Take a pulse of China with social media data

Introduction to Weibo Mobility Index Data

Lingbo Liu[1,2], Wendy Guan[1], Devika Kakkar[1], Shuming Bao[3]

[1] Center for Geographic Analysis, Harvard University
[2] School of Urban Design, Wuhan University
[3] China Data Institute

Contents

A **Social Media Dataset—CGA**
Devika Kakkar

B **Weibo Mobility Index Data**
Lingbo Liu

C **Take Pulse of China with WMI Data**
Shuming Bao

SINA-Weibo
Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Social Media Dataset Project
Harvard CGA

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

CGA's Geotweets Archive

- A global record of tweets spanning time, geography, and language; extending from **2010** to the present; about **10.3 Billion**
- The primary purpose is to make a comprehensive collection of geolocated tweets available to the research community
- Stored on Harvard's High Performance Computing Cluster to provide easy access to both the dataset and processing tools in an integrated environment for Harvard users
- Extremely useful not only for geospatial research but also research involving Social Science, Business and Humanitarian studies
- More details here: https://gis.harvard.edu/geotweet-archive-v20
- To request access to this data: https://gis.harvard.edu/geotweet-request-form

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Enrichment of the Geotweets Archive

- Harvard CGA and MIT SUL have collaborated to conduct research on the use of social media data to study the effects of climate change on people's well being
- To achieve this objective every tweet in was enriched with two important variables: **Sentiment** and **Geography**
- Advanced GIS Data Science (GPU database OmniSci) and Machine Learning techniques were used on Harvard's FAS Research Computing Cluster for processing
- First time enrichment has been achieved on this large scale involving about 10 Billion tweets
- Useful for a wide range of research applications involving the use of Sentiment and Geography on Twitter data

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Future Plans: Real-time Collection and Processing of Weibo Data

- Build a high-performance system for real-time collection and processing of data from Sina-Weibo, second largest social media platform in China

Objectives of this projects are:

- Develop a comprehensive global social media data archive using Twitter and Sina-Weibo; fill the spatial gap in our archive
- Develop a High-Performance Computing environment to enhance the processing capabilities for big geospatial datasets
- Develop social media based case studies to explore ways in which this data can be applied to policy making, event response and global comparison studies
- Develop training programs for using data science and high-performance computing for handling big geospatial datasets

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Weibo User Historical Geotagged Posts Dataset

- MIT SUL has constructed a unique dataset which contains geotagged posts from Sina-Weibo, the biggest social media platform in China
- This panel data contains 9.95 million posts generated by a cohort of 447 thousand active users
- 3 years of data ranging from January 1st 2018 and June 30 2021
- More information of this dataset can be found here
- Researchers may request the access to the dataset via filling out this contact form

B

Weibo Mobility Index Data

Why Weibo Mobility Index?

- An open index for multi-scale mobility in China
- Baidu Qianxi Index （discontinuous time span）
- Combined with Tweet for global Mobility Index

