LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Essays in the Economics of Digital Transformation

Moritz Goldbeck

Thesis defense

July 1, 2024

Introduction

- knowledge work drives technological and economic progress
- knowledge worker productivity impacted by increasing digitization
- closely linked to the spatial (re-)organization of knowledge work
  - physical infrastructure provision = basic precondition for participation
  - potential countering force against strong agglomeration effects

▶ **This thesis: interplay between digital transformation, knowledge work, and geography.**

Four papers on the economic geography of digital knowledge work:

1 Effects of physical infrastructure provision on growth and structural change.

2 + 3 New facts on spatial collaboration in digital knowledge work.

4 How digital technology shapes labor markets and public good production.

Digital Infrastructure and Local Economic Growth

Early Internet in Sub-Saharan Africa

(with Valentin Lindlacher)

Motivation

- physical internet infrastructure is the basic precondition to participate in the digital economy
- strong evidence of significant growth effects in developed countries
- investments (public and private) in internet infrastructure in Sub-Saharan Africa (SSA), but:
  - rural areas of developing countries with agricultural economies
  - low-skilled workforce, large informal sector, low adoption

▶ **Does internet access facilitate economic development in lagging regions, as well?**

Source: L'Afrique des Ideés, 2016

Empirical approach

- measure town-level economic growth using night-time light emissions
- exploit quasi-experimental variation arising from particularities of infrastructure roll-out in a difference-in-differences setup
  - arrival of first-generation sub-marine cables for nationwide shock
  - incidentally-connected towns along priority routes of national network expansion to generate treatment and control group

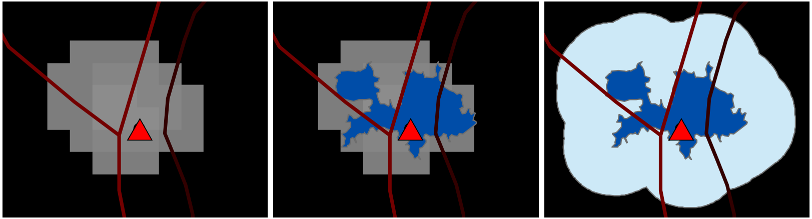


Figure: Data example from Dassa-Zoumè, Benin (2004)

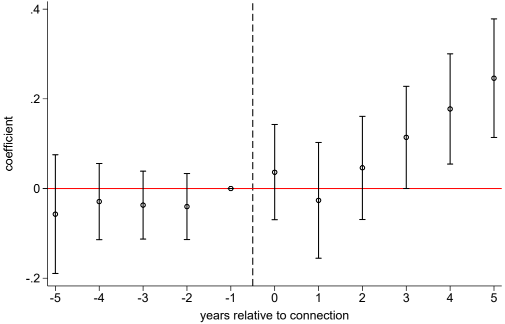
Results

Main effect

- +11% NTL → +3.3 p.p. economic growth

Mechanism

- productivity growth, not migration
- complementarity to market access
- simultaneous shift from agriculture to manufacturing in regional employment



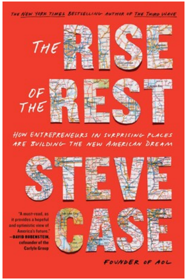
Bit by Bit

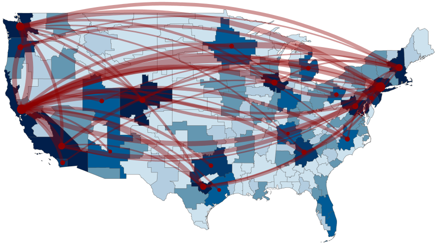
Colocation and the Death of Distance in Software Developer Networks

Motivation

- historically, innovative activities are highly clustered in space
- ICT and digitization potentially relax geographic friction
  - knowledge diffusion
  - team formation
  - communication cost
- little empirical evidence of such effects

▶ **Compare collaboration in digital knowledge work to less digital benchmarks.**





Empirical approach

Non-parametric CEF estimation via residualized binscatter

$$\mathbb{E}[\,\text{links}_{i,j} \mid \mathbf{X}_i,\ \mathbf{X}_j,\ \mathbf{X}_{i,j}\,]$$

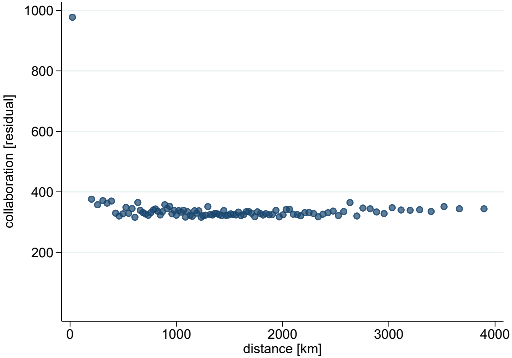
- disentangle effect of distance from agglomeration effects
- plot CEF after partialling out controls for collaboration potential and cluster size

Gravity-type modelling via OLS

$$\text{links}_{i,j} = \beta_0 + \beta_1\,\text{coloc}_{i,j} + \beta_2\,\text{dist}_{i,j} + \mathbf{X_i}\boldsymbol{\beta_3} + \mathbf{X_j}\boldsymbol{\beta_4} + \mathbf{X_{i,j}}\boldsymbol{\beta_5} + \epsilon_{i,j}$$

$\text{links}_{i,j}$	number of links between regions
$\text{coloc}_{i,j}$	colocation indicator
$\text{dist}_{i,j}$	geographic distance (centroid-based)
$\mathbf{X_i}, \mathbf{X_j}$	origin/destination controls or fixed effects
$\mathbf{X_{i,j}}$	origin-destination pair controls
$\epsilon_{i,j}$	error term

Results: colocation effect

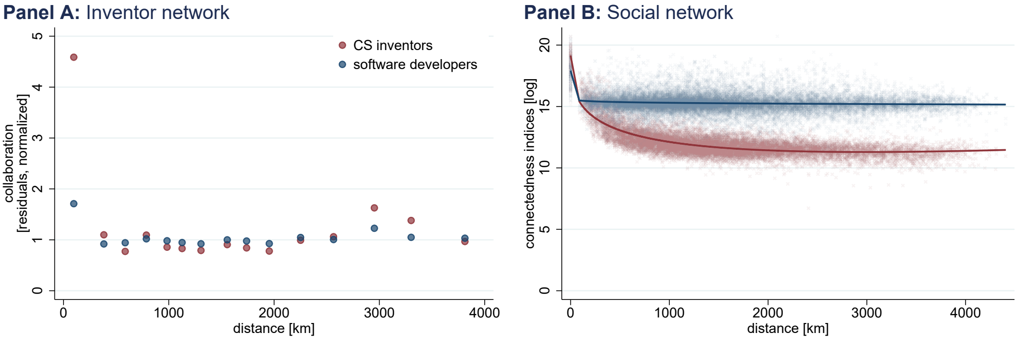


- median residualized collaboration per percentile bin
- discontinuity in CEF for colocated collaboration: colocation effect
- otherwise neglectable relation to increased distance

Results: gravity estimation

Collaboration [log]	(1)	(2)	(3)	(4)	(5)	(6)
Colocation	2.825***	2.354***	2.298***	2.371***	2.286***	2.329***
	(0.223)	(0.176)	(0.177)	(0.171)	(0.153)	(0.071)
Distance	0.024***	-0.006***	-0.006***	-0.001	-0.006***	-0.004***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Users		×	×	×	×	
Users, multiplied			×	×	×	×
GDPs				×	×	
Populations					×	
Origin FE						×
Destination FE						×
Observations	31,329	31,329	31,329	31,329	31,329	31,329
Adj. R^2	0.016	0.409	0.409	0.469	0.595	0.922
$\exp(\hat{\beta}_{\text{colocation}}) - 1$	15.87	9.53	8.96	9.71	8.83	9.26

Results: less digital benchmarks



- **inventor network:** colocation effect **2-3** times higher
- **social network:** colocation effect ≈**4** times higher + continued spatial decay

Results: heterogeneity in digital work

Smaller **colocation effect**

- within large organizations
  200+ users: –15%; big tech –35%
- higher-quality projects
  forks: -19%; followers –28%; stars: –59%
- large projects
  team: -77%; commits: –31%; age: –72%

Higher **colocation effect**

- for intense collaboration
  projects: +230%; commits: +970%
- inexperienced users
  platform tenure: +62%

Virtually Borderless?

