ifo WORKING PAPERS

407
2024
March 2024

# Virtually Borderless?

## Cultural Proximity and International Collaboration of Developers

*Lena Abou El-Komboz and Moritz Goldbeck*

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Imprint:

ifo Working Papers
Publisher and distributor: ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5, 81679 Munich, Germany
Telephone + 49(0)89 9224 0, Telefax +49(0)89 985369, email ifo@ifo.de
www.ifo.de

An electronic version of the paper may be downloaded from the ifo website
www.ifo.de

*ifo Working Paper* No. 407

# Virtually Borderless?
## Cultural Proximity and International Collaboration of Developers

Lena Abou El-Komboz* Moritz Goldbeck†

March 13, 2024
Latest Version

**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 repository 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 robust and consistent empirical findings in international economics. Border effects (or home bias) are present, for example, in trade (e.g., Anderson and van Wincoop, 2003; McCallum, 1995), investment (e.g., Chan et al., 2005; Strong and Xu, 2003; French and Poterba, 1991) and innovative activity (e.g., 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 settings of the digital economy, traditional explanations for the presence of border effects, such as trade or transportation costs, do not apply (Blum and Goldfarb, 2006).

In this paper, we therefore ask if a border effect is present in virtual collaboration, as well, and explore its relationship with cultural factors. Using unique data on the inter-regional collaboration of around 144,000 European developers on the largest online code repository platform, *GitHub*, we estimate the border effect in virtual collaboration in a parsimonious region-level gravity framework. We then 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.

The setting of developers collaborating online is particularly suitable here as it not only represents an important and representative sector of the knowledge economy (Korkmaz et al., 2024), but at the same time also precludes many of the traditional explanations driving border effects for multiple reasons. First, online code projects technically allow for fully virtual interaction and IT professionals' adoption of such technologies is high. Second, code development is not affected by transportation costs nor are open-source developers constrained by tariffs or bureaucratic barriers. Third, programming is codified to a higher degree compared to other knowledge work, which facilitates cross-border communication. And lastly, language barriers are likely less important as many developers speak English and use similar (universal) programming languages.

We find a significant digital border effect for developer collaboration in Europe of –16.4% after accounting for collaboration potential and geographic factors in the baseline gravity framework. Although this effect is sizable, it is five to six times smaller as in goods trade. The border effect is particularly high when at least one of the involved countries is small in terms of hosted users. 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 while religious proximity and most political circumstances are unrelated to the border effect. Investigating several widely-used frameworks of cross-country cultural differences shows some relation of the border effect to preferences and interests. There is a particularly strong relation to shared interests in non-local business. In contrast, 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 language differences are largely absent, suggests that indeed culture is a main driver of the international border effect since the domestic border effect is much smaller.

This work entails several contributions that have important managerial and policy implications. It is one of few studies to investigate digital border effects, i.e., border effects in collaboration that technically can be shifted completely into the virtual space. To the best of our knowledge, we are the first to thoroughly examine border effects in software developer collaboration on a online platform. Estimated digital border effects are several magnitudes smaller compared to goods trade, where border effects are studied extensively. Generally, this points to fewer and less important barriers to international collaboration. While existing works on international collaboration are mainly concerned with travel costs or geographic factors, we relate the observed border effect to cultural factors. As geography increasingly becomes less relevant in the knowledge economy, the importance of cultural factors for international collaboration in the digital economy increases. We demonstrate which among the many dimensions of culture, broadly defined, are most strongly related to the digital border effect among software developers in Europe.

The remainder of this paper is organized as follows. We discuss the related literature in Section 2. Section 3 introduces the data. In Section 4, we discuss the empirical model. Results are presented in Section 5 and Section 6 concludes.

## 2 Related literature

**ICT and remote collaboration** This study contributes to three related strands of literature. First, there is a growing literature in economics on the impact of ICT on remote collaboration. Existing work shows that ICT tends to foster inter-regional trade (Visser, 2019; Steinwender, 2018; Jensen, 2007), research and innovation (Forman and van Zeebroeck, 2019; Agrawal and Goldfarb, 2008), and entrepreneurship (Agrawal et al., 2015). However, geographically close exchange tends to increase disproportionately (Akerman et al., 2022; Agrawal and Goldfarb, 2008), in line with theoretical considerations that ICT and geographic proximity are complements (Gaspar and Glaeser, 1998). In knowledge work, colocation is especially important (see, e.g., Goldbeck, 2023; Urry, 2002; Olson and Olson, 2000) and average collaborator distance in teams increases with ICT adoption (Adams et al., 2005). In non-collaborative office settings, remote work is feasible and may even increase productivity (Choudhury et al., 2021; Bloom et al., 2015). Yet, studies find that face-to-face interaction opportunity remains valuable in many settings (e.g., Gibbs et al., 2023; Atkin et al., 2022; Brucks and Levav, 2022; Yang et al., 2022; Pentland, 2012), partly due to improved learning (Emanuel et al., 2023; van der Wouden and Youn, 2023; Eckert et al., 2022; Akcigit et al., 2018; De La Roca and Puga, 2017; Glaeser and Mare, 2001). Still, Chen et al. (2022) find that the costs of distributed teams tend to fall over time as remote collaboration technology improves and learning effects materialize and Forman and van Zeebroeck (2012) show internet adoption leads to more geographically dispersed inventor teams.

**Geography, gravity, and border effects** There is a large literature examining the determinants of geographic distribution of economic activity. Large parts of this literature center around the gravity model (Tinbergen, 1962; Bergstrand, 1985) that considers geographic distance and size to empirically explain economic exchange, most prominently trade (Anderson, 1979; Eaton and Kortum, 2002; Disdier and Head, 2008; Head and Mayer, 2010), but also knowledge flows (Bahar et al., 2022; Montobbio and Sterzi, 2013; Picci, 2010), foreign aid (Alesina and Dollar, 2000), online behaviour (Steegmans and de Bruin, 2021), or migration (van der Kamp, 1977; Lewer and van den Berg, 2008). For trade, the impact of distance has fallen steadily over time (Yotov, 2012), especially between rich countries (Brun et al., 2005). Blum and Goldfarb (2006) were first to show that the gravity model holds even for digital goods, where there are no trade costs, but also find no distance effect for non-taste dependent products such as software. Hanson and Xiang (2011) confirm gravity for movie exports, another product with no trade or transport costs. In contrast, Lendle et al. (2016) find distance irrelevant in e-commerce. Virtual proximity is positively associated with services trade (Hellmanzik and Schmitz, 2016, 2015) and investment (Hellmanzik and Schmitz, 2017). Recent evidence from gravity applications for developer collaboration shows smaller effects of distance globally when compared to trade (Fackler and Laurentsyeva, 2020) and a negligible distance effect for the US but significant colocation effects (Goldbeck, 2023).

Within the gravity framework, McCallum (1995) was first to explicitly estimate border effects for trade and Anderson and van Wincoop (2003) refines the empirical model and provides theoretical foundations. There is vast empirical evidence on border effects in trade (e.g., Head and Mayer, 2021; Havranek and Irsova, 2017; Anderson et al., 2014; Millimet and Osang, 2007; Chen, 2004; Helliwell and Verdier, 2001; Wolf, 2000) with recent work on European international borders (Santamaría et al., 2023a,b) pointing to still very large effects. In comparison, investigations of the border effect in collaboration and knowledge flows are relatively scant. Singh and Marx (2013) find significant but diminishing border effects in patent collaboration. However, Li (2014) shows that the decrease over time is driven by age effects. Griffith et al. (2011) point out that the speed of patent citations as measure for knowledge spillovers steadily increased with improved ICT and travel cost reductions.

**Cultural proximity in the knowledge economy** A growing strand of literature studies the role of cultural factors as deep determinants of economic activity (Alesina and Giuliano, 2015; Guiso et al., 2006). Considering cultural factors in gravity applications is widely established. Deardorff (1998) distinguishes trade barriers related to transport costs and unfamiliarity. Since then, the gravity literature routinely found cross-country cultural factors important determinants of trade (e.g., Gokmen, 2017; Felbermayr and Toubal, 2010; Boisso and Ferrantino, 1997) and other economic outcomes including innovation (e.g., Gorodnichenko and Roland, 2017), collaboration (e.g., Bercovitz and Feldman, 2011; Cummings and Kiesler, 2007; Hinds and Bailey, 2003), and productivity (e.g., Stewart and Gosain, 2006). Since culture is a fuzzy concept, the literature investigates more tractable sub-dimensions of culture such as preferences (Kondo et al., 2021; Guiso et al., 2009; Huang, 2007), institutions (Hoekman et al., 2010; Acemoglu et al., 2005), shared history (Alesina and Dollar, 2000), social ties (Bailey et al., 2021; Agrawal et al., 2006), or language (Visser, 2019; Falck et al., 2012; Melitz, 2008; Baier and Bergstrand, 2007).

