CPGIS Educational Webinar Series
"Spatiotemporal Study of Urban Dynamics"

Co-sponsors:

International Association of Chinese Professionals in Geographical Information Sciences
NSF Spatiotemporal Innovation Center
Urban Institute, Tsinghua University
School of Geography, Jiangxi Normal University
China Data Institute & Future Data Lab
Journal of Computational Urban Science

9:00PM-10:00 PM, Thursday, Feb 25-May 6, 2021 (US Eastern Time)
10:00AM-11:00 AM, Friday, Feb 26-May 7, 2021 (Beijing Time)

Measuring Place Connectivity using Big Social Media Data

Zhenlong Li

Associate Professor, Department of Geography

Director, Geoinformation and Big Data Research Laboratory (GIBD)

Big Data Health Science Center, Hazards and Vulnerability Research Institute

University of South Carolina, USA

http://gis.cas.sc.edu/gibd

Tao Hu

Postdoc Research Fellow, Center for Geographic Analysis

Harvard University, USA

taohu@fas.Harvard.edu

Agenda

- A brief introduction to the GIBD lab and social media analytics
- ODT flow system for extracting, analyzing, and sharing multi-source multi-scale human mobility
- Measuring global multi-scale place connectivity based on movement derived from geotagged tweets
- Workflow demonstrations of using derived mobility and connectivity datasets shared through APIs (Dr. Tao Hu)
- Discussion

Geospatial Big Data Research at GIBD Lab @ USC

Big Social Media Data

Big Remote Sensing Data

Big Health Data

Big Climate Data

…

Innovations

Methods, Algorithms, Framework, Tools

Combine GIS with Big Data Analytics

Disaster Management

Human Mobility

Public Health

Climate Analysis

…

Sponsors:

Big Data Health Science Center
University of South Carolina

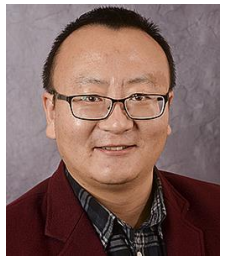
HVRI
HAZARDS & VULNERABILITY RESEARCH INSTITUTE
University of South Carolina

http://gis.cas.sc.edu/gibd

GIBD

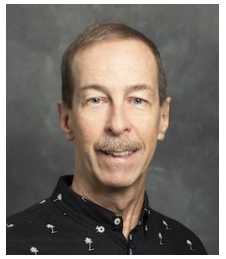
Enabled by a Strong Interdisciplinary Team

Lab Core Faculty


Dr. Zhenlong Li

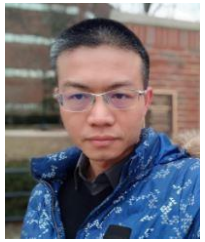
Dr. Susan Cutter

Dr. Cuizhen Wang


Dr. Michael Hodgson

Students

Yuqin Jiang


Huan Ning

Seth Church

Grayson Morgan

Alumni

Dr. Xiao Huang
(UARK)

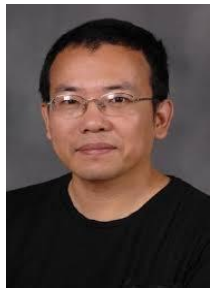
Dr. Yago Martin
(UCF)

Dong Xu
(ECNU)

Collaborators
(not a full list)

Dr. Xiaoming Li
(USC)

Dr. Dwayne Porter
(USC)

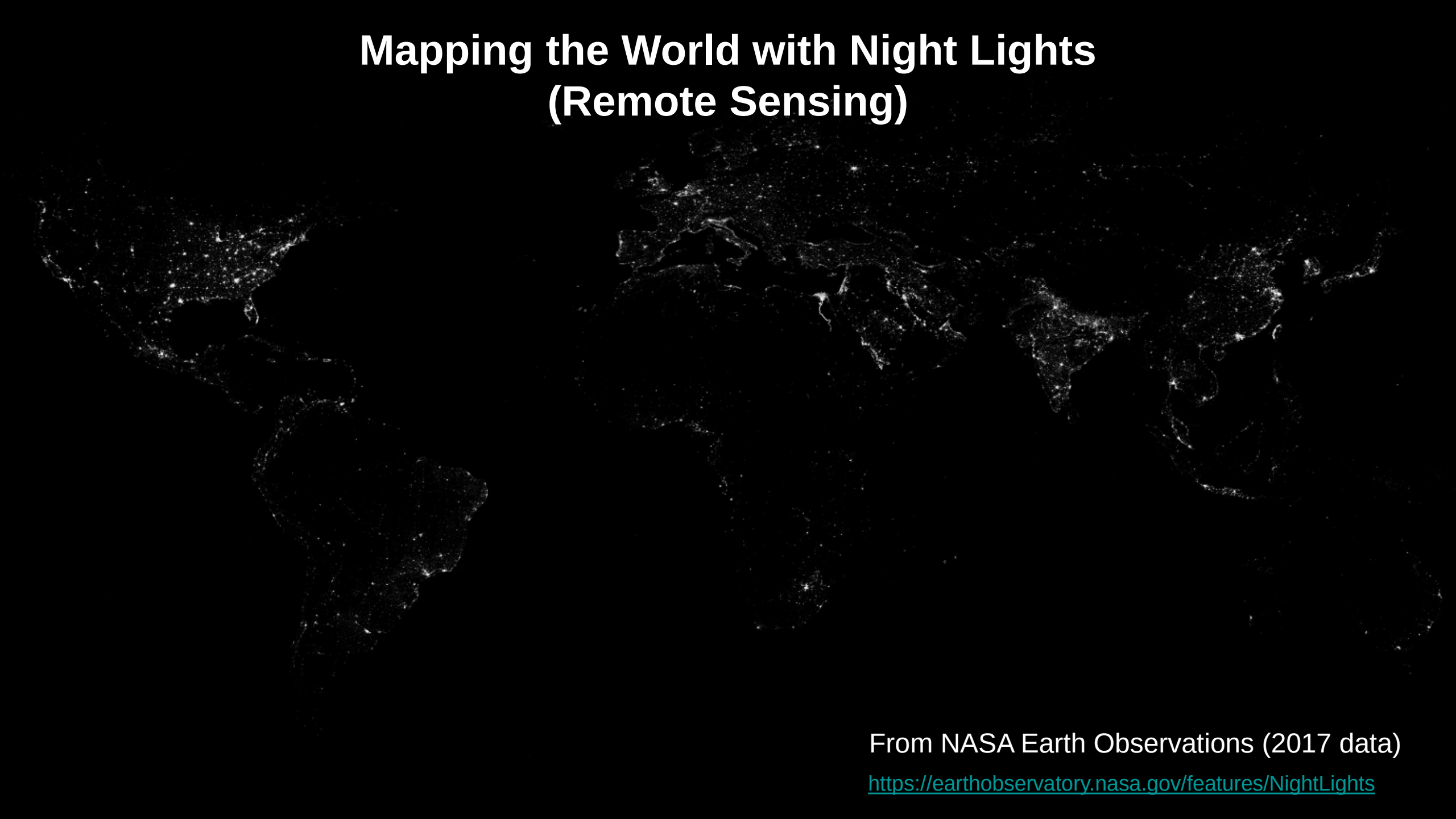

Dr. Xinyue Ye
(Texas A&M)

Dr. Tao Hu
(Harvard)

Dr. Chris Emrich
(UCF)

Dr. Sathish Kumar
(CCU)

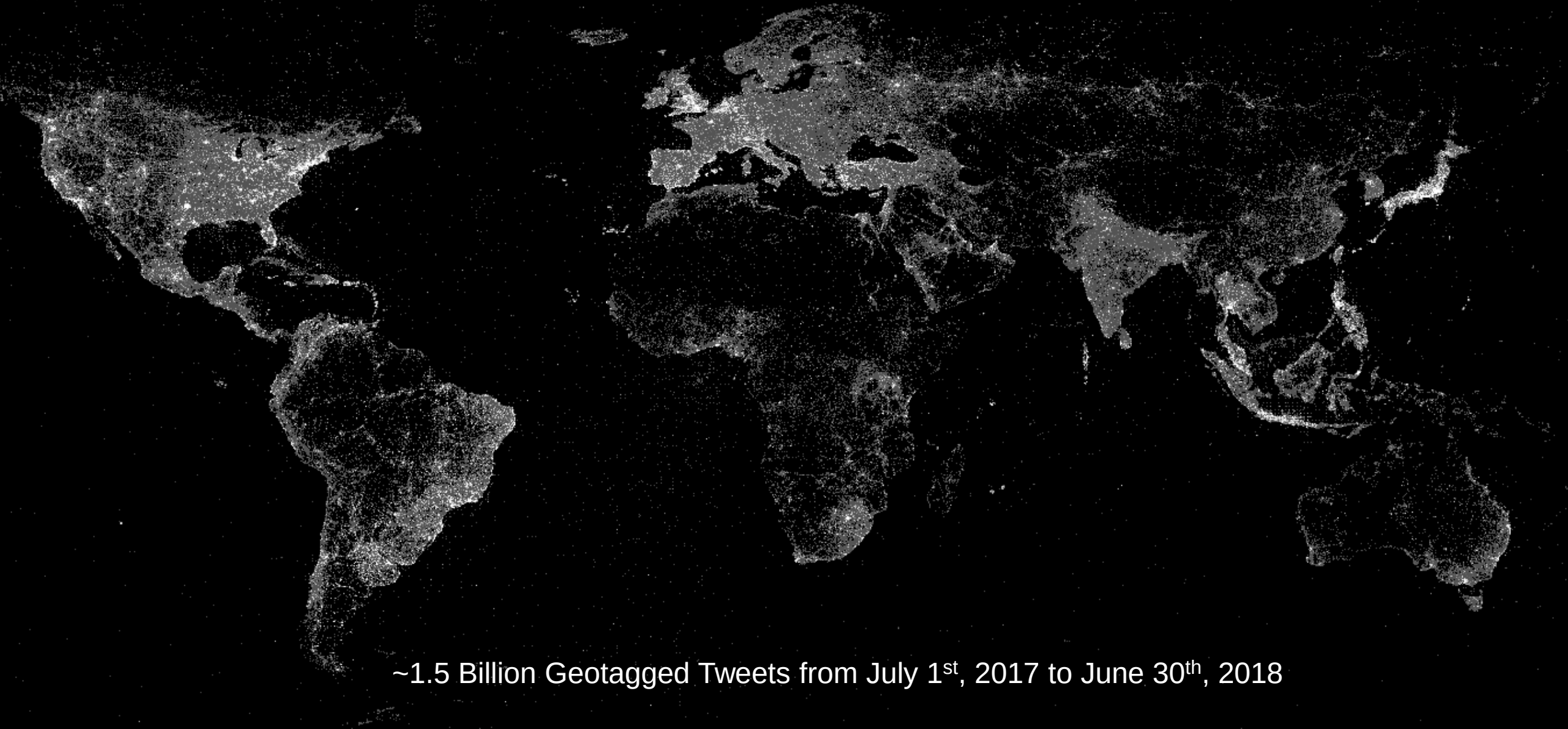
Mapping the World with Night Lights (Remote Sensing)



From NASA Earth Observations (2017 data)

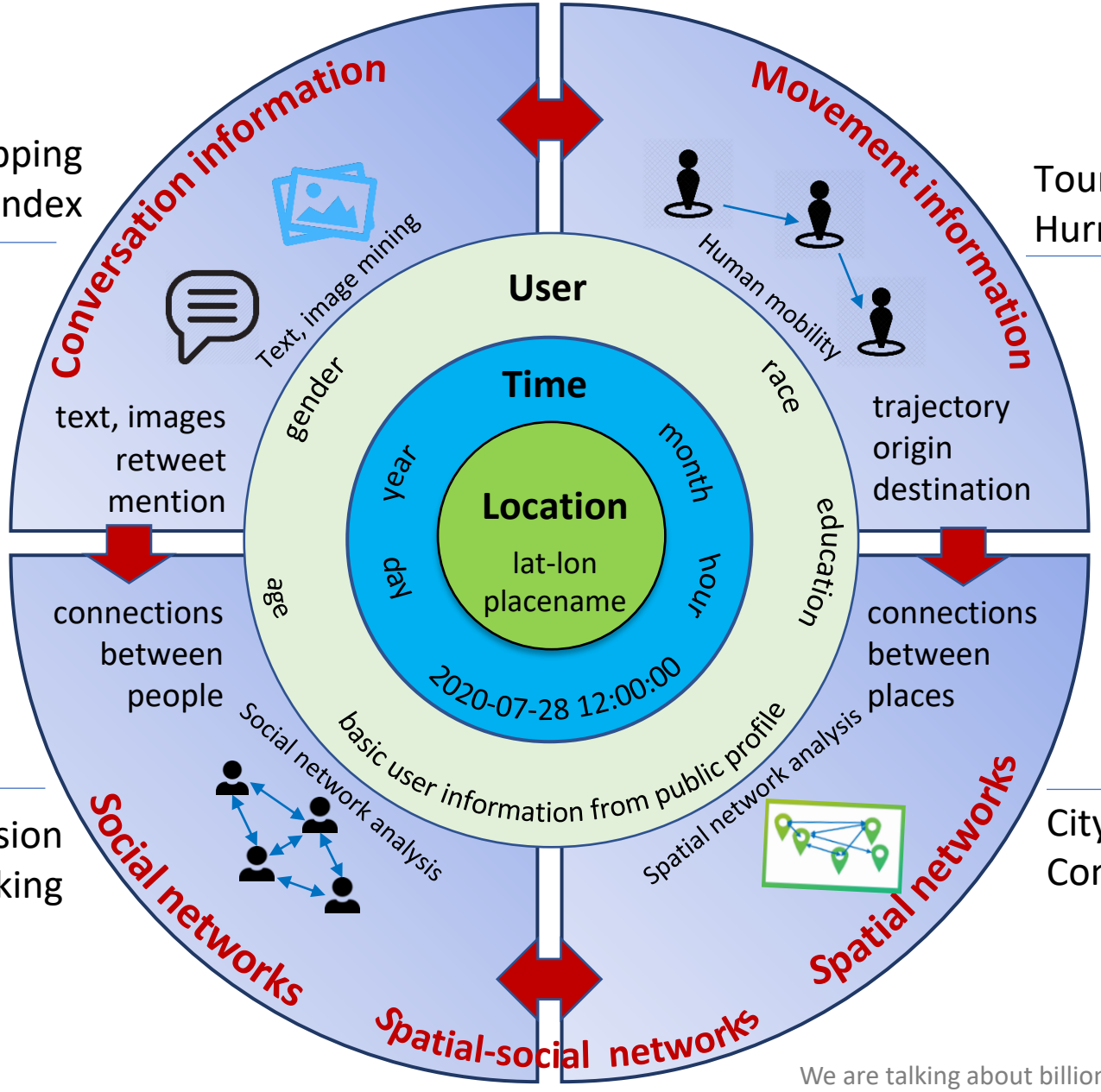
https://earthobservatory.nasa.gov/features/NightLights

Mapping the World with Social Media Data (Social Sensing)



~1.5 Billion Geotagged Tweets from July 1st, 2017 to June 30th, 2018

What information can we extract from geotagged social media?



Flood mapping
HIV risk behavior index

Tourist movement
Hurricane evacuation compliance

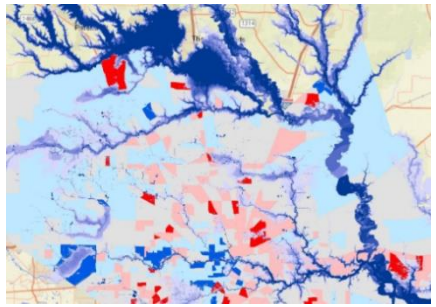
Information diffusion
Evacuation decision making

City/place connectivity
Commute patterns

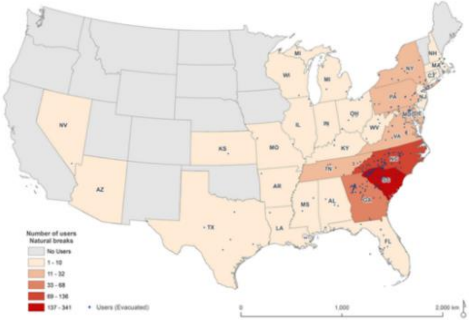
We are talking about billions of geotagged tweets and millions of Twitter users.

What research have we done with social media data?

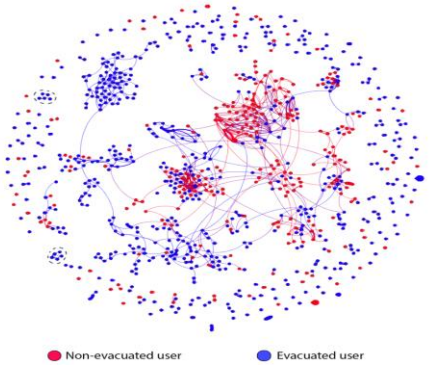
Rapid flood mapping for situational awareness



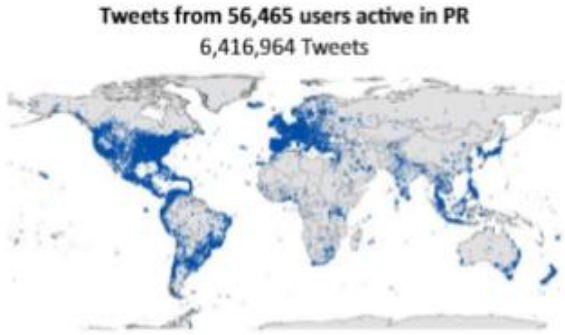
Gauge hurricane evacuation compliance



Understand evacuation decision making

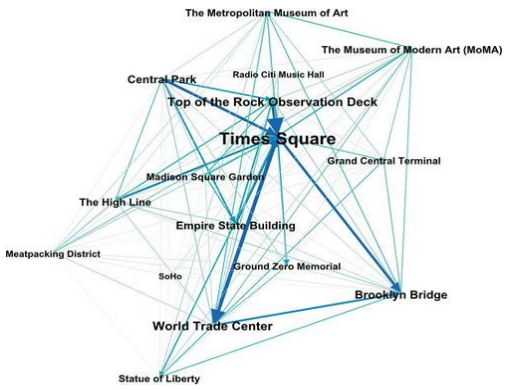


Population migration caused by disasters

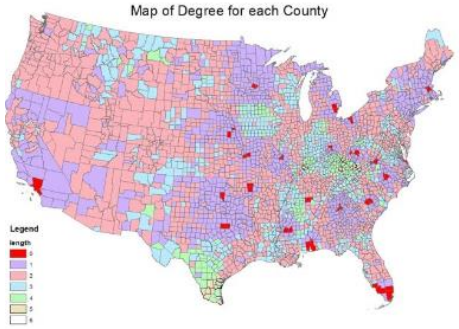


Spatiotemporal Studies of Urban Dynamics

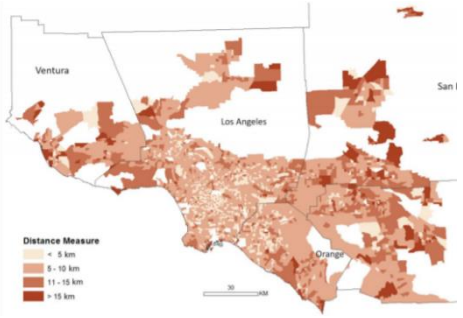
Study tourist movement patterns (NYC)



Measure inter-city spatial networks



Measure and model human activity space (LA)



Human mobility during disruptive events

Find out more at http://gis.cas.sc.edu/gibd/research

Some background: Human Movement and Place Connectivity

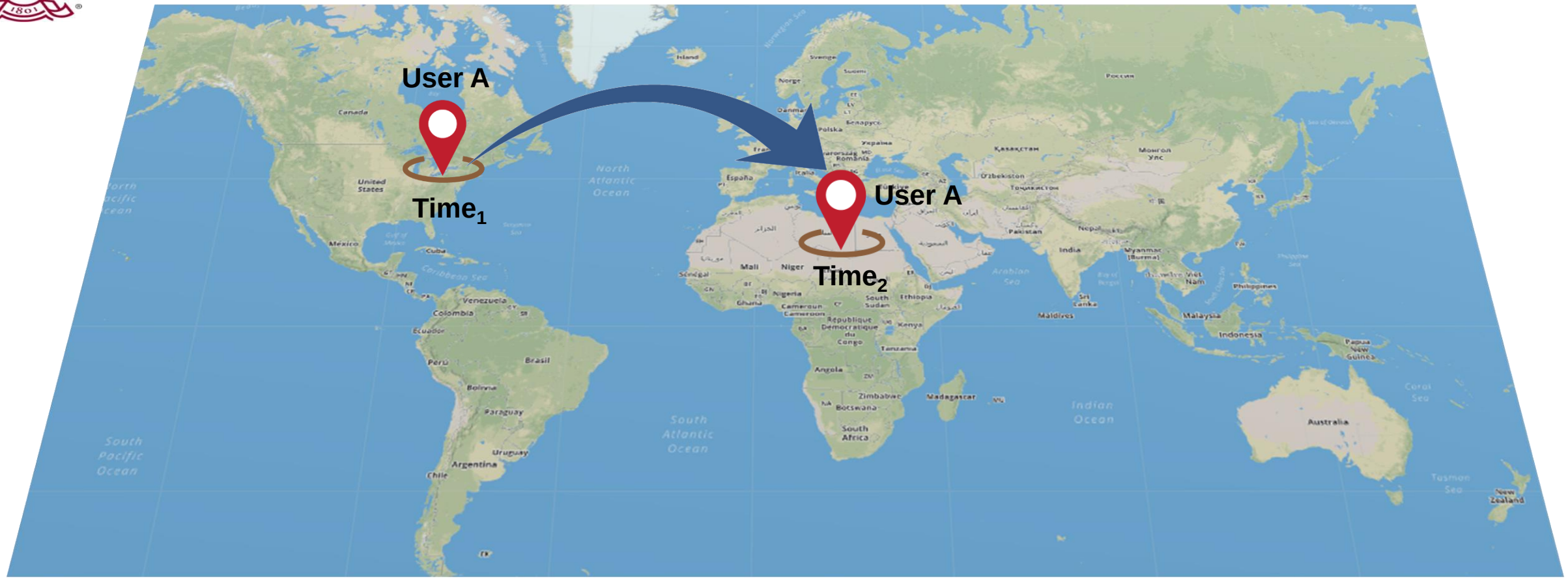
- Understanding human mobility dynamics among places benefits a wide range of applications that in need of knowledge in human spatial interactions.
- People move across places (due to attractions of places, social and physical →forces). Such movement shapes the relationships (i.e., connectivity) among places., which can be observed and quantified.
- Human movement, city/place connections have been studied by many (e.g., Salt, 1987; Derudder and Witlox, 2008; Xu and Harriss, 2008; Barcus and Brunn, 2010; Liu et al., 2014; Boyle, 2014; Hu et al., 2017; Lin et al., 2019; Ye et al., 2020…).