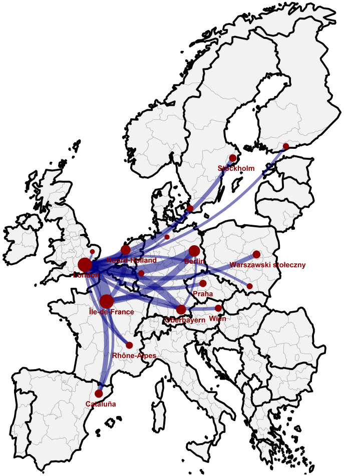
Cultural Proximity and International Collaboration of Developers

(with Lena Abou El-Komboz)

Motivation

- fully virtual collaboration is possible
- little impact of geographic distance
- European software industry is lagging, despite large market size
- border effects are a potential barrier to international collaboration

▶ **Are national borders an impediment to collaboration in the digital knowledge economy?**



Empirical approach

- explore association of border effects with cross-country differences
- to what extent is a potential border effect associated with cultural proximity

Estimation of gravity-type model via OLS

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

$y_{i,j}$	number of links between regions
$\text{crossborder}_{i,j}$	cross-border indicator
$\text{coloc}_{i,j}$	colocation indicator
$\text{dist}_{i,j}$	geographic distance (centroid-based)
$\mathbf{X_{c(i),c(j)}}$	country-level differences
$\mathbf{X_i}, \mathbf{X_j}$	origin/destination controls or fixed effects
$\epsilon_{i,j}$	error term

Results: digital border effect

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%

- border effect ≈**5-6x smaller** compared to goods trade

Results: 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

Role of organizations

- decomposition of interest overlap
- strongest association with professional interests

Additional controls

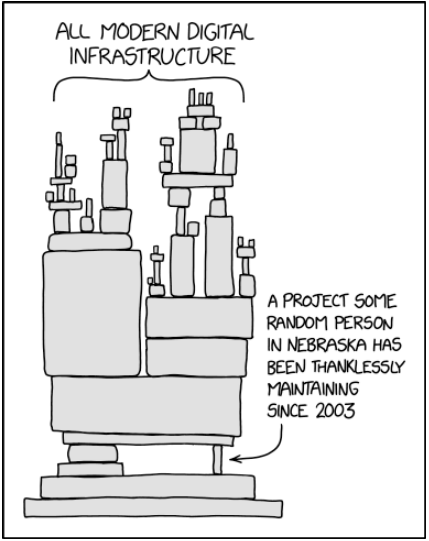
- political factors
- history controls
- diplomatic relations
- legal systems

Career Concerns as Public Good

The Role of Signaling for Open Source Software Development

(with Lena Abou El-Komboz)

Motivation



- open source software (OSS) is a valuable public good
  - 96% of software codebases contain OSS
  - equiv. 7.2% of software investment (USD37bn/yr)
- decentralized community of volunteer developers
- motivation to contribute hard to rationalize

► **Are OSS developers motivated by labor market signaling incentives?**

Empirical approach

Difference-in-differences

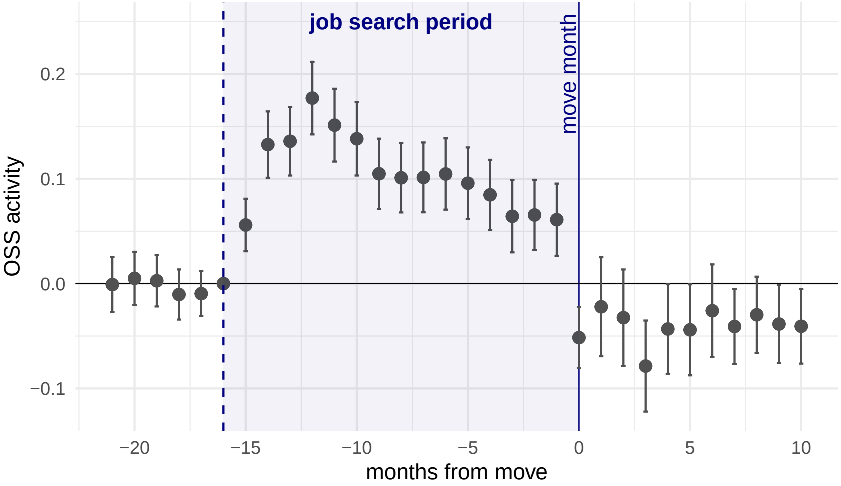
- look at job changers and their activity in the job search period
- compare job movers versus other movers

Event study specification

$$y_{it} = \beta_1 + \sum_{j=\underline{T}}^{\overline{T}} \left[\beta_j \left(t_j \times \text{JobChanger}_i \right) \right] + \delta_i + \delta_{s(t)} + \delta_{a(i)t} + e_{it},$$

$y_{i,j}$	number of user *i*'s commits in month ***t*** (IHS)
δ_i	user fixed effects
$\delta_{s(t)}$	month fixed effects
$\delta_{a(i)t}$	user experience fixed effects
e_{it}	error term

Results: signaling activity



Results: difference-in-differences

IHS(single commits)	(1)	(2)	(3)
Job mover × job search	0.2595***	0.2230***	0.1177***
	(0.0088)	(0.0093)	(0.0091)
Job mover × post move	-0.2154***	-0.1738***	-0.0813***
	(0.0120)	(0.0131)	(0.0123)
User FE	×	×	×
Month FE		×	×
Experience FE			×
Adjusted R^2	0.139	0.154	0.217
Observations	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896

- back-of-the-envelope calculation → **≈4.9% of overall OSS production**

Results: heterogeneity

- signaling projects focus less on (direct) community use-value (stars, forks)
- signaling activity concentrates on labor market value and external visibility
  - higher-valued programming languages
  - in web development and data engineering
  - keywords for coding and (personal) website
- users' signaling activity
  - higher for international/-continental movers
  - higher when moving to academia
  - lower when moving to big tech

Overall, results highlight the inclusionary effects of the digital transformation, while pointing out remaining geography-related barriers.

- digital technology makes it easier to *participate from anywhere*
- *intense interaction* remains *local*; *large organizations* seem to facilitate *remote* collaboration
- *cultural barriers* are relatively more important in the digital economy
- technology allows to *signal skill from anywhere* while generating a *public good*

Thanks,

what are your questions?

moritz@goldbeck.net

moritz.goldbeck.net

moritzgoldbeck

Appendix

Appendix

Chapter 1 Internet Infrastructure and Economic Growth

▸ Reg. eq. ▸ Internet adoption ▸ On-route towns ▸ Balance POIs ▸ Balance geography ▸ Population density ▸ Access points ▸ SMC arrivals

▸ NTL results ▸ Market access ▸ Employment ▸ Migration ▸ Spillovers ▸ Mobile internet ▸ Placebos ▸ Infrastructure

Chapter 2 Colocation and Distance in Digital Knowledge Work

▸ Sample ▸ Measurement ▸ Sum stats ▸ Representative ▸ Benchmark data ▸ GHCI/SCI ▸ Het. method ▸ All inventors

▸ Local ▸ Binscatter ▸ GHCI-SCI corr. ▸ GHCI/SCI hist. ▸ Simil. inventors ▸ Network overlap ▸ CI and dist.

▸ Inventors ▸ Regional ▸ Link het. ▸ Dynamics ▸ Model flex. ▸ Individual-level ▸ Non-para. ▸ Indiv. non-para. ▸ Relatedness

Chapter 3 Digital Border Effects and Cultural Proximity

▸ Sample ▸ Size heterogeneity ▸ Decomposition interest overlap ▸ States USA ▸ Santamaria viz ▸ Residual plots

Chapter 4 Signaling of Open Source Software Developers

▸ Sample ▸ Summary stats ▸ Move map ▸ Domestic moves ▸ Move timing ▸ Organizations ▸ Cities ▸ Job transitions ▸ Keywords

▸ Stars/forks ▸ Category ▸ Keyword-based ▸ Languages ▸ International ▸ Affiliation ▸ Project age ▸ Followers ▸ Model ▸ Own/forks

Appendix 1

Internet Infrastructure and Economic Growth

Regression equation

Difference-in-differences

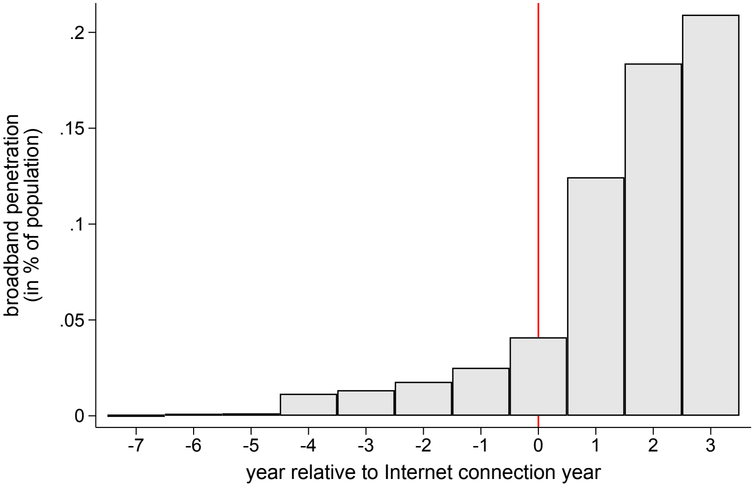
$$y_{it} = \beta_0 + \beta_1 \left(\text{connect}_{c(i)t} \times \text{access}_i\right) + \beta_2 \text{GSM}_{it} + \beta_3 (\mathbf{X}_i' \times \text{connect}_{c(i)t}) + \alpha_i + \alpha_{c(i)t} + \varepsilon_{it}$$

Dynamic specification

$$y_{it} = \mu_0 + \sum_{j(c(i))=\underline{T}}^{\overline{T}} \left[\mu_{1j} \left(t_{j(c(i))} \times \text{access}_i\right)\right] + \mu_2 \text{GSM}_{it} + \mu_3 (\mathbf{X}_i' \times \text{connect}_{c(i)t}) + \delta_i + \delta_{c(i)t} + e_{it}$$

y_{it}	town-level NTL emissions
$\text{connect}_{c(i)t}$	post SMC arrival indicator
access_i	access point at SMC arrival time
GSM_{it}	GSM signal coverage
$\mathbf{X}_i$	geography controls
$\alpha_i,\ \delta_i$	town FE
$\alpha_{c(i)t},\ \delta_{c(i)t}$	country-year FE