Cultural factors play an important role in knowledge-intensive and innovative sectors, as well. Several studies identify common language as important, e.g., for effective team communication (Koçak and Puranam, 2022), research performance (Cao et al., 2024), or knowledge transfer (Parrotta et al., 2014). Gomez-Herrera et al. (2014) study e-commerce and also find linguistic border important but no difference in the border effect compared to offline trade. 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). Organizational links (Duede et al., 2024; Fadeev, 2023; Adams et al., 2005) as well as immigration (Tadesse and White, 2010) attenuate negative border effects associated with culture. Specifically for (open-source) software development, existing works in the organizational economics literature study culture extensively. For example, Engelhardt and Freytag (2013) shows that cultural and institutional factors explain software developers' open-source software (OSS) activity differences across countries. OSS activity differences are partly driven by social identity (Bagozzi and Dholakia, 2006) and intellectual property rights (O'Mahony, 2003), and Stewart and Gosain (2006) show shared values make OSS teams more effective. Furthermore, culturally diverse teams are associated with improved performance (Ren et al., 2016; Daniel et al., 2013; Page, 2010; van Knippenberg and Schippers, 2007) and creativity (Jang, 2017), at least up to a certain threshold (Ren et al., 2016; van Knippenberg and Schippers, 2007).

# 3 Data

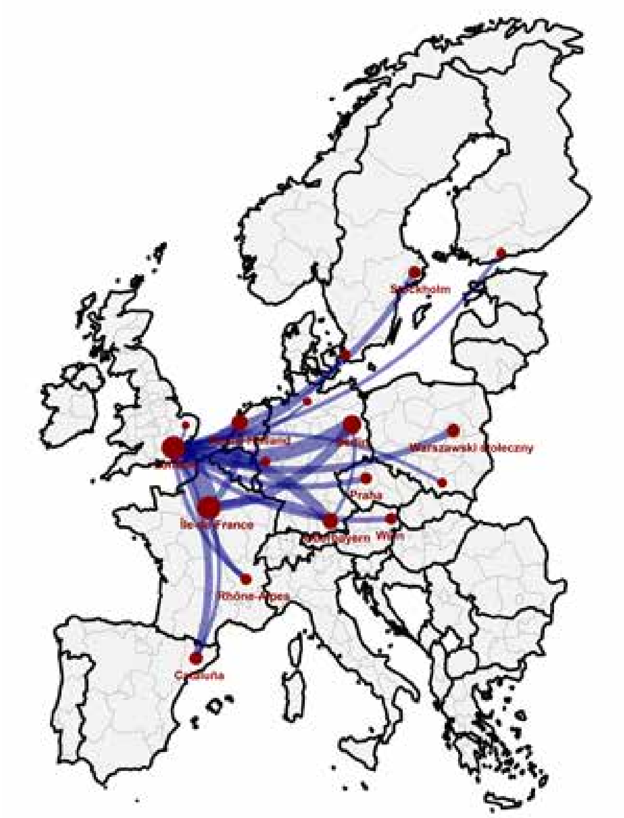
**Virtual collaboration** We compute regional collaboration networks of software developers on *GitHub*, the by far largest online code repository platform with about 73 million users worldwide in 2021 (GitHub, 2021). To this end, we draw on the *GHTorrent* database by Gousios (2013), which mirrors the data publicly available via the *GitHub* API and generates a queryable relational database in irregular time intervals.[1] This paper relies on ten *GHTorrent* snapshots dated between 09/2015 and 03/2021, which contain data from public user profiles and repositories as well as a detailed activity stream capturing all contributions to and events in open-source repositories.[2] *GitHub* projects ("repositories") are maintained using the integrated version control software *git*. Importantly, the nature of the *git* version control system allows us to observe

[1] *GHTorrent* data contains potentially sensitive personal information. Information considered sensitive (e.g., e-mail address or user name) has been de-identified (i.e., recoded as numeric identifiers) by data center staff prior to data analysis by the author. Data from the *GHTorrent* project is publicly available at `ghtorrent.org`.

[2] Snapshots are dated 2015/09/25, 2016/01/08, 2016/06/01, 2017/01/19, 2017/06/01, 2018/01/01, 2018/11/01, 2019/06/01, 2020/07/17, and 2021/03/06.

each users' activity and collaborators in public repositories. Additionally, users can indicate their location on their *GitHub* profile. We assign users to cities via exact matching to city names in the *World Cities Database*. Goldbeck (2023) validates the location information using various benchmarks, finding no systematic bias at the regional and region-pair level. Defining a collaboration as active contribution during the observation period to at least one joint project, we compute the regional collaboration network at the NUTS2 level.[3]

**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.

Overall, our data contains 290 NUTS2 regions in 34 European countries[4] and captures the activity in open-source repositories of 144,121 active, geolocated, and collaborating users. Users are highly concentrated

[3] 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.

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

in space with 39% of users in the ten largest regions.[5] The London metro area is by far the biggest region with more than 19,000 users, followed by Paris metro (Île-de-France) with 11,496 and Amsterdam metro (Noord-Holland) with 4,794. The left map in Figure A.2 shows the spatial distribution of users across European regions. Generally, this pattern is also reflected in the regional collaboration patterns depicted in Figure 1, which shows the most important nodes and edges in the regional collaboration network. The red nodes are scaled by the total number of collaborations and edge width represents bi-regional collaboration intensity. London as the central hub for software development in Europe is clearly visible and we observe most collaborations between the large cities in terms of the number of software developers. We are interested in the border effect, i.e., the relation of international versus national collaborations after controlling for geographic factors in a gravity framework. Figure A.1 plots distance histograms for cross-border and within-country network edges and shows there is a large region of common support in the distributions to facilitate robust estimation.

**Cultural proximity** We associate potential border effects to various measures of cultural proximity, drawing on multiple data sources. First, we use a composite measure of cultural proximity derived from detailed data on online behaviour (Obradovich et al., 2022). This large-scale data collection effort systematically queries the *Facebook* marketing API to dissect societies' interests along hundreds of thousands dimensions. The API offers insights derived from users' self-reported interests, clicking behaviours and likes on the platform, as well as software downloads and behaviour on other websites employing *Facebook* ads. Due to the large number of active users on *Facebook* and the representativeness of in-sample users to the general population (Bailey et al., 2018), this source provides insight into cultural differences at unprecedented scale. Specifically, from the universe of *Wikipedia* articles on *DBpedia*, Obradovich et al. (2022) extract 60,000 interest dimensions with at least 500,000 users worldwide to create a composite as well as sub-indices for cultural proximity as cosine distance between the interest vectors of populations $k$ and $l$

$$\cos\text{dist}(k,l) = 1 - \cos(\theta) = 1 - \frac{S_k * S_l}{\|S_k\|\|S_l\|} \tag{1}$$

where $S_k$ denotes a $n$-dimensional vectors with components $s_{ik}$ that measure the share of population k holding a particular interest $i = 1, \ldots, n$ and $\theta$ is the angle between $S_k$ and $S_l$. Consequently, the resulting index is independent of $n$. Obradovich et al. (2022) validate this composite index using traditional composite measures of culture and find a high overlap. Still, their index improves in granularity and represents a bottom-up approach in contrast to top-down measurement along few select dimensions. We use the cross-country composite measure as well as the sub-indices for the 14 main interest dimensions.

Second, we relate border effects to genetic distance, a well-established proxy for cultural factors associated with ethnicity (Spolaore and Wacziarg, 2009; Creanza et al., 2015). We use the cross-country genetic distance data from Creanza et al. (2015), which measures the degree of similarity in vertically transmitted

[5] Note, however, that this concentration is much less pronounced than in the US where this number is 79% (Goldbeck, 2023).

characteristics as aggregated differences in allele frequencies for highly predictive parts of a chromosome. In particular, we follow the literature and use the co-ancestor coefficients (also: $F_{ST}$ distance) that is based on heterozygosity, i.e., the probability of two specific areas of genes being different. By this measure, we proxy for co-ancestral distance between national populations, a measure found highly relevant for economic outcomes (see, e.g., Bove and Gokmen, 2018; Spolaore and Wacziarg, 2009).

Third, we account for important cultural factors traditionally used in the gravity literature and captured in the *CEPII Gravity Database* (Conte et al., 2022). As language is commonly found to be an important factor for collaboration, we use the indicator for common spoken language (Melitz and Toubal, 2014). Likewise, we control for religious proximity measured as the product of the shares of Catholics, Protestants, and Muslims in origin and destination countries (Disdier and Mayer, 2007; La Porta et al., 1999). As measures for a shared history we account for two factors: whether countries ever were part of the same nation, and whether they have a colonial history, both sourced form the *CEPII GeoDist Database* (Mayer and Zignago, 2011).

Fourth, we assess the relationship to traditional survey-based cultural dimensions as measured in the Hofstede model Hofstede (2011) and the *Global Preferences Survey* (Falk et al., 2018). The Hofstede model measures national cultural dimensions quantitatively along six dimensions: power distance, uncertainty avoidance, individualism/collectivism, achievement and success, long/short-term orientation, and indulgence/restraint. The *Global Preferences Survey* elicits cross-country differences in preferences along the six dimensions patience, risk taking, positive/negative reciprocity, altruism, and trust.