SINA-Weibo
Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

SINA-Weibo
Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Contents

1 Data source

2 Methodology

3 Comparison

4 Applications

5 Conclusion

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Background

SINA Weibo

China's second most popular social media platform

User

Content

Photos

Response

Sample page of Weibo

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

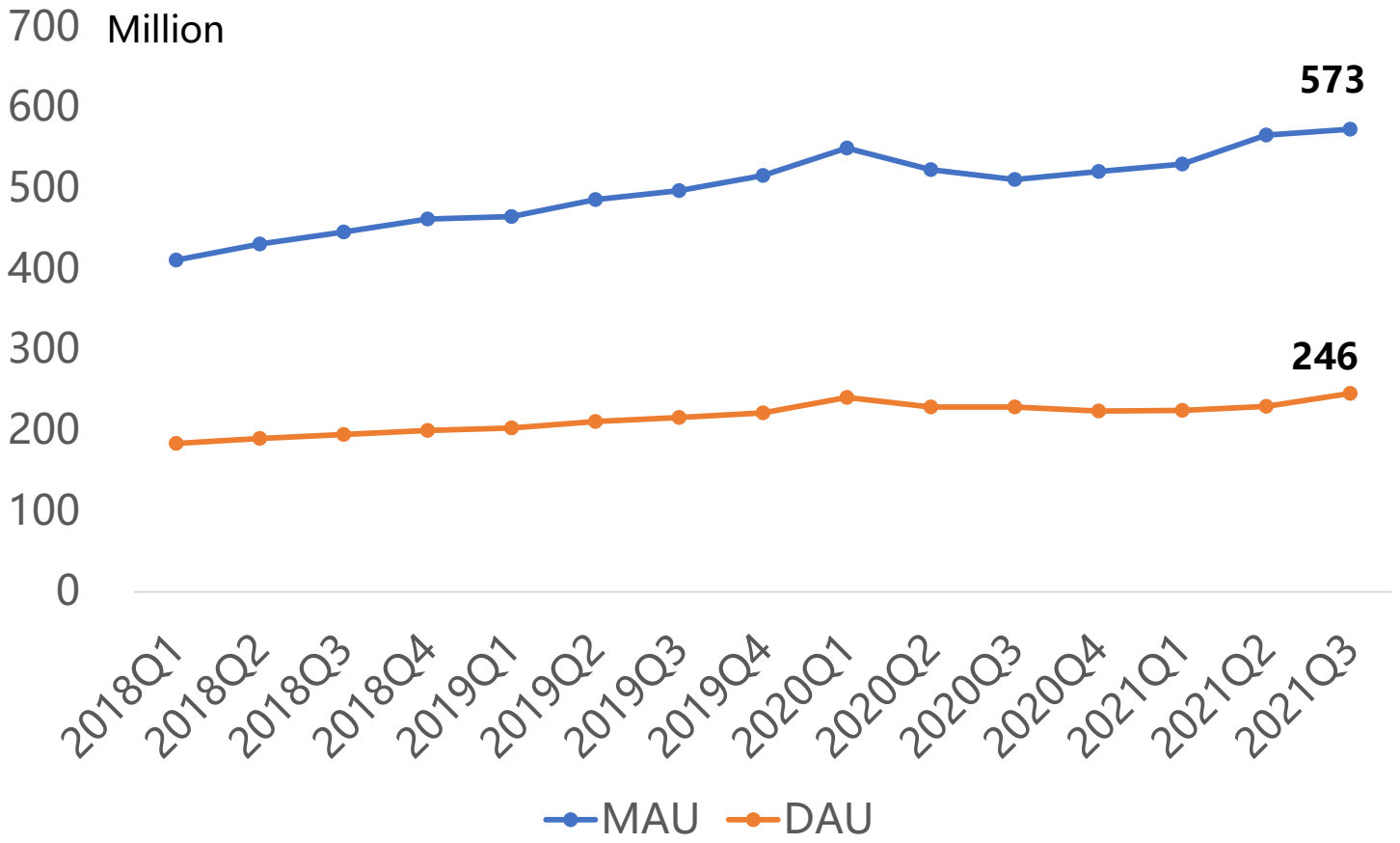
Center for Geographic Analysis
Harvard University

Background

SINA Weibo

Monthly Active User(MAU)

Daily Active User(DAU)



Twitter

DAU in Millions

211

US

International

Data Source

https://www.statista.com/statistics/1058070/china-sina-weibo-dau/
https://www.statista.com/statistics/795303/china-mau-of-sina-weibo/
https://www.socialmediatoday.com/news/twitter-rises-to-211-million-active-users-though-longer-term-growth-target/608958/

1

Data source

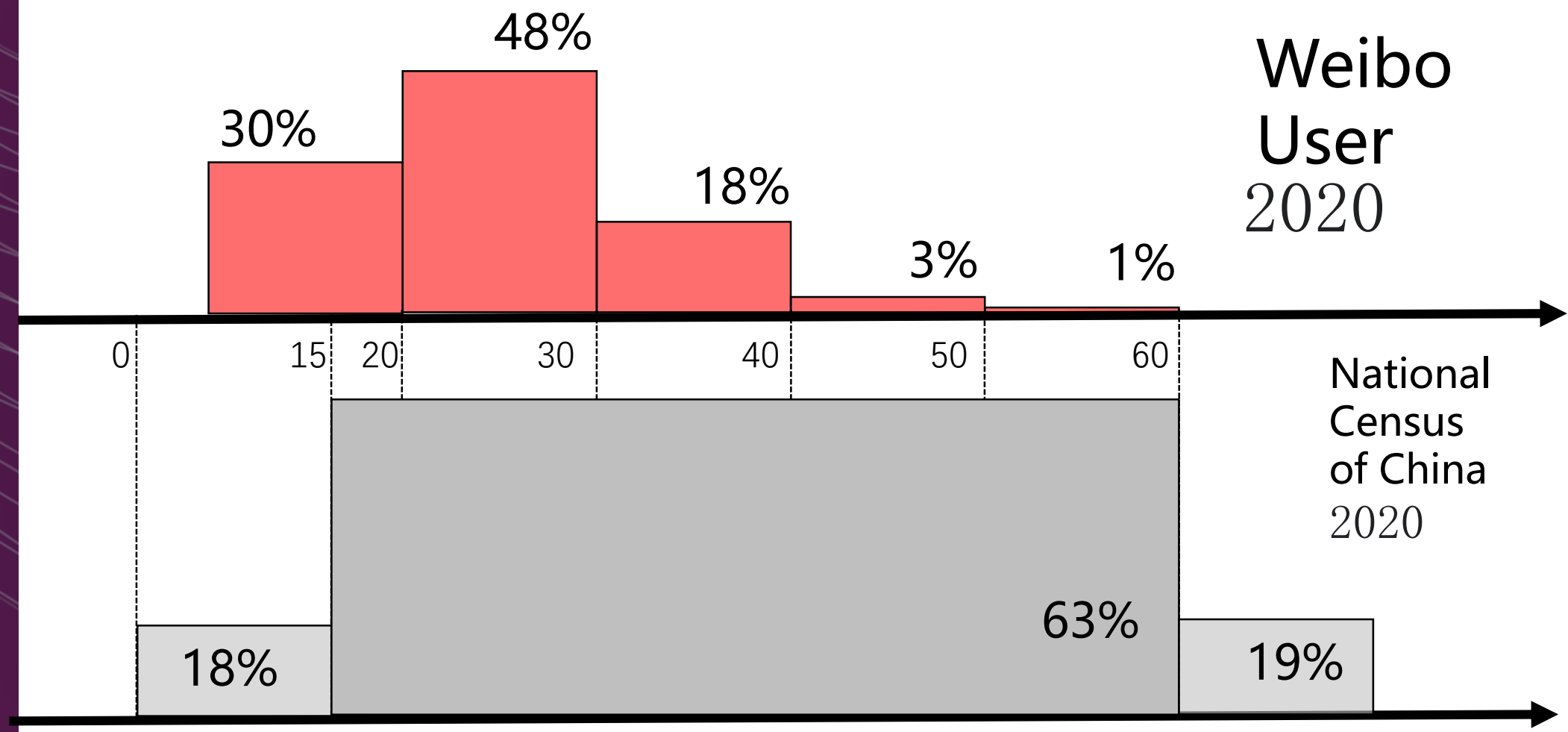
SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Background