Internet adoption



Difference-in-differences

	NTL growth			NTL growth margin	
	(1) composite	(2) composite	(3) composite	(4) intensive	(5) extensive
Connection × access	0.129***	0.134***	0.109***	0.0769***	0.0817**
	(0.0427)	(0.0433)	(0.0383)	(0.0237)	(0.0330)
Town FE	×	×	×	×	×
Country × year FE	×	×	×	×	×
GSM coverage		×	×	×	×
Geography controls × connection			×	×	×
Observations	2,310	2,310	2,310	2,310	2,310
Countries	10	10	10	10	10
Towns	210	210	210	210	210
Share treated	*.462*	*.462*	*.462*	*.462*	*.462*
Adjusted R^2	0.936	0.936	0.948	0.923	0.919
Economic growth effect	3.90	4.06	3.26	–	–

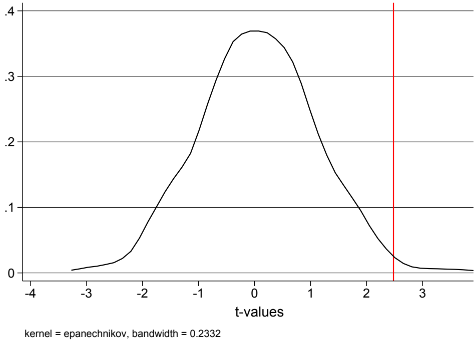
Market access

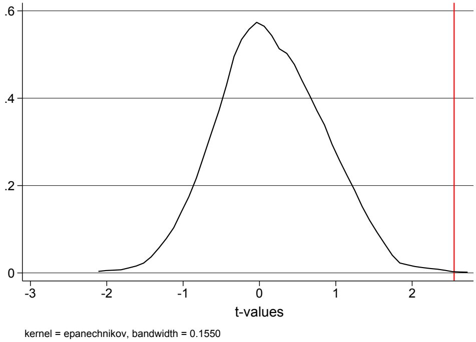
	(1)	(2)	(3)
Connection × access	0.110***	0.101***	0.205***
	(0.0378)	(0.0367)	(0.0721)
Connection × access × distance port	-0.0667*		
	(0.0400)		
Connection × access × market access		0.0369**	
		(0.0175)	
Connection × access × landlocked			-0.145*
			(0.0807)
Town FE	×	×	×
Country × year FE	×	×	×
GSM coverage	×	×	×
Geography controls × connection	×	×	×
Observations	2,310	2,310	2,310
Countries	10	10	10
Towns	210	210	210
Share treated	*.462*	*.462*	*.462*
Adjusted R^2	0.943	0.942	0.942

Regional employment shares

	(1)	(2)	(3)
Sector:	agriculture	manufacturing	services
Connection × access	-0.0194	0.0129*	0.00642
	(0.0163)	(0.0074)	(0.0107)
Region FE	×	×	×
Country × year FE	×	×	×
GSM coverage	×	×	×
Observations	956,454	956,454	956,454
Countries	5	5	5
Regions	99	99	99
Share treated	*.208*	*.208*	*.208*
Adjusted R^2	0.128	0.039	0.100

Access and connection placebos





Infrastructure heterogeneity

	(1)	(2)	(3)	(4)	(5)
Connection × access	0.115**	0.107**	0.115***	0.119***	0.0949**
	(0.0490)	(0.0440)	(0.0437)	(0.0451)	(0.0466)
Connection × access ×					
distance roads	0.0306				
	(0.120)				
distance railroads		-0.0224			
		(0.0302)			
distance electricity grid			0.0765		
			(0.0492)		
distance border				-0.0421	
				(0.0508)	
distance capital					-0.0246
					(0.0541)
Town FE	×	×	×	×	×
Country × year FE	×	×	×	×	×
GSM coverage	×	×	×	×	×
Geography controls × connection	×	×	×	×	×
Observations	2,310	2,310	2,310	2,310	2,310
Countries	10	10	10	10	10
Towns	210	210	210	210	210
Share treated	*.462*	*.462*	*.462*	*.462*	*.462*
Adjusted R^2	0.943	0.942	0.942	0.942	0.942

Electricity access heterogeneity

Sample:	extended		capital and landing		all nodal	
Dep. var.: electricity access	(1)	(2)	(3)	(4)	(5)	(6)
Connection × access	0.000387	-0.0359	0.0411	0.0579	-0.0731	-0.0914
	(0.103)	(0.0688)	(0.114)	(0.0766)	(0.211)	(0.173)
Town FE	×	×	×	×	×	×
Country × year FE	×	×	×	×	×	×
GSM coverage	×	×	×	×	×	×
Geography controls × connection	×	×	×	×	×	×
Weights		×		×		×
Observations	270	270	250	250	102	102
Countries	6	6	6	6	4	4
Towns	94	94	88	88	37	37
Share treated	.351	.351	.307	.307	.351	.351
Adjusted R^2	0.680	0.806	0.675	0.784	0.720	0.814

Population growth

Dep. var.: population Time window:	(1) baseline	(2) 2000 - (SMC + 3)	(3) incl. 1995	(4) excl. 1995	(5) pre/post
Connection × access	0.0116 (0.0183)	-0.00283 (0.00805)	0.0218 (0.0374)	0.0124 (0.0277)	0.0102 (0.0191)
Town FE	×	×	×	×	×
Country × year FE	×	×	×	×	×
GSM coverage	×	×	×	×	×
Geography controls × connection	×	×	×	×	×
Observations	2,310	1,765	830	610	440
Countries	10	10	10	10	10
Towns	210	210	210	210	210
Share treated	*.462*	*.462*	*.462*	*.462*	*.462*
Adjusted R^2	0.999	1.000	0.997	0.999	1.000

X

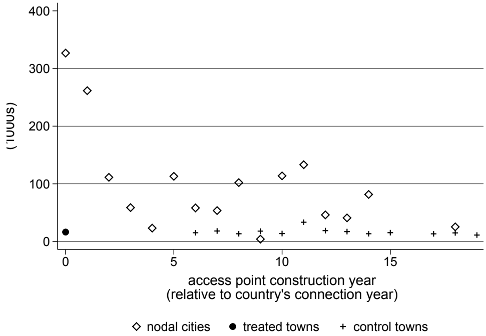
Spillovers

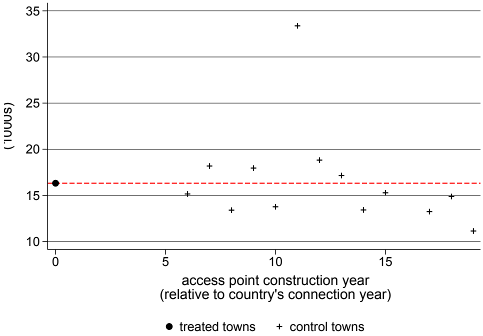
	(1)	(2)
Connection × access point ∈ (0km, 10km]	0.147***	0.119***
	(0.0511)	(0.0385)
Connection × access point ∈ (10km, 30km]	0.0925	0.0863**
	(0.0606)	(0.0367)
Connection × access point ∈ (30km, 50km]	0.0489	0.0280
	(0.0545)	(0.0369)
Town FE	×	×
Country × year FE	×	×
GSM coverage	×	×
Geography controls × connection	×	×
Untreated controls		×
Observations	2,310	4,114
Countries	10	12
Towns	210	374
Share treated	.462	.27
Adjusted R^2	0.942	0.927

Mobile internet connectivity

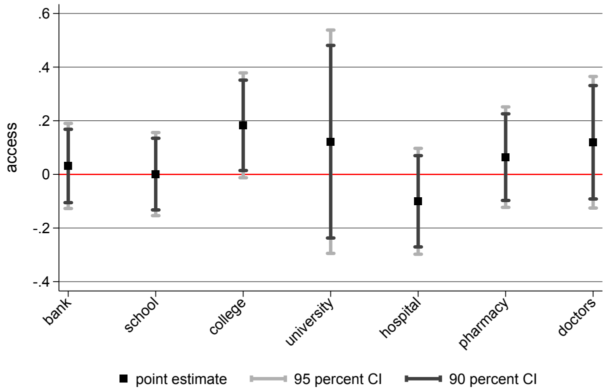
	(1)	(2)	(3)	(4)	(5)	(6)
Connection × access	0.109***	0.110***	0.105***	0.106***	0.105***	0.102***
	(0.0383)	(0.0378)	(0.0384)	(0.0373)	(0.0373)	(0.0381)
GSM coverage	0.0539					
	(0.0380)					
GSM coverage (lag 1)		0.0758*				
		(0.0402)				
GSM coverage (lag 2)			-0.0161			
			(0.0399)			
GSM coverage (lag 3)				0.0510		
				(0.0327)		
GSM coverage (lag 4)					0.0518*	
					(0.0311)	
GSM coverage (lag 5)						0.0434
						(0.0335)
Town FE	×	×	×	×	×	×
Country × year FE	×	×	×	×	×	×
Geography controls × connection	×	×	×	×	×	×
Observations	2,310	2,310	2,310	2,310	2,310	2,310
Countries	10	10	10	10	10	10
Towns	210	210	210	210	210	210
Share treated	.462	.462	.462	.462	.462	.462
Adjusted R^2	0.942	0.942	0.942	0.942	0.942	0.942

