance_{i,j})$, 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. The results are essentially unchanged in this specification.

Columns 4 and 5 interact our measures of the rule of law in the destination and origin countries with the social connectedness across country pairs. 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. This magnitude implies that, 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.[14] This finding suggests that social connections mitigate some of the contract enforcement frictions that have been shown to reduce trade in countries with weak institutions. 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 variation in the trade elasticity to social connectedness across the rule of law of the trading 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 at least two channels. First, by helping to alleviate information asymmetries and search costs, especially for non-exchange-traded products where information asymmetries are the highest. Second, by improving contract enforcement when the quality of institutions in importing countries is low.

Social Connectedness and Prices. To further understand the mechanisms through which social connectedness influences trade patterns, we next study the relationship between social connectedness and the unit prices of traded goods. This test is motivated by empirical evidence that search costs intro-

[14]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.

duce trade frictions by reducing traders' ability to find the best import prices (Jensen, 2007; Aker, 2010; Allen, 2014; Startz, 2016). These frictions should be larger for products where prices are more obscure, such as those that are not traded on formal exchanges. In the presence of these search frictions, social connections are one potential channel through which traders can obtain information on products and prices (Chaney, 2014). By exploring how equilibrium prices vary with the social connectedness between countries, and how this differs across goods with different levels of price transparency, we can provide evidence consistent with this mechanism.

We analyze unit prices at the 6-digit HS96 level as provided by Berthou and Emlinger (2011) — to our knowledge, the data set provides the most disaggregated publicly available information on unit prices in international trade. Following a large literature that works to disentangle quality from price (for example, Schott, 2004; Hallak, 2006; Hallak and Schott, 2011; Khandelwal, 2010; Feenstra and Romalis, 2014), the data construction starts from product-level trade data at a high level of disaggregation to control for differences in quality. The data are then converted into common-weight units and outliers are removed in the cross-section and time-series, before prices are aggregated to the 6-digit HS96 level. With these price data, we estimate the following regression:

$$\log(P_{i,j,k}) = \beta_1 \log(SCI_{i,j}) + \beta_2 \log(SCI_{i,j}) \cdot \mathbb{1}_k^{ET} + \beta_3 G_{i,j,k} + \epsilon_{i,j,k}. \tag{4}$$

In this regression, $P_{i,j,k}$ is the the per-unit value of exports of product category k from country i to country j. Prices are exclusive of transport costs and tariffs. The variable $\mathbb{1}_k^{ET}$ is an indicator if the products in category k are traded on an exchange, and $G_{i,j,k}$ are a set of gravity variables and fixed effects. We include importer $\times$ product fixed effects and exporter $\times$ product fixed effects to control for differences in the average price of products in each category that are imported or exported by different countries; this adjusts for heterogeneity across countries in which precise products within each product category they are trading. We also control for the same other gravity variables as in Tables 1 and 2.

The results are presented in Table 3. Column 1 shows that trade unit prices are lower when countries are more socially connected: a doubling of social connectedness lowers prices of traded goods by about 11%. In column 2, we interact the control for social connectedness with an indicator for exchange traded goods. We see that social connectedness is mostly associated with lower prices for goods not traded on organized exchanges. This finding provides additional evidence that trade prices are affected by informational frictions such as search costs, and that social connectedness can help alleviate these frictions. In columns 3 and 4, we add flexible product-specific controls for the geographic distance between the countries, to account for the fact that prices often increase in the geographic distance between locations (see Hummels and Skiba, 2004; Baldwin and Harrigan, 2011). When we control non-linearly for distance, we still find that a doubling of social connectedness leads to an approximately 6% decrease in unit prices. Additionally, in column 4 we find that the effect of social connectedness on prices is entirely driven by goods which are not traded on organized exchanges.

Groups of Socially Connected Countries. We next explore the effect on trade of two countries sharing connections to a similar set of other countries, over and above the direct effect of the bilateral connections between those two countries. There are several mechanisms through which being in the same social

Table 3: Price Regressions

	Dependent variable: log(Bilateral Trade Unit Price)			
	(1)	(2)	(3)	(4)
log(SCI)	-0.114***	-0.115***	-0.069***	-0.070***
	(0.001)	(0.001)	(0.001)	(0.001)
log(SCI) × Exchange Traded		0.062***		0.060***
		(0.004)		(0.007)
Origin Country × Product FE	Y	Y	Y	Y
Destination Country × Product FE	Y	Y	Y	Y
Other Gravity Controls	Y	Y	Y	Y
Distance Group × Product FE			Y	Y
R^2	0.839	0.839	0.851	0.851
N	3,715,782	3,715,782	3,697,160	3,697,160
N - Explained by FE	169,387	169,387	188,009	188,009

Note: Table shows results from regression 4. The dependent variable is the log per-unit value of exports (exclusive of transport costs and tariffs) of product category k from country i to country j. 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 distance groups (dividing distance into 500 quantiles) interacted with product classifications in columns 3 and 4. Exchange Traded is an indicator that is one if the product is classified as exchange traded (based on the Rauch classification) for a product category. 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 include 111 origin countries, 174 destination countries and 4,830 product categories, with 3,885,169 non-missing observations. Observations that are fully explained by the fixed effects are dropped before the estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

cluster of countries 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 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. If country A and B are both connected to country C, traders in country A can can use existing links to country C in order to establish connections with potential trading partners in country B. This channel is closely related to the existing evidence 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 described in Section 1 to build two measures of whether countries share social connections with a similar set of other countries. Our first measure is simply a dummy variable for whether countries are in the same one of the 30 clusters from Section 1, though our results are robust to variations 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.[15] 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.

[15] 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.

Table 4: Gravity Regressions - The Role of Social Clusters

	Dependent variable: Aggregate Exports						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(SCI)	0.325*** (0.030)	0.313*** (0.030)	0.321*** (0.029)	0.310*** (0.031)			
Same Cluster (30 Gr.)		0.188** (0.084)	-0.472** (0.193)		0.241*** (0.082)	-0.189 (0.235)	
Join Same Cluster - Middle				0.100 (0.082)			0.180*** (0.069)
Join Same Cluster - Early				0.241** (0.119)			0.246*** (0.095)
Same Cluster (30 Gr.) $\times$ log(Cluster Size)			0.326*** (0.109)			0.203* (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 2. The dependent variable is total exports from country i to country j. "Same Cluster (30 Gr.)" is a dummy variable indicating that the two countries are in the same cluster when we create 30 clusters using the hierarchical clustering algorithm described in Sectino 1. "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).

We present the results from this analysis in Table 4. For comparison, in column 1 we present our baseline specification for aggregate trade, corresponding to column 7 of Table 1. 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. 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 on trade of being in the same cluster increases when the cluster includes a larger number of other countries. 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 a non-linear relationship of bilateral social connectedness and distance on trade, columns 5-7 introduce controls for 500 quantiles of both 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.

2.3 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 explanations for the aggregate relationship between social connectedness and trade flows. The first alternative interpretation is reverse causality, whereby trade may cause social connections rather than social connections causing trade. The second alternative interpretation is the existence of an omitted variable that is correlated with both trade and social connectedness. For example, countries that are strongly connected 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 discussed in Section 1 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 to provide evidence against the quantitative 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 it is possible that trade 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 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 directly interact with trading partners in foreign countries. 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.[16] In addition, only a subset of connections of links formed via

[16] 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. 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 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 1, 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.

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.

3 Trade and Subnational 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 this section, we further analyze the mechanism through which social connectedness influences trade patterns. To do so, we exploit the granular nature of our social connectedness measure and study trade and connectedness across subnational European regions. By focusing on Europe, we can zoom in on the patterns of social connections that influence trade in specific products and relate them to the geographic distribution of production throughout the continent. Our findings also allow us to address a number of potential alternative interpretations for the observed relationship between social connectedness on trade.

We conduct two exercises. In Section 3.1, we construct product-specific measures of across-country social connectedess. These measures weight the connectedness of subnational 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 of 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 correlation between social connectedness and trade flows at the country level is driven by either reverse causality or by similar preferences between individuals in more connected countries.

In Section 3.2, we link regional social connectedness to information about rail freight volumes between those regions as a proxy for subnational 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 that the border effect is primarily driven by declines in social connectedness across borders. Finally, we present evidence suggesting that social connectedness and trust act as substitutes in their effect on trade.