**Supplementary data** We further use regional-level social connectedness measures derived from *Facebook* (Bailey et al., 2018) to investigate potential mechanisms of collaboration. For better comparability, we compute the *GH Connectedness Index* (GHCI) (Goldbeck, 2023) similarly to the *Social Connectedness Index* (SCI) as the relative probability of connection between users in two regions

$$\text{index}_{i,j} = \frac{\text{links}_{i,j}}{\text{users}_i * \text{users}_j}, \tag{2}$$

and scale between 1 and 1,000,000,000. Note that these indices are independent of regions size by design. Furthermore, we use various additional variables traditionally used in gravity applications from the *CEPII Gravity Database* (Conte et al., 2022). In addition, we use *Fraser Institute's Economic Freedom of the World Index* and the *Freedom House Index of Political Rights* from Graafland and de Jong (2022) and compute bilateral differences in these indices.

# 4 Empirical model

To estimate border effects in software developer collaboration, we deploy the gravity model, which is widely used to explain economic outcomes like migration, trade, and FDI between countries (see, e.g., van der

Kamp, 1977; Anderson, 1979; Frankel and Rose, 2002). In the innovation literature, the gravity model is applied to describe knowledge flows and collaboration measured through patenting activity (e.g., Bahar et al., 2022; Montobbio and Sterzi, 2013; Picci, 2010). While traditionally applied in cross-country settings the model is equally suitable at the sub-national regional level, where it is used to estimate border effects (e.g., Anderson and van Wincoop, 2003; Wolf, 2000; McCallum, 1995). Note that border effects gravity models are theory-consistent and, because they feature domestic flows by design, even more so than traditional cross-country gravity (Yotov, 2022). In our context, the gravity model, in its simplest form, states that regional collaboration is proportional to the product of the regions' masses (measured by the number of local users) and inversely proportional to the distance between the regions. We take the parsimonious gravity model from McCallum (1995), which includes an indicator for cross-border collaboration, as starting point for estimating the border effect:

$$\ln(y_{i,j}) = \beta_0 + \beta_1 \text{crossborder}_{i,j} + \beta_2 \text{coloc}_{i,j} + \beta_3 \ln(\text{dist}_{i,j}) + \delta_i + \delta_j + \varepsilon_{i,j} \tag{3}$$

where $y_{i,j}$ represents the number of bilateral collaborations between regions $i$ and $j$ including domestic collaborations $i = j$. The dummy variable $\text{crossborder}_{i,j}$ indicates if region $i$ is located in a different country than region $j$, and $\text{dist}_{i,j}$ denotes the geographic distance between the regions. We further add a colocation indicator, $\text{coloc}_{i,j}$, to account for strong colocation effects in collaboration (Goldbeck, 2023; Urry, 2002; Olson and Olson, 2000). Origin and destination fixed effect $\delta_i$ and $\delta_j$ account for unobserved regional determinants of collaboration common across all partner regions. The coefficient $\beta_2$ captures the elasticity of collaboration with respect to geographic distance, which we expect to be negative from theory. The border effect is given by our coefficient of interest $\beta_1$, which we expect to be negative or zero, depending on the presence of a border effect in the population.

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 3 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 this broadly defined border effect we extend the baseline model to include variables at the country-pair level measuring different cultural dimensions that vary across borders:

$$\ln(y_{i,j}) = \beta_0 + \beta_1 \text{crossborder}_{i,j} + \beta_2 \text{coloc}_{i,j} + \beta_3 \ln(\text{dist}_{i,j}) + \mathbf{X}'_{c(i),c(j)}\boldsymbol{\beta_4} + \mathbf{X}'_{i,j}\boldsymbol{\beta_5} + \delta_i + \delta_j + \varepsilon_{i,j} \tag{4}$$

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, i.e., $c(i) = c(j)$. Thus, the coefficients $\beta_4$ capture the part of the border effect that is attributable to a particular cross-border difference while $\beta_1$ is the residual part of the average border effect not explained by the included variables in $\mathbf{X}_{c(i),c(j)}$. $\mathbf{X}_{i,j}$ is a vector of region-pair level determinants of collaboration and $\beta_5$ are the related coefficients.

As in the baseline model, the main assumption for causal interpretation of the coefficients $\beta_4$ is that there are no omitted factors related to $\mathbf{X}_{c(i),c(j)}$ that determine inter-regional collaboration. Note that the cross-border indicator isolates the remaining part of the border effect and therefore provides indication for the presence of omitted variables 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 $\beta_4$.

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 determinants as it points to dimensions that are statistically associated with the border effect. Plausible, theory-guided selection of explanatory variables is therefore paramount to avoid spurious correlation issues. We return to this discussion in subsection 5.3. Note that Equations 3 and 4 are partial equilibrium models and, as such, estimated border effects should not be misconstrued as counterfactual for border removal, as wiedely acknowledged in the literature (see, e.g., Santamaría et al., 2023a; Havranek and Irsova, 2017).

# 5 Results

## 5.1 Digital border effect

Table 1 reports estimation results of the border effect for online collaboration among software developers in Europe. We start with a model that does not consider gravity and subsequently control for size and geographic distance. The raw correlation in model (1) suggests a large border effect of 60% less collaborations. Controlling for size in terms of logarithms of multiplied user bases in origin and destination regions halves the effect. The large positive coefficient on multiplied user bases demonstrates the importance of collaboration potential. Model (3) drops the functional form assumption for the size effect and instead includes unobserved regional characteristics using origin and destination region FE. This more flexible model slightly increases 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. We include logarithmic distance between origin and destination region centroids as specified in Equation 3. Since our data features within-region collaborations and Goldbeck (2023) finds colocation hugely important for collaboration, we also add a colocation indicator. As expected, results show a highly significant negative relation of collaboration and distance and a substantial collaboration premium for colocation.

**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.

There still is a border effect in our preferred baseline specification, with 16.4% fewer collaborations for region-pairs that are located in different countries compared to within-country pairs. While the border effect is economically significant, it is much smaller than for trade. The meta-analysis by Havranek and Irsova (2017) aggregates 263 estimates for the EU from similar gravity model specifications and finds a border effect of –91.5%[6], a slightly smaller effect size than the original estimates of McCallum (1995) and nearly identical to the border effect for Europe in Santamaría et al. (2023b) of –90.4%[7] estimated from recent granular freight data. Thus, a comparison to their results suggests a five to six times larger border effect in (goods) trade compared to (online) 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. Still, there is significant heterogeneity in the border effect. Table A.2 demonstrates that the border effect is systematically related to the number of country-wide users. Model (2) shows the border

[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 $(\exp(-2.34) - 1) * 100$.

effect roughly doubles when a small country is involved, defined as hosting an above-median number of users. Model (3) shows the effect does not differ depending on whether both countries are small or just one, meaning there is a smaller border effect among large countries.

### 5.2 The role of culture

As there still is a significant border effect present in virtual collaboration, we investigate potential channels through which cross-border collaboration of software developers might be affected. We elicit association of various cultural factors with the border effect and collaboration by including appropriate cross-country level variables as specified in Equation 4.

Table 2 reports the results of variations of our baseline model 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, 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 interests on *Facebook* and validated using traditional, mostly survey-based, metrics of culture. Second, we control for genetic distance from Spolaore and Wacziarg (2009) as a well-established proxy for cultural factors associated with ethnicity. The coefficient estimates of both distance measures have the expected negative sign. Cultural distance is strongly negatively associated with collaobration 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 that have been identified as relevant in the previous literature, namely common language, religious distance, and a common history reflected by same country or colonial history. Religious distance is statistically and economically insignificantly related to collaboration.[8] In contrast, there appears to be a sizable benefit from common spoken language of around 8.4% more collaborations, although imprecisely estimated. On the one hand, this makes sense as it eases communication. On the other hand, most knowledge work professionals speak English and code projects in software development are written in computer code. Reassuringly, 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).[9] A shared colonial history is often highly predictive in gravity models but does not explain collaboration today. This is likely due to the few colonial relationships within Europe. History as a same country is associated negatively with collaboration, which is surprising only at first and likely relates to the fact that this indicator

[8]Note that this might reflect that religious differences in Europe are generally small.

[9]Cf. column (2) in Table 3 of Melitz and Toubal (2014). Note that estimate magnitudes for common (spoken) language in log-specifications are generally quite robust in the trade literature (Melitz, 2008). Yet, most semi-elasticities refer to a worldwide sample. Still, estimates for European samples are comparable in size (see, e.g., Fidrmuc and Fidrmuc, 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.

captures mostly historical occupations in the former Yugoslavia and Austria-Hungary that lead to disrupted relationships until today (e.g., Kešeljević and Spruk, 2023).

