Inferring Flow Patterns from Sequential Snapshots of Spatial Distributions

Dr. Di Zhu
Assistant Professor, Geospatial Data Intelligence (GeoDI) Lab
Department of Geography, Environment & Society
University of Minnesota, Twin Cities

U-Spatial Support for Spatial Research

University of Minnesota Driven to Discover®

China Data Institute

dizhu@umn.edu
https://dizhu-gis.com

University of Minnesota Driven to Discover®

Outline

- Background
- Methodology
- Case 1: Migration flows during Chinese Spring Festival
- Case 2: Spatial shifting patterns of CovID-19 pandemic in the U.S.
- Workflow implementation

Spatial Prediction?

Consider locations (geographic units) …

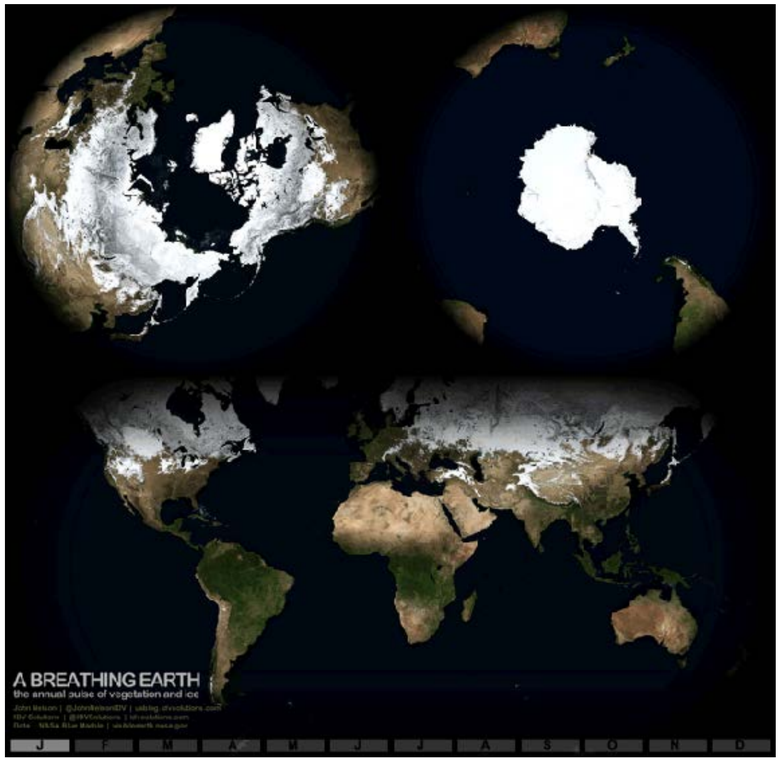
Two tasks:

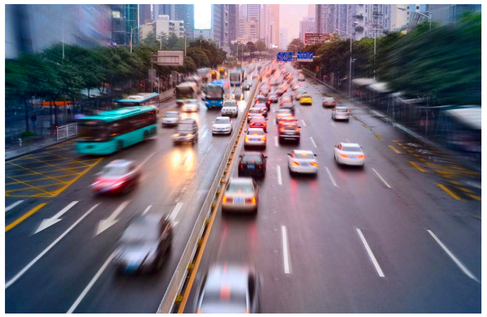
Predict for unsampled locations

Predict for unknown connections

Data-driven geospatial analysis?

- They said "… *geo-tagged data collected from various sensors offers us the opportunity to understand the complex geographic process from a data-driven perspective* …"
- **A truth or a fairy tale?**





source: http://images.google.com

Unknown Flows

- A spatio-temporal problem:

- The fact is:

- The acquisition of high quality flow (spatial interaction) data is perplexing due to:
  - Privacy
  - Preprocessing

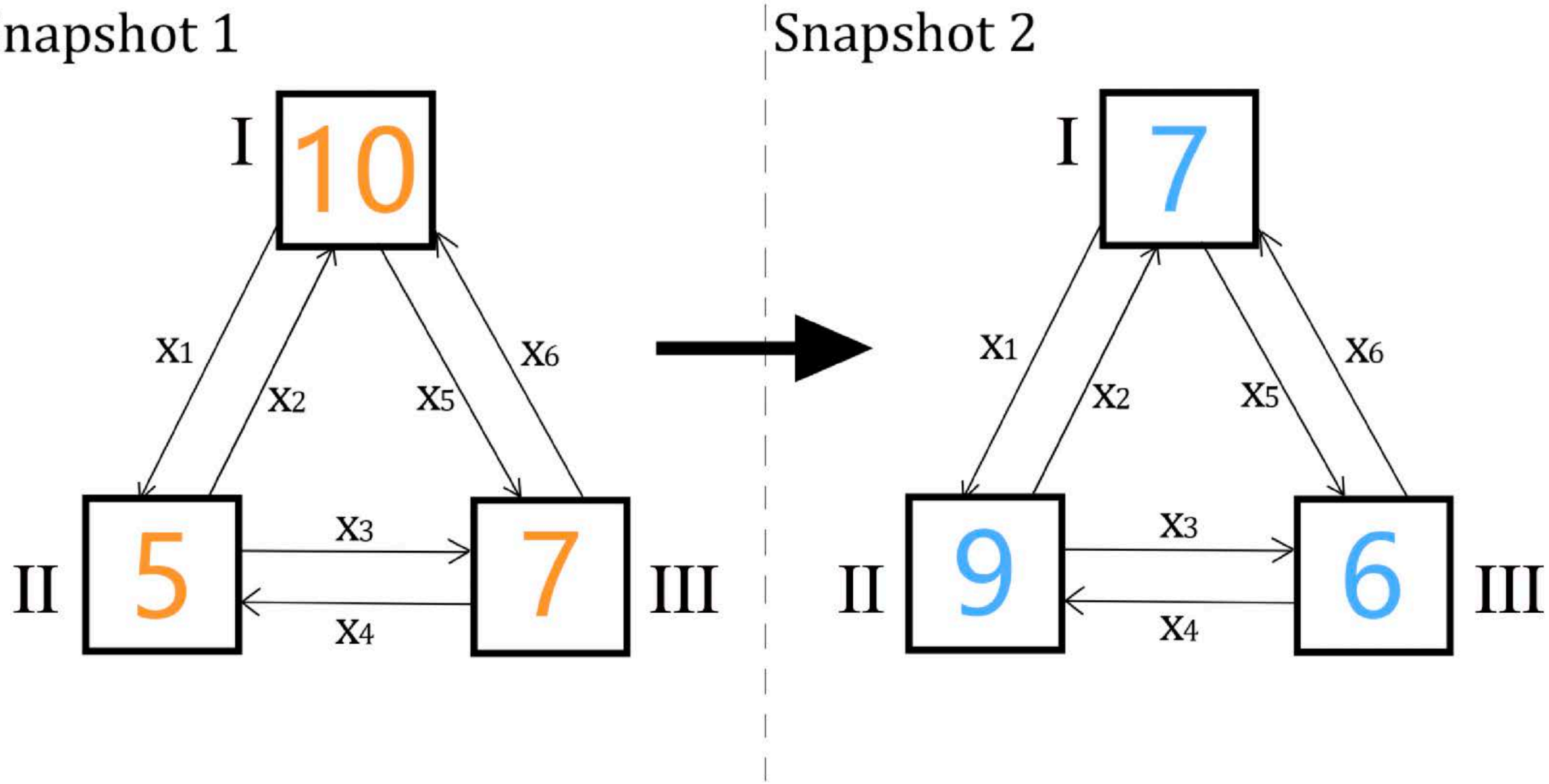
- Temporal snapshot sequence of spatial distribution data is more accessible:
  - Remote sensing
  - Geo-tagged big data

- In most cases, we have access to high quality distribution datasets. However, the detailed flows behind sequential snapshots are hard to acquire.

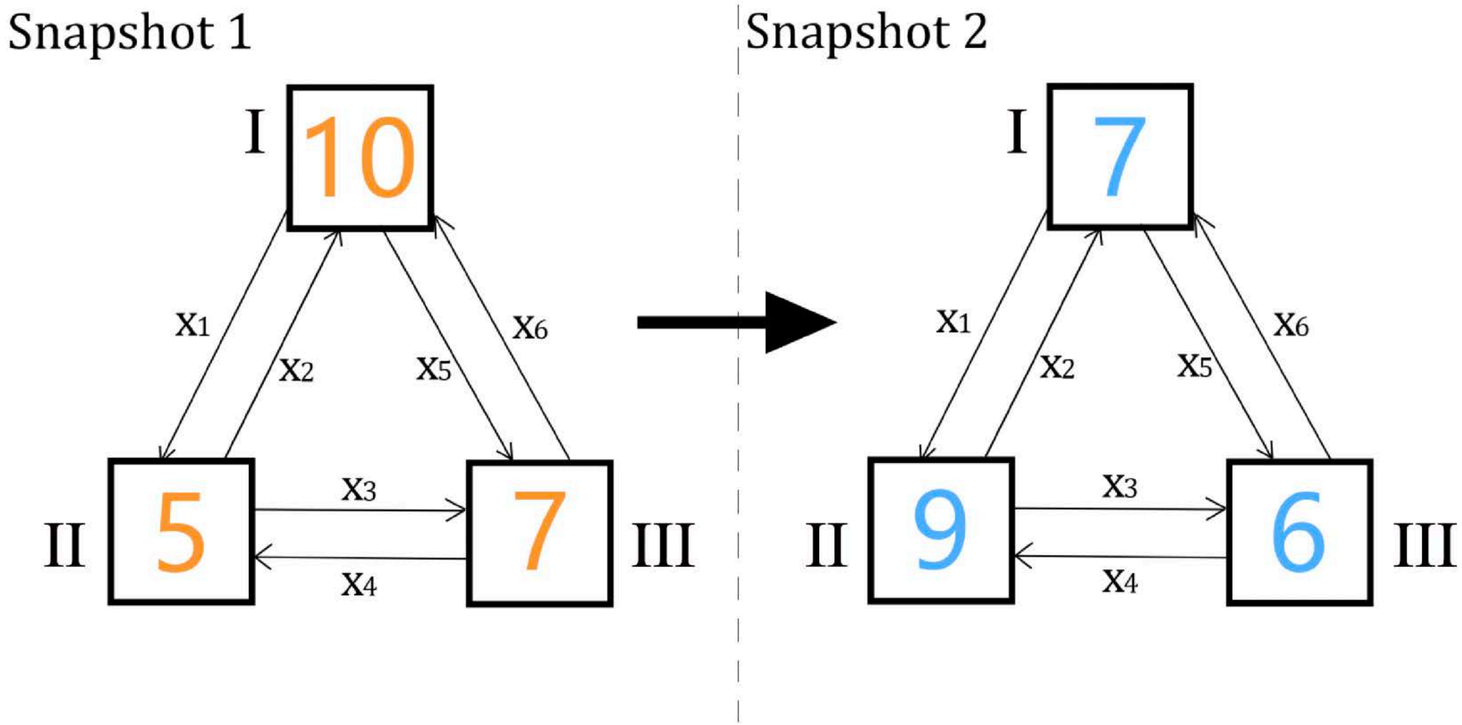
Outline

- Background
- Methodology
- Case 1: Migration flows during Chinese Spring Festival
- Case 2: Spatial shifting patterns of CovID-19 pandemic in the U.S.
- Workflow implementation

A simplified example



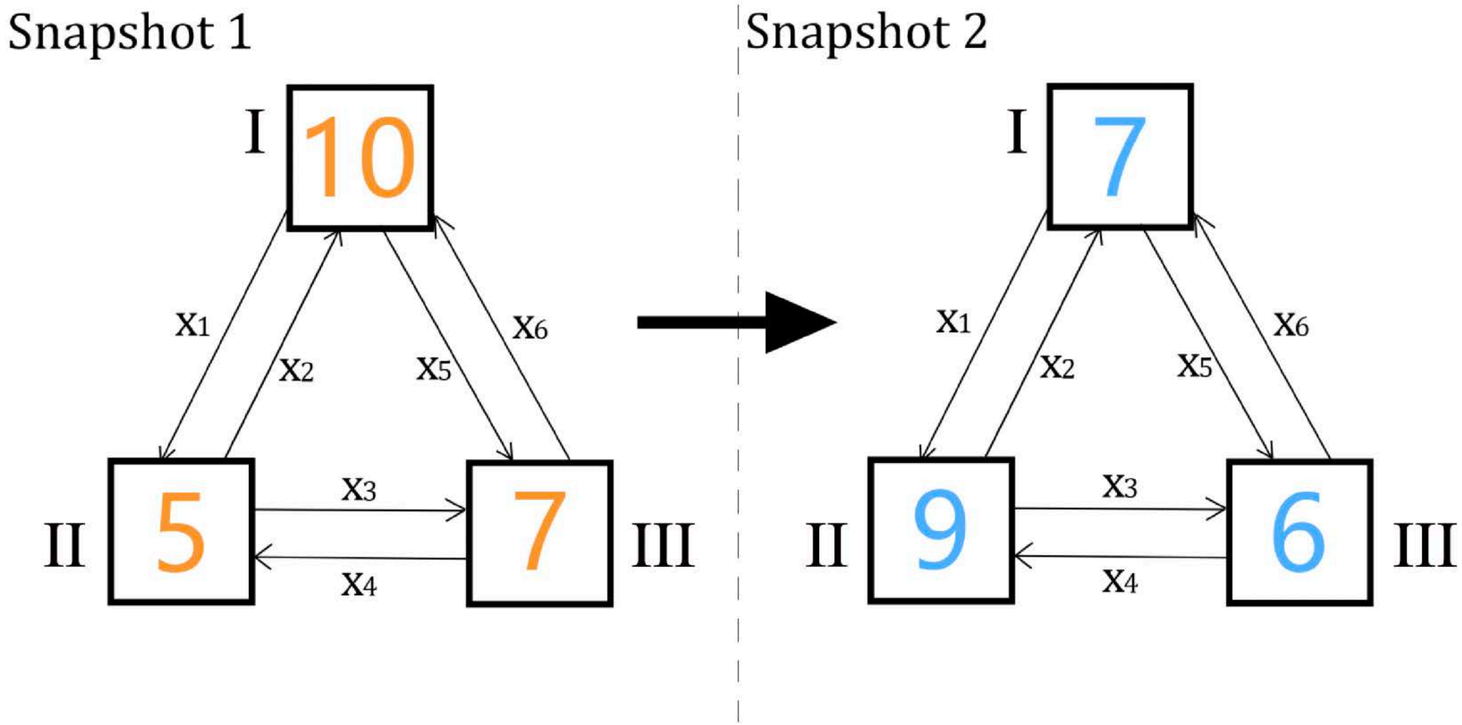
A simplified example



$$
\begin{aligned}
&\textit{minimize} && x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \\
&\textit{subject to} && -x_1 + x_2 \qquad\qquad - x_5 + x_6 = -3 \\
& && x_1 - x_2 - x_3 + x_4 \qquad\qquad = 4 \\
& && \qquad + x_3 - x_4 + x_5 - x_6 = -1 \\
& && x_i \geqslant 0 \quad for \quad i = 1, 2, ..., 6
\end{aligned}
$$

$$
\begin{aligned}
&\textit{minimize} && \mathbf{c}^T\mathbf{x} \\
&\textit{subject to} && A^{re}\mathbf{x} = \mathbf{b}^{re} \\
& && \mathbf{x} \geqslant \mathbf{0}
\end{aligned}
$$

A simplified example



minimize $x_1 + x_2 + x_3 + x_4 + x_5 + x_6$

subject to $x_1 - x_2 + x_5 - x_6 = 3$

$+ x_3 - x_4 + x_5 - x_6 = -1$

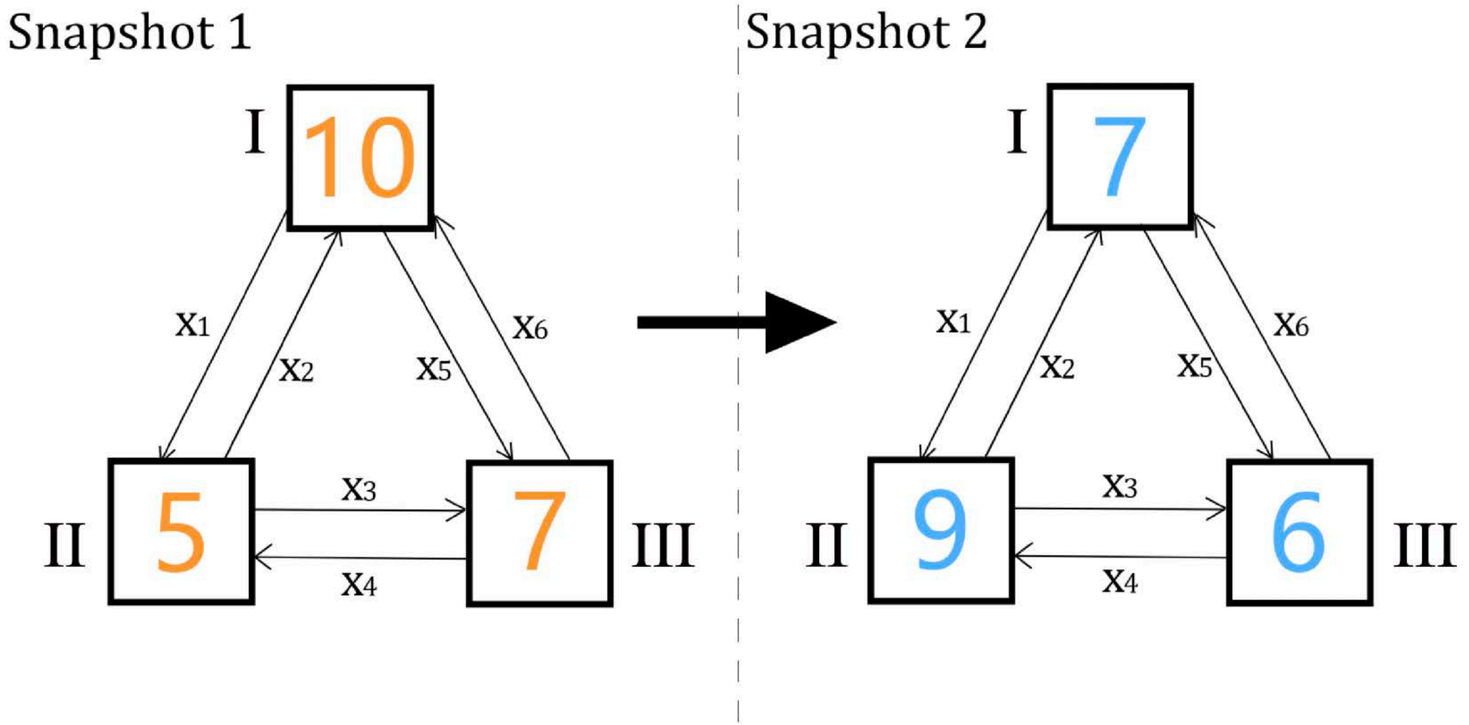
$x_i \geq 0$ *for* $i = 1, 2, \ldots, 6.$

- e.g. Select column indices {1,6}, we have:

$$x_1 = 3 + x_6$$
$$x_6 = 1$$
$$z = x_1 + x_6.$$

$$z = 5$$

A simplified example



minimize $x_1 + x_2 + x_3 + x_4 + x_5 + x_6$

subject to $x_1 - x_2 + x_5 - x_6 = 3$

$+ x_3 - x_4 + x_5 - x_6 = -1$

$x_i \geq 0$ *for* $i = 1, 2, \ldots, 6.$

- e.g. Select column indices {1,4}, we have:

$$x_1 = 3$$
$$x_4 = 1$$
$$z = x_1 + x_4.$$

$$z = 4$$

A generalized model for n units

University of Minnesota

Inferring Interactions from Distribution Snapshots (IIDS)

Input

- Two consecutive spatial distribution snapshots:

$$S_1 = [s_1, s_2, ..., s_i, ..., s_{n-1}, s_n]$$
$$S_2 = [s_1^*, s_2^*, ..., s_i^*, ..., s_{n-1}^*, s_n^*]$$

Input

- Predefined costs for interactions:

$$C = \begin{pmatrix} 0 & c_{1,2} & \cdots & c_{1,n-1} & c_{1,n} \\ c_{2,1} & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & c_{i,j} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n-1,1} & \cdots & \cdots & \cdots & \cdots \\ c_{n,1} & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

Output

- Inferred flows matrix:

$$F = \begin{pmatrix} 0 & f_{1,2} & \cdots & f_{1,n-1} & f_{1,n} \\ f_{2,1} & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & f_{i,j} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n-1,1} & \cdots & \cdots & \cdots & \cdots \\ f_{n,1} & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

$$\begin{aligned} &minimize && C^T \times F \\ &subject\ to && s_i - \sum_{j \in \mathbf{N}} f_{i,j} + \sum_{j \in \mathbf{N}} f_{j,i} = s_i^*, \quad \forall i \in \mathbf{N} \\ & && f_{i,j} \in \mathbb{Z}^n, \quad \forall i, j \in \mathbf{N} \end{aligned}$$

Outline

- **Background**
- **Methodology**
- **Case 1: Migration flows during Chinese Spring Festival**
- **Case 2: Spatial shifting patterns of CovID-19 pandemic in the U.S.**
- **Workflow implementation**

***Zhu D**, Huang Z, Shi L, et al. Inferring spatial interaction patterns from sequential snapshots of spatial distributions[J]. **International Journal of Geographical Information Science**, 2018, 32(4): 783-805.*

https://github.com/dizhu-gis/IIDS-Inferring_Interactions_from_Distribution_Snapshots