Incidental connection

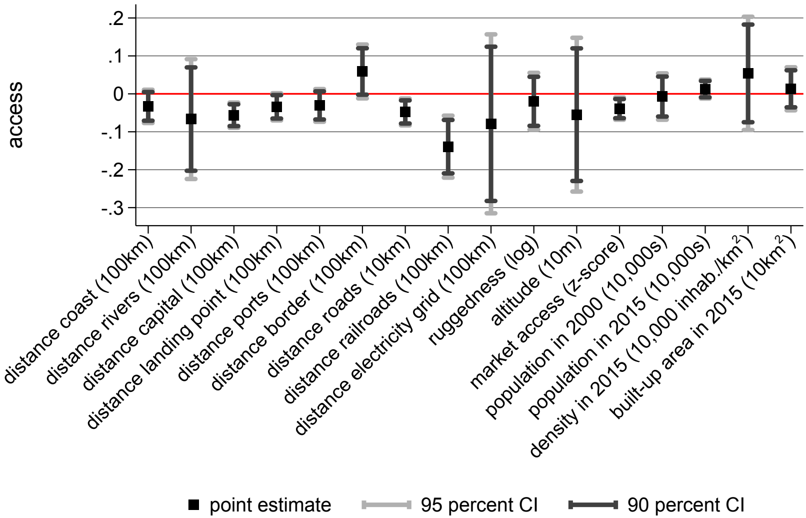




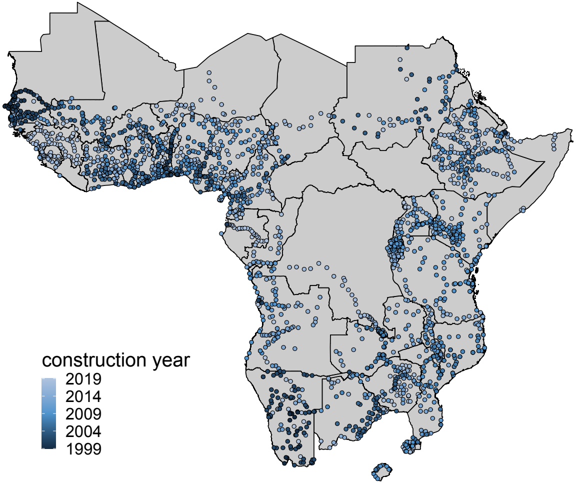
Sample balance: OSM POIs



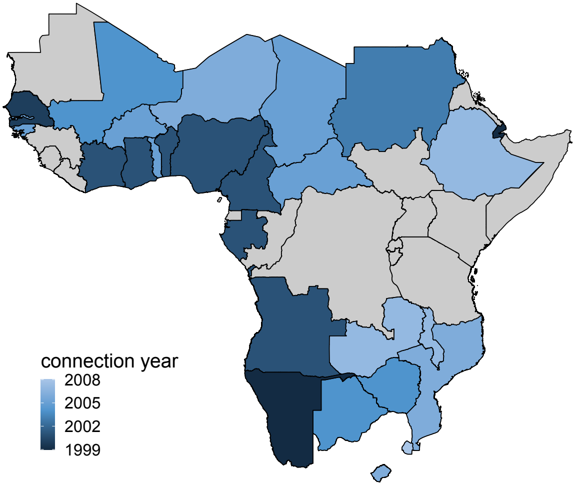
Sample balance: geography



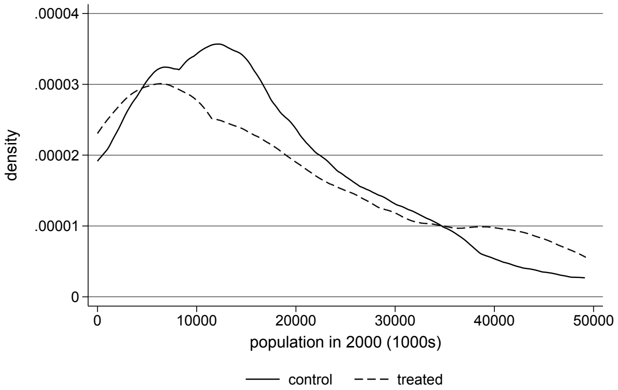
Data: access points



Data: SMC years



Population density



Appendix 2

Colocation and Distance in Digital Knowledge Work

Sample

User base: active, geolocated in U.S., and collaborating

- 10 snapshots 2015-2021 (≈ every 7 months) ▸ Snapshots
- overall ≈ 44,131k users worldwide
  - thereof ≈ 2,299k (5.2%) with self-reported geolocation
  - thereof ≈ 778k (34%) users have a geolocation in United States
  - thereof ≈ 354k (46%) users are active users
  - thereof ≈ 191k (54%) users have in-sample collaborations (= sample of analysis) ▸ Descriptives

Users' activity stream

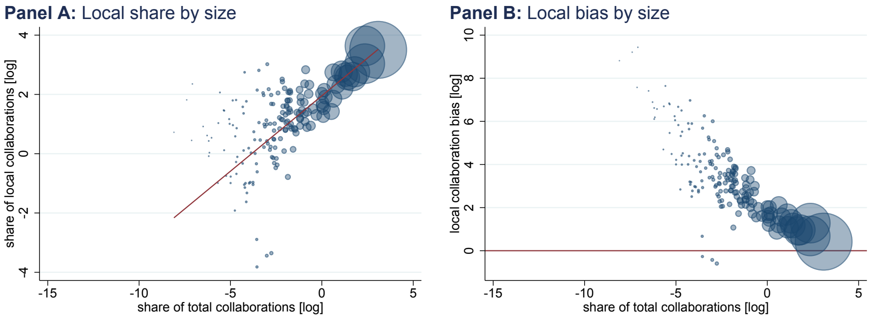
- 4,289k repositories (= code/software projects)
- 97,304k commits (= code contributions)
- mostly in Javascript, Python, and versions of C ▸ Languages

Measurement

- collaboration/links defined via joint projects
  - contribution to project during observation period
  - links to other (active) project members
- geographic unit are *Bureau of Economic Analysis*' economic areas
  - "relevant regional markets around metropolitan statistical areas"
  - essentially MSAs, but larger for big cities
  - sufficient users in each area

▶ **unit of analysis are economic-area pairs**
  - number of links between economic areas
  - centroid-based (geodesic) distance

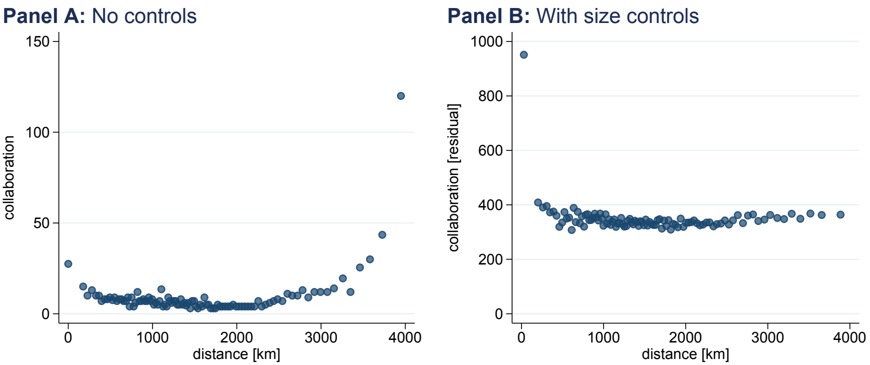
Collaboration tends to be local



- the bigger the economic area, the higher the share of local links
- but: given local cluster size, collaboration in small economic areas are disproportionately local

▸ Links to hubs ▸ Local share ▸ Local bias

Colocation and size matter, but not distance



- most links are between large clusters (opposite coasts)
- after controlling for size, colocation effect clearly visible

Distance histogram

Benchmarks

1. **Inventor network**
   - U.S. computer science inventors, 2015–2021
   - data on collaborative patents from *PatStat*
   - geolocation from Seliger et al. (2019)
   - N $\approx$ 17,000

   - high-skilled professionals
   - more innovative, creative, and novel work

2. **Social network**
   - U.S. friendship/aquantanceship networks, 2016
   - geolocated *Facebook* data from Bailey et al. (2018)
   - N(guessed) $\approx$ 180mn; 87% (18-26y) $\rightarrow$ 56% (>65y)

   - general population
   - more face-to-face, less digital

Inventor gravity

Collaboration	all		connected	
	(1) inventors	(2) developers	(3) inventors	(4) developers
Colocation	3.373***	2.329***	3.292***	2.478***
	(0.138)	(0.071)	(0.102)	(0.081)
Distance	-0.009***	-0.004***	-0.018***	-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)
Users, multiplied	×	×	×	×
Origin FE	×	×	×	×
Destination FE	×	×	×	×
Observations	31,329	31,329	6,662	6,662
Adj. R^2	0.566	0.922	0.593	0.975
$\exp(\hat{\beta}_{\text{colocation}}) - 1$	28.18	9.26	25.90	10.91
Relative effect size	3.04		2.37	

Social network benchmark

- social network only available as connectedness index

GitHub/Social Connectedness Index

- SCI introduced by Bailey et al. (2018) using Facebook data
- relative probability of a link between users in two regions

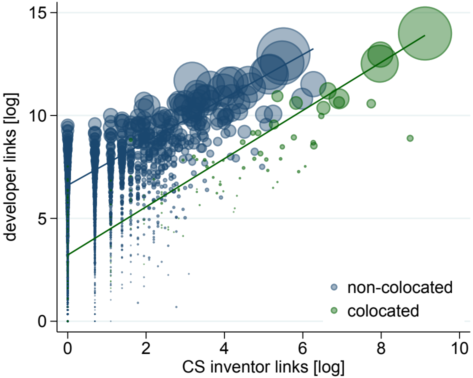
$$GHCI_{i,j} = \frac{links_{i,j}}{users_i * users_j}$$