Model (4) additionally adds social connectedness between regions as explanatory variable for collaboration. Social connectedness is highly statistically and economically significantly and positively related to collaboration. Controlling for social connectedness leads to irrelevance of geographic distance and a smaller 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. Note, however, that social connectedness might constitute a bad control in our setting as it likely is determined by cultural factors, as well. Therefore our preferred specification is model (3). While the relevance of colocation remains highly important throughout all specifications, geographic distance is statistically significant at a lower level and the coefficient size shrinks considerably. This is in line with empirical evidence on knowledge worker collaboration suggesting a high relevance of face-to-face meeting possibility (e.g., Emanuel et al., 2023; Atkin et al., 2022) but irrelevance of geographic distance otherwise (cf. 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 (see, e.g., Waldinger, 2012; Azoulay et al., 2010).

We further investigate the relation between culture and international collaboration using established frameworks for particular cultural dimensions. First, we exploit 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 with respect to industries and companies. It is, however, unclear if common professional interests are responsible for increased collaboration or if the presence and relation to local industries are a common driver of both collaboration and interests. Existing literature points toward an important role of organizations in shaping software developer collaboration (e.g., Duede et al., 2024; Goldbeck, 2023). Other subcategories are relatively unimportant, but mostly show positive associations. This points to cultural differences not being unidimensionally negatively related to collaboration but rather paints a more nuanced picture that some cultural differences, e.g. with respect to food or lifestyle, might in fact spur collaboration.

Second, we explore how cross-country differences in preferences relate to international collaboration. To this end, we use the six preference dimensions from the *Global Preferences Survey*: patience, risk taking, trust, altruism as well as positive and negative reciprocity. Table A.4 reports the results and shows that especially patience and positive reciprocity are negatively related to collaboration. Negative reciprocity explains collaboration to a lesser extent and is only weakly significant and the other dimensions are statistically insignificant, although point estimates are negative throughout. Generally, cross-country differences in preferences partly explain the border effect but only to a small extent.

Third, we use the established traditional cross-country measures of culture by Hofstede (2011) to study possible associations with collaboration. Of the six standard dimensions (power distance, individualism, achievement and success, uncertainty avoidance, long-term orientation, and indulgence), only power distance is significantly and negatively related to collaboration as shown in Table A.5. Individualism is also negatively related to collaboration but only weakly significant. Overall, the Hofstede cultural dimensions do not prove useful to explain the border effect as the point estimate is only slightly reduced when including differences in the six cultural dimensions.

### 5.3 Robustness

We demonstrate the robustness of the digital border effect estimated in Table 1 in multiple ways. First, we follow the methodology in Santamaría et al. (2023b) and compute an independence benchmark that disregards everything but the size component of gravity. This essentially corresponds to a theory in which all user-pairs feature equal probability of collaboration independent of their locations. We then relate observed collaborations to the benchmark in Panel (a) of Figure A.9 and distinguish cross-border, within-country, and colocated links. This shows the strong predictive power of the logarithmic multiplication of region size in terms of users. It is reassuring that the relationship between collaboration potential measured by multiplied user size is not significantly different between cross-border and within-country collaborations. Importantly, the analysis confirms that collaboration probability is significantly increased for within-country compared to cross-border collaborations, depicted by a shift to the right of the distribution in Panel (b) of Figure A.9.

Second, we plot residuals of fixed-effects models disregarding the cross-border indicator in Figure A.10. Panels (a) and (b) plot the averages and distributions of residuals for cross-border and within-country collaborations for the baseline fixed-effects model without and with geography controls, respectively. We generally observe well-behaved residual distributions, which is reassuring of our model specification. The significant right-shift of the residual distribution for within-country collaborations points to omitted variables bias in models that disregard border effects and, therefore, the presence of border effects in virtual collaboration. The narrowing of this gap between the distributions in Panel (b) compared to Panel (a) while still retaining statistical significance shows that geographic factors are important but do not fully explain the raw border effect. This is corroborated by models featuring a non-parametric distance specification. Figure A.5 compares non-parametric models with and without the cross-border indicator. Results show that considering the cross-border indicator significantly flattens the distance gradient and decreases the colocation effect.

Third, we calculate the size-independent *GH Connectedness Index* (GHCI) (Goldbeck, 2023), which is similar to the *Social Connectedness Index* (SCI) by Bailey et al. (2018), and directly plot the relation to distance for within-country and cross-border links, respectively, in Figure A.6. As depicted in Figure A.7, GHCI and SCI feature similar distributional shapes, but are unrelated at the region-pair level (Figure A.8). Generally, the relationships of the within-country and cross-border GHCI to distance are largely overlapping, i.e., have significant common support, and a border effect for software developers is not clearly visible. This is due to the relatively small size of the border effect that, in fact, is statistically highly significant. In contrast, for the SCI there is a magnitudes larger and visually easily identifiable upwards shift for within-country collaborations. In line with expectations, this comparison suggests that the border effect in virtual collaboration of knowledge workers is much smaller compared to the border effect present in social networks, which is reassuring of our analysis.

Although cultural factors explain the border effect in Europe well, our parsimonious gravity model does not allow causal interpretation. Still, model fit and explanatory power point to cultural proximity as important driver of virtual collaboration. To strengthen the conjecture that culture plays an important role as deep determinant of (online) collaboration, we compare border effects in software developer collaboration for European nations and US states (Figure A.4). The idea is that there are far fewer and less pronounced cultural differences across populations in different US states than in culturally much more diverse European countries. Thus, we use the same data on regional collaboration in the US at the economic-area level from Goldbeck (2023) and estimate the state border effect using the same approach as in Table 1. Table A.6 reports the results. The raw border effect, disregarding geographic factors, in the US in model (1) is 0.69 of the European estimate. Similarly, the preferred specification that takes into account size and distance in the US is 0.58 the size of the border effect in Europe, as shown by model (4). This is in line with expectations of cultural factors such as language barriers as a key determinant of the digital border effect.

Further, we assess the robustness of the coefficient estimates for the culture variables in Tables A.7 and A.8. We demonstrate that all estimates remain stable when we include various other potential control variables, e.g., regarding historical and political circumstances. Table A.7 shows robustness with respect to inclusion of contiguity, an indicator for a common border, a common control variable in gravity models that theoretically should be irrelevant in our setting. Models (2) through (7) demonstrate that all estimates remain stable when controlling for a common legal origin and shared communist history. Coefficients are similarly stable when including further control variables for political circumstances in Table A.8. For example, we account for a diplomatic disagreement score, EU membership, regional trade agreements, hegemonic relationship, relationships between monarchies as well as differences in economic and press freedom scores. Again, our coefficient estimates remain robust throughout all specifications.

In Table A.9, we examine different alternative measures for language and religion. Similarly to the trade literature (e.g., Melitz and Toubal, 2014), where continuous language proximity variables show weaker relation to trade, we find only common spoken language relevant to collaboration. Various other metrics such as other binary indicators like common native language but also continuous metrics of linguistic proximity are insignificant. This is in line with expectations that only speaking the exact same language benefits collaboration and closely related but still different languages have no impact. Model (7) in Table A.9 switches to an alternative continuous metric for religion that uses a different methodology but is also insignificantly related to collaboration. Importantly, the other coefficients remain robust and largely unchanged throughout all specifications.

# 6 Discussion and 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 digital border effect is particularly high whenever a small country, in terms of hosted users, is involved. Generally, the remaining border effect in software developer collaboration in Europe is entirely explained by cultural factors, especially shared 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 settings 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 hard to causally explore the intricate patterns of mediation and co-determination among the countless cultural factors. Further, our data contains information on public repositories only. While the geographical collaboration pattern is representative of the entire population of software developers (Goldbeck, 2023), it is less clear if the relationship between cultural factors and collaboration differs between open- and closed-source developers. Ideally, the measurement of culture is conducted on a more granular scale both population-wise and geographically as, e.g., software developers might be different to the general population.

Our work has several practical implications relevant to management and policy makers. Importantly, we show that there is a significant border effect for international collaboration of developers on online code repository platforms. Still, the digital border effect is much smaller compared to other outcomes, which generally points to improved feasibility of international collaboration in digital knowledge work. Since the digital border effect is entirely explained by cultural factors, they merit more attention. Together with decreasing role of geography in ICT-intensive settings of the knowledge economy this suggests that management and policy makers should shift their attention to cultural barriers to collaboration as they are relatively more important in the digital economy when fully virtual collaboration is technically possible.

## References

**Acemoglu, Daron, Simon Johnson, and James A. Robinson**, "Institutions as a Fundamental Cause of Long-run Growth," *Handbook of Economic Growth*, 2005, *1*, 385–472.

**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.

_ **, Christian Catalini, and Avi Goldfarb**, "Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions," *Journal of Economics & Management Strategy*, 2015, *24* (2), 253–274.

_ **, 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.

**Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi**, "Dancing with the Stars: Innovation through Interactions," *NBER Working Paper*, 2018.