18 Jan 2016 — 22 Jan 2016
(Before Spring Festival)

7 Feb 2016 — 11 Feb 2016
(During Spring Festival)

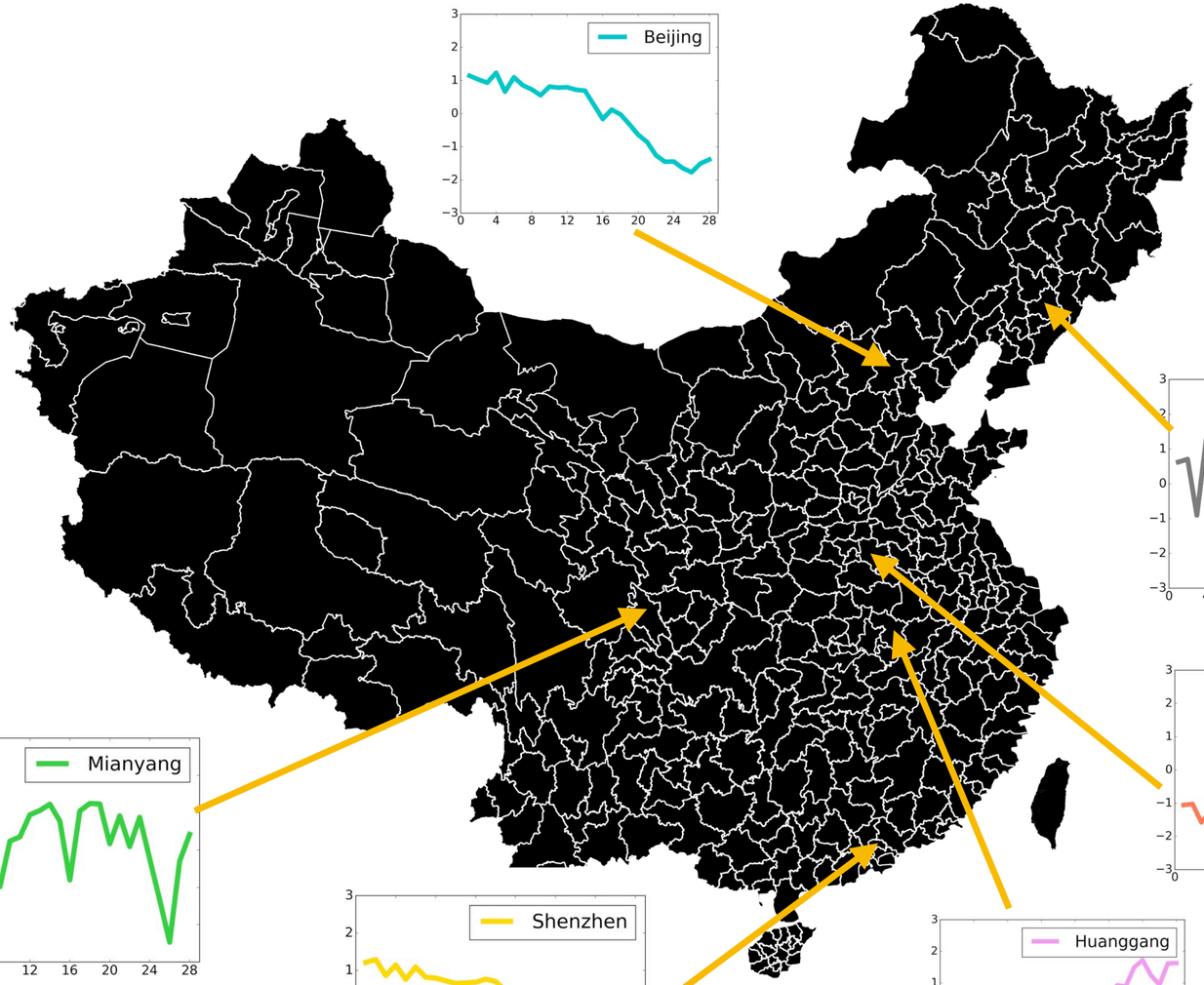
A Synthetic Map

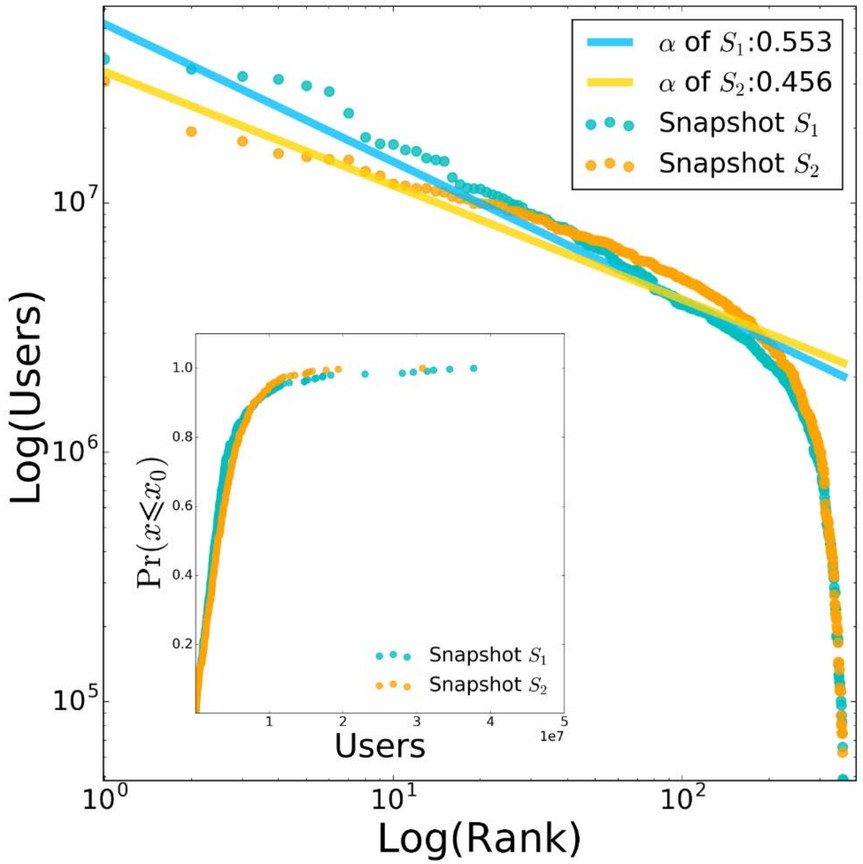
R: Before SF
G: During SF
B: Night Light

Case descriptions

Noting that our study units do not include Hong Kong, Macao and cities in Taiwan due to the bias of the dataset and all the maps we illustrated in this paper do not delineate the islands in the South China Sea for cartographic convenience.

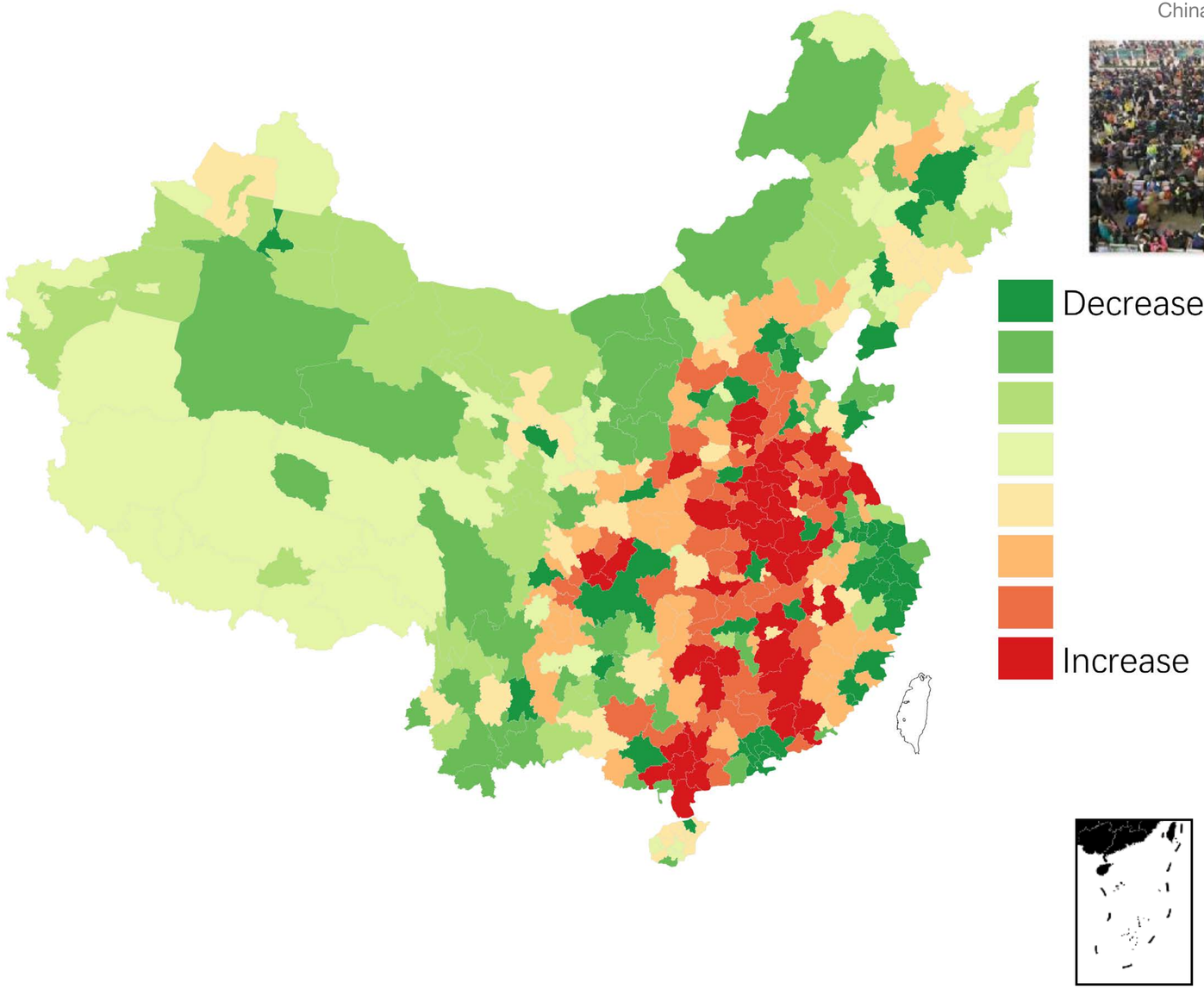
"The Spring Rush"

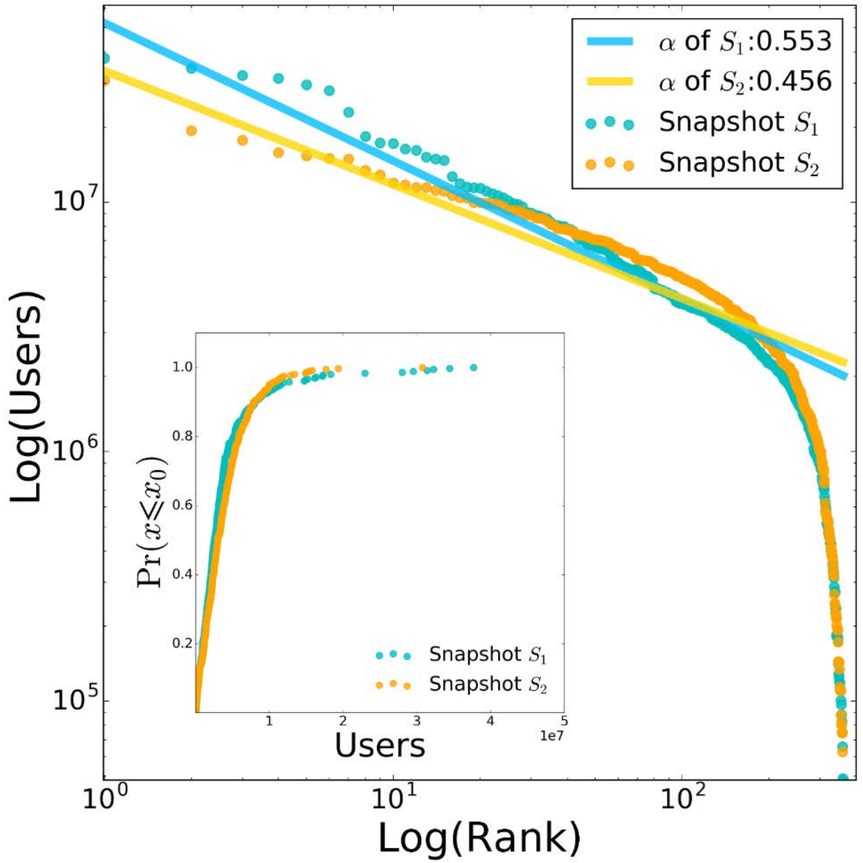




Case descriptions

Noting that our study units do not include Hong Kong, Macao and cities in Taiwan due to the bias of the dataset and all the maps we illustrated in this paper do not delineate the islands in the South China Sea for cartographic convenience.

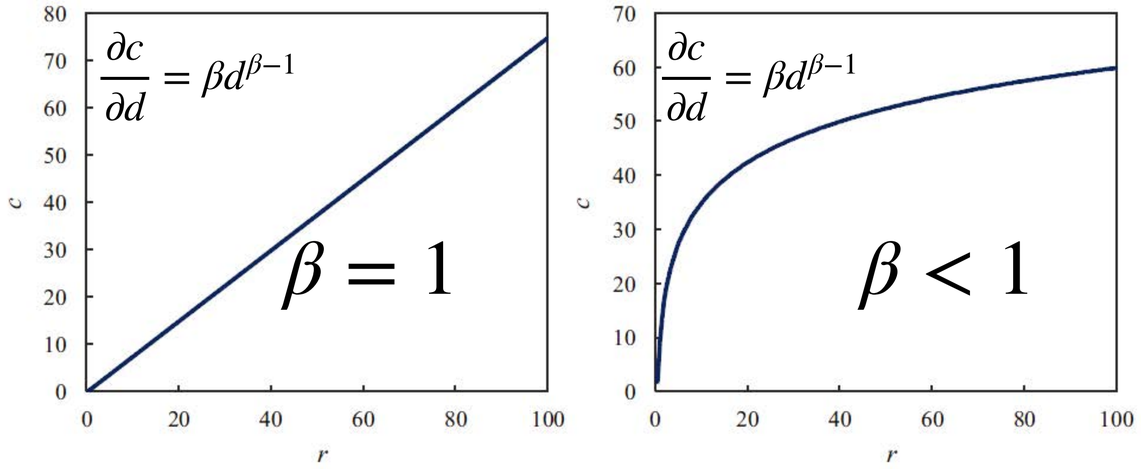


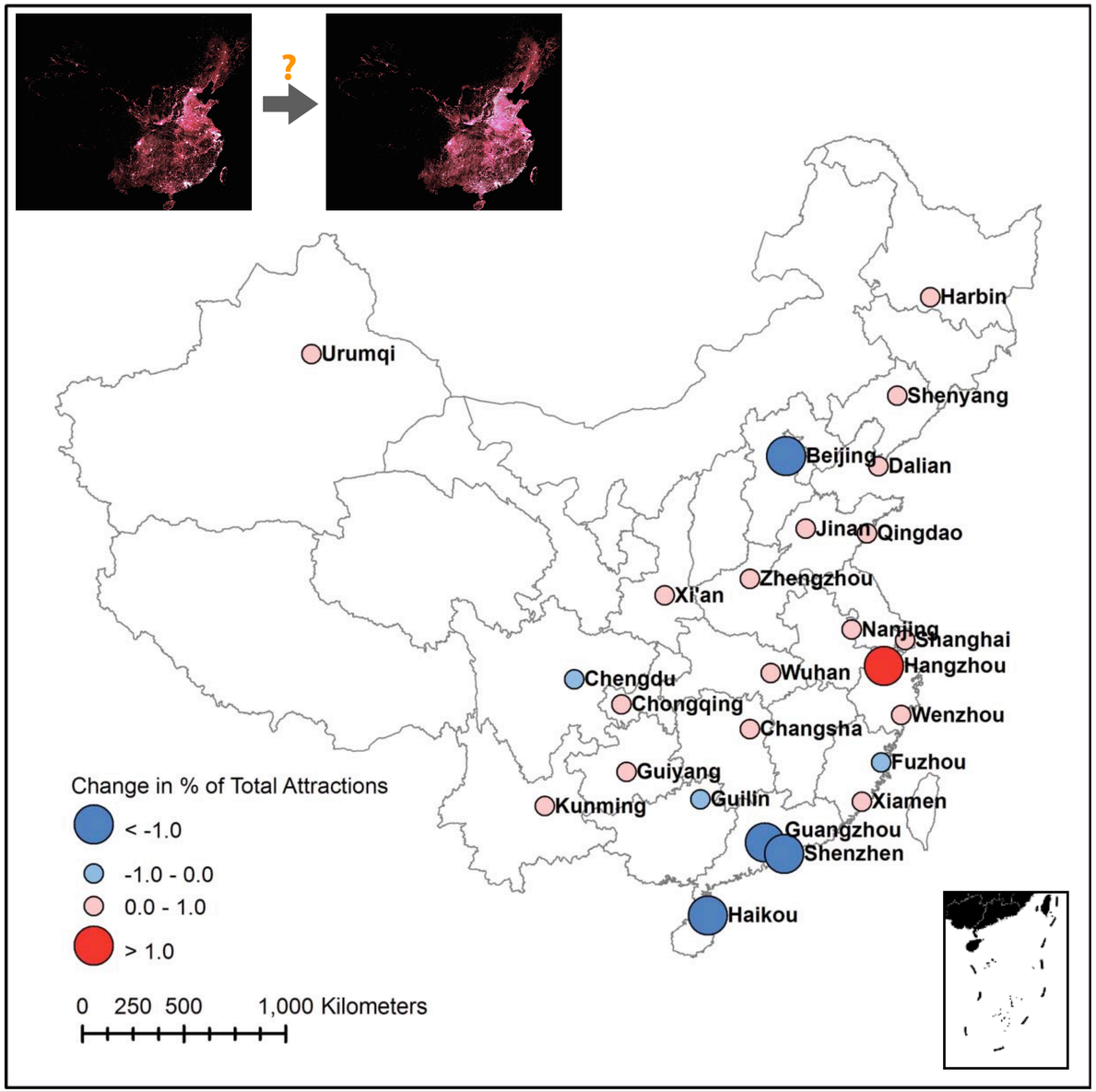


How to define the migration cost?

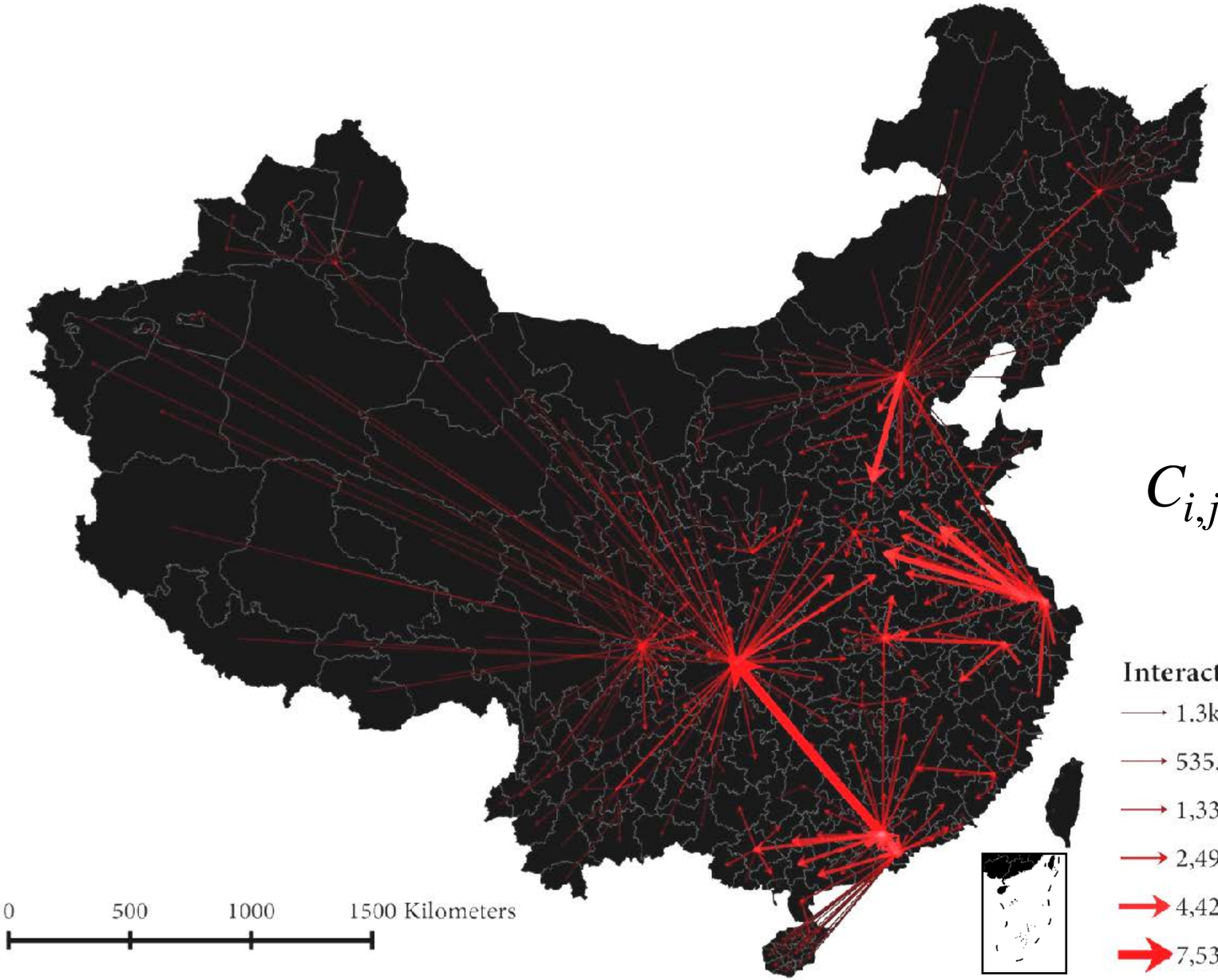
- **Home-Work displacement**
- **Demand for returning home**
- **Gravity model**
  - Job opportunities as cities' attractions
  - Preference for closer cites as the distance decay

