Virtually Borderless?
Cultural Proximity and International Collaboration of Developers

Lena Abou El-Komboz* Moritz Goldbeck†

September 2, 2024

Abstract

Are national borders an impediment to online collaboration in the knowledge economy? Unlike in goods trade, knowledge workers can collaborate fully virtually, such that border effects might be eliminated. Here we study collaboration patterns of some 144,000 European developers on the largest online code management platform, *GitHub*. To assess the presence of border effects we deploy a gravity model that explains developers' inter-regional collaboration networks. We find a sizable border effect of –16.4%, which is, however, five to six times smaller than in trade. The border effect is entirely explained by cultural factors such as common language, shared interests, and historical ties. The international border effect in Europe is much larger than the state border effect in the US, where cross-border cultural differences are much less pronounced, further strengthening our conjecture that culture is a main driver of the border effect in virtual collaboration.

Keywords: digitization; software development; knowledge work; culture; language

JEL Codes: F66; J61; O31; O33; O36

*University of Munich & ifo Institute.

†Corresponding author: goldbeck@ifo.de. University of Munich & ifo Institute.

We thank Raunak Mehrotra for excellent research assistance and gratefully acknowledge public funding through DFG grant number 280092119.

1 Introduction

Border effects, the reduction of economic exchange that flows across international borders, are one of the most consistent empirical findings in international economics. Border effects (or home bias) are present, for example, in trade (Anderson and van Wincoop, 2003; McCallum, 1995) and innovative activity (Peri, 2005; Maurseth and Verspagen, 2002). Today however, digital exchange enabled by modern information and communication technologies (ICT) accounts for a sizable part of economic activity. In such digital settings, traditional explanations for the presence of border effects, such as trade or transportation costs, do not apply (Blum and Goldfarb, 2006).

In fact, evidence suggests virtual collaboration can be effective (Choudhury et al., 2021; Bloom et al., 2015), and ICT has been shown to promote inter-regional collaboration in innovative sectors (Chen et al., 2022; Forman and van Zeebroeck, 2019, 2012; Agrawal and Goldfarb, 2008; Adams et al., 2005). Still, cultural factors such as language (Koçak and Puranam, 2022; Cao et al., 2024; Gomez-Herrera et al., 2014) and social ties (Diemer and Regan, 2022; Agrawal et al., 2006) potentially cause a significant border effect even in knowledge worker collaboration (Gorodnichenko and Roland, 2017; Bercovitz and Feldman, 2011; Cummings and Kiesler, 2007; Hinds and Bailey, 2003). But while there is ample evidence on large border effects in trade, production-side investigations of the border effect in collaboration within the digital economy are scant.[1]

In this paper, we therefore ask if a border effect is present in digital production by analyzing virtual collaboration of software developers. We further examine the relation between the border effect and cultural factors. Using unique data on the inter-regional collaboration of around 144,000 European developers on the largest online code management platform, *GitHub*, we estimate the border effect in virtual collaboration in a parsimonious region-level gravity framework. We assess potential drivers of the border effect via the inclusion of a large set of potential cultural determinants while controlling for confounding factors. As a reference, we estimate the border effect using the same model and data for US state borders, where cross-border cultural differences are much less pronounced compared to national borders in Europe.

Our analysis reveals a significant digital border effect of -16.4% for developer collaboration in Europe, after accounting for collaboration potential and geographic factors. Although this effect is sizable, it is five to six times smaller than that observed in goods trade. Our results further suggest cultural factors fully explain the digital border effect. Specifically, common interests, a common spoken language and a shared history are significantly associated with the border effect. Social ties do not explain much of the border effect but rather the distance gradient. Comparison with the state border effect in the US, a setting where cultural and

[1] Recent studies for trade include, e.g., Head and Mayer (2021); Havranek and Irsova (2017); Anderson et al. (2014). Likewise, Santamaría et al. (2023a,b) find large border effects in Europe. For patent collaborations, Singh and Marx (2013) find significant but diminishing border effects, and Li (2014) shows that the decrease over time is driven by age effects.

language differences are largely absent, suggests that indeed culture is a main driver of the international border effect.

2 Data

We compute regional collaboration networks of software developers on *GitHub*, the by far largest code management platform. To this end, we tap data from *GHTorrent* (Gousios, 2013), which comprises public user profiles and repositories as well as a detailed activity stream capturing users' contributions between 09/2015 and 03/2021. We assign users that provide a location to cities via exact matching to city names in the *World Cities Database*. Defining a collaboration as contributing to at least one joint project, we compute the regional collaboration network at the NUTS2 level.[2] NUTS2 regions reflect the data resolution (users typically state metro-area level locations, e.g., "Munich area") and sample size, allowing for sufficient observations in each region. Overall, our data contains 290 NUTS2 regions in 34 European countries[3] and captures the activity of 144,121 active, geolocated, and collaborating users. Users are highly concentrated in space, reflected in the regional collaboration patterns in Figure 1 with most collaborations between the large cities.

We associate potential border effects to cultural proximity. First, we use a composite measure derived from detailed data on online behavior (Obradovich et al., 2022). This large-scale data collection effort systematically queries the *Facebook* marketing API capturing users' online behavior. Specifically, Obradovich et al. (2022) extract 60,000 interest dimensions with at least 500,000 worldwide users to create a composite as well as sub-indices for cultural proximity as cosine distance between the interest vectors of two populations. Second, we relate border effects to genetic distance, a well-established proxy for cultural factors associated with ethnicity (Spolaore and Wacziarg, 2009). We use the cross-country genetic distance data from Creanza et al. (2015), which measures the co-ancestoral distance between national populations. Third, we account for important cultural factors used in gravity models (Conte et al., 2022). We use common spoken language (Melitz and Toubal, 2014) and religious proximity (Disdier and Mayer, 2007) as well as shared history, i.e., whether countries ever were part of the same nation or have a colonial history (Mayer and Zignago, 2011).