SINA Weibo
Proportion of Population By Age



Data Source

http://www.stats.gov.cn/tjsj/zxfb/202105/t20210510_1817181.html
https://data.weibo.com/report/reportDetail?id=456

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Geotagged SINA Weibo Data: the Chines Twitter

Data Source： Sina Weibo (http://Weibo.com)

Data Collection: Harvard FASRC

Raw Data：https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8FOPNO

Covid19 Data：https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DULFFJ

Weibo mobility index：https://dataverse.harvard.edu/dataverse/weibomobilityindex

Time: 2018-01-01, 2021-07-21

Year	2018	2019	2020
Number of Geotag-Weibo Tweets	2,247,163	2,468,591	2,247,983
Number of Unique User ID	226,422	309,935	238,649
Monthly Active Users(MAUs)	462M	516M	521M
Daily Active Users (DAUs)	200M	222M	225M
Total Population in China	1.403B	1.408B	1.411B
Ratio of Sample Users to DAU	1.132‰	1.396‰	1.061‰
Ratio of Sample Users to MAU	0.490‰	0.601‰	0.458‰
Ratio of Sample Users to Total Population	0.161‰	0.219‰	0.169‰

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Geotagged SINA Weibo Tweets

Sample Data Structure

Fields	Contents
userid	Unique ID for Users
mblogid	Unique ID for Weibo tweets
Created_at	Time for posting Weibo tweets
poiid	Tweets with POI information
source	Device for posting Weibo tweets
emoji	Emoji in Text
Text	Text content
Tag	Hashtag
Latitude	Y coordinates WGS 84
Longitude	X coordinates WGS 84

L userid	L mblogid	S bid	S created_at	S city	S poiid	S location	S source	S text	S emoji	S tag	S ot...	S isLon...	S poi_type	S poi_city	S page_...	D latitude	D longi...
1074895510	4191451673790...	FCn4HfUq6	2018-01-01T18:43:35	北京	8008611010900000000	北京·潭柘寺	微博视频号	2018年新年...	['给力', '心']	[]	[]	False	行政单元	北京	门头沟区	39.937	116.105
1683696901	4201450554490...	G0BJBiQ7s	2018-01-29T08:55:35	北京	B2094557DA69A7FD...	北京·BAD...	iPhone客户端	bad famers ...	[]	['#脏脏包#']	[]	False	?	北京	BAD FARME...	39.939	116.454
1811641001	4192546378867...	FCPylBb5P	2018-01-04T19:13:33	北京	8008611010500000000	北京·潘...	HUAWEI P1...	心情不爽到...	[]	[]	[]	False	行政单元	北京	朝阳区	39.921	116.486
1811641001	4191313735870...	FCjudoEYa	2018-01-01T09:35:28	北京	8008611010500000000	北京·双井	HUAWEI P1...	不念过往，...	[]	[]	[]	False	行政单元	北京	朝阳区	39.921	116.486
2479274000	4194412801960...	FDC6I8eKI	2018-01-09T22:50:03	北京	B2094654D66DA7F5...	北京·军...	iPhone客户端	估计某人这...	[]	[]	[]	False	一般地点标志	北京	军都山滑雪场	40.234	116.331
2479274000	4194408666200...	FDC02q2cE	2018-01-09T22:33:37	北京	8008611011404000000	北京·崔村镇	iPhone客户端	即使是摔倒...	['嘻嘻']	[]	[]	False	行政单元	北京	崔村镇	40.245	116.337
2420524011	4198758415460...	FFr9LmVyE	2018-01-21T22:37:58	北京	8008611000000000000	北京	红米4	又回来了 北京	[]	[]	[]	False	行政单元	北京	北京	39.905	116.405
3717015912	4199140562610...	FFB68aYmJ	2018-01-22T23:56:29	北京	B2094654D06BA4FA...	北京·世...	三星GALAXY A	北京·世奥国际	['吃瓜']	[]	[]	False	楼宇	北京	世奥国际	39.995	116.438
3717015912	4198610218560...	FFniJzWAN	2018-01-21T12:49:05	北京	B2094654D06BA4FA...	北京·世...	三星GALAXY A	继前任3后，...	['吃瓜', '顶']	[]	[]	False	楼宇	北京	世奥国际	39.995	116.438
3717015912	4198416382790...	FFig6bIr8	2018-01-20T23:58:51	北京	B2094654D06BA4FA...	北京·世...	三星GALAXY A	淘宝是个让...	['并不简单']	[]	[]	False	楼宇	北京	世奥国际	39.995	116.438
3717015912	4198371687700...	FFh60wjNS	2018-01-20T21:01:15	北京	B2094554D764A0F9...	北京·檐...	三星GALAXY A	北京·檐下...	['哈欠']	[]	[]	False	中餐厅	北京	檐下风铃酒馆	39.934	116.409
2087030813	4194679061170...	FDJ2a4Wyf	2018-01-10T16:28:04	北京	B2094757D06FA2FC...	北京·北...	?回来的iP...	#全世界最好...	['心']	['#全世界...	[]	False	五星级宾馆	北京	北京王府...	39.915	116.416

