

Social Connectedness in Europe

Michael Bailey (Facebook)

Drew Johnston (Harvard)

Theresa Kuchler (NYU Stern, CEPR)

Dominic Russel (NYU Stern)

Bogdan State (Facebook)

Johannes Stroebel (NYU Stern, CEPR, NBER)

October 8, 2020

Motivation

- Social networks shape important aspects of European society
- Understanding what factors shape these networks informative for a wide range of social science questions
- **Challenge:** The geographic structure of social networks is difficult to measure on a national or global scale
- **Solution:** Aggregated measure of connections between region pairs from de-identified Facebook social graph
 - Facebook global social network = 394 million active users in Europe
 - Limit on friends & required consent of both parties → more likely to capture real-world connections than other online networks

Social Connectedness Index

- Social Connectedness Index (Bailey et al., 2018)

$$\text{Social Connectedness}_{i,j} = \frac{\text{FB Connections}_{i,j}}{\text{FB Users}_i * \text{FB Users}_j}$$

- Normalized number of Facebook friendship links between regions
- Captures relative probability of friendship between Facebook users in regions i and j
- Focus on NUTS2 regions (standardized data collection)
- Data are widely available to other researchers!

<https://data.humdata.org/dataset/social-connectedness-index>

Two-step approach to understanding the factors that shape European social connectedness:

1. Unsupervised exploratory analyses

- Case studies
- Generate socially connected communities

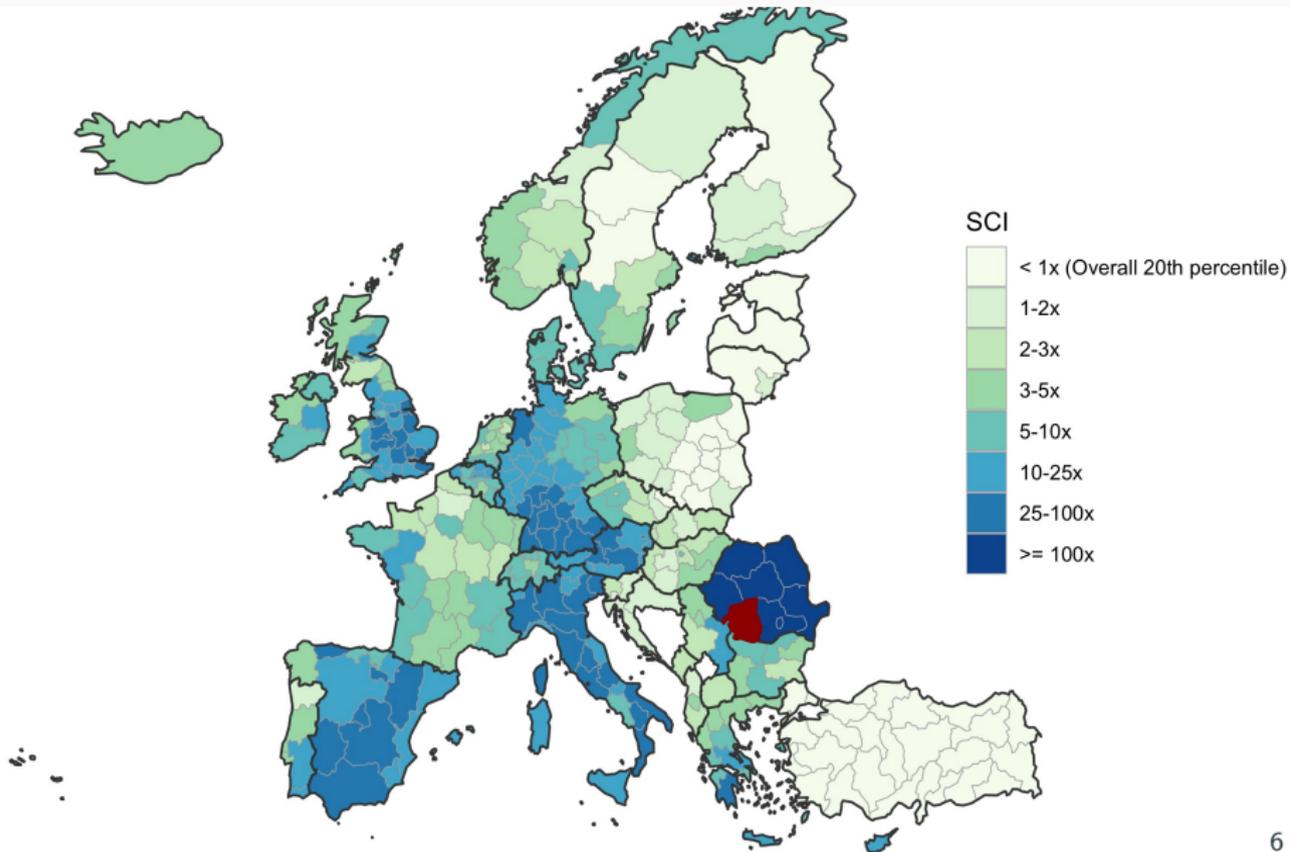
2. Regression analyses

- Test potential factors found in exploration

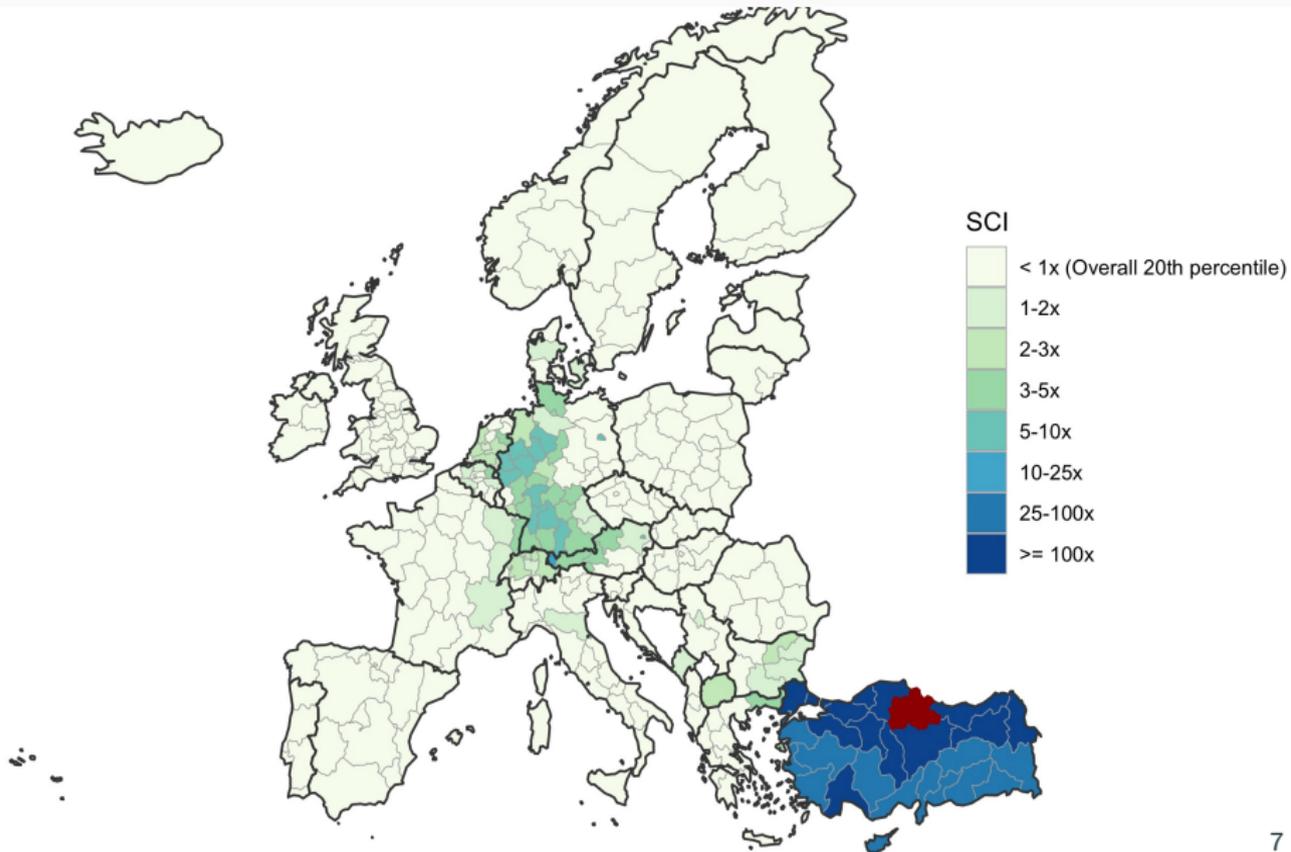
Case Study 1/2: Romania & Turkey

- South-West Oltenia, Romania
 - Romania joined EU in 2007
 - 3-5 million Romanians live & work abroad (1/5 of country's pop.)
 - Top destinations = Italy, Spain, Germany, US, and UK
- Samsun Subregion, Turkey
 - Turkey not an EU member state
 - Turkey + West Germany 1961-1973 labor recruitment agreement (*Anwerbeabkommen*) → many workers re-settling

Case Study 1/2: South-West Oltenia, Romania (RO41)