- independent of regions' size
- scaled between 1 and 1,000,000,000

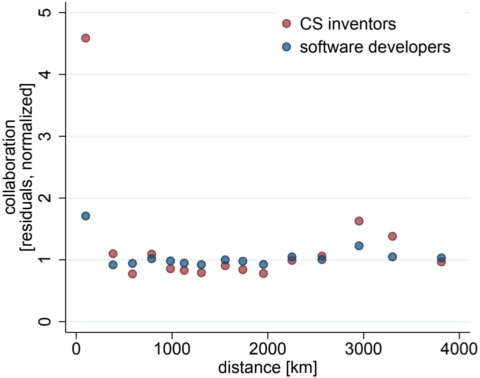
▸ Histograms

Results: inventor network benchmark

Panel A: Network overlap

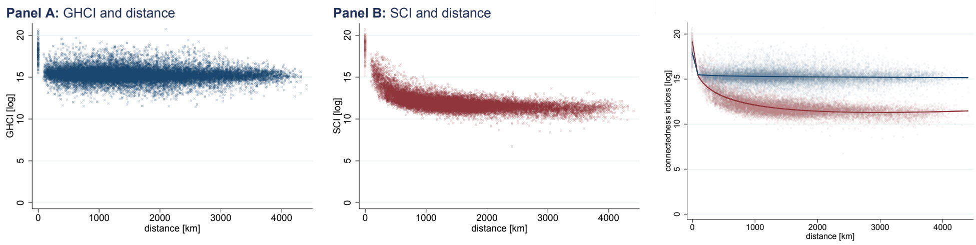


Panel B: Binscatter



- quantification via gravity: **2-3 times higher** colocation effect for inventors

Results: social network benchmark



Colocation effect $= \frac{\hat{CI}_{dist=0}}{\hat{CI}_{\mathbf{min}\ (dist \neq 0)}}$

- GHCI: **11.2**x; SCI: **41.4**x
- **3.7x larger effect** in social network

Heterogeneity

- by **regional characteristics**
  - data on firm presence from *County Business Patterns*
  - number of local users (i.e., cluster size)

  ▶ In which regions is the colocation effect strong/weak?

- by **link characteristics**
  - compute above- and below-threshold networks for link-level metrics
  - organizations, quality, users, projects, intensity

  ▶ Which types of collaborations colocate more/less?

Heterogeneity: regional characteristics

Collaboration	(1) baseline	# local users (2) $\geq$ median	# local users (3) Top 10	large firm presence (4) tech $\geq$ median	large firm presence (5) software $\geq$ median
Colocation	2.329***	2.478***	2.430***	2.498***	2.430***
	(0.071)	(0.113)	(0.068)	(0.074)	(0.069)
Distance	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Colocation interaction		-0.295**	-1.978***	-1.026***	-1.595***
		(0.142)	(0.446)	(0.183)	(0.386)
Observations	31,329	31,329	31,329	31,329	31,329
Adj. R^2	0.922	0.923	0.923	0.923	0.923
$\exp(\hat{\beta}_{\text{colocation}}) - 1$	9.26	10.91	10.36	11.16	10.36
$\exp(\hat{\beta}_{\text{colocation}} + \hat{\beta}_{\text{interaction}}) - 1$	–	7.87	0.57	3.36	1.31

Heterogeneity: link characteristics

Dimension	colocation effect	relative effect	relative to baseline
Panel A: Organizations			
intra-organization	5.26	1.41	0.57
inter-organization	3.73		0.40
within big-tech firm	0.13	0.65	0.01
big-tech firm involved	0.20		0.02
within multi-establishment firm	3.48	0.99	0.38
multi-establishment firm involved	3.51		0.38
within large firm	0.59	0.76	0.06
large firm involved	0.78		0.08
Panel B: Quality			
above-median followers	6.64	0.72	0.72
below-median followers	9.16		0.99
above-median forks	8.97	0.81	0.97
below-median forks	11.07		1.20
with stars	6.49	0.41	0.70
no stars	15.80		1.71

Heterogeneity: link characteristics

Dimension	colocation effect	relative effect	relative to baseline
Panel C: User type			
above-median user experience	6.00	0.62	0.65
below-median user experience	9.75		1.05
above-median experience differential	4.36	0.39	0.47
below-median experience differential	11.08		1.20
common programming language	8.02	0.99	0.87
no common programming language	8.13		0.88
Panel D: Collaboration intensity			
strong tie, via project	11.23	1.57	1.21
weak tie, via project	7.16		0.77
above-median project commits	13.00	4.36	1.40
below-median project commits	2.98		0.32
strong tie, via commits	13.05	2.54	1.41
weak tie, via commits	5.12		0.55

▶ Relatedness

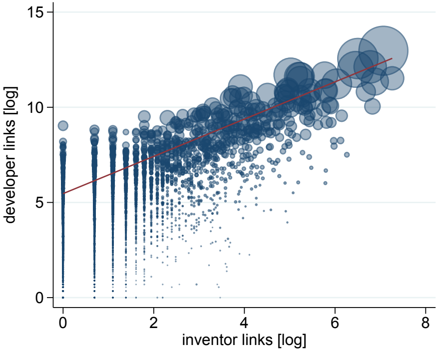
Heterogeneity: link characteristics

Dimension	colocation effect	relative effect	relative to baseline
Panel E: Project type			
above-median users	6.13	0.33	0.66
below-median users	18.47		1.99
above-median commits	8.64	0.69	0.93
below-median commits	12.47		1.35
above-median project age	6.38	0.38	0.69
below-median project age	16.99		1.83

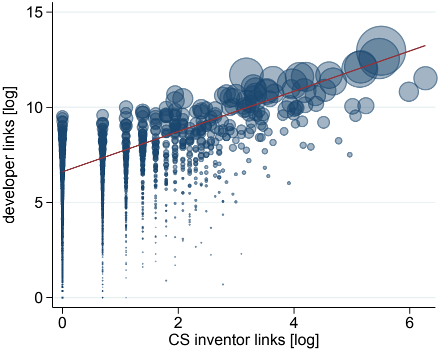
▸ Relatedness

Inventor network overlap

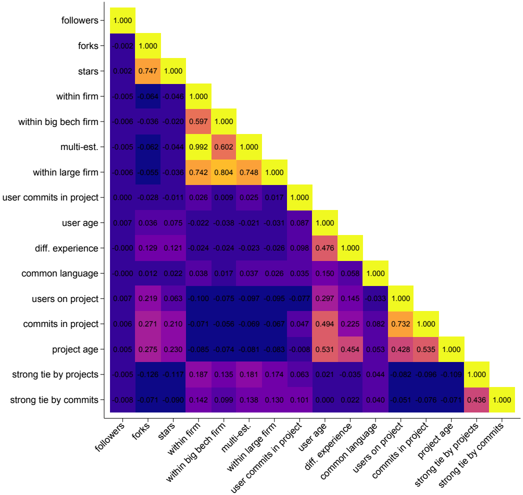
Panel A: All inventors



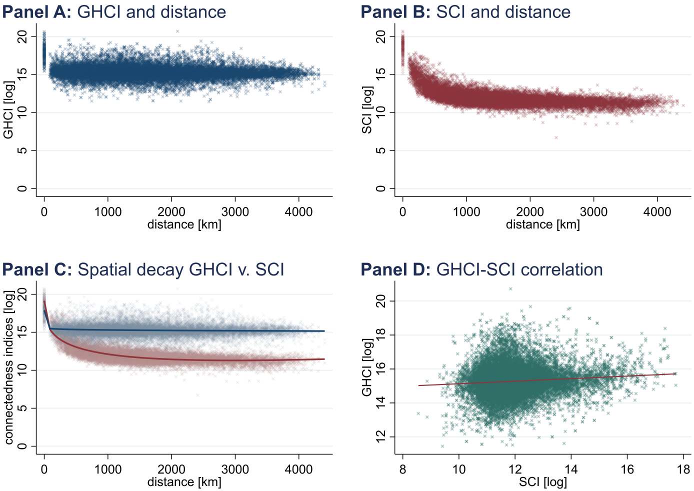
Panel B: CS inventors



Relatedness



Correlation SCI/GHCI



Summary statistics

Statistic	Mean	Median	Min	Max	N
Users					
Projects per user	28.51	14	1	46,508	190,637
Links per user	123.65	7	1	14,739	190,637
Commits per user	510.42	156	1	388,287	190,637
Commits per user-project	18.40	3	1	364,397	5,286,886
Projects					
Commits per project	22.64	3	1	364,397	4,298,045
per personal project	13.97	3	1	364,397	3,867,611
per team project	100.52	18	2	209,214	430,435
Users per team project	3.64	2	2	147,236	430,435
Economic areas					
Users per economic area	1,895	302	2	53,818	179
Projects per economic area	26,924	3,328	4	831,728	179
Links per economic area	130,562	15,329	1	5,175,727	179
Links per economic-area pair	930	23	1	1,550,463	25,135
Commits per economic area	543,600	69,185	19	19,165,952	179

Model specification

Collaboration [log]	(1)	(2)	(3)	(4)
Colocation	2.295***	2.353***	2.433***	2.277***
	(0.075)	(0.082)	(0.074)	(0.079)
Distance	-0.022***	-0.004***	-0.004***	-0.020***
	(0.002)	(0.001)	(0.001)	(0.002)
Distance, squared	0.001***			0.000***
	(0.000)			(0.000)
Users, multiplied	×	×	×	×
Users, multiplied (squared)			×	×
GDPs, multiplied		×		×
GDPs, multiplied (squared)				×
Populations, multiplied		×		×
Populations, multiplied (squared)				×
Origin FE	×	×	×	×
Destination FE	×	×	×	×
Observations	31,329	31,329	31,329	31,329
Adj. R^2	0.913	0.915	0.913	0.917
$\exp(\hat{\beta}_{colocation}) - 1$	8.92	9.52	10.39	8.74