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

In Section 2, 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, 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 Facebook friendship links between individuals in regions r_i and r_j, let Pop_{r_i} denote the total (Facebook) 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:

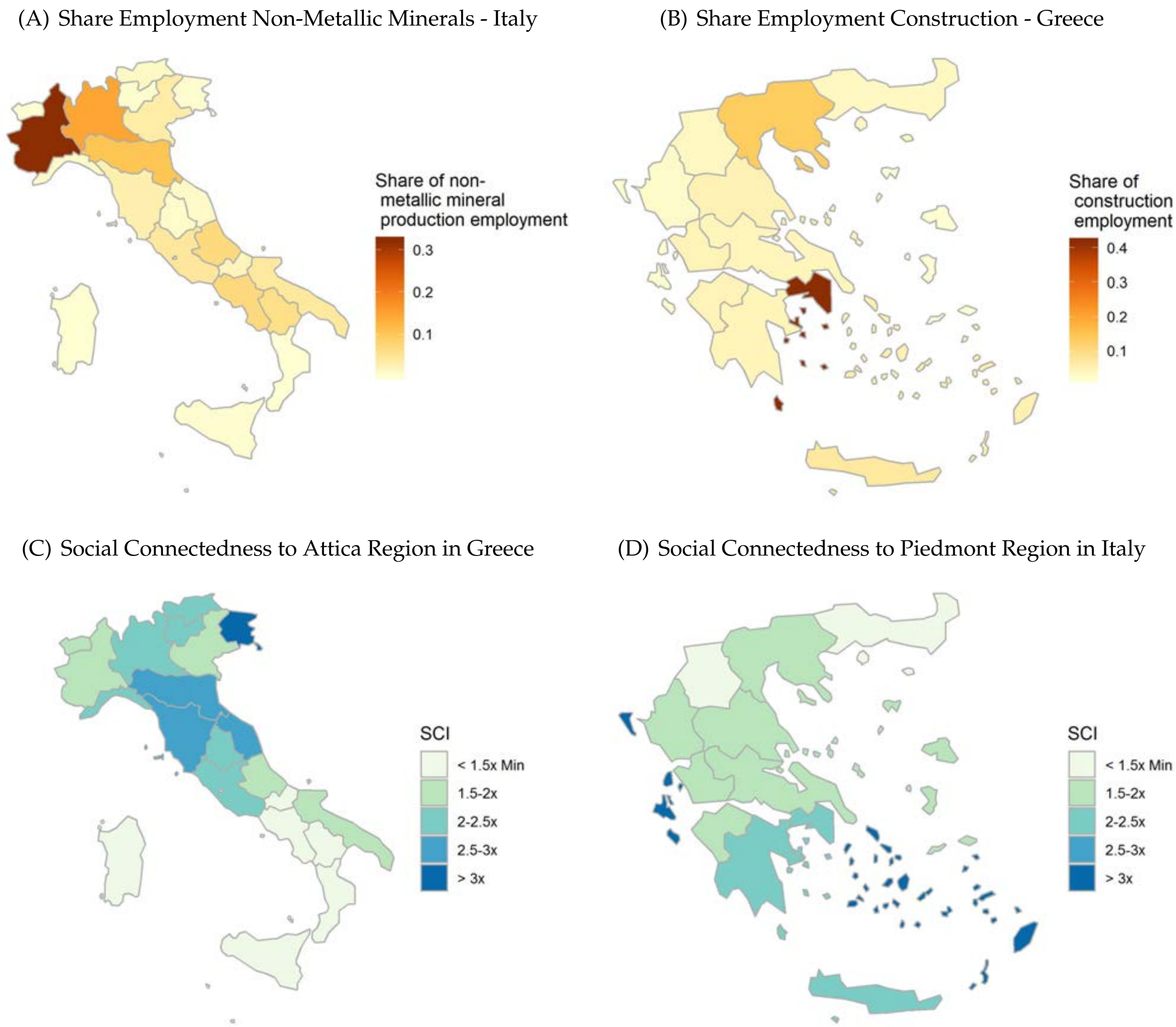
$$
\begin{aligned}
SCI_{i,j} &= \frac{Friendships_{i,j}}{Pop_i \times Pop_j} = \frac{\sum\limits_{r_i \in R(i)} \sum\limits_{r_j \in R(j)} Friendships_{r_i,r_j}}{\left(\sum\limits_{r_i \in R(i)} Pop_{r_i}\right) \times \left(\sum\limits_{r_j \in R(i)} Pop_{r_j}\right)} \\
&= \sum_{r_i \in R(i)} \sum_{r_j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

In other words, in exploring the role of $SCI_{i,j}$ as a determinant of trade between two countries, we implicitly imposed 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 4 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 share is in the northwestern Piedmont region, which includes the city of Torino and a number of major industrial sites. Similarly, Panel B of Figure 4 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, the 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 4 shows that there is substantial variation in which regions in Italy are connected to the Attica region in Greece (Panel C), and which regions 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 in that product. To conduct this exercise, we construct, for each country i × country j × 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

Figure 4: Regional Employment Shares And Social Connectedness



Note: Panel A plots the regional shares 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 heat maps of social connectedness from the Attica Region in Greece to Italian NUTS2 regions, and from the Piedmont Region in Italy to Greek NUTS2 regions.

trade data are at the product level, while the employment and input-output data are 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).[17] 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 input-output-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:

$$SCI_{i,j}^{p} = \sum_{r_i \in R(i)} EmpShare_{p,r_i} \times \left[\sum_{p' \in P} IO_{i,j}^{p,p'} \times \left(\sum_{r_j \in R(j)} EmpShare_{p',r_j} \times SCI_{r_i,r_j} \right) \right]. \qquad (6)$$

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.[18] Similarly, we construct a product-specific measure of the input-output-weighted 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:

$$Distance_{i,j}^{p} = \sum_{r_i \in R(i)} EmpShare_{p,r_i} \times \left[\sum_{p' \in P} IO_{i,j}^{p,p'} \times \left(\sum_{r_j \in R(j)} EmpShare_{p',r_j} \times Distance_{r_i,r_j} \right) \right]. \qquad (7)$$

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 $\times$ 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 than the average region in a country pair. This can be also seen by comparing equations 5 and 6, in which employment shares must be different from population shares to generate any variation in $SCI_{i,j}^{p}$ over $SCI_{i,j}$. 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 between these countries. These differences provide the identifying variation in the following regressions.

Next, we explore how trade in different products correlates with the product-specific measures of

[17] 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.

[18] 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.

social connectedness, $SCI^p_{i,j}$, and geographic distance, $Distance^p_{i,j}$, by estimating the following regression:

$$X_{i,j,p} = \exp[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(Distance_{i,j}) + \beta_3 \log(SCI^p_{i,j}) + \beta_4 \log(Distance^p_{i,j}) + \delta_{i,j,p}] \cdot \epsilon_{i,j,p}. \quad (8)$$

Here, $X_{i,j,p}$ denotes the total value of exports of product p from country i to country j. We also include the logarithm of population-weighted measures of social connectedness and distance as controls; these are the same covariates that we used in Section 2. 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 5 shows results from Regression 8. 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 2. 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 regional variation in social connectedness is explained by a component that is common for all region pairs in a country pair.

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 other observable or unobservable factors that might have been correlated with both social connectedness and trade flows for a given country pair, and which would have thus caused 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 input-output-weighted 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 5 is that the correlation between country-level social connectedness and trade documented in Section 2 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 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 connected-

Table 5: 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 $\times$ Product FE	Y	Y	Y	Y	Y	Y
Destination Country $\times$ Product FE	Y	Y	Y	Y	Y	Y
Undir. Country Pair FE				Y	Y	
log(Distance^p) Group FE					Y	Y
Undir. Country Pair $\times$ 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

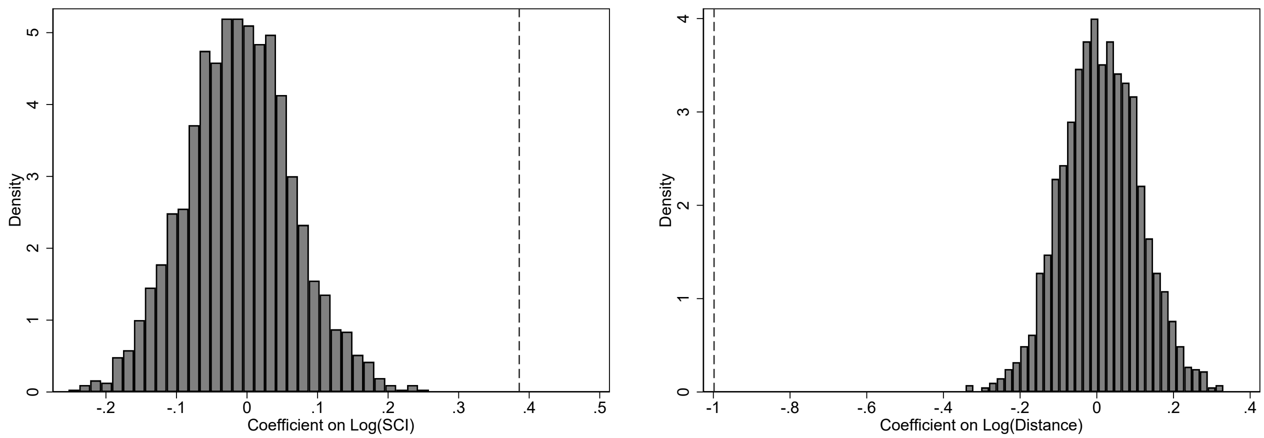
Note: Table shows the results from regression 8. The dependent variable is exports of product category k from country i to country j. The variable $\text{SCI}_{i,j}$ is the population-weighted average of NUTS2 region-level social connectedness. The variable $\text{SCI}^p_{i,j}$ is an employment share weighted measure that uses information on input-output trade, as defined in equation 6. The measures $\text{Distance}_{i,j}$ and $\text{Distance}^p_{i,j}$ are constructed in the same way as the corresponding social connectedness measures. All specifications include exporter $\times$ product and importer $\times$ 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($\text{Distance}^p_{i,j}$) 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 include 28 countries and 20 products leading to 15,120 (= 28 $\times$ 27 $\times$ 20) observations. Observations that are fully explained by the fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

ness 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 2, the population-weighted social connectedness across regions would be the most appropriate measure of the similarity of preferences between the populations of two 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 preferences between more connected countries does not constitute a quantitatively important determinant of trade. An alternative way of exploring the statistical significance of the estimates in column 4 of Table 5 is