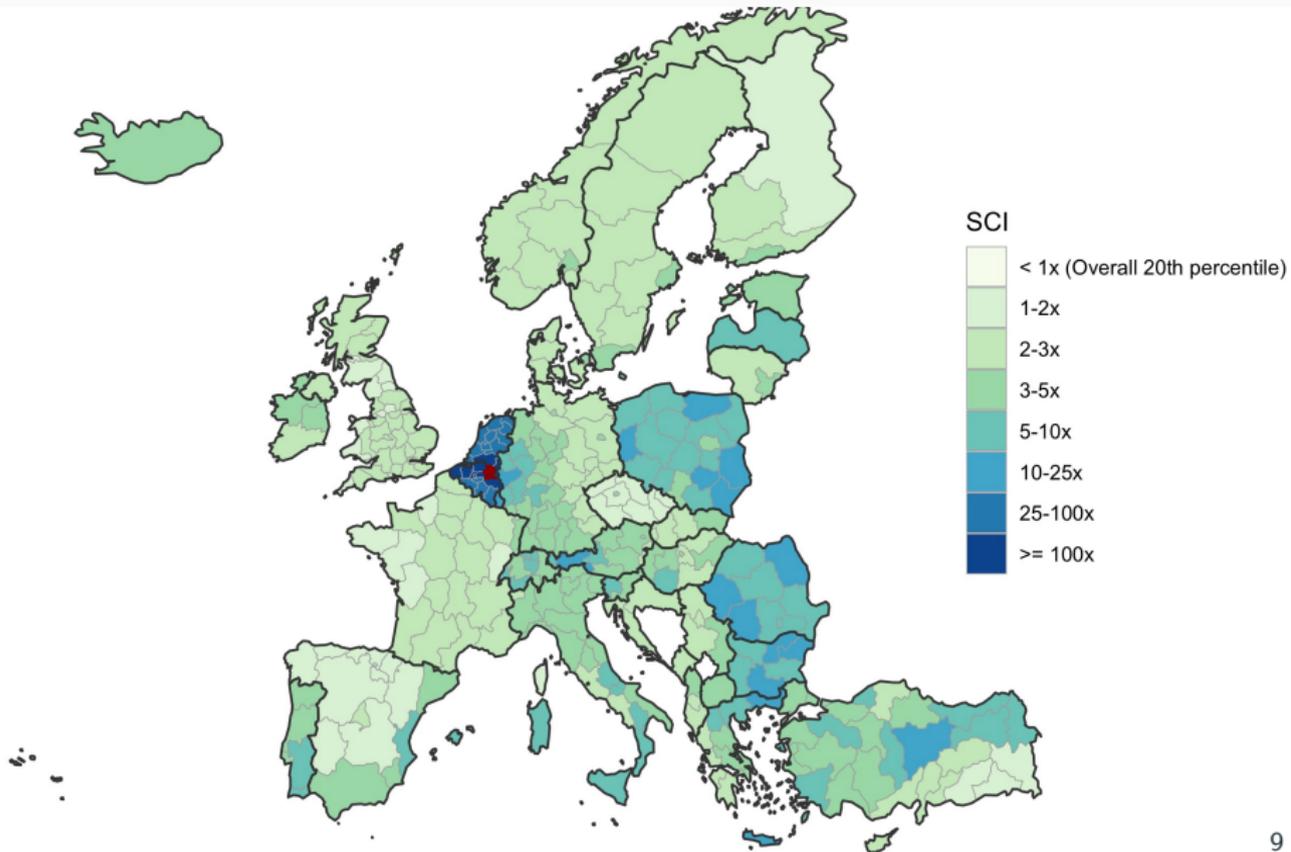
Case Study 1/2: Samsun Subregion, Turkey (TR83)



