

# The Determinants and Effects of Social Connectedness in Europe

Michael Bailey\*   Theresa Kuchler<sup>†</sup>   Dominic Russel<sup>‡</sup>   Bogdan State<sup>§</sup>   Johannes Stroebe<sup>¶</sup>

May 12, 2020

## Abstract

We use aggregated data from Facebook to study the structure of social networks across European regions. Social connectedness declines strongly in geographic distance and at country borders. Historical borders and unions — such as the Austro-Hungarian Empire, Czechoslovakia, and East/West Germany — shape present-day social connectedness over and above today's political boundaries. All else equal, social connectedness is stronger between regions with residents of similar ages and education levels, as well as between those that share a language and religion. In contrast, region-pairs with *dissimilar* incomes tend to be more connected, likely due to increased migration from poorer to richer regions. We find more socially connected region-pairs to have more passenger train trips between them, even after controlling for distance and travel time. We also find that regions with a higher share of connections to other countries have higher rates of trust in the E.U. and lower rates of voting for anti-E.U. political parties.

**Keywords:** Social Connectedness, Europe, Homophily, Border Effects, Migration

**JEL Codes:** D72, J61, O52, R23, Z13

---

\*Facebook. Email: [mcbailey@fb.com](mailto:mcbailey@fb.com)

<sup>†</sup>New York University, Stern School of Business. Email: [tkuchler@stern.nyu.edu](mailto:tkuchler@stern.nyu.edu)

<sup>‡</sup>New York University, Stern School of Business. Email: [drussel@stern.nyu.edu](mailto:drussel@stern.nyu.edu)

<sup>§</sup>Facebook. Email: [bogdanstate@fb.com](mailto:bogdanstate@fb.com)

<sup>¶</sup>New York University, Stern School of Business, NBER, and CEPR. Email: [johannes.stroebe@nyu.edu](mailto:johannes.stroebe@nyu.edu)

Social networks shape many aspects of global society including patterns of migration and travel, social mobility, and political preferences. In turn, social networks reflect both past and present political borders and migration patterns, as well as geographic proximity, culture, and other factors. While understanding the determinants and effects of these networks across regions and countries can be informative for a wide range of questions in the social sciences, researchers have traditionally been limited by the availability of large-scale representative data on regional social connections.

In this paper, we investigate the spatial structure of social networks in Europe. We measure social networks using aggregated data from Facebook, a global online social network.<sup>1</sup> We construct a measure of social connectedness across NUTS2 regions — regions with between 800,000 and 3 million inhabitants — which captures the probability that Facebook users located in these regions are Facebook friends with each other. Europe consists of a number of proximate nations, has a relatively high population density, and includes a diversity of areas with distinct cultural and linguistic identities. Each of these factors differentiates Europe from the U.S., which has been the focus of prior research on social connectedness. This paper documents the important role that these and other factors play in shaping connections, furthering our understanding of the determinants and effects of social networks.

We begin by discussing a number of case studies that show the relationship of European social connections with patterns of migration, political borders (past and present), geographic distance, language, and other demographics. We then explore the association between social connectedness and these factors more formally. We find that social connectedness strongly declines in geographic distance: a 10% increase in distance is associated with a 13% decline in social connectedness. Social connectedness also drops off sharply at country borders. Controlling for geographic distance, 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. Furthermore, using a number of major 20th century European border changes, we find that this relationship between political borders and connectedness can persist decades after boundaries change. For example, we find higher social connectedness across regions that were originally part of the Austro-Hungarian empire, even after controlling for distance, current country borders, and a number of other relevant factors.

In addition to distance and political borders, we find that regions more similar along demographic measures such as language, religion, education, and age are more socially connected. In particular, social connectedness between two regions with the same most common language is about 4.5 times larger than for two regions without a common language, again controlling for same and border country effects, distance, and other factors. Somewhat surprisingly, we see that pairs of regions with *dissimilar* incomes are more connected. This finding may be explained by patterns of migration from regions with low incomes to regions with high income. This finding in Europe contrasts with prior research that finds a positive relationship between connectedness and income similarity across U.S. counties (Bailey et al., 2018b) and New York zip codes (Bailey et al., 2019a).

We next look at the relationship between social connectedness and travel. We find that a 10% increase in social connectedness between two regions is associated with a 12% to 17% increase in the

---

<sup>1</sup>The European social connectedness data that we compile and use in this project is accessible to researchers and policy makers by emailing a 1-paragraph proposal to [sci\\_data@fb.com](mailto:sci_data@fb.com). See Bailey et al. (2018b) for information on county-level U.S. social network data and Bailey et al. (2019a) for zip code-level data in the New York Combined Statistical area.

number of passengers that travel between the regions by train. This result persists even after controlling for the geographic distance and travel time, by train and car, between the central points of the regions. We highlight that this result provides empirical evidence to support a number of theoretical models suggesting social networks play an important role in individuals' travel decisions (see e.g., Axhausen, 2008; Carrasco and Miller, 2009; Páez and Scott, 2007). It also provides strong evidence that the patterns of social connectedness correspond to real-world social connections.

In the final part of the paper, we study variation in the degree of connectedness of European regions to other countries. We find that at the 10th percentile of the distribution, less than 4.1% of connections are to individuals living in different countries, compared to over 19.7% of connections at the 90th percentile. We also study the relationship between this variation and the share of a region's residents that hold Eurosceptic beliefs or that vote for Eurosceptic political parties. According to both measures, we find that Euroscepticism decreases with the share of a region's connections that are international. Specifically, a 1 percentage point increase in the share of a region's connections that are to individuals outside of their home country is associated with a 0.5 percentage point increase in the share of residents who trust the E.U. and a 0.76 percentage point decrease in the share that voted for an anti-E.U. political party. These results persist, but become weaker (0.25 and -0.54 percentage points, respectively), after adding controls for the share of residents living in the region who are born in other European countries as well as the regional average income and unemployment rate, and the shares of employment in manufacturing, construction, and professional sectors.

Throughout this paper, our focus is on documenting and describing salient patterns of social connectedness throughout Europe, not to make any conclusive causal inferences on the determinants and effects of the observed social structures. We hope that the patterns presented in this paper can help advance our understanding of social connectedness in Europe and highlight the significant opportunities for future research on social networks across regions and countries.

## 1 Data

Our measures of social connectedness across locations comes from administrative data from Facebook, a global online social networking service. Facebook was created in 2004 and, by the fourth quarter of 2019, had about 2.5 billion monthly active users globally, including 394 million in Europe. One independent resource estimates 80% of European social media site visits from September 2018 to September 2019 were to Facebook (StatCounter, 2019). A separate study found that the number of active accounts on the most used social network in each country, as a share of population, was 66% in Northern Europe, 56% in Southern Europe, 54% in Western Europe, and 45% in Eastern Europe (We Are Social and Hootsuite). Another 2018 survey found that the share of adults who used any social networking site in 10 European countries was between 40% and 67% (Poushter et al., 2018).

Establishing a Facebook friendship link requires the consent of both individuals, and the total number of friends for a person is limited to 5,000. As a result, Facebook connections are primarily between real-world acquaintances. Indeed, one independent survey of American Facebook users revealed that only 39% of users reported being Facebook friends with someone that they have never met in person (Duggan et al., 2015). In contrast, Facebook users reported that they were generally Facebook friends

with individuals for which they have real-world connections: 93% said they were Facebook friends with family other than parents or children, 91% said they were connected with current friends, 87% said they were connected to past friends (such as former classmates), and 58% said they were connected to work colleagues (Duggan et al., 2015). As a result, networks formed on Facebook 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.

We observed a snapshot of all active Facebook users from July 2019. We focus on those users who reside in one of 37 European countries and who had interacted with Facebook over the 30 days prior to the date of the snapshot. The 37 countries are the members of the European Union and European Free Trade Association, as well as European Union candidate countries as of 2016; these countries were selected because they have standardized administrative boundaries at the NUTS2 (Nomenclature of Territorial Units for Statistics level 2) level.<sup>2</sup> NUTS2 regions contain between 800,000 and 3 million people, and are generally based on existing sub-national administrative borders. For example, NUTS2 corresponds to 21 “regions” in Italy, 12 “provinces” in the Netherlands, and a single unit for all of Latvia.

To measure social connections between NUTS2 regions, we construct our measure of Social Connectedness,  $SocialConnectedness_{ij}$  as follows (see Bailey et al., 2018b, for the first use of this measure of social connectedness in the United States):

$$SocialConnectedness_{ij} = \frac{FB\_Connections_{ij}}{FB\_Users_i * FB\_Users_j} \quad (1)$$

Here,  $FB\_Connections_{ij}$  is the total number of connections between individuals living in NUTS2 region  $i$  and individuals living in NUTS2 region  $j$ .  $FB\_Users_i$  and  $FB\_Users_j$  are the number of eligible Facebook users in each region. Dividing by the product of regional Facebook users allows us to take into account the fact that we will see more friendship links between regions with more Facebook users. As a result, we can say that if  $SocialConnectedness_{ij}$  is twice as large, a Facebook user in region  $i$  is about twice as likely to be connected with a given Facebook user in region  $j$ .<sup>3</sup> In prior work this measure has been shown to be useful to describe real world networks. For example, social connectedness as measured through Facebook friendship links is strongly related to patterns of COVID-19 spread (Kuchler et al., 2020b), international trade (Bailey et al., 2020), and investment decisions (Kuchler et al., 2020a).<sup>4</sup>

As one way of better understanding the connections underlying this measure, we compare it to similar measures constructed using restricted sets of friendships. Table 1 presents the cross-correlation of measures limited to connections made during certain periods of time (e.g., recent friendships vs old friendships) and to friendships between individuals with certain shared characteristics (e.g., ages less than 5 years apart). Each of these measures is highly correlated with the others. This provides more

<sup>2</sup>Specifically the list of countries is: Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Germany, Greece, Finland, France, Hungary, Iceland, Ireland, Italy, Latvia, Lichtenstein, Lithuania, Luxembourg, Malta, Montenegro, the Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Spain, Slovakia, Slovenia, Sweden, Switzerland, Turkey, and the United Kingdom.

<sup>3</sup>Put another way, the measure represents the relative probability of a Facebook friendship link between a user in region  $i$  and a user in region  $j$ .

<sup>4</sup>See Bailey et al. (2017, 2018a,b, 2019a,b) for additional discussion of the evidence that friendships observed on Facebook serve as a good proxy for real-world social connections.

evidence that Facebook connections resemble full real-world networks and not, for example, primarily recently formed online-only connections.

**Table 1:** Correlation of Social Connectedness Constructed from Select Friendship Pairs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) All	1.000							
(2) Added <1 Year Ago	0.946	1.000						
(3) Added 1-5 Years Ago	0.991	0.967	1.000					
(4) Added <5 Years Ago	0.985	0.982	0.998	1.000				
(5) Added >5 Years Ago	0.970	0.848	0.932	0.915	1.000			
(6) Both Female	0.982	0.910	0.960	0.953	0.975	1.000		
(7) Both Male	0.986	0.940	0.982	0.977	0.949	0.955	1.000	
(8) Ages Within 5 Years	0.997	0.942	0.982	0.977	0.974	0.983	0.979	1.000

*Note:* Table presents correlations between social connectedness constructed in Equation 1 and similar measures constructed from restricted sets of connections. Rows and Columns 2-5 limit to connections made less than a year ago, between 1 and 5 years ago, less than five years ago, and more than 5 years ago. Rows and Columns 5-8 limit to connections between females, males, and individuals with ages 5 or fewer years apart.

## 2 Exploration of European Social Connectedness

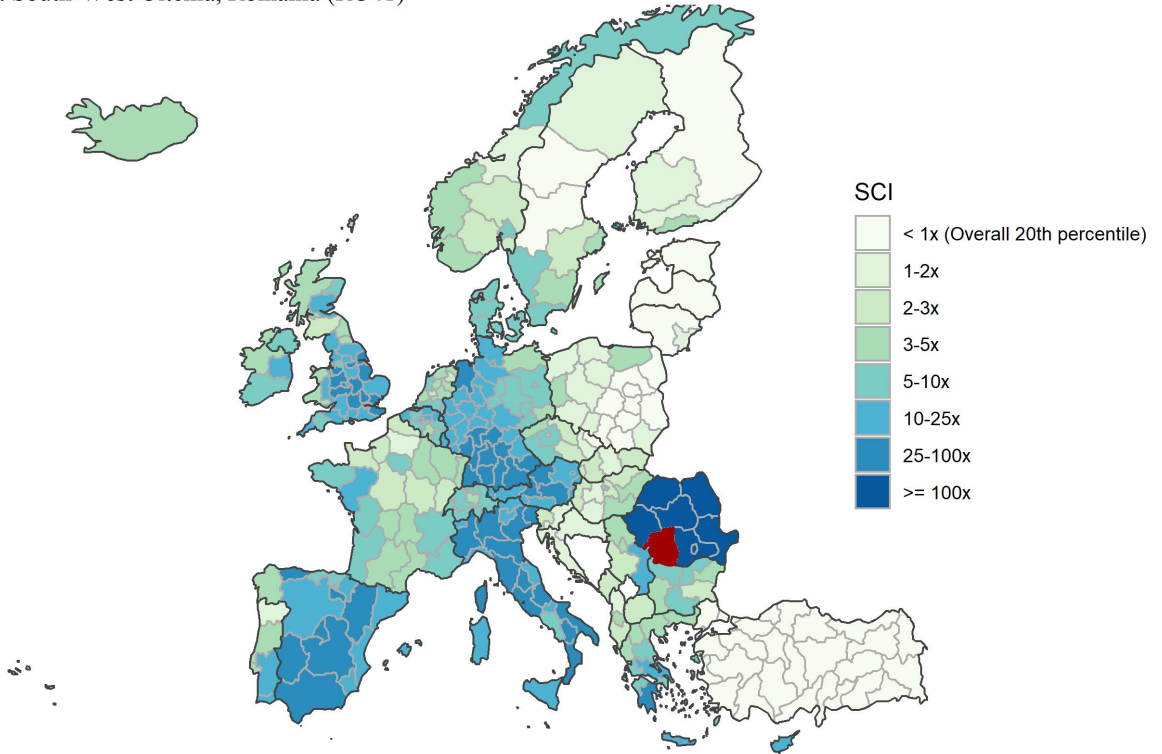
To illustrate the data and explore what factors shape social connections within Europe, we first highlight the geographic structure of social connections in a few selected European regions. We provide additional cases studies in Appendix A.

Figure 1 maps the social network of South-West Oltenia in Romania in Panel A and the Samsun Subregion in Turkey in Panel B; darker shading indicates greater connectedness. In both examples, the strongest social connections are to nearby regions in the same country. Residents of South-West Oltenia have relatively strong social connections throughout Europe, especially with Italy, Spain, Germany, and the United Kingdom. This is likely related to patterns of migration. Romania became a member of the European Union in 2007, which entitled its citizens to certain freedoms to travel and work in other EU member states. According to a report by the World Bank, between 3 and 5 million Romanians currently live and work abroad, representing around a fifth of the country’s population. The top destination countries in 2017 were Italy, Spain, Germany, the United States, and the United Kingdom. By contrast, Panel B shows that the connections between the Samsun Subregion in Turkey, which is not an EU member state, and other European regions are much weaker. The strongest connections between the Samsun Subregion and other countries are concentrated in western Germany and Berlin, with substantially weaker connections in eastern Germany (former German Democratic Republic). These connections likely reflect the lasting impacts of the West Germany’s 1961-1973 labor recruitment agreement (*Anwerbeabkommen*) with Turkey, which resulted in many Turkish workers re-settling in West Germany (Aydin, 2016).