$$C_{i,j} \propto \frac{d_{i,j}^{\beta}}{A_i A_j}$$



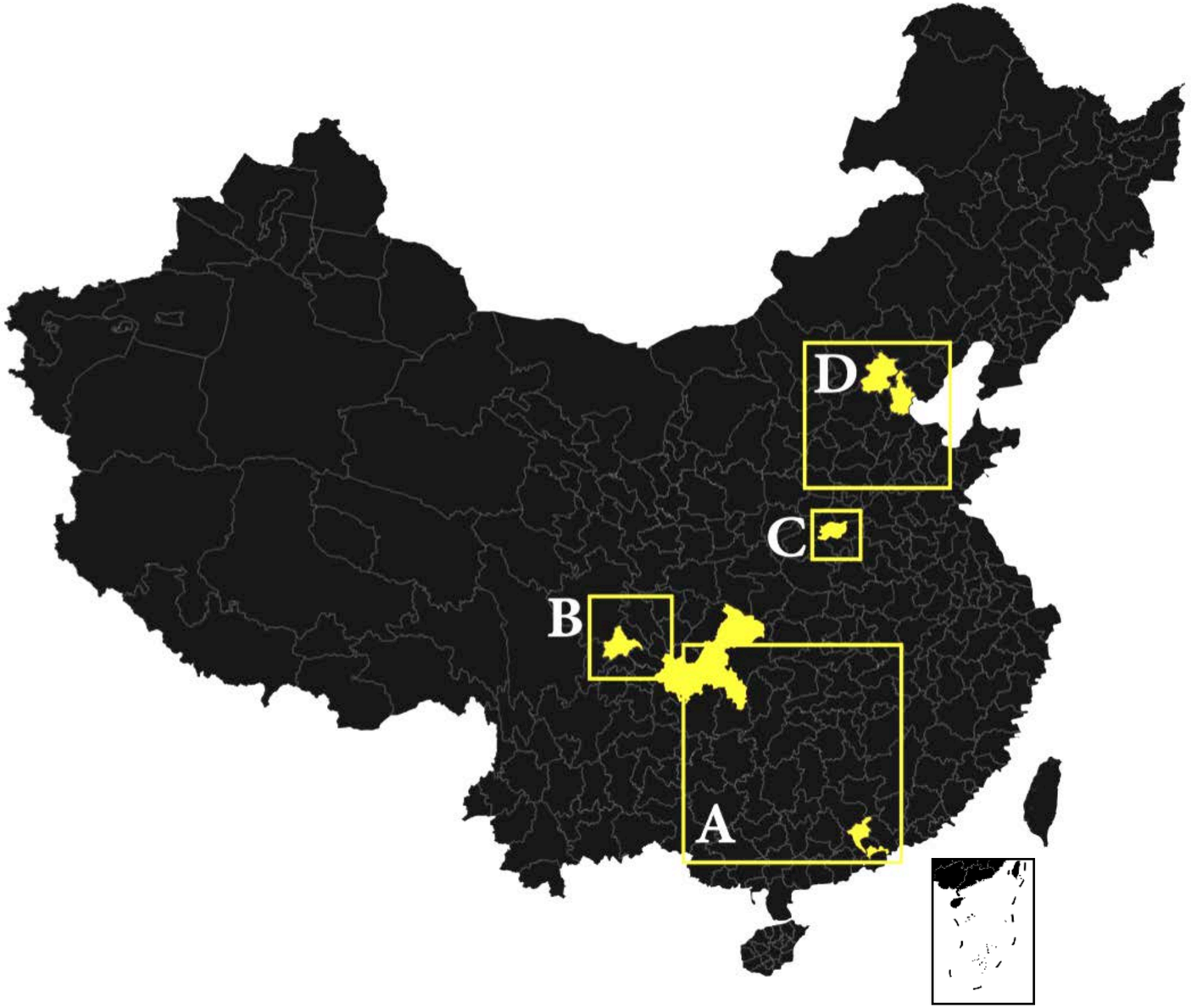


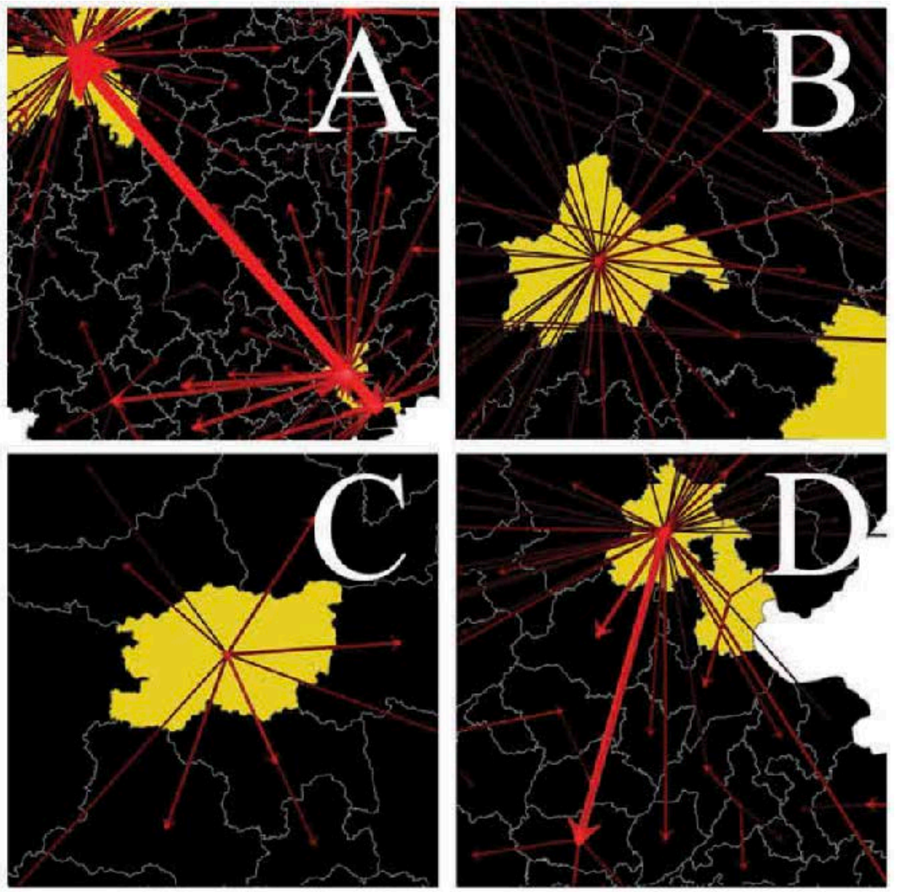
Global flow pattern



$$C_{i,j} = \alpha \frac{d_{i,j}^{\beta}}{A_i A_j} \quad \beta = 1.0$$

Local flow patterns



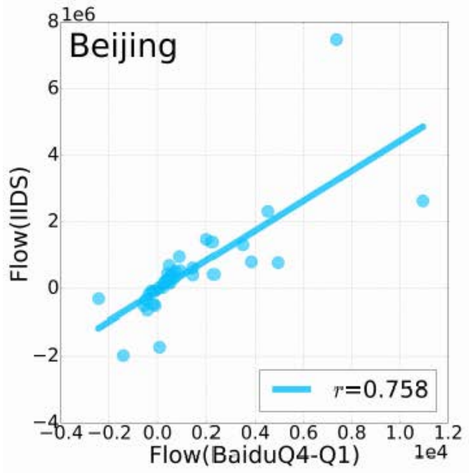


$$C_{i,j} = \alpha \frac{d_{i,j}^{\beta}}{A_i A_j} \quad \beta = 1.0$$

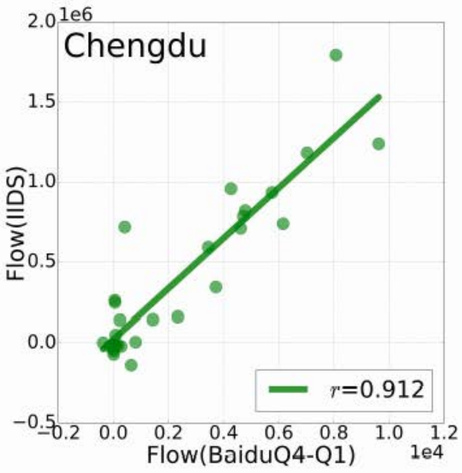
Evaluation based on a ground-truth

UNIVERSITY OF MINNESOTA

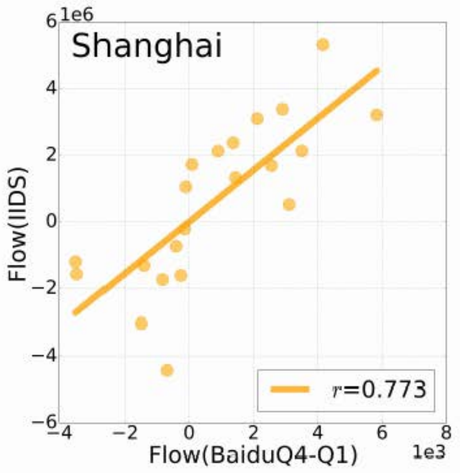
Inferred flows v.s. Baidu's quarterly migration (sampled)



(a)

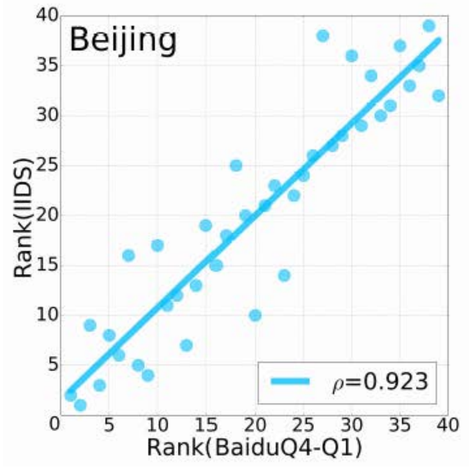


(c)

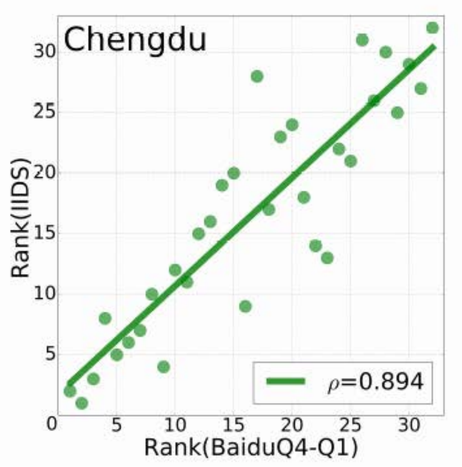


(e)

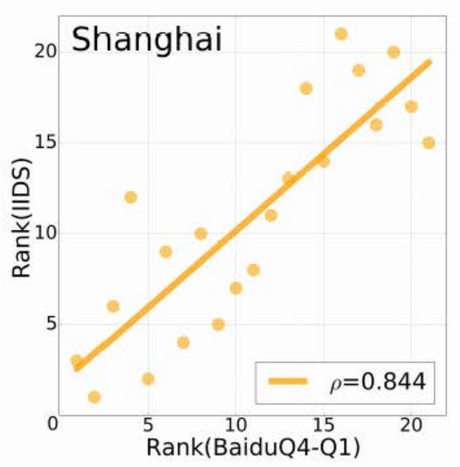
..... **Pearson's** r



(b)



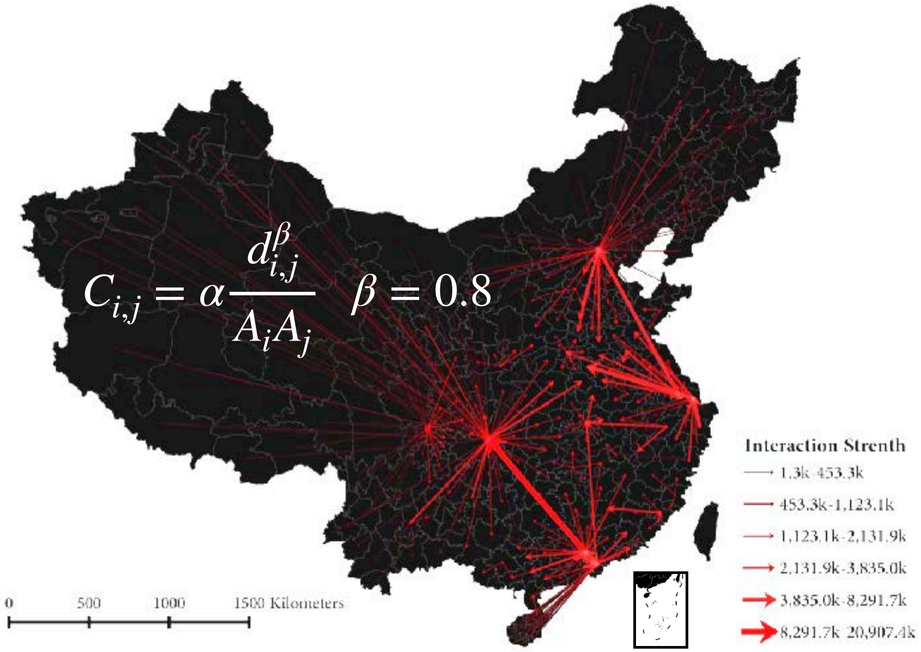
(d)

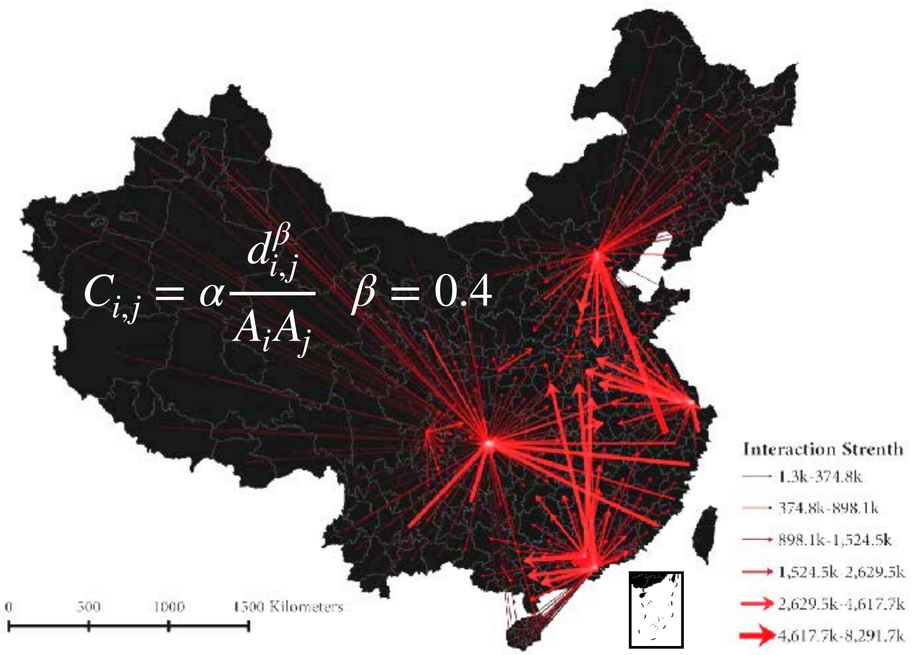


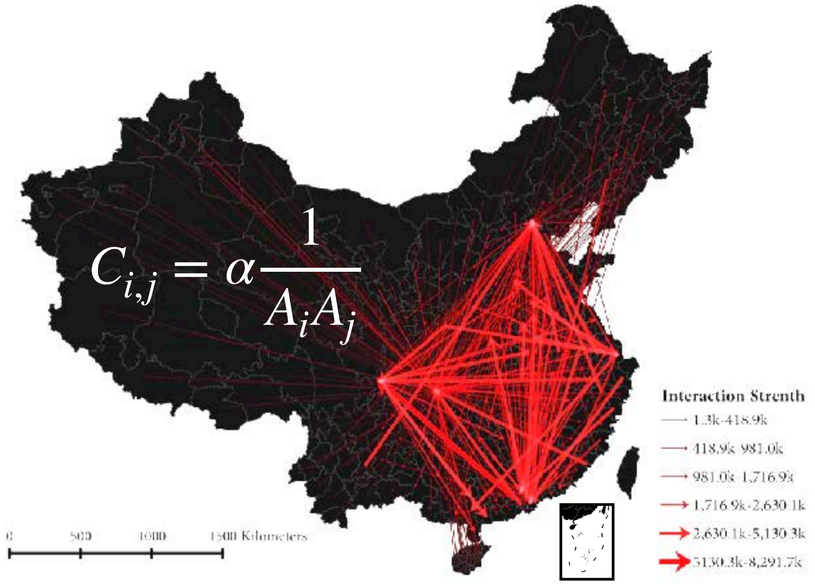
(f)

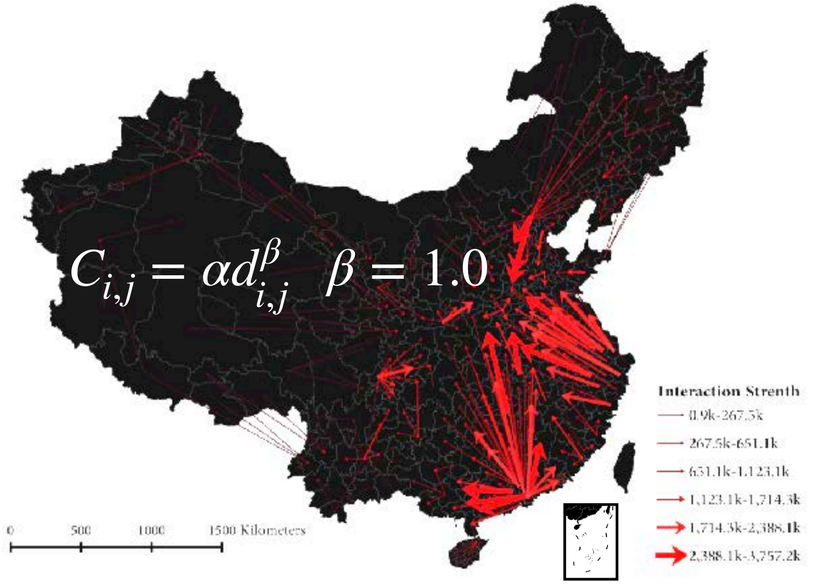
..... **Spearman's** ρ

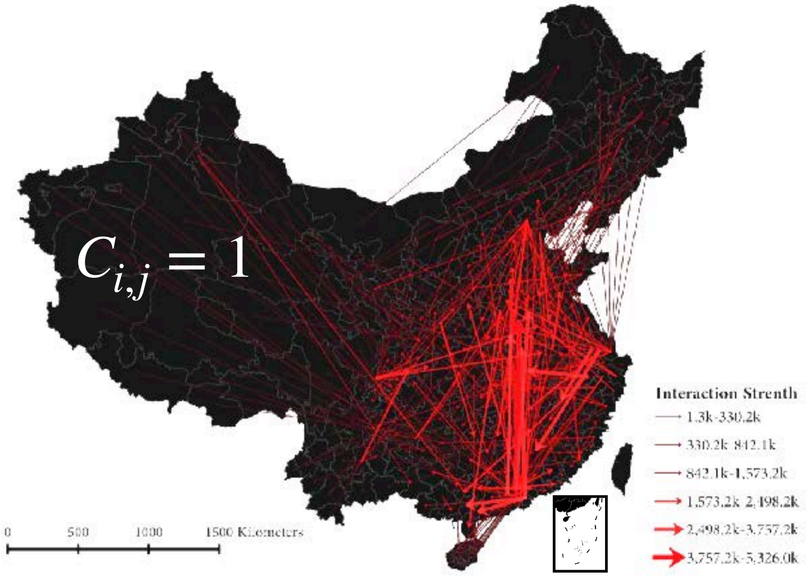
Definition of the cost $C_{i,j}$











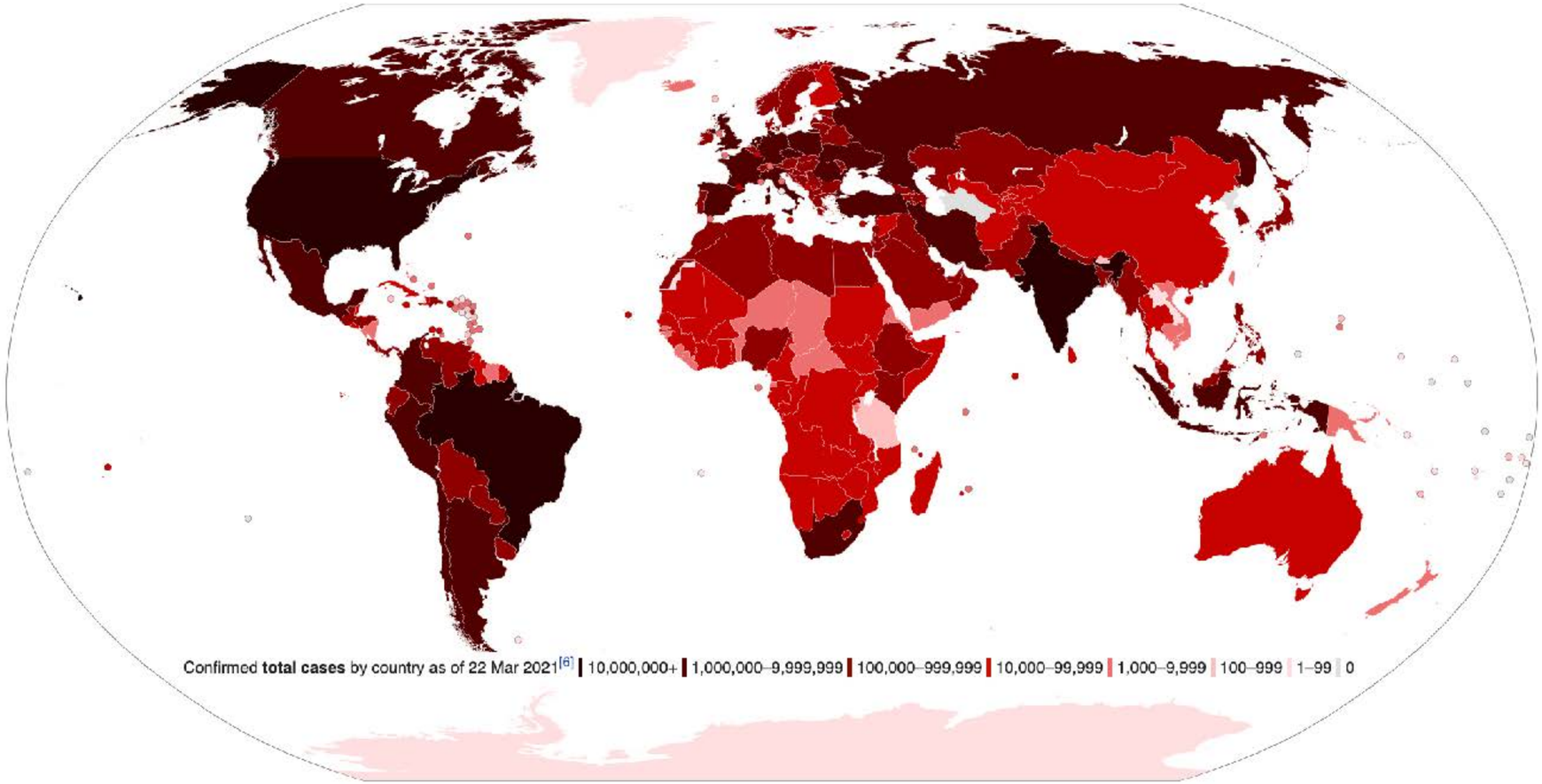
Outline

- **Background**
- **Methodology**
- **Case 1: Migration flows during Chinese Spring Festival**
- **Case 2: Spatial shifting patterns of CovID-19 pandemic in the U.S.**
- **Workflow implementation**