Figure 5: Randomization Inference

(A) Coefficient on $\log(SCI_{i,j}^{p})$

(B) Coefficient on $\log(Distance_{i,j}^{p})$



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 5; 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 $\log(SCI_{i,j}^{p})$ and $\log(Distance_{i,j}^{p})$ 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.

the following. For each country $i \times$ country $j \times$ product p triplet, we assign a value of $\log(SCI_{i,j}^{p})$ and $\log(Distance_{i,j}^{p})$ 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 5 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 of Table 5. 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 of Table 5, we replace the control for $\log(Distance_{i,j}^{p})$ with 500 quantiles for distance. The coefficient on $\log(SCI_{i,j}^{p})$ remains unaffected, ruling out concerns that the loading on social connectedness may be picking up non-linearities in the relationship between the input-output-weighted 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 $SCI_{i,j}^{p}$ 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 $SCI_{i,j}^{p}$ 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 large driver of the findings in Section 2.

Ruling out Reverse Causality. Another benefit of exploring the input-output-weighted social connectedness is that it allows us to further address the concern 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 that are 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 $SCI_{i,j}^{p}$:

$$SCI\_Divergence_{i,j}^{p} = \frac{SCI_{i,j}^{p} - SCI_{i,j}}{SD(SCI_{i,j}^{p})}. \tag{9}$$

Figure 6 shows a histogram of $SCI\_Divergence_{i,j}^{p}$ 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.[19]

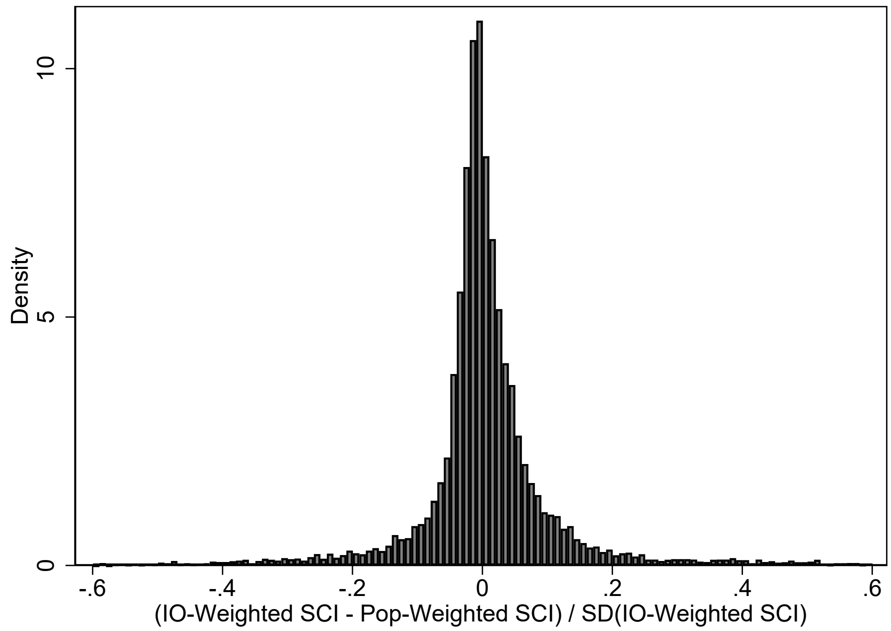
3.2 Subnational Social Connectedness and Rail Freight Flows

The challenge for exploring subnational trade patterns is the absence of trade data at the subnational level. However, within Europe, we observe data on rail freight tonnage shipped between pairs of NUTS2 regions for a number of countries.[20] 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 explore the relationship between rail freight flows and social connectedness across European NUTS2

[19] 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 $SCI\_Divergence_{i,j}^{p}$ towards being larger than zero.

[20] 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.

Figure 6: 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 $SCI^p_{i,j}$, as specified in equation 9.

regions using the following regression:

$$RailFreight_{r_i,r_j} = \exp[\beta_1 \log(SCI_{r_i,r_j}) + \beta_2 \log(Distance_{r_i,r_j}) + \delta_{r_i,r_j}] \cdot \epsilon_{r_i,r_j}. \tag{10}$$

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

Table 6 presents the results from regression 10. 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 2 only on countries included in the rail freight analysis yields elasticities similar to those in the baseline regression in Table 1. Instead, 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 be that the importance of social connectedness varies with the means of transportation.

In column 2 and all following columns of Table 6, we replace our controls for geographic distance with dummy variables for 500 quantiles 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 6 we remove our controls for the measure of social connectedness between re-

Table 6: Subnational 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 10. The dependent variable is the rail freight shipped from NUTS2 region r_i to NUTS2 region r_j. "Same Country" is a dummy variable indicating whether the rail shipment is between 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. All specifications include origin region and destination region fixed effects. Columns 2 to 9 include dummy variables for 500 quantiles of the distance distribution. Column 5 adds country pair fixed effects that do no distinguish the direction of trade. Columns 6 to 9 add origin region $\times$ destination country and destination region $\times$ origin country fixed effects. Standard errors are clustered by NUTS2 origin region and NUTS2 destination region. Observations that are fully explained by the fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

gions, and instead include a "same country" dummy variable. Conditional on the geographic distance between two regions, trade is about seven times larger 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 six 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 regions 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 social connections can affect 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 between 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).[21] The correlation between trust and social connectedness across country pairs is 0.35.

In column 7 of Table 6, 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 separate SCI coefficients for

[21]Guiso et al. (2009) construct data on trust between countries by using a survey 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 United Kingdom.

high-trust and low-trust region pairs. We find 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.

4 Conclusion

In this paper, we use anonymized data from Facebook to construct a comprehensive measure of the social connectedness between countries as well as between European regions. We use this measure to study the relationships between patterns of social connectedness and trade flows.

We hope that our easily accessible measures of social connectedness will also be used to broaden our understanding of the economic effects of social networks across a range of settings beyond international trade. For example, Appendix D provides evidence that international asset holdings also vary with social connectedness. Our social connectedness data may additionally help overcome measurement challenges in other subfields of economics. For example, 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). Our data can help researchers test these and other predictions on the relationship between social structure and socio-economic outcomes.

Beyond economics, researchers in international relations have debated whether the relationships between citizens of countries or the relationships between political leaders are important for the maintenance of peaceful relationships between countries. Our measure of international social connectedness could be used in tests hoping to distinguish between those theories. 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).

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.

Cristian Badarinza, Tarun Ramadorai, and Chihiro Shimizu. Gravity, counterparties, and foreign investment. *C.E.P.R. Discussion Papers*, 2019.

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, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. House price beliefs and mortgage leverage choice. Technical Report 6, 2019a.

Michael Bailey, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. Social connectedness in urban areas. *Revision requested at the Journal of Urban Economics*, 2019b.

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

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

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. *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. *Annual Review of Economics*, 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, 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, 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.

David Hummels and Alexandre Skiba. Shipping the good apples out? an empirical confirmation of the alchian-allen conjecture. *Journal of Political Economy*, 112(6):1384–1402, 2004.

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. *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, 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. *Working Paper*, 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

In Section A.1, we explore a number of case studies of the social connectedness of individual countries. In Section A.2, we formally analyze the role of geography and other factors such as similarity of language, history, religion, and economic development in shaping international social connections.

A.1 Case Studies of International Social Connectedness

Figures A.1 to A.5 present heat maps of the social connections of several countries in addition to the heat map of Portugal's international connections in Panel A of Figure 1. As before, darker colors correspond to closer connections.