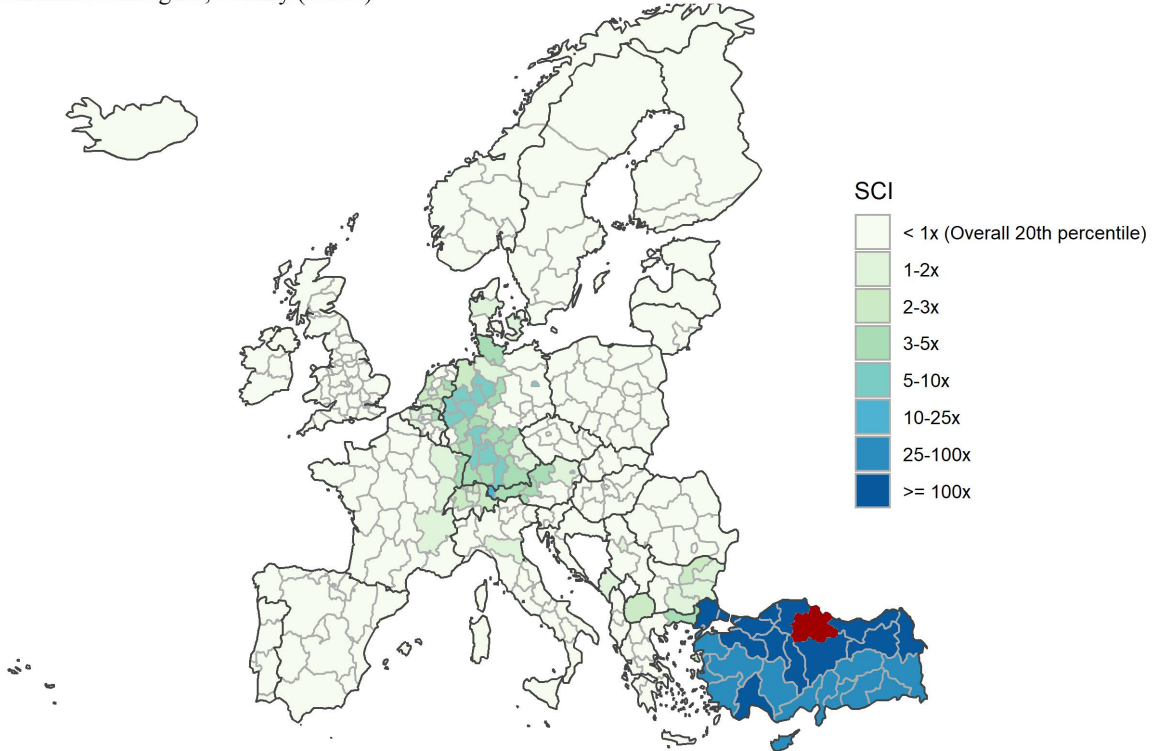
Figure 2 maps the social network of two regions in Belgium: Limburg, in Panel A, and Namur, in Panel B. Again, both regions are most strongly connected to nearby regions within their own coun-

**Figure 1: Social Network Distributions in Romania and Turkey**

A: South-West Oltenia, Romania (RO41)



B: Samsun Subregion, Turkey (TR83)

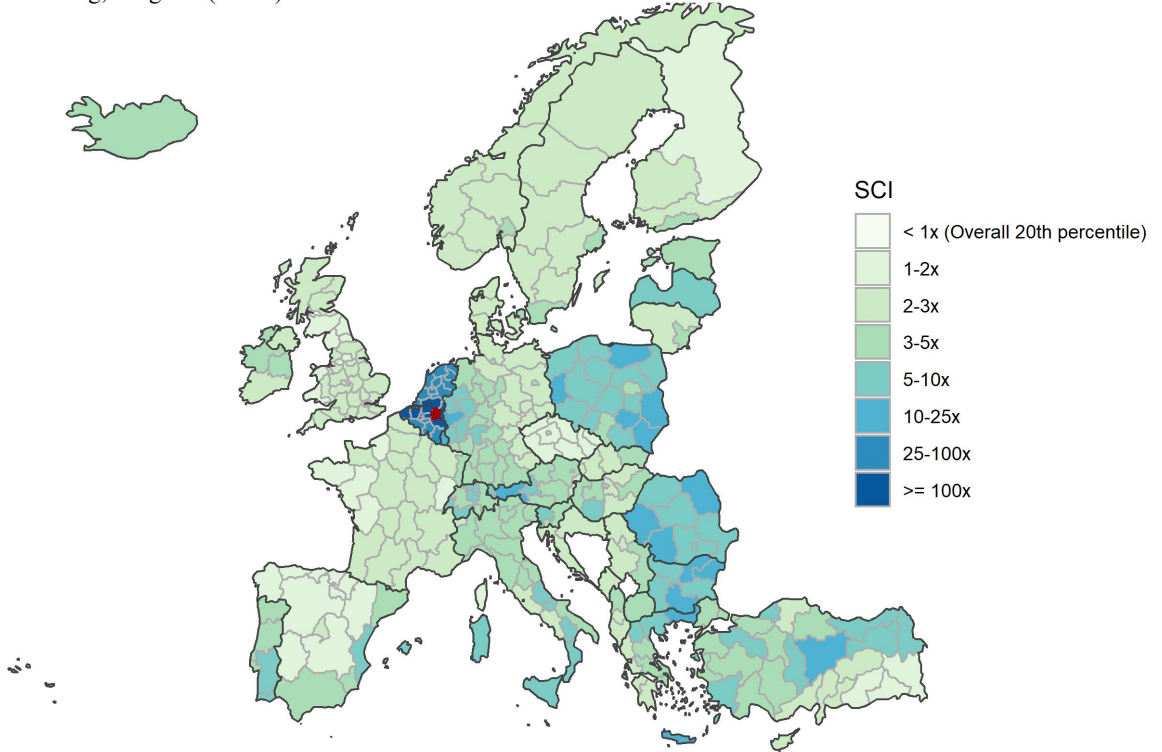


*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions  $j$  with two regions  $i$ : South-West Oltenia, RO (Panel A) and Samsun Subregion, TR (Panel B). The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.

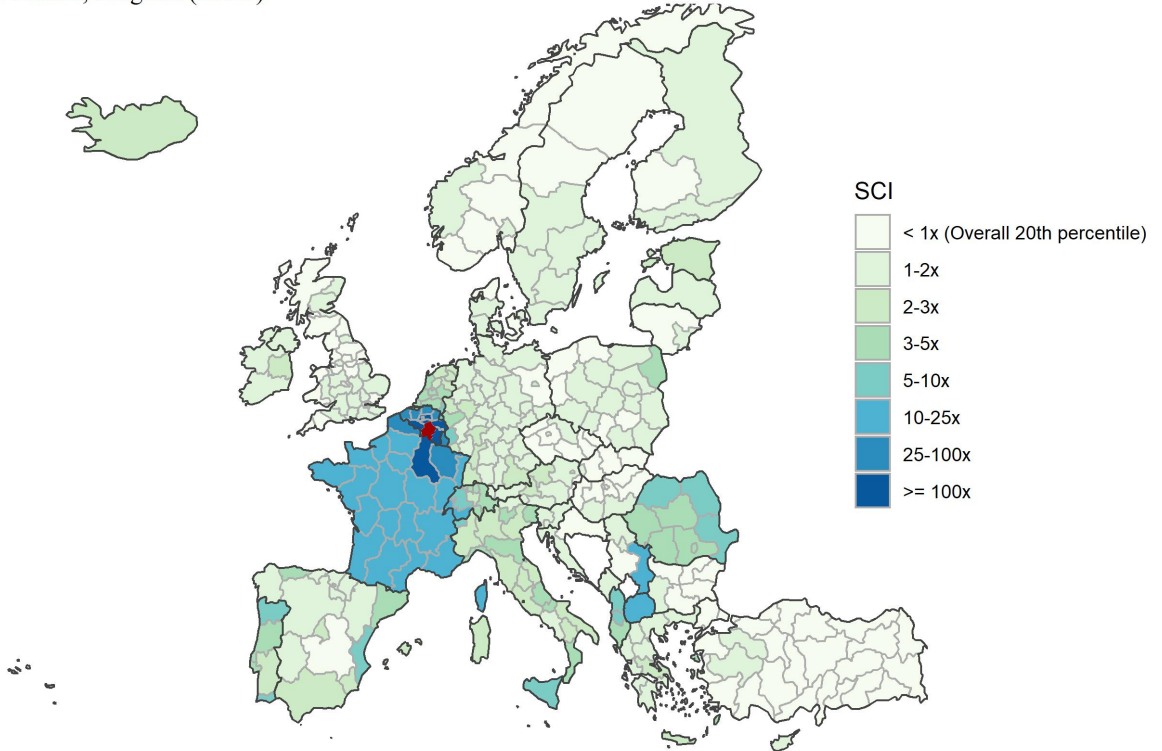


**Figure 2: Social Network Distributions in Belgium**

A: Limburg, Belgium (BE22)



B: Namur, Belgium (BE35)



*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions  $j$  with two regions  $i$ : Limburg, BE (Panel A) and Namur, BE (Panel B). The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.

try. Yet, while the capitals of the two regions (Hasslet and Namur, respectively) are less than 70km apart, the two regions' connections outside Belgium differ substantially. The official and most commonly spoken language in Limburg is Dutch, whereas in Namur it is French. Accordingly, Limburg is more strongly connected to the entire Netherlands to the north and Namur is more strongly connected to areas throughout all of France to the south.

The examples in Figures 1 and 2 suggest that social connections are substantially stronger within countries than between countries, but that there is also substantial within-country variation based on certain regional characteristics. We next seek to understand how these patterns would be reflected if we created communities of regions with strong connections to each other. To do so, we create clusters that maximize within-cluster pairwise social connectedness using hierarchical agglomerative linkage clustering.<sup>5</sup> Figure 3 shows the results when we use this algorithm to group Europe into 20 and 50 communities, instead of the 37 existing countries.

In Panel A, the 20 unit map, nearly all of the community borders (denoted by a change in area color) line up with country borders (denoted by large black lines), consistent with strong intra-country social connectedness. The only exception is in the United Kingdom, where one region, Outer London West and North West, is grouped together with Romania. This area of London has welcomed a large number of Romanian immigrants in recent years and includes Burnt Oak, a community nicknamed "Little Romania" (Harrison, 2018). Furthermore, the borders of cross-country communities mostly line up with historical political borders. For example, every region in the countries that made up Yugoslavia until the early 1990s for which we have data—Slovenia, Croatia, Serbia, North Macedonia, and Montenegro—are grouped together in one community. The same is true of the regions in the Czech Republic and Slovakia, which were united as Czechoslovakia until 1993. Other cross-country communities line up with even older historical borders: the United Kingdom and Ireland, Denmark and Iceland, Sweden and Norway, and Austria and Hungary all split from political unions in the first half of the 20th century but remain grouped together by present-day social connectedness.

While much of the alignment between country and community borders persists in the 50 unit map in Panel B of Figure 3, countries begin to break apart internally. Most of these resulting sub-country communities are spatially contiguous, consistent with distance playing a strong role in social connectedness within countries. One notable exception is Île-de-France, which is home to Paris and accounts for nearly 30% of French GDP. The region is grouped with France's southeastern coast (often referred to as French Riviera) and Corsica, popular vacation destinations for the well-heeled. We also see linguistic communities form: Belgium splits into a French and a Dutch speaking community, and Catalan and Andalusian Spanish communities emerge in Spain. The southeastern corner of Turkey, which has a much higher concentration of Kurds than the rest of the country, also forms its own community. Finally, historical country borders remain influential: the regions of former Czechoslovakia, former Yugoslavia, and Denmark and Iceland still form single communities, while former East Germany forms its own community.

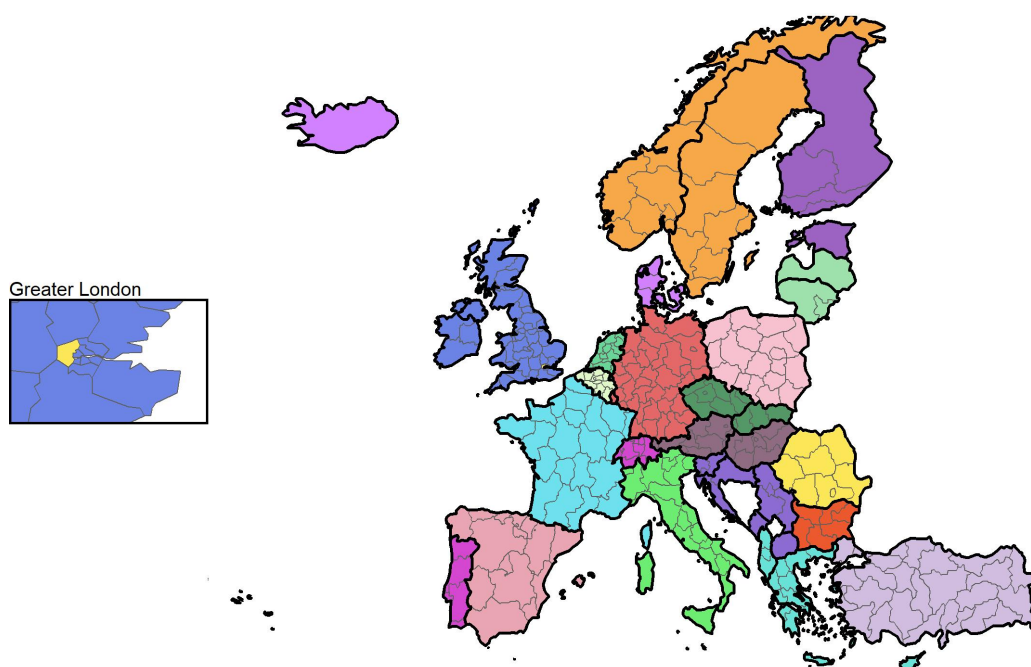
---

<sup>5</sup>There are a number of possible algorithms that could construct such clusters. Conceptually, the agglomerative clustering algorithm starts by considering each of the  $N$  countries as a separate community of size one. It then iteratively combines clusters that are "closest" together. Here, our measure of distance is  $1 / \text{SocialConnectedness}_{i,j}$ . The distance between clusters is the average of the pairwise distances between regions in the clusters.