- This work aims to quantify the global multi-scale population movement, and using such movement to further derive a spatiotemporal-continuous place connectivity index, and to promote data sharing with the hope to facilitate reproducibility & replicability in the Big Data research.

Agenda

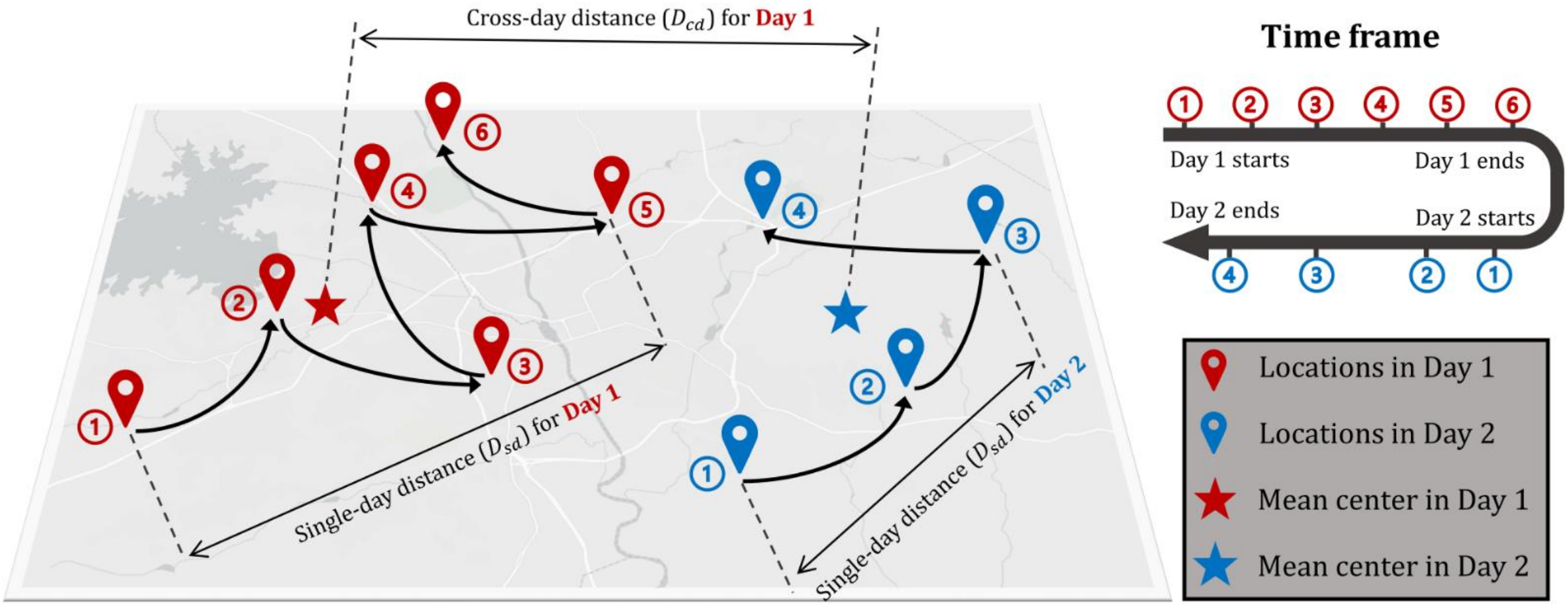
- A brief introduction to GIBD lab and social media analytics
- **ODT flow system for extracting, analyzing, and sharing multi-source multi-scale human mobility**
- Measuring global multi-scale place connectivity based on movement derived from geotagged tweets
- Workflow demonstrations of using derived mobility and connectivity datasets shared through APIs (Dr. Tao Hu)
- Discussion

Extract population flows from social media users



We have ~**1 million unique Twitter users** on a daily basis.

Single-day and Cross-day Movements based on Geotagged Tweets

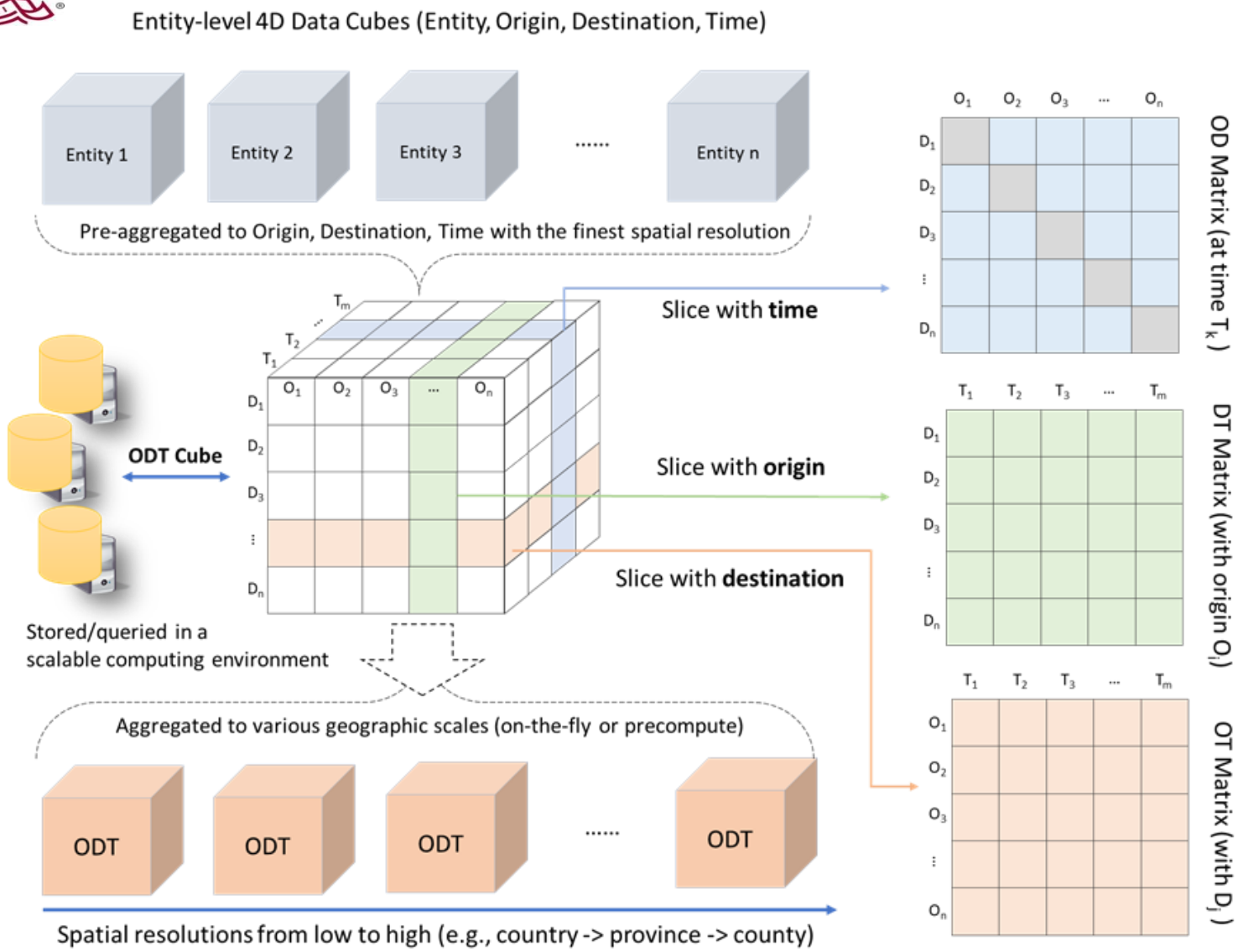


Works on other sparse trajectory data as well.

Data-intensive and computing-intensive process

Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PloS one, 15(11), e0241957.

Addressing Big Data Challenges



Origin-Destination-Time (ODT) data model coupled with a scalable big data computing cluster

Apache Hadoop, Hive, Impala, and Esri GIS Tools for Hadoop

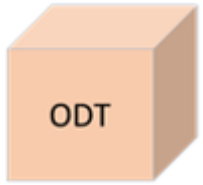
Li Z., Huang X., Hu T., Ning H., Ye X., Li X., (2021). ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility, preprint

ODT-based big data approach enable us to handle different data sources in a unified way

- We computed the daily OD flows for 2019 and 2020 using worldwide geotagged tweets.
- We further computed the daily OD flows from SafeGraph Social Distancing Metrics data[1].

Statistics of the derived daily OD flows from Twitter data and SafeGraph data

	Twitter-derived OD Flow	**SafeGraph-derived OD Flow**
Spatial coverage	Worldwide	U.S.
Temporal coverage	2019-2020 (daily)	2019-2020 (daily)
Original data records	2,695,552,594 geotagged tweets by 24,863,844 Twitter users	160,301,510 SafeGraph social distancing metrics records
Derived Entity-ODT	636,984,772	11,108,696,071
World country/territory	1,253,291	—
World 1st level division	9,333,761	—
U.S. state	809,741	1,958,450
U.S. county	10,206,119	439,790,381
U.S. census tract	—	6,710,889,890



[1]https://docs.safegraph.com/docs/social-distancing-metrics

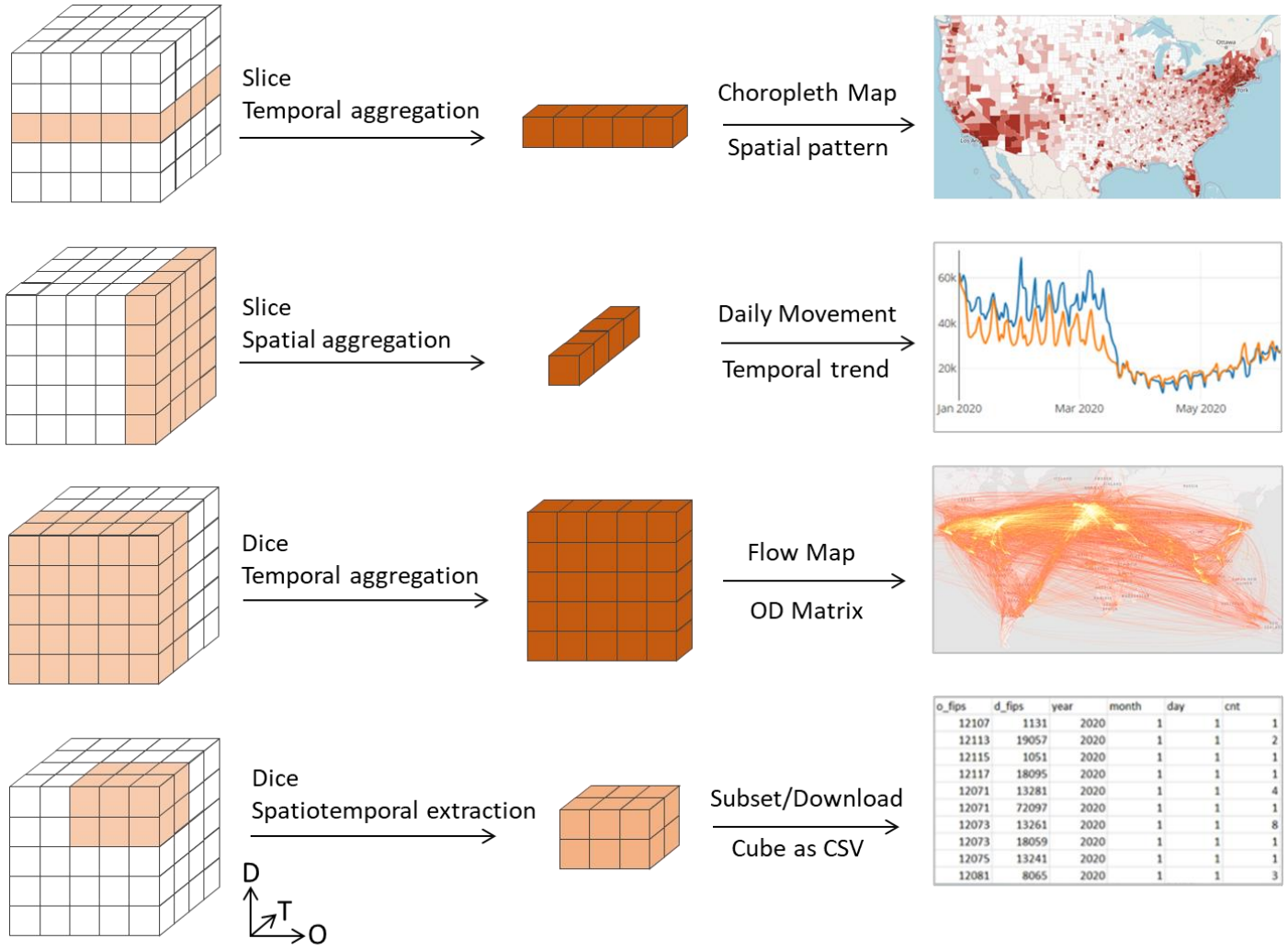
ODT-based human mobility analysis powered by HPC

Four application scenarios exemplifying how ODT coupled with the traditional data cube operations and HPC can help analyze big mobility data.

Apache Impala

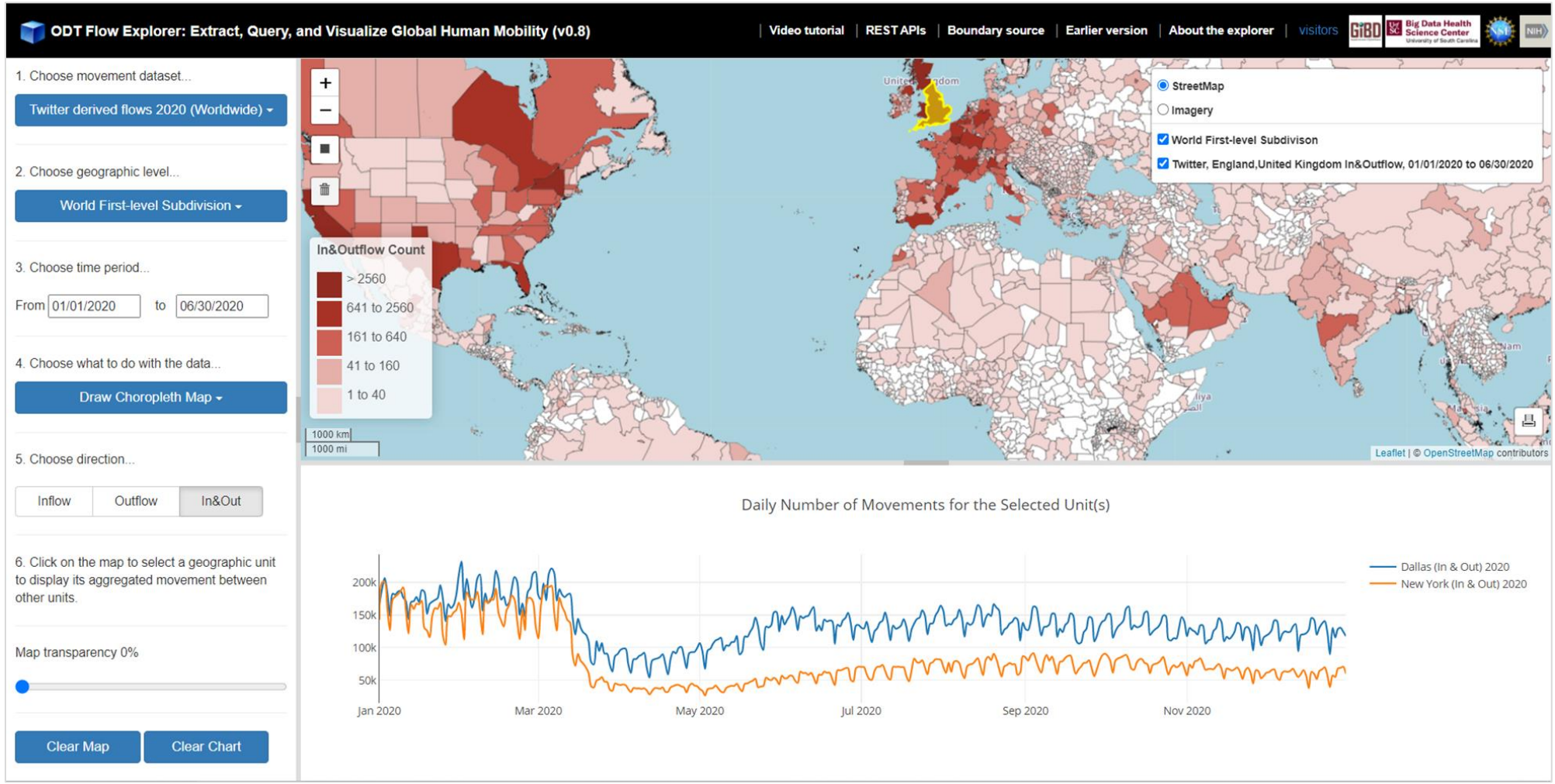
Providing low latency and high concurrency for BI/analytic queries on Hadoop.
Open source.