1

Data source

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Geotagged SINA Weibo Tweets

Base Maps:

- 2010 Province boundary Map (34)
- 2000-2010 Prefecture city boundary map (342)

Liuzhou Diqu; Direct Units （Hubei, Xinjiang, Hainan）

Chaohu (Anhui)； Laiwu(Shandong)

2

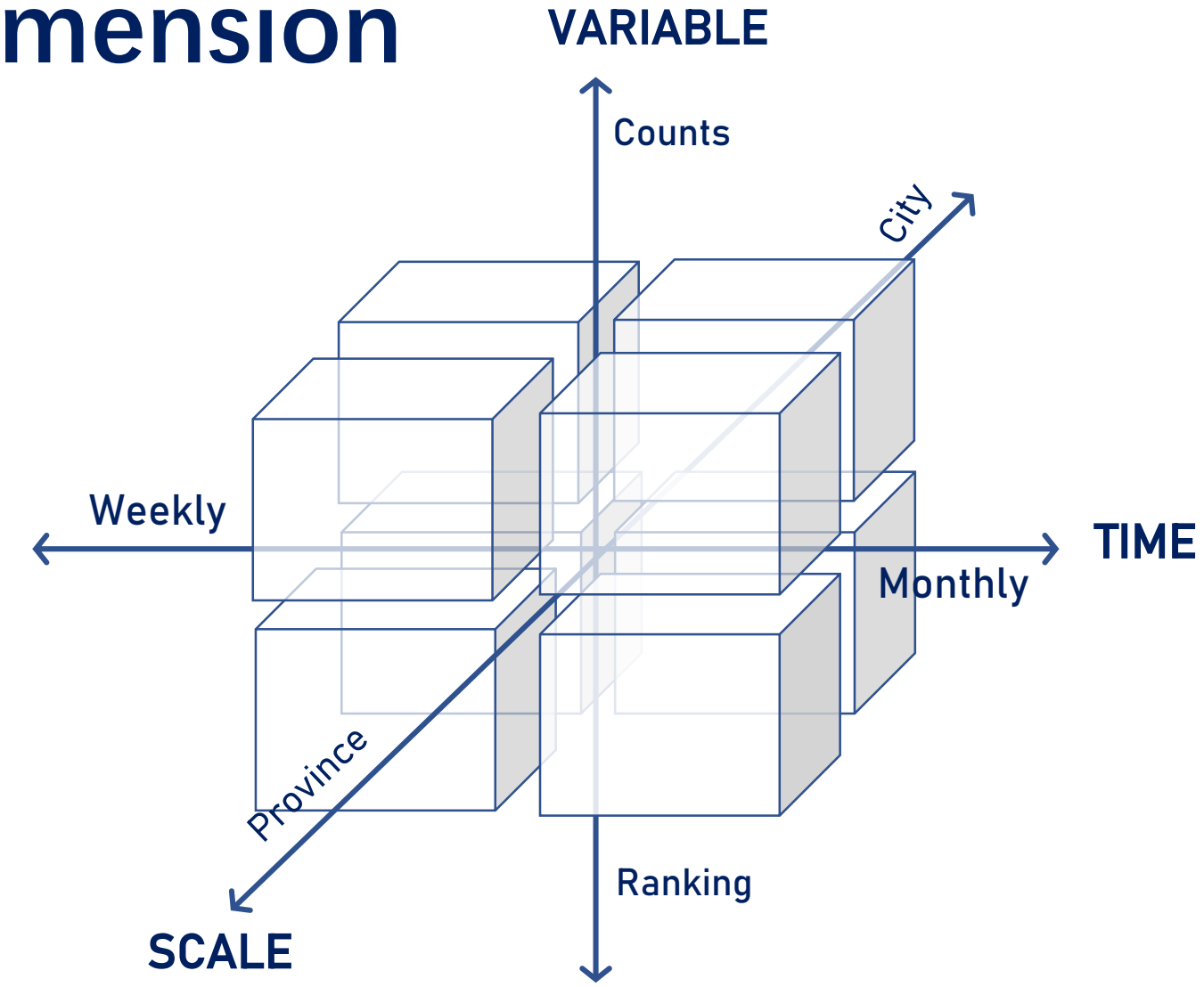
Methodology

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Index Dimension