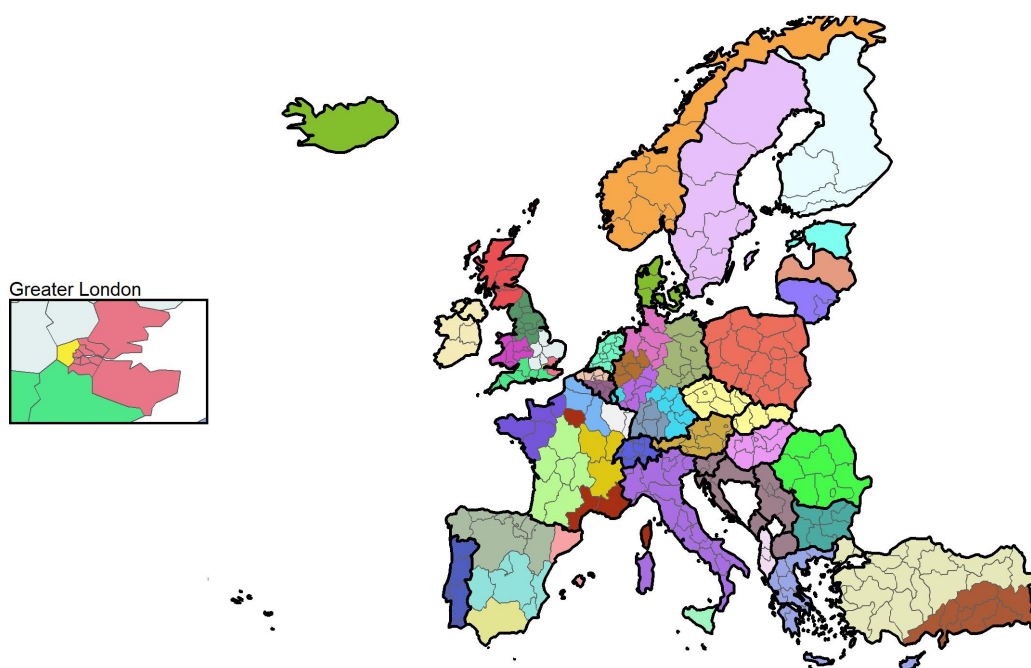


**Figure 3: Socially Connected Communities within Europe**

A: 20 Units



B: 50 Units



*Note:* Figure shows European regions grouped together when we use hierarchical agglomerative linkage clustering to create 20 (Panel A) and 50 (Panel B) distinct groups of regions.

### 3 Determinants of European Social Connectedness

The previous section suggests a relationship between social connectedness, political borders (past and present), geographic distance, language, and other demographic characteristics. We now assess the role of these factors in a regression framework.

**Regional Demographic and Socioeconomic Data.** Information on demographic and socioeconomic characteristics of each region, such as educational attainment, median age, average income, and unemployment rate, is available from Eurostat. We calculate a measure of region-to-region industrial composition similarity using employment data from Eurostat’s Structural Business Statistics series. For regional data on language and religion, we use the European Social Survey. In particular, the survey asks respondents which language they speak most often at home and — if the respondent considers him- or herself religious — their religious affiliation (European Social Survey).<sup>6</sup> Our analysis focuses on the regional pairs within the smaller set of countries for which we have a full set of control data.<sup>7</sup>

**Assessing Potential Determinants of Social Connectedness.** To estimate the relationship between various factors and social connectedness between European regions, we estimate the following equation:

$$\log(\text{SocialConnectedness}_{ij}) = \beta_0 + \beta_1 \log(d_{ij}) + X_{ij} + \psi_i + \psi_j + \epsilon_{ij} \quad (2)$$

The unit of observation is a pair of NUTS2 regions. The dependent variable is the log of Social Connectedness (defined in Equation 1) between regions  $i$  and  $j$ . The geographic distance is denoted by  $\log(d_{ij})$ . The vector  $X_{ij}$  includes measures of similarity and dissimilarity along the following demographic and socioeconomic factors: education (the difference in the share of the population that has only lower secondary education or less),<sup>8</sup> age (the difference in the median age), income (the difference in the average household income), religion (an indicator for whether the regions have the same most common religion), unemployment (the difference in the average unemployment rate for persons aged 15 to 74 from 2009-2018), language (an indicator for whether the regions have the same language most commonly spoken at home), and industry similarity (the cosine distance between vectors of industry employment shares). In some specifications, we also include indicators that are set equal to one if the two regions are in the same or border countries, respectively. All specifications include fixed effects  $\psi_i$  and  $\psi_j$  for regions  $i$  and  $j$ ; this allows us to control for average differences across regions in Facebook usage patterns.

---

<sup>6</sup>Most of our data come from Wave 8 of the survey, which was conducted in 2016. However, when a country was not included in Wave 8, we use the most recent wave of the survey for which the country was included. These countries are Denmark (Wave 7, 2014), Albania, Bulgaria, Cyprus and Slovakia (Wave 6, 2012), Croatia and Greece (Wave 5, 2010), and Latvia and Romania (Wave 4, 2008). In addition, Malta was not surveyed but is comprised of a single NUTS2 region. According to other survey data, 97% of the population considers Maltese their “mother tongue” (European Commission, a) and 95% identifies as Roman Catholic (Sansone, 2018). We include these as the most common language and religion.

<sup>7</sup>We do not include Albania, Iceland, Lichtenstein, Luxembourg, Montenegro, North Macedonia, Serbia, Switzerland, and Turkey; most of these countries are excluded because they are not part of the European Union, and therefore do not participate in many of the data collection efforts we use to construct our data.

<sup>8</sup>Specifically, the International Standard Classification of Education levels 0-2.

**Table 2: Determinants of Social Connectedness Across Region Pairs**

	Dependent Variable: log(SocialConnectedness)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)	-0.572*** (0.038)	-0.737*** (0.027)	-1.177*** (0.032)	-0.591*** (0.031)
Same Country		2.896*** (0.077)	1.651*** (0.124)				
Border Country			0.285*** (0.044)	0.340*** (0.046)			
$\Delta$ Share Pop Low Edu (%)			-0.013*** (0.002)	-0.012*** (0.002)	-0.002 (0.001)	-0.007** (0.003)	-0.000 (0.001)
$\Delta$ Median Age			-0.017*** (0.004)	-0.021*** (0.004)	0.000 (0.003)	-0.014*** (0.005)	0.001 (0.002)
$\Delta$ Avg Income (k €)			0.053*** (0.003)	0.055*** (0.003)	0.015*** (0.002)	0.025*** (0.006)	0.012*** (0.002)
$\Delta$ Unemployment (%)			-0.000 (0.005)	0.004 (0.005)	0.006* (0.003)	0.021** (0.010)	0.007* (0.004)
Same Religion			0.027 (0.031)	0.049* (0.025)	0.044*** (0.013)	0.127*** (0.040)	0.029** (0.013)
Same Language			1.493*** (0.097)	1.548*** (0.120)	1.529*** (0.216)	2.279*** (0.133)	1.909*** (0.107)
Industry Similarity			0.128 (0.169)	0.044 (0.158)	0.528*** (0.107)	0.242 (0.199)	0.633*** (0.109)
NUTS2 FEs	Y	Y	Y	Y	Y	Y	Y
Indiv. Same Country FEs				Y			
All Country Pair FEs					Y	Y	Y
Sample						Same country	Diff. country
$R^2$	0.490	0.669	0.745	0.775	0.906	0.927	0.839
Number of Observations	75,900	75,900	75,900	75,900	75,900	5,266	70,634

Note: Table shows results from Regression 2. The unit of observation is a NUTS2 region pair. The dependent variable in all columns is the log of *SocialConnectedness<sub>ij</sub>*. Column 1 includes the log of distance and region fixed effects. Column 2 adds a control for regions in the same country. Column 3 incorporates demographic and socioeconomic similarity measures, as well as a control for regions in countries that border. Column 4 adds fixed effect for each same-country pair. Column 5 adds fixed effects for each country pair. Columns 6 and 7 limit the observations to pairs in the same country and pairs in different countries, respectively. Standard errors are double clustered by each region  $i$  and region  $j$  in a region pair. Significance levels: \*( $p < 0.10$ ), \*\*( $p < 0.05$ ), \*\*\*( $p < 0.01$ ).

Table 2 shows regressions estimates of Equation 2. Column 1 includes only the distance measure,  $\log(d_{ij})$ , and the region fixed effects. A 10% increase in the distance between two regions is associated with a 13.2% decline in the connectedness between those regions. This elasticity is comparable to that

observed for U.S. county pairs in Bailey et al. (2018b). However, the amount of variation in connectedness that distance alone is able to explain is substantially lower in Europe than it is in the United States — in Europe, distance explains 36% of the variation in social connectedness not explained by region fixed effects, while the same number is 65% for the United States. In column 2, we add the variable indicating whether both regions are in the same country. This “same country effect” explains an additional 18% of the cross-sectional variation in region-to-region social connectedness. The estimated elasticity is larger in magnitude than for same-state indicators in the Bailey et al. (2018b) U.S. county regressions, suggesting that there is a greater drop-off in social connectedness at European national borders than at U.S. state borders.

In column 3, we add differences in demographics and socioeconomic outcomes and an indicator for regions that are in bordering countries as explanatory variables. Regions with the same language and those where residents are more similar in terms of educational attainment and age are more connected to each other. Such “homophily” – more friendship links between similar individuals, regions or countries – has been documented in prior work (Lazarsfeld and Merton, 1954; Zipf, 1949; Verbugge, 1983; Marmaros and Sacerdote, 2006; Bailey et al., 2018a,b, 2019a; State et al., 2015). Our estimates suggest that social connectedness between two regions with the same language is about 4.5 times larger than for two regions without the same language, even after controlling for same country and border country effects, geographic distance, and other demographic and socioeconomic factors. When we include language and demographic factors, the estimated effect of being in the same country falls (from a coefficient estimate of 2.9 to 1.6) suggesting that some—but not all—of the higher in intra-country connectedness is due to common language and other demographic similarities.

Somewhat surprisingly, we see higher connectedness between regions with larger differences in income, even after controlling for country-pair fixed effects, and both limiting to regions within the same country and limiting to regions in different countries. In some of these specifications, we also see a positive relationship between connectedness and differences in unemployment. These relationships run contrary to findings from prior research that finds positive relationships between connectedness and income similarity across U.S. counties and New York zip codes (Bailey et al., 2019a, 2018b).

A possible explanation is related to the migration patterns suggested by our case studies: migrants are particularly likely to move from regions with low income (or higher unemployment) to regions with higher income (or lower unemployment) and comparatively less likely to move to other low or middle income regions. Hence, we see more migration and more connections between regions with large differences in income versus those with more similar levels of income or unemployment. This finding is particularly interesting in light of a recent and substantial literature on intra-U.S. migration that documents a general decline in moves over the past three decades and the importance of opportunistic moves for the U.S. labor market (for example, Karahan and Li, 2016; Kaplan and Schulhofer-Wohl, 2012; Molloy et al., 2017; Yagan, 2014). By contrast, much less is known about regional migration flows within Europe, largely due to a lack of comprehensive data. The existing prior research has focused on country-to-country flows (International Organization for Migration), the intensity of within country migration (Bell et al., 2015; Esipova et al., 2013), or regional net-migration (Sardadvar and Rocha-Akis, 2016). Our unique data set on connectedness provides insights into region-to-region migration patterns through-

out the continent. For example, existing data show that within country moves in Europe are generally less common than in the United States; however, the positive relationship we observe between income dissimilarity and connectedness, compared to the negative relationship observed in the U.S., suggest that there may be higher rates of migration in Europe from less prosperous to more prosperous regions. These are exactly the opportunistic moves that increase labor market dynamism.

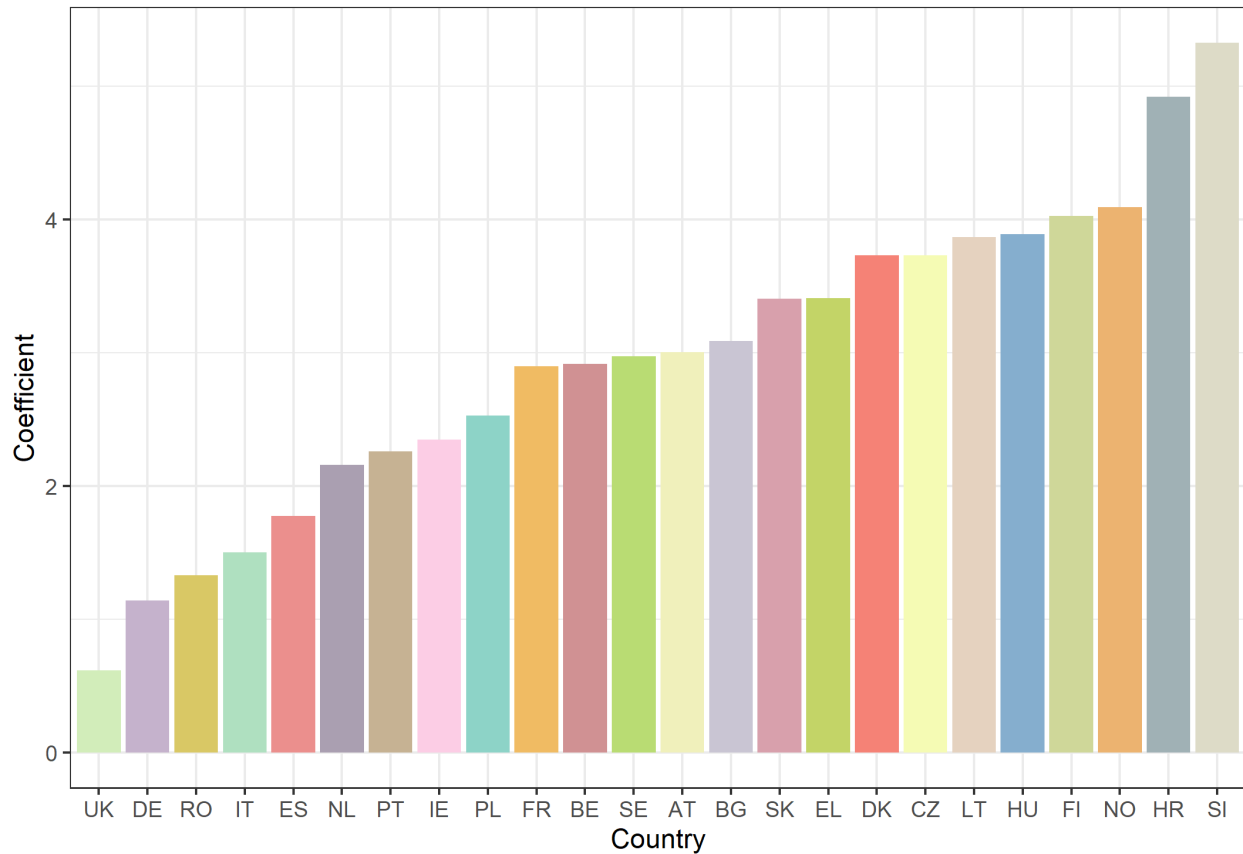
Column 4 adds fixed effects for each same-country pair, and column 5 adds fixed effects for every country pair. The magnitude of the coefficient on income dissimilarity falls, consistent with country-level migration flows explaining some of the connectedness between regions with dissimilar incomes; however, even holding average connectedness across country pairs fixed, social connectedness is stronger between regions with more different incomes. Columns 6 and 7 limit to pairs of regions in the same and different countries, respectively. Social connectedness declines in geographic distance more within countries than across countries: a 10% greater geographic distance between regions within the same country implies a 11.7% decrease in social connectedness, whereas a 10% greater geographic distance between regions in different countries implies only a 5.9% decrease in connectedness (conditional on other controls).

**Strength of Within-Country Connectedness.** So far, we have shown that, on average, regions in the same country are more connected than regions in different countries that are similarly far apart. We next explore the extent of heterogeneity in this within-country effect on connectedness. We do so by comparing the coefficients on the individual same country effects estimated in column 4 of Table 2, which capture the additional connectedness associated with two regions being part of the same country. Figure 4 shows these coefficients plotted for all countries with two or more NUTS2 regions. Higher values are indicative of stronger within-country connectedness. Within-country connectedness is generally stronger for countries with smaller populations, such as Slovenia and Croatia, than for countries with larger populations, such as the United Kingdom and Germany. But there are noticeable differences between countries of similar sizes. For example, the United Kingdom and France have roughly equal populations, yet two regions in France are on average 18 times more connected than two similarly situated regions in Europe, whereas two regions in the United Kingdom are only 1.8 times more connected. This pattern also aligns with the number of within-country communities formed in the 50 unit map of Figure 3 — 6 in France and 8 in the UK. There are several possible reasons for such differences, such as historical patterns (e.g., did the nations unite at different times?), geography (e.g., are there physical barriers that separate parts of the nations?), or modern government structures (e.g., do sub-regional governments have greater autonomy in some countries than in others?).