Parallel query analytics



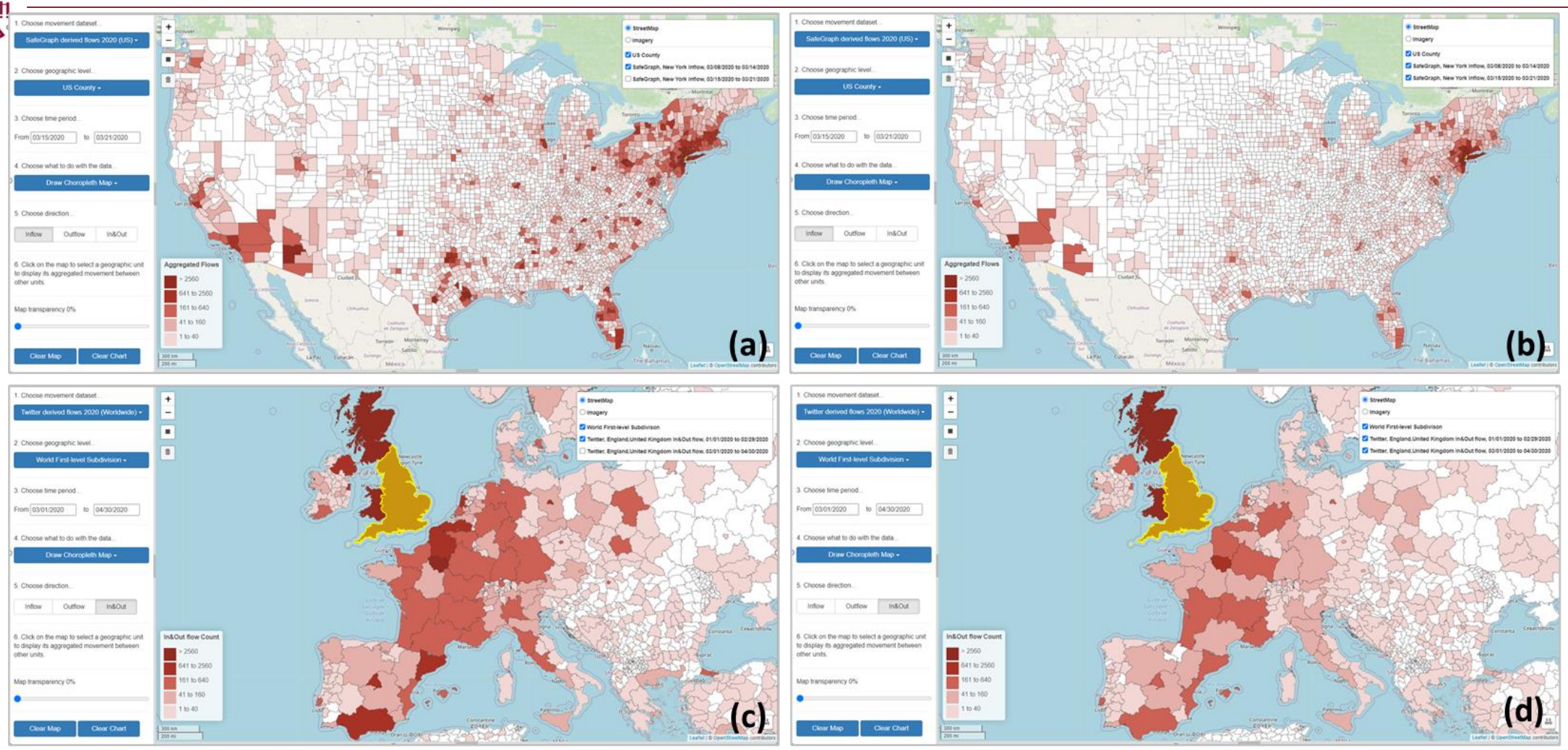
Explore and visualize the derived multi-source multi-scale OD flows

ODT Flow Explorer: an interactive spatial web portal for on-demand querying, slicing, aggregating, and visualizing the billion-level OD flows (backed by HPC)

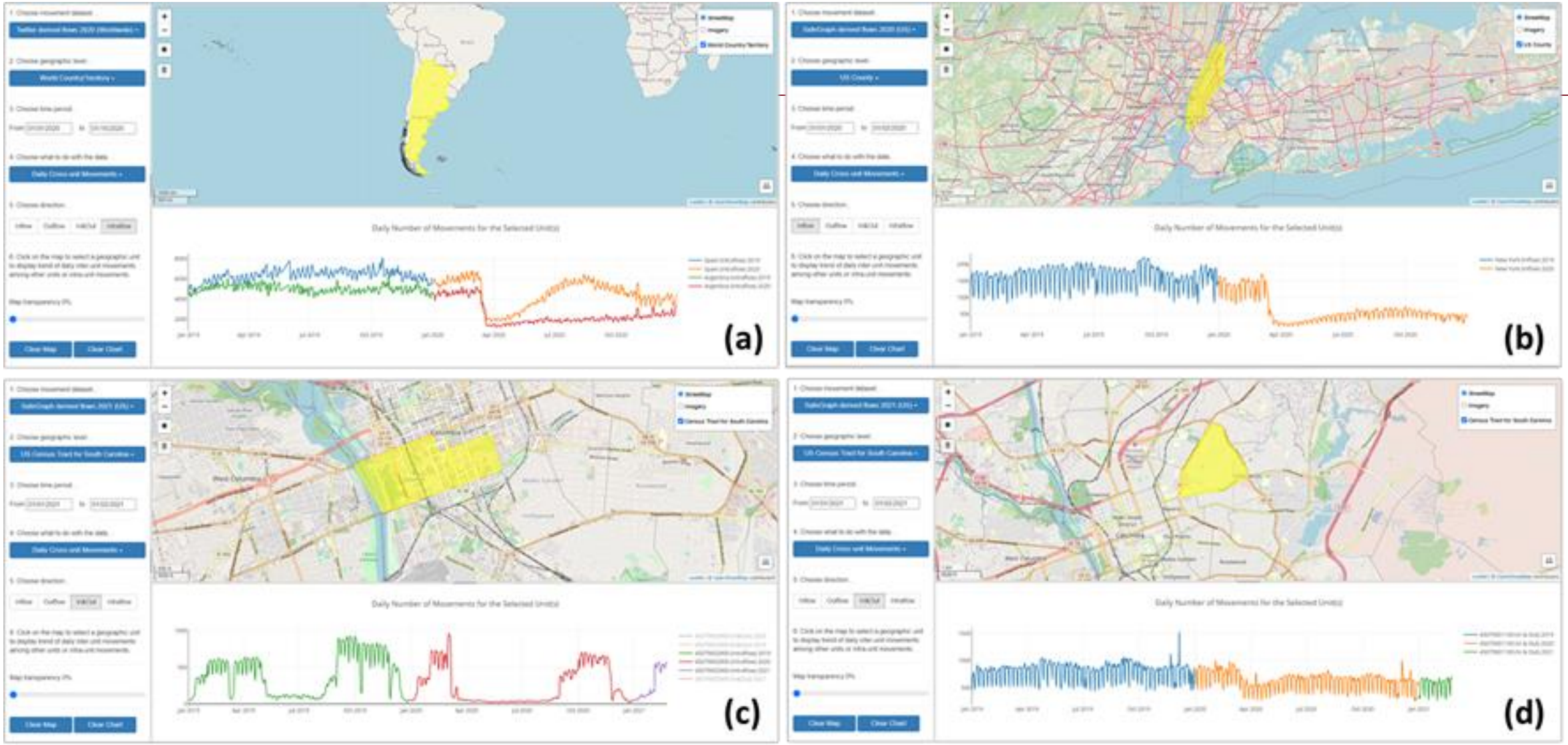


http://gis.cas.sc.edu/GeoAnalytics/od.html

Visual analytics of the COVID-19 impact on human mobility

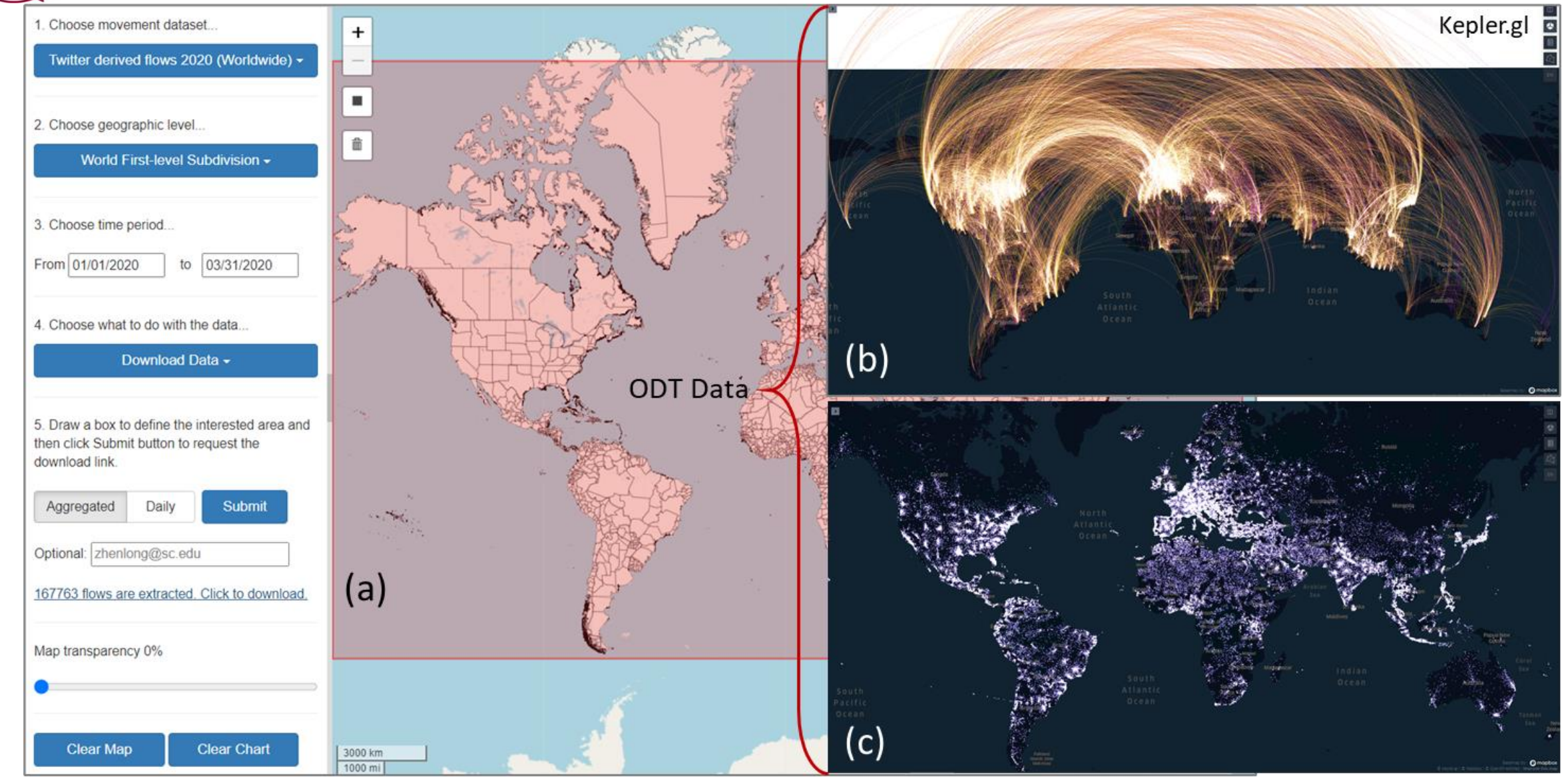


- SafeGraph-derived county population flows **to** New York County from (a) 03/08/2020 to 03/14/2020 and (b) for the following week (03/15/2020 to 03/21/2020).
- Twitter-derived in & out flows between England, UK and other first-level administrative units in the Europe area for (c) 01/01/2020 to 02/29/2020, and (d) 03/01/2020 to 04/30/2020



- (a) Intraflow for Spain (top line) and Argentina (bottom line) in 2019 and 2020;
- (b) Inflow for New York County, U.S. in 2019 and 2020;
- (c) Intraflow for a census tract in Columbia, South Carolina (mainly located within the USC) from 01/01/2019 to 02/24/2021;
- (d) Intraflow for a census tract in a residential area of Columbia from 01/01/2019 to 02/24/2021.

Extract and download flow data with user-defined spatiotemporal constraints



Access flow data programmatically using the ODT Flow REST APIs

ODT Flow REST APIs

Each API performs a specific task such as aggregating the flows for a selected place and downloading flow data for a selected geographic area. All APIs return data in CSV (comma-separated values) format. The API is specified in the "operation" parameter in the request (see examples below).

APIs

- **get_flow_by_place**

Return the aggregated movement between the selected place and other places.

- **get_daily_movement_by_place**

Return the daily inter-unit movements between the selected place and other places or the selected place's daily intra-unit movements.

- **get_daily_movement_for_all_places**

Return the daily movements for all places of a specific geographic level (currently return intra movement).

- **extract_odt_data**

Return the selected OD flows in either temporally aggregated format or daily format. The study area can be specified by a bbox. For SafeGraph daily flows, the days selected need be less than 31.

- **extract_odt_data_url**

Same as extract_odt_data, but returns a download URL and number of records instead of directly returning the csv data. Works better for extracting large amounts of flows.

extract_odt_data

Return the selected OD flows in either temporally aggregated format or daily format. The study area can be s
selected need be less than 31.

In [11]:
```
# set the parameters of your interested data, including operation, scale, source, place..
params = {"operation": "extract_odt_data",
          "source": "twitter",
          "scale": "us_county",
          "begin": "04/01/2019" ,
          "end": "04/15/2019",
          "bbox": "-90,90,-180,180",
          "type": "daily"}

# obtain data using REST APIs
q = r'http://gis.cas.sc.edu/GeoAnalytics/REST'
r = requests.get(q, params=params)

# put the data into a Pandas DataFrame
df = pd.read_csv(StringIO(r.text))
df
```

Out[11]:

	o_place	d_place	year	month	day	cnt	o_lat	o_lon	d_lat	d_lon
0	21115	21115	2019	4	8	5	37.811	-82.816	37.811	-82.816
1	1099	1001	2019	4	7	1	31.523	-87.335	32.576	-86.681
2	36029	36121	2019	4	12	1	42.969	-78.582	42.867	-78.362
3	17109	17031	2019	4	9	1	40.460	-90.674	42.020	-87.772
4	51550	51550	2019	4	10	100	36.761	-76.289	36.762	-76.294
5	51041	51760	2019	4	12	13	37.441	-77.531	37.532	-77.493
6	49057	49011	2019	4	10	4	41.201	-111.990	41.128	-111.997
7	13121	39001	2019	4	2	1	33.740	-84.449	38.906	-83.347
8	18127	18167	2019	4	8	1	41.499	-87.067	39.486	-87.409
9	26125	42091	2019	4	8	1	42.491	-83.143	40.124	-75.458
10	39003	39095	2019	4	9	1	40.887	-83.899	41.657	-83.575
11	24013	24003	2019	4	3	1	39.577	-76.998	39.133	-76.625
12	37135	37101	2019	4	15	1	35.927	-79.087	35.723	-78.418

https://github.com/GIBDUSC/ODT_Flow

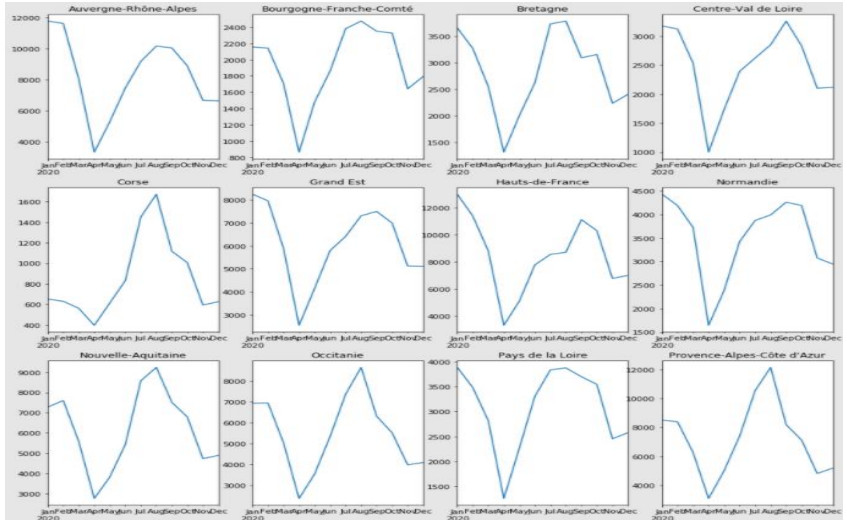
ODT Flow APIs and Jupyter Notebook Case study 1
Visual analytics of the impact of COVID-19 on human mobility in France in 2020

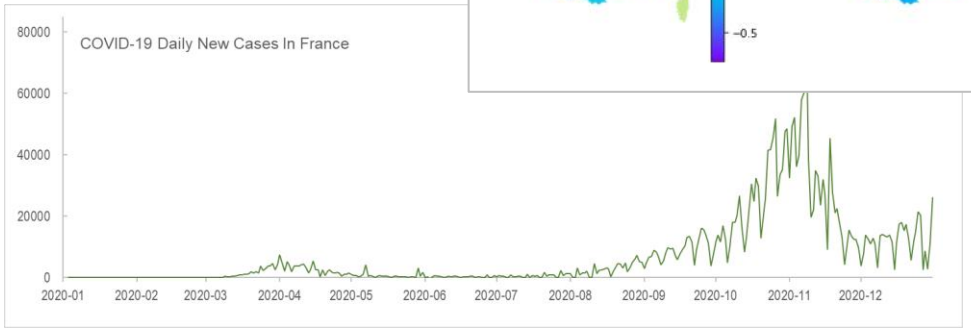
Read the boundary file

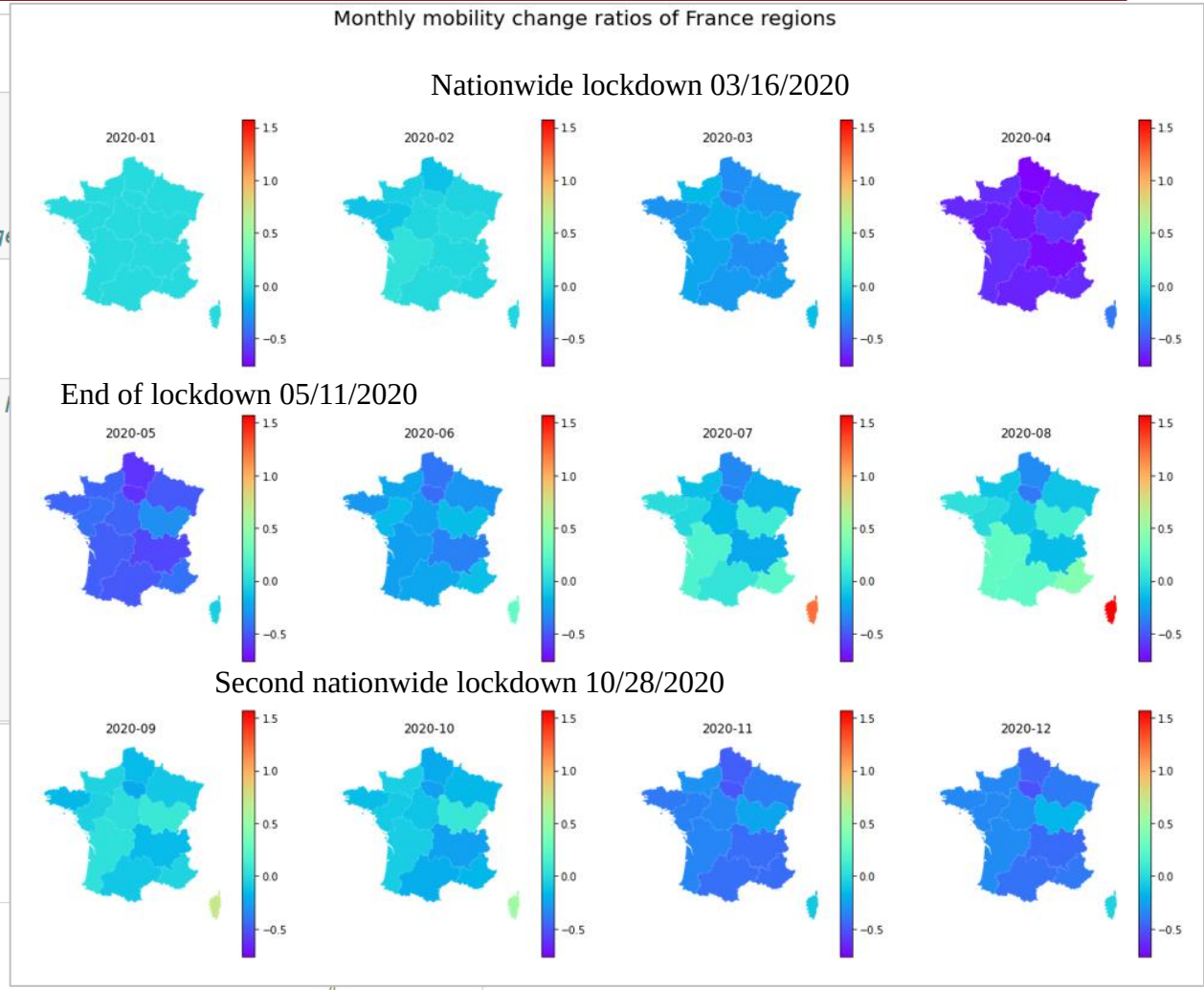
```
subdivision_file = r'gadm01_simplified/gadm36_1.shp'
gdf = gpd.read_file(subdivision_file)

target_place = r'FRA'    # set France as the target place (ISO code)
gdf_country = gdf[gdf['GID_1'].str[:3] == target_place] # Extract the boundary of the targe
```

Obtain 2020 flow data using the ODT Flow API

```
q = r'http://gis.cas.sc.edu/GeoAnalytics/REST'  #Set query url and parameters for the ODT
params = {"operation": "get_daily_movement_for_all_places",
          "scale": "world_first_level_admin",
          "source": "twitter",
          "begin": "01/01/2020",
          "end": "12/31/2020"
      }
r = requests.get(q, params=params) #Submit request
df = pd.read_csv(StringIO(r.text))

df = df[df['place'].str[:3] == target_place] # Extract flows of the target place
```