**Akerman, Anders, Edwin Leuven, and Magne Mogstad**, "Information Frictions, Internet, and the Relationship between Distance and Trade," *American Economic Journal: Applied Economics*, 2022, *14* (1), 133–163.

**Alesina, Alberto and David Dollar**, "Who Gives Foreign Aid to Whom and Why?," *Journal of Economic Growth*, 2000, *5*, 33–63.

_ **and Paola Giuliano**, "Culture and Institutions," *Journal of Economic Literature*, 2015, *53* (4), 898–944.

**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.**, "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 1979, *69* (1), 106–116.

_ **, 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.

**Bagozzi, Richard P. and Utpal M. Dholakia**, "Open Source Software User Communities: A Study of Participation in Linux User Groups," *Management Science*, 2006, *52* (7), 1099–1115.

**Bahar, Dany, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W. Koo**, "Innovation on Wings: Nonstop Flights and Firm Innovation in the Global Context," *Management Science*, 2022.

**Baier, Scott L. and Jeffrey H. Bergstrand**, "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 2007, *71* (1), 72–95.

**Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel**, "International Trade and Social Connectedness," *Journal of International Economics*, 2021, *129*, 103418.

**_ , 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.

**Bergstrand, Jeffrey H.**, "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *The Review of Economics and Statistics*, 1985, pp. 474–481.

**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.

**Boisso, Dale and Michael Ferrantino**, "Economic Distance, Cultural Distance, and Openness in International Trade: Empirical Puzzles," *Journal of Economic Integration*, 1997, pp. 456–484.

**Bove, Vincenzo and Gunes Gokmen**, "Genetic Distance, Trade, and the Diffusion of Development," *Journal of Applied Econometrics*, 2018, *33* (4), 617–623.

**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.

**Brucks, Melanie S. and Jonathan Levav**, "Virtual Communication Curbs Creative Idea Generation," *Nature*, 2022, *605* (7908), 108–112.

**Brun, Jean-Francois, Céline Carrère, Patrick Guillaumont, and Jaime De Melo**, "Has Distance Died? Evidence from a Panel Gravity Model," *The World Bank Economic Review*, 2005, *19* (1), 99–120.

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

**Chan, Kalok, Vicentiu Covrig, and Lilian Ng**, "What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide," *The Journal of Finance*, 2005, *60* (3), 1495–1534.

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

**Chen, Natalie**, "Intra-national Versus International Trade in the European Union: Why Do National Borders Matter?," *Journal of International Economics*, 2004, *63* (1), 93–118.

**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.

**Daniel, Sherae, Ritu Agarwal, and Katherine J. Stewart**, "The Effects of Diversity in Global, Distributed Collectives: A Study of Open Source Project Success," *Information Systems Research*, 2013, *24* (2), 312–333.

**De La Roca, Jorge and Diego Puga**, "Learning by Working in Big Cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

**Deardorff, Alan**, "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?," in "The Regionalization of the World Economy," University of Chicago Press, 1998, pp. 7–32.

**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 Keith Head**, "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *The Review of Economics and Statistics*, 2008, *90* (1), 37–48.

— **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.

**Eaton, Jonathan and Samuel Kortum**, “Technology, Geography, and Trade,” *Econometrica*, 2002, *70* (5), 1741–1779.

**Eckert, Fabian, Mads Hejlesen, and Conor Walsh**, “The Return to Big-City Experience: Evidence from Refugees in Denmark,” *Journal of Urban Economics*, 2022, *130*, 103454.

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

**Fackler, Thomas and Nadzeya Laurentsyeva**, “Gravity in Online Collaborations: Evidence from GitHub,” *CESifo Forum*, 2020, *21* (03), 15–20.

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

**Falck, Oliver, Stephan Heblich, Alfred Lameli, and Jens Südekum**, “Dialects, Cultural Identity, and Economic Exchange,” *Journal of Urban Economics*, 2012, *72* (2-3), 225–239.

**Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde**, “Global Evidence on Economic Preferences,” *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

**Felbermayr, Gabriel and Farid Toubal**, “Cultural Proximity and Trade,” *European Economic Review*, 2010, *54* (2), 279–293.

**Fidrmuc, Jan and Jarko Fidrmuc**, “Foreign Languages and Trade,” *CESifo Working Paper*, 2014.

**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).

**Frankel, Jeffrey and Andrew Rose**, “An Estimate of the Effect of Common Currencies on Trade and Income,” *The Quarterly Journal of Economics*, 2002, *117* (2), 437–466.

**French, Kenneth R. and James M. Poterba**, “Investor Diversification and International Equity Markets,” *American Economic Review*, 1991, *81* (2), 222.

**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.

**Gaspar, Jess and Edward L. Glaeser**, "Information Technology and the Future of Cities," *Journal of Urban Economics*, 1998, *43* (1), 136–156.

**Gibbs, Michael, Friederike Mengel, and Christoph Siemroth**, "Work from Home and Productivity: Evidence from Personnel and Analytics Data on Information Technology Professionals," *Journal of Political Economy: Microeconomics*, 2023, *1* (1), 7–41.

**GitHub**, "The 2021 State of the Octoverse," 2021.

**Glaeser, Edward L. and David C. Mare**, "Cities and Skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

**Gokmen, Gunes**, "Clash of Civilizations and the Impact of Cultural Differences on Trade," *Journal of Development Economics*, 2017, *127*, 449–458.

**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.

**Griffith, Rachel, Sokbae Lee, and John van Reenen**, "Is Distance Dying at Last? Falling Home Bias in Fixed-effects Models of Patent Citations," *Quantitative Economics*, 2011, *2* (2), 211–249.

**Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 2006, *20* (2), 23–48.

— , — , **and** — , "Cultural Biases in Economic Exchange?," *The Quarterly Journal of Economics*, 2009, *124* (3), 1095–1131.

**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.

**Hanson, Gordon and Chong Xiang**, "Trade Barriers and Trade Flows with Product Heterogeneity: An Application to US Motion Picture Exports," *Journal of International Economics*, 2011, *83* (1), 14–26.

**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**, "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade," *The Gravity Model in International Trade*, 2010.

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

**Helliwell, John F. and Genevieve Verdier**, "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada," *Canadian Journal of Economics*, 2001, pp. 1024–1041.

**Hellmanzik, Christiane and Martin Schmitz**, "Virtual Proximity and Audiovisual Services Trade," *European Economic Review*, 2015, *77*, 82–101.

— **and** — , "Gravity and International Services Trade: The Impact of Virtual Proximity," *European Economic Review*, 2016, *77*, 82–101.

— **and** — , "Taking Gravity Online: The Role of Virtual Proximity in International Finance," *Journal of International Money and Finance*, 2017, *77*, 164–179.

**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.

**Hoekman, Jarno, Koen Frenken, and Robert J. W. Tijssen**, "Research Collaboration at a Distance: Changing Spatial Patterns of Scientific Collaboration Within Europe," *Research Policy*, 2010, *39* (5), 662–673.

**Hofstede, Geert**, "Dimensionalizing Cultures: The Hofstede Model in Context," *Online Readings in Psychology and Culture*, 2011, *2* (1), 8.

**Huang, Rocco**, "Distance and Trade: Disentangling Unfamiliarity Effects and Transport Cost Effects," *European Economic Review*, 2007, *51* (1), 161–181.

**Jang, Sujin**, "Cultural Brokerage and Creative Performance in Multicultural Teams," *Organization Science*, 2017, *28* (6), 993–1009.

**Jensen, Robert**, "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *Quarterly Journal of Economics*, 2007, *122* (3), 879–924.

**Kešeljević, Aleksandar and Rok Spruk**, "Long-term Effects of the Yugoslav War," *Defence and Peace Economics*, 2023, *34* (4), 410–436.

**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.

**Kondo, Jiro, Danielle Li, and Dimitris Papanikolaou**, "Trust, Collaboration, and Economic Growth," *Management Science*, 2021, *67* (3), 1825–1850.

**Korkmaz, Gizem, J Bayoán Santiago Calderón, Brandon L. Kramer, Ledia Guci, and Carol A. Robbins**, "From GitHub to GDP: A Framework for Measuring Open Source Software Innovation," *Research Policy*, 2024, *53* (3).

**Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina**, "There Goes Gravity: eBay and the Death of Distance," *Economic Journal*, 2016, *126* (591), 406–441.

**Lewer, Joshua J. and Hendrik van den Berg**, "A Gravity Model of Immigration," *Economics Letters*, 2008, *99* (1), 164–167.

**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**, "Language and Foreign Trade," *European Economic Review*, 2008, *52* (4), 667–699.

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

**Millimet, Daniel L. and Thomas Osang**, "Do State Borders Matter for US Intranational Trade? The Role of History and Internal Migration," *Canadian Journal of Economics*, 2007, *40* (1), 93–126.

**Montobbio, Fabio and Valerio Sterzi**, "The Globalization of Technology in Emerging Markets: A Gravity Model on the Determinants of International Patent Collaborations," *World Development*, 2013, *44*, 281–299.