**Relationship Between Historical Borders and Connectedness.** Next, we take a more detailed look at the relationship between historical political boundaries and today's social connectedness. Information on national borders in 1900, 1930, 1960, and 1990 comes from files provided by the Max Planck Institute for Demographic Research Population History GIS Collection (MPIDR and CGG).<sup>9</sup> In Appendix B, we provide maps that show the region of each country in each year of interest. Table 3 adds additional

<sup>9</sup>In cases when a modern NUTS2 region spans two historical countries, we classify the region as part of the country for which it had a greater land area overlap.

**Figure 4: Connectedness within European Countries**



*Note:* Figure shows the coefficients of the individual same-country effects from the regression reported in column 4 of Table 2. The coefficients are roughly the additional connectedness that is associated with two regions being part of the same country, for each country. Higher values are indicative of stronger within-country connectedness (after controlling for certain demographic and socioeconomic effects). The labels on the x-axis are the two-letter prefix of each country's NUTS codes.

variables based on these historical borders to the analysis in Table 2. Column 1 uses all of the same controls as column 4 in Table 2 except  $\log(\text{Distance})$ ; throughout this table, we instead use 100 dummy variables representing percentiles of the distribution of distance to avoid picking up non-linearities in the relationship between geographic distance and historical borders.<sup>10</sup> The remaining columns, columns 2 to 5, add indicators based on national borders at the start of 1990, 1960, 1930, and 1900, respectively.

We look at several major European border changes dating back to the early 20th century, showing that present connectedness is higher between regions that have been part of the same country in the past. This result is *in addition* to the effects of being in the same country today, being in bordering countries today, region-to-region distance, and all the demographic and socioeconomic controls in Table 2. The largest increases in connectedness from having been part of the same country are associated with the most recent border changes. For example, two regions in former Czechoslovakia (which split in 1993) are more than 19 times more connected on average than similar region pairs in other countries. Likewise, two regions in former Yugoslavia (which split in the early 1990s) are more than 13 times more

<sup>10</sup>In general, the historical coefficients in Table 3 do not change, or become slightly larger, when using  $\log(\text{Distance})$ .



connected. Patterns of social connectedness are also related to country borders prior to the 1990-1991 fall of the Soviet Union. Specifically, connectedness between regions that were both within East Germany is more than 2 times higher than connectedness between other similar region pairs in Germany. Pairs of regions in the three countries in our data that were former republics of the Soviet Union—Estonia, Latvia, and Lithuania—are also 6 times more connected than similar region pairs.

Borders dating back to earlier in the 20th century appear to have weaker, though still economically and statistically significant relationships with present-day social connectedness. In the early 20th century, the United Kingdom controlled both Malta and Cyprus (the two became independent in 1960 and 1964, respectively). A pair of regions in Malta, Cyprus, or the UK are twice as connected as a similarly situated regional pair, again, over and above modern country borders. The borders of Germany in 1930 were also different than today: the country included the Liege region in modern Belgium and a number of regions in modern Poland, but not the Saarland—a formerly independent nation within modern Germany. We find a 46% increase in connectedness between regions in 1930 Germany. (As discussed below, this falls when we add the 1900 German borders.)

Finally, we look at three national borders that changed before or shortly after the first World War: the Austro-Hungarian Empire, the German Empire, and the United Kingdoms of Sweden-Norway. In 1900, the Austro-Hungarian Empire stretched across much of central and eastern Europe, encompassing part or all of modern Austria, Hungary, Czech Republic, Slovakia, Slovenia, Croatia, Romania, Poland, and Italy. After adding our present-day controls, we find that two regions within this empire are more than 90% more connected than a pair of otherwise similar regions. Compared to modern Germany, the German Empire in 1900 controlled large parts of modern Poland (even more so than 1930 Germany) and the Alsace region of France. We find that being part of the German Empire in 1900 is associated with a 50% increase in connectedness, again controlling for both the effects of the modern German borders and 1930 German borders. It is interesting that the regression primarily loads on the older 1900 borders, while the coefficient for the 1930 borders decreases. One possible explanation is the period of time the borders were in effect: whereas the 1930 German borders were effective only in the 20 year interwar period (and indeed changed even during that period), the 1900 borders essentially remained unchanged for nearly 50 years between 1871 to 1918. Lastly, from 1814 to 1905 the lands of present-day Sweden and Norway were united under a common monarch as the United Kingdoms of Sweden and Norway. A pair of regions within this union are more than 7 times more connected today than similarly situated regions in bordering countries. As with all of our analyses, the historical patterns we observe are correlations rather than necessarily causal and may also capture the effect of other factors that relate to historical borders that we do not explicitly control for.

**Table 3: Historical Determinants of Social Connectedness**

	Dependent Variable: log(SocialConnectedness)				
	(1)	(2) 1990	(3) 1960	(4) 1930	(5) 1900
Border Country	0.418*** (0.045)	0.399*** (0.045)	0.392*** (0.045)	0.372*** (0.045)	0.310*** (0.043)
Both Czechoslovakia		3.525*** (0.217)	3.529*** (0.217)	3.541*** (0.216)	2.945*** (0.217)
Both Yugoslavia		3.108*** (0.105)	3.110*** (0.105)	3.123*** (0.105)	2.616*** (0.114)
Both West Germany		0.006 (0.046)	0.005 (0.046)	0.015 (0.044)	-0.005 (0.043)
Both East Germany		1.088*** (0.053)	1.092*** (0.053)	1.072*** (0.055)	1.124*** (0.050)
Both Soviet Union		1.884*** (0.080)	1.874*** (0.081)	1.882*** (0.081)	2.052*** (0.077)
Both United Kingdom 1960			1.015*** (0.155)	1.016*** (0.156)	0.998*** (0.157)
Both Germany 1930				0.465*** (0.104)	0.159** (0.063)
Both Austro-Hungarian Empire 1900					0.920*** (0.111)
Both German Empire 1900					0.492*** (0.074)
Both United Sweden-Norway					2.057*** (0.123)
Distance Controls	Y	Y	Y	Y	Y
Table 1 Controls	Y	Y	Y	Y	Y
NUTS2 FEs	Y	Y	Y	Y	Y
Indiv. Same Country FEs	Y	Y	Y	Y	Y
$R^2$	0.784	0.790	0.791	0.792	0.801
Number of Observations	75,900	75,900	75,900	75,900	75,900

Note: Table shows results from Regression 2 with added historical country borders controls  $X_{ij}$ . The unit of observation is a NUTS2 region pair. The dependent variable in all columns is the log of  $SocialConnectedness_{ij}$ . Every column includes controls for same country, region  $i$ , and region  $j$  effects. Column 1 is the same as column 4 of Table 2, except with 100 dummy variables representing percentiles of distance instead of log(distance). Columns 2, 3, 4, and 5 add controls for certain historical borders in 1990, 1960, 1930, and 1900, respectively. Coefficients for the demographic and socioeconomic controls in Table 2 are excluded for brevity. Standard errors are double clustered by each region  $i$  and region  $j$  in a region-pair. Significance levels: \*( $p < 0.10$ ), \*\*( $p < 0.05$ ), \*\*\*( $p < 0.01$ ).

## 4 Social Connectedness and Travel

Social connectedness between two regions may be related to certain economic and social interactions. For example, Bailey et al. (2018b) documents correlations between social connectedness, trade flows, patent citations, and migration patterns between U.S. counties. Bailey et al. (2019a) looks more specifically at transportation, highlighting the relationship between transportation infrastructure and urban social networks. In this section, we look at the relationship between social connectedness and European passenger train flows. Our unique data allow us to add a continental-level empirical analysis to an existing literature of theoretical and survey-based studies that explore the influence of social networks on travel decisions (for an overview, see Kim et al., 2018).

**Regional Data on European Travel.** For our dependent variable, we use the number of passenger train trips between each regional pair  $i$  and  $j$ . We discuss the availability of these data, as well as our procedure for cleaning and standardization, in Appendix C. We also use information on rail and drive travel times between region geographic centers. Rail travel times come from the European Transport Information System 2010 “observed” data.<sup>11</sup> Drive times were generated using the Open Source Routing Machine, an OpenStreetMap-based routing service.<sup>12</sup>

**Relationship between Travel and Connectedness.** As the outcome variable is non-negative and contains many zeros, we follow Correia et al. (2019) and estimate a Poisson Pseudo-Maximum Likelihood regression model. Specifically, our equation of interest is:

$$PassengerTrain_{ij} = \exp[\beta_0 \log(SocialConnectedness_{ij}) + X_{ij} + \alpha \log(D_{ij}) + \psi_i + \psi_j] \cdot \epsilon_{ij} \quad (3)$$

Here, the definitions of  $\log(SocialConnectedness_{ij})$ ,  $X_{ij}$ ,  $\psi_i$ , and  $\psi_j$  remain unchanged from Equation 2. The variable  $PassengerTrain_{ij}$  is the number of rail passengers that travel from region  $i$  to region  $j$  in a given year. The vector for “distance,” represented by  $D_{ij}$ , includes the geographic distance, as well as the rail and driving travel times in minutes between the central points of the regions.

Table 4 shows the results from Regression 3. Due to differences in data availability between years, we present results using the most recent year for which passenger train flows for a given  $i, j$  pair is available (in columns 1-5) and, separately, results using 2015, 2010, and 2005 (columns 6, 7, and 8). Column 1 includes the log of  $SocialConnectedness_{ij}$  and the NUTS2 region fixed effects as the only explanatory variables. We find that a 10% increase in connectedness between a pair of regions is associated with a 17% increase in passenger rail traffic between the two. Column 2 adds the log of the geographic distance between the regions, which, intuitively, has a negative relationship with the number of passenger train trips. In column 3, we add controls for the travel time, separately by train and car, between the central points of the regions. In general, we find that these travel times have a stronger negative relationship with passenger rail traffic than distance alone. Column 4 adds all country pair fixed effects;

<sup>11</sup>Data available at: <http://ftp.demis.nl/outgoing/etisplus/datadeliverables/TextFiles/>.

<sup>12</sup>OSRM relies on only open-source software and does not use real-time data (e.g., traffic). Both of these factors increase replicability compared to other online-mapping services. For more information, see: <http://project-osrm.org/>.

column 5 adds all the demographic/soceioeconomic controls from Table 2; and columns 6, 7, and 8 repeat these analyses for the years 2015, 2010, and 2005, respectively. Adjusting for geographic distance, travel time by car and train, and country fixed effects, we find that, on average, a pair of regions with 10% higher social connectedness will have between 11.7 and 13.6% more rail passengers travel between them. Overall, our results provide large-scale empirical evidence to support existing models that suggest social networks play an important role in individuals' travel decisions (see e.g., Axhausen, 2008; Carrasco and Miller, 2009; PÄqez and Scott, 2007).

**Table 4: Social Connectedness and Passenger Train Travel**

Dependent Variable: log(Passenger Rail Trips)								
	(1) Most recent	(2) Most recent	(3) Most recent	(4) Most recent	(5) Most recent	(6) 2015	(7) 2010	(8) 2005
log(SocialConnectedness)	1.711*** (0.052)	1.506*** (0.080)	1.416*** (0.101)	1.290*** (0.112)	1.364*** (0.099)	1.241*** (0.056)	1.167*** (0.071)	1.236*** (0.093)
log(Distance in KM)		-0.361** (0.180)	0.259 (0.327)	0.041 (0.354)	0.118 (0.284)	0.109 (0.172)	0.627*** (0.198)	0.076 (0.252)
log(Rail Time in Mins)			-0.493 (0.442)	-0.004 (0.528)	0.104 (0.543)	-1.096*** (0.162)	-1.071*** (0.251)	0.036 (0.490)
log(Drive Time in Mins)			-0.396** (0.201)	-0.446** (0.198)	-0.448*** (0.170)	-0.013 (0.110)	-0.562*** (0.172)	-0.409** (0.165)
NUTS2 FEs	Y	Y	Y	Y	Y	Y	Y	Y
All Country Pair FEs				Y	Y	Y	Y	Y
Table 1 Controls					Y	Y	Y	Y
pseudo- $R^2$	0.964	0.965	0.965	0.972	0.975	0.980	0.980	0.975
Number of Observations	85,390	85,390	77,390	77,390	61,810	27,650	32,176	58,462
N Explained by FEs	30,698	30,698	23,784	49,000	36,517	19,146	16,269	29,659

*Note:* Table shows results from Regression 3. The unit of observation is a NUTS2 region pair. The dependent variable in all columns is the number of passenger rail trips in 2015, 2010, or 2005 from region  $i$  to region  $j$ . In columns 1-5, we use the most recent year for which these data are available for a given  $i, j$  pair. In columns 6, 7, and 8, we use only the 2015, 2010, and 2005 data, respectively. Column 1 shows the results from using the log of *SocialConnectedness<sub>ij</sub>* and NUTS2 region fixed effects as the only explanatory variables. Column 2 adds the log of the geographic distance between the regions. Column 3 adds the log of the travel time, by train and car, between the central points of the regions. Column 4 adds all country-pair fixed effects. Columns 5-8 add the demographic/soceioeconomic controls from Table 2. Observations that are fully explained by fixed effects are dropped before the PPML estimation. Standard errors are double clustered by each region  $i$  and region  $j$  in a region pair. Significance levels: \*( $p < 0.10$ ), \*\*( $p < 0.05$ ), \*\*\*( $p < 0.01$ ).

## 5 International Connections and Views on European Union

A central goal of the European Union is to enhance cohesion and solidarity across European countries, and a variety of programs, such as the Erasmus exchange student program, explicitly exist to foster this connectivity (European Union; European Commission). However, in recent years there has been a decline in trust in the EU and a rise of anti-EU voting, that have lead to, for example, the United Kingdom's 2016 vote to exit the European Union. A number of studies explore correlates with, and potential origins of, support for Eurosceptic political parties (for example Algan et al., 2017; Becker et al., 2017;

Colantone and Stanig, 2018; Inglehart and Norris, 2016). While much of this research emphasizes either economic insecurity or a cultural backlash against multiculturalism/progressive values, a related strand of research explores the role of personal connections in shaping political preferences. Early evidence was provided by Lazarsfeld et al. (1944), who documents the influence of friends on U.S. voters. More recently, McLaren (2003) finds that intimate contact with members of minority groups can reduce individuals' willingness to expel immigrants and Algan et al. (2019) finds that the formation of friendships between students with different political opinions causes their views to converge. Motivated by this literature, we explore whether differences in how socially connected a region is to other European countries is associated with higher trust in the European Union and a reduced propensity to vote for Eurosceptic political parties.

European regions vary considerably on the share of social connections to individuals living in foreign countries.<sup>13</sup> For the median NUTS2 region, 9.8% of European connections are to individuals in different countries. However, in the least internationally connected decile of regions, less than 4.1% of connections are to individuals living in different countries, compared to over 19.7% in the most internationally connected decile. Figure 5 maps this share for every region.