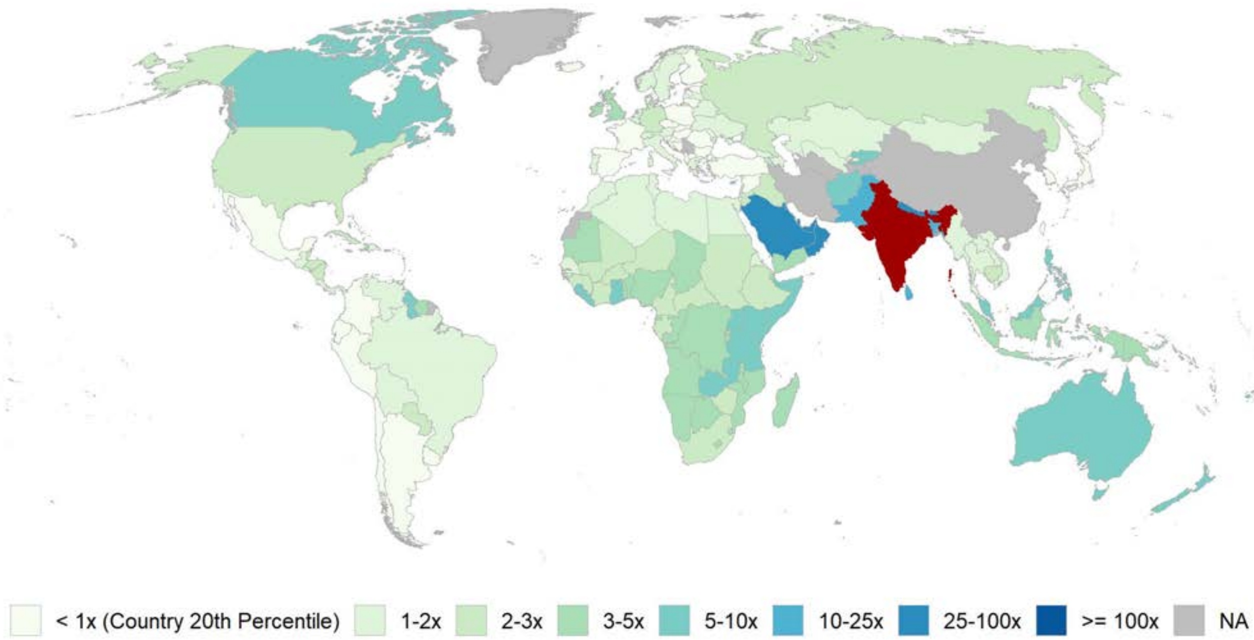
India and Malaysia. Figure A.1 shows the social connectedness of two Asian countries, India (Panel A) and Malaysia (Panel B). Both countries are strongly connected to the countries on the Arab peninsula, likely a result of migrant workers from India and Malaysia moving to work in these countries in recent 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 predominantly 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.

Argentina. Figure A.2 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 weaker, 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.

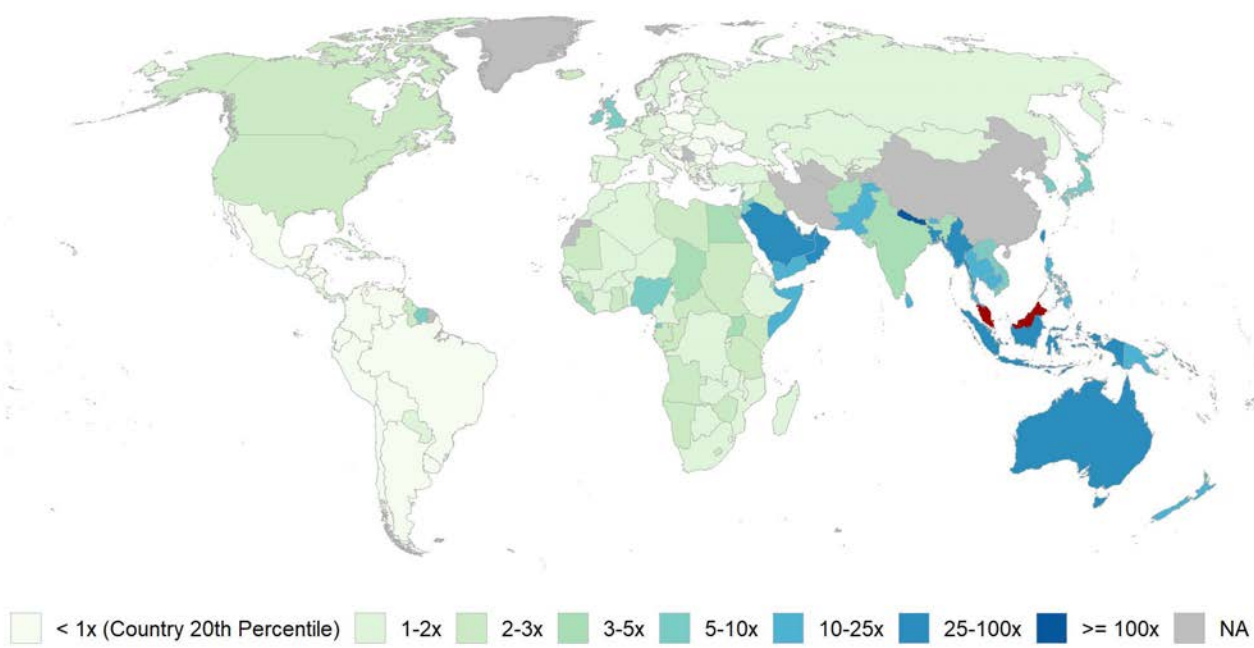
Figure A.1: Social Connectedness of India and Malaysia

(A) Social Connectedness of India

t



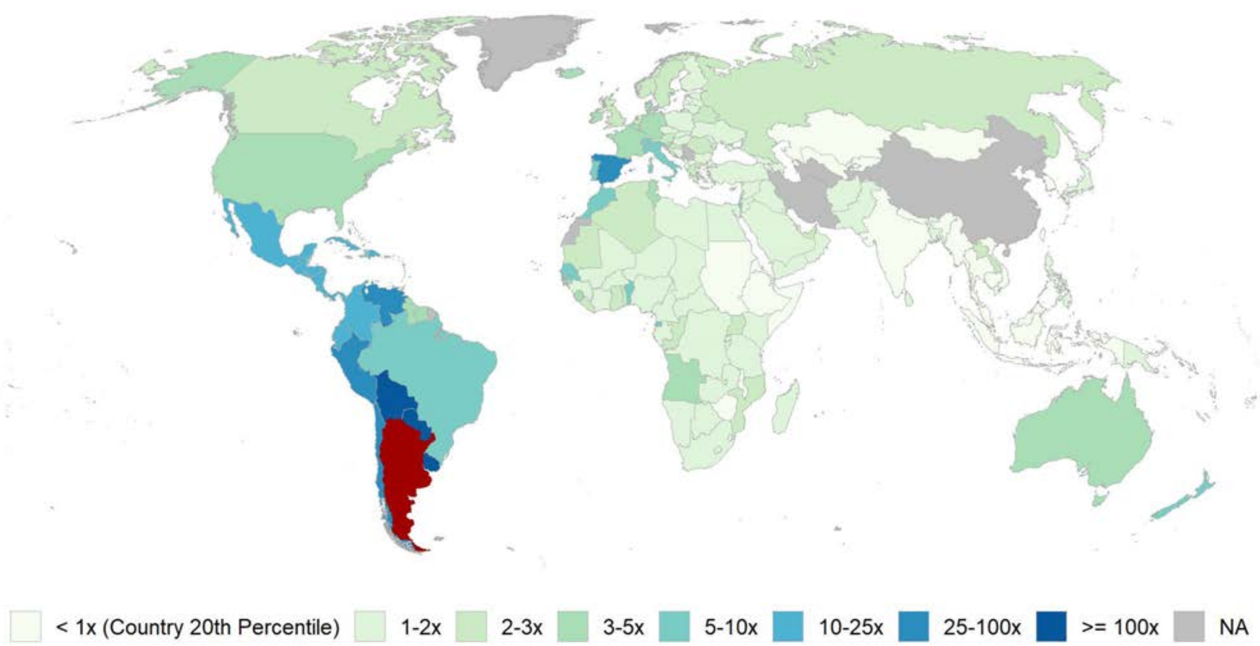
(B) Social Connectedness of Malaysia



Note: Figures show a heat map of the social connectedness of India (Panel A) and Malaysia (Panel B). For each country in our data, the colors highlight connections of the focal country, given in red. The lightest color corresponds to the 20th percentile of the connectedness to the focal country; darker colors correspond to closer connections.