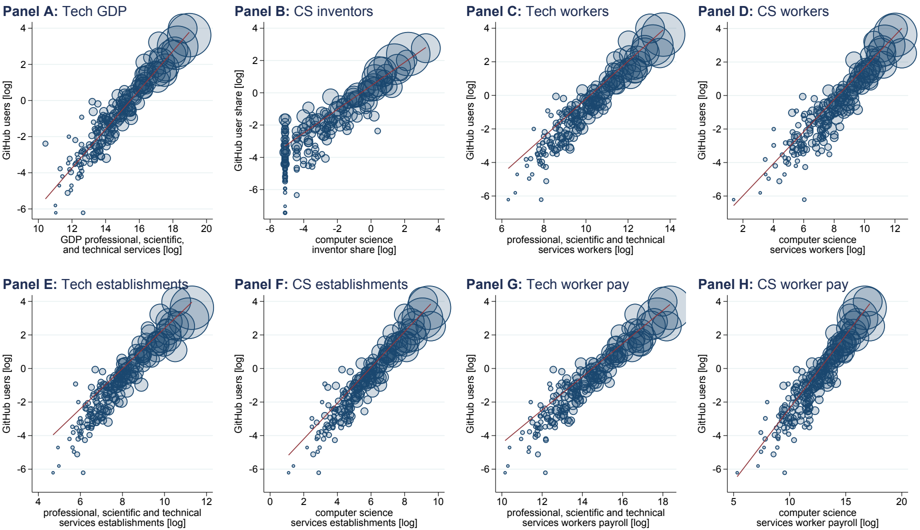
Individual-level probability models

Collaboration	(1) LPM	(2) PPML	(3) Probit
$<$ 100 km	0.00139***	0.226***	0.080***
	(0.00006)	(0.010)	(0.003)
100 – 400 km	0.00019***	0.036***	0.013***
	(0.00007)	(0.012)	(0.004)
400 – 1200 km	-0.00005	-0.008	-0.003
	(0.00004)	(0.007)	(0.003)
1200 – 2400 km	-0.00009*	-0.019**	-0.006**
	(0.00005)	(0.009)	(0.003)
2400 – 3200 km	-0.00011**	-0.020**	-0.007**
	(0.00005)	(0.009)	(0.003)
Origin FE	×	×	×
Destination FE	×	×	×
Observations	33,183,717	33,179,297	33,179,297
Users (random sample)	10,726	10,726	10,726
Sample share	0.056	0.056	0.056
(Pseudo) Adj. R^2	0.0003	0.0046	0.0046

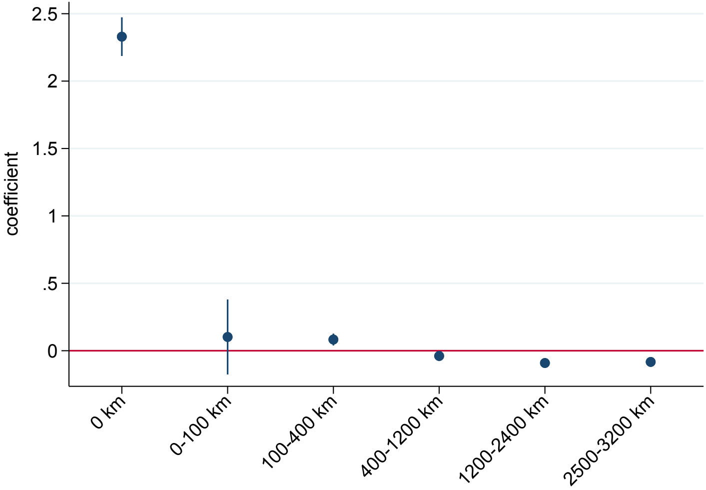
Regional heterogeneity and organizations

Collaboration		# local users		large firm presence	
	(1)	(2)	(3)	(4)	(5)
	baseline	$\geq$ median	Top 10	tech $\geq$ median	software $\geq$ median
Colocation	2.329***	2.478***	2.430***	2.498***	2.430***
	(0.071)	(0.113)	(0.068)	(0.074)	(0.069)
Distance	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Colocation interaction		-0.295**	-1.978***	-1.026***	-1.595***
		(0.142)	(0.446)	(0.183)	(0.386)
Observations	31,329	31,329	31,329	31,329	31,329
Adj. R^2	0.922	0.923	0.923	0.923	0.923
$\exp(\hat{\beta}_{\text{colocation}}) - 1$	9.26	10.91	10.36	11.16	10.36
$\exp(\hat{\beta}_{\text{colocation}} + \hat{\beta}_{\text{interaction}}) - 1$	–	7.87	0.57	3.36	1.31

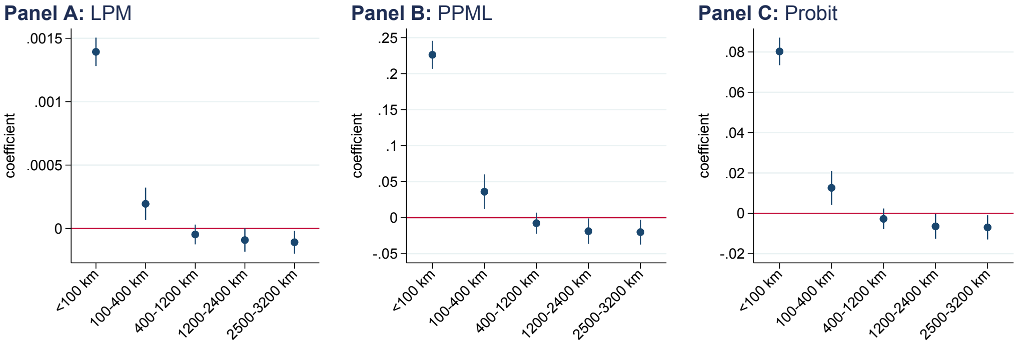
Representativeness



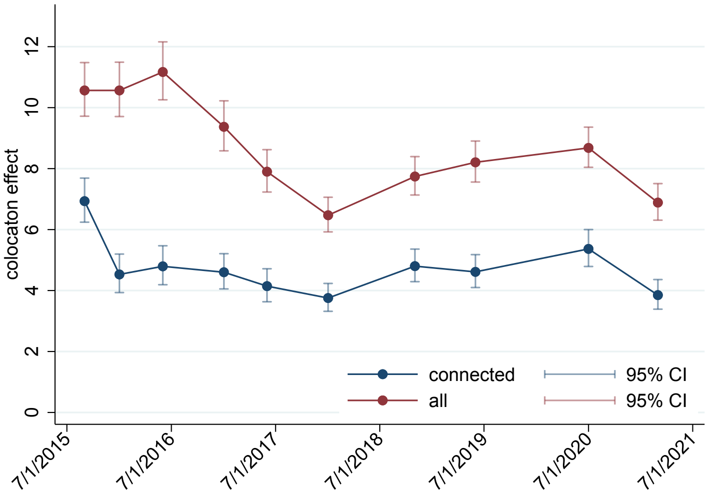
Non-parametric distance



Distance in individual-level models

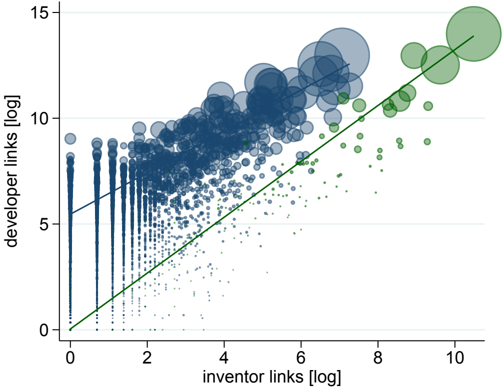


Colocation dynamics

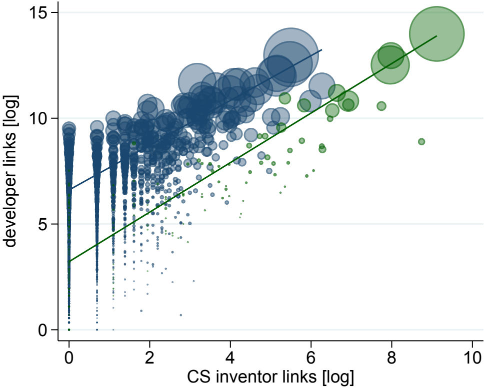


(Computer science) inventors

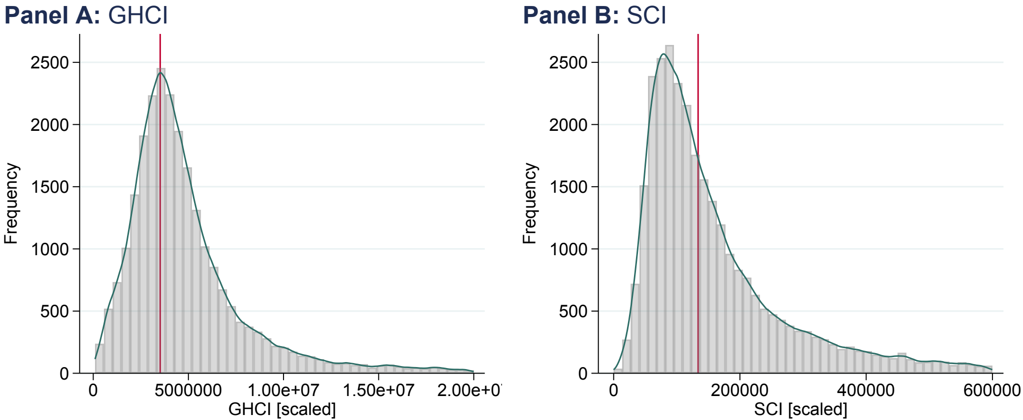
Panel A: All inventors



Panel B: CS inventors



Histogram GHCI/SCI

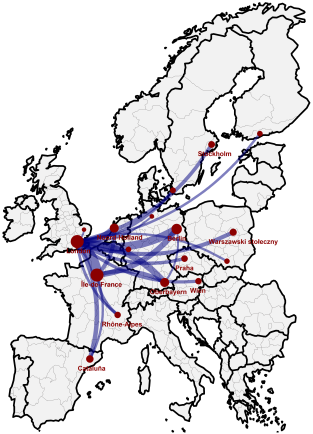


Appendix 3

Digital Border Effects and Cultural Proximity

Sample

- 10 snapshots (2015-2021)
- active, geolocated, collaborating
- 144k European users