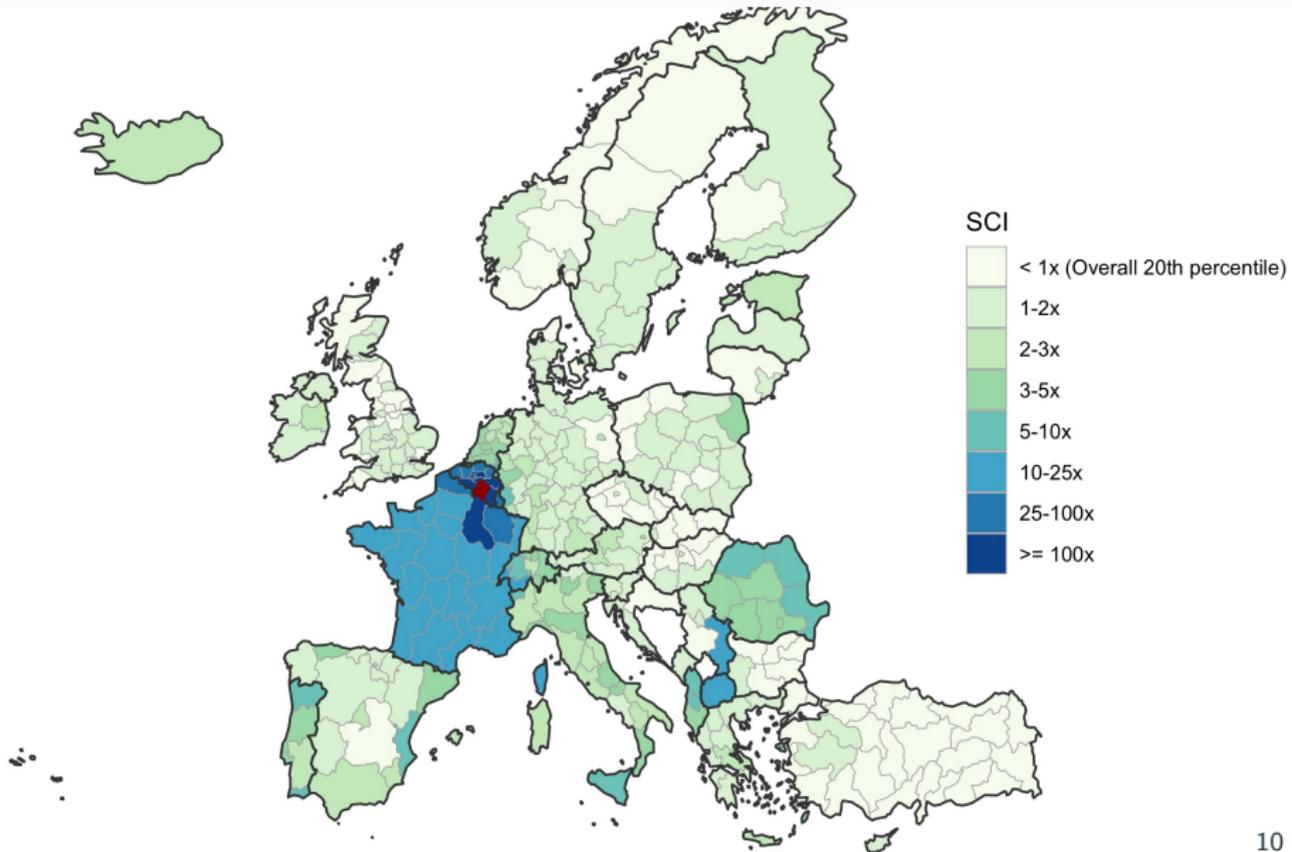
Case Study 2/2: Belgium

- Limburg, Belgium: Official language = Dutch
- Namur, Belgium: Official language = French
- Capitals of two regions (Hasslet and Namur) less than 70km apart

Case Study 2/2: Limburg, Belgium (BE22)



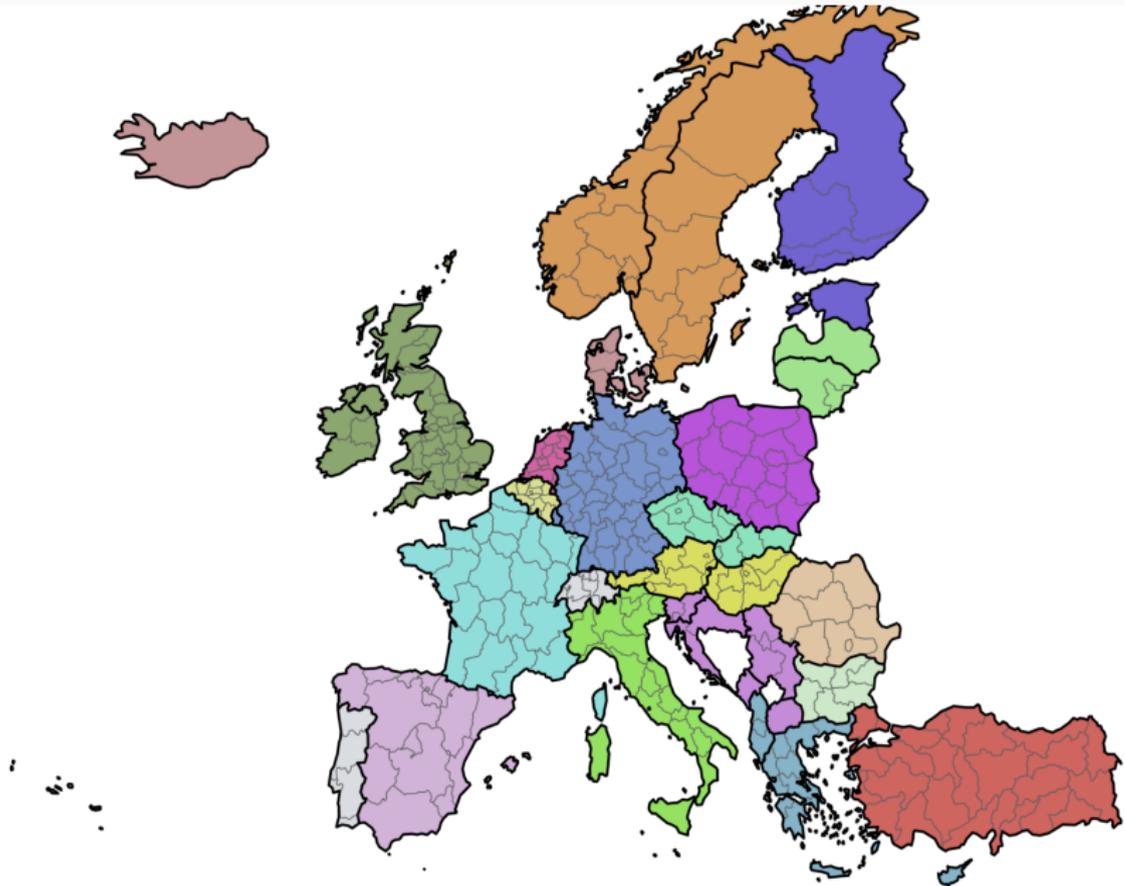
Case Study 2/2: Namur, Belgium (BE35)