2

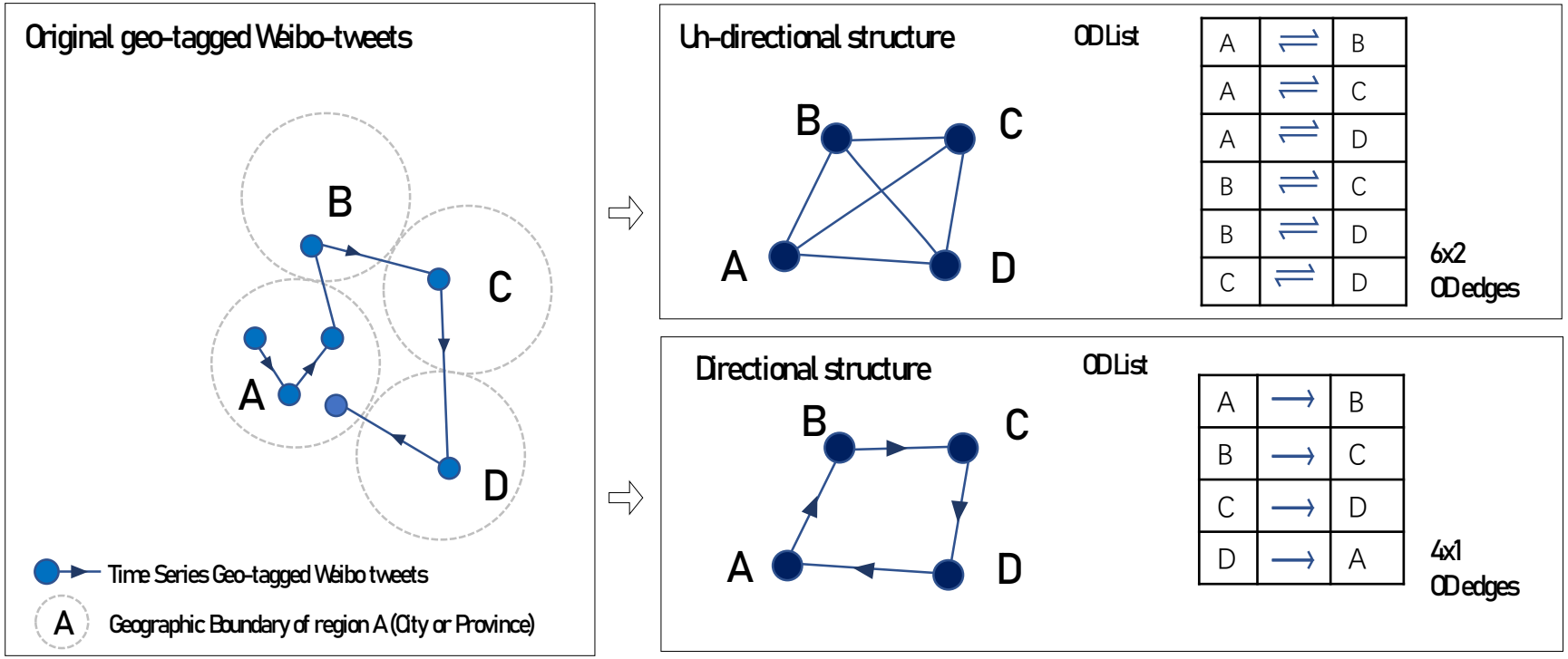
Methodology

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Flow Aggregate



2

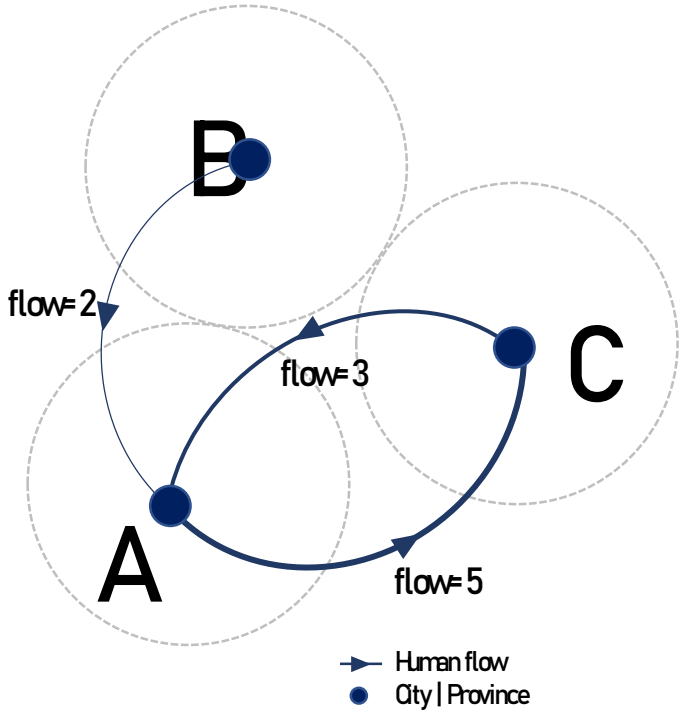
Methodology

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Index structure



(a) Example of Inter-regional human flow

Inflow index	Total Inflow	Portion of Origin		
		A	B	C
A	5	0	0	100%
B	0	0	0	0
C	5	40%	60%	0