**Regional Data on Euroscepticism.** We analyze two measures of Euroscepticism. Our first measure is a survey measure of trust in the EU. As part of the “Eurobarometer” series of public opinion surveys undertaken for the European Commission, the *Flash Eurobarometer 472: Public opinion in the EU regions* asks respondents about their trust in the EU. Specifically, the survey asks respondents “I would like to ask you a question about how much trust you have in the European Union. Could you tell me if you tend to trust it or tend not to trust it?” (European Commission, b). The survey was conducted between October and November 2018, and had 61,968 respondents (around 300 per region). Second, we use the vote share for anti-EU parties. Regional-level data on parliamentary and presidential elections in Europe from 2000 to 2017 are available from Algan et al. (2017).<sup>14</sup> As described in the paper, the determination of whether a particular party is “Anti-EU” is primarily based on the Chapel Hill Expert Survey, with some additional analysis of party platforms by the paper’s authors. From this data, we use the average share of the electorate that voted for Anti-EU parties in elections between 2009 and 2017. In addition to these two measures of Euroscepticism, we use demographic and socioeconomic data from Eurostat, including the share of each region’s population that was born in a different European country.

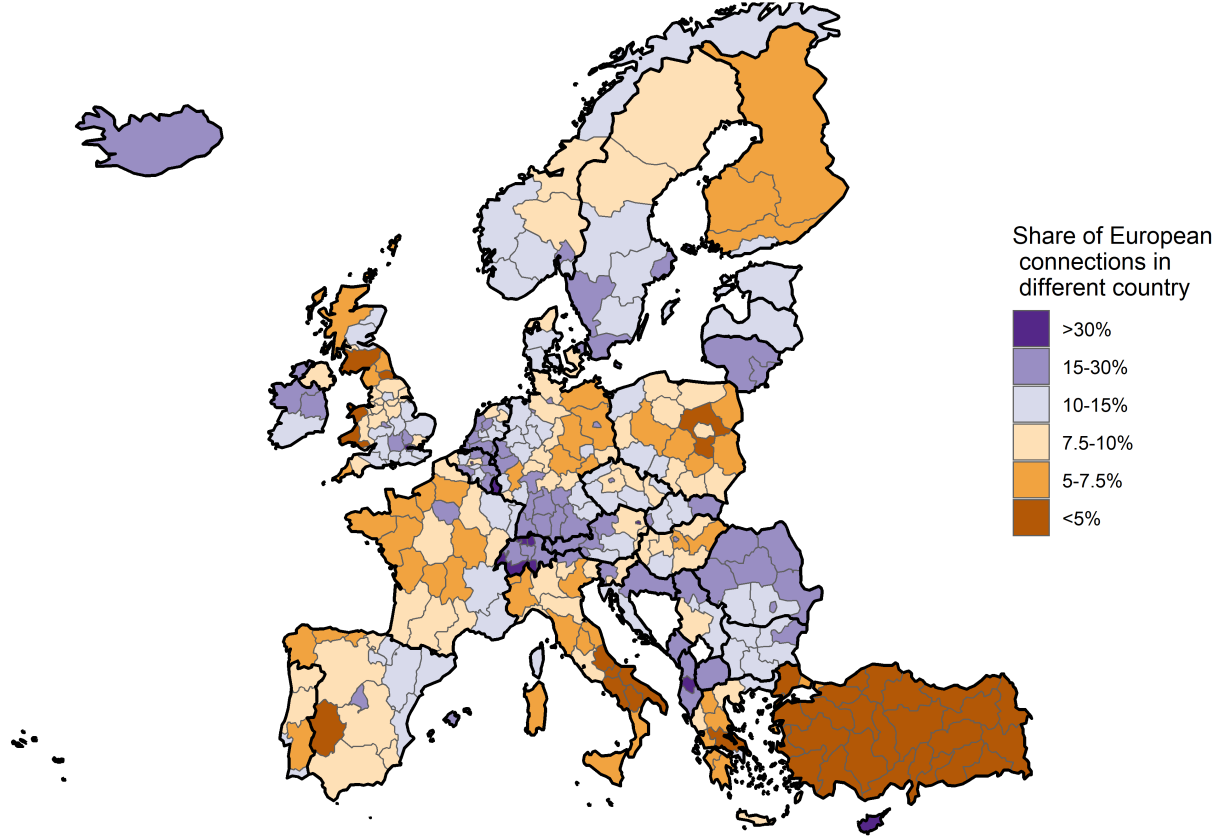
**Relationship between Euroscepticism and Connectedness.** To systematically explore the relationship between regional views on the European Union and international friendships, we use the equation:

$$Share\_EU\_View_i = \beta_0 * Share\_Connections\_Foreign_i + X_i + \zeta_{c(i)} + \epsilon_i \quad (4)$$

<sup>13</sup>In our data, we focus on connections between individuals living within Europe. Therefore, (as we describe) our measure is the share of *European* connections to individuals in different *European* countries, rather than the share of *all* connections in *any* different country.

<sup>14</sup>Data and programs are available at: <https://www.brookings.edu/bpea-articles/the-european-trust-crisis-and-the-rise-of-populism/>.

**Figure 5: International Connectedness within Europe**



*Note:* Figure shows the share of European friendship links from users in each region to users in a different country. Darker orange regions have a lower share of European friends in different countries and darker purple regions have a higher share.

Here,  $Share\_EU\_View_i$  is either the share of individuals responding that they trust the European Union in the Eurobarometer survey or the average share of the electorate that voted for Anti-EU parties between 2009 and 2017. Our unit of observation for this regression is the level at which we observe these two outcomes of interest. For both data sets, this includes a mixture of NUTS1 and NUTS2 regions. The first explanatory variable,  $Share\_Connections\_Foreign_i$ , is the share of the region's European connections that are to individuals in foreign countries.  $X_i$  is a set of regional socioeconomic characteristics. These are average income, unemployment rate, and the shares of employment in manufacturing, construction, and professional sectors. They may also include the share of residents living in the region who are born in other European countries. The controls are indicators based on the deciles of each measure, to capture any non-linear relationships. Some specifications include country fixed effects, denoted here by  $\zeta_{c(i)}$ .

Table 5 shows the results of this regression. Columns 1 and 5 show the relationships between the outcomes of interest and the share of the region's European connections that are to individuals in foreign countries: a 1 percentage point increase in the share of connections that are to other European countries is associated with a 0.50 percentage point increase in the share of residents who trust the EU and a 0.76 percentage point decrease in the share of votes for an Anti-EU political party. The second



and fifth columns show that when adding the socioeconomic controls listed above, the magnitude of the relationship between the trust in the EU decreases and becomes insignificant, but the magnitude of the relationship with anti-EU voting slightly increases. When we also add the share of the population that is born in other European countries (columns 3 and 7), the magnitude of the relationship with trust increases and the magnitude of the relationship with voting decreases. There is a complicated relationship between demographic and socioeconomic factors, and views on the European Union; however, the directionally persistent relationship between Euroscepticism and international social connectedness at the regional level suggest foreign connections might play a role in shaping views on the EU, a result that would be consistent with the existing literature on personal connections impacting political preferences. Columns 4 and 8 show that, after adding country-level fixed effects, increases in a region’s foreign connections continue to imply a decrease in the share of residents that trust the EU, while the effect on the share that vote for anti-EU parties becomes insignificant. This suggests that country-specific factors (e.g. the organizational resources of anti-EU political parties within the country) dominate possible effects of international connections in shaping Eurosceptic voting patterns, but not in shaping individuals’ views.

**Table 5: International Connections and Views on the European Union**

	Share Trust in EU (%)				Share Vote for Anti-EU Parties (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of European friends outside of country (%)	0.499*** (0.101)	0.161 (0.127)	0.254* (0.153)	0.326** (0.155)	-0.762*** (0.187)	-0.819*** (0.230)	-0.541** (0.273)	0.454 (0.303)
Socioeconomic Controls		Y	Y	Y		Y	Y	Y
Share Foreign Born Controls			Y	Y			Y	Y
Country Fixed Effects				Y				Y
$R^2$	0.109	0.503	0.571	0.891	0.071	0.493	0.568	0.818
Number of Observations	201	201	201	201	221	221	221	221

*Note:* Table shows results from Regression 4. The dependent variable in columns 1-4 is the share of individuals responding that they trust the European Union in a survey conducted for the European Commission. The dependent variable in columns 5-8 is the average share of the electorate that voted for “Anti-EU” parties between 2009 and 2017. The unit of observation is the level at which we observe each dependent variable (either a NUTS2 or NUTS1 region). Columns 1 and 5 include only one explanatory variable: the share of the region’s European connections that are to individuals in foreign countries. Columns 2 and 6 add a set of demographic and economic controls. Columns 3 and 7 add controls for the share of the region’s population that was born in a different country. Columns 4 and 8 add country-level fixed effects. Significance levels: \*( $p < 0.10$ ), \*\*( $p < 0.05$ ), \*\*\*( $p < 0.01$ ).

## 6 Conclusion

We use aggregated data from Facebook to better understand social connections in Europe. We find sizeable connectedness decreases in geographic distance and at country borders. Using a number of European border changes in the 20th century (such as the breakups of the Austro-Hungarian Empire and Czechoslovakia), we find that the relationship between political borders and connectedness can persist decades after the boundaries change. We also find evidence of homophily in Europe, as social connections are stronger between regions with residents of similar ages and education levels, as well as between those that share a language and religion. However, region pairs with *dissimilar* incomes are more connected, likely due to migration from poorer to richer regions. We find that more socially

connected region-pairs have more train trips between them, even after controlling for distance and travel time. Furthermore, we explore the variation in overall international connectedness between European regions. We find that regions with a higher share of connections to other countries have higher rates of trust in the E.U. and lower rates of voting for anti-E.U. political parties.

## References

- Yan Algan, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. The european trust crisis and the rise of populism. Brookings Papers on Economic Activity, 2017.
- Yann Algan, Nicolò Dalvit, Quoc-Anh Do, Alexis Le Chapelain, and Yves Zenou. Friendship networks and political opinions: A natural experiment among future french politicians. Working Paper 1294, Research Institute of Industrial Economics, 2019.
- K W Axhausen. Social networks, mobility biographies, and travel: Survey challenges. Environment and Planning B: Planning and Design, 35(6):981–996, 2008.
- Yasar Aydin. The germany-turkey migration corridor: Refitting policies for a transnational age. Report, Migration Policy Institute, 2016.
- Michael Bailey, Eduardo D’Ávila, Theresa Kuchler, and Johannes Stroebe. House price beliefs and mortgage leverage choice. Working Paper 24091, National Bureau of Economic Research, 2017.
- Michael Bailey, Rachel Cao, Theresa Kuchler, and Johannes Stroebe. The economic effects of social networks: Evidence from the housing market. Journal of Political Economy, 126(6):2224–2276, 2018a.
- Michael Bailey, Rachel Cao, Theresa Kuchler, Johannes Stroebe, and Arlene Wong. Social connectedness: Measurements, determinants, and effects. Journal of Economic Perspectives, 32(3):259–80, 2018b.
- Michael Bailey, Patrick Farrell, Theresa Kuchler, and Johannes Stroebe. Social connectedness in urban areas. Working Paper 26029, National Bureau of Economic Research, 2019a.
- Michael Bailey, Drew M Johnston, Theresa Kuchler, Johannes Stroebe, and Arlene Wong. Peer effects in product adoption. Working Paper 25843, National Bureau of Economic Research, 2019b.
- Michael Bailey, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebe. International trade and social connectedness. Working Paper 26960, National Bureau of Economic Research, 2020.
- Sascha O Becker, Thiemo Fetzer, and Dennis Novy. Who voted for Brexit? A comprehensive district-level analysis. Economic Policy, 32(92):601–650, 2017.
- Martin Bell, Elin Charles-Edwards, Philipp Ueffing, John Stillwell, Marek Kupiszewski, and Dorota Kupiszewska. Internal migration and development: Comparing migration intensities around the world. Population and Development Review, 41(1):33–58, 2015.
- Juan-Antonio Carrasco and Eric J. Miller. The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. Transportation Research Part A: Policy and Practice, 43(1):90–104, 2009.
- Italo Colantone and Piero Stanig. Global competition and brexit. American Political Science Review, 112(2):201–218, 2018.

- Sergio Correia, Paulo Guimaraes, and Tom Zylkin. Ppmlhdf: Fast poisson estimation with high-dimensional fixed effects, 2019.
- Maeve Duggan, Nicole B Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. Social media update 2014. Report, Pew Research Center, 2015.
- Neli Esipova, Anita Pugliese, and Julie Ray. The demographics of global internal migration. Migration Policy Practice, 3(2):3–5, 2013.
- European Comission. Special eurobarometer 386: Europeans and their languages. Report, 2012a.
- European Comission. Flash eurobarometer 472: Public opinion in the eu regions. Report, 2018b.
- European Commission. What is erasmus+? URL [https://ec.europa.eu/programmes/erasmus-plus/about\\_en](https://ec.europa.eu/programmes/erasmus-plus/about_en).
- European Social Survey. European Social Survey Cumulative File, ESS 1-8 (2018). Data file edition 1.0. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS-CUMULATIVE.
- European Union. The eu in brief: Goals and values of the eu. URL [https://europa.eu/european-union/about-eu/eu-in-brief\\_en#goals-and-values-of-the-eu](https://europa.eu/european-union/about-eu/eu-in-brief_en#goals-and-values-of-the-eu).
- Eurostat. Reference manual on rail transport statistics version 10.1. Technical report, 2019.
- Facebook. Facebook form 10-q, quarter 2, 2019. URL [https://s21.q4cdn.com/399680738/files/doc\\_financials/2019/FACEBOOK\\_INC\\_10Q\\_20190725.pdf](https://s21.q4cdn.com/399680738/files/doc_financials/2019/FACEBOOK_INC_10Q_20190725.pdf).
- George Harrison. Inside little bucharest: How romanian immigration changed this small london suburb beyond recognition. The Sun, 2018.
- Ronald Inglehart and Pippa Norris. Trump, brexit, and the rise of populism: Economic have-nots and cultural backlash. Working Paper RWP16-026, Harvard Kennedy School, August 2016.
- International Organization for Migration. World Migration Report 2018. 2017.
- Greg Kaplan and Sam Schulhofer-Wohl. Understanding the long-run decline in interstate migration. Working Paper 18507, National Bureau of Economic Research, November 2012.
- Faith Karahan and Darius Li. What caused the decline in interstate migration in the united states? In Liberty Street Economics (blog). Federal Reserve Bank of New York, 2016.
- Jinhee Kim, Soora Rasouli, and Harry J. P. Timmermans. Social networks, social influence and activity-travel behaviour: a review of models and empirical evidence. Transport Reviews, 38(4):499–523, 2018.
- Theresa Kuchler, Lin Peng, Johannes Stroebe, Yan Li, and Dexin Zhou. Social proximity to capital: Implications for investors and firms. Technical report, 2020a. Working paper.