Clustering Communities by Connectedness

- Another exploratory approach: generate communities with strong intra-community connections
- Do this by maximizing within-community pairwise SCI
- Use simple hierarchical agglomerative clustering (dist. = $1/SCI$)

20 Socially Connected Communities



50 Socially Connected Communities



Socially Connected Communities

- 20 Communities
 - Line-up very well with country borders
 - Only one non-contiguous cluster: Outer London West w/ Romania
 - Includes Burnt Oak, large Romanian immigrant community
 - Cross-country connections line-up with historical borders: Yugoslavia; Czechoslovakia; UK & Ireland; Denmark & Iceland
- 50 Communities
 - Sub-national linguistic communities
 - Belgium into French and Dutch speaking
 - Spain into Catalan and Andalusian speaking
 - Well-heeled residents in Ile-de-France & French Riviera
 - Czechoslovakia and Yugoslavia remain; East & West Germany split

Determinants of Social Connectedness: Regression Framework

- Exploration suggests importance of: migration, political borders (past & present), geographic distance, language, other demographics
- Next, look systematically using regression framework

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

- d_{ij} = Geographic distance
- X_{ij} = Similarity across demographics; same/border country
- ψ_i, ψ_j = FEs by region (control for regional FB usage)

Determinants of Social Connectedness

	Dependent Variable: log(SocialConnectedness)		
	(1)	(2)	(3)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)
Same Country		2.896*** (0.077)	1.651*** (0.124)
Border Country			0.285*** (0.044)
Δ Share Pop Low Edu (%)			-0.013*** (0.002)
Δ Median Age			-0.017*** (0.004)
Δ Avg Income (k €)			0.053*** (0.003)
Δ Unemployment (%)			-0.000 (0.005)
Same Religion			0.027 (0.031)
Same Language			1.493*** (0.097)
Industry Similarity			0.128 (0.169)
NUTS2 FEs	Y	Y	Y
R^2	0.490	0.669	0.745
Number of Observations	75,900	75,900	75,900

Determinants of Social Connectedness

Dependent Variable: log(SocialConnectedness)			
	(1)	(2)	(3)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)
Same Country		2.896*** (0.077)	1.651*** (0.124)
Border Country			0.285*** (0.044)
Δ Share Pop Low Edu (%)			-0.013*** (0.002)
Δ Median Age			-0.017*** (0.004)
Δ Avg Income (k €)			0.053*** (0.003)
Δ Unemployment (%)			-0.000 (0.005)
Same Religion			0.027 (0.031)
Same Language			1.493*** (0.097)
Industry Similarity			0.128 (0.169)
NUTS2 FEs	Y	Y	Y
R ²	0.490	0.669	0.745
Number of Observations	75,900	75,900	75,900

- Takeaway 1: ↑ Geographic distance → ↓ Connectedness

Determinants of Social Connectedness

	Dependent Variable: log(SocialConnectedness)		
	(1)	(2)	(3)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)
Same Country		2.896*** (0.077)	1.651*** (0.124)
Border Country			0.285*** (0.044)
Δ Share Pop Low Edu (%)			-0.013*** (0.002)
Δ Median Age			-0.017*** (0.004)
Δ Avg Income (k €)			0.053*** (0.003)
Δ Unemployment (%)			-0.000 (0.005)
Same Religion			0.027 (0.031)
Same Language			1.493*** (0.097)
Industry Similarity			0.128 (0.169)
NUTS2 FEs	Y	Y	Y
R ²	0.490	0.669	0.745
Number of Observations	75,900	75,900	75,900

- Takeaway 2: Country borders predict connectedness (above dist.)

Determinants of Social Connectedness

	Dependent Variable: log(SocialConnectedness)		
	(1)	(2)	(3)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)
Same Country		2.896*** (0.077)	1.651*** (0.124)
Border Country			0.285*** (0.044)
Δ Share Pop Low Edu (%)			-0.013*** (0.002)
Δ Median Age			-0.017*** (0.004)
Δ Avg Income (k €)			0.053*** (0.003)
Δ Unemployment (%)			-0.000 (0.005)
Same Religion			0.027 (0.031)
Same Language			1.493*** (0.097)
Industry Similarity			0.128 (0.169)
NUTS2 FEs	Y	Y	Y
R ²	0.490	0.669	0.745
Number of Observations	75,900	75,900	75,900