Figure A.2: Social Connectedness of Argentina

(A) Social Connectedness of Argentina



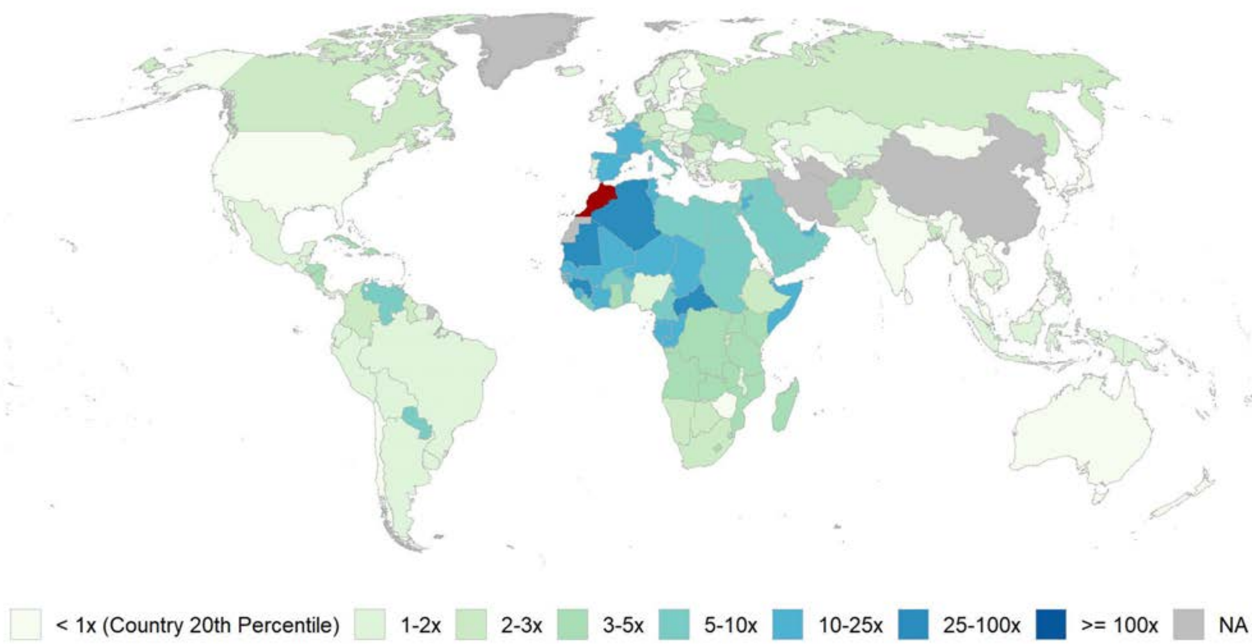
Note: Figure shows a heat map of the social connectedness of Argentina. For each country in the world, the colors highlight connections of the focal country, given in red. The lightest color corresponds to the 20th percentile of the connectedness to the focal country; darker colors correspond to closer connections.

Morocco and Mauritania. Figure A.3 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 Haratin and West African ethnicity. These patterns suggest that ethnic ties are important in shaping friendships across countries.

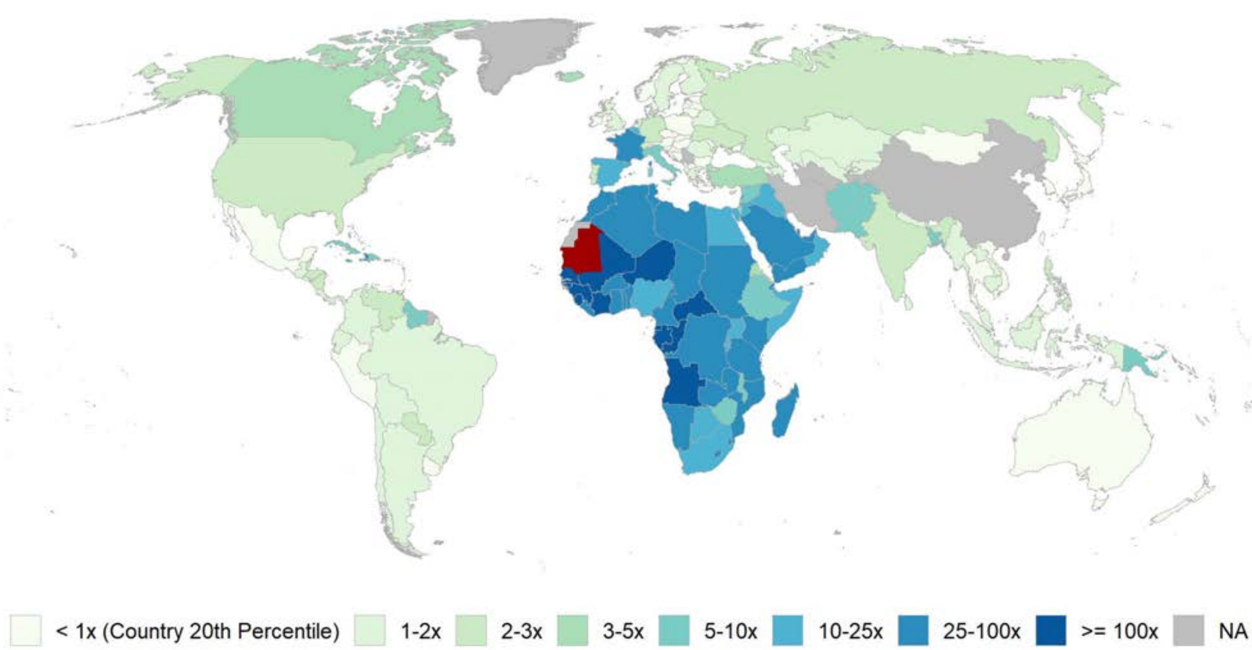
Azerbaijan and Turkey. Figure A.4 shows the social connectedness of Azerbaijan (Panel A) and Turkey (Panel B), two countries in Northwest Asia that share a short border. Both are strongly connected to each other, the nearby Caucasus countries of Armenia and Georgia, and Central European countries which 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.3: Social Connectedness of Morocco and Mauritania