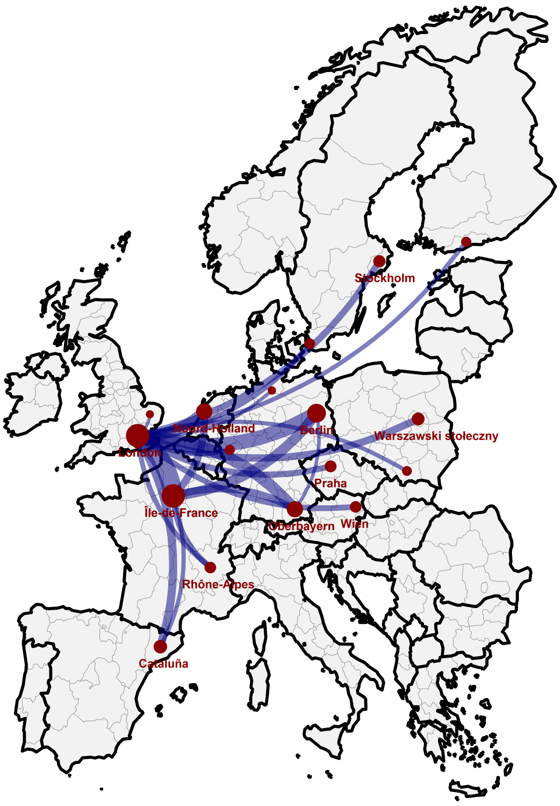
3 Empirical model

To estimate border effects, we deploy the gravity model. While traditionally applied in cross-country settings the model is equally suitable at the sub-national level, where it is used to estimate border effects (Anderson and van Wincoop, 2003; McCallum, 1995). In our context, the gravity model states that regional collaboration is proportional to the product of the regions' masses and inversely proportional to the distance

[2] We merge the NUTS2 regions for London, UKI3 through UKI7, to increase comparability, as this is the only capital city metro area that is split into multiple NUTS2 regions.

[3] Table A.1 reports user numbers by country.

Figure 1: Regional collaboration network



Notes: Map shows the structure of the European software developer collaboration network. Important edges of the network, defined as links between economic areas above 25,000 connections, are shown in blue and scaled by the logarithm of the number of links. Economic areas shown in gray with their centroids as nodes in red, scaled by overall links to other economic areas. Ireland not shown. *Sources:* GHTorrent, own calculations.

between the regions. We take the parsimonious gravity model from McCallum (1995) as starting point for estimating the border effect:

$$\ln(y_{i,j}) = \beta_0 + \beta_1 \text{crossborder}_{i,j} + \beta_2 \text{coloc}_{i,j} + \beta_3 \ln(\text{dist}_{i,j}) + \delta_i + \delta_j + \varepsilon_{i,j} \tag{1}$$

where $y_{i,j}$ represents the number of bilateral collaborations between regions i and j, including domestic collaborations $i = j$. The variable crossborder$_{i,j}$ indicates if region i is located in a different country than region j, and dist$_{i,j}$ denotes the centroid-based geographic distance between the regions.[4] We further add a colocation indicator for within-region collaborations, coloc$_{i,j}$, to account for colocation effects in collabo-

[4] Figure A.1 depicts the distance histogram for within-country and cross-border collaboration.

ration (Goldbeck, 2023; Urry, 2002). Origin and destination fixed effects δ_i and δ_j account for unobserved regional determinants of collaboration. The coefficient β_2 captures the elasticity of collaboration with respect to geographic distance, which we expect to be negative from theory. The border effect is given by β_1, which we expect to be negative or zero, depending on the presence of a border effect.

It is important for the interpretation of the effect to clarify how the border effect is conceptualized in the model. The key identifying assumption for the border effect in the gravity model is that there are no third factors related to the border indicator driving collaboration. The plausibility of this assumption depends on how we think of the border effect. If we think of the border effect narrowly in the sense that the border *itself* causes collaboration to decrease, this assumption is clearly implausible. However, if we conceptualize the border effect as a proxy measure of *all* things that vary across borders and possibly determine collaboration, it is plausible yet tautological. Put differently, the border effect estimated from Equation 1 represents a quantification of how much inter-regional collaborations decline on average for cross-border links as compared to within-country links. Therefore, it should rather be interpreted as descriptive proxy measure of many potential deeper determinants rather than causal estimate of the effect of the border itself.

To assess the specific drivers of the border effect, we extend the baseline model to include variables at the country-pair level measuring different cultural dimensions that vary across borders:

$$\ln(y_{i,j}) = \beta_0 + \beta_1 \text{crossborder}_{i,j} + \beta_2 \text{coloc}_{i,j} + \beta_3 \ln(\text{dist}_{i,j}) + \mathbf{X}'_{c(i),c(j)} \boldsymbol{\beta_4} + \mathbf{X}'_{i,j} \boldsymbol{\beta_5} + \delta_i + \delta_j + \varepsilon_{i,j} \quad (2)$$

where $\mathbf{X}_{c(i),c(j)}$ is a vector of variables that measure differences between the respective country of region i, $c(i)$, and the country of region j, $c(j)$. By definition, these differences are zero if region i and j belong to the same country, $c(i) = c(j)$. Thus, the coefficients $\boldsymbol{\beta_4}$ capture the part of the border effect that is attributable to a particular cross-border difference while β_1 is the residual part of the border effect not explained by variables included in $\mathbf{X}_{c(i),c(j)}$.[5] $\mathbf{X}_{i,j}$ is a vector of region-pair level determinants of collaboration.