- Takeaway 3: Homophily by edu, age, religion, language, industry

Determinants of Social Connectedness

	Dependent Variable: log(SocialConnectedness)		
	(1)	(2)	(3)
log(Distance in KM)	-1.318*** (0.046)	-0.558*** (0.053)	-0.582*** (0.041)
Same Country		2.896*** (0.077)	1.651*** (0.124)
Border Country			0.285*** (0.044)
Δ Share Pop Low Edu (%)			-0.013*** (0.002)
Δ Median Age			-0.017*** (0.004)
Δ Avg Income (k €)			0.053*** (0.003)
Δ Unemployment (%)			-0.000 (0.005)
Same Religion			0.027 (0.031)
Same Language			1.493*** (0.097)
Industry Similarity			0.128 (0.169)
NUTS2 FEs	Y	Y	Y
R ²	0.490	0.669	0.745
Number of Observations	75,900	75,900	75,900

- Takeaway 4: ↓ Income/employment similarity → ↑ Connectedness
 - Somewhat surprising (due to migration?)

Historical Determinants of Social Connectedness

- In exploratory analyses, historical borders seemed important in shaping *present* connections
- A regression approach:
 - Map present NUTS2 region to past borders (using max area overlap)
 - Add 10 major historical European border changes to our regression
- As with previous results, somewhat hard to pin-down causality

Historical Determinants of Social Connectedness

- In exploratory analyses, historical borders seemed important in shaping *present* connections
- A regression approach:
 - Map present NUTS2 region to past borders (using max area overlap)
 - Add 10 major historical European border changes to our regression
- As with previous results, somewhat hard to pin-down causality
- **Takeaway:** Historical borders — even back to early 1900s! — correlate w/ modern patterns of connectedness (full results in paper)

Conclusion

- *Social Connectedness Index* = unique measure to overcome measurement challenges in many social science applications
- Geographic distance, political borders (past & present), migration, and homophily shape patterns of European social connectedness
- Online Appendix explores some *effects* of European connectedness
 - Pairwise-connectedness predicts train travel flows
 - Share of connections within country predicts anti-EU sentiment
- Many opportunities for future research
 - Data available for U.S. counties; Europe NUTS3; GADM1 or GADM2 in much of the rest of the world

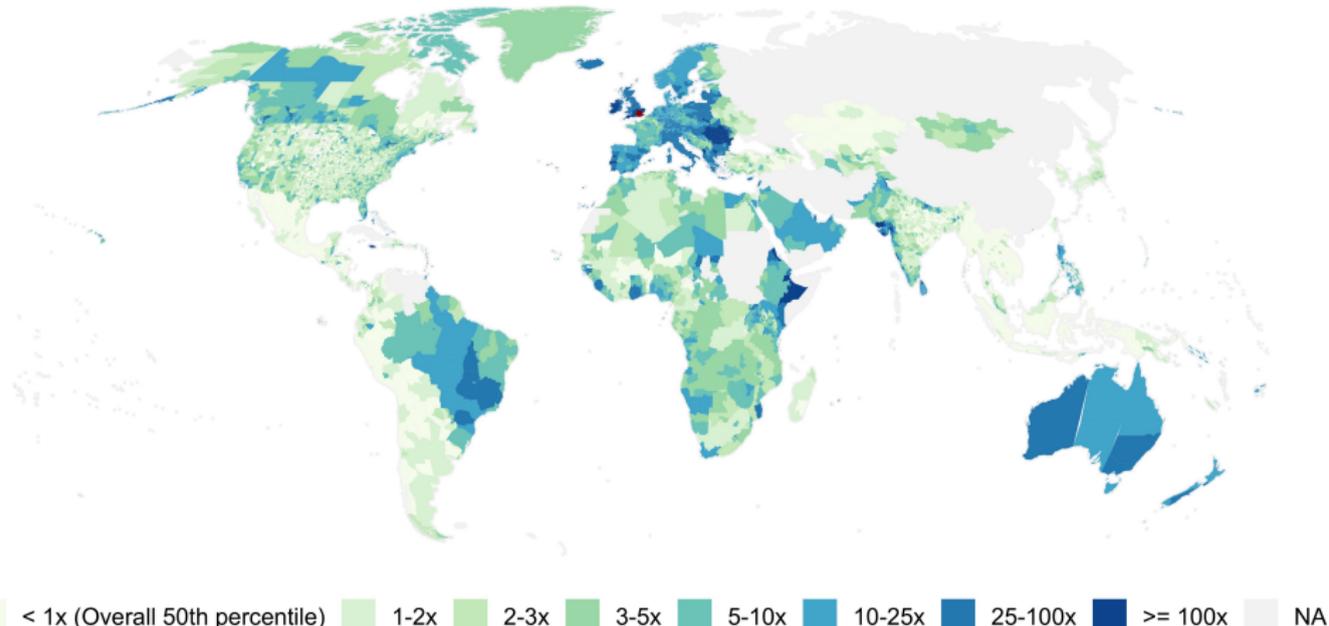
Data: <https://data.humdata.org/dataset/social-connectedness-index>