- Theresa Kuchler, Dominic Russel, and Johannes Stroebel. The geographic spread of covid-19 correlates with structure of social networks as measured by facebook. Working Paper 26990, National Bureau of Economic Research, 2020b.
- Paul Lazarsfeld and Robert King Merton. Friendship as a social process: A substantive and methodological analysis. In Morroe Berger, Theodore Abel, and Charles H Page, editors, Freedom and Control in Modern Society, pages 18–66. 1954.
- Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet. The People’s Choice: How the Voter Makes up His Mind in a Presidential Campaign. Columbia University Press, 1944.
- David Marmaros and Bruce Sacerdote. How do friendships form? The Quarterly Journal of Economics, 121(1):79–119, 2006.
- Lauren M. McLaren. Anti-immigrant prejudice in europe: Contact, threat perception, and preferences for the exclusion of migrants. Social Forces, 81(3):909–936, 2003.
- Raven Molloy, Christopher L. Smith, and Abigail Wozniak. Job changing and the decline in long-distance migration in the united states. Demography, 54(2):631–653, Apr 2017.
- MPIDR and CCG. Max Planck Institute for Demographic Research and Chair for Geodesy and Geoinformatics, University of Rostock. MPIDR Population History GIS Collection – Europe (partly based on EuroGeographics for the administrative boundaries), 2013.
- Antonio Páez and Darren M Scott. Social influence on travel behavior: A simulation example of the decision to telecommute. Environment and Planning A: Economy and Space, 39(3):647–665, 2007.
- Jacob Poushter, Caldwell Bishop, and Hanyu Chwe. Social media use continues to rise in developing countries but plateaus across developed ones. Report, Pew Research Center, 2018.
- Kurt Sansone. Maltatoday survey: Maltese identity still very much rooted in catholicism. Report, 2018.
- Sascha Sardadvar and Silvia Rocha-Akis. Interregional migration within the european union in the aftermath of the eastern enlargements: a spatial approach. Review of Regional Research, 36(1):51–79, Feb 2016.
- Alexandra Schmidt. Why are croatians moving to ireland? Croatia Week, 2017.
- StatCounter. Social media stats europe, 2019. URL <https://gs.statcounter.com/social-media-stats/all/europe>.
- Bogdan State, Patrick Park, Ingmar Weber, and Michael Macy. The mesh of civilizations in the global network of digital communication. PloS one, 10(5):e0122543, 2015.
- Mark Thomas. Number of croatians moving to ireland increase tenfold. The Dubrovnik Times, 2019.
- Lois M. Verbugge. A research note on adult friendship contact: A dyadic perspective. Social Forces, 62(1):78–83, 1983.

We Are Social and Hootsuite. Digital in 2018. Report, 2018. URL <https://wearesocial.com/blog/2018/01/global-digital-report-2018>.

World Bank. Romania systematic country diagnostic: Migration background note. Technical report, World Bank Group, 2018.

Danny Yagan. Moving to opportunity? migratory insurance over the great recession. Job market paper, 2014.

George Kingsley Zipf. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley Press, 1949.



# Appendices

## A Additional Case Studies of European Social Connectedness

**Continental Croatia.** Figure 6 shows the connectedness of Continental Croatia, one of Croatia's two NUTS2 regions. The connections of the region are strongest with other countries that were part of Yugoslavia—including Slovenia, Serbia, Montenegro, and North Macedonia—suggestive of the importance of past country borders in shaping connectedness. There are also strong connections between Croatia and Ireland, likely related to a pattern of migration since Croatia's 2013 accession into the European Union (see, for example, Thomas, 2019; Schmidt, 2017). Similarly, there are connections to central European countries, which are geographically close and have also welcomed Croatian migrants.

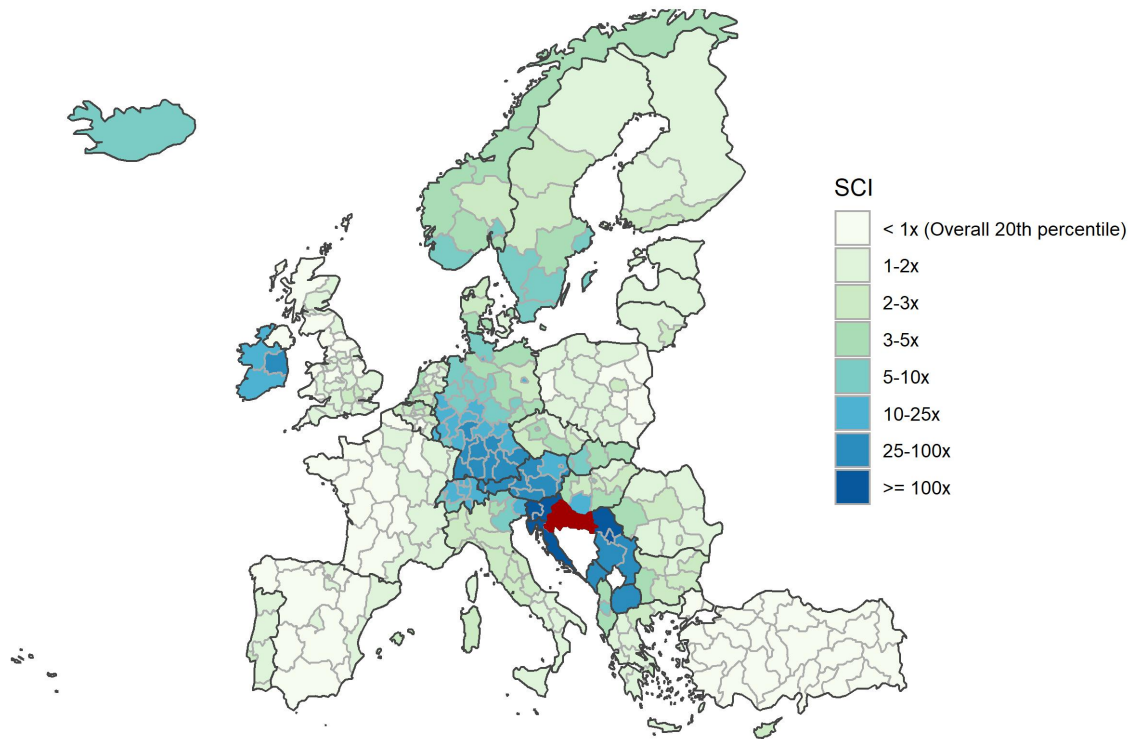
**Lake Geneva and Central Switzerland.** Figure 7 shows the social connectedness of two neighboring regions in Switzerland, the Lake Geneva Region (Panel A) and Central Switzerland (Panel B). The Lake Geneva Region includes the Swiss cantons of Geneva, Vaud, and Valais, all of which are primarily French speaking. Central Switzerland includes the cantons of Lucerne, Uri, Schwyz, Obwalden, Nidwalden, and Zug, all of which are primarily German speaking. Accordingly, the connections of the Lake Geneva Region are strong throughout all of France to the west, whereas the connections of Central Switzerland are stronger throughout all of Germany to the north. These patterns provide more evidence of language playing an important role in shaping social connectedness.

**Aydıń and Ağrı.** Figure 8 shows the connections of two regions in Turkey, the Aydıń Subregion (Panel A) and the Ağrı Subregion (Panel B). Aydıń's connections to Germany appear similar to the Samsun Subregion in Figure 1, with connectedness dropping-off at the former East German border. This pattern can be seen in most Turkish regions, likely reflecting West Germany's 1961-1973 labor recruitment agreement with Turkey. However, there is no similar drop-off for the Ağrı Subregion in Panel B, which shows relatively strong connections to all of Germany. The Ağrı Subregion is home to a sizeable Kurdish population, including the largely Kurdish city of Diyarbakır. Armed conflict between the Kurdish Workers' Party and the Turkish government in the 1990s led many Kurds to seek refuge in Germany (Aydın, 2016). Much of this migration happened after the reunification of East and West Germany, likely explaining the Ağrı's connectedness to the entire country.

**Picardy and Île-de-France.** Figure 9 shows the connections of two neighboring regions in France, Picardy (Panel A) and Île-de-France (Panel B). Picardy is most strongly connected to other regions in France, with connectedness dropping off sharply at country borders. Indeed, only 6.2% of the region's European connections are to individuals located outside of France. By contrast, Île-de-France, which includes Paris and its metropolitan area, shows fairly strong connections to a number of regions outside of France—17.3% of the region's European connections are international. In particular, Île-de-France's connections to regions with other capital cities, including Madrid, Berlin, Prague, Warsaw, Oslo, and Stockholm, appear comparatively much stronger. This is an example of the differences in international

connectedness between European regions.

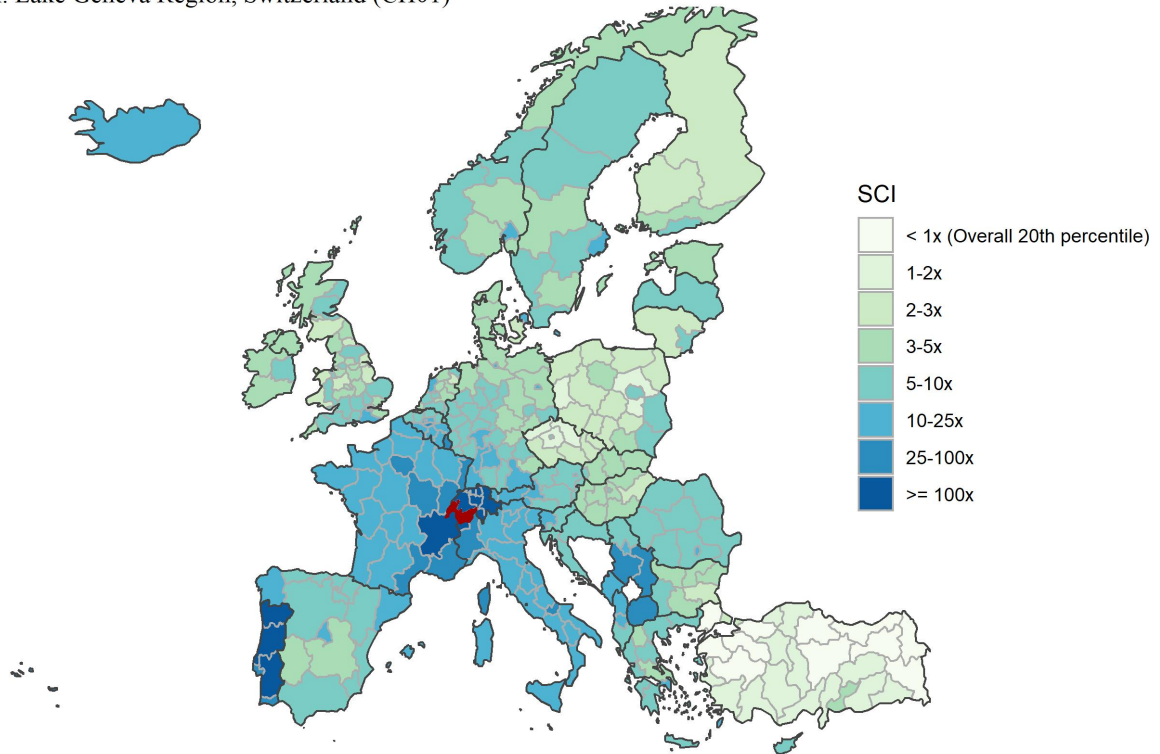
**Figure 6: Social Network Distributions in Continental Croatia (HR04)**



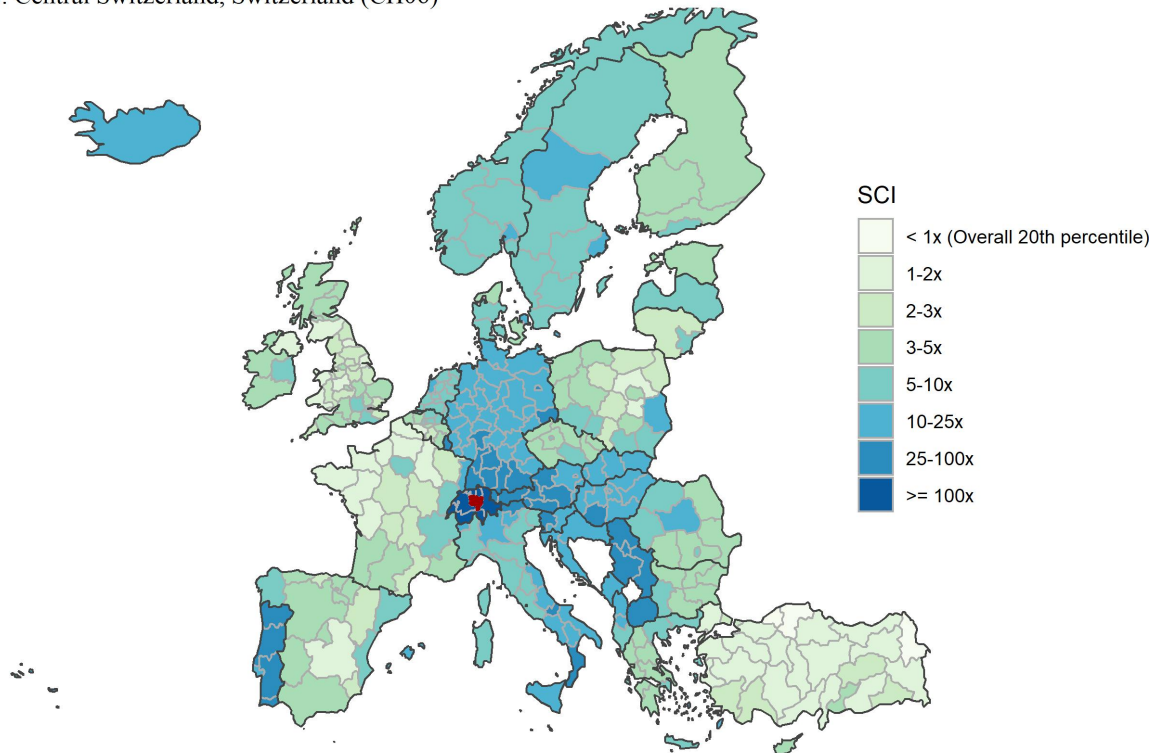
*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions with Continental Croatia. The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.