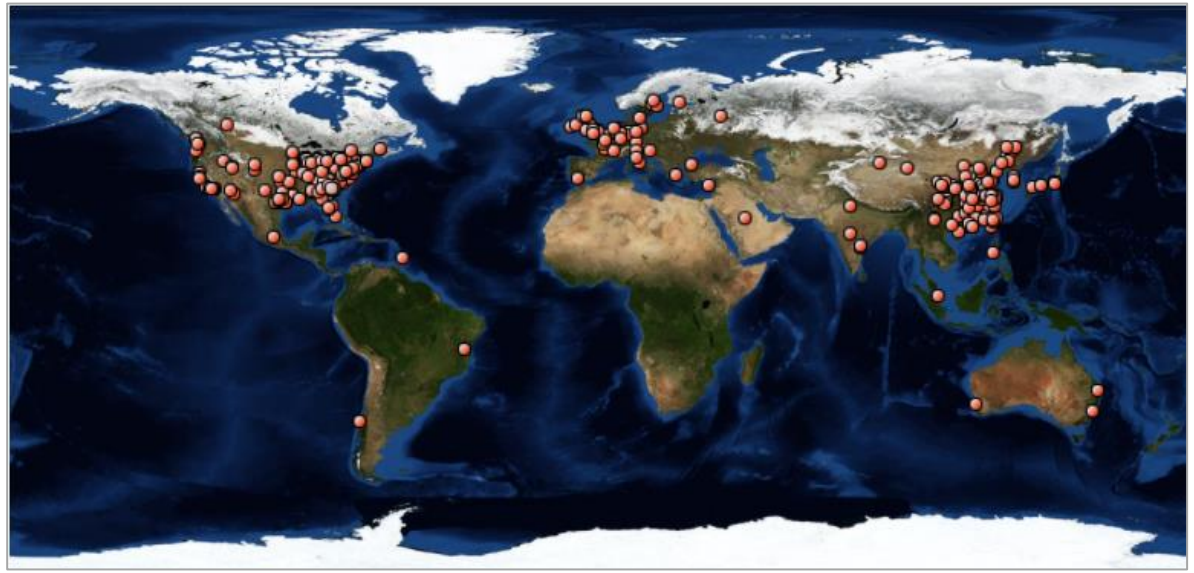
ODT Flow APIs and Jupyter Notebook Case study 2
Interactive visualization of massive flows using ODT Flow APIs and Kepler.gl

```
import os
import pandas as pd
import numpy as np
import requests
from io import StringIO
import pandas as pd
from keplergl import KeplerGl
import geopandas as gpd
import json
```

```
q = r'http://gis.cas.sc.edu/GeoAnalytics/REST' #Set query url and parameters for the ODT REST API
params = {"operation": "extract_odt_data",
       "source": "twitter",
        "scale": "world_first_level_admin",
       "begin": "01/01/2020" ,
       "end": "01/05/2020",
       "bbox": "-90,90,-180,180",
       "type": "aggregated"}

r = requests.get(q, params=params) #Submit request
df = pd.read_csv(StringIO(r.text))

flowMap = KeplerGl(height=500,config=flowConfig) # Create a Kepler.gl map
flowMap.add_data(data=df, name='Interactive Flow Map')
flowMap # Show the interactive map
```

The ODT flow data have been used by other researchers around the world

The ODT Flow Explorer has attracted over 2100 visits worldwide from 31 countries, served about 1 billion flow extractions



http://gis.cas.sc.edu/GeoAnalytics/od.html

WorldPop

UNIVERSITY OF Southampton

February 24th, 2021

Preliminary risk analysis of the international spread of new COVID-19 variants, lineage B.1.1.7, B.1.351 and P.1

Shengjie Lai, Jessica Floyd, Andrew Tatem

WorldPop, School of Geography and Environmental Science, University of Southampton, UK

Email: Shengjie.Lai@soton.ac.uk; or A.J.Tatem@soton.ac.uk

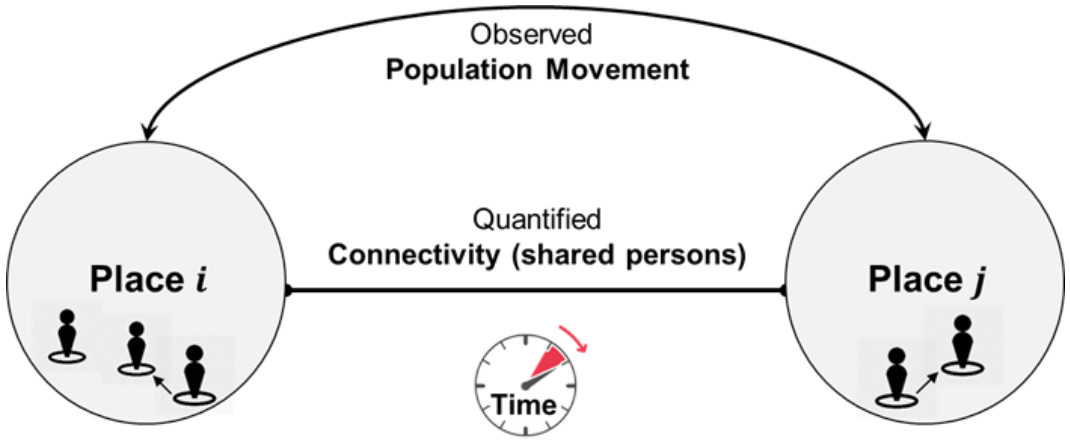
The novel SARS-CoV-2 variants have raised serious concerns about a new wave of the pandemic and the effectiveness of vaccines, particularly concerning the recent new strains (lineages B.1.1.7, B.1.351, and P.1) found in the UK, South Africa, and Brazil [1-3]. We used de-identified and aggregated worldwide population movement data from billions of geotagged tweets and SafeGraph data from October – December 2020, derived using the ODT Flow Explorer [4], to explore the main destinations of international travellers departing from the UK, South Africa, and Brazil

https://www.worldpop.org/events/covid_variants

Agenda

- A brief introduction to GIBD lab and social media analytics
- ODT flow system for extracting, analyzing, and sharing multi-source multi-scale human mobility
- **Measuring global multi-scale place connectivity based on movement derived from geotagged tweets**
- Workflow demonstrations of using derived mobility and connectivity datasets shared through APIs (Dr. Tao Hu)
- Discussion

Place Connectivity Index (PCI) based on shared social media users



S_i Number of twitter users in place i

S_j Number of twitter users in place j

S_{ij} Number of shared twitter users between place i and j

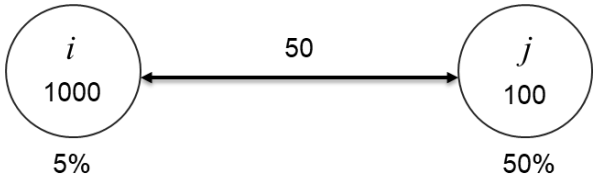
PCI between two places (e.g., county) for a specified time period (t) is defined as the normalized number of **shared users** based on the users from each area (controlling for population).

PCI can be computed in varying spatial scales, e.g., county, state, country, tract, covering a continuous space.

Following the general geometric average and normalization strategy (e.g., Liu et al., 2014; Bailey et al., 2018; Lin et al., 2019)

$$PCI_{ij} = \frac{S_{ij}}{\sqrt{S_i S_j}} \qquad i, j \in [1, n]$$

Capture asymmetrical impact of the shared users on different places



$$PCI_{i \to j} = \frac{S_{ij}}{S_j} \qquad i, j \in [1, n]$$

$$PCI_{j \to i} = \frac{S_{ij}}{S_i} \qquad i, j \in [1, n]$$

Li, Z., Huang, X., Ye, X., Jiang, Y., Martin, Y., Ning, H., ... & Li, X. Measuring Global Multi-Scale Place Connectivity using Geotagged Social Media Data, preprint

Compute PCI based on the Entity-ODT Cubes using 2019 worldwide geotagged tweets

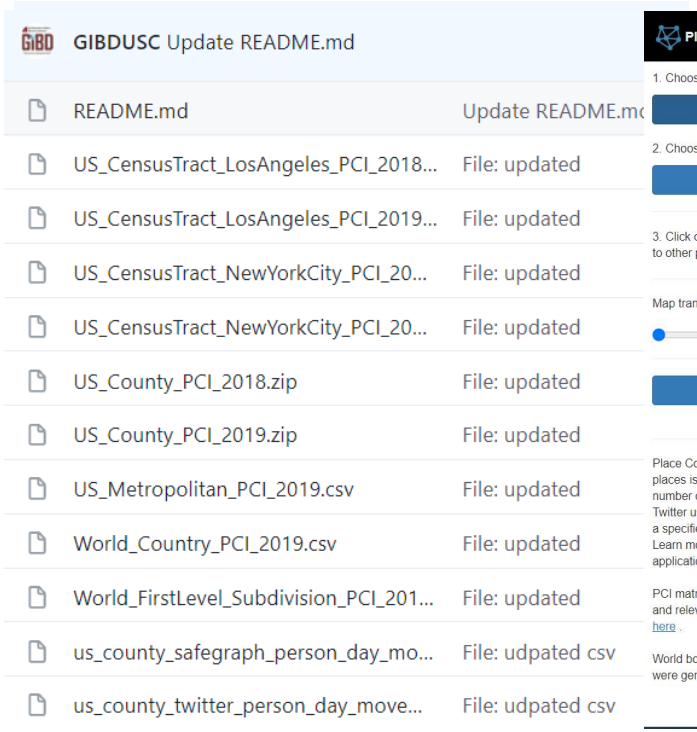
Over 1.4 billion worldwide geotagged tweets in 2019

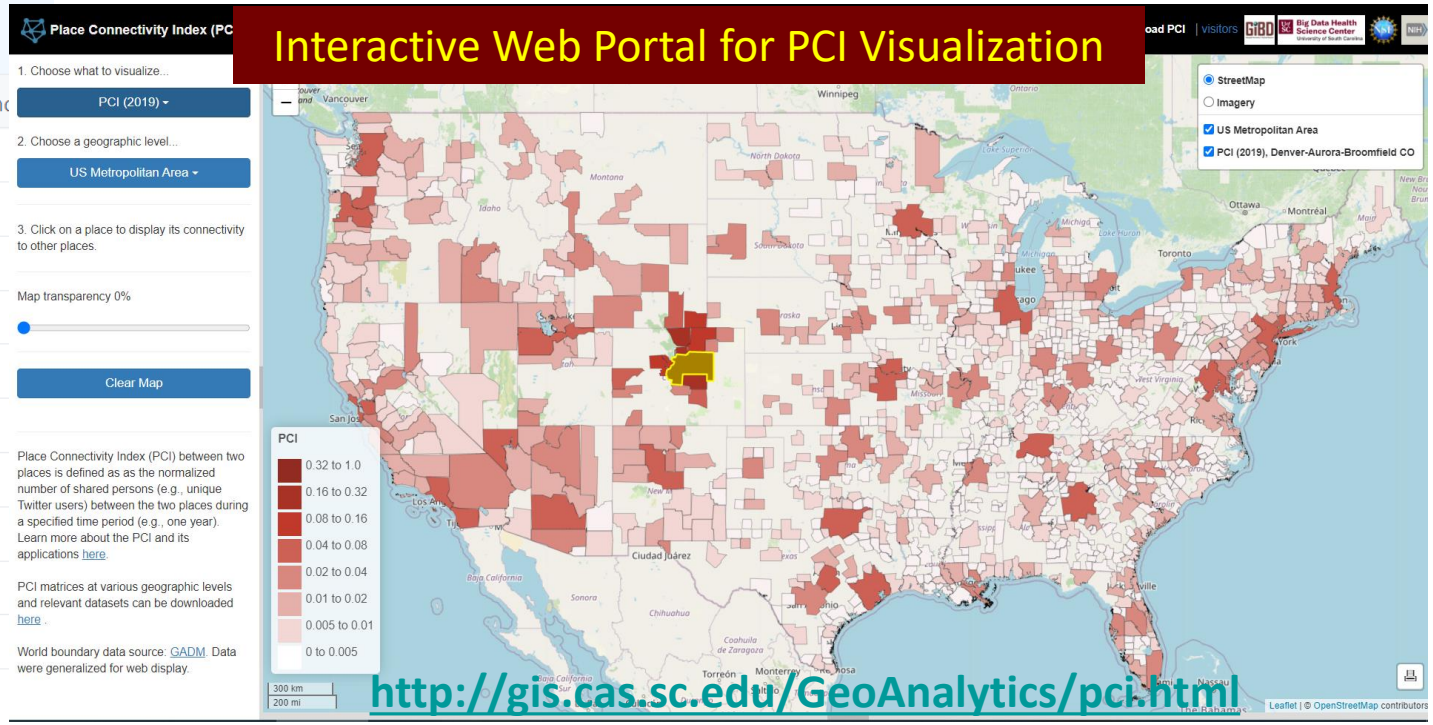
Entity-ODT Cubes

Geographic Scale	Number of PCI place pairs	
World Country/Territory	45,287	Inter-country/territory
World 1st-level Subdivision	2,718,537	Inter-state/province
US Metropolitan Area	634,612	Inter-metro
US County	3,405,113	Inter-city
US Tract (LA & NYC)	2,661,032	Intra-city

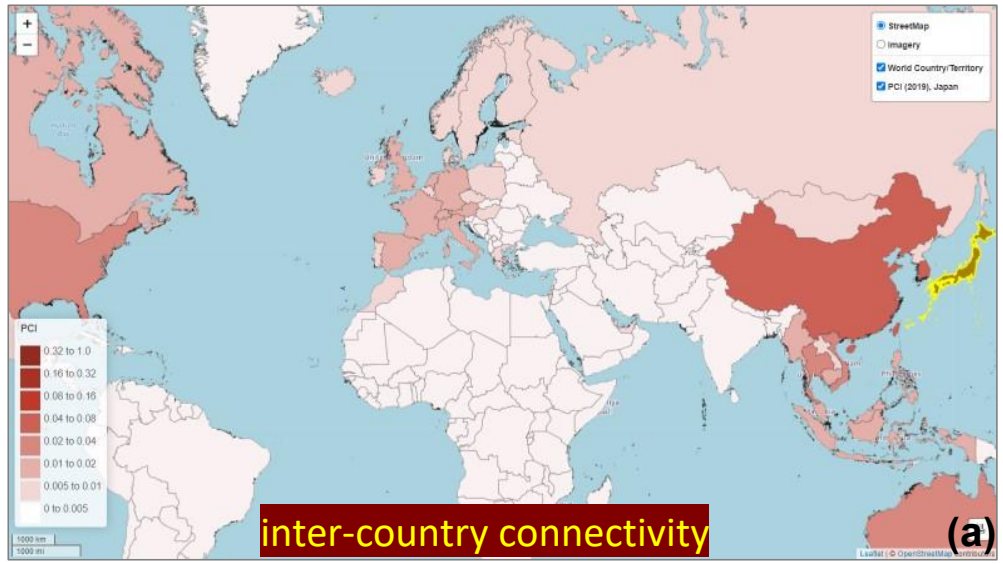
All PCI datasets are shared on GitHub and can also be accessed via REST APIs

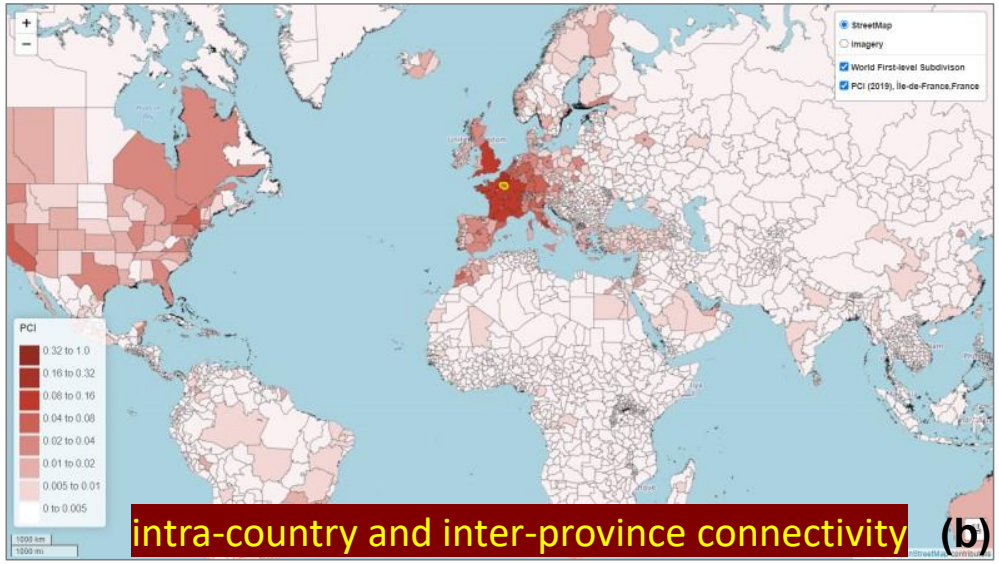
https://github.com/GIBDUSC/Place-Connectivity-Index

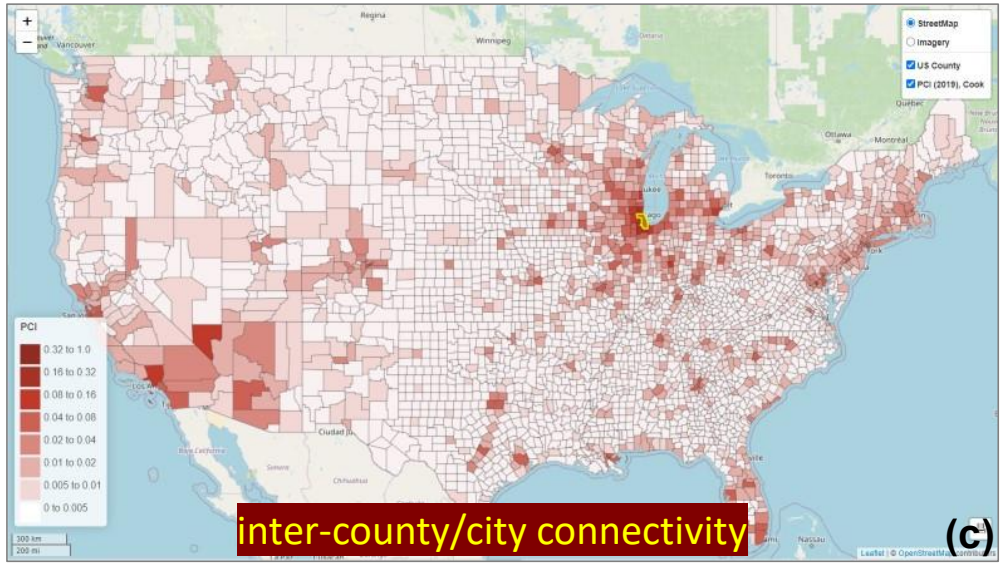


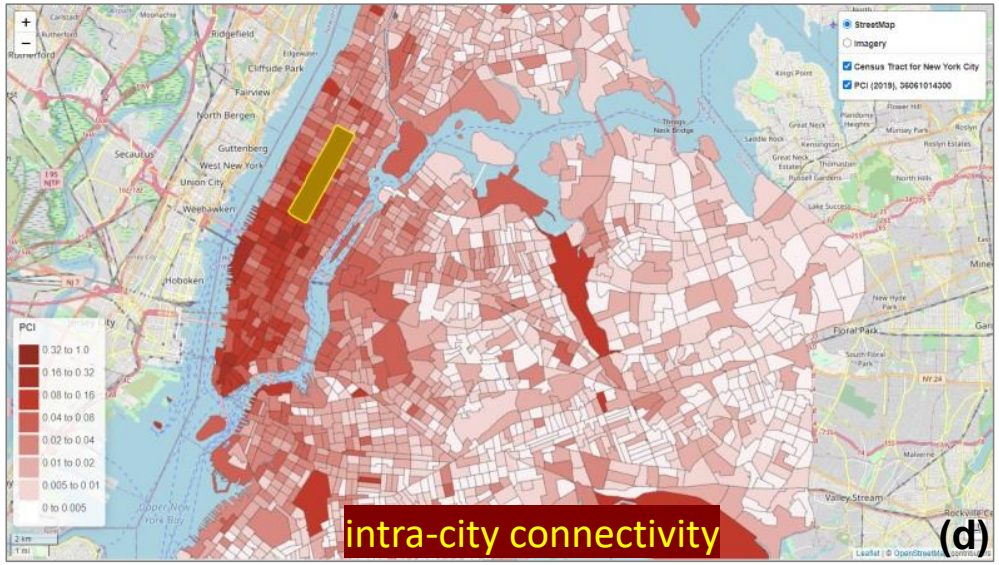


Demonstration of PCI computed at four geographic levels using 2019 twitter data









(a) World country level PCI for Japan showing the inter-country connectivity;