Outflow index	Total Outflow	Portion of Destination		
		A	B	C
A	5	0	0	100%
B	2	100%	0	0
C	3	100%	0	0

Total flow Flow matrix

(b) Data Structure of Baidu Qianxi Index

Inflow	Scale
C← A	5
A← C	3
A← B	2

Outflow	Scale
A→ C	5
C→ A	3
B→ A	2

(c) OD List

2

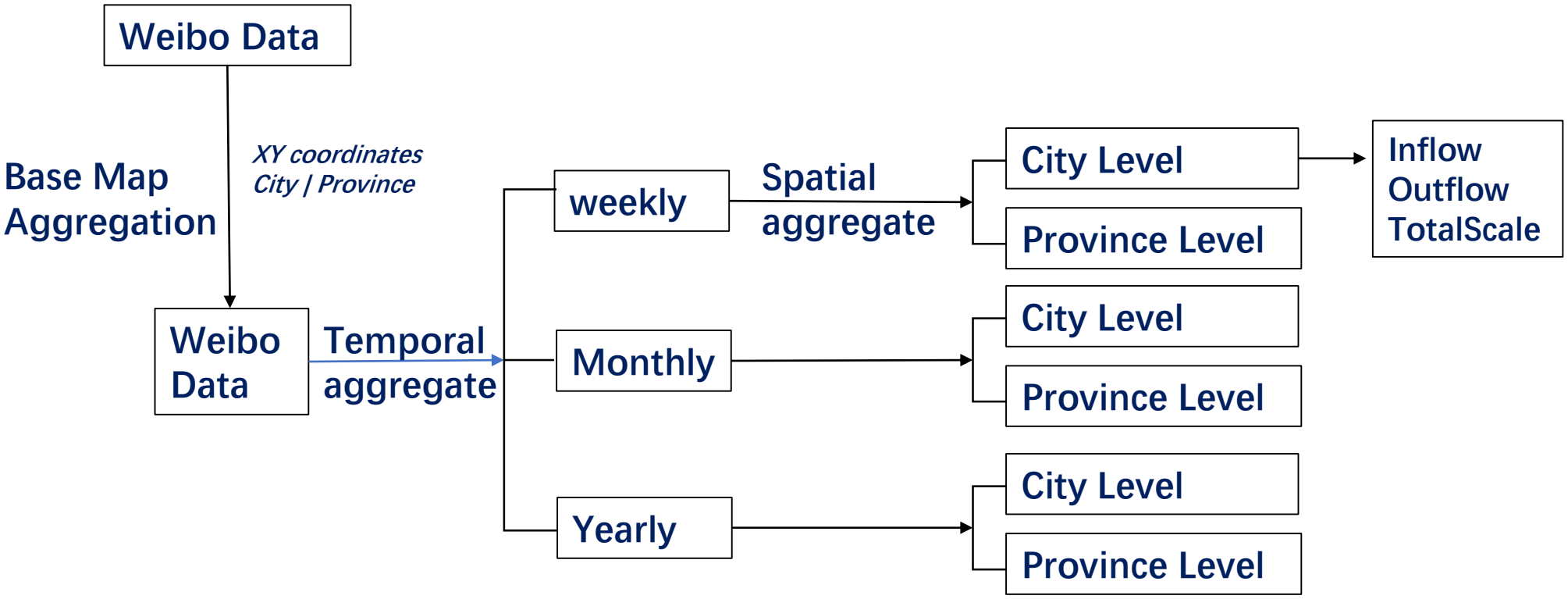
Methodology

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Data Preprocessing Flowchart