Zhu D, *Ye X, Manson S. Revealing the spatial shifting pattern of COVID-19 pandemic in the United States[J].* ***Scientific Reports****, 2021, 11(1): 8396.*

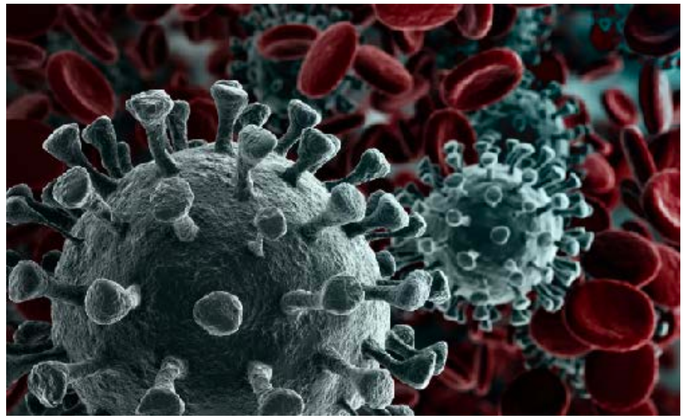
https://github.com/dizhu-gis/CovIDSpatialShifts

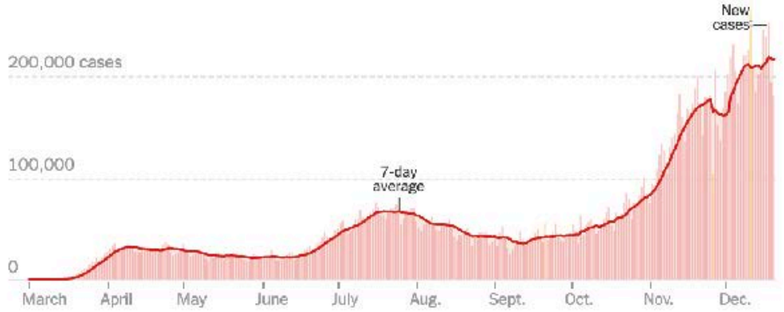
Epidemic snapshot maps



https://en.wikipedia.org/wiki/COVID-19_pandemic_by_country_and_territory

- Spatial variation
- Total confirmed cases (CC) by Mar. 22, 2021





	TOTAL REPORTED	ON DEC. 20	14-DAY CHANGE
Cases	17.8 million+	179,801	+10% →
Deaths	317,800	1,422	+19% →
Hospitalized		113,663	+13% →

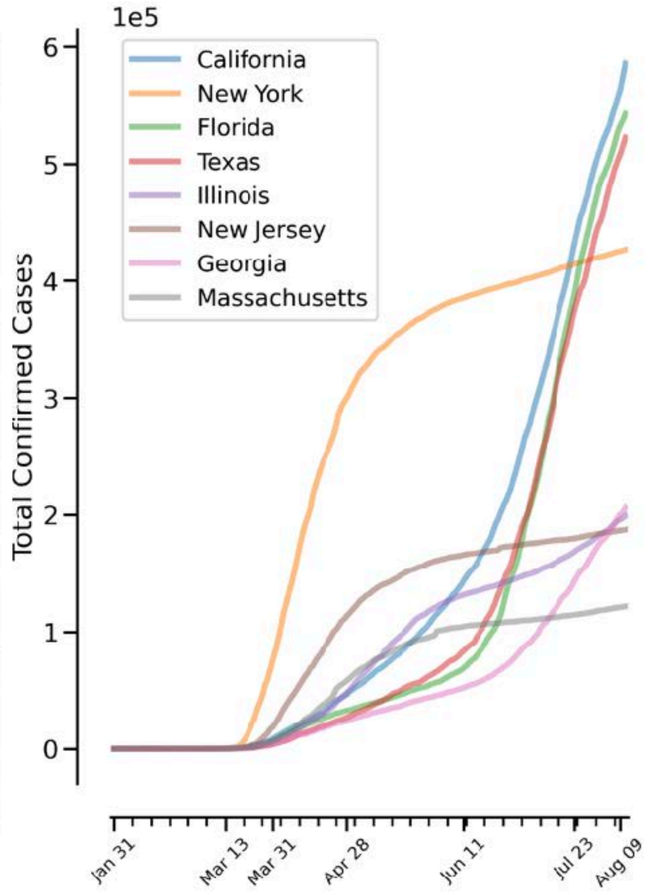
https://www.nytimes.com/

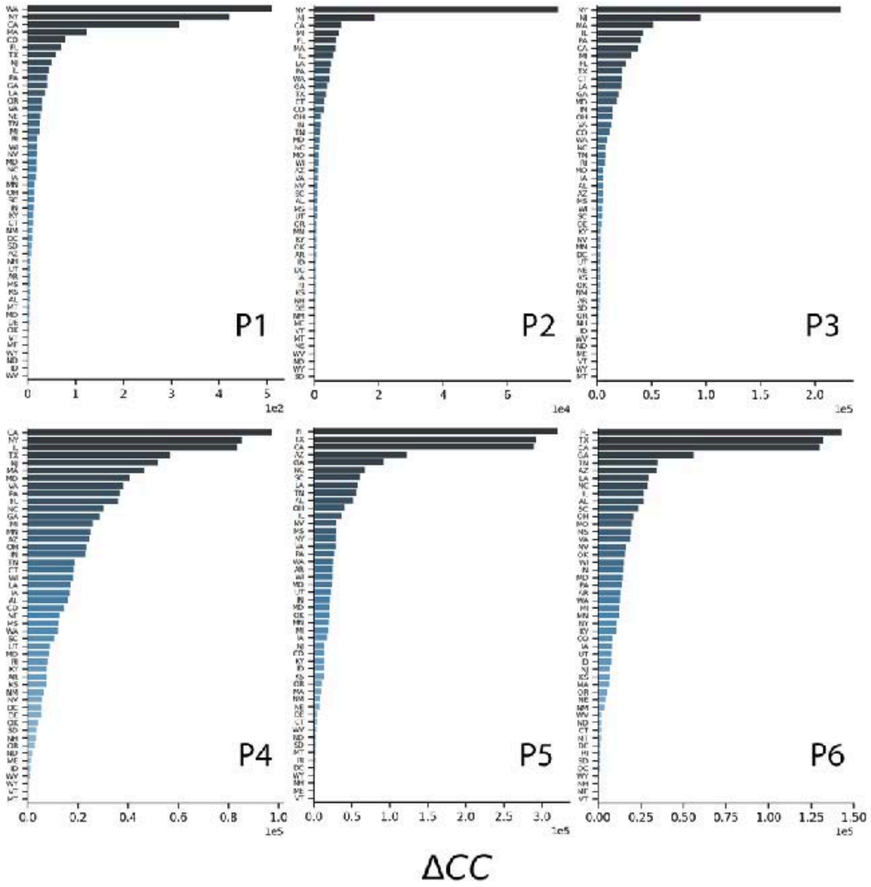
- Temporal variation
- Dynamic tracking of CovID-19

Case description

- Key events to determine the pandemic phases
- Temporal signature of total CC
- Rank-size of increased CC

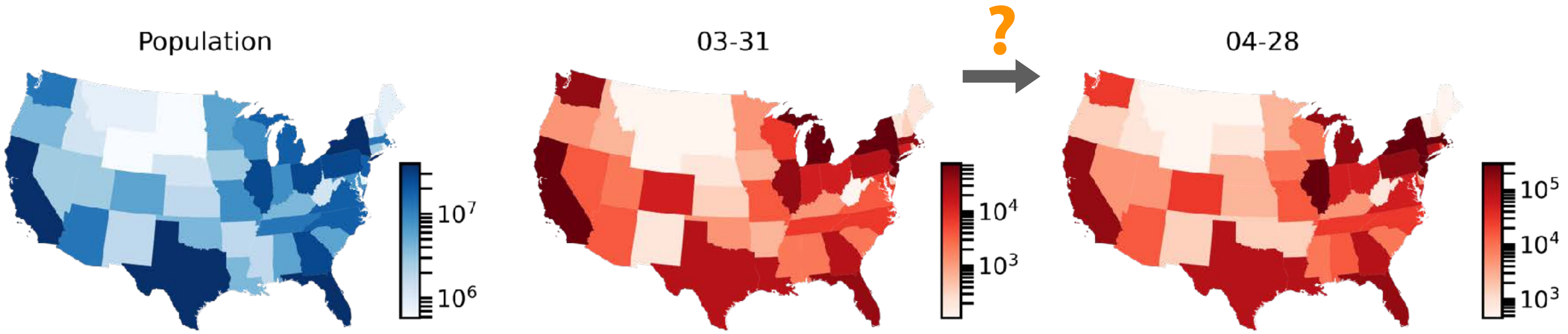
Date	Events	Phase
2020. 1. 21	Officials in Washington state confirm the first case on US soil.	N/A
2020. 1. 30	The United States reports its first confirmed case of person-to-person transmission of the coronavirus. WHO determines that the outbreak constitutes a PublicHealth Emergency of International Concern (PHEIC).	N/A
2020. 1. 31	The U.S. government announces it will deny entry to foreign nationals who have traveled in China in the last 14 days.	P1
2020. 2. 11	WHO names the coronavirus Covid-19.	
2020. 3. 11	WHO declares the novel coronavirus outbreak to be a pandemic. The U.S is restricting travel from Europe to slow the spread of coronavirus.	
2020. 3. 13	The U.S. declares a national emergency to free up $50 billion in federal resources to combat coronavirus.	P2
2020. 3. 31	Most states have reacted to the stay-at-home order.	P3
2020. 4. 8	China reopens Wuhan after a 76-day lockdown.	
2020. 4. 28	The total number of confirmed cases reaches one million.	P4
2020. 5. 25	The death of George Floyd sparked civil right protests and anti-lock down protests across the United States.	
2020. 6. 11	The total number of confirmed cases reaches two million.	P5
2020. 7. 7	The U.S. administration notifies Congress and the UN that the US is formally withdrawing from WHO.	
2020. 7. 23	The total number of confirmed cases reaches four million.	P6
2020. 8. 09	The total number of confirmed cases reaches five million.	





CovID-19 pandemic is **dynamic in space and time**, but how to tell the story beyond epidemic snapshot maps?

Case description



- Data:
  - CovID-19 daily reports (from 01-31 to 08-09) — New York Times
  - Socioeconomic attributes (pop., household, GDP,…) — Census
  - Twitter movements — Twitter API

- How did the pandemic centres shift spatially over time under a mix of interventions?

Case description

1/21: first case

P1 **1/31**: ban travel from China

2/29: first death

3/11: WHO declares pandemic

3/13: national emergency

P2 **3/31**: stay-at-home order one-

P3 by-one for all states

4/28: 1m cases

P4 5/25: George Floyd a 46-year-old black man was killed in Minneapolis, Black lives matter protests start

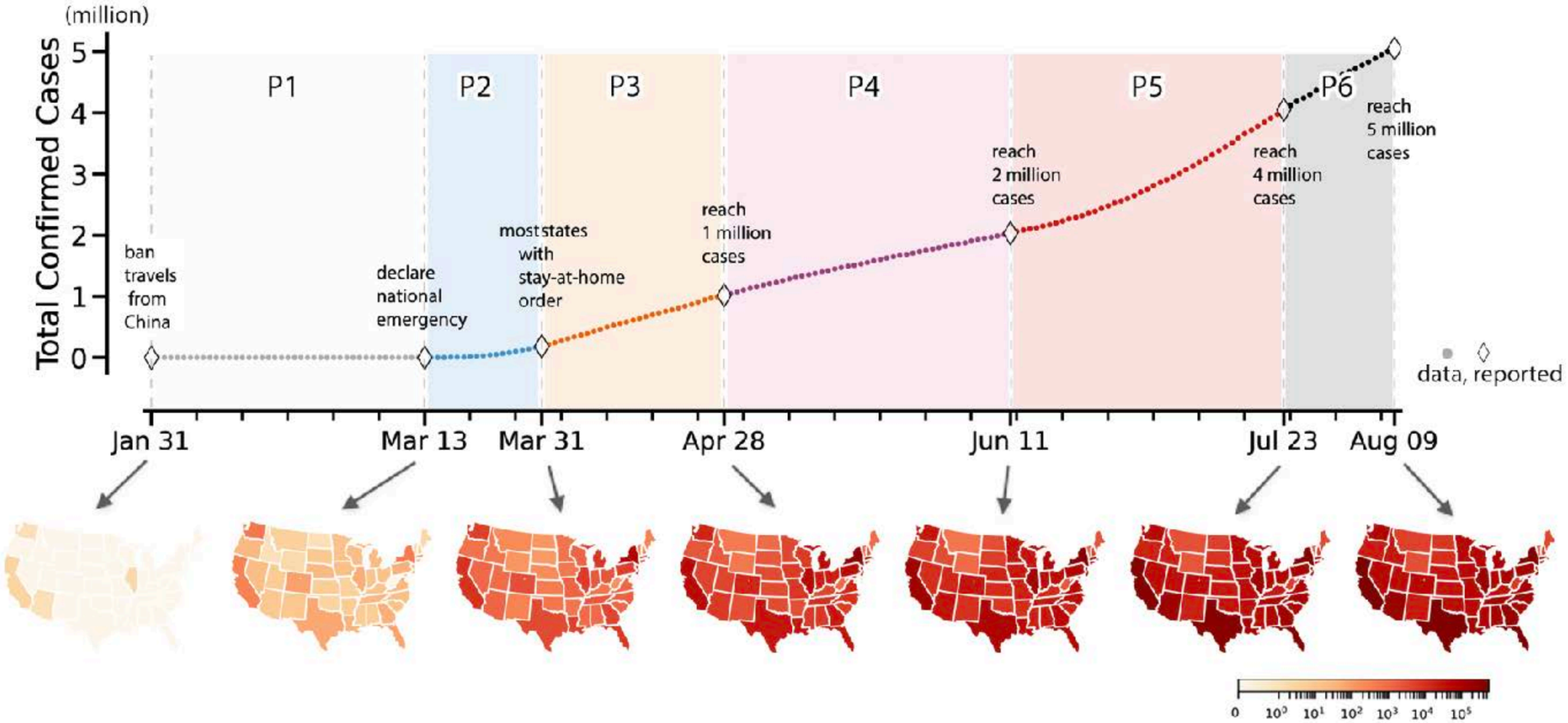
5/27: 100,000 deaths

P5 **6/11**: 2m cases

7/23: 4 million cases

P6 7/29: 150000 deaths

8/9: 5m cases



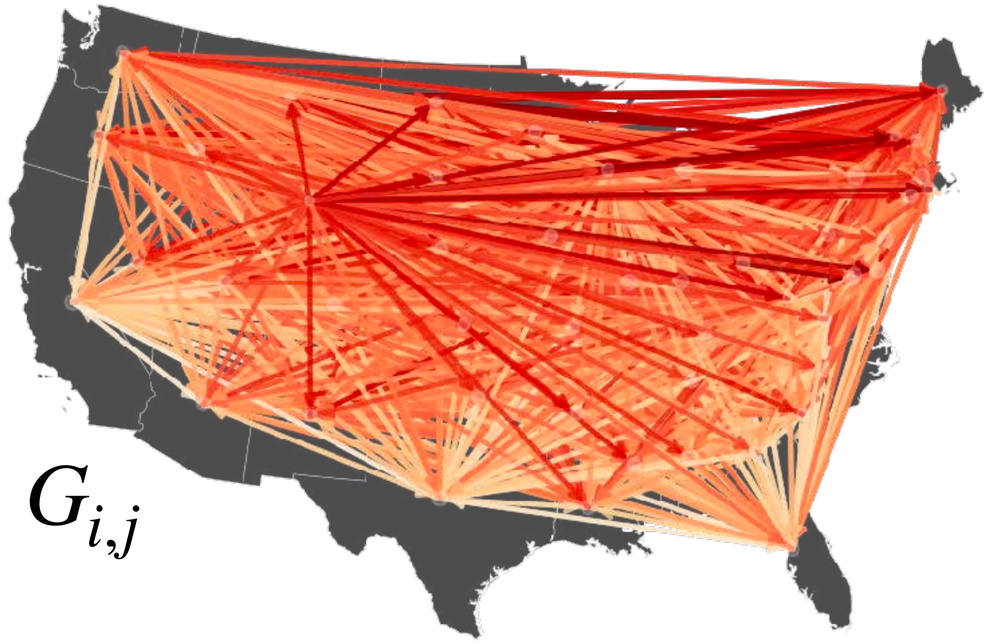
Spatial segregation & social distancing

University of Minnesota

Using IIDS algorithm:

- Variation of total confirmed case as the constraints
- Minimizing the system's total cost

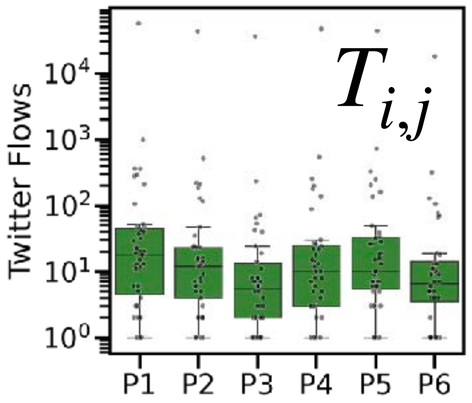
$$
\begin{aligned}
&minimize && C^T \times X \\
&subject\ to && -\sum_{j \in \mathbf{N}} x_{i,j} + \sum_{j \in \mathbf{N}} x_{j,i} = \Delta cc_i, \quad \forall i \in \mathbf{N} \\
& && x_{i,j} \in \mathbb{R}, \quad x_{i,j} \geq 0 \quad \forall i, j \in \mathbf{N}
\end{aligned}
$$

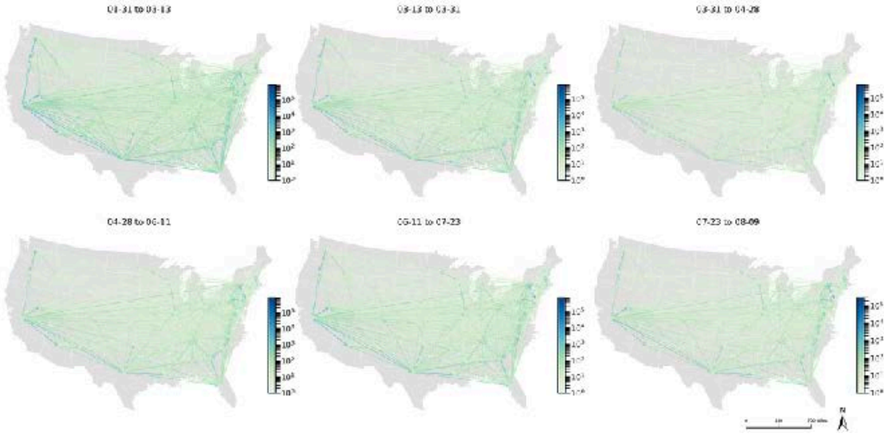


- Modeling the spatial shift costs as integrated effects
  - Gravity as the spatial segregation $G_{i,j}$
  - Twitter movements as the social distancing $T_{i,j}$