Example Code: <https://github.com/social-connectedness-index>

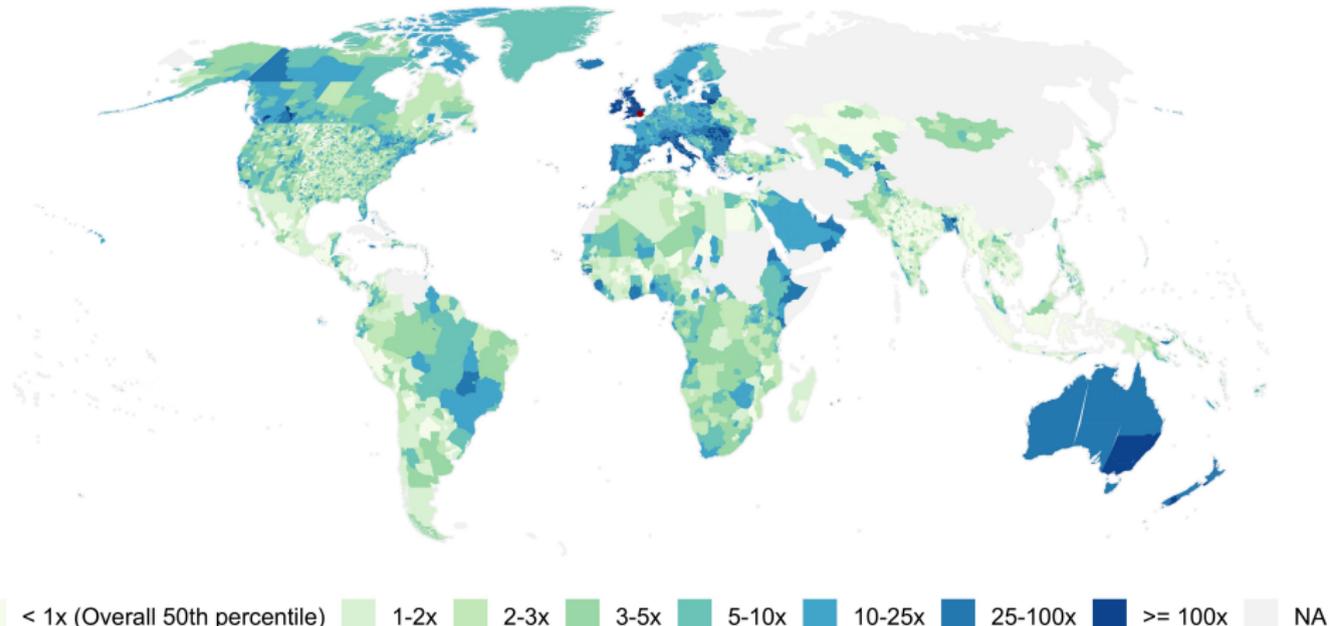
Other SCI examples: London

- Brent, London
 - Includes West Oak, a densely Romanian immigrant community
 - Some of largest communities from Brazil and Western India in UK
- Tower Hamlets, London
 - Largest Bangladeshi community in UK (32% of borough)