**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.

**Olson, Gary M. and Judith S Olson**, "Distance Matters," *Human-Computer Interaction*, 2000, *15* (2-3), 139–178.

**O'Mahony, Siobhán**, "Guarding the Commons: How Community Managed Software Projects Protect Their Work," *Research Policy*, 2003, *32* (7), 1179–1198.

**Page, Scott E.**, *Diversity and Complexity*, Princeton University Press, 2010.

**Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova**, "Labor Diversity and Firm Productivity," *European Economic Review*, 2014, *66*, 144–179.

**Pentland, Alex Sandy**, "The New Science of Building Great Teams," *Harvard Business Review*, 2012, *90* (4), 60–69.

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

**Picci, Lucio**, "The Internationalization of Inventive Activity: A Gravity Model Using Patent Data," *Research Policy*, 2010, *39* (8), 1070–1081.

**Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny**, "The Quality of Government," *Journal of Law, Economics, and Organization*, 1999, *15* (1), 222–279.

**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.

**Ren, Yuqing, Jilin Chen, and John Riedl**, "The Impact and Evolution of Group Diversity in Online Open Collaboration," *Management Science*, 2016, *62* (6), 1668–1686.

**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.

**Steegmans, Joep and Jonathan de Bruin**, "Online Housing Search: A Gravity Model Approach," *PLOS One*, 2021, *16* (3), e0247712.

**Steinwender, Claudia**, "Real Effects of Information Frictions: When the States and the Kingdom Became United," *American Economic Review*, 2018, *108* (3), 657–96.

**Stewart, Katherine J. and Sanjay Gosain**, "The Impact of Ideology on Effectiveness in Open Source Software Development Teams," *MIS Quarterly*, 2006, pp. 291–314.

**Strong, Norman and Xinzhong Xu**, "Understanding the Equity Home Bias: Evidence from Survey Data," *Review of Economics and Statistics*, 2003, *85* (2), 307–312.

**Tadesse, Bedassa and Roger White**, "Cultural Distance as a Determinant of Bilateral Trade Flows: Do Immigrants Counter the Effect of Cultural Differences?," *Applied Economics Letters*, 2010, *17* (2), 147–152.

**Tinbergen, Jan**, "An Analysis of World Trade Flows," *Shaping the World Economy*, 1962, *3*, 1–117.

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

**v. Engelhardt, Sebastian and Andreas Freytag**, "Institutions, Culture, and Open Source," *Journal of Economic Behavior & Organization*, 2013, *95*, 90–110.

**van der Kamp, John**, "The Gravity Model and Migration Behaviour: An Economic Interpretation," *Journal of Economic Studies*, 1977, *4* (2), 89–102.

**van der Wouden, Frank and Hyejin Youn**, "The Impact of Geographical Distance on Learning through Collaboration," *Research Policy*, 2023, *52* (2), 104698.

**van Knippenberg, Daan and Michaela C. Schippers**, "Work Group Diversity," *Annual Review of Psychology*, 2007, *58*, 515–541.

**Visser, Robin**, "The Effect of the Internet on the Margins of Trade," *Information Economics and Policy*, 2019, *46*, 41–54.

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

**Wolf, Holger C.**, "Intranational Home Bias in Trade," *Review of Economics and Statistics*, 2000, *82* (4), 555–563.

**Yang, Longqi, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, and Jamie Teevan**, "The Effects of Remote Work on Collaboration Among Information Workers," *Nature Human Behaviour*, 2022, *6* (1), 43–54.

**Yotov, Yoto V.**, "A Simple Solution to the Distance Puzzle in International Trade," *Economics Letters*, 2012, *117* (3), 794–798.

_ , "On the Role of Domestic Trade Flows for Estimating the Gravity Model of Trade," *Contemporary Economic Policy*, 2022, *40* (3), 526–540.

# 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:** Border effect and country size

| Collaboration | (1) | (2) | (3) |
|---|---|---|---|
| Cross-border | -0.180*** | -0.133*** | -0.269*** |
| | (0.014) | (0.014) | (0.022) |
| Cross-border × small involved | | -0.155*** | |
| | | (0.012) | |
| Cross-border × both small | | | 0.034 |
| | | | (0.022) |
| Cross-border × both large | | | 0.129*** |
| | | | (0.020) |
| Colocation | 0.862*** | 0.879*** | 0.888*** |
| | (0.068) | (0.068) | (0.068) |
| Distance [log] | -0.129*** | -0.119*** | -0.120*** |
| | (0.007) | (0.007) | (0.007) |
| Origin FE | × | × | × |
| Destination FE | × | × | × |
| Observations | 84,100 | 84,100 | 84,100 |
| Adj. $R^2$ | 0.922 | 0.922 | 0.922 |

*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, 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:** Collaboration and preferences

| Collaboration | (1) | (2) | (3) |
|---|---|---|---|
| Cross-border | -0.361*** | -0.229*** | -0.158*** |
| | (0.011) | (0.012) | (0.017) |
| Colocation | | 1.310*** | 1.360*** |
| | | (0.066) | (0.068) |
| Distance [log] | | -0.044*** | -0.033*** |
| | | (0.007) | (0.007) |
| Patience | | | -0.118*** |
| | | | (0.017) |
| Risk taking | | | -0.036 |
| | | | (0.049) |
| Positive reciprocity | | | -0.094*** |
| | | | (0.034) |
| Negative reciprocity | | | -0.040** |
| | | | (0.017) |
| Altruism | | | -0.033 |
| | | | (0.027) |
| Trust | | | -0.015 |
| | | | (0.020) |
| Origin FE | × | × | × |
| Destination FE | × | × | × |
| Observations | 48,888 | 48,888 | 48,888 |
| Adj. $R^2$ | 0.951 | 0.954 | 0.955 |

*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, Falk et al. (2018), CEPII, own calculations.

**Table A.5:** Collaboration and cultural dimensions

| Collaboration | (1) | (2) | (3) |
|---|---|---|---|
| Cross-border | -0.396*** | -0.248*** | -0.221*** |
| | (0.011) | (0.012) | (0.016) |
| Colocation | | 1.312*** | 1.317*** |
| | | (0.066) | (0.067) |
| Distance [log] | | -0.048*** | -0.047*** |
| | | (0.006) | (0.007) |
| Power distance | | | -0.034*** |
| | | | (0.006) |
| Individualism | | | -0.022* |
| | | | (0.012) |
| Achievement and success | | | 0.002 |
| | | | (0.004) |
| Uncertainty avoidance | | | 0.010* |
| | | | (0.006) |
| Long-term orientation | | | -0.001 |
| | | | (0.006) |
| Indulgence | | | 0.001 |
| | | | (0.006) |
| Origin FE | × | × | × |
| Destination FE | × | × | × |
| Observations | 67,828 | 67,828 | 67,828 |
| Adj. $R^2$ | 0.939 | 0.941 | 0.941 |

*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, Hofstede (2011), own calculations.

**Table A.6:** 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.7:** 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.8:** 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.

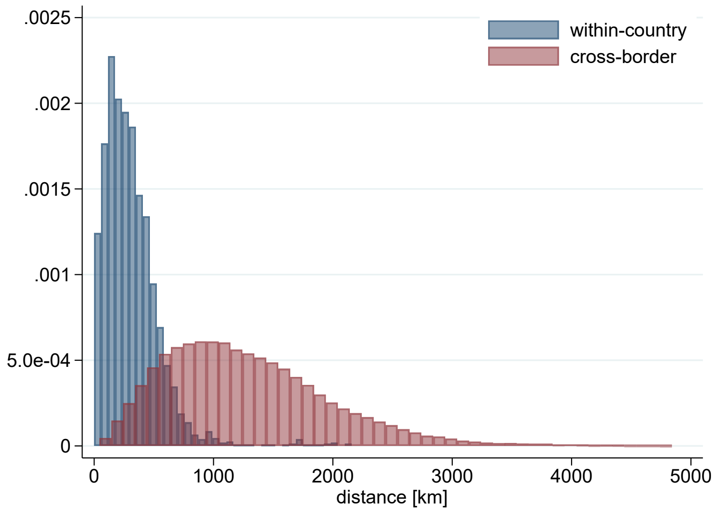
**Table A.9:** Collaboration, language, and religion

| 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.