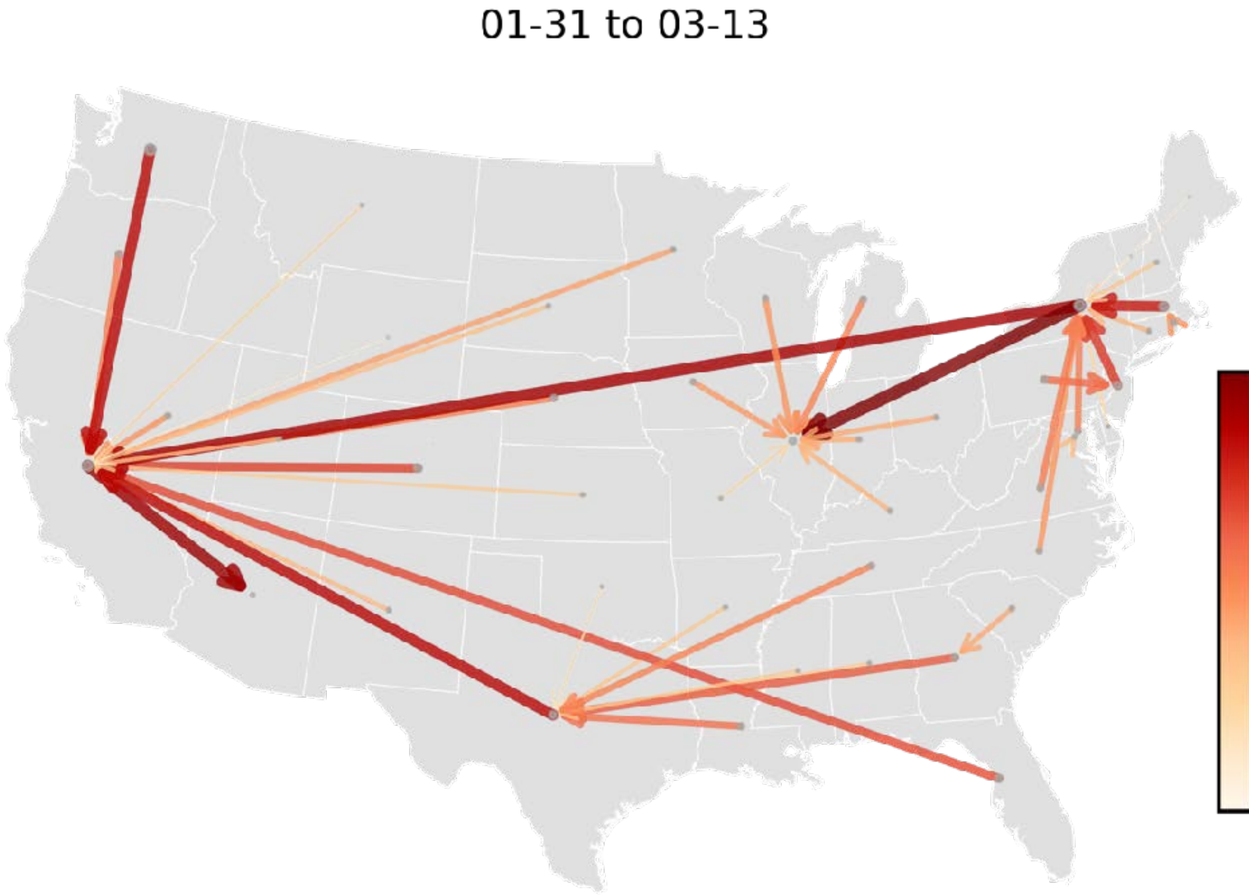
$$
c_{i,j} \propto \frac{G_{i,j}}{T_{i,j}} = \frac{k \frac{d_{i,j}^{\beta}}{A_i A_j}}{log_{10}(m_{i,j} + \delta)}
$$

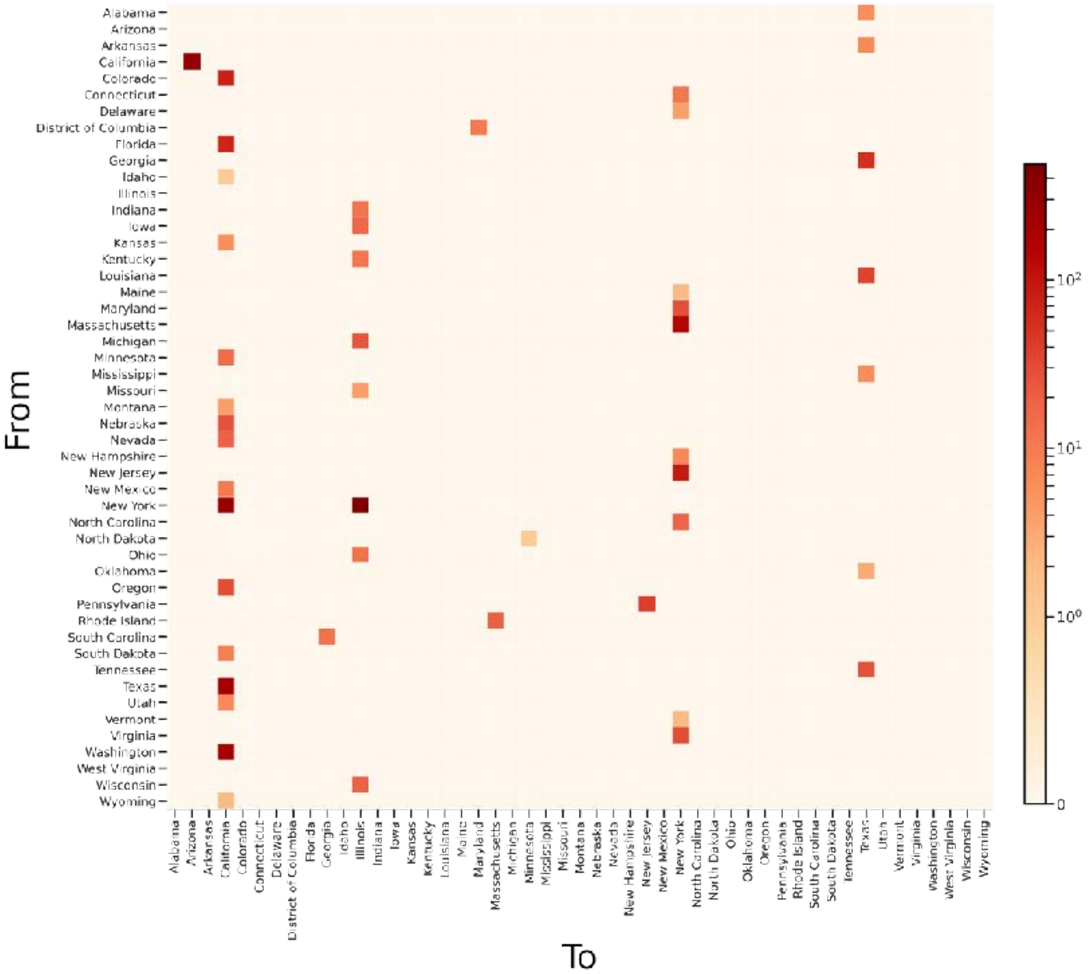
$$
\beta = 0.8 \quad k = 10^8, \delta = 1
$$





Results (P1)

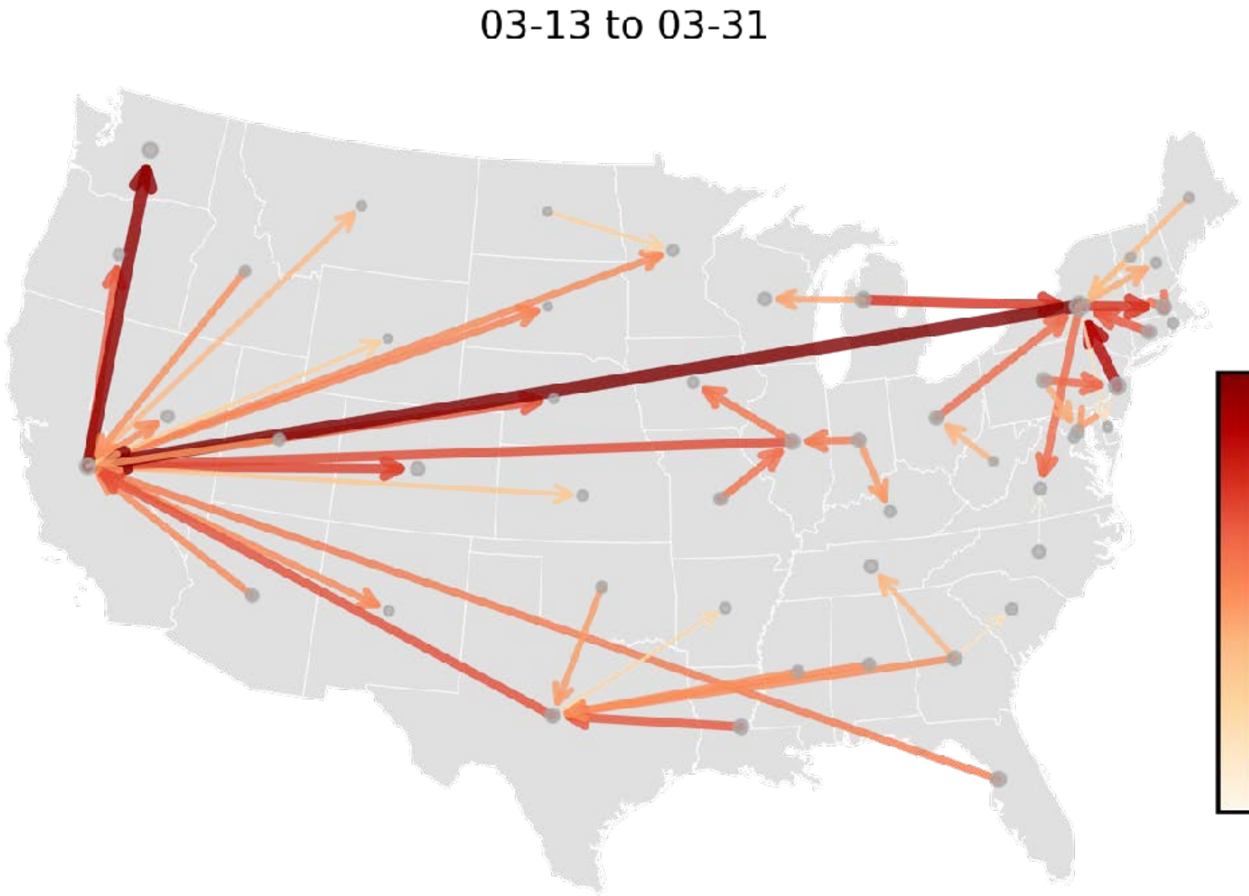


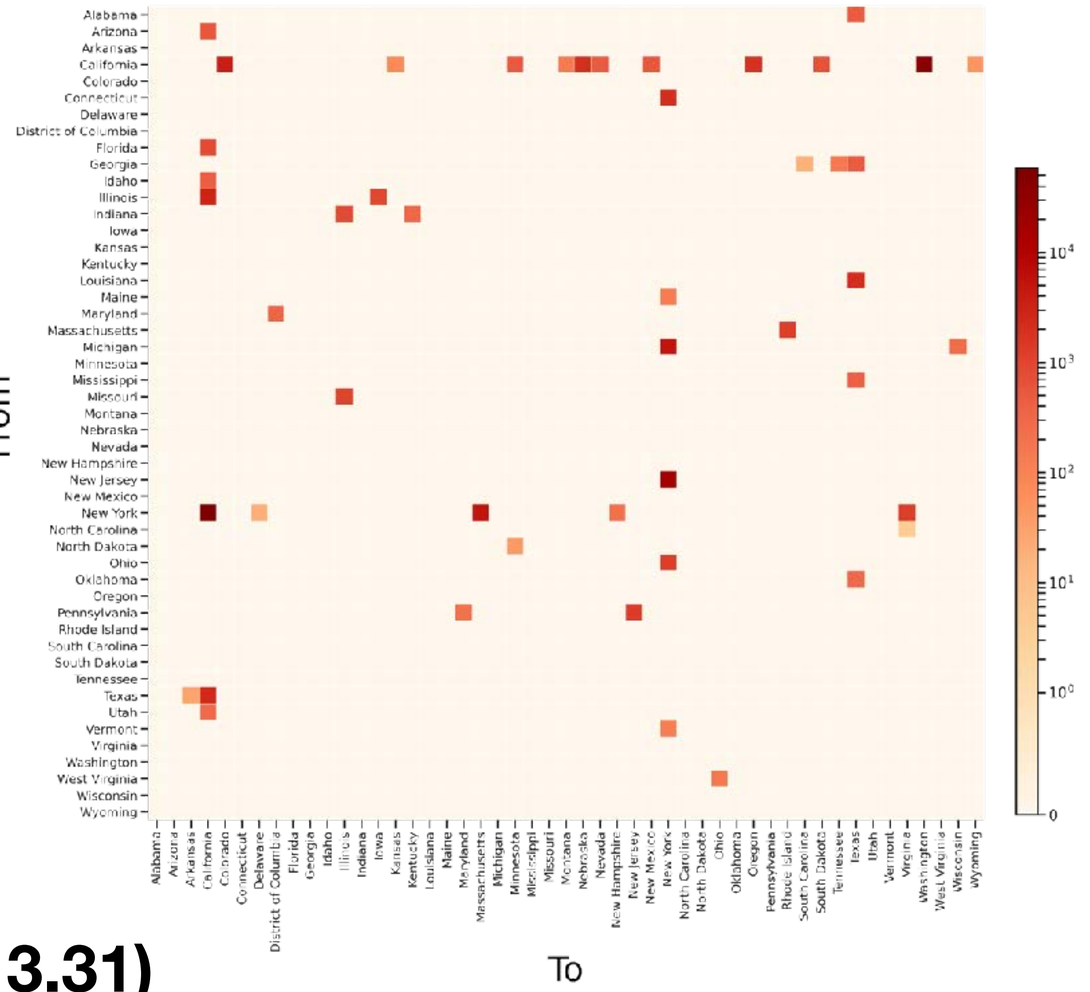


Internal Spreading (1.31 to 3.13)

1. California, Illinois, New York, Texas are the major gravity centers for pandemic shifting at the beginning
2. Strong concentration of the pandemic shifts
3. No restrictions for domestic trips and activities
4. The risk is not very high (up to ~10^2), but long-range shifts exist

Results (P2)

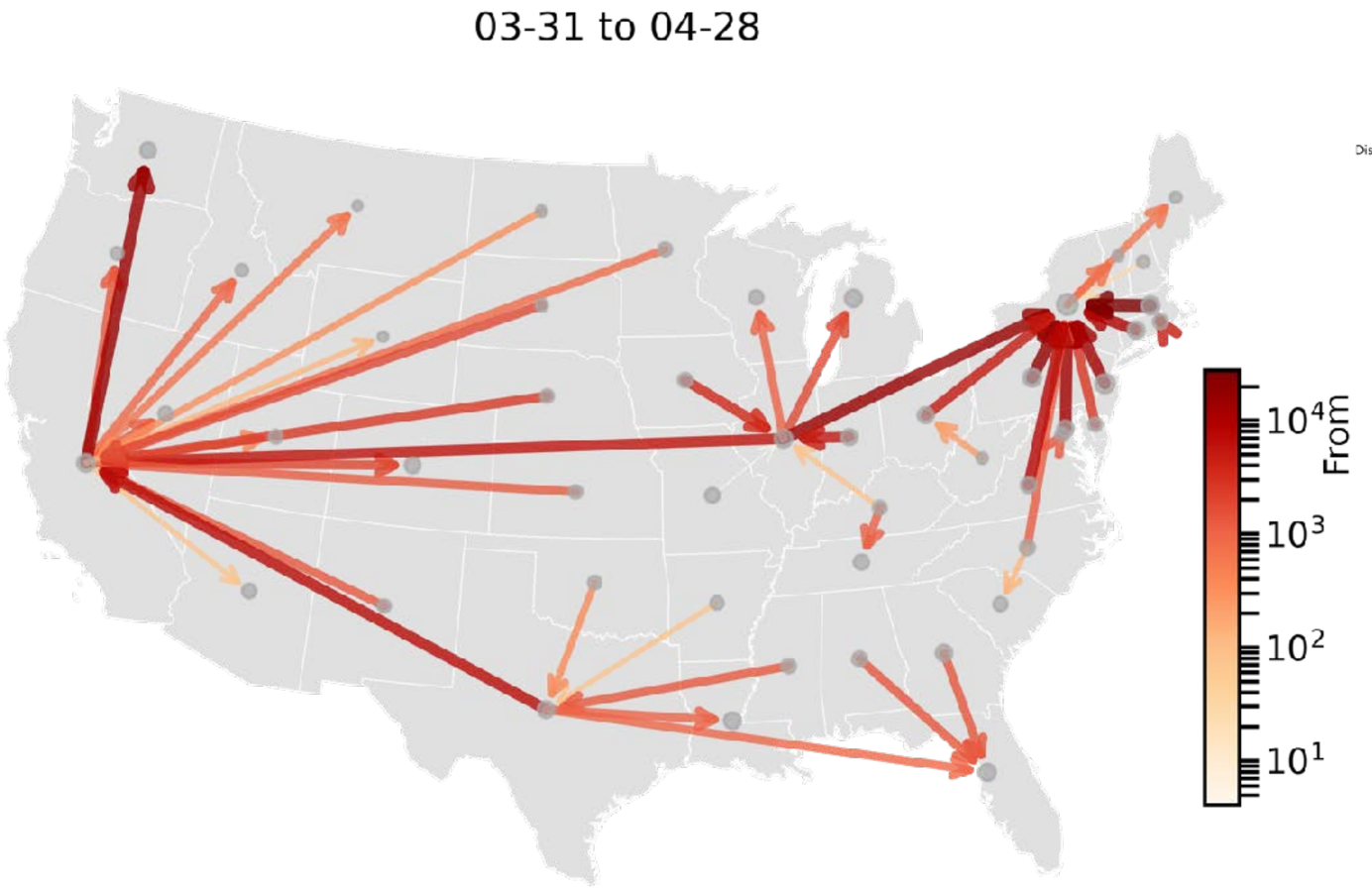


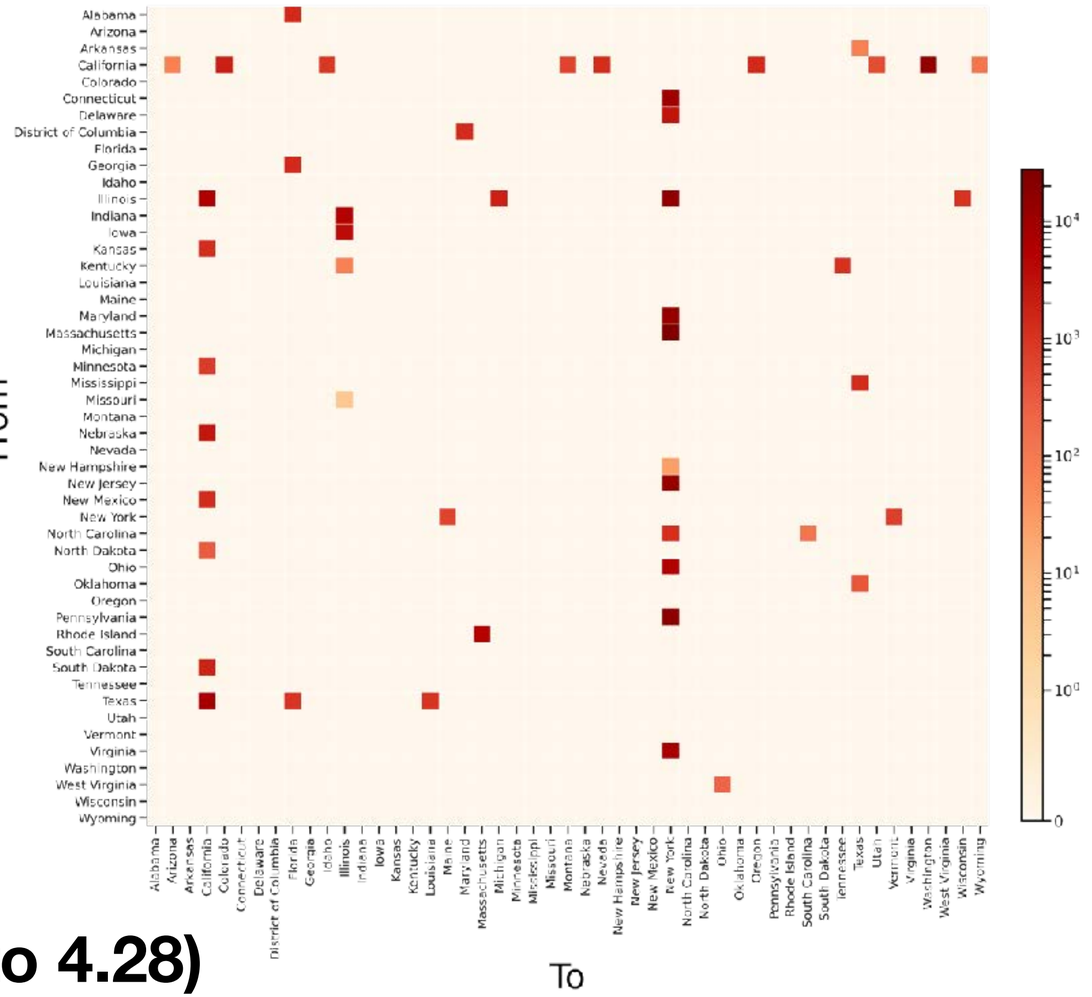


Stay-at-home order state-by-state (3.13 to 3.31)

1. The shifted risks are higher (up to ~10^4)
2. California began to affect other states as both in-shifts and out-shifts existed.
3. Stay-at-home order is working as there are more short-range shifts.