Country size heterogeneity

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

Cultural proximity decomposition

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)

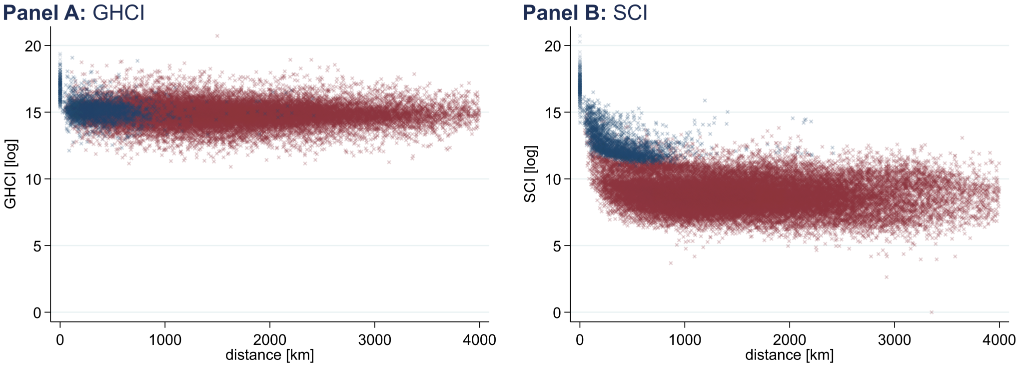
Cultural proximity decomposition

Collaboration	(1)	(2)	(3)
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

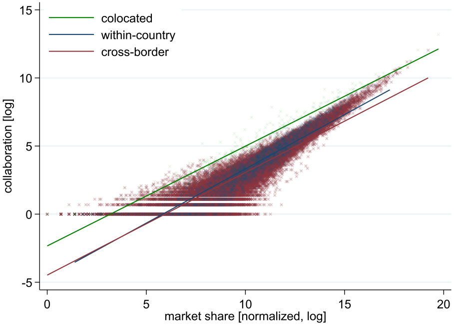
Border effect USA

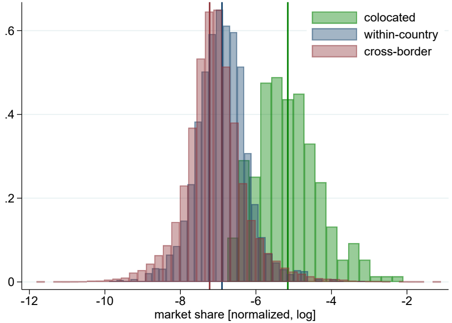
Collaboration	(1)	(2)	(3)	(4)
Cross-border	-0.527***	-0.429***	-0.502***	-0.100***
	(0.098)	(0.041)	(0.037)	(0.033)
Users, multiplied [log]		0.750***		
		(0.004)		
Colocation				2.191***
				(0.073)
Distance [log]				-0.060***
				(0.011)
Origin FE			×	×
Destination FE			×	×
Observations	32,041	32,041	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

Border effect social connections

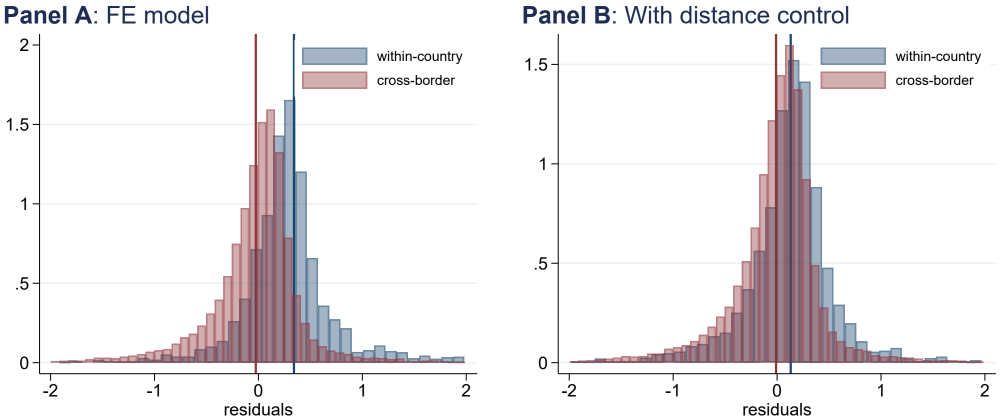


Digital border effect





Border effect residuals



Appendix 4

Signaling of Open Source Software Developers

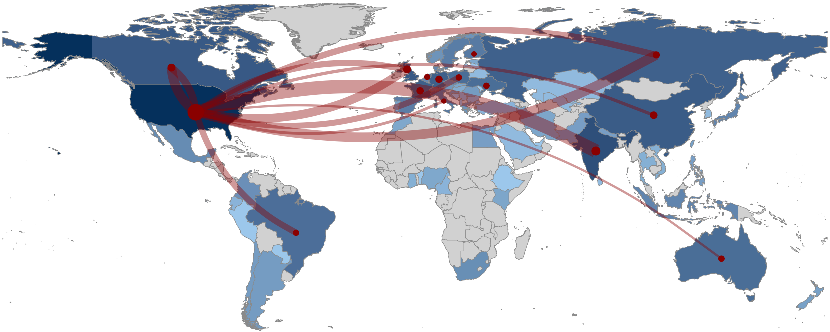
Sample

- high-skilled professionals often move for job (≈2/3 in our data)
- move could confound our effect, so we focus on movers only
- users who move for job (= simultaneously change affiliation) v. relocate for other reasons (and keep their affiliation)
- same snapshots as before, worldwide geolocated users who move once during observation period
- results in 22,896 users worldwide

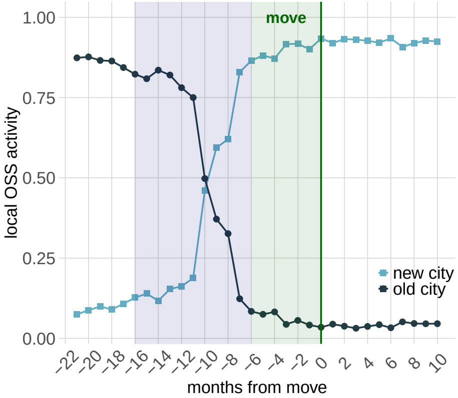
Summary statistics

Median	Movers			
	job	other	Δ	%Δ
Activity				
Commits	163	188	−25	13.3%
commits single projects	*72*	*76*	*−4*	*5.3%*
commits team projects	*59*	*80*	*−21*	*26.3%*
Experience	37	42	−5	11.9%
Collaboration				
Projects	14	16	−2	12.5%
single projects	*9*	*9*	*0*	*0.0%*
team projects	*5*	*6*	*−1*	*16.7%*
Project members	2.21	2.82	−0.61	21.6%
Quality				
Followers	5	5	0	0.0%
Stars	1.10	1.88	−0.78	41.5%
stars single projects	*0.09*	*0.12*	*−0.03*	*25.0%*
Forks	0.62	1.11	−0.49	44.1%
forks single projects	*0*	*0*	*0*	*0.0%*

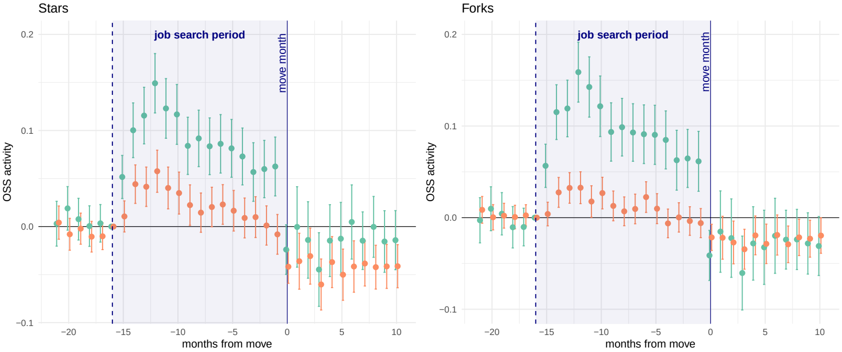
User relocations worldwide, but mostly within the U.S.