(A) Social Connectedness of Morocco



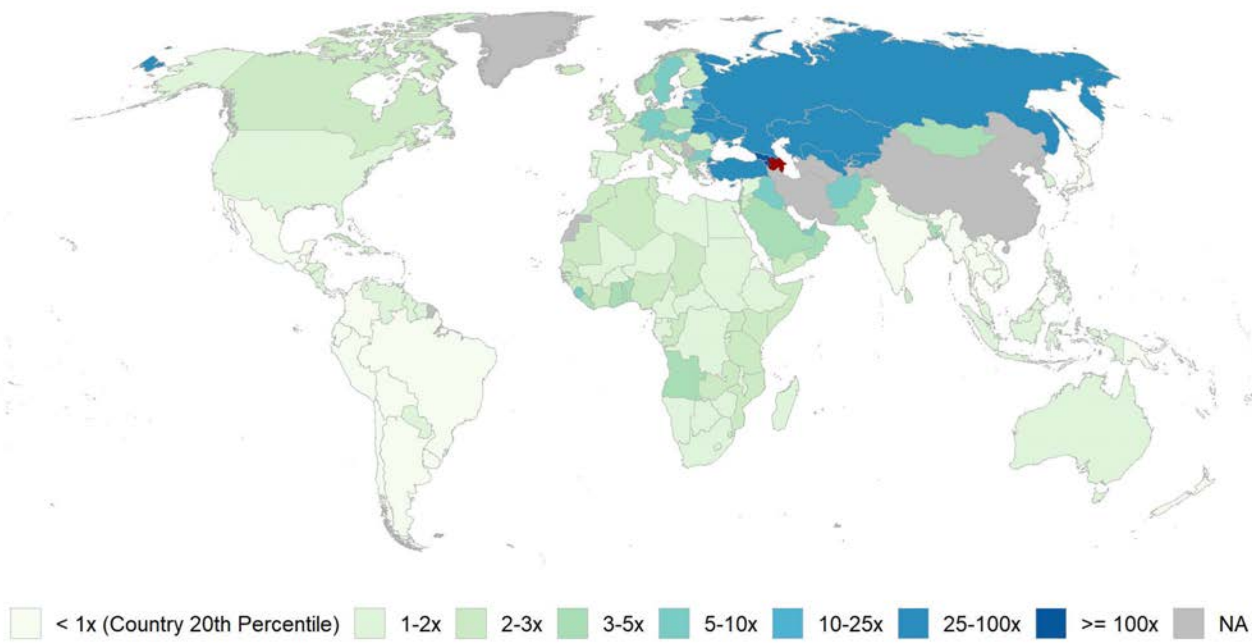
(B) Social Connectedness of Mauritania



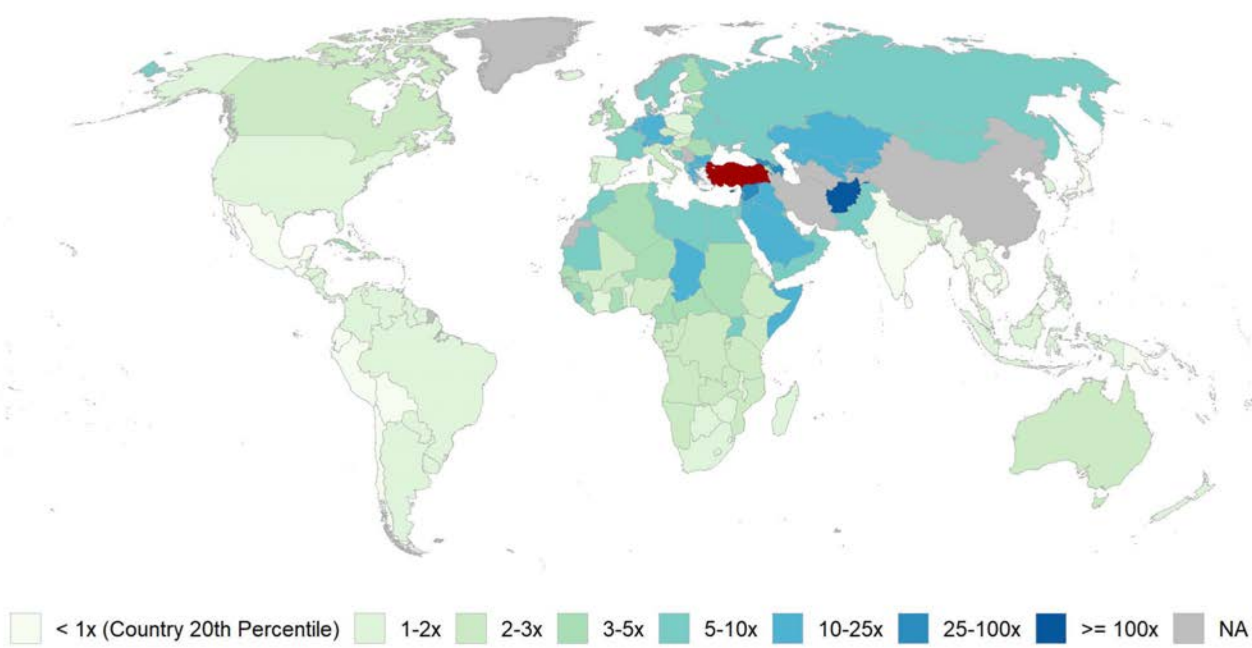
Note: Figures show a heat map of the social connectedness of Morocco (Panel A) and Mauritania (Panel B). For each country in the world, the colors highlight connections of the focal country, given in red. The lightest color corresponds to the 20th percentile of the connectedness of the focal country; darker colors correspond to closer connections.

Figure A.4: Social Connectedness of Azerbaijan and Turkey

(A) Social Connectedness of Azerbaijan



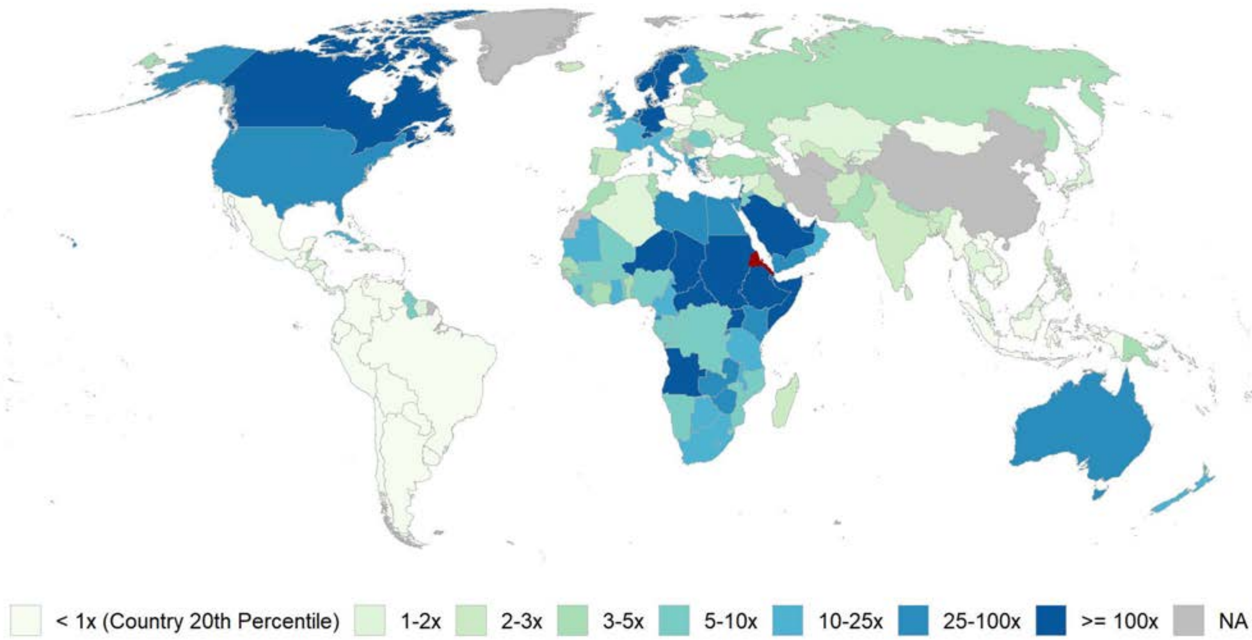
(B) Social Connectedness of Turkey



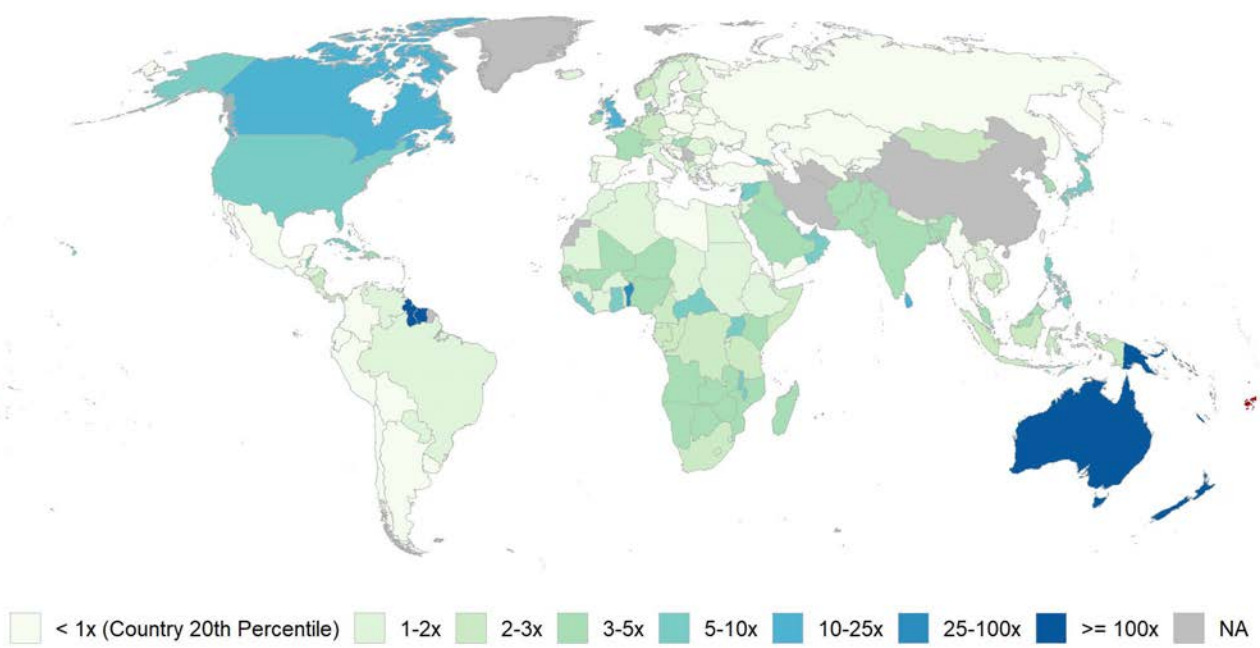
Note: Figures show a heat map of the social connectedness of Azerbaijan (Panel A) and Turkey (Panel B). For each country in the world, the colors highlight connections of the focal country, given in red. The lightest color corresponds to the 20th percentile of the connectedness of the focal country; darker colors correspond to closer connections.

Figure A.5: Social Connectedness of Eritrea and Fiji

(A) Social Connectedness of Eritrea



(B) Social Connectedness of Fiji



Note: Figures show a heat map of the social connectedness of Eritrea (Panel A) and Fiji (Panel B). For each country in the world, the colors highlight connections of the focal country, given in red. The lightest color corresponds to the 20th percentile of the connectedness to the focal country; darker colors correspond to closer connections.

Eritrea. Panel A of Figure A.5 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.

Fiji. Panel B of Figure A.5 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 previously. 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.2 Regression Analysis

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

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

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.

A.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 A.1. The binscatter plots in Figure A.6 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 A.2 explores the role of distance using regression A.1. Column 1 includes only 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 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.