Results (P3)

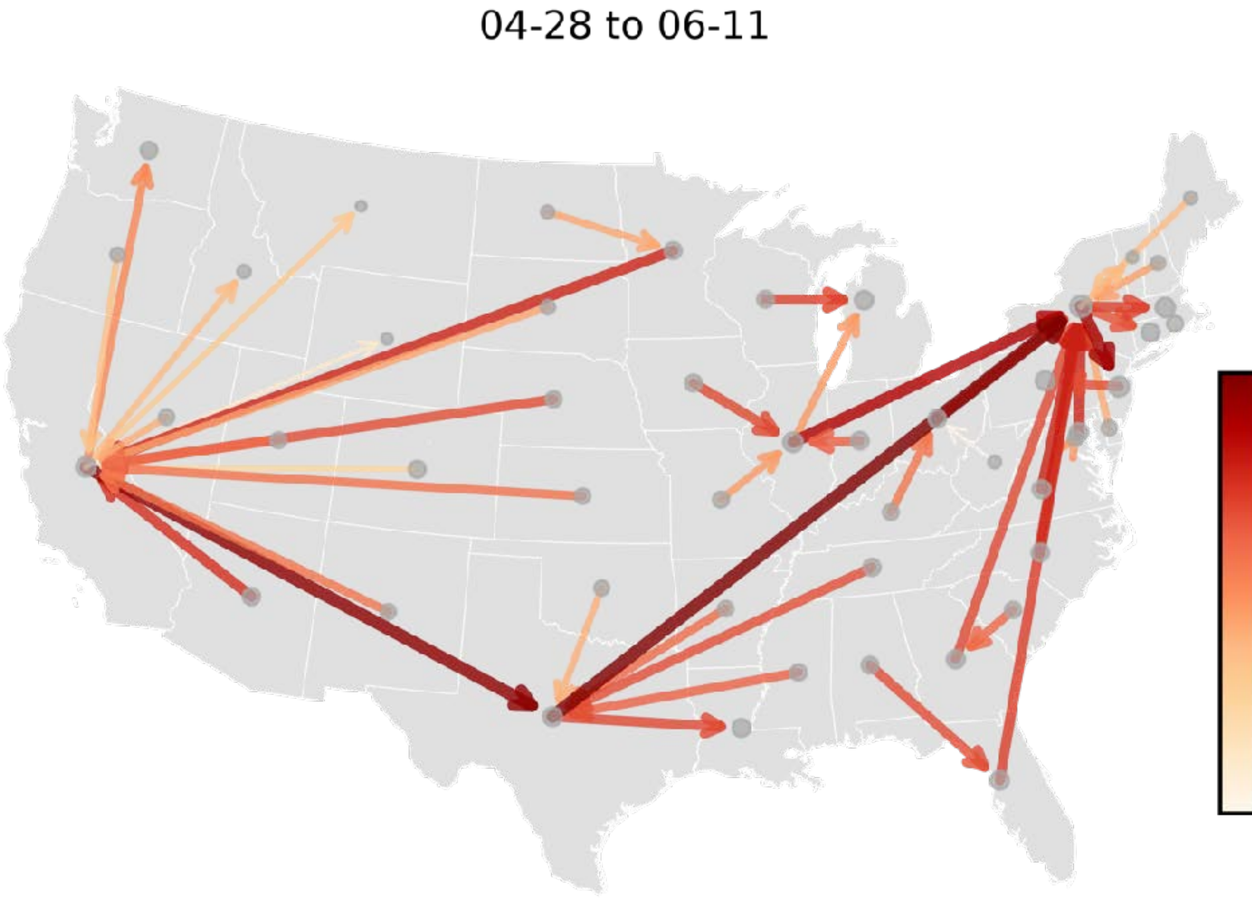


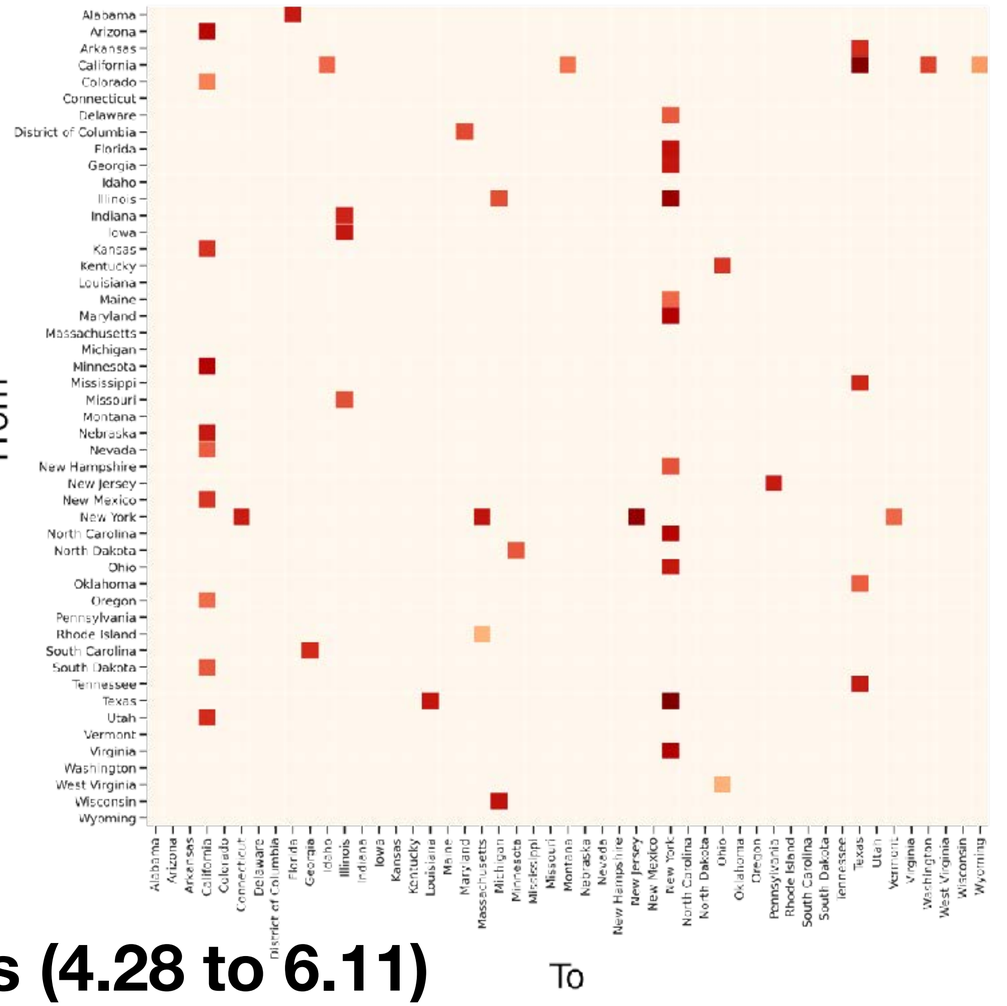


Reaching 1 Million Confirmed Cases (3.31 to 4.28)

1. The shifted risks are even higher
2. California remained as a gravity center for both in-shifts and out-shifts
3. New York started to attracting pandemic shifts from nearby states significantly
4. Stay-at-home order started to lose efficacy as local shifts are stronger than those of P2 (see the colors)

Results (P4)

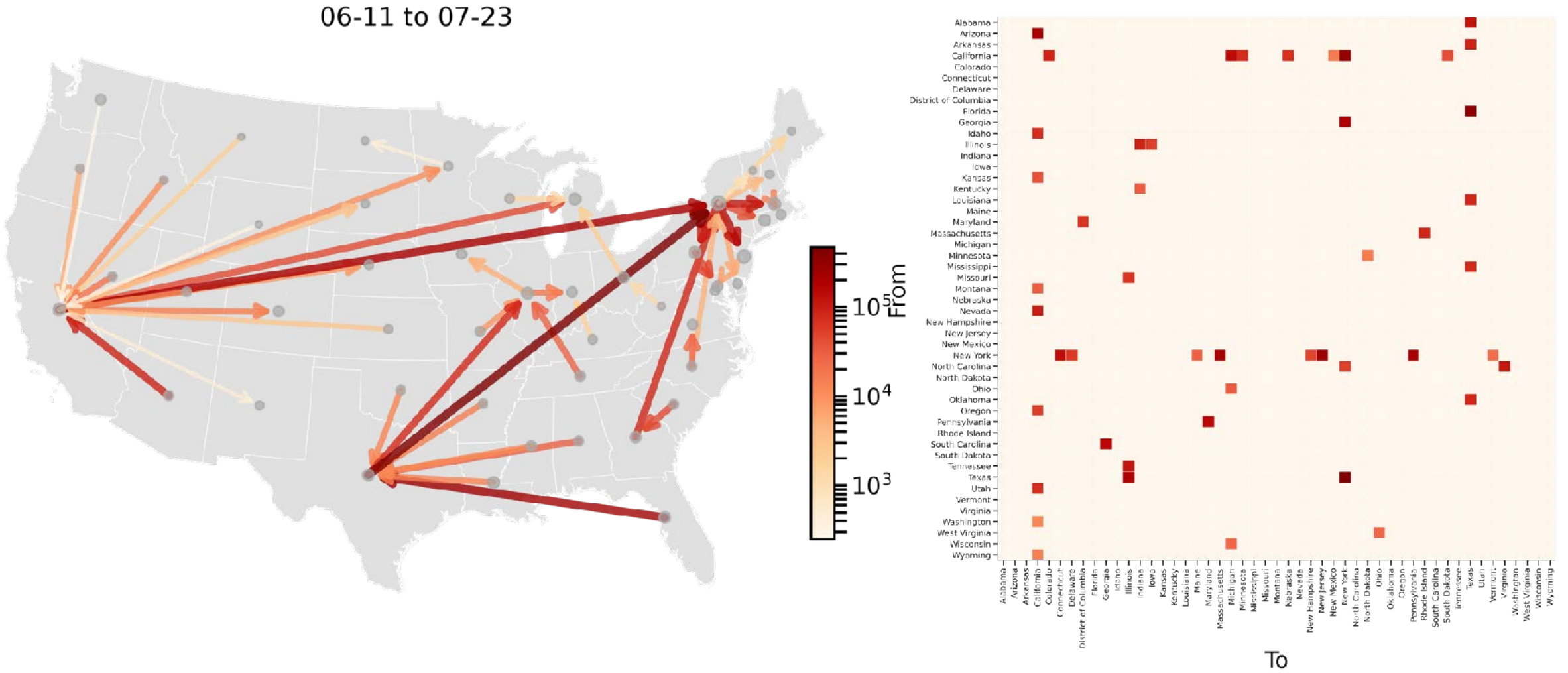




From 1 Million to 2 Million Confirmed Cases (4.28 to 6.11)

1. Strong shifts from New York to New Jersey
2. Many long & strong shifts occurs: lose control of the pandemic? (BLM protests)
3. Back to a free pandemic shifting pattern
4. Texas, Florida Illinois are attracting risks

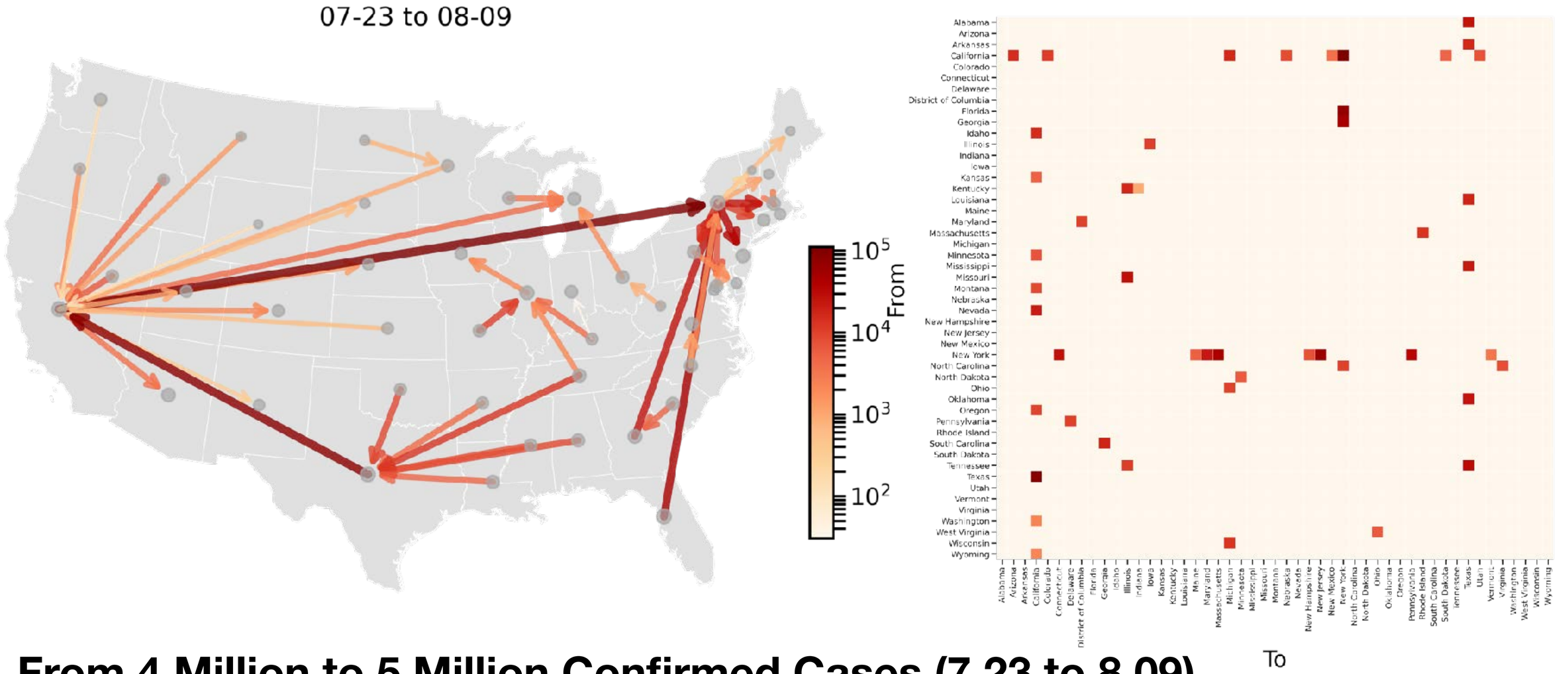
Results (P5)



From 2 Million to 4 Million Confirmed Cases (6.11 to 7.23)

1. Diffusive pattern for New York and California: a sign of deconcentration and further break out
2. Local shifts are more complex than before

Results (P6)



From 4 Million to 5 Million Confirmed Cases (7.23 to 8.09)

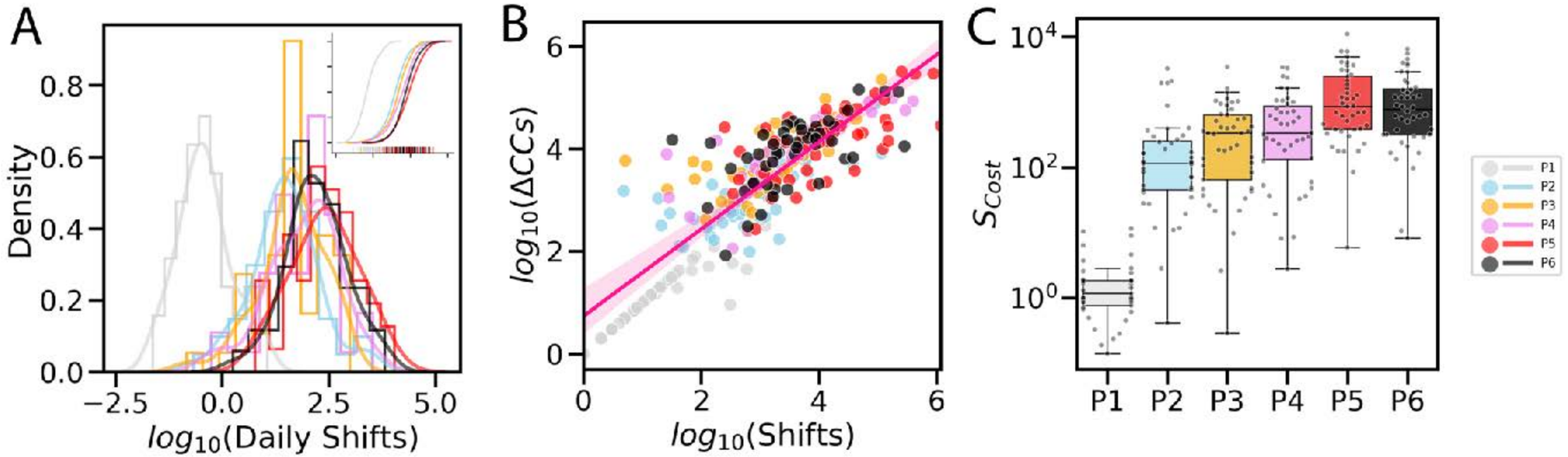
1. Similar pattern to [6.11-7.23]: A stable pattern of pandemic shifting in near future
2. Georgia and Florida emerge in the flow maps of P4, P5 and P6 with strong shifts, implying potential outbreaks in later phases.

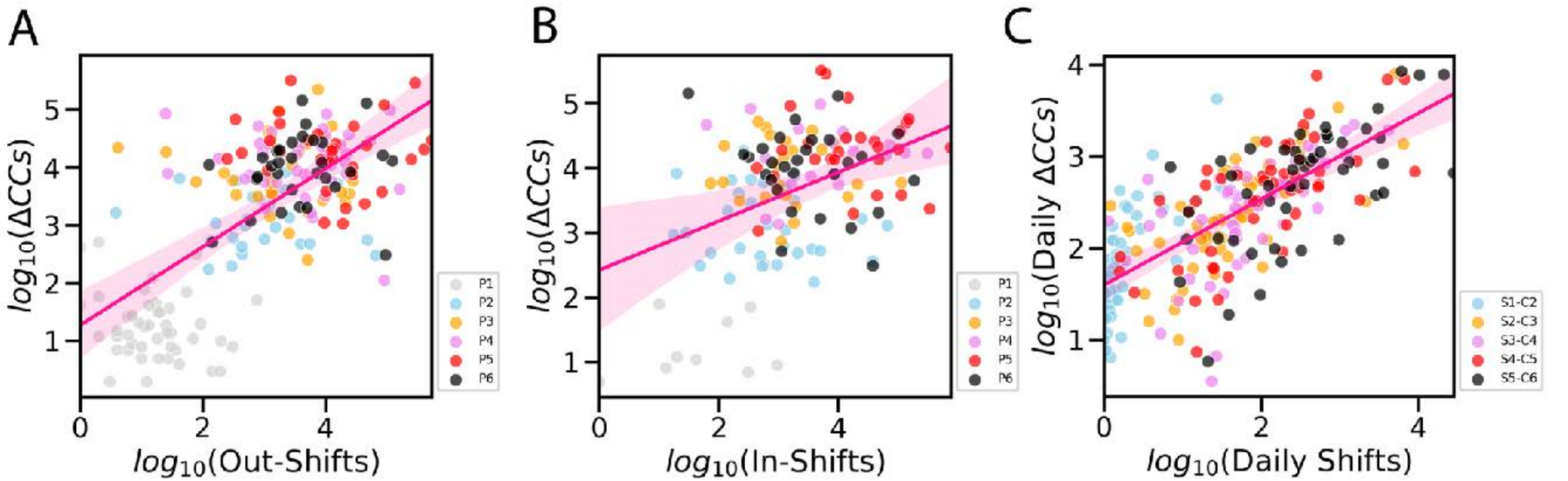
Statistical metrics

- Daily shifts
- $x_{ij}^{*(t)} = x_{ij}^{(t)} / \Gamma^{(t)}$
- the overall intensities of daily shifts are stable

- Shifts v.s. Cases
- Pearson R for total shifts: 0.86
- Out-shifts (0.68); In-shifts (0.42)
- Cross-phase relationships (0.72)

- Daily cost of shifts
- $S_{cost}(ij) = x_{ij}^{*} * c_{ij}$
- A hybrid indicator to measure the severity and complexity of the pandemic
- How strong the shifts are, and how difficult it is for the shifts to occur.