As in the baseline model, the main assumption for causal interpretation of the coefficients $\boldsymbol{\beta_4}$ is that there are no omitted factors related to $\mathbf{X}_{c(i),c(j)}$ determining inter-regional collaboration. Note that the cross-border indicator isolates the remaining part of the border effect and therefore provides indication of omitted variable bias when significant. Nonetheless, country-pair explanatory variables that are related to unobserved determinants of collaboration are a threat to identification. Together with potential measurement error, especially in related explanatory variables, this cautions us of a narrow interpretation of the separate coefficients in $\boldsymbol{\beta_4}$.

Especially since cultural factors are often interrelated and can have common deep determinants, a narrow causal interpretation is likely inappropriate. Rather, the model provides some indication of possible deter-

[5] Note that while cross-country differences are sufficient to elicit the relation to the border effect as captured by the cross-country indicator, differences in the underlying populations of such metrics and software developers potentially introduce a downward bias in our coefficient estimates, leading to underestimation of the effect of cultural factors.

minants as it points to dimensions that are statistically associated with the border effect. Plausible, theory-guided selection of a statistically manageable amount of explanatory variables is therefore paramount to avoid spurious correlation or multicollinearity issues. Table A.2 reports the correlation matrix of our main variables. Further, note that Equations 1 and 2 are partial equilibrium models and, as such, estimated border effects should not be misconstrued as counterfactual for border removal, as widely acknowledged in the literature (see, e.g., Santamaría et al., 2023a; Havranek and Irsova, 2017).

4 Results

4.1 Digital border effect

Table 1 reports estimation results of the border effect. We start with a model that does not consider gravity. This raw correlation in model (1) suggests a large border effect of –60%. Controlling for size in terms of logarithms of multiplied user bases halves the effect. Model (3) drops the functional form assumption on size and instead includes origin and destination region fixed effects, slightly increasing the estimate of the border effect. Finally, our preferred specification in model (4) resembles a typical parsimonious gravity model that additionally controls for geographic distance and colocation. Results show a highly significant negative relation of collaboration and distance and a substantial collaboration premium for colocation. Importantly, there is a significant border effect, with 16.4% fewer collaborations for region-pairs that are located in different countries.

While the border effect is economically significant, it is much smaller than for trade. The meta-analysis by Havranek and Irsova (2017) finds an average border effect of –91.5%[6] for trade, nearly identical to the border effect for Europe in Santamaría et al. (2023b) of –90.4%[7] estimated from recent granular freight data. Thus, the border effect is five to six times larger for trade than in software developer collaboration. This is generally in line with our conjecture that national borders should play a minor or no role for virtual collaboration of software developers.

4.2 The role of culture

We elicit association of various cultural factors with the border effect and collaboration by including appropriate cross-country level variables as specified in Equation 2.

Table 2 reports the results of models that consider cross-country cultural differences. Note that the metrics for culture are available only for a subset of countries. For consistency, we estimate all models on the same,

[6]Cf. the unweighted mean coefficient for the EU in Table 1 in Havranek and Irsova (2017), expressed as home bias of $\exp(2.55) - 1 \approx 11.8$, translated into a percentage border effect as defined here via $\left(\frac{1}{\exp(2.55)-1} - 1\right) * 100$.

[7]Cf. the border effect coefficient in Table 1 column (2) of Santamaría et al. (2023b), translated into a percentage border effect as defined here via $(\exp(-2.34) - 1) * 100$.

Table 1: Border effect in collaboration

Collaboration	(1)	(2)	(3)	(4)
Cross-border	-0.906***	-0.371***	-0.446***	-0.180***
	(0.041)	(0.016)	(0.012)	(0.014)
Users, multiplied [log]		0.755***		
		(0.002)		
Colocation				0.862***
				(0.068)
Distance [log]				-0.129***
				(0.007)
Origin FE			×	×
Destination FE			×	×
Observations	84,100	84,100	84,100	84,100
Adj. R^2	0.011	0.837	0.919	0.922
Border effect	-59.6%	-31.0%	-36.0%	-16.4%

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Users, multiplied, is the natural logarithm of the multiplication of the number of users in origin and destination. Robust standard errors are reported in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. *Sources:* GHTorrent, own calculations.

reduced sample that features a slightly higher baseline border effect in model (1). In model (2), we add two distinct composite measures of culture. First, we take the cultural distance metric from Obradovich et al. (2022) derived from common online interests. Second, we control for genetic distance from Spolaore and Wacziarg (2009). The coefficient estimates of both distance measures have the expected negative sign. Cultural distance is strongly negatively associated with collaboration while genetic distance is much less relevant and also features weaker significance. Importantly, the border effect is entirely explained by these cultural distance composite measures, as shown by the insignificant point estimate close to zero of the border effect coefficient.

In model (3), we further add specific cultural factors, namely common language, religious distance, and same country or colonial history. Religious distance is statistically and economically insignificantly related to collaboration. In contrast, there appears to be a sizable relation with common spoken language of around 8.4% more collaborations, although imprecisely estimated. The magnitude of the language effect is almost 14 times smaller compared to trade, where the corresponding semi-elasticity is 0.775 (Melitz and Toubal, 2014).[8] A shared colonial history does not explain collaboration today, likely due to the few colonial relationships within Europe. History as a same country is associated negatively with collaboration, which is

[8]Cf. column (2) in Table 3 of Melitz and Toubal (2014).