Table A.1: Summary Statistics and Data Sources

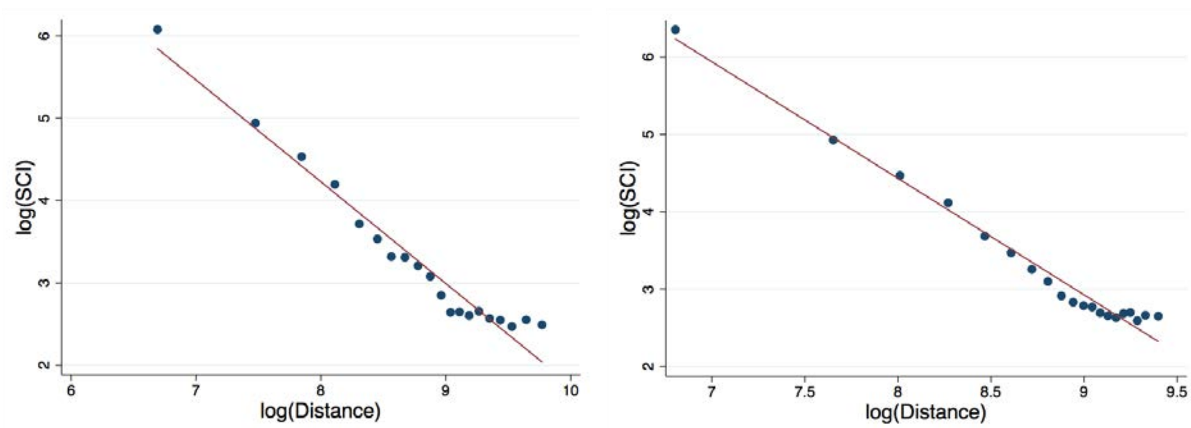
	Mean	P10	P25	P50	P75	P90	N	Source
log(SCI)	3.30	1.61	2.08	2.94	4.16	5.54	32,220	Facebook
log(Distance)	8.75	7.68	8.39	8.92	9.30	9.58	32,220	CEPII
log(1+Migrant Population)	2.49	0.00	0.00	0.00	4.93	8.13	31,862	United Nations
Common Colonizer	0.11	0.00	0.00	0.00	0.00	1.00	32,220	CEPII
Colonial Relationship	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 A. 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 A.6: Social Connectedness vs. Geographic Distance

(A) No Country Fixed Effects

(B) Country Fixed Effects



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

Table A.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	$<$ 6000km	$>$ 6000km	All	All	All
R^2	0.291	0.652	0.672	0.647	0.551	0.657	0.678	0.690
N	32,220	32,220	32,220	12,478	19,742	32,220	32,220	32,220

Note: Table shows results of regression A.1. 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).

Column 2 adds $\log(Distance_{i,j})$ as a control to the country fixed effects; this specification corresponds to the binscatter plots in Panel B of Figure A.6. 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 A.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 A.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 about 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 A.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 all else equal. Being on the same continent (but not the same subcontinent) is associated with an 62% 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.

A.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 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 similarity, common language, religion, and similarity in GDP. Table A.1 shows summary statistics on these variables, as well as the data sources; not all variables are available for all country pairs. Table A.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 A.4 contains the results from Regression A.1 when controlling for these variables. We first explore the relationships between the gravity variables and social connectedness separately for each variable, before analyzing them jointly in a multivariate analysis.

Migration. It is likely that migrants retain many friends and family ties to their countries 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 Na-

Table A.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 A. 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.

tions.[1] The estimates in column 1 of Table A.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 fixed effects increases the R^2 of the regression by 6.7 percentage points.

Common Colonial History. The case studies in Section A.1 suggested that countries with a common colonial history 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 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

[1]Most of the data are based on population 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 and the country-level fixed effects.

Table A.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
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 A.1. The dependent variable is the logarithm of the social connectedness for 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. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

distance following Spolaore and Wacziarg (2018).[2] Table A.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 A.3. Column 4 of Table A.4 confirms the strong relationship. 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. In addition, people of the same religion may be more likely to meet each other across countries, for instance when traveling for pilgrimage. To explore the effect of religion on social connectedness, we measure religious distance between countries following Spolaore and Wacziarg (2016).[3] The estimates in column 5 of Table A.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 A.4 shows that differences in GDP correlate with social connectedness. A ten thousand dollar higher absolute difference in GDP per capita 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 A.4 have shown the relationship between each regressor and social connectedness separately (in addition to controls for distance and country fixed effects). However, many of these variables are correlated with each other, as shown in Table A.3, and they often capture related aspects of across-country similarity. Column 7 explores how much of the variation

[2] This 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.

[3] 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 follow 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.

in social connectedness can be jointly explained by these variables. 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%.

As shown in Table A.3, a number of the control variables in Table A.4 are highly correlated. For example, the correlation between genetic distance and our measure of migration is -0.44. Therefore, another interesting metric is how much explanatory power each variable contributes when controlling for all other variables. For each variable, we run two regressions with the specification equal to column 7 of Table A.4; one where we exclude the variable of interest and one where we include it. We then compare the R^2 of the two regressions, giving us an estimate of the incremental explanatory power of the variable beyond all other explanatory variables. The results of this exercise are reported in Table A.5. We find that migration has the largest incremental R^2 with 3.74% followed by distance with 1.97%. Genetic distance (1.93%) and the common official language dummy (0.80%) also add sizable explanatory power. Colonial heritage, religious distance, a common border, and the same subcontinent dummies have very little explanatory power once we control for other factors.

Table A.5: Incremental Explanatory Power of Regressors for SCI

Incremental R^2	Dependent variable: log(SCI)
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: 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 A.4. To get the incremental R^2, we compute the difference between the R^2 of a regression when 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.

A.3 Groups of Socially Connected Countries

In Section 1, we used a hierarchical clustering algorithm to construct 30 clusters of countries with high within-cluster social connectedness. Table A.6 lists the countries in the 30 clusters described.

Table A.6: 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

In this section, we present material complementary to our analysis of country-level trade in Section 2. Section B.1 explains the construction of the data, Section B.2 shows robustness to using OLS estimation instead of PPML, Section B.3 estimates time-varying elasticities of trade to social connections, and Section B.4 runs a horse race between different gravity variables in explaining international trade.

B.1 Construction of Trade and SCI Data

For the analyses in Section 2, we merge data on 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. These countries are Botswana, Lesotho, Luxembourg, Namibia, Sudan, and Swaziland. In the original data, products are classified according to the 6-digit HS96 classification into 4,914 product categories.

In order to construct total bilateral trade, used in Section 2.1, we aggregate trade across all products for each exporter-importer pair. We replace missing trade values with zero trade values. Our final data that contain both social connectedness and aggregate bilateral trade include 174 countries and 30,102 observations.

In Section 2.2, we use data on product-level trade instead of total trade. For computational reasons, we aggregate trade flows up to the first two digits of the HS96 product code for each exporter-importer pair.[4] This procedure results in 96 product categories. We replace missing trade values with zeros. 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 lack four countries: Cook Islands, New Caledonia, Palestine, and French Polynesia. The data that contain product-level trade, international SCI and the rule of law measure include 170 countries and 2,758,080 (=170 x 169 x 96) observations.

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

Table 2 in Section 2.1 showed results from regressing aggregate bilateral trade on social connectedness and other gravity variables. The regression was 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 relationships using OLS. We focus on the intensive margin of trade in order to avoid problems with zero-trade observations. We estimate the following regression:

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

The dependent variable $log(X_{i,j})$ denotes the logarithm of the total value of exports from country i to country j.

The results are reported in Table B.1. 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

[4]The first two 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".

fixed effects, nearly as high as the within-R^2 of 28.9% for distance (see column 3). Similarly to Table 2, 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). Overall, these results show that our baseline results are not an artefact of the specific estimation method.

Table B.1: 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	Y	Y	Y	Y	Y	Y	Y	Y
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

Note: Table shows results from regression B.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).

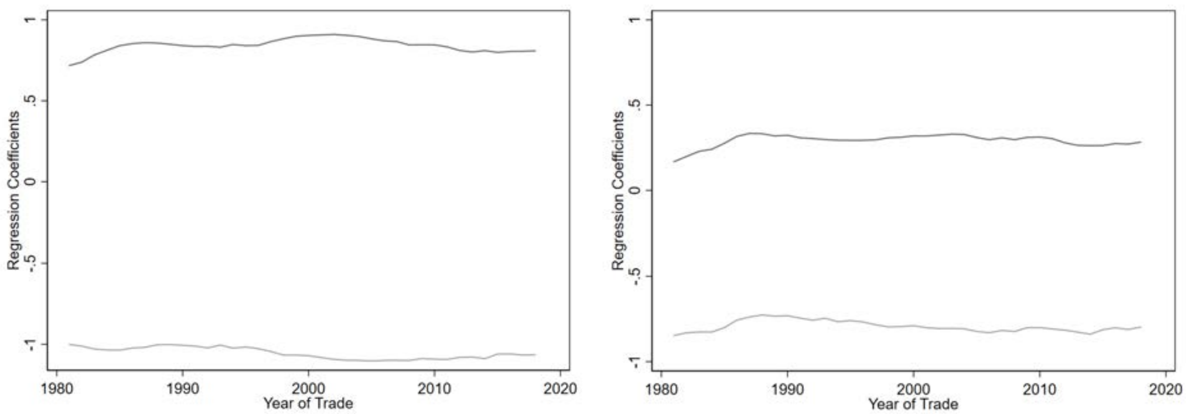
B.3 Time-variation in Trade Elasticities

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 contain 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 aggregate bilateral trade on our distance measure and our social connectedness measure from March 2019. This allows to estimate year-specific regression coefficients that are shown in Figure B.1. These findings highlight that the elasticity of trade to both distance and social connectedness has been highly constant over the past fourty years.