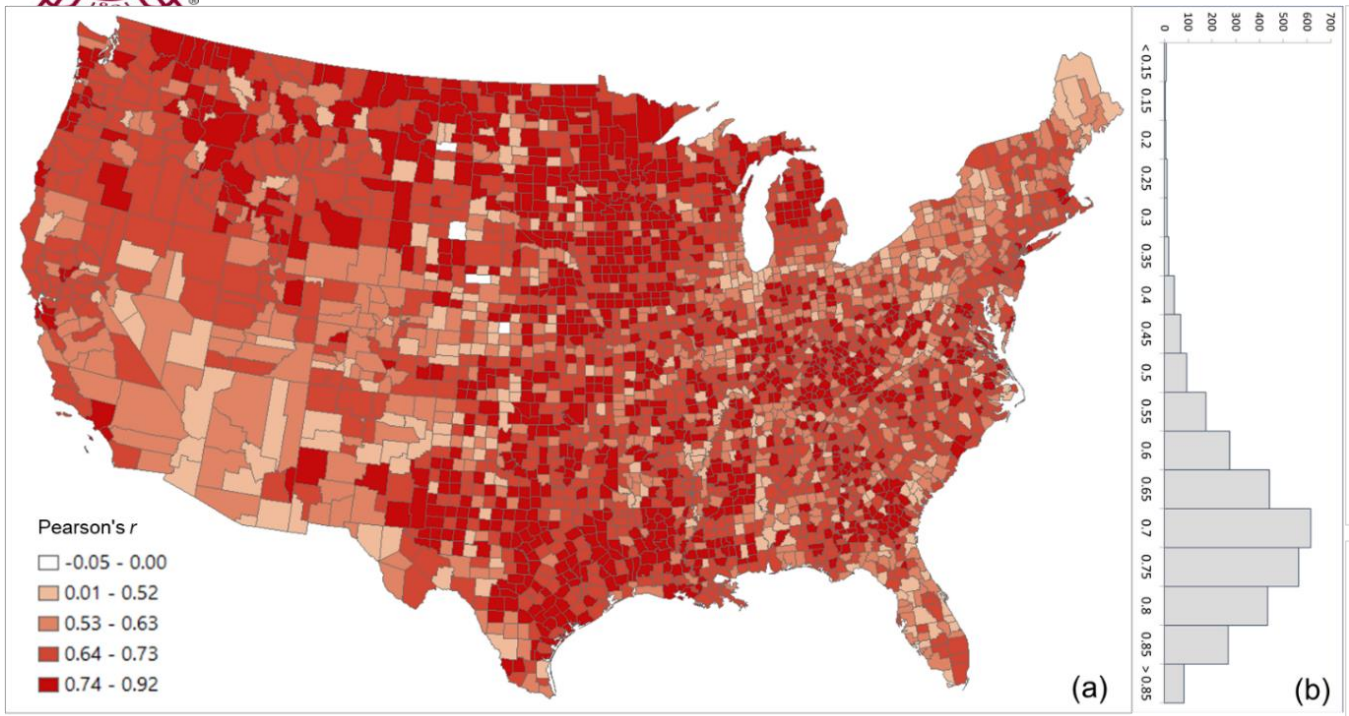
(b) World first-level subdivision PCI for Ile-de-France (surrounding Paris), France showing the inter-country and intra-country connectivity at the state or province level;

(c) US county level PCI for Cook County (Chicago) showing the inter-county/city connectivity; and

(d) US census tract level PCI for Central Park, New York City showing the intra-city connectivity.

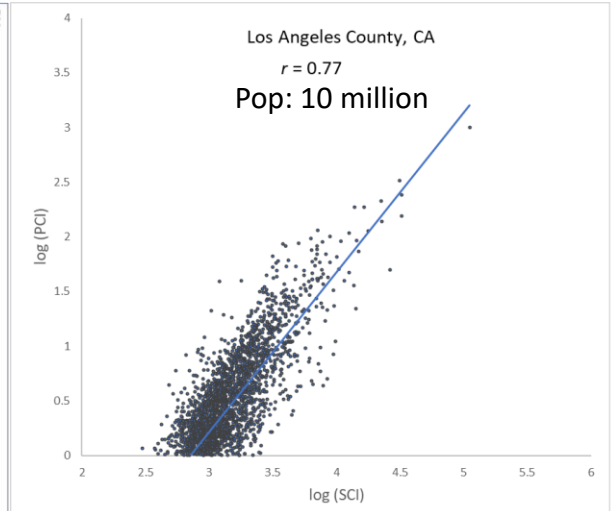
DEMO

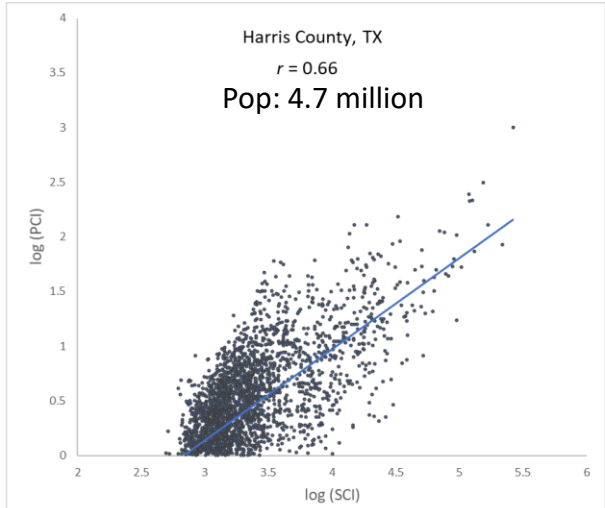
Comparing PCI with Facebook SCI (Social Connectedness Index, Bailey et al., 2018)

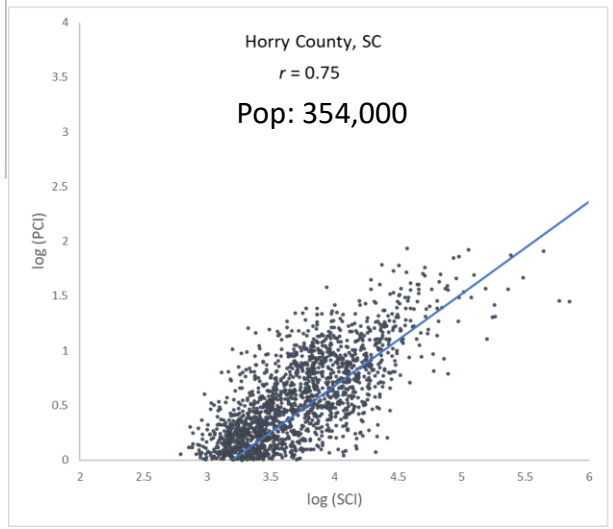


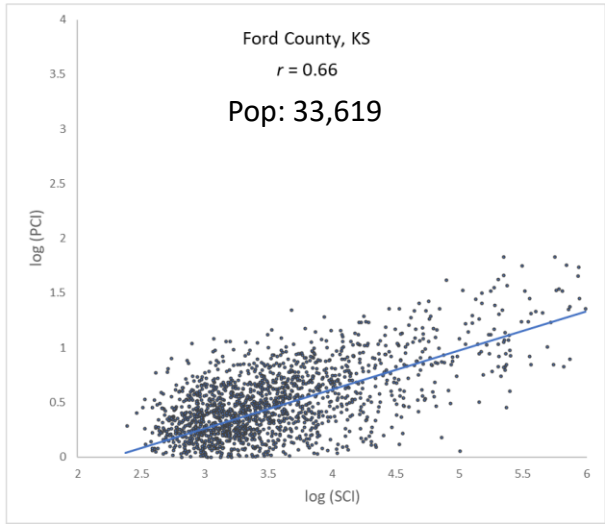
Distribution of the Pearson's *r* between log PCI and log SCI for all counties (a) Spatial distribution; (b) histogram

This comparison allows us to evaluate the hypothesis that places connected through (social media) friendship links are likely to have more physical interactions (e.g., population movement). This hypothesis has already been suggested in recent studies (Kuchler et al., 2021) but not corroborated using SCI data.







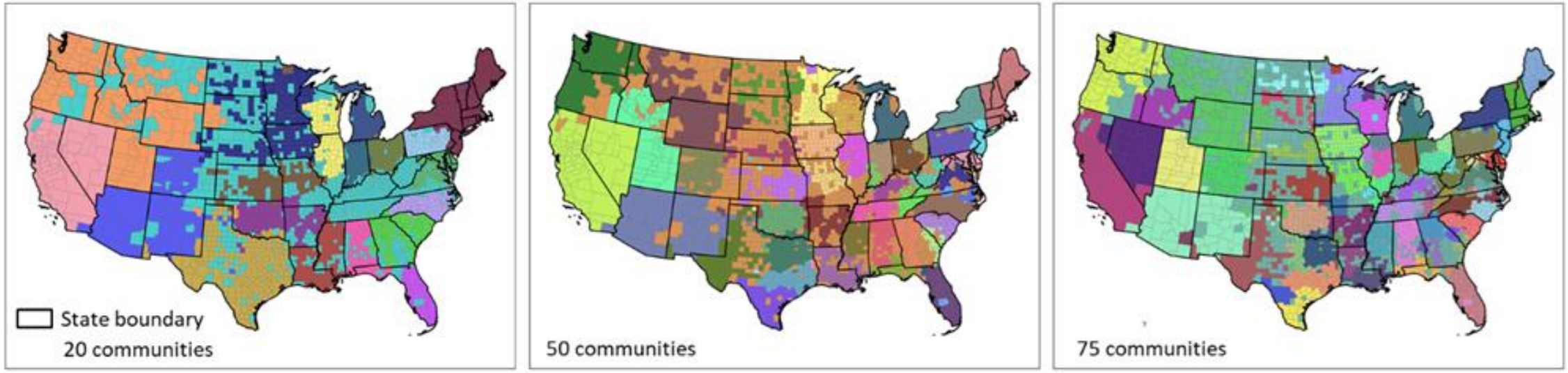


The overall *r* of 0.62 indicates a strong linear association between social and spatial connections (*n* = 1,702,531).

Examine the boundary effect of PCI: US county level

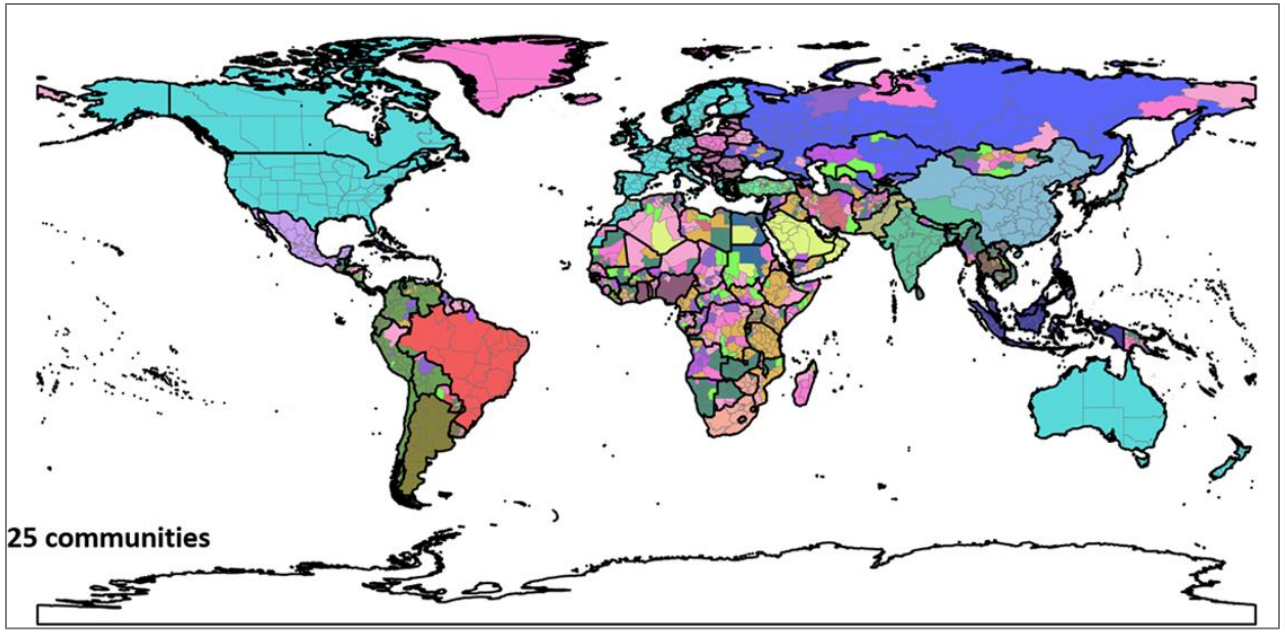
In a general sense, people tend to travel to their adjacent counties more frequently than non-adjacent counties (distance decay). However, do the residents near the state border prefer the in-state counties as their destinations rather than the adjacent county across the state border?

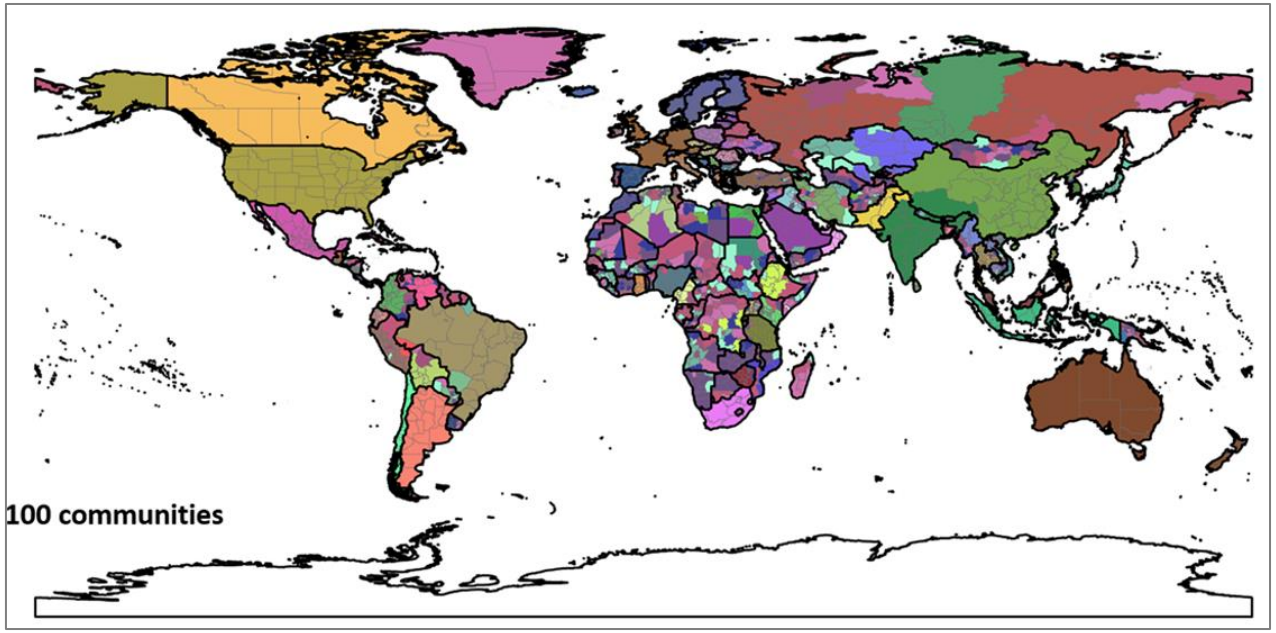
Results of the hierarchical agglomerative clustering of PCI with three different targeted numbers of communities for the contiguous US. Each color depicts a community.



The different regions identified in the US using the agglomerative clustering not only demonstrate the boundary effect of PCI, but also suggest that PCI can be potentially used **as a tool in regionalization analysis** to reveal how places are connected and regions are formed at different geographic scales.

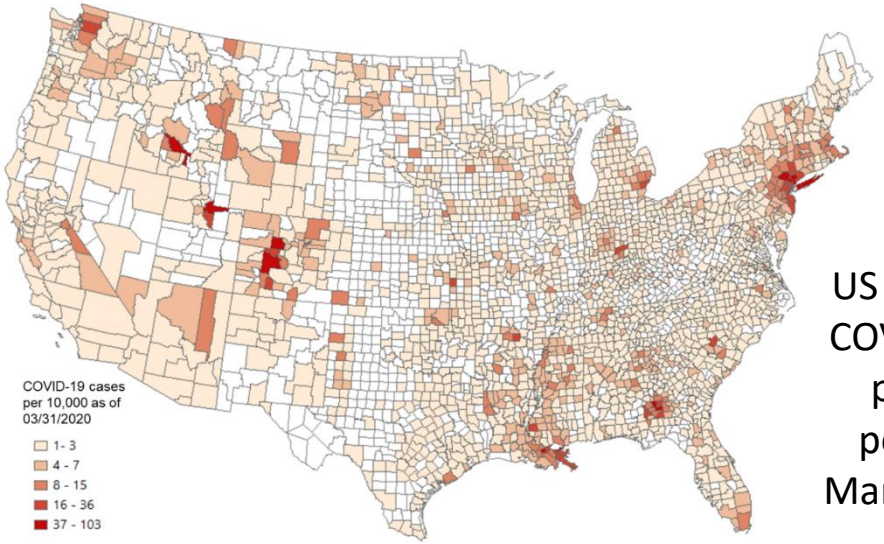
Examine the boundary effect of PCI: World 1st level Subdivision



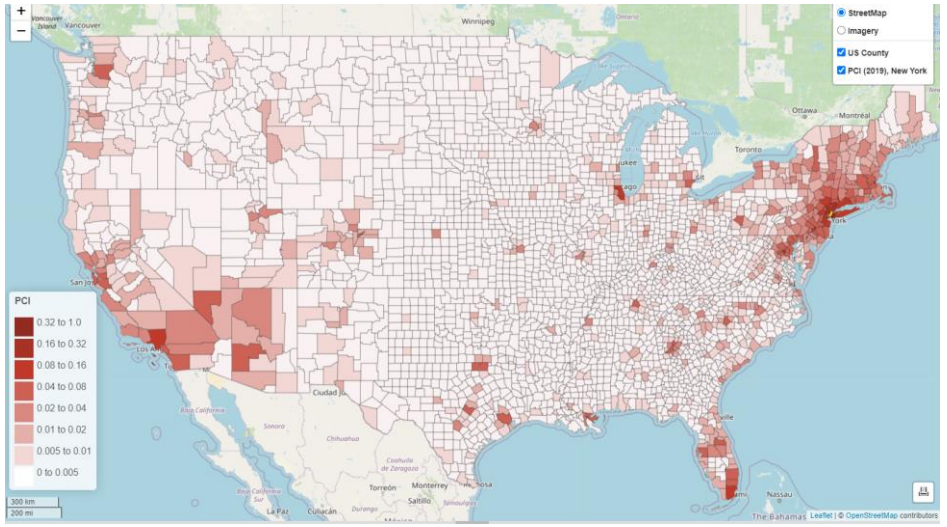


The results reveal that the groupings with the 25 communities are consistent with what many people perceive as connected regions (e.g., US with Canada and Europe). However, once into the 100 level, the divisions between east and west start to emerge. Another interesting finding is how unconnected the regions in Africa are, though the country boundary effect is still observable. However, it should be cautious that whether such disconnection resulted from the sparsity of Twitter data in Africa countries needs further investigation.

PCI as a factor in predicting the spatial spread of COVID-19 during the early stage



US county level COVID-19 cases per 10,000 people as of March 31, 2020.