Table 2: Collaboration and cultural proximity

Collaboration	(1)	(2)	(3)	(4)
Cross-border	-0.233***	-0.009	-0.014	0.013
	(0.012)	(0.035)	(0.037)	(0.038)
Colocation	1.341***	1.485***	1.476***	1.472***
	(0.066)	(0.069)	(0.070)	(0.070)
Distance [log]	-0.046***	-0.016**	-0.018**	-0.009
	(0.007)	(0.008)	(0.008)	(0.008)
Cultural distance		-0.097***	-0.081***	-0.080***
		(0.016)	(0.017)	(0.017)
Genetic distance		-0.001**	-0.001*	-0.001*
		(0.000)	(0.000)	(0.000)
Common language			0.082**	0.062*
			(0.034)	(0.034)
Religious distance			-0.005	-0.007
			(0.020)	(0.020)
Same country history			-0.071**	-0.078***
			(0.028)	(0.028)
Colonial history			0.011	0.001
			(0.016)	(0.016)
Social connectedness				0.013***
				(0.004)
Origin FE	×	×	×	×
Destination FE	×	×	×	×
Observations	55,169	55,169	55,169	55,169
Adj. R^2	0.947	0.947	0.947	0.947

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. *Sources:* GHTorrent, Obradovich et al. (2022), Creanza et al. (2015), Bailey et al. (2018), CEPII, own calculations.

surprising only at first as it likely relates to the fact that this indicator captures mostly historical occupations in the former Yugoslavia and Austria-Hungary that lead to disrupted relationships until today.

A large strand of literature examines the role of social ties on knowledge worker collaboration (e.g., Bercovitz and Feldman, 2011) and knowledge flows (e.g., Diemer and Regan, 2022; Reagans et al., 2005). As social ties are closely related to geographic distance (Bailey et al., 2018; Breschi and Lissoni, 2009) they are an important channel to explain the robust distance effect in gravity applications (Diemer and Regan, 2022; Garmendia et al., 2012; Bercovitz and Feldman, 2011; Breschi and Lissoni, 2009) as well as for collaboration success more generally (Hahn et al., 2008; Cowan et al., 2007; Grewal et al., 2006). Model (4) additionally adds social connectedness between regions as explanatory variable for collaboration. Social connectedness is highly statistically and economically significantly related to collaboration. Inclusion diminishes the geographic distance and language effect, but otherwise does not significantly alter the results.

This points to the distance effect being driven by social connections and is reassuring toward the other effects. This is in line with empirical evidence on knowledge worker collaboration suggesting a high relevance of face-to-face meeting possibility (Emanuel et al., 2023; Atkin et al., 2022) but irrelevance of geographic distance otherwise (Goldbeck, 2023) and feeds into the discussion that geography, in most models, is to a large extent merely a proxy for deeper determinants of outcomes (Waldinger, 2012; Azoulay et al., 2010).

We further investigate the relation between culture and international collaboration using the decomposition of the cultural interest composite measure by Obradovich et al. (2022) into 14 subcategories of interest. The results reported in Table A.3 reveal that especially different interests in the category *non-local business* explain the border effect. This means that international software developer collaboration is associated with overlapping professional interests. This is in line with existing evidence that organizational links attenuate negative border effects associated with culture (Duede et al., 2024; Fadeev, 2023; Adams et al., 2005). Other subcategories are relatively unimportant, but mostly show positive associations. This implies cultural differences are not unidimensionally negatively related to collaboration and some cultural differences, e.g. other food or lifestyle, might in fact spur collaboration.

To assess the robustness of our findings, we include a variety of additional explanatory variables into our model. Results are reported in Table A.5 for additional historical circumstances, Table A.6 for political factors, and Table A.7 for more nuanced measures of linguistic proximity and an alternative measure of religious distance. Inclusion does not significantly alter our main effects regarding the border effect, cultural proximity, history, or language. Additionally, we estimate the border effect for US state borders in Table A.4, where cultural differences are much less pronounced. The international border effect in Europe is much larger than the state border effect in the US, further strengthening our conjecture that culture is a main driver of the border effect in virtual collaboration.

5 Conclusion

We provide evidence of border effects in virtual collaboration that are, however, five to six times smaller compared to trade. This is consistent with trade and transportation costs being largely absent in the digital economy. The border effect in software developer collaboration in Europe is entirely explained by cultural factors, especially shared (professional) interest, a common language, and history. Most other political and historical circumstances are unrelated to the digital border effect. Compared to the digital border effect at the domestic borders between US states, where cultural differences are comparably negligible, the European digital border effect is about twice as large.

This study has limitations that open up avenues for further research. Notably, our setting lacks a quasi-experimental approach where stronger identification could be achieved. Yet, already few settings exist where border effects can be estimated at all, as estimation requires domestic flow data. Opportunities to causally estimate border effects are extremely rare (e.g., Santamaría et al., 2023a). Additionally, culture evolves endogenously, which makes it difficult to causally explore the intricate patterns of mediation and co-determination among the countless cultural factors in a comprehensive framework. Further, the measurement of culture is ideally conducted on a more granular scale both population-wise and geographically as software developers might be different to the general population along these dimensions.