Data Preprocessing with KNIME

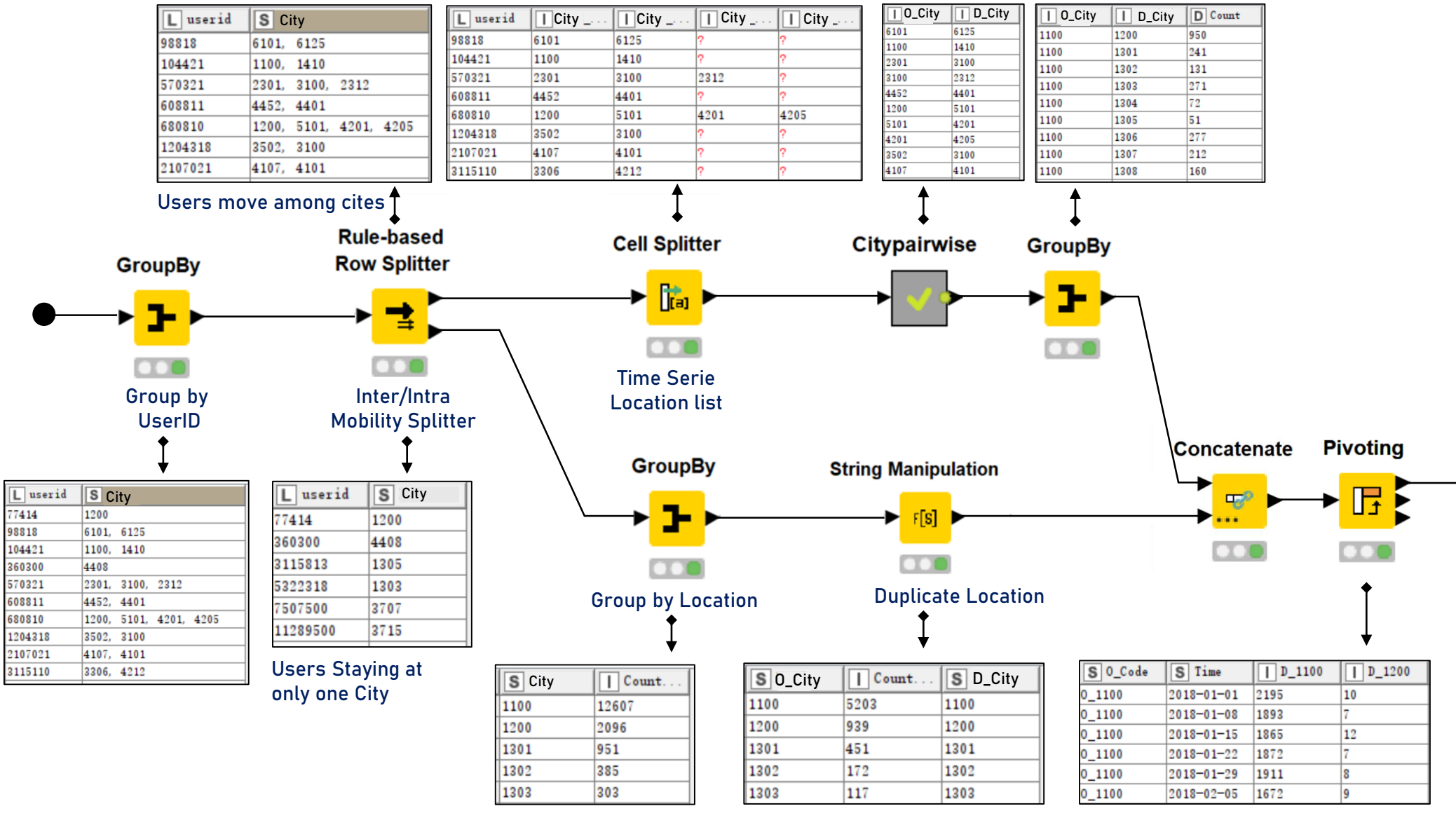
2

Methodology

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University



userid	City
98818	6101, 6125
104421	1100, 1410
570321	2301, 3100, 2312
608811	4452, 4401
680810	1200, 5101, 4201, 4205
1204318	3502, 3100
2107021	4107, 4101

userid	City _	City _	City _	City _
98818	6101	6125	?	?
104421	1100	1410	?	?
570321	2301	3100	2312	?
608811	4452	4401	?	?
680810	1200	5101	4201	4205
1204318	3502	3100	?	?
2107021	4107	4101	?	?
3115110	3306	4212	?	?

O_City	D_City
6101	6125
1100	1410
2301	3100
3100	2312
4452	4401
1200	5101
5101	4201
4201	4205
3502	3100
4107	4101

O_City	D_City	Count
1100	1200	950
1100	1301	241
1100	1302	131
1100	1303	271
1100	1304	72
1100	1305	51
1100	1306	277
1100	1307	212
1100	1308	160

userid	City
77414	1200
98818	6101, 6125
104421	1100, 1410
360300	4408
570321	2301, 3100, 2312
608811	4452, 4401
680810	1200, 5101, 4201, 4205
1204318	3502, 3100
2107021	4107, 4101
3115110	3306, 4212

userid	City
77414	1200
360300	4408
3115813	1305
5322318	1303
7507500	3707
11289500	3715

City	Count
1100	12607
1200	2096
1301	951
1302	385
1303	303

O_City	Count	D_City
1100	5203	1100
1200	939	1200
1301	451	1301
1302	172	1302
1303	117	1303

O_Code	Time	D_1100	D_1200
0_1100	2018-01-01	2195	10
0_1100	2018-01-08	1893	7
0_1100	2018-01-15	1865	12
0_1100	2018-01-22	1872	7
0_1100	2018-01-29	1911	8
0_1100	2018-02-05	1672	9