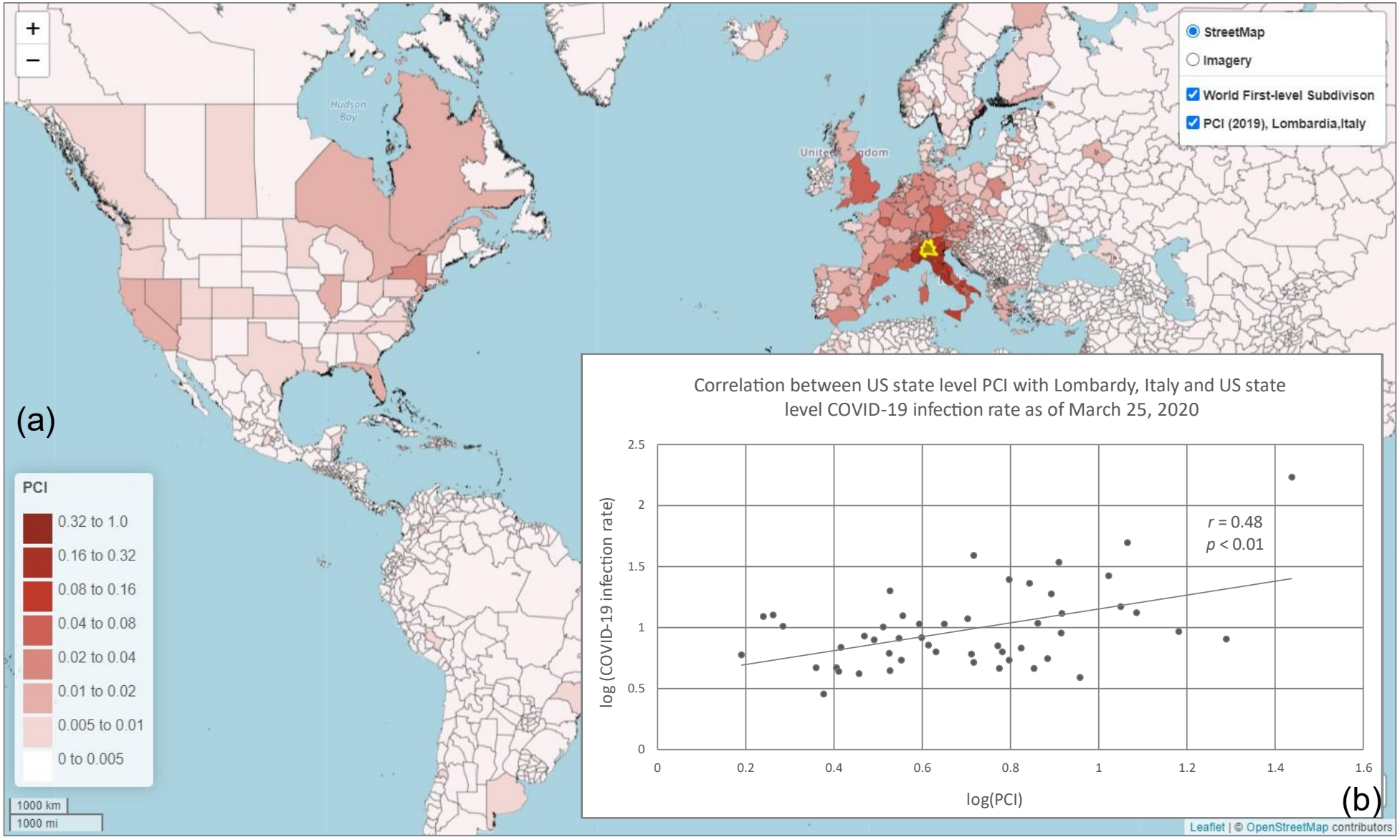
2019 PCI from NYC to other counties

Regression result using *COVID-19 infection rate* as the dependent variable, and *PCI, SCI*, or *SafeGraph* as the predictor variable controlling for *distance*.

	2019 PCI		2018 PCI		Facebook (2020) SCI		SafeGraph (2020) Population movement	
	Coefficients	SE	Coefficients	SE	Coefficients	SE	Coefficients	SE
Intercept	0.73883***	0.22272	0.67638***	0.23186	2.19947***	0.23329	2.43267***	0.24346
PCI/SCI/SafeGraph	0.23706***	0.00960	0.22307***	0.00931	0.00013***	0.00001	0.00040***	0.00003
Distance	0.00087***	0.00019	0.00091***	0.00019	0.00007	0.00020	0.00004	0.00022
Adjusted R^2	0.25		0.25		0.08		0.13	
Observations	1847		1755		1847		1497	

*p < 0.1 **p < 0.05 ***p < 0.01

PCI as a factor in predicting the spatial spread of COVID-19 during the early stage



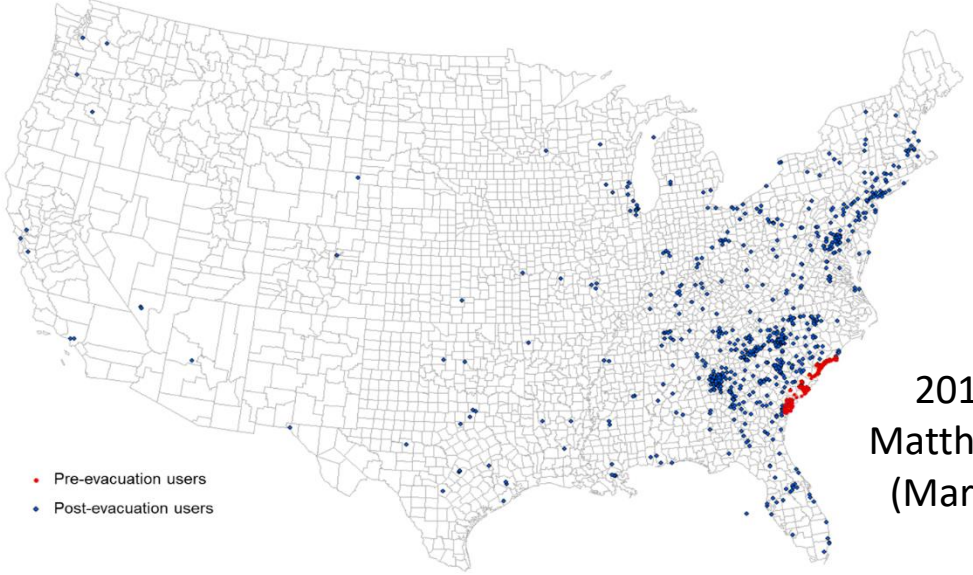
Global scale analysis of PCI and COVID-19 infection rate.

(a) Map showing the 2019 world first-level subdivision PCI between Lombardy, Italy and US states (and other parts of the world);

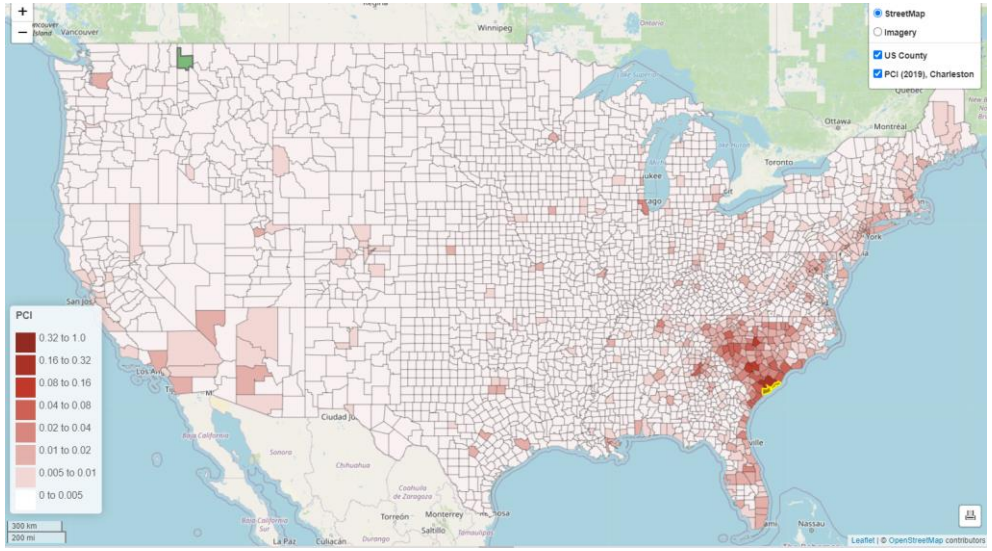
(b) Correlation between the log US state level PCI with Lombardy, Italy and log US state level COVID-19 infection rate as of March 25, 2020.

Lombardy, Italy was an epicenter of the COVID-19, with the first cluster of cases detected on February 21, 2020. The travel restrictions between the US and Europe were not in place until March 12, 2020.

PCI as a factor in predicting hurricane evacuation destination choices



2016 Hurricane Matthew evacuation (Martin, Li, Cutter, 2017)



2019 PCI from Charleston to other counties

Regression results for the number of evacuated users in the destination counties (dependent variable) and PCI (or SCI) of the county pairs between evacuation county and each of the destination counties.

	Charleston County PCI		Charleston County SCI		Chatham County PCI		Chatham County SCI	
	Coefficients	SE	Coefficients	SE	Coefficients	SE	Coefficients	SE
Intercept	0.28676	0.23274	1.63200***	0.32877	-1.66645**	0.63781	2.57697***	0.69610
PCI/SCI	0.09611***	0.00606	0.00003***	0.00000	0.19071***	0.01927	0.00001	0.00001
Distance	0.00019	0.00030	-0.00047	0.00047	0.00115	0.00083	-0.00155	0.00112
Adjusted R^2	0.71		0.29		0.47		0.04	
Observations	118		118		120		120	

*p < 0.1 **p < 0.05 ***p < 0.01

Agenda

- A brief introduction to GIBD lab and social media analytics
- ODT flow system for extracting, analyzing, and sharing multi-source multi-scale human mobility
- Measuring global multi-scale place connectivity based on movement derived from geotagged tweets
- **Workflow demonstrations of using derived mobility and connectivity datasets shared through APIs (Dr. Tao Hu)**
- Discussion

Reproducibility, Replicability, and Generalization

REPRODUCIBILITY, REPLICABILITY, AND GENERALIZATION IN THE SOCIAL, BEHAVIORAL, AND ECONOMIC SCIENCES

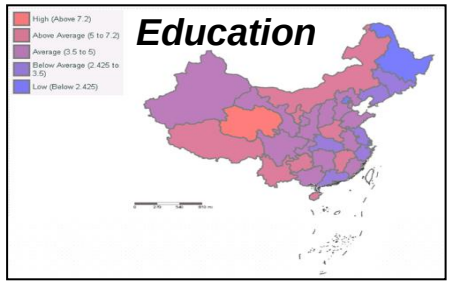
REPRODUCIBILITY: the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator.

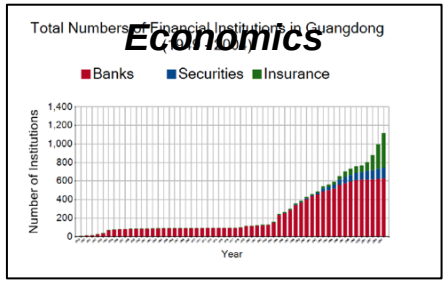
REPLICABILITY: the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected.

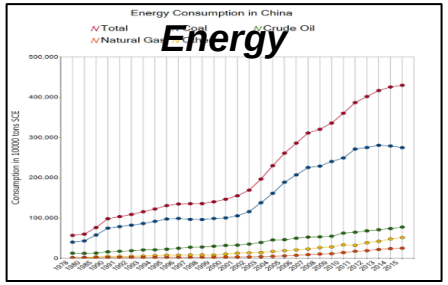
GENERALIZABILITY: whether the results of a study apply in other contexts or populations that differ from the original one

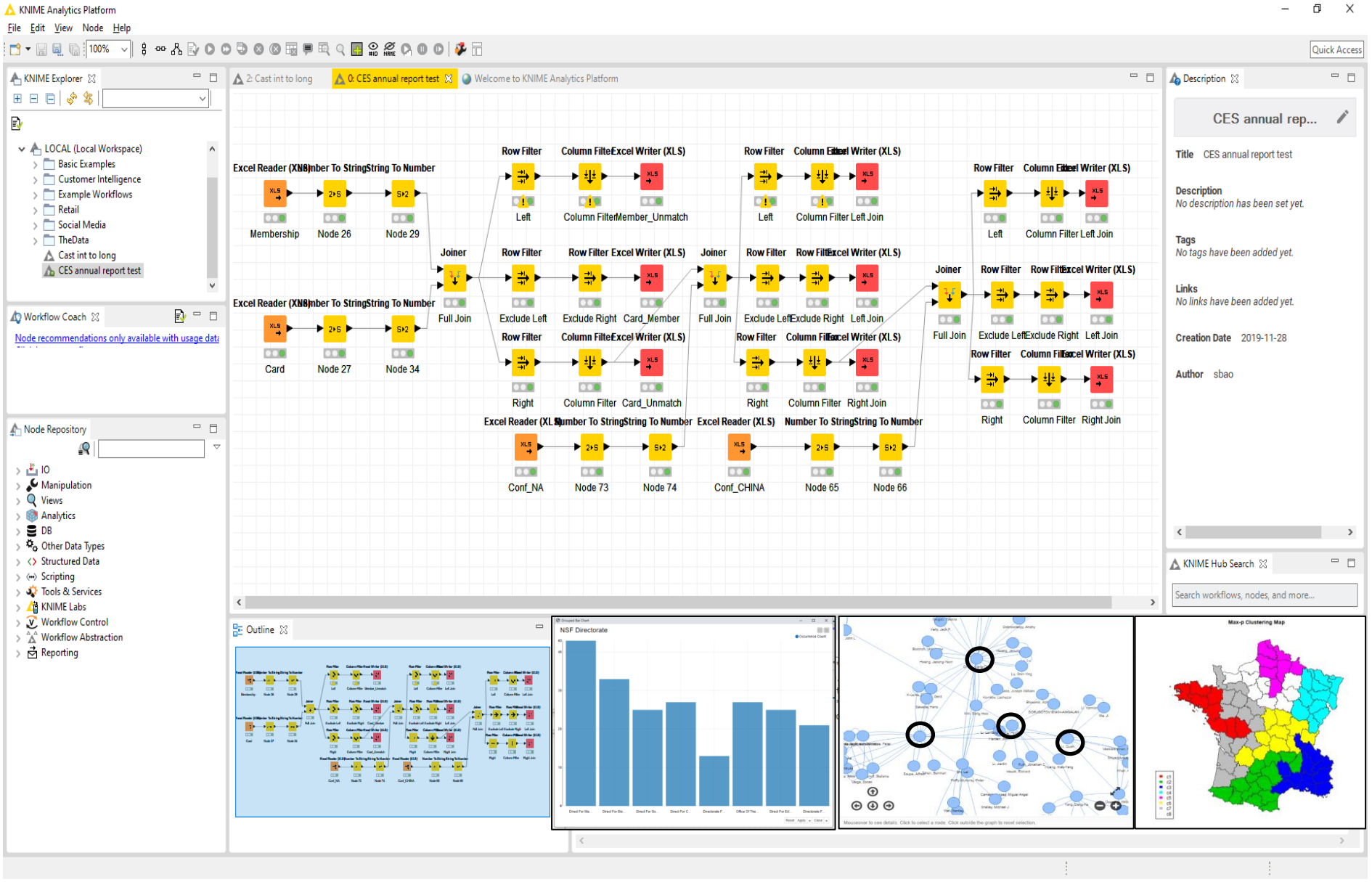
Report of the Subcommittee on Replicability in Science of the SBE Advisory Committee to the National Science Foundation
13 May 2015 Presentation
at SBE AC Spring Meeting by K. Bollen.