Our work has several implications relevant to policy makers and management. The digital border effect is relatively small, which points to improved feasibility of international collaboration in digital knowledge work. Generally, relative to market integration on the consumer side, production-side barriers in the digital economy are largely overlooked albeit their increasing importance for digital business. This seems especially important in Europe, where the workforce is geographically distributed across nation states and international collaboration is required to exploit size advantages of labor markets in the era of remote work. Importantly, together with decreasing role of geography in ICT-intensive settings of the knowledge economy, our results suggest that management and policy makers should shift their attention to cultural barriers to collaboration as they are relatively more important in the digital economy.

References

Adams, James D., Grant C. Black, J. Roger Clemmons, and Paula E. Stephan, "Scientific Teams and Institutional Collaborations: Evidence from US Universities, 1981–1999," *Research Policy*, 2005, *34* (3), 259–285.

Agrawal, Ajay and Avi Goldfarb, "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 2008, *98* (4), 1578–1590.

_ , Iain Cockburn, and John McHale, "Gone But Not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships," *Journal of Economic Geography*, 2006, *6* (5), 571–591.

Anderson, James and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

Anderson, James E., Catherine A. Milot, and Yoto V. Yotov, "How Much Does Geography Deflect Services Trade? Canadian Answers," *International Economic Review*, 2014, *55* (3), 791–818.

Atkin, David, M. Keith Chen, and Anton Popov, "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley," *NBER Working Paper*, 2022.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, "Superstar Extinction," *The Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–80.

Bercovitz, Janet and Maryann Feldman, "The Mechanisms of Collaboration in Inventive Teams: Composition, Social Networks, and Geography," *Research Policy*, 2011, *40* (1), 81–93.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying, "Does Working from Home Work? Evidence from a Chinese Experiment," *The Quarterly Journal of Economics*, 2015, *130* (1), 165–218.

Blum, Bernardo and Avi Goldfarb, "Does the Internet Defy the Law of Gravity?," *Journal of International Economics*, 2006, *70* (2), 384–405.

Breschi, Stefano and Francesco Lissoni, "Mobility of Skilled Workers and Co-invention Networks: An Anatomy of Localized Knowledge Flows," *Journal of Economic Geography*, 2009, *9* (4), 439–468.

Cao, Yihui, Robin C. Sickles, Thomas P. Triebs, and Justin Tumlinson, "Linguistic Distance to English Impedes Research Performance," *Research Policy*, 2024, *53* (4), 104971.

Chen, Chinchih, Carl Benedikt Frey, and Giorgio Presidente, "Disrupting Science," *Working Paper*, 2022.

Choudhury, Prithwiraj, Cirrus Foroughi, and Barbara Larson, "Work-from-Anywhere: The Productivity Effects of Geographic Flexibility," *Strategic Management Journal*, 2021, *42* (4), 655–683.

Conte, Maddalena, Pierre Cotterlaz, and Thierry Mayer, *The CEPII Gravity Database*, CEPII, 2022.

Cowan, Robin, Nicolas Jonard, and Jean-Benoit Zimmermann, "Bilateral Collaboration and the Emergence of Innovation Networks," *Management Science*, 2007, *53* (7), 1051–1067.

Creanza, Nicole, Merritt Ruhlen, Trevor J. Pemberton, Noah A. Rosenberg, Marcus W. Feldman, and Sohini Ramachandran, "A Somparison of Worldwide Phonemic and Genetic Variation in Human Populations," *Proceedings of the National Academy of Sciences*, 2015, *112* (5), 1265–1272.

Cummings, Jonathon N. and Sara Kiesler, "Coordination Costs and Project Outcomes in Multi-University Collaborations," *Research Policy*, 2007, *36* (10), 1620–1634.

Diemer, Andreas and Tanner Regan, "No Inventor is an Island: Social Connectedness and the Geography of Knowledge Flows in the US," *Research Policy*, 2022, *51* (2), 104416.

Disdier, Anne-Célia and Thierry Mayer, "Je t'Aime, Moi Non Plus: Bilateral Opinions and International Trade," *European Journal of Political Economy*, 2007, *23* (4), 1140–1159.

Duede, Eamon, Misha Teplitskiy, Karim Lakhani, and James Evans, "Being Together in Place as a Catalyst for Scientific Advance," *Research Policy*, 2024, *53* (2), 104911.

Emanuel, Natalia, Emma Harrington, and Amanda Pallais, "The Power of Proximity: Training of Tomorrow or Productivity Today?," *Working Paper*, 2023.

Fadeev, Evgenii, "Creative Construction: Knowledge Sharing and Cooperation Between Firms," *Working Paper*, 2023.

Forman, Chris and Nicolas van Zeebroeck, "From wires to partners: How the Internet has fostered R&D collaborations within firms," *Management science*, 2012, *58* (8), 1549–1568.

__ **and** __ , "Digital Technology Adoption and Knowledge Flows Within Firms: Can the Internet Overcome Geographic and Technological Distance?," *Research Policy*, 2019, *48* (8).

Garmendia, Aitor, Carlos Llano, Asier Minondo, and Francisco Requena, "Networks and the Disappearance of the Intranational Home Bias," *Economics Letters*, 2012, *116* (2), 178–182.

Goldbeck, Moritz, "Bit by Bit: Colocation and the Death of Distance in Software Developer Networks," *Working Paper*, 2023.

Gomez-Herrera, Estrella, Bertin Martens, and Geomina Turlea, "The Drivers and Impediments for Cross-border E-commerce in the EU," *Information Economics and Policy*, 2014, *28*, 83–96.

Gorodnichenko, Yuriy and Gerard Roland, "Culture, Institutions, and the Wealth of Nations," *Review of Economics and Statistics*, 2017, *99* (3), 402–416.