Brent, London to the World



Tower Hamlets, London to the World

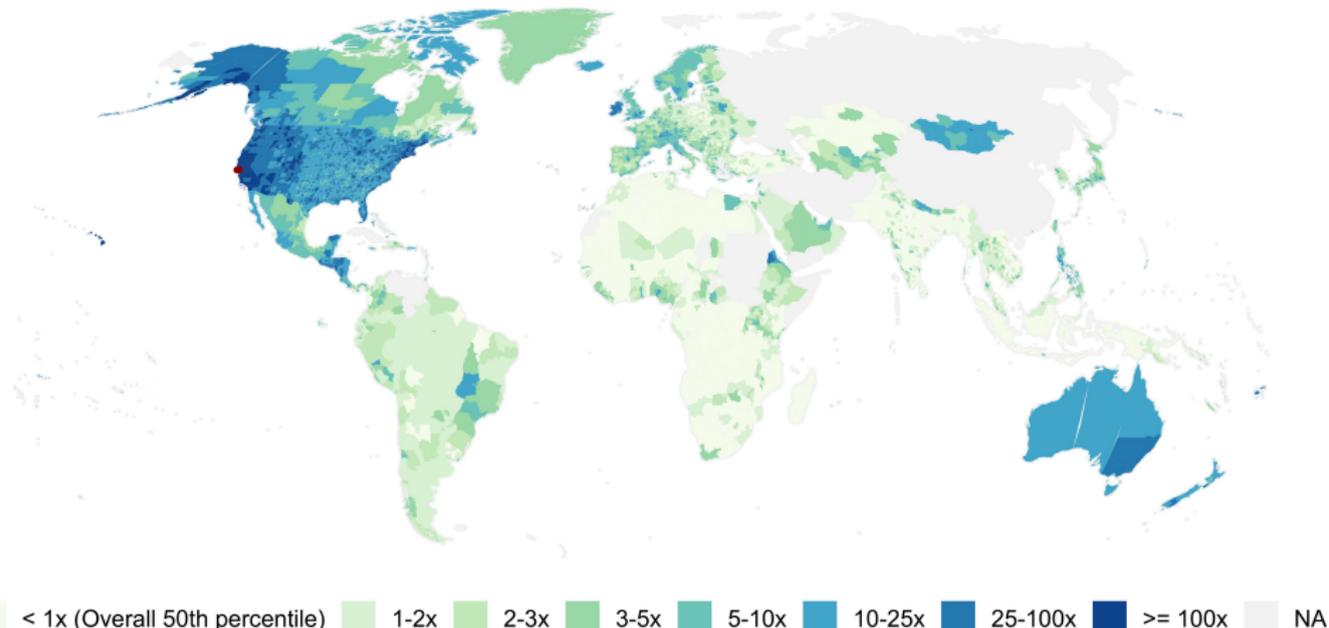


Other SCI examples: San Francisco & Kern Counties

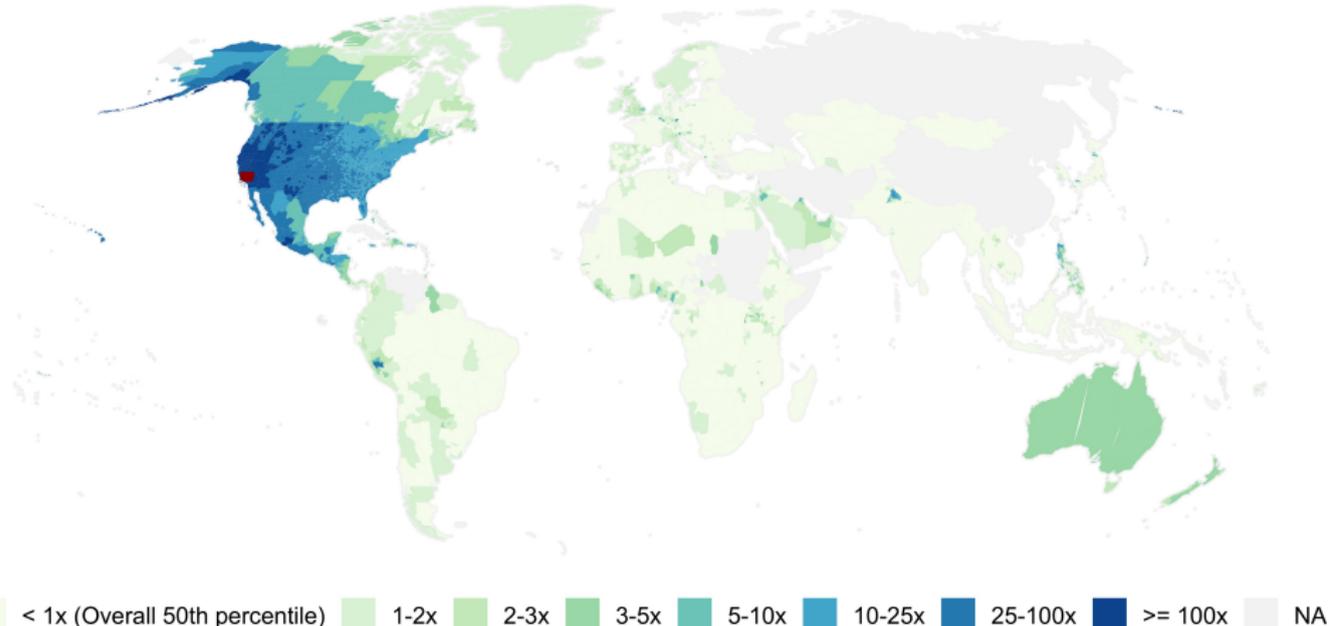
- San Francisco County, CA
 - Median Household Income: \$72,947
 - Median age: 39 years
 - Share non-Hispanic White: 41.9%
 - Share Hispanic: 15.1%
 - Share Black: 6.1%
 - Share Asian: 33.3%

- Kern County, CA
 - Median Household Income: \$48,021
 - Median Age: 32 years
 - Share non-Hispanic White: 49.5%
 - Share Hispanic: 38.4%
 - Share Black: 6.0%
 - Share Asian: 3.4%

Other SCI examples: San Francisco County to the World



Other SCI examples: Kern County to the World



Other SCI examples: SCI in San Francisco & Kern Counties

- San Francisco County, CA
 - Stronger connections to US east coast, western Europe (esp. Ireland), Australia, and Mongolia
- Kern County, CA
 - Stronger connections to western Mexico (consistent with large Hispanic population) and close-by areas in California
 - Connections to Oklahoma (Dust Bowl migration) and North Dakota (oil boom)
 - Generally less connected to rest of US and world

Historical Determinants of Social Connectedness - Results

	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)
All Table 1 Controls	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

- Persistent relationship between political borders and connectedness

References



Bailey, Michael and Cao, Rachel and Kuchler, Theresa and Stroebel, Johannes and Wong, Arlene

Social connectedness: Measurements, determinants, and effects.

Journal of Economic Perspectives, 2018.