Reproducibility, Replicability, and Generalization with Workflow







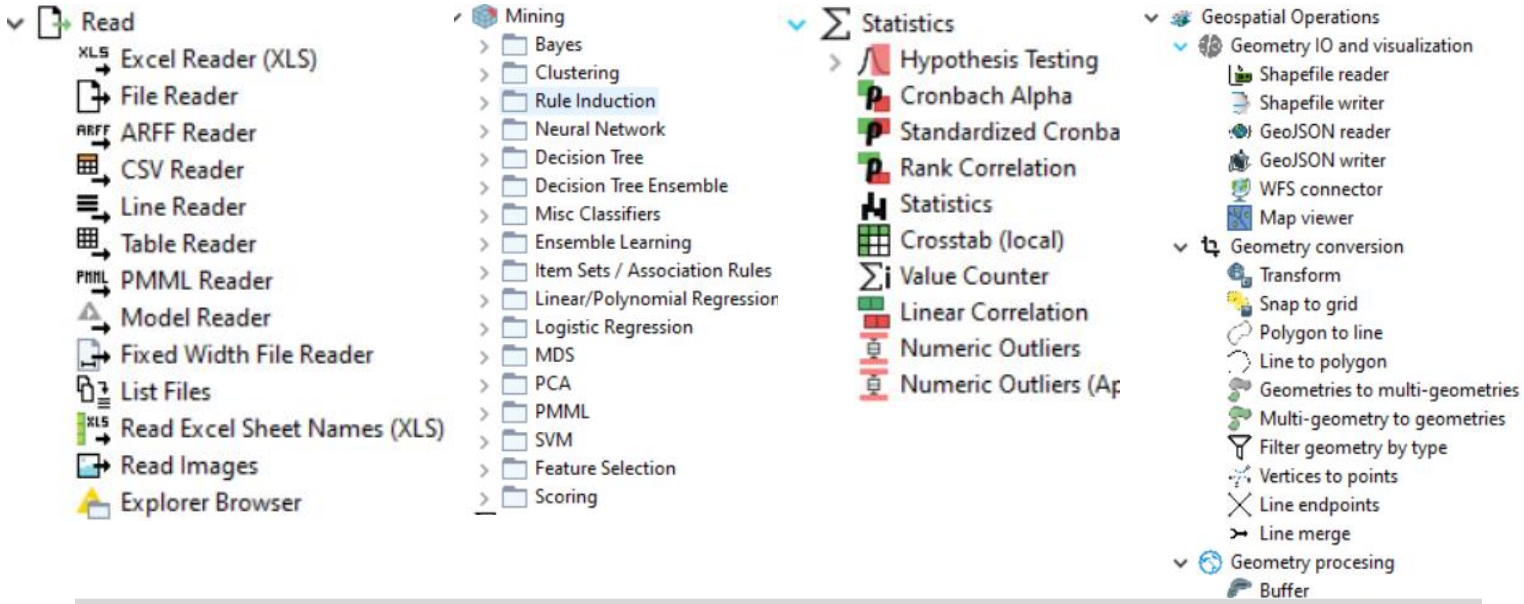


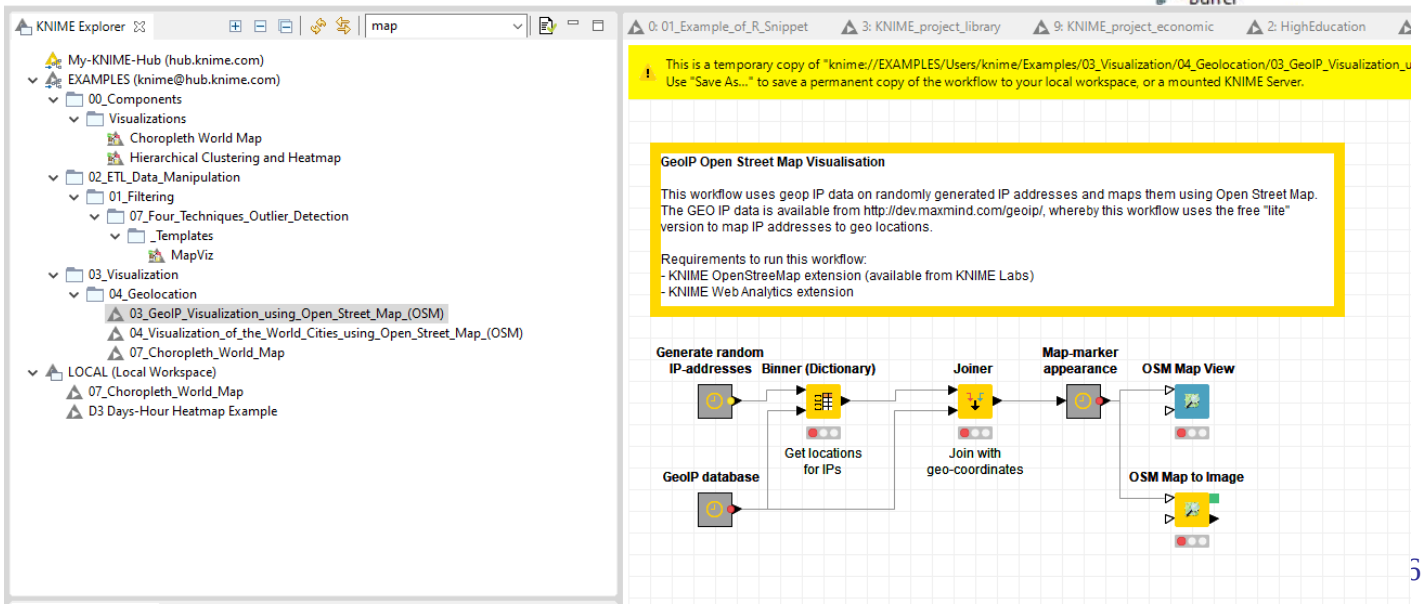
Data Science Workflow Tool KINME

Open for Innovation KNIME

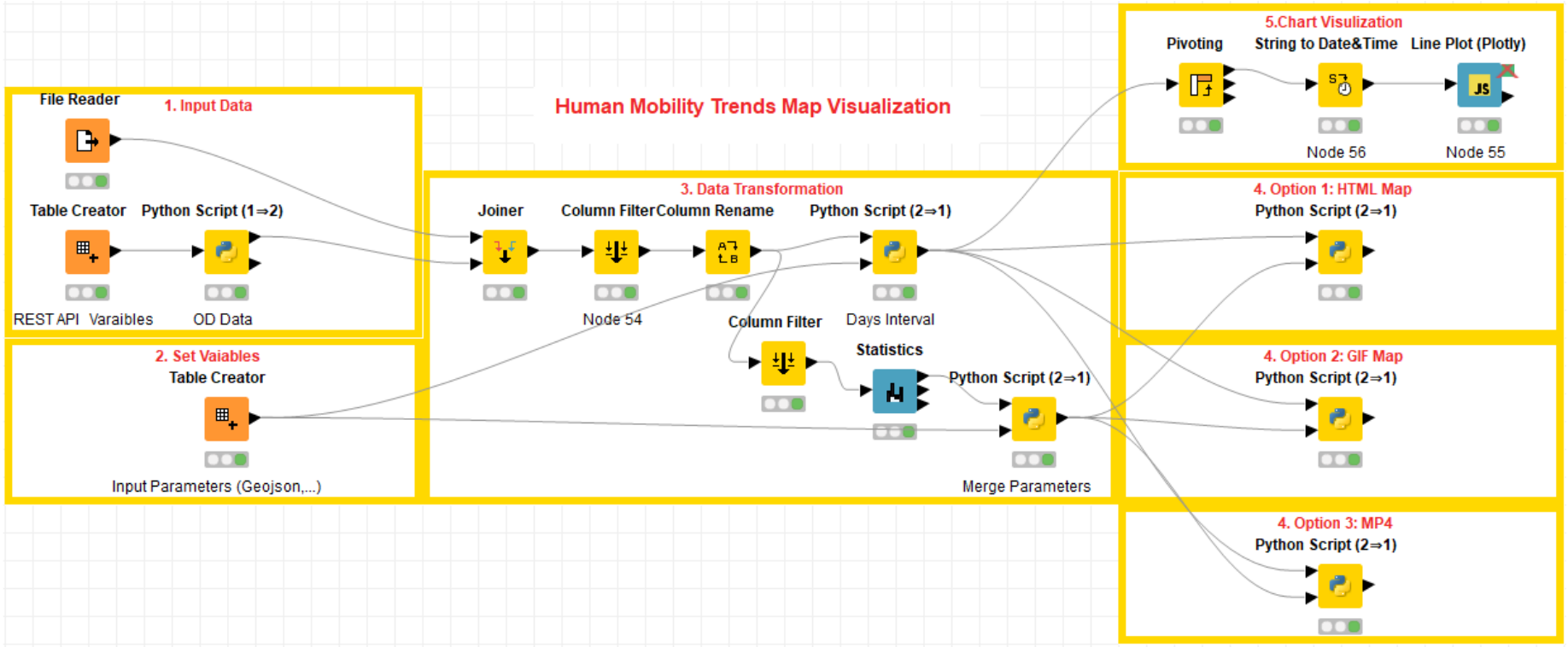
https://www.knime.com/

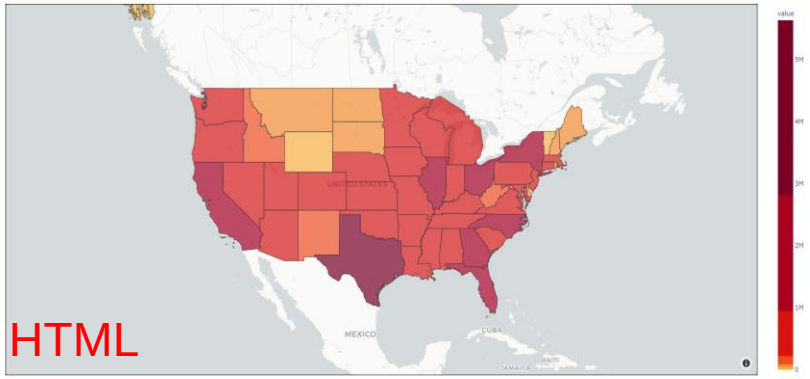
- Workflow is displayed as connected nodes which makes it easy to troubleshoot and visualize
- Easy to use without much knowledge of coding
- Great extensions for data preprocessing, analysis, and visualization
- Connection to other languages, such as JS, R, Python, etc.
- Open-source
- Cross platform interoperability
- Has a decent size community that supports Q&A.

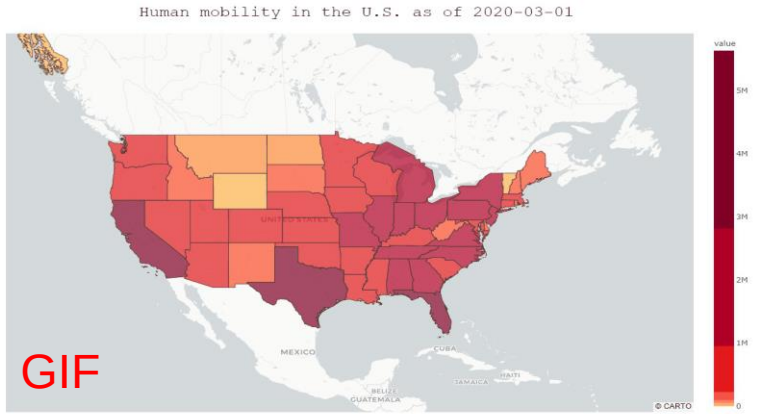


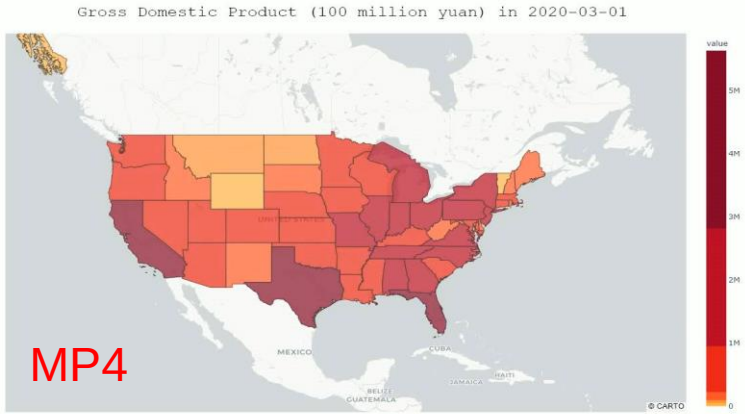


Case Study 1: Human Mobility Trends Visualization with Dynamic Map

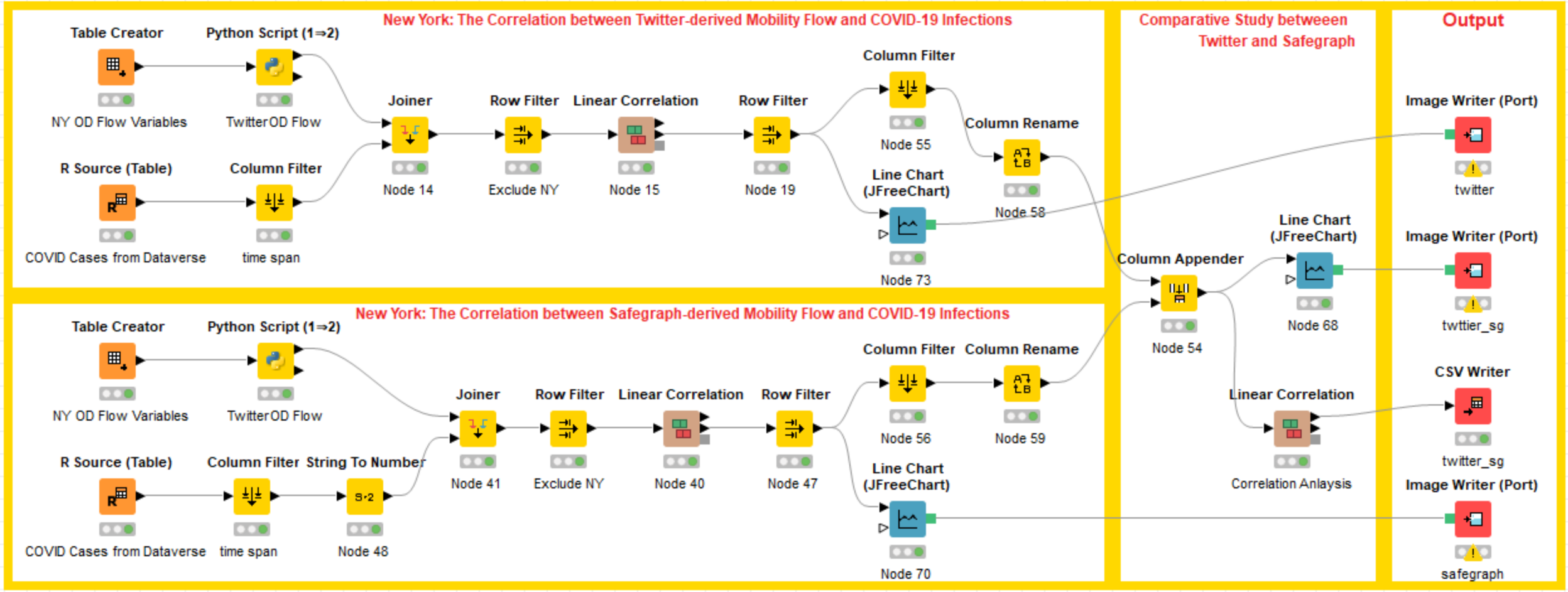




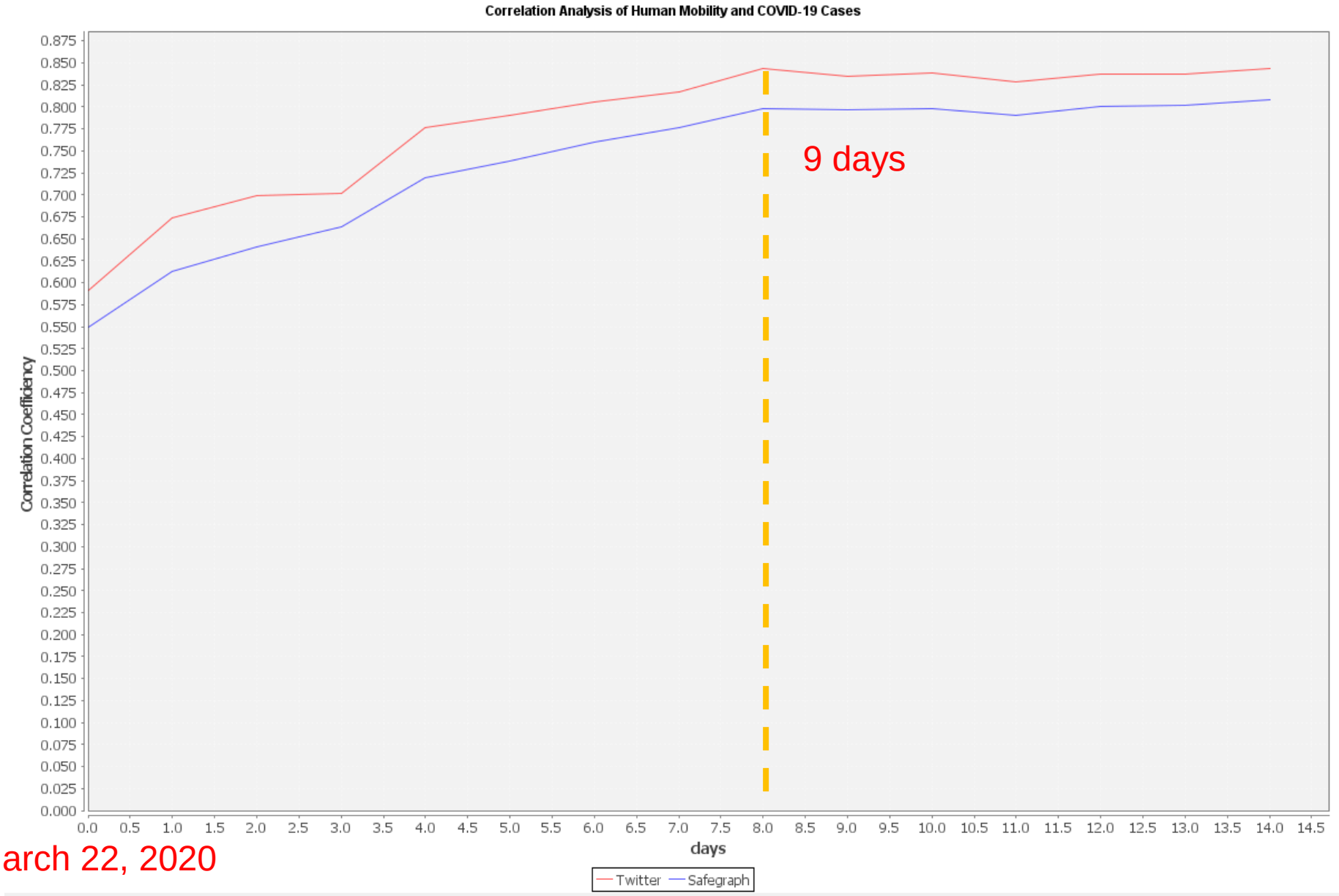




Case Study 2: Correlation Analysis between Mobility and COVI-19 Cases



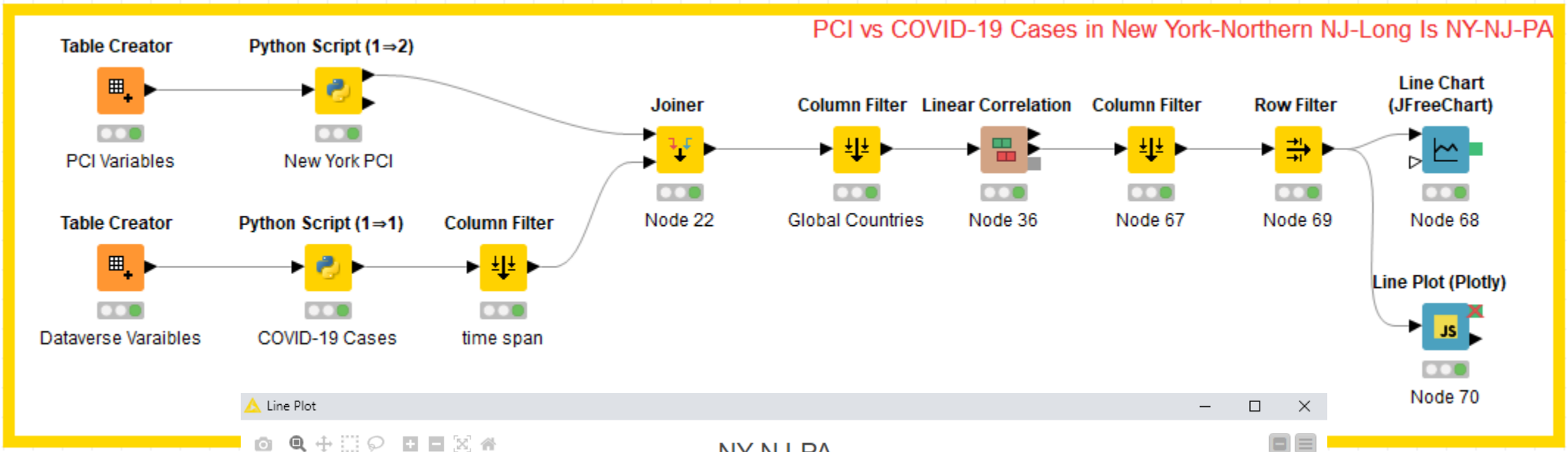
Case Study 2: Correlation Analysis between Mobility and COVI-19 Case

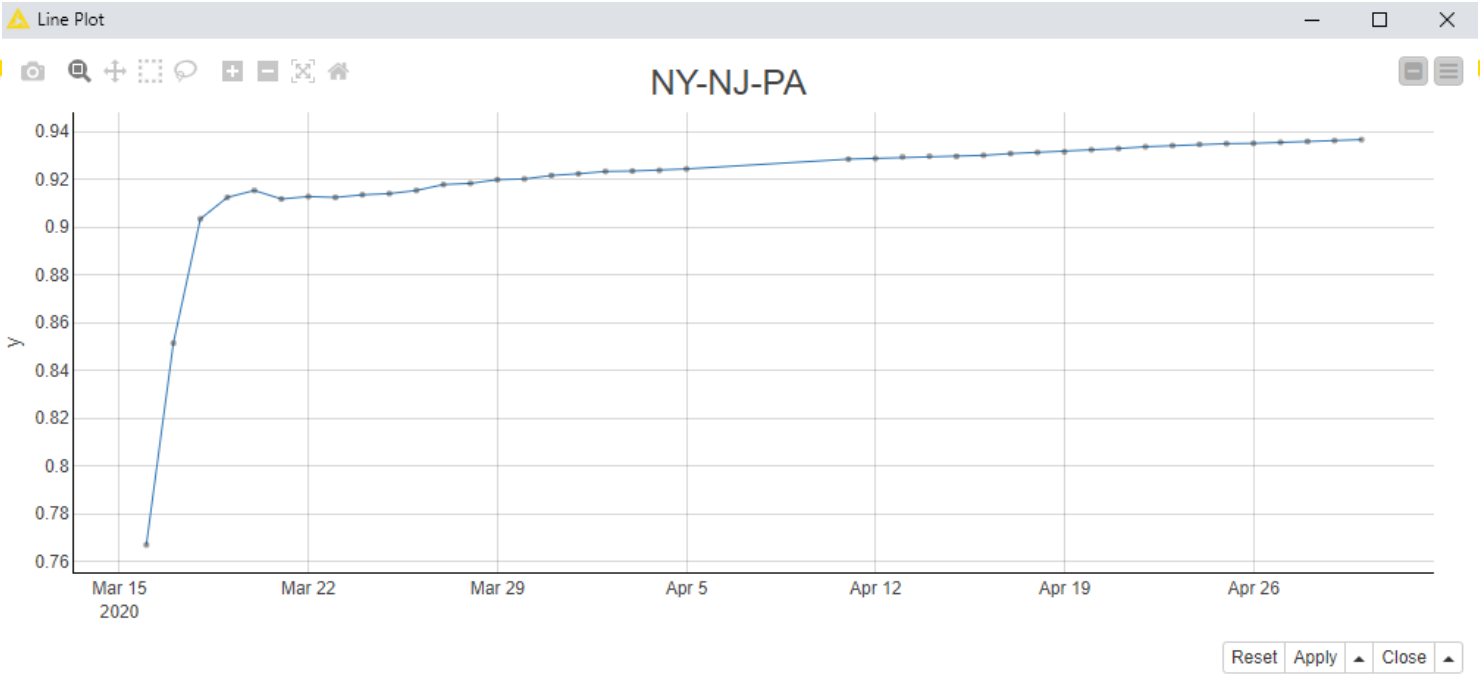


R=0.995

March 22, 2020

Case Study 3: PCI-based COVID-19 Infection Cases Analysis





Data, Code, Workflow, Case Study and Publication Sharing

Dataverse Project About Community Best Practices Software Contact

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts.* A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal Dataverse collection?

Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data.* Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. Want to find out more about journal Dataverse collections?

Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. Want to install a Dataverse repository?

Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization* and *exploration tools,* or other research and data archival systems with the Dataverse Project. Want to contribute?

HARVARD Dataverse — Add Data | Search | About | User Guide | Support | Data Lab China 26

Harvard Dataverse > China Data Lab Dataverse > Resources for COVID-19 > Workflows >

OD Flow Explorer Case Studies

Version 4.2

Li, Zhenlong; Hu, Tao; Ning, Huan; Huang, Xiao; Ye, Xinyue, 2021, "OD Flow Explorer Case Studies", https://doi.org/10.7910/DVN/GL3HAB, Harvard Dataverse, V4, UNF:6:yU0ynDI/tA3SR8wQdEUIWg== [fileUNF]

Cite Dataset — Learn about Data Citation Standards.

Access Dataset | Edit Dataset | Link Dataset | Contact Owner | Share

Dataset Metrics

48 Downloads

Description The Geoinformation and Big Data Research Laboratory (GIBD) at the University of South Carolina developed the OD Flow Explorer (http://gis.cas.sc.edu/GeoAnalytics/od.html), which provides free human mobility data derived by global wide Geotagged Tweets and the US Safegraph data. To make the data analysis reproducible, replicable, and expandable, we created three case studies based on workflow to visualize data in dynamic maps, and analyze human mobilities' impact on COVID-19 disease transmission in the US and Europe.