Gousios, Georgios, "The GHTorent Dataset and Tool Suite," *IEEE 10th Working Conference on Mining Software Repositories (MSR)*, 2013, pp. 233–236.

Graafland, Johan and Eelke de Jong, "The Moderating Role of Culture on the Benefits of Economic Freedom: Cross-country Analysis," *Journal of Comparative Economics*, 2022, *50* (1), 280–292.

Grewal, Rajdeep, Gary L. Lilien, and Girish Mallapragada, "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems," *Management Science*, 2006, *52* (7), 1043–1056.

Hahn, Jungpil, Jae Yun Moon, and Chen Zhang, "Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties," *Information Systems Research*, 2008, *19* (3), 369–391.

Havranek, Tomas and Zuzana Irsova, "Do Borders Really Slash Trade? A Meta-analysis," *IMF Economic Review*, 2017, *65*, 365–396.

Head, Keith and Thierry Mayer, "The United States of Europe: A Gravity Model Evaluation of the Four Freedoms," *Journal of Economic Perspectives*, 2021, *35* (2), 23–48.

Hinds, Pamela J. and Diane E. Bailey, "Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams," *Organization Science*, 2003, *14* (6), 615–632.

Koçak, Özgecan and Phanish Puranam, "Separated by a Common Language: How the Nature of Code Differences Shapes Communication Success and Code Convergence," *Management Science*, 2022, *68* (7), 5287–5310.

Li, Yao Amber, "Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? Evidence from Patent Citations," *European Economic Review*, 2014, *71*, 152–172.

Maurseth, Per Botolf and Bart Verspagen, "Knowledge Spillovers in Europe: A Patent Citations Analysis," *Scandinavian Journal of Economics*, 2002, *104* (4), 531–545.

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures: The GeoDist Database," 2011.

McCallum, John, "National Borders Matter: Canada-US Regional Trade Patterns," *American Economic Review*, 1995, *85* (3), 615–623.

Melitz, Jacques and Farid Toubal, "Native Language, Spoken Language, Translation and Trade," *Journal of International Economics*, 2014, *93* (2), 351–363.

Obradovich, Nick, Ömer Özak, Ignacio Martín, Ignacio Ortuño-Ortín, Edmond Awad, Manuel Cebrián, Rubén Cuevas, Klaus Desmet, Iyad Rahwan, and Ángel Cuevas, "Expanding the Measurement of Culture with a Sample of Two Billion Humans," *Journal of the Royal Society Interface*, 2022, *19* (190), 20220085.

Peri, Giovanni, "Determinants of Knowledge Flows and Their Effect on Innovation," *Review of Economics and Statistics*, 2005, *87* (2), 308–322.

Reagans, Ray, Linda Argote, and Daria Brooks, "Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together," *Management Science*, 2005, *51* (6), 869–881.

Santamaría, Marta, Jaume Ventura, and Uğur Yeşilbayraktar, "Borders Within Europe," *Working Paper*, 2023.

_ , _ , and _ , "Exploring European Regional Trade," *Journal of International Economics*, 2023, p. 103747.

Singh, Jasjit and Matt Marx, "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity," *Management Science*, 2013, *59* (9), 2056–2078.

Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," *The Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

Urry, John, "Mobility and Proximity," *Sociology*, 2002, *36* (2), 255–274.

Waldinger, Fabian, "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, 2012, *79* (2), 838–861.

A Appendix

A.1 Tables

Table A.1: Users by country

ISO2	Country	Users	Share
UK	United Kingdom	32,914	22.8%
FR	France	23,516	16.3%
DE	Germany	21,211	14.7%
PL	Poland	10,293	7.1%
NL	Netherlands	9,371	6.5%
ES	Spain	7,104	4.9%
IT	Italy	5,167	3.6%
CZ	Czech Republic	3,701	2.6%
SE	Sweden	3,692	2.6%
FI	Finland	3,660	2.5%
DK	Denmark	3,227	2.2%
AT	Austria	3,021	2.1%
CH	Switzerland	2,637	1.8%
BE	Belgium	2,136	1.5%
NO	Norway	1,897	1.3%
RO	Romania	1,863	1.3%
EL	Greece	1,682	1.2%
PT	Portugal	1,534	1.1%
HR	Croatia	965	0.7%
RS	Serbia	740	0.5%
	Other	3,790	2.6%
	Total	144,121	100%

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Distance is scaled in 100km. Users, GDPs, and Populations refers to the respective variables for both origin and destination. Users, multiplied, is the multiplication of the number of users in origin and destination. Collaboration with Anchorage, AK, and Honolulu, HI, are excluded. Robust standard errors are reported in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. *Sources:* GHTorrent, Bureau of Economic Analysis, own calculations.

Table A.2: Correlation matrix of main explanatory variables

	colocation	distance	language	genetic	culture	SCI	history	religion	colonial
colocation	1.0000								
distance	-0.5599	1.0000							
language	0.1264	-0.5082	1.0000						
genetic	-0.0916	0.4875	-0.5678	1.0000					
culture	-0.1872	0.6308	-0.7188	0.6097	1.0000				
SCI	0.3525	-0.6535	0.5777	-0.3775	-0.6415	1.0000			
history	-0.0099	-0.1708	0.0079	0.0050	0.0071	0.1046	1.0000		
religion	0.1535	-0.4448	0.3232	-0.3314	-0.5368	0.5736	0.0262	1.0000	
colonial	-0.0134	-0.0798	-0.0283	0.0124	0.0050	0.0998	0.1785	-0.0355	1.0000

Notes: Table reports correlations between main explanatory variables in model (4) of Table 2. *Sources:* GHTorrent, Obradovich et al. (2022), Creanza et al. (2015), Bailey et al. (2018), CEPII, own calculations.