**Figure 7: Social Network Distributions in Switzerland**

A: Lake Geneva Region, Switzerland (CH01)



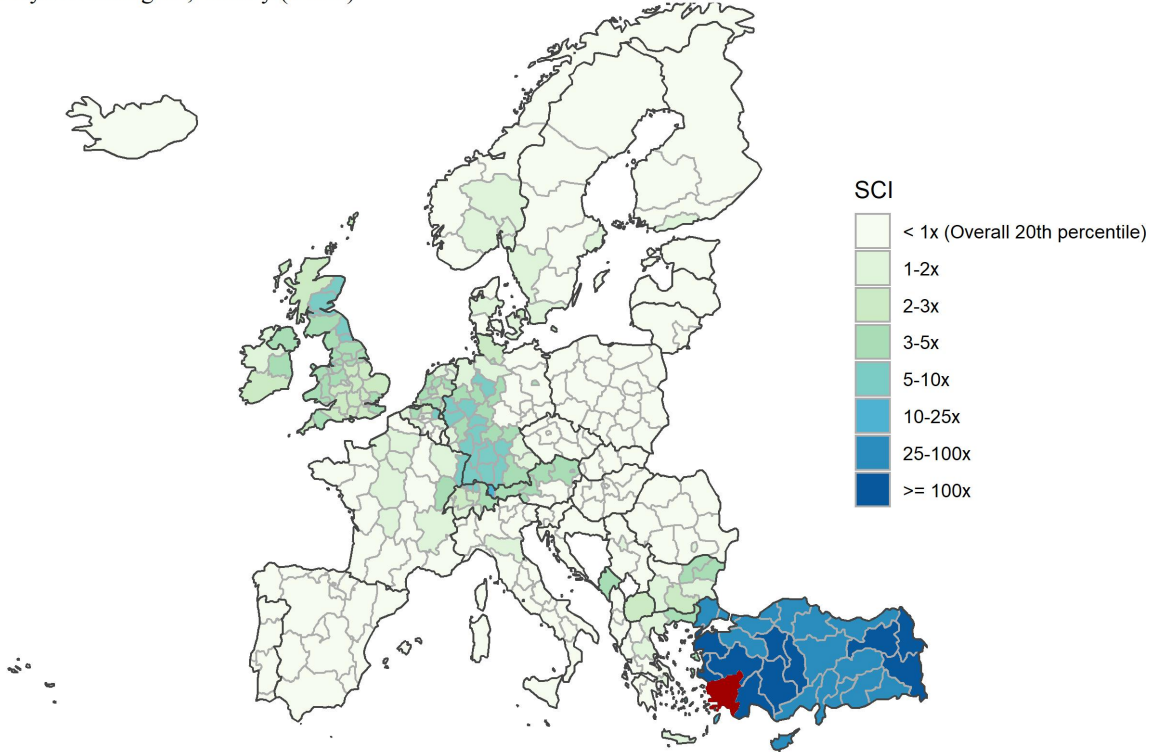
B: Central Switzerland, Switzerland (CH06)



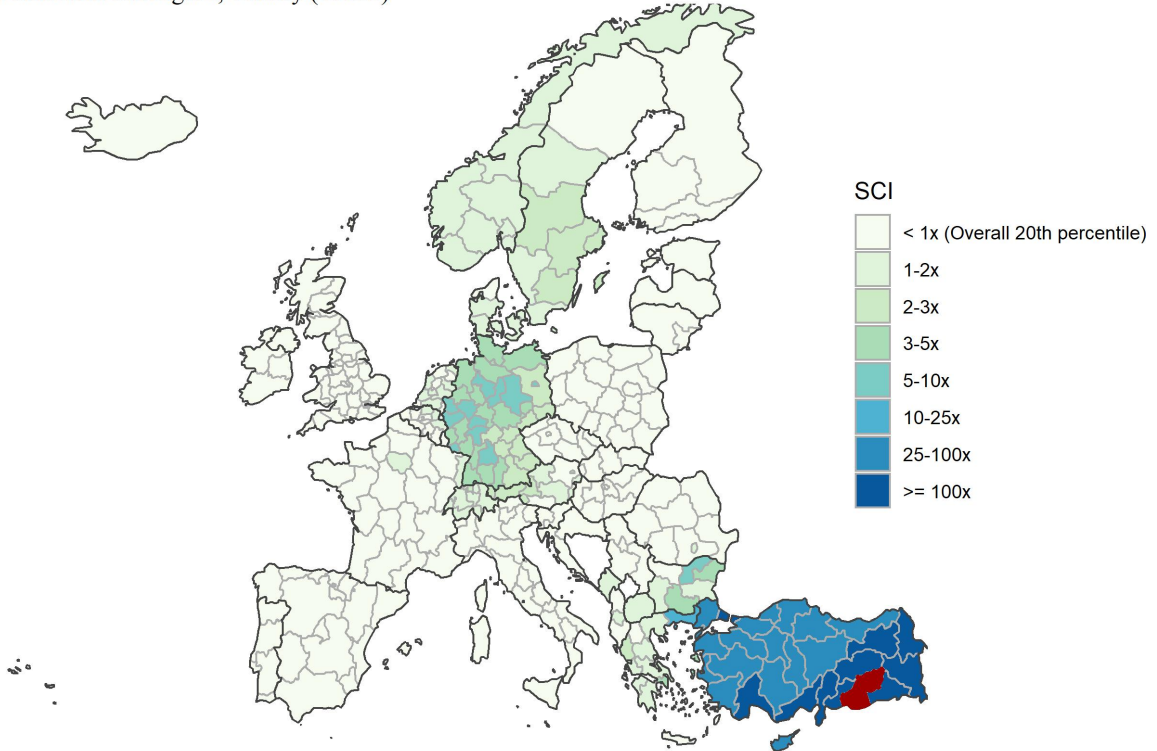
*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions  $j$  with two regions  $i$ : Lake Geneva Region, CH (Panel A) and Central Switzerland, CH (Panel B). The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.

**Figure 8: Social Network Distributions in Turkey**

A: Aydın Subregion, Turkey (TR23)



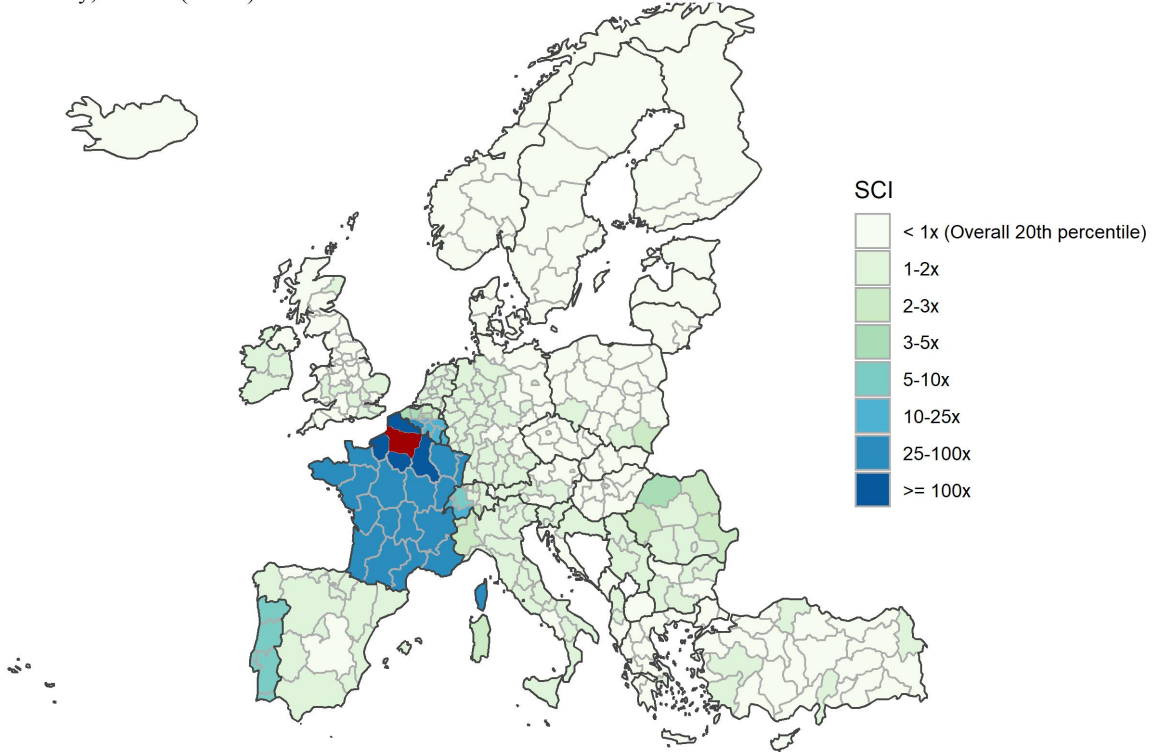
B: Sanliurfa Subregion, Turkey (TRC2)



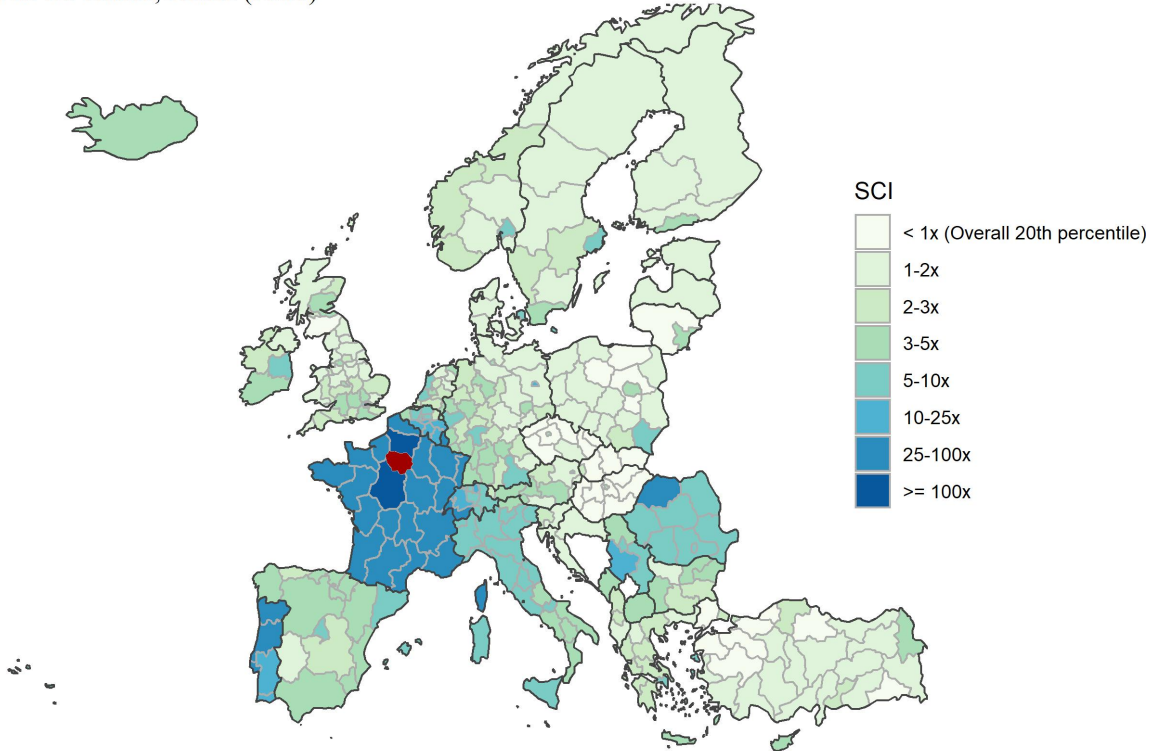
*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions  $j$  with two regions  $i$ : the Aydın Subregion, TR (Panel A) and the Sanliurfa Subregion, TR (Panel B). The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.

**Figure 9: Social Network Distributions in France**

A: Picardy, France (FRE2)



B: Ile-De-France, France (FR10)



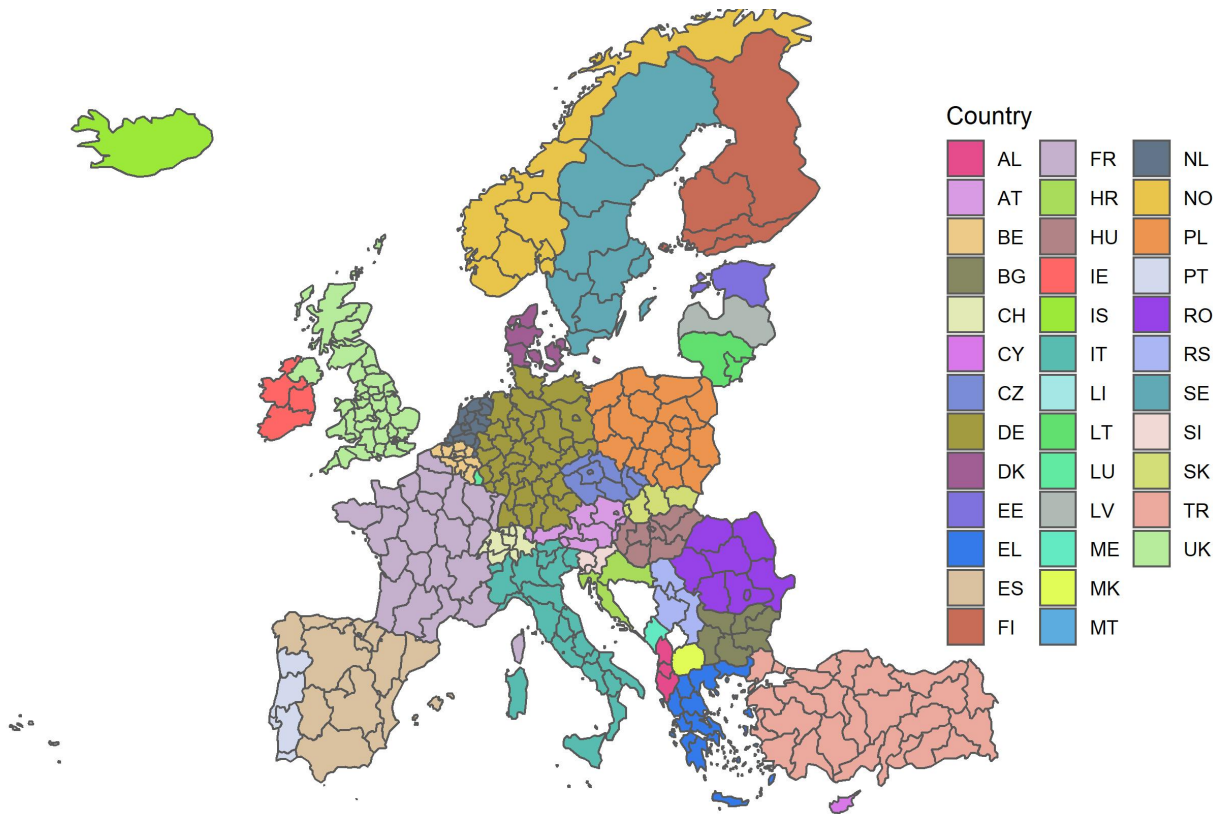
*Note:* Figure shows the relative probability of connection, measured by  $SocialConnectedness_{ij}$ , of all European regions  $j$  with two regions  $i$ : Picardy, FR (Panel A) and Île-de-France (Panel B). The measures are scaled from the 20th percentile of all  $i, j$  pairs in Europe. Darker regions have a higher probability of connection.



# B Historical Europe Maps

The analyses that produce Table 3 use information on the country that each modern NUTS2 region was a part of in the years 1900, 1930, 1960, and 1990. The maps in this appendix show these country classifications for each year. The data largely come from files provided by the Max Planck Institute for Demographic Research Population History GIS Collection (MPIDR and CGG). In cases when a modern region spans two historical countries, we classify the region as part of the country for which it had a greater land area overlap.

**Figure 10: 2019 countries of Europe**

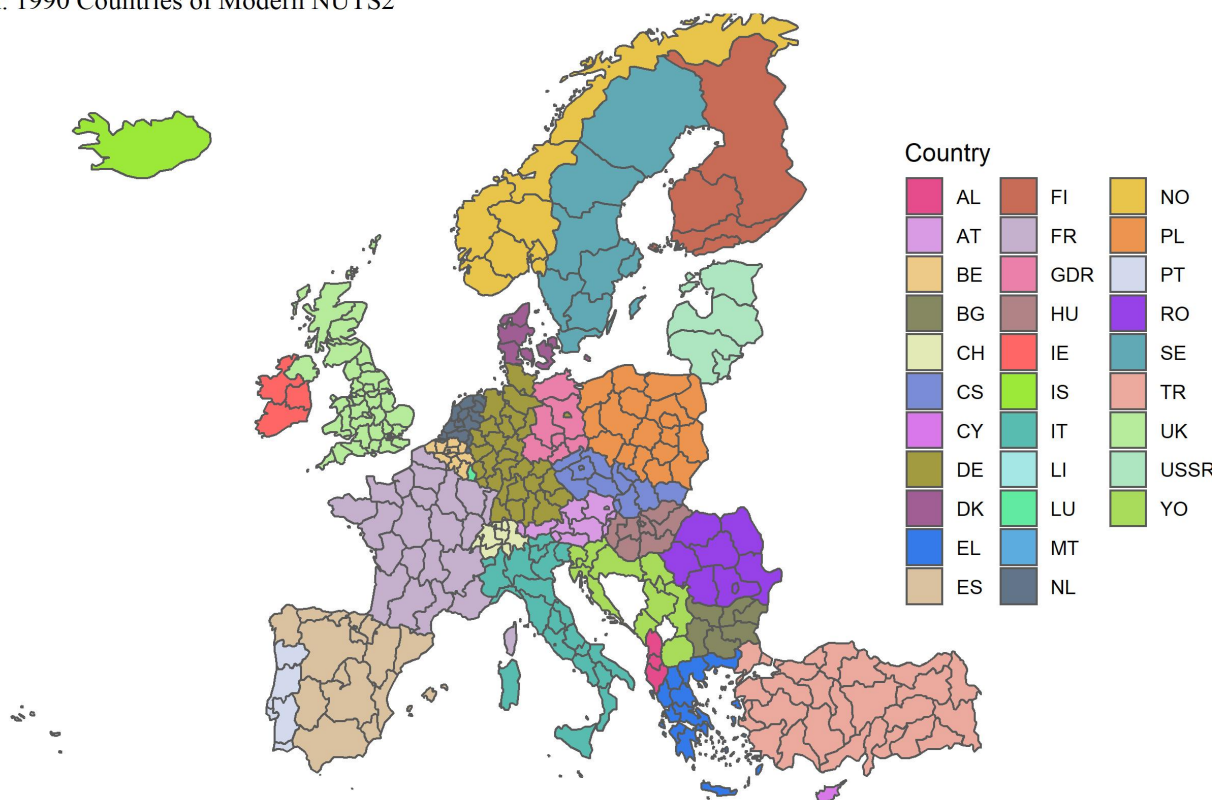


*Note:* Figure shows the 2019 country of each present-day NUTS2 region.