Figure B.1: Time-variation in Elasticity of Trade to Social Connectedness

(A) Univariate Regressions

(B) Multivariate Regression



Note: Figures show year-specific regression coefficients from regressing trade on $\log(SCI)$ and $\log(Distance)$ as specified in Equation 2. Panel A shows the regression coefficients obtained in univariate regressions, Panel B shows the regression coefficients obtained in multivariate regressions. 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 2 of the main paper, we regressed aggregate bilateral trade on the logarithm of social connectedness and 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 separately. In this section, we conduct this analysis using both, PPML regressions and OLS regressions on the intensive margin of trade. We report the results of this "horse race of predictors" in Table B.2. Columns 1 to 6 are estimated using PPML, while columns 7 to 12 are estimated using OLS. 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. A similar results emerges from the OLS regression: Including distance and social connectedness raises the R^2 significantly compared to all other gravity variables.

Table B.2: Gravity Regressions - Horse Race

	Dependent variable: Aggregate Bilateral Trade											
	PPML						OLS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log(Distance)	-0.996***						-1.706***					
	(0.06)						(0.063)					
log(SCI)		0.683***						0.813***				
		(0.04)						(0.03)				
Common Border			1.900***						4.205***			
			(0.204)						(0.230)			
Common Official Language				0.936***						1.809***		
				(0.146)						(0.163)		
Common Colonizer					1.439***						1.564***	
					(0.156)						(0.199)	
Colonial Relationship						0.607*						1.544***
						(0.325)						(0.404)
Orig. and Dest. Country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R^2	0.929	0.919	0.887	0.845	0.839	0.835	0.772	0.769	0.703	0.699	0.688	0.680
N	30,102	30,102	30,102	30,102	30,102	30,102	20,054	20,054	20,054	20,054	20,054	20,054

Note: Table shows results from regression 2 (PPML) and regression B.1 (OLS). We estimate the regression using PPML in columns 1–6 and using OLS in columns 7–12. The dependent variable is total exports from country i to country j for the PPML regressions, and the logarithm thereof for the OLS regressions. 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. The OLS regressions use only non-zero trade observations (intensive margin of trade). Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

C Trade and Subnational Social Connectedness in Europe

In this section, we explain the construction of the data used in Section 3 when analyzing trade and subnational social connectedness across European regions.

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

Our analyses in Section 3.1 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.[5] 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.[6] In each region, we use the most recent year in which the data were available starting with 2017. 64% of the data come from 2017, 26% from 2016, 1.6% from 2015, 1.3% from 2014, and 3.22% from between 2013 and 2008. 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.[7] 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 these 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 (HS96) 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 C.1.

C.2 Construction of Rail Freight Data

Our analyses in Section 3.2 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. 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 prepare them for our analyses as described below. This process was informed by the "Reference Manual on Rail Transport Statistics" and our correspondence with the Eurostat data providers.

[5] 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.

[6] 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.

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

Table C.1: 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 3.1. Industry descriptions come from the NACE Rev. 2 European statistical classifications of economic activities.

We first restrict the data to observations that are at the NUTS2 region level, including country-level data for countries which consist of only a single NUTS2 region. We exclude all pairs that include a region with the unknown indicator "XX" or the extra-regio territory indicator "ZZ".[8] 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.[9] 2) There are a number of hypothetical region pairs missing, even between countries that did report data elsewhere. 3) For some international region pairs, there are data reported from both countries on the same train flows, and the reported 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 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 C.2. 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

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

[9]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.

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 region 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 region pair, there still remain 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 region pair (challenge 3). In these instances, we take the average of the two reports.[10] 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.[11] 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 (i.e. we assume that tons of goods transported in each of these regions is proportional to the 2015 population).

Table C.2: Rail Freight Data Availability By Reporting Country

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

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

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

D International Asset Holdings and Social Connectedness

In this section, we examine the relationship between international asset holdings and social connectedness. We obtain asset holdings from the Coordinated Portfolio Investment Survey (CPIS) released by the IMF. The CPIS data contain information on cross-border holdings of portfolio securities (equities and debt securities). The value of asset holdings is reported in USD. We focus on data reported for the end of the year 2018 and replace missing values with the corresponding value at the end of year 2017 if available. In the original CPIS data, there are 87 countries that report their foreign asset holdings. From this set of countries we drop 8 countries or regions that are not in our trade data set; namely Aruba, Netherlands Antilles, Bermuda, China, Guernsey, Gibraltar, Jersey, Isle of Man. Furthermore, we drop 6 countries – Albania, Bangladesh, Bolivia, Liberia, Palau, Singapore – from the list of asset holding countries, because these countries report very few asset holdings.[12] We follow the literature and also exclude 11 financial offshore centers: the Bahamas, Bahrain, Cayman Islands, Cyprus, Lebanon, Luxembourg, Macao, Malta, Mauritius, Panama, Vanuatu. As a last step, we replace missing values with zeros. Our final data set contains 62 countries that report their assets in 161 countries. The regression specification we estimate is similar to the one for aggregate trade:

$$A_{i,j} = \exp\left[\beta_1 \log(SCI_{i,j}) + \beta_2 \log(Distance_{i,j}) + \beta_3 G_{i,j} + \delta_i + \delta_j\right] \cdot \epsilon_{i,j}, \tag{D.1}$$

where $A_{i,j}$ denotes the asset holdings of country i in country j. $G_{i,j}$ is a vector of controls, δ_i is a reporting country fixed effect, and δ_j is a fixed effect for the country of the underlying assets.

The results from estimating regression D.1 are presented in Table D.1. Column 1 shows that the country fixed effects alone explain 95% of the variation in bilateral portfolio holdings. In column 2, we introduce the social connectedness between each country pair as an additional regressor. The elasticity of asset holdings with respect to social connectedness is 0.60, which is close the elasticity we estimated for bilateral trade (0.68). This suggest that a 1% increase in the social connectedness is associated with a 0.60% increase in cross-border portfolio holdings. Column 3 excludes social connectedness, but controls for the geographical distance between countries. The explained variation is slightly below the variation that is explained using only social connectedness. The same is true when we use other gravity variables in column 4. When we simultaneously control for geographical distance and social connectedness in column 5, the elasticity of asset holdings with respect to social connectedness drops from 0.60 (column 2) to 0.37 (column 5). Similarly, the elasticity with respect to distance drops from -0.57 to -0.30. In column 7 we include the full set of control variables. The elasticity with respect to social connectedness remains fairly stable and is highly economically significant. The estimates suggest that a doubling of social connectedness between two countries is associated with a 27% increase in bilateral asset holdings controlling for a host of control variables.

[12] We drop these countries, because we do not want to make the assumption that all unreported asset holdings of these countries are zero.

Table D.1: Gravity Regressions - International Asset Holdings

	Dependent variable: Bilateral Asset Holdings						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(SCI)		0.604***			0.366***		0.273***
		(0.078)			(0.082)		(0.071)
log(Distance)			-0.566***		-0.300***	-0.433***	-0.266***
			(0.069)		(0.087)	(0.079)	(0.079)
Common border				1.005***		0.336***	0.266***
				(0.102)		(0.083)	(0.067)
Common official language				0.212		0.308**	0.171
				(0.159)		(0.157)	(0.154)
Common colonizer				1.503***		0.947**	0.857***
				(0.477)		(0.391)	(0.326)
In colonial relationship				0.168		-0.139	-0.205
				(0.198)		(0.487)	(0.439)
Rep. and Asset Country FE	Y	Y	Y	Y	Y	Y	Y
R^2	0.950	0.968	0.967	0.965	0.970	0.970	0.971
N	9,920	9,920	9,920	9,920	9,920	9,920	9,920
N - Explained by FE	404	404	404	404	404	404	404

Note: Table shows results from regression D.1. The dependent variable is the value of assets of country j that country i holds. 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 reporting country and the asset country. Standard errors are clustered by the reporting country and the asset country. Our final data contain 62 countries that report their assets in 161 countries, which yields 9,920 observations. Observations that are fully explained by the fixed effects are dropped before the PPML estimation. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).