Table A.3: Collaboration and interests

Collaboration	(1)	(2)	(3)
Cross-border	-0.414***	-0.212***	-0.004
	(0.011)	(0.013)	(0.032)
Colocation		1.132***	1.436***
		(0.067)	(0.070)
Distance [log]		-0.084***	-0.025***
		(0.007)	(0.008)
Business and Industry			0.918**
			(0.409)
Education			0.000
			(0.164)
Family and Relationships			-0.700***
			(0.185)
Fitness and Wellness			1.704***
			(0.552)
Food and Drink			1.153**
			(0.473)
Hobbies and Activities			2.089***
			(0.372)
Lifestyle and Culture			3.788***
			(0.427)
News and Entertainment			6.952***
			(0.795)
Non-local Business			-17.013***
			(2.024)
People			0.287***
			(0.068)
Shopping and Fashion			0.595
			(0.435)
Sports and Outdoors			0.152
			(0.163)
Technology			1.035***
			(0.299)
Travel, Places and Events			1.074***
			(0.266)
Other			-1.000
			(0.737)
Origin FE	×	×	×
Destination FE	×	×	×
Observations	77,284	77,284	77,284
Adj. R^2	0.929	0.932	0.933

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. *Sources:* GHTorrent, Obradovich et al. (2022), own calculations.

Table A.4: Border effect in the United States

Collaboration	(1)	(2)	(3)	(4)
Cross-border	-0.527***	-0.429***	-0.502***	-0.100***
	(0.098)	(0.041)	(0.037)	(0.033)
Users, multiplied [log]		0.750***		
		(0.004)		
Colocation				2.191***
				(0.073)
Distance [log]				-0.060***
				(0.011)
Origin FE			×	×
Destination FE			×	×
Observations	32,041	32,041	32,041	32,041
Adj. R^2	0.002	0.856	0.917	0.922
Border effect	-41.0%	-34.9%	-39.4%	-9.5%
Δ(Europe – USA)	-18.6 p.p.	+3.9 p.p.	+3.4 p.p.	-6.9 p.p.
BE_{USA} / BE_{Europe}	0.69	1.13	1.09	0.58

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Distance is scaled in 100km. Users, GDPs, and Populations refers to the respective variables for both origin and destination. Users, multiplied, is the multiplication of the number of users in origin and destination. Collaboration with Anchorage, AK, and Honolulu, HI, are excluded. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. *Sources:* GHTorrent, Bureau of Economic Analysis, Goldbeck (2023), own calculations.

Table A.5: Collaboration and history

Collaboration	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cross-border	0.000	0.032	-0.010	-0.010	-0.008	-0.006	0.048
	(0.037)	(0.042)	(0.041)	(0.041)	(0.041)	(0.037)	(0.043)
Colocation	1.469***	1.441***	1.447***	1.447***	1.473***	1.465***	1.490***
	(0.069)	(0.070)	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)
Distance [log]	-0.007	-0.011	-0.008	-0.008	-0.006	-0.007	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cultural distance	-0.068***	-0.073***	-0.059***	-0.059***	-0.065***	-0.065***	-0.064***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.017)	(0.017)
Genetic distance	-0.001*	-0.001	-0.001*	-0.001*	-0.001*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Common language	0.069**	0.078**	0.103***	0.103***	0.073**	0.066**	0.071**
	(0.033)	(0.033)	(0.034)	(0.034)	(0.034)	(0.033)	(0.033)
Religious distance	-0.000	0.002	0.016	0.016	-0.001	0.004	-0.001
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)
Same country history	-0.081***	-0.078***	-0.072***	-0.072***	-0.080***	-0.116***	-0.091***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Colonial history	0.001	0.011	0.023	0.023	0.001	0.005	0.007
	(0.015)	(0.016)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)
Social connectedness	0.016***	0.017***	0.019***	0.019***	0.016***	0.015***	0.018***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Contiguity		-0.020*					
		(0.010)					
Common legal origin			-0.037***				
			(0.009)				
Common legal origin (post-transformation)				-0.037***			
				(0.009)			
Common legal origin (pre-transformation)					-0.003		
					(0.011)		
Communist history						0.141***	
						(0.041)	
Iron curtain							0.059**
							(0.027)
Origin FE	×	×	×	×	×	×	×
Destination FE	×	×	×	×	×	×	×
Observations	54,702	54,702	54,630	54,630	54,630	54,702	54,702
Adj. R^2	0.949	0.949	0.949	0.949	0.949	0.949	0.949

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. *Sources:* GHTorrent, Obradovich et al. (2022), Creanza et al. (2015), Bailey et al. (2018), CEPII, own calculations.