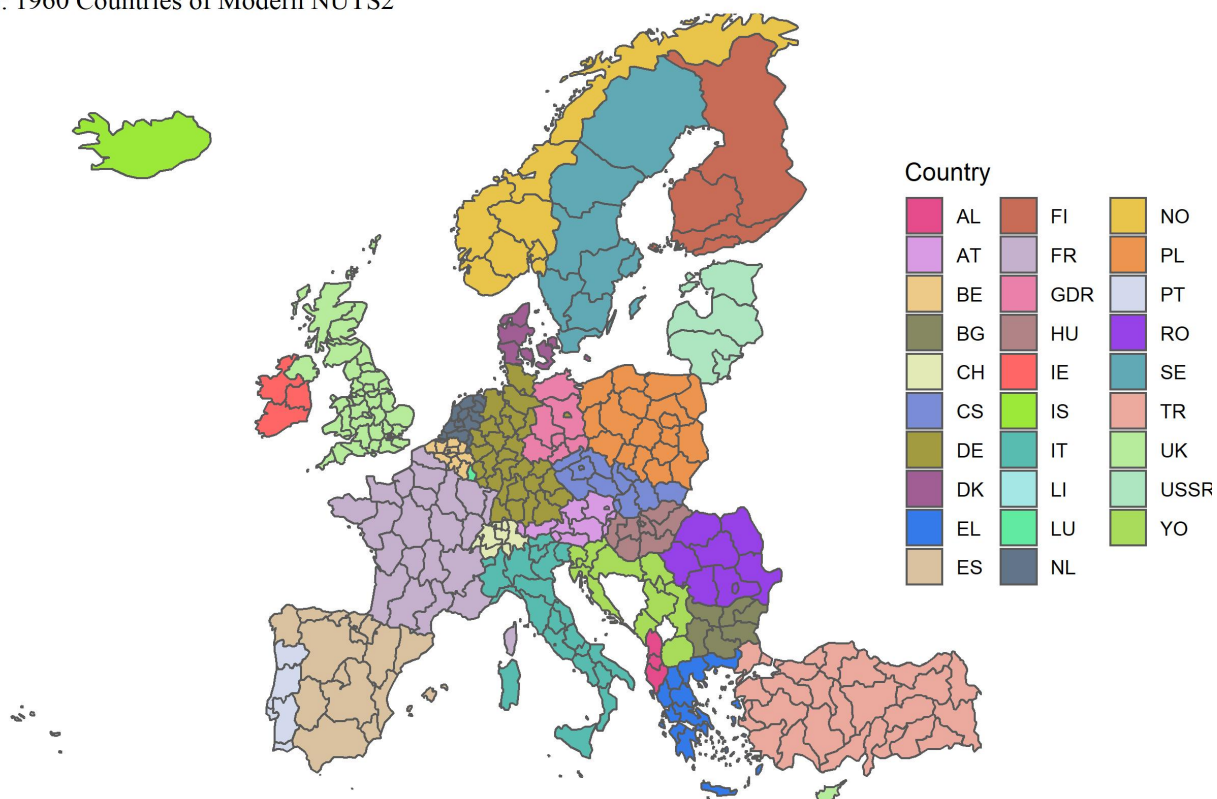


**Figure 11: 1990 and 1960 countries of Europe**

A: 1990 Countries of Modern NUTS2



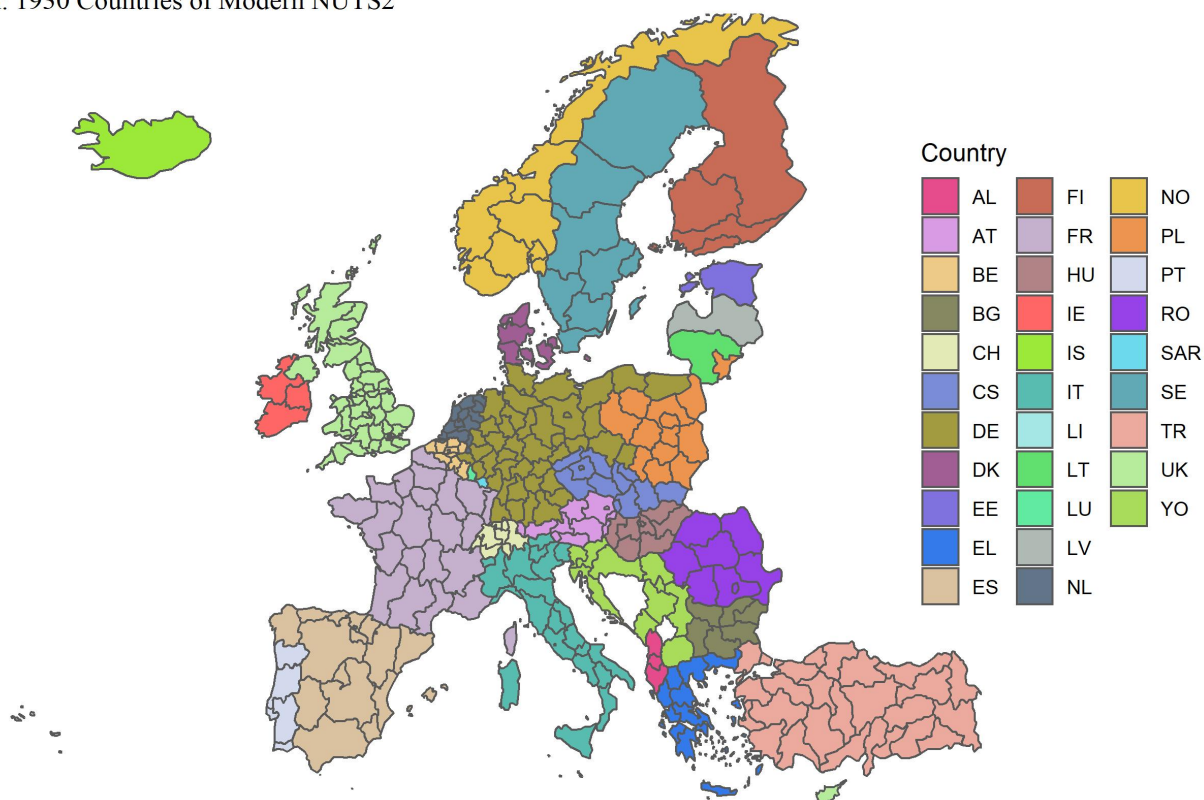
B: 1960 Countries of Modern NUTS2



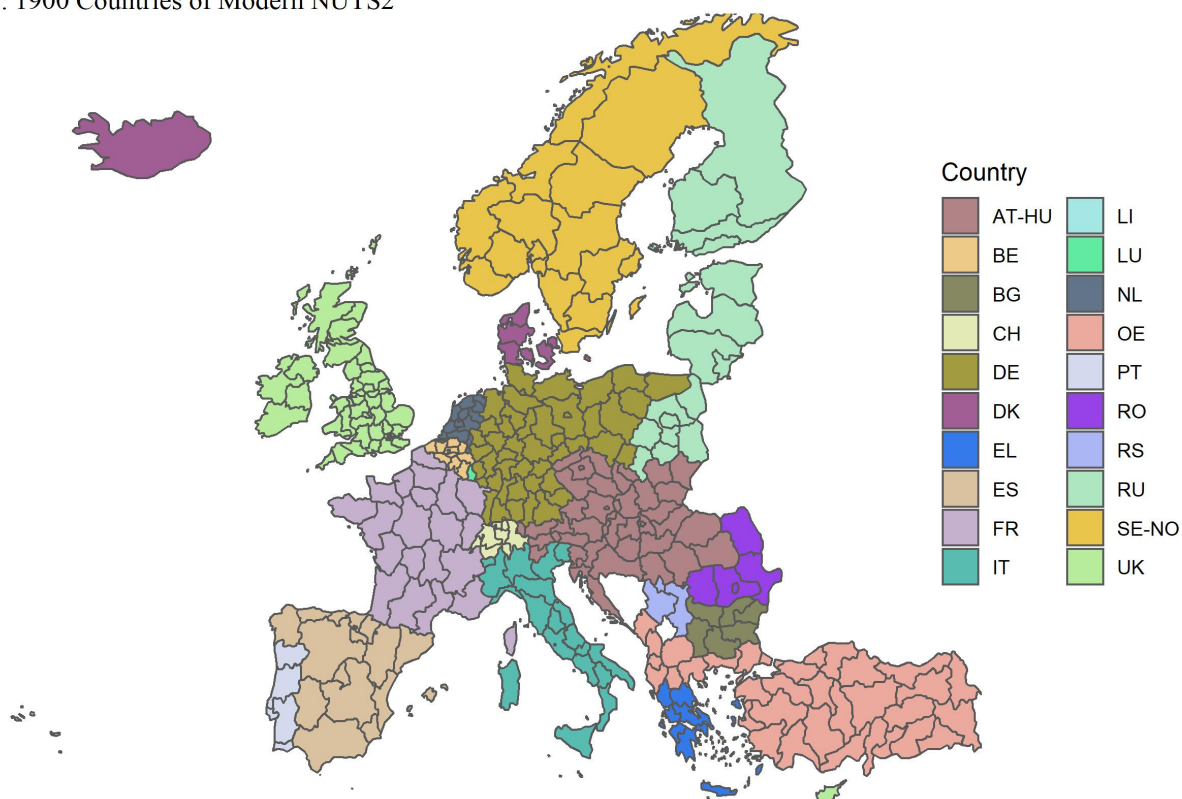
*Note:* Figure shows the 1990 country (Panel A) and 1960 country (Panel B) of each present-day NUTS2 region.

**Figure 12: 1930 and 1900 countries of Europe**

A: 1930 Countries of Modern NUTS2



B: 1900 Countries of Modern NUTS2



*Note:* Figure shows the 1930 country (Panel A) and 1900 country (Panel B) of each present-day NUTS2 region.

## C Passenger Train Travel Data

Our analyses in Section 4 use information on region-to-region passenger train travel made available by Eurostat in the series “tran\_r\_rapa.” The data are based on individual reports from Member States of the European Union or European Free Trade Association, as well as European Union Candidate Countries. For each observation, the reporting country is identified. The data are reported for 2005, 2010, and 2015, in accordance with Regulation 91/2003 of the European Commission and subsequent regulatory updates. We take a number of steps to clean the data to prepare it for our analyses. This process was informed by both the “Reference Manual on Rail transport statistics” (Eurostat) and correspondence with the Eurostat data providers.

We first restrict our data to observations at the NUTS2 level, removing any country-level observations (we do, however, keep country-level data for countries which have only a single NUTS2 region). We also exclude all pairs that include a region with the unknown indicator “XX” or the extraregio territory indicator “ZZ.” These pairs make up around 1.3% of passenger trips in the data. 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 traffic or that the relevant country did not provide the data.<sup>15</sup> 2) There are a number of hypothetical regional pairs missing, even between countries that did report data elsewhere. 3) For some international regional pairs, there are data reported from both countries on the same train flows, and often the number of passengers does not match. 4) NUTS2 classifications changed in 2006, 2010, and 2013. Each year of data is reported using the NUTS2 classification that was relevant at that particular point in time.

With respect to challenges 1 and 2, we found that each country reports data to Eurostat in two intermediate data sets: one for domestic passenger travel and another for international passenger travel. 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 6. When “non-available” values are reported by a country that *did not* report data elsewhere in the year/domestic-international group, we treat this as missing and exclude it. When “non-available” values are reported by a country that *did* report data elsewhere in the group, we treat this value as a zero (no traffic). Additionally, for countries that reported data in a particular group, we fill any missing regional pairs (i.e. pairs that are not in the data) in the group with zeros. Together, these assumptions handle challenges 1 and 2.

For each international regional pair, there still remains two possible reports: one from each of the regions’ home countries in the pair. In instances when only one country reports the data, we take the non-missing value from the reporting country. However, there are a number of instances when each country reports data for the same international regional pair (challenge 3). In these instances, we take the average of the two reports. 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.<sup>16</sup> In instances when an

---

<sup>15</sup>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.

<sup>16</sup>Available at: <https://ec.europa.eu/eurostat/web/nuts/history>

older NUTS2 region split into multiple new regions, we set the number of passengers in each row that includes a new region equal to the corresponding old row's number of passengers, multiplied by the new region's population share of the old region population (i.e. we assume that passenger train travel in each of these regions is proportional to population).

**Table 6:** Passenger train travel data availability

reporter	domestic2015	international2015	domestic2010	international2010	domestic2005	international2005
AL	0	0	0	0	0	0
AT	0	0	0	0	0	0
BE	0	0	1	1	1	1
BG	1	1	1	1	0	0
CH	1	1	1	1	0	0
CY	0	0	0	0	0	0
CZ	1	1	1	1	1	1
DE	1	0	1	0	1	1
DK	0	0	0	0	1	1
EE	1	0	1	0	1	0
EL	0	0	1	0	1	0
ES	1	0	1	0	1	0
FI	1	0	1	0	1	0
FR	0	0	0	0	0	0
HR	1	1	1	1	0	0
HU	0	0	1	1	1	1
IE	1	1	1	1	1	1
IS	0	0	0	0	0	0
IT	1	0	0	0	1	0
LI	0	0	0	0	0	0
LT	1	1	1	1	1	1
LU	1	1	1	1	1	1
LV	1	1	1	1	1	1
ME	0	0	0	0	0	0
MK	0	0	0	0	0	0
MT	0	0	0	0	0	0
NL	0	0	0	0	1	1
NO	1	1	1	0	1	0
PL	1	1	1	1	1	1
PT	1	0	1	0	1	0
RO	0	0	0	0	0	0
RS	0	0	0	0	0	0
SE	0	0	0	0	0	0
SI	1	1	1	1	1	1
SK	1	1	1	1	1	1
TR	1	0	1	0	0	0
UK	0	0	0	0	0	1

*Note:* Table shows the regional passenger train travel data availability by reporting country, year, and whether the travel is domestic or international. 0s indicate the data were not available and 1s indicate the data were available. Reporting country is given by the two-letter prefix of each country's NUTS codes. The table only shows whether any data from a particular *reporter* were available, not whether any regions from this country are included in the final analysis. For example, although Austria did not report international data in 2015, pairs that include an Austrian region and a region in a country that did report international data in 2015 are included.