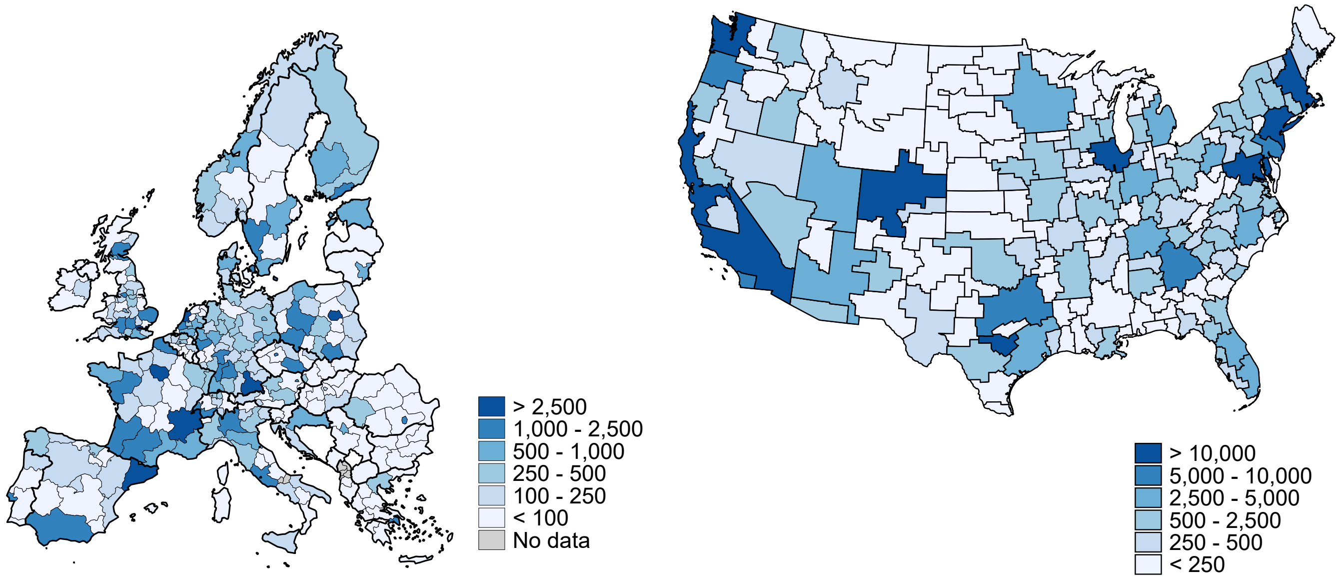
## A.2 Figures

**Figure A.1:** Distance histogram



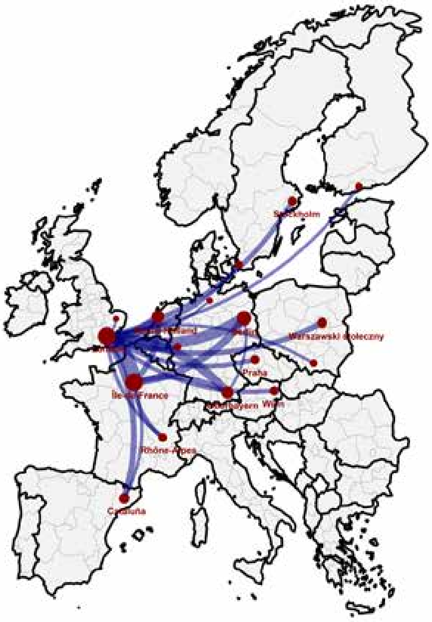

*Notes:* Figure shows histograms of within-country and cross-border distances based on NUTS2 centroids, respectively. *Sources:* GHTorrent, own calculations.

**Figure A.2:** Geographic user distribution

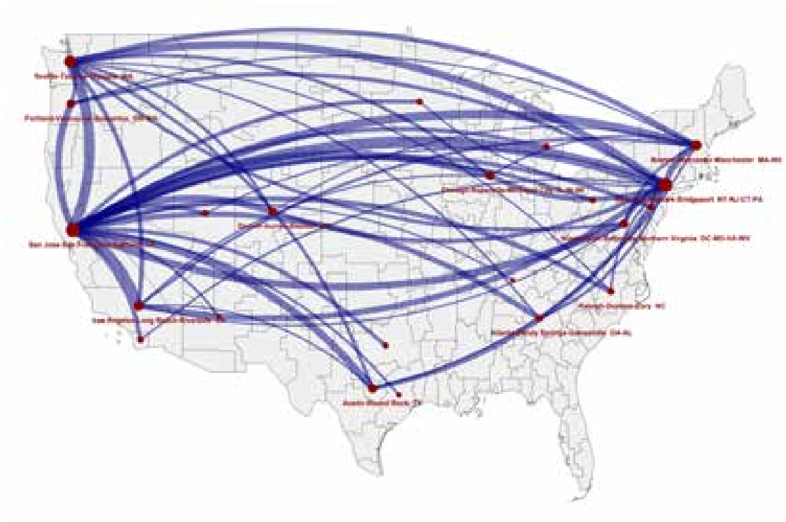



*Notes:* Maps show the number of (in-sample) users per NUTS2 region and economic area, respectively. The remote economic areas Anchorage, AK, and Honolulu, HI, as well as Ireland are not shown. *Sources:* GHTorrent, Bureau of Economic Analysis, Goldbeck (2023), own calculations.

**Figure A.3:** Inter-regional collaboration



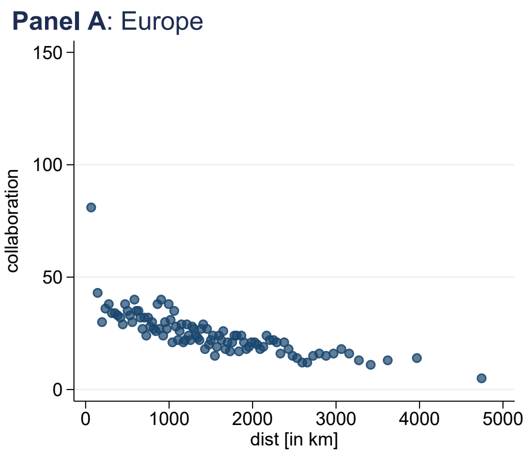

(**a**) Europe



(**b**) USA

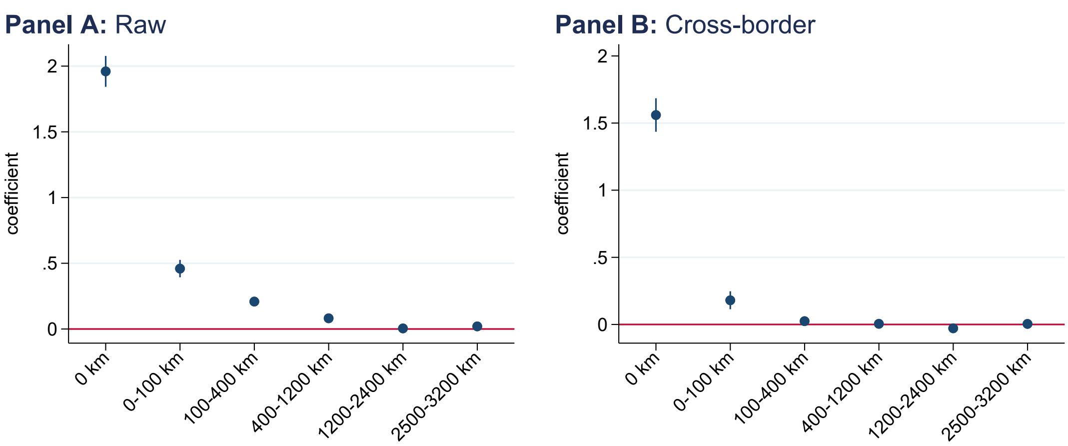
*Notes:* Maps show the structure of the European and US software developer collaboration networks, respectively. Important edges of the network, defined as links between economic areas above 25,000 connections, are shown in blue and scaled by the logarithm of the number of links. Regions are shown in gray with their centroids as nodes in red, scaled by overall links to other economic areas. The remote economic areas Anchorage, AK, and Honolulu, HI, as well as Ireland are not shown. *Sources:* GHTorrent, Bureau of Economic Analysis, Goldbeck (2023), own calculations.

**Figure A.4:** Collaboration and distance



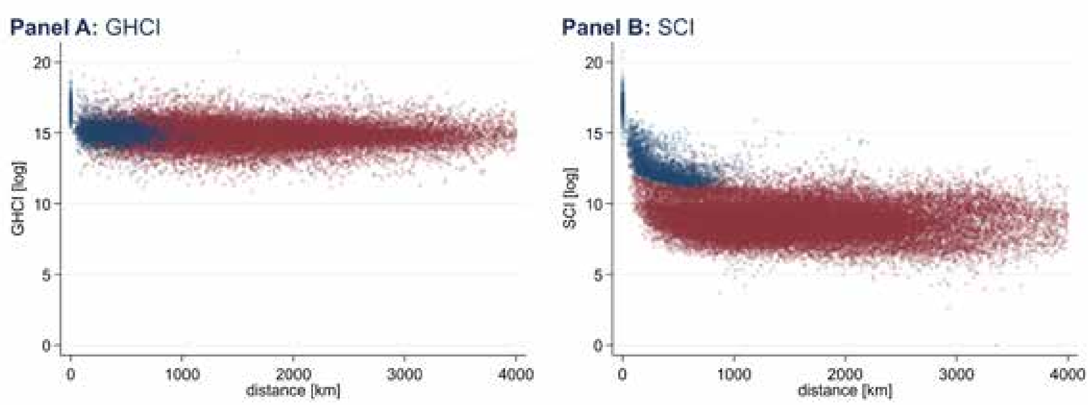

*Notes:* Panels A and B show binned scatter plots of the median number of collaborations and the geographic distance between economic-area pairs in Europe and the US, respectively. The number of bins is 100, i.e., each point represents one percentile of economic-area pairs. *Sources:* GHTorrent, own calculations.