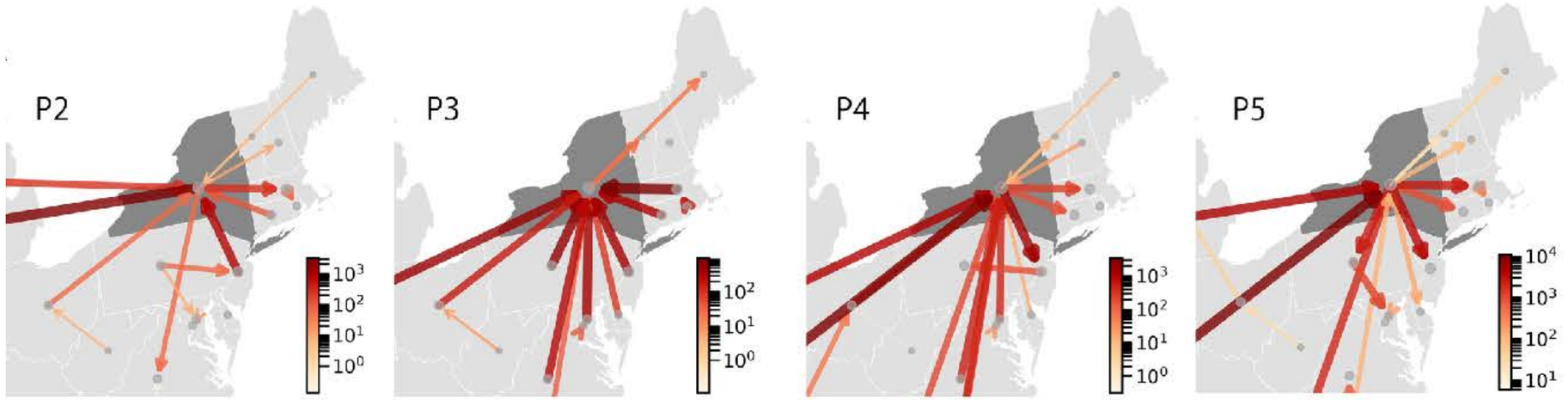


Local shift patterns

UNIVERSITY OF MINNESOTA

Changing New York

1. **From black-hole to volcano**
2. **Long-term risks**



No sign of becoming a dominant pandemic center

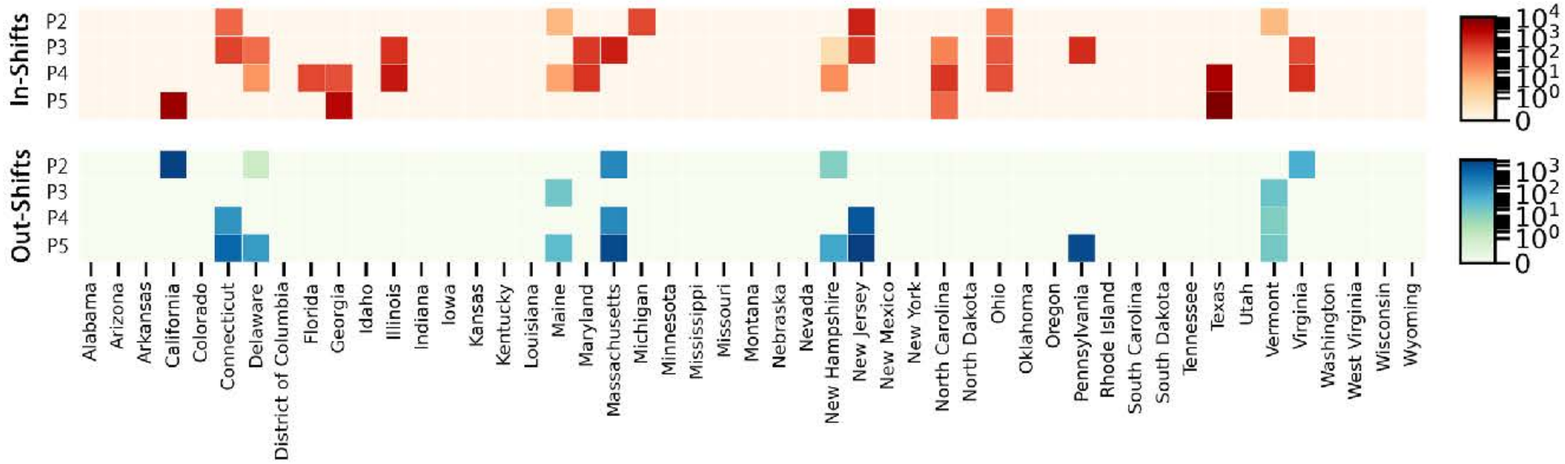
Far more in-shifts than out-shifts : a black-hole pattern

Start to output pandemic risks to nearby states

More out-shifts than in-shifts: a volcano pattern

1/21: first case
1/31: ban travel from China
2/29: first death
3/11: WHO declares pandemic
3/13: national emergency
3/31: stay-at-home order one-by-one for all states
4/28: 1m cases
5/25: Black Lives Matter protests triggered
5/27: 100,000 deaths
6/11: 2m cases
7/23: 4 million cases
7/29: 150000 deaths
8/9: 5m cases

P1 P2 P3 P4 P5 P6



Outline

- Background
- Methodology
- Case 1: Migration flows during Chinese Spring Festival
- Case 2: Spatial shifting patterns of CovID-19 pandemic in the U.S.
- Workflow implementation

https://github.com/dizhu-gis/CovIDSpatialShifts

Introduction to Workflow Tool Knime

UNIVERSITY OF MINNESOTA

Open for Innovation

KNIME

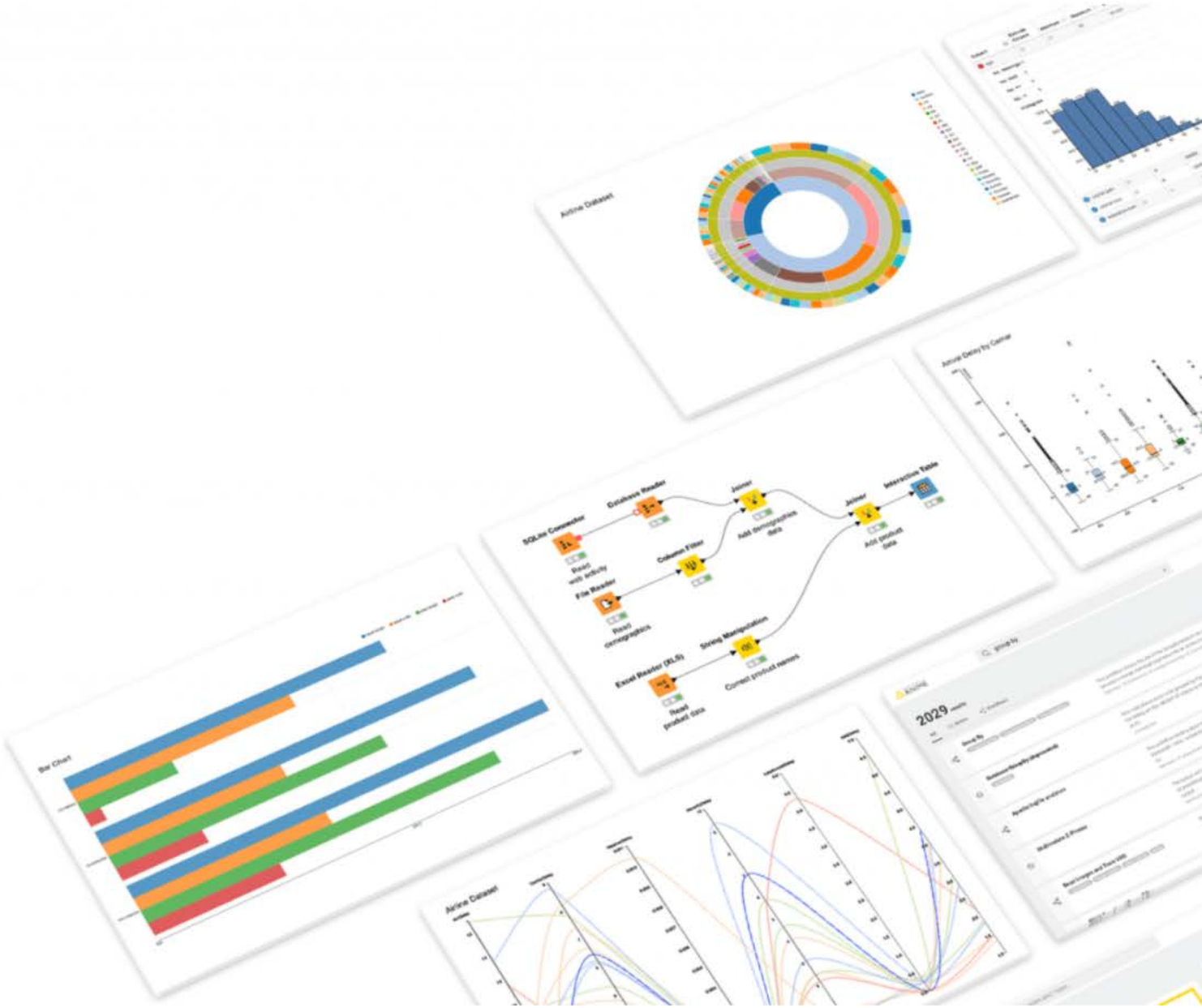
Home

https://www.knime.com/

Ru Wang

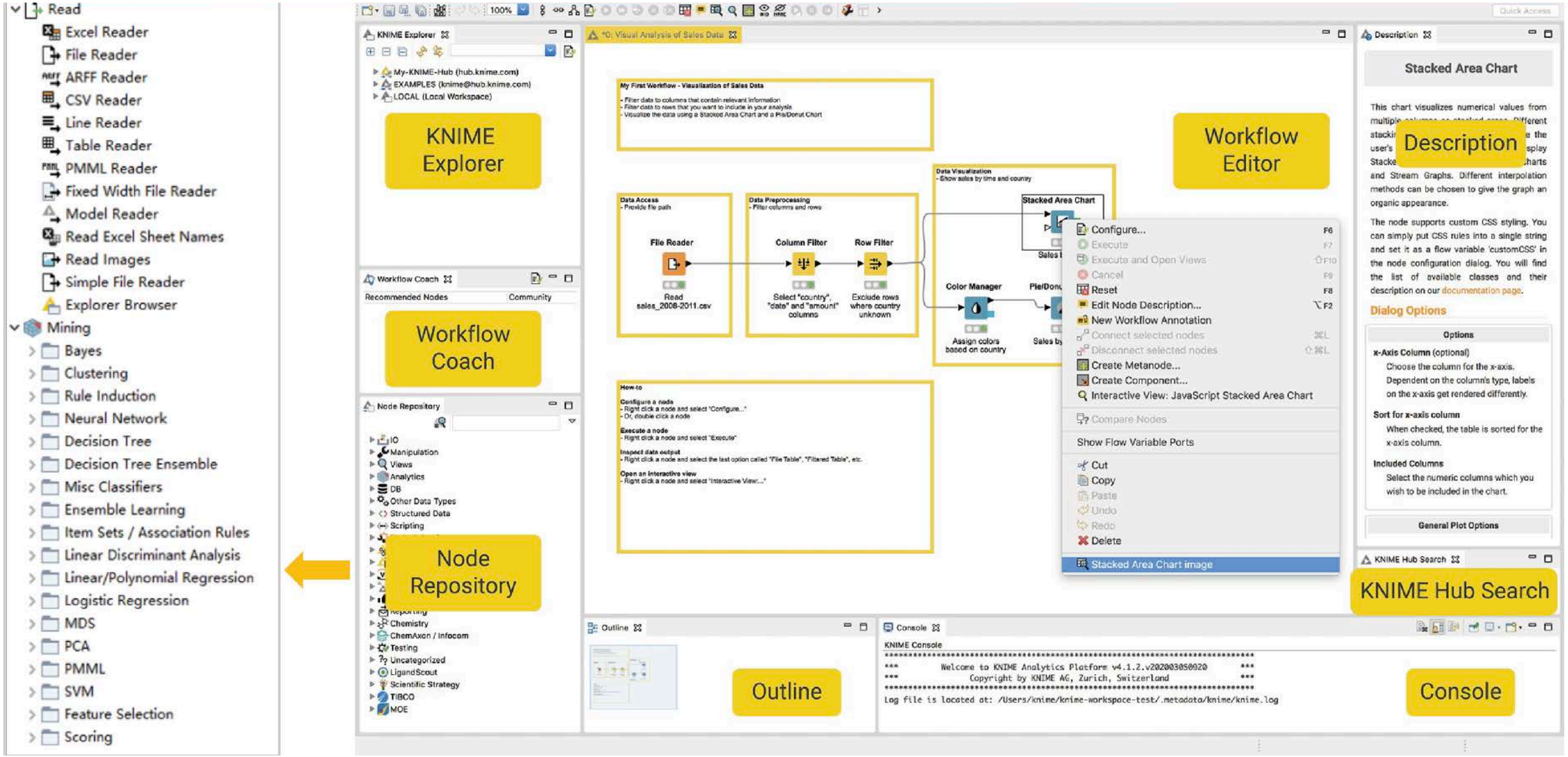
School of Urban Design

Wuhan University



Introduction to Workflow Tool Knime

University of Minnesota



Introduction to Workflow Tool Knime

https://www.knime.com/

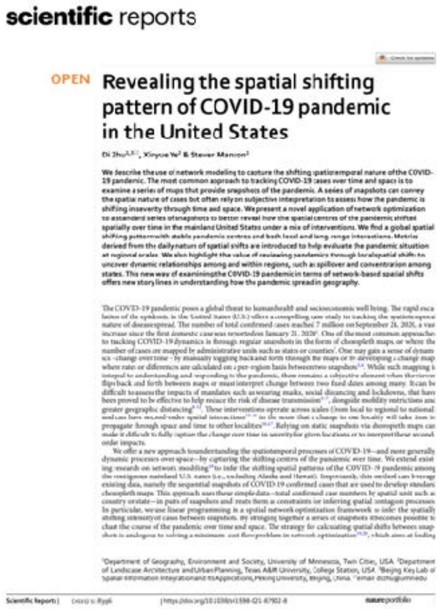
- Workflow is displayed as connected nodes which makes it easy to troubleshoot and visualize
- Easy to use without much knowledge of coding
- Great extensions for data preprocessing, analysis, and visualization
- Connection to other languages, such as JS, R, Python, etc.
- Open-source
- Cross platform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

Paper Replication by Workflow

University of Minnesota

*Zhu D. Ye X. Manson S. Revealing the spatial shifting pattern of COVID-19 pandemic in the United States[J]. **Scientific Reports**, 2021, 11(1): 8396.*

https://github.com/dizhu-gis/CovIDSpatialShifts


scientific reports

OPEN

Revealing the spatial shifting pattern of COVID-19 pandemic in the United States

Di Zhu[1,2*], Xinyue Ye[2] & Steven Manson[3]

We describe the use of network modelling to capture the shifting patio-temporal nature of the COVID-19 pandemic. The most common approach to tracking COVID-19 cases over time and space is to examine a series of maps that provide snapshots of the pandemic. A series of maps can convey the spatial nature of cases but often rely on subjective interpretation to assess how the pandemic is shifting in severity through time and space. We present a novel application of network optimization to standard series of snapshots to better reveal how the spatial centres of the pandemic shifted spatially over time in the mainland United States under a mix of interventions. We find a global spatial shifting pattern with stable pandemic centres and both local and long-range interactions. Metrics derived from the daily nature of spatial shifts are introduced to help evaluate the pandemic situation at regional scales. We also highlight the value of reviewing pandemics through local-spatial shifts to uncover dynamic relationships among and within regions, such as spillover and concentration among states. This new way of examining the COVID-19 pandemic in terms of network-based spatial shifts offers new story lines in understanding how the pandemic spreads in geography.

The COVID-19 pandemic poses a global threat to human health and socioeconomic well being. The rapid escalation of the syndrome, to the United States (U.S.) offers a compelling case study to tracking the spatio-temporal nature of disease spread. The number of total confirmed cases reached 7 million on September 21, 2020, a vast increase since the first domestic case was reported on January 21, 2020[1]. One of the most common approaches to tracking COVID-19 dynamics is through regular snapshots in the form of choropleth maps, or where the number of cases are mapped by administrative units such as states or counties[2]. One may gain a sense of dynamics (change over time) by manually toggling back and forth through the maps or by developing a change map where rates or differences are calculated on a per-region basis between two snapshots[3,4]. While such mapping is integral to understanding and responding to the pandemic, these remains a subjective element when the viewer flips back and forth between maps or must interpret change between two fixed dates among many. It can be difficult to assess the impacts of mandates such as wearing masks, social distancing and lockdowns, that have been proved to be effective to help reduce the risk of disease transmission[5–7], alongside mobility restrictions and greater geographic distancing[8–11]. These interventions operate across scales (from local to regional to national), and can have second-order spatial interactions[12–14] in the sense that a change to one locality will take time to propagate through space and time to other localities[15,16]. Relying on static snapshots via choropleth maps can make it difficult to fully capture the change over time in severity for given locations or to interpret these second-order impacts.

We offer a new approach to understanding the spatiotemporal processes of COVID-19—and more generally dynamic processes over space—by capturing the shifting centres of the pandemic over time. We extend existing research on network modelling[17,18] to infer the shifting spatial patterns of the COVID-19 pandemic among the contiguous mainland U.S. states (i.e., excluding Alaska and Hawaii). Empirically, this method can leverage existing data, namely the sequential snapshots of COVID-19 confirmed cases that are used to develop standard choropleth maps. This approach uses these simple data—total confirmed case numbers by spatial unit such as county or state—in pairs of snapshots and treats them as constraints or inferring spatial contagion processes. In particular, we use linear programming in a spatial network optimization framework in order to infer the spatially shifting intensity of cases between snapshots. By stringing together a series of snapshots it becomes possible to chart the course of the pandemic over time and space. The strategy for calculating spatial shifts between snapshots is analogous to solving a minimum cost flow problem in network optimization[19,20], which aims at finding

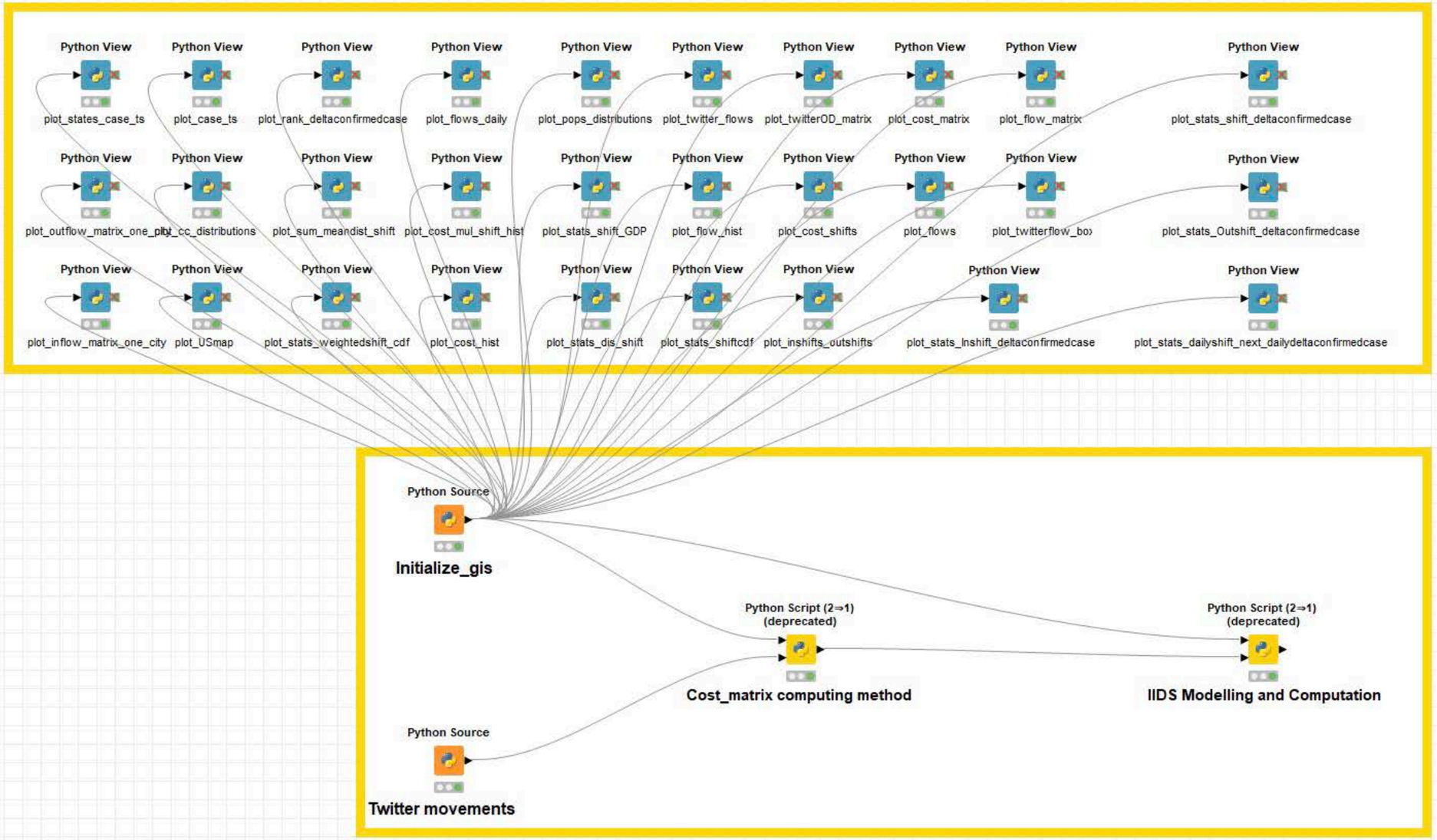
[1]Department of Geography, Environment and Society, University of Minnesota, Twin Cities, USA. [2]Department of Landscape Architecture and Urban Planning, Texas A&M University, College Station, USA. [3]Beijing Key Lab of Spatial Information Integration and Its Applications, Peking University, Beijing, China. *email: dizhu@umn.edu

Scientific Reports | (2021) 11:8396 | https://doi.org/10.1038/s41598-021-87902-8 nature portfolio

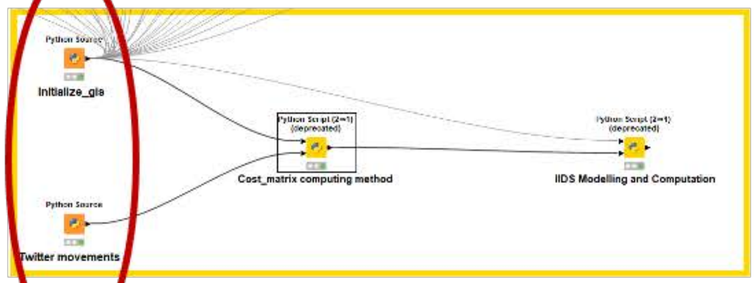
Steps:

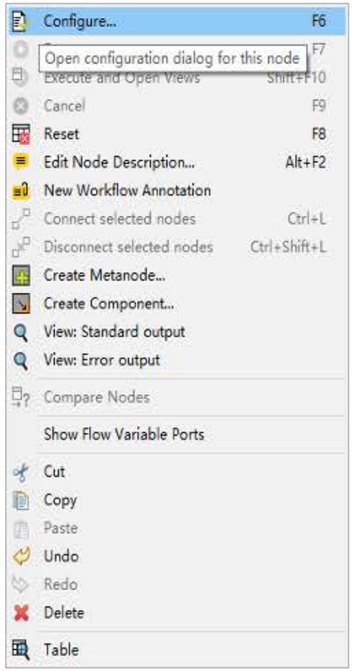
1: Download data
2: Download workflow
3: Open KNIME
4: Import workflow file
5: Set work path
6: Click Run function