Subject Earth and Environmental Sciences; Computer and Information Science; Social Sciences

Related Publication Zhenlong Li, Xiao Huang, Tao Hu, Huan Ning, Xinyue Ye, and Xiaoming Li. (2021) ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility Flows. Preprint. https://www.researchgate.net/publication/350342301_ODT_FLOW_A_Scalable_Platform_for_Extracting_Analyzing_and_Sharing_Multi-source_Multi-scale_Human_Mobility_Flows

Files | Metadata | Terms | Versions

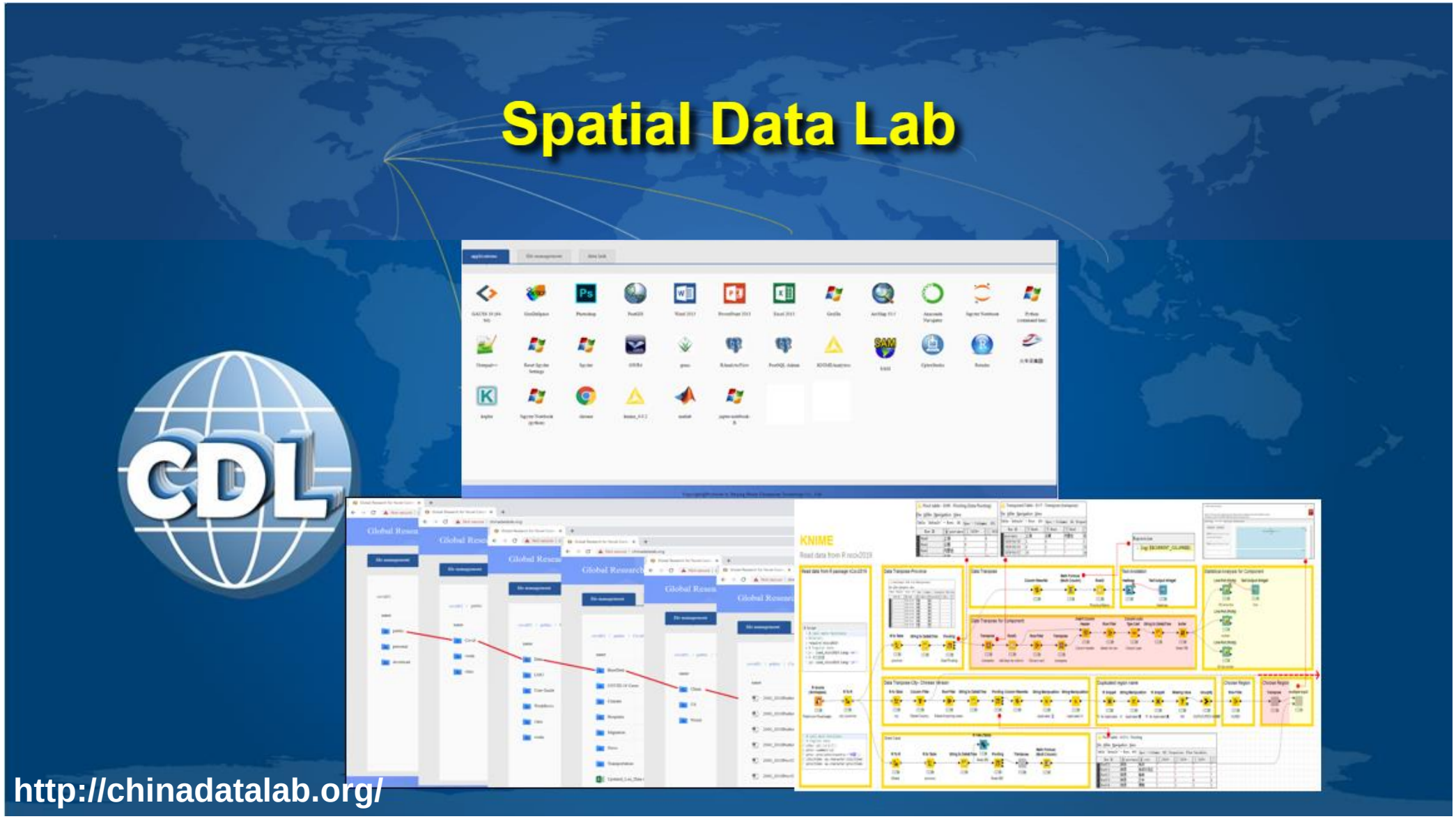
Search this dataset... Find — Upload Files

Filter by: File Type: All — Access: All — Sort

1 to 9 of 9 Files — Edit Files — Download

- API with Jupyter Notebook Case study 1.tar — TAR Archive - 20.5 MB - Mar 23, 2021 - 5 Downloads — MD5: 9ccb6944c3a4ba54a3653a2f0010a917
- API with Jupyter Notebook Case study 2.tar — TAR Archive - 23.5 KB - Mar 23, 2021 - 4 Downloads — MD5: d05265b50979acc4796421b769e88c25
- API with Jupyter Notebook Case study 3.tar — TAR Archive - 26.9 MB - Mar 23, 2021 - 4 Downloads — MD5: 32c3cc2a7f0343cf893cf22b9e87219a

Spatial Data Lab: an online data sharing and management platform



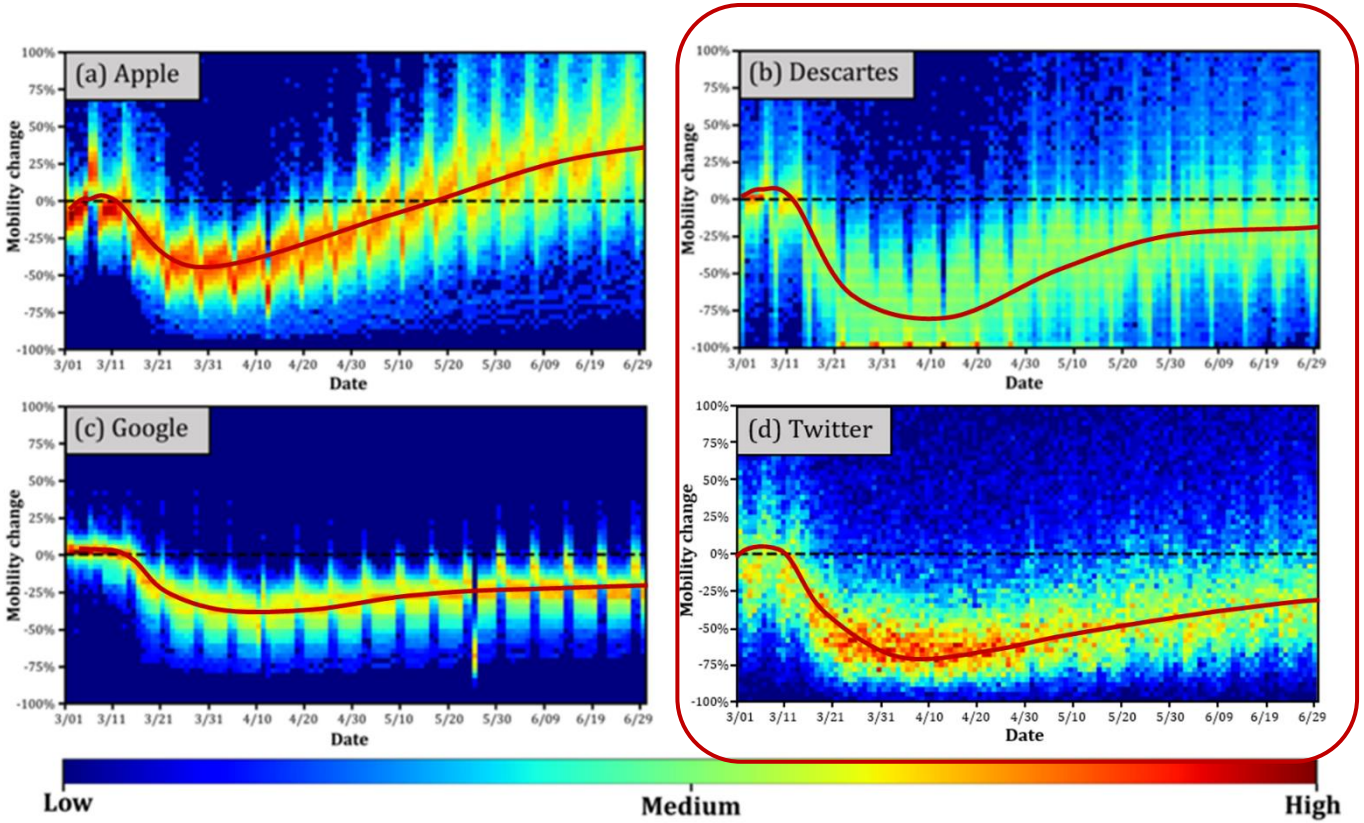
Agenda

- A brief introduction to GIBD lab and social media analytics
- ODT flow system for extracting, analyzing, and sharing multi-source multi-scale human mobility
- Measuring global multi-scale place connectivity based on movement derived from geotagged tweets
- Workflow demonstrations of using derived mobility and connectivity datasets shared through APIs (Dr. Tao Hu)
- **Discussion**

Twitter (social media) data limitations

- Twitter is not proportionally used by different population groups and thus shows demographic and socioeconomic biases.
- Geotagged tweets are unevenly distributed across space, due to a variety of factors such as population density, Internet access, and governmental policies etc.
- The dynamics of people's Twitting activities, as well as the changing of Twitter's internal API, affect the daily number of tweets being collected.
- Studies using twitter-derived data should be aware of the limitations when interpreting the results.

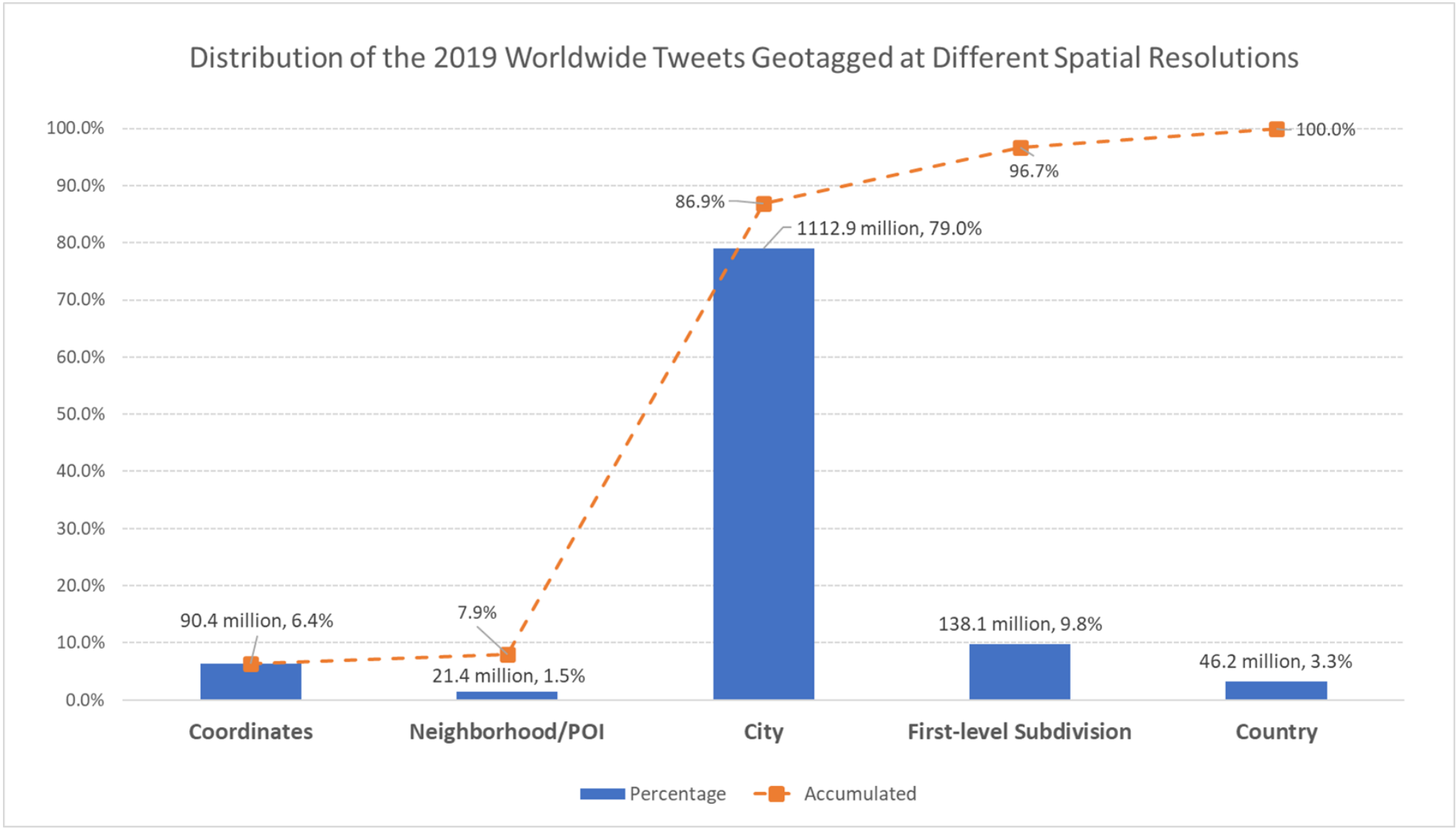
Compare Twitter data with other three mobility data sources: Apple, Google, and Descartes Labs (mobile phone data).



Huang X., Li Z., Jiang Y., Ye X., Deng C., Zhang J., Li X., (2021), The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S., *International Journal of Digital Earth*

Statistics of geotagging resolution of worldwide geotagged tweets in 2019

For 2019 data, a total of 1,409,404,996 geotagged tweets posted by 17,013,612 unique Twitter were selected after removing non-human tweets.



For more details about the work, please check our preprints

ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility

Zhenlong Li[1*], Xiao Huang[2], Tao Hu[3], Huan Ning[1], Xinyue Ye[4], Xiaoming Li[5]

[1]*Geoinformation and Big Data Research Lab, Department of Geography, University of South Carolina, Columbia, SC, USA*

[2]*Department of Geosciences, University of Arkansas, Fayetteville, AR, USA*

[3]*Center for Geographic Analysis, Harvard University, Cambridge, MA, USA*

[4]*Department of Landscape Architecture & Urban Planning, Texas A&M University, TX, USA*

[5]*Department of Health Promotion, Education, and Behavior, University of South Carolina, Columbia, SC, USA*

https://www.researchgate.net/publication/350342301_ODT_FLOW_A_Scalable_Platform_for_Extracting_Analyzing_and_Sharing_Multi-source_Multi-scale_Human_Mobility

Measuring Global Multi-Scale Place Connectivity using Geotagged Social Media Data

Zhenlong Li[1*], Xiao Huang[2], Xinyue Ye[3], Yuqin Jiang[1], Yago Martin[4], Huan Ning[1], Michael E. Hodgson[1], and Xiaoming Li[5]

[1] *Geoinformation and Big Data Research Laboratory, Department of Geography, University of South Carolina, SC, USA*

[2] *Department of Geosciences, University of Arkansas, AR, USA*

[3] *Department of Landscape Architecture & Urban Planning, Texas A&M University, TX, USA*

[4] *School of Public Administration, University of Central Florida, FL, USA*

[5] *Department of Health Promotion, Education, and Behavior, University of South Carolina, SC, USA*

https://www.researchgate.net/publication/349124701_Measuring_Global_Multi-Scale_Place_Connectivity_using_Geotagged_Social_Media_Data

Moving forward…

- Integrate other mobility data sources in the ODT system (i.e., NYC taxi data) and develop more APIs for server-side mobility analysis (leverage the HPC).
- Develop new REST APIs and tools to share twitter data with appropriate aggregations (we have ~10 billion geo-tweets collected since 2015).
- We encourage researchers to try out, evaluate, and use the ODT flow datasets and PCI datasets in their research.
- Call for more efforts on data/tool/model sharing to facilitate reproducibility and replicability.

Funding

- 2020-2022, RAPID: Monitoring the Spatial Spread of COVID-19 through the Lens of Human Movement using Big Social Media Data, National Science Foundation (NSF), 2028791, $108,717
- 2020-2023, Big Data Driven Clinical Informatics & Surveillance – A Multimodal Database Focused Clinical, Community, & Multi-Omics Surveillance Plan for COVID-19, National Institutes of Health (NIH), 3R01AI127203-04S1, $1,252,550
- 2020-2021, RAPID: Building a Spatiotemporal Platform For Rapid Response To COVID-19, , National Science Foundation (NSF), 2027540, $100,000
- 2020-2022, A Preliminary Study of using Social Media to Monitor the Spatial Propagation of COVID-19 and Quantify the Effectiveness of the Control Measures, USC COVID-19 Internal Funding Initiative, $13,017
- 2017-2018, Enhancing Situational Awareness by Mining Big Social Media Data in Near-real Time for Disaster Management: A CyberGIS Approach, USC OVPR, ASPIRE, $15,019

NSF NIH NIAID Uof SC

References

Derudder, B., & Witlox, F. (2008). Mapping world city networks through airline flows: Context, relevance, and problems. Journal of Transport Geography, 16(5), 305–312.

- Liu, Y., Wang, F., Kang, C., Gao, Y., & Lu, Y. (2014). Analyzing Relatedness by Toponym Co-O ccurrences on Web Pages. Transactions in GIS, 18(1), 89-107.
- Salt, J. (1987). Contemporary trends in international migration study. International Migration (Geneva, Switzerland), 25(3), 241–251.
- Boyle, P., Halfacree, K., & Robinson, V. (2014). Exploring contemporary migration. Routledge.
- Xu, Z., & Harriss, R. (2008). Exploring the structure of the US intercity passenger air transportation network: A weighted complex network approach. GeoJournal, 73(2), 87.
- Barcus, H. R., & Brunn, S. D. (2010). Place elasticity: Exploring a new conceptualization of mobility and place attachment in rural America. Geografiska Annaler: Series B, Human Geography, 92(4), 281–295.
- Hu, Y., Ye, X., & Shaw, S. L. (2017). Extracting and analyzing semantic relatedness between cities using news articles. International Journal of Geographical Information Science, 31(12), 2427-2451.
- Lin, J., Wu, Z., & Li, X. (2019). Measuring inter-city connectivity in an urban agglomeration based on multi-source data. International Journal of Geographical Information Science, 33(5), 1062-1081.
- Ye, X., Gong, J., & Li, S. (2020). Analyzing Asymmetric City Connectivity by Toponym on Social Media in China. Chinese Geographical Science, 1-13.
- Li Z., Huang X., Hu T., Ning H., Ye X., Li X., (2021),ODT FLOW: A Scalable Platform for Extracting, Analyzing, and Sharing Multi-source Multi-scale Human Mobility, preprint
- Hu T., Wang S., She B., Zhang M., Huang X., Cui Y., …, Li Z., (2021) Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges, preprint, 10.13140/RG.2.2.32549.42728
- Li Z., Huang X., Ye X., Jiang Y., Martin Y., Ning H., Hodgson M., Li X., (2021), Measuring Global Multi-Scale Place Connectivity using Geotagged Social Media Data, preprint, arXiv:2102.03991
- Hu, T., Guan, W. W., Zhu, X., Shao, Y., Liu, L., Du, J., ... & Bao, S. (2020). Building an open resources repository for COVID-19 research. Data and Information Management, 4(3), 130-147.
- Huang X., Li Z., Jiang Y., Ye X., Deng C., Zhang J., Li X., (2021), The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S., International Journal of Digital Earth, https://doi.org/10.1080/17538947.2021.1886358
- Li Z., Li X., Porter D., Zhang J., Jiang Y., Olatosi B., Weissman S. (2020) Monitoring the Spatial Spread of COVID-19 and Effectiveness of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics, JMIR Research Protocols, https://doi.org/10.2196/24432
- Huang X., Li Z., Jiang Y., Li X., Porter D., (2020) Twitter reveals human mobility dynamics during the COVID-19 pandemic, PloS One, https://doi.org/10.1371/journal.pone.0241957
- Yang C., Sha D., Liu S., Li Y., Lan H., Guan W., Hu T., Li Z., Zhang Z., Thompson J., Wang Z., Wong D., Ruan S., Yu M., Richardson D., et al., (2020) Taking the pulse of COVID-19: A spatiotemporal perspective, International Journal of Digital Earth, https://doi.org/10.1080/17538947.2020.1809723

Thank you !

Questions/Comments?

For more information about our COVID-19 related research, please visit our lab website

http://gis.cas.sc.edu/gibd/covid-19