**Figure A.5:** Non-parametric distance



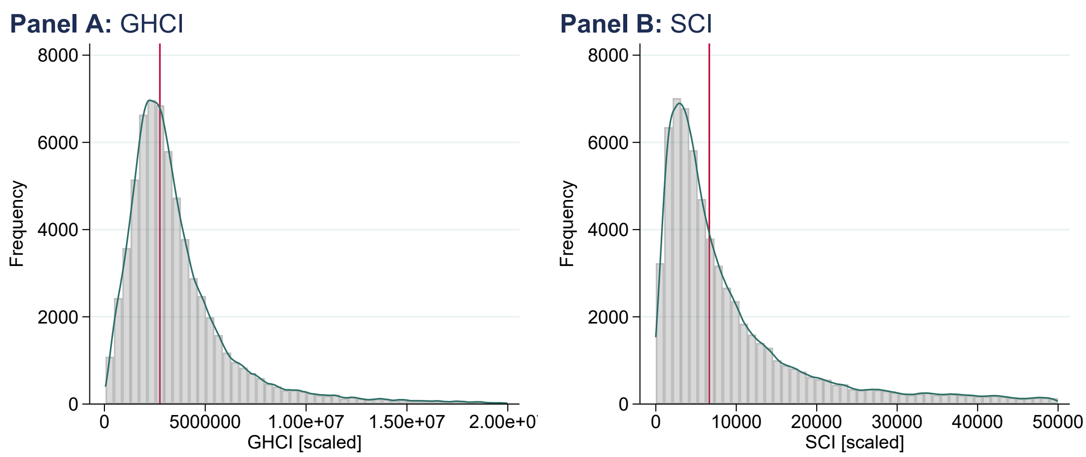

*Notes:* Plot shows coefficient point estimates and confidence intervals for the baseline fixed effects model specification with non-parametric distance. Panel A (Panel B) shows results from a specification without (with) cross-border indicator. The indicator for distances above 3,200 km is omitted. Blue bars show 95% confidence intervals from robust standard errors. *Sources:* GHTorrent, own calculations.

**Figure A.6:** Border effect



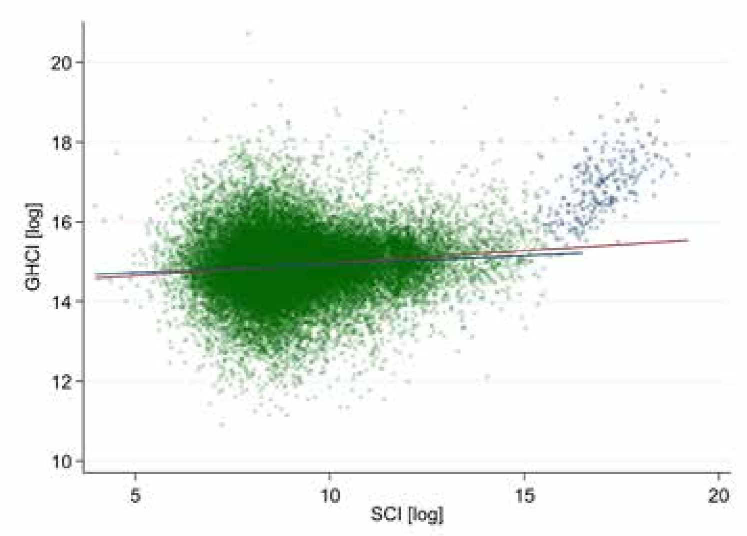

*Notes:* Figure shows scattered values of scaled GHCI (Panel A) and scaled SCI (Panel B) after logarithmic transformation. Both indices are scaled between 1 and 1,000,000,000. Scaled SCI from Bailey et al. (2018) is mean-aggregated from county-county level weighted by multiplied populations of each county-pair and rescaled between 1 and 1,000,000,000. Within-country (cross-border) observations are shown in blue (red). *Sources:* GHTorrent, Bailey et al. (2018), own calculations.

**Figure A.7:** Distribution connectedness indices



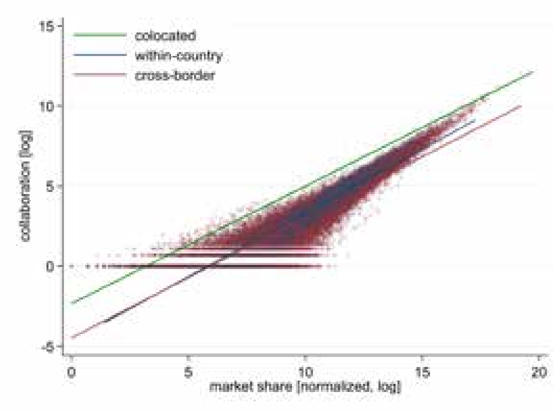

*Notes:* Plots show the distribution of scaled GHCI and SCI regional connectedness indices. The horizontal red lines indicate medians of 6,650 for the SCI and 2,750,304 for the GHCI. The blue curves represent the Epanechnikov kernel density estimates. Both indices are scaled between 1 and 1,000,000,000. Scaled SCI from Bailey et al. (2018) is mean-aggregated from county-county level weighted by multiplied populations of each county-pair and rescaled between 1 and 1,000,000,000. As indices are highly skewed, we restrict the y-axes to maximum values of 20,000,000 for GHCI and 50,000 for SCI to achieve meaningful visualization. Scaled GHCI values of one, representing no links, are excluded from the histogram but not from the median. *Sources:* GHTorrent, Bailey et al. (2018), own calculations.

**Figure A.8:** Relatedness GHCI and SCI

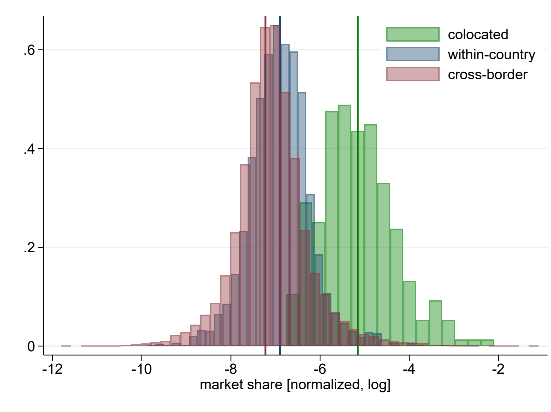



*Notes:* Figure shows the correlation between scaled GHCI and SCI after logarithmic transformation with within-regional collaborations excluded. Colocated collaborations are colored blue. *Sources:* GHTorrent, Bailey et al. (2018), own calculations.

**Figure A.9:** Independence benchmark



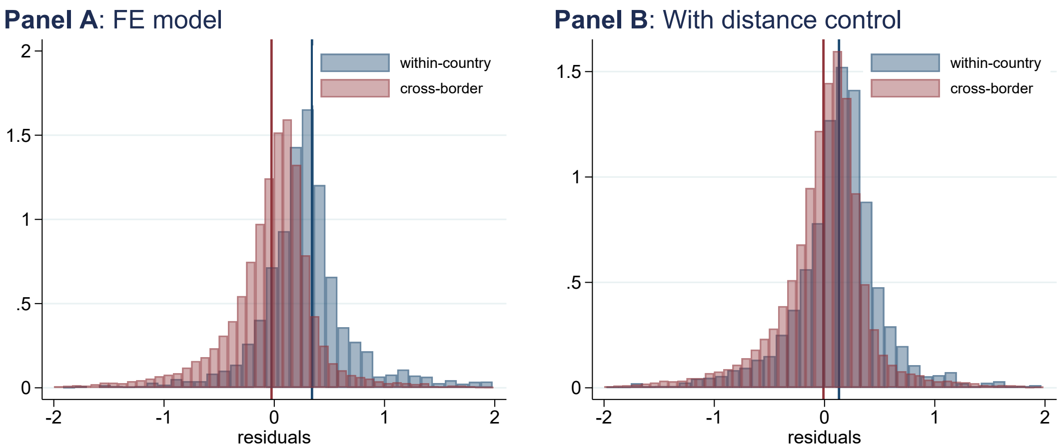

**(a)** market share




**(b)** distribution

*Note:* Figure shows the independence benchmark following Santamaría et al. (2023b) for colocated (green) within-country (blue) and cross-border (red) collaboration, respectively. *Sources:* GHTorrent, own calculations.

**Figure A.10:** Fixed-effect model residuals




*Notes:* Figure shows residual histograms for within-country and cross-border collaboration, respectively. Panel A (Panel B) depicts residuals from the baseline fixed-effects model without (with) controls. *Sources:* GHTorrent, own calculations.