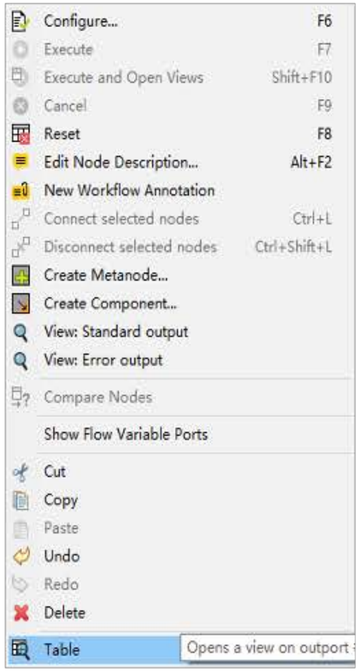
* Install python extension
Conda install packages

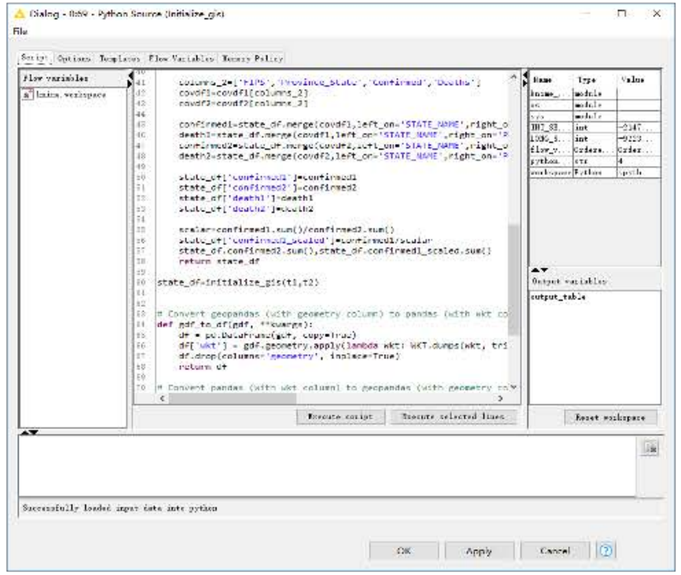


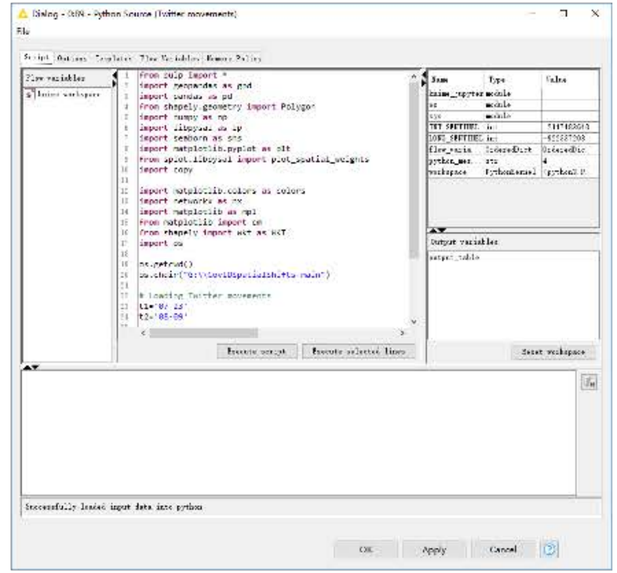
Paper Replication by Workflow

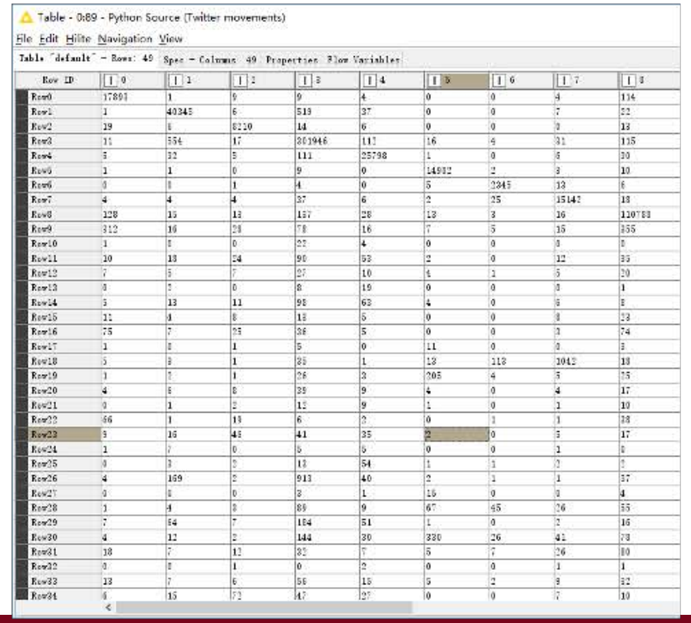


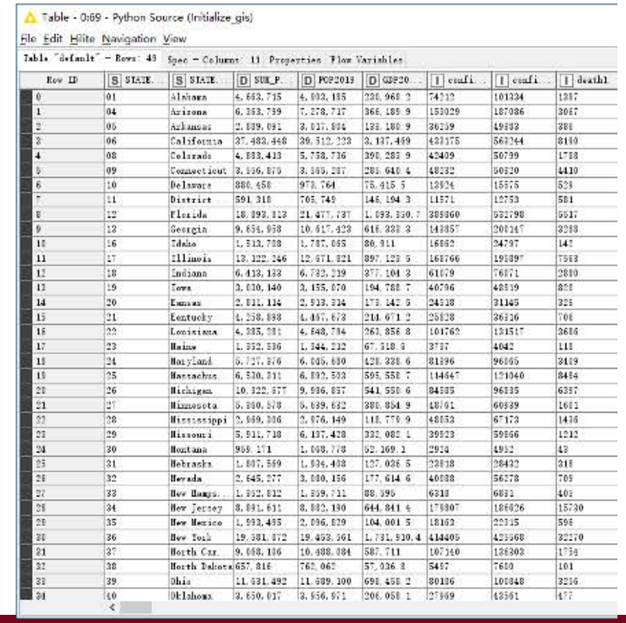






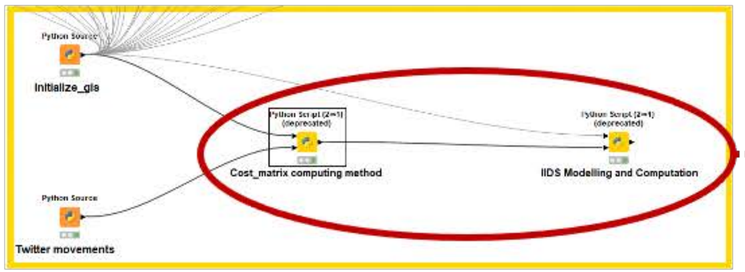


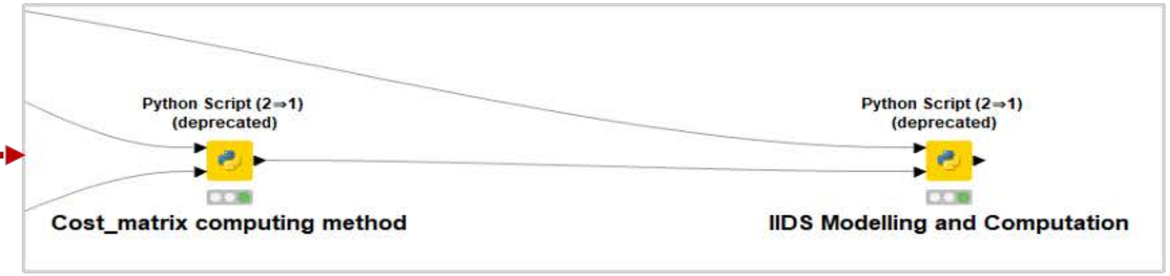


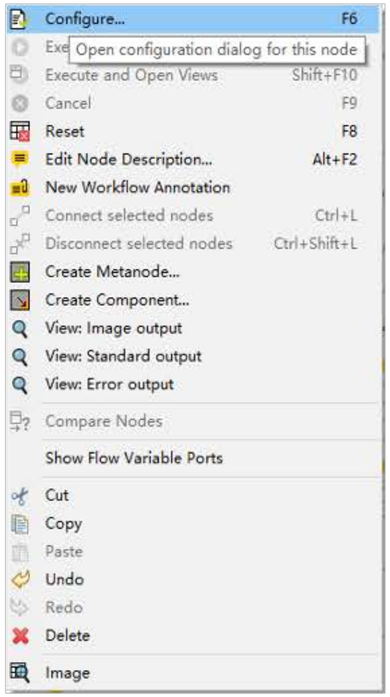


Paper Replication by Workflow

UNIVERSITY OF MINNESOTA







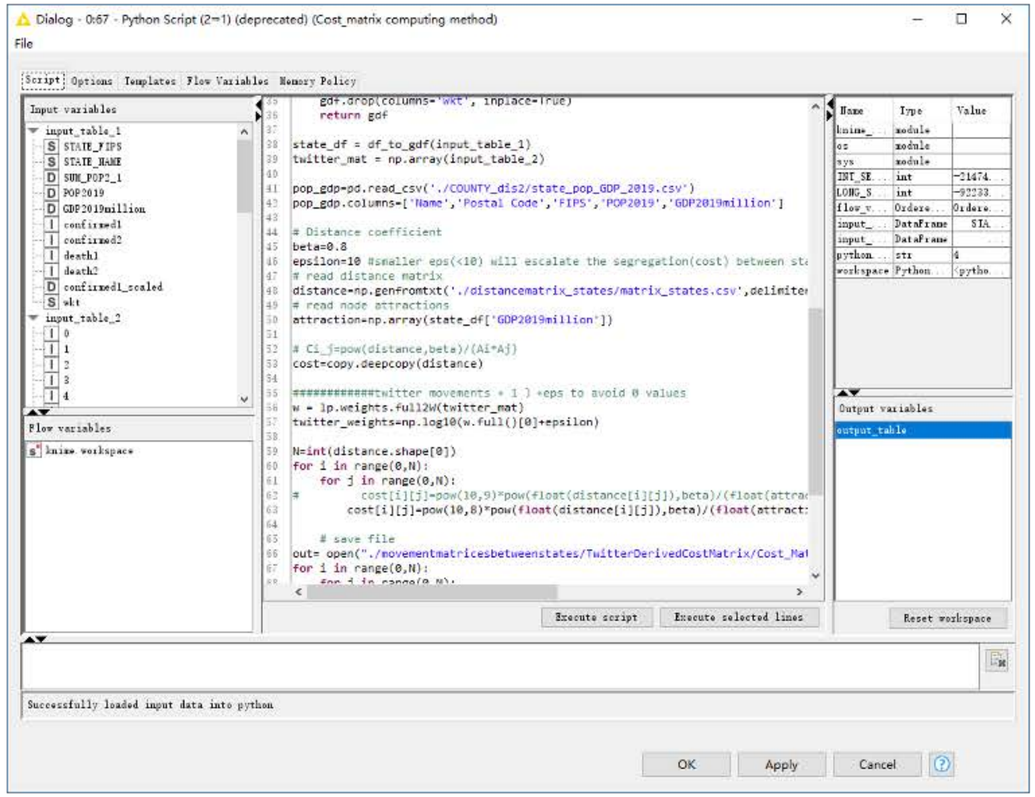


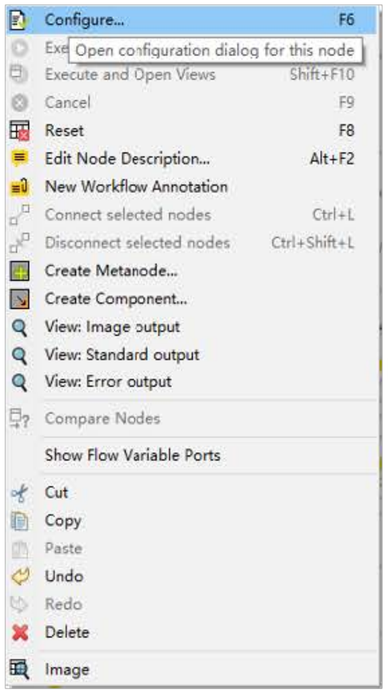
Table - 0:67 - Python Script (2=1) (deprecated) (Cost_matrix computing method)

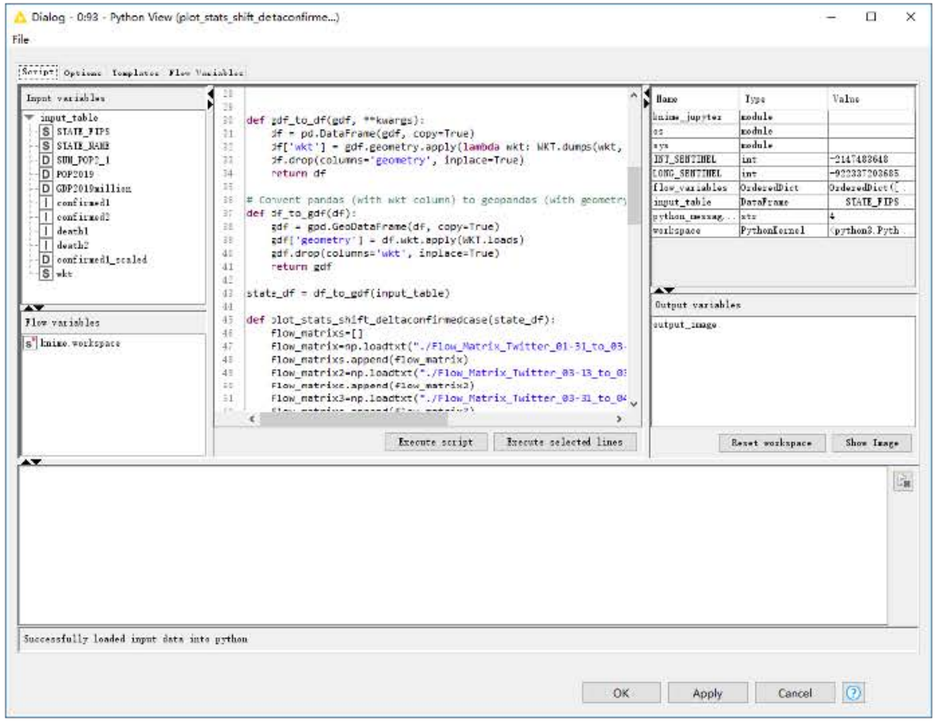
File Edit Hilite Navigation View

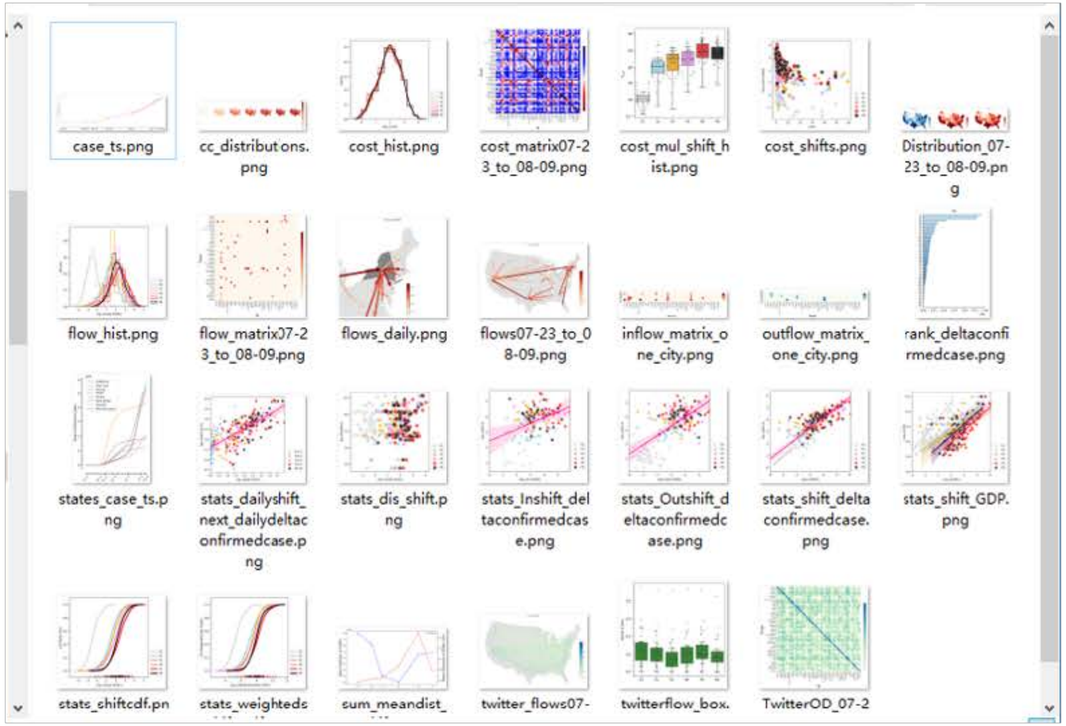
Table "default" – Rows: 49 Spec – Columns: 49 Properties Flow Variables

Row ID	Alabama	Arizona	Arkansas	Calif...	Colorado	Conne...	Delaware	Distr...	Florida	Georgia
Alabama	0	146	92	15	96	135	420	173	6	7
Arizona	146	0	184	1	24	162	586	238	24	44
Arkansas	80	184	0	21	122	255	828	295	30	39
California	15	1	21	0	2	14	64	23	2	4
Colorado	94	25	124	2	0	116	431	177	17	33
Connecticut	120	156	255	16	121	0	116	73	23	40
Delaware	420	586	795	64	431	107	0	83	79	112
District o...	173	255	344	22	177	76	73	0	32	43
Florida	6	25	30	2	17	22	86	32	0	1
Georgia	6	44	39	4	32	36	112	44	1	0
Idaho	719	248	1,104	18	175	688	2,535	1,275	179	286
Illinois	19	24	20	2	12	29	102	35	4	5
Indiana	45	76	67	7	45	58	199	81	11	14
Iowa	165	134	210	15	55	167	583	283	43	62
Kansas	151	101	151	9	36	190	752	304	39	60
Kentucky	65	141	113	15	92	114	344	124	17	19
Louisiana	30	93	58	9	71	145	476	206	9	17
Maine	667	733	1,216	77	552	126	1,015	582	139	227
Maryland	58	89	129	7	69	20	18	1	11	13
Massachusetts	66	74	124	6	54	2	65	40	9	17
Michigan	59	55	83	4	34	38	163	70	12	17
Minnesota	111	76	149	7	37	87	333	159	21	37
Mississippi	56	248	81	29	167	297	922	441	24	35
Missouri	54	61	40	6	31	88	316	128	14	21
Montana	1,086	323	1,661	38	254	1,035	3,810	1,839	272	432
Nebraska	279	174	325	19	51	283	995	471	69	104
Nevada	287	18	446	1	48	312	1,170	586	46	90
New Hampshire	483	535	885	59	382	58	612	360	95	157
New Jersey	50	60	91	4	39	5	14	13	6	10
New Mexico	328	68	442	9	54	469	1,718	789	71	134
New York	18	19	37	1	11	1	13	7	2	3
North Caro...	25	57	63	5	42	27	65	22	4	4
North Dakota	812	512	1,125	63	282	670	2,477	1,189	195	318
Ohio	25	44	52	3	26	22	70	25	5	7
Oklahoma	107	81	69	9	44	198	672	266	27	41

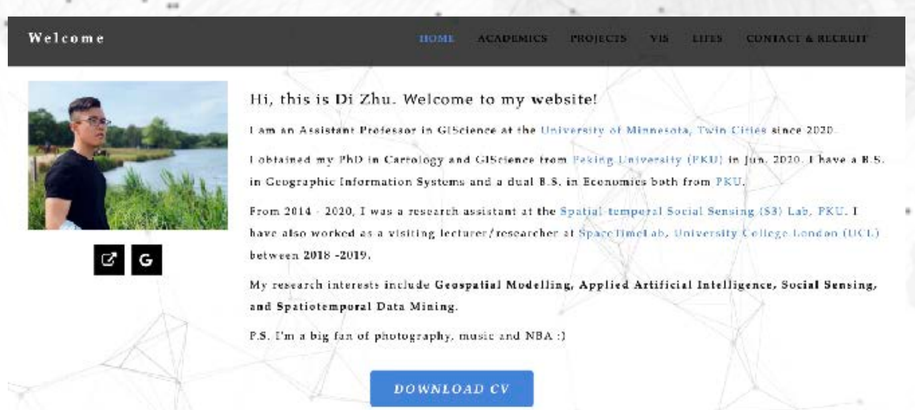
Paper Replication by Workflow







- **Acknowledgements:**
  - New Faculty Set-Up Funding, College of Liberal Art, University of Minnesota;
  - The National Institutes of Health supported Minnesota Population Center (R24 HD041023)
  - The National Spatiotemporal Population Research Infrastructure (2R01HD057929-11)
  - The National Science Foundation for Distinguished Young Scholars of China, no. 41625003;
  - The Major Program of the National Natural Science Foundation of China, no. 41830645;
  - The National Key Research and Development Program of China, no. 2017YFB0503600;



https://cla.umn.edu/geography

2019 U.S. News Global Rank 41
2020 U.S. News Global Rank 47
2020 ARWU Global Rank 40
2021 CWUR Global Rank 47

dizhu@umn.edu
https://dizhu-gis.com

University of Minnesota
Driven to Discover®