Table A.6: Collaboration and political systems

Collaboration	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cross-border	0.013	0.008	0.013	0.047	-0.003	0.008	0.003	0.000
	(0.038)	(0.038)	(0.038)	(0.044)	(0.044)	(0.037)	(0.037)	(0.037)
Colocation	1.472***	1.464***	1.471***	1.462***	1.472***	1.449***	1.469***	1.469***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.069)	(0.069)
Distance [log]	-0.009	-0.010	-0.009	-0.010	-0.009	-0.014*	-0.006	-0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Cultural distance	-0.080***	-0.081***	-0.080***	-0.076***	-0.081***	-0.077***	-0.068***	-0.068***
	(0.017)	(0.017)	(0.017)	(0.019)	(0.018)	(0.017)	(0.017)	(0.017)
Genetic distance	-0.001*	-0.001**	-0.001*	-0.001*	-0.001*	-0.001**	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Common language	0.062*	0.055	0.062*	0.066*	0.061*	0.070**	0.068**	0.069**
	(0.034)	(0.035)	(0.034)	(0.034)	(0.034)	(0.034)	(0.033)	(0.033)
Religious distance	-0.007	-0.005	-0.007	0.001	-0.007	0.003	-0.002	-0.001
	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.020)
Same country history	-0.078***	-0.079***	-0.078***	-0.076***	-0.080***	-0.073***	-0.081***	-0.081***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)
Colonial history	0.001	0.002	0.001	0.003	0.017	0.004	0.001	0.001
	(0.016)	(0.016)	(0.016)	(0.016)	(0.033)	(0.016)	(0.015)	(0.015)
Social connectedness	0.013***	0.013***	0.013***	0.012***	0.013***	0.011**	0.017***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Diplomatic disagreement		0.017						
		(0.018)						
EU			-0.020					
			(0.048)					
RTA				-0.044***				
				(0.013)				
Hegemon					-0.019			
					(0.033)			
Monarchies						-0.045***		
						(0.015)		
Δ economic freedom							-0.008	
							(0.018)	
Δ political rights								0.007
								(0.037)
Origin FE	×	×	×	×	×	×	×	×
Destination FE	×	×	×	×	×	×	×	×
Observations	55,169	55,169	55,169	55,097	55,169	55,169	54,702	54,702
Adj. R^2	0.947	0.947	0.947	0.947	0.947	0.947	0.949	0.949

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Robust standard errors are reported in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. *Sources:* GHTorrent, Obradovich et al. (2022), Creanza et al. (2015), Bailey et al. (2018), Graafland and de Jong (2022), CEPII, own calculations.

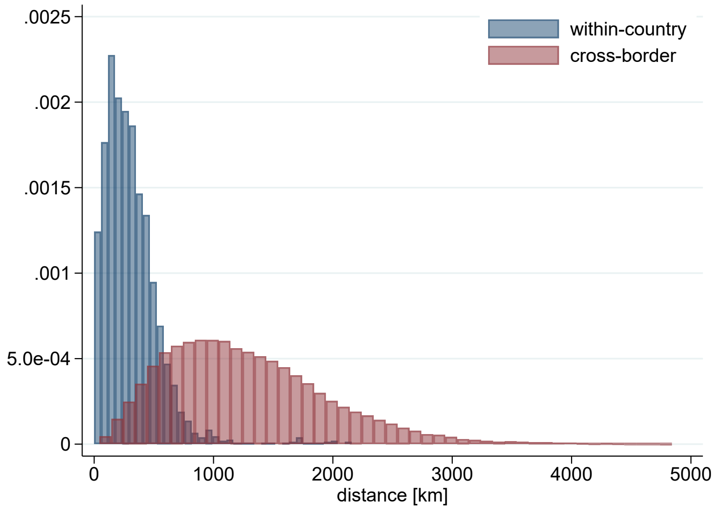
Table A.7: Collaboration, language, and religion

Collaboration	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cross-border	0.013	0.027	0.023	0.033	0.024	0.024	0.021
	(0.038)	(0.037)	(0.043)	(0.048)	(0.037)	(0.037)	(0.040)
Colocation	1.472***	1.460***	1.461***	1.462***	1.462***	1.463***	1.477***
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Distance [log]	-0.009	-0.010	-0.010	-0.009	-0.010	-0.010	-0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Cultural distance	-0.080***	-0.090***	-0.092***	-0.092***	-0.089***	-0.089***	-0.079***
	(0.017)	(0.018)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)
Genetic distance	-0.001*	-0.001***	-0.001**	-0.001**	-0.001**	-0.001**	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Same country history	-0.078***	-0.081***	-0.080***	-0.081***	-0.081***	-0.081***	-0.077***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Colonial history	0.001	-0.000	0.001	0.003	0.002	0.002	0.003
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Social connectedness	0.013***	0.014***	0.014***	0.014***	0.014***	0.014***	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Common spoken language	0.062*						0.064*
	(0.034)						(0.035)
Common native language		0.013					
		(0.025)					
Linguistic proximity (Tree)			0.001				
			(0.003)				
Linguistic proximity (ASJP)				0.002			
				(0.004)			
Common Language Index [log]					0.018		
					(0.028)		
Common Language Index [level]						0.019	
						(0.028)	
Religious distance	-0.007	-0.009	-0.012	-0.013	-0.011	-0.011	
	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	(0.020)	
Religious proximity [Fearon weighted]							0.003
							(0.008)
Origin FE	×	×	×	×	×	×	×
Destination FE	×	×	×	×	×	×	×
Observations	55,169	55,169	55,097	55,097	55,169	55,169	54,702
Adj. R^2	0.947	0.947	0.947	0.947	0.947	0.947	0.947

Notes: The outcome variable is the natural logarithm of collaborations between two economic areas plus one. Colocation indicates collaboration between users in the same economic area. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. *Sources:* GHTorrent, Obradovich et al. (2022), Creanza et al. (2015), Bailey et al. (2018), CEPII, own calculations.

A.2 Figures

Figure A.1: Distance histogram



Notes: Figure shows histograms of within-country and cross-border distances based on NUTS2 centroids, respectively. *Sources:* GHTorrent, own calculations.