User relocation and collaboration



Results: community use value



Project category

IHS(single commits)	(1) low-level	(2) data eng.	(3) app dev.	(4) web dev.	(5) routine	(6) other
Job mover × job search	0.0082**	0.0336***	0.0145***	0.0423***	0.0176***	0.0202***
	(0.0036)	(0.0045)	(0.0029)	(0.0065)	(0.0041)	(0.0036)
Job mover × post move	-0.0034	-0.0146*	-0.0069	-0.0624***	-0.0086	0.0032
	(0.0044)	(0.0057)	(0.0046)	(0.0089)	(0.0052)	(0.0046)
User FE	×	×	×	×	×	×
Month FE	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×
Adjusted R^2	0.15111	0.13901	0.13828	0.16824	0.15890	0.16948
Observations	1,717,200	1,717,200	1,717,200	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896	22,896	22,896	22,896

Programming languages

IHS(single commits)	listed		
	(1) top 30	(2) other	(3) not listed
Job mover × job search	0.0573***	0.0355***	0.0259***
	(0.0055)	(0.0061)	(0.0040)
Job mover × post move	-0.0205	-0.0484***	-0.0136*
	(0.0074)	(0.0084)	(0.0.0051)
User FE	×	×	×
Month FE	×	×	×
Experience FE	×	×	×
Adjusted R^2	0.13533	0.15462	0.13667
Observations	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896

International movers

IHS(single commits)	international		upward moves	
	(1) international	(2) inter-continental	(3) income group	(4) GDP p. c.
Job mover × job search	0.1013***	0.1041***	0.1148***	0.1158***
	(0.0101)	(0.0097)	(0.0094)	(0.0093)
Job mover × job search × indicator	0.0561**	0.0739***	0.0320	0.0398
	(0.0167)	(0.0199)	(0.0263)	(0.0302)
Job mover × post move	-0.0817***	-0.0815***	-0.0816***	-0.0815***
	(0.0123)	(0.0123)	(0.0123)	(0.0123)
User FE	×	×	×	×
Month FE	×	×	×	×
Experience FE	×	×	×	×
Adjusted R^2	0.21735	0.21736	0.35945	0.35945
Observations	1,717,200	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896	22,896

Affiliation moves

IHS(single commits)	destination			origin	
	(1) median	(2) big tech	(3) academia	(4) median	(5) academia
Job mover × job search	0.1354***	0.1255***	0.1115***	0.1204***	0.1115***
	(0.0129)	(0.0093)	(0.0093)	(0.0093)	(0.0342)
Job mover × job search × indicator	-0.0280*	-0.1085***	0.0702**	-0.0155	0.0702**
	(0.0151)	(0.0270)	(0.0273)	(0.0429)	(0.0273)
Job mover × post move	-0.0815***	-0.0818***	-0.0813***	-0.1082**	-0.0813***
	(0.0123)	(0.0123)	(0.0123)	(0.0429)	(0.0123)
User FE	×	×	×	×	×
Month FE	×	×	×	×	×
Experience FE	×	×	×	×	×
Adjusted R^2	0.21733	0.21736	0.21734	0.21875	0.21734
Observations	1,717,200	1,717,200	1,717,200	1,249,275	1,717,200
Users	22,896	22,896	22,896	22,896	22,896

Job transitions

Affiliation	all movers	job movers	other movers	Δ
Largest 100 firms	28.9 %	28.9 %	27.2 %	+1.7 p.p.
Big tech	*7.2 %*	*7.3 %*	*4.9 %*	*+2.4 p.p.*
Academic	8.9 %	9.0 %	6.3 %	+2.7 p.p.
Other	55.1 %	54.8 %	61.6 %	-6.8 p.p.
Job transitions	anytime	origin	destination	Δ
Largest 100 firms	28.9 %	20.3 %	26.8 %	+6.5 p.p.
Big tech	*7.2 %*	*2.0 %*	*7.1 %*	*+5.1 p.p.*
Academic	8.9 %	9.1 %	7.2 %	-2.0 p.p.
Other	55.1 %	68.6 %	58.9 %	-9.6 p.p.

Top origin/destination cities

Origin	Users	Share	Destination	Users	Share
New York, USA	650	2.84 %	San Francisco, USA	1,307	5.71 %
San Francisco, USA	618	2.70 %	New York, USA	936	4.09 %
London, UK	421	1.84 %	London, UK	763	3.33 %
Bangalore, India	325	1.42 %	Seattle, USA	708	3.09 %
Chicago, USA	311	1.36 %	Bangalore, India	559	2.44 %
Boston, USA	305	1.33 %	Los Angeles, USA	379	1.66 %
Los Angeles, USA	305	1.33 %	Austin, USA	345	1.51 %
Moscow, Russia	305	1.33 %	Toronto, Canada	331	1.45 %
Seattle, USA	273	1.19 %	Chicago, USA	318	1.39 %
Paris, France	247	1.08 %	Boston, USA	315	1.38 %
Cumulative share		15.09 %	Cumulative share		26.05 %

Domestic moves

Country	Users	Share	
		all	domestic
United States	10,348	45.20 %	63.49 %
India	1,219	5.32 %	7.48 %
United Kingdom	638	2.79 %	3.91 %
Canada	620	2.71 %	3.80 %
China	522	2.28 %	3.20 %
France	436	1.90 %	2.68 %
Germany	417	1.82 %	2.56 %
Russia	375	1.64 %	2.30 %
Poland	195	0.85 %	1.20 %
Australia	194	0.85 %	1.19 %
		65.36 %	91.81 %

Frequent affiliations

Origin	Share	Destination	Share
Student	0.92 %	Microsoft	2.08 %
Microsoft	0.72 %	Google	2.00 %
University of Washington	0.62 %	Amazon	1.37 %
Freelancer	0.51 %	Facebook	1.00 %
IBM	0.41 %	Red Hat	0.64 %
New York University	0.41 %	Shopify	0.44 %
University of California	0.41 %	IBM	0.37 %
University of Florida	0.41 %	Stanford University	0.31 %
University of Oxford	0.41 %	LinkedIn	0.28 %
Amazon	0.31 %	Apple	0.26 %
	5.13 %		8.75 %

Model specification

Model class:	OLS				LPM	NB	PPML
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	log	ihs	ihs	ihs	dummy	count	count
Sample:	full	full	geo	change	full	full	full
Job mover × job search	0.0942***	0.1177***	0.1182**	0.1044***	0.0534***	0.5919***	0.1553**
	(0.0075)	(0.0091)	(0.0091)	(0.0362)	(0.0030)	(0.0290)	(0.0614)
Job mover × post move	-0.0663***	-0.0813***	-0.0804***	-0.1744***	-0.0294***	-0.3090***	-0.2914**
	(0.0102)	(0.0123)	(0.0123)	(0.0526)	(0.0041)	(0.0459)	(0.0695)
User FE	×	×	×	×	×	×	×
Month FE	×	×	×	×	×	×	×
Experience FE	×	×	×	×	×	×	×
Adjusted R^2	0.21493	0.21732	0.21720	0.24840	0.21681		
Observations	1,717,200	1,717,200	1,717,200	66,375	1,717,200	1,401,002	1,401,002
# User FE	22,896	22,896	22,838	885	22,896	22,896	22,896

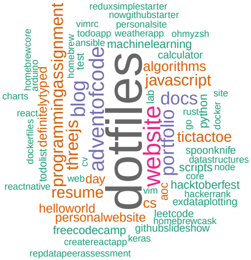
Initial forks and ownership

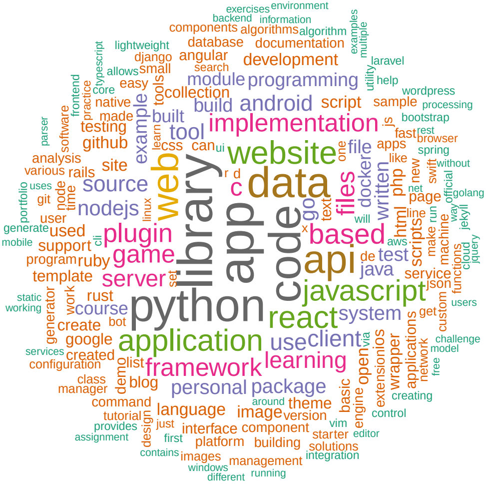
IHS(single commits)	project owner		
	(1) own	(2) non-own	(3) no initial forks
Job mover × job search	0.0980***	0.0227***	0.1036***
	(0.0087)	(0.0035)	((0.0091)
Job mover × post move	-0.0845***	0.0227	-0.0815***
	(0.0116)	(0.0035)	((0.0122)
User FE	×	×	×
Month FE	×	×	×
Experience FE	×	×	×
Adjusted R^2	0.20781	0.14464	0.19764
Observations	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896

Keyword-based project classification

IHS(single commits)	(1) education	(2) data science	(3) website	(4) code
Job mover × job search	0.0050***	0.0028***	0.0090***	0.0403***
	(0.0014)	(0.0007)	(0.0038)	(0.0054)
Job mover × post move	-0.0027*	-0.0003	-0.0038	-0.0439***
	(0.0016)	(0.0008)	(0.0031)	(0.0075)
User FE	×	×	×	×
Month FE	×	×	×	×
Experience FE	×	×	×	×
Adjusted R^2	0.03966	0.04912	0.11611	0.14795
Observations	1,717,200	1,717,200	1,717,200	1,717,200
Users	22,896	22,896	22,896	22,896

Description keywords





User popularity

Project age


job search period
move month
OSS activity
0.06
0.03
0.00
−0.03
−20
−15
−10
−5
0
5
10
months from move

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

Seliger, Florian, Jan Kozak, and Gaétan de Rassenfosse, "Geocoding of Worldwide Patent Data," 2019.