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

20210201-20210630

Geo-tagged

Weibo Data

User Trajectory Mapping

Daily Weekly Monthly

GUI

Harvard CGA

Global Urban Impulse Index

Daily

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

2021-02-01

GUI

Harvard CGA

Global Urban Impulse Index

Weekly

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

Week00

GUI

Harvard CGA

Global Urban Impulse Index

Monthly

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

202102

3

Comparison

SINA-Weibo
Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

20210201-20210630

Geo-tagged

Weibo Data

Aggregated Human Flow

Daily Weekly Monthly

GUI

Harvard CGA

Global Urban Impulse Index

Monthly

PROVINCE-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

202102

GUI

Harvard CGA

Global Urban Impulse Index

Weekly

PROVINCE-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

Week00

GUI

Harvard CGA

Global Urban Impulse Index

Weekly

CITY-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

Week00

GUI

Harvard CGA

Global Urban Impulse Index

Monthly

CITY-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Weibo Data

20210201-20210630

202102

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

20210201-20210630

Geo-tagged

Baidu Data

Aggregated Human Flow

Daily Weekly Monthly City Province

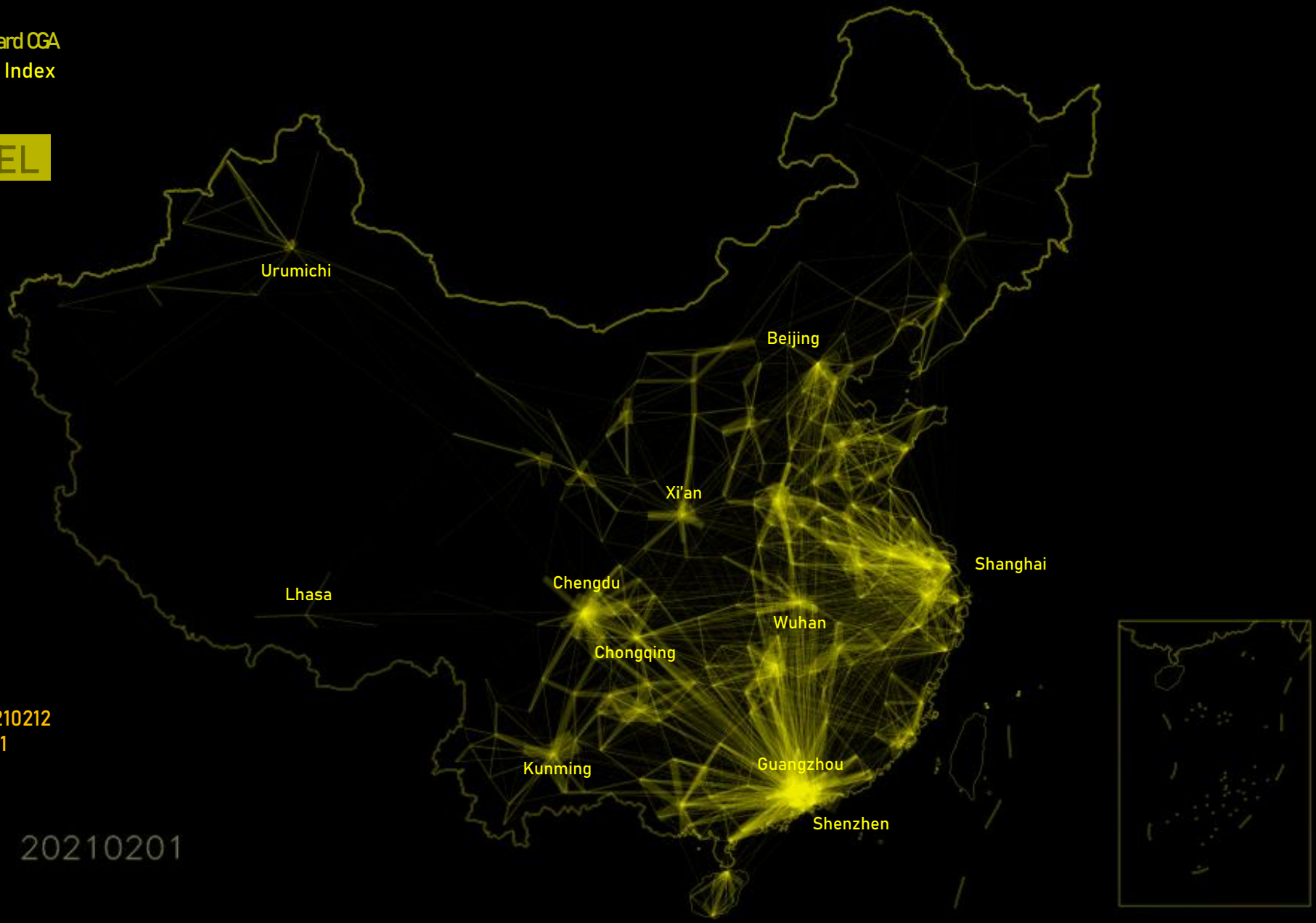
GUI

Harvard CGA

Global Urban Impulse Index

Daily

CITY-LEVELEL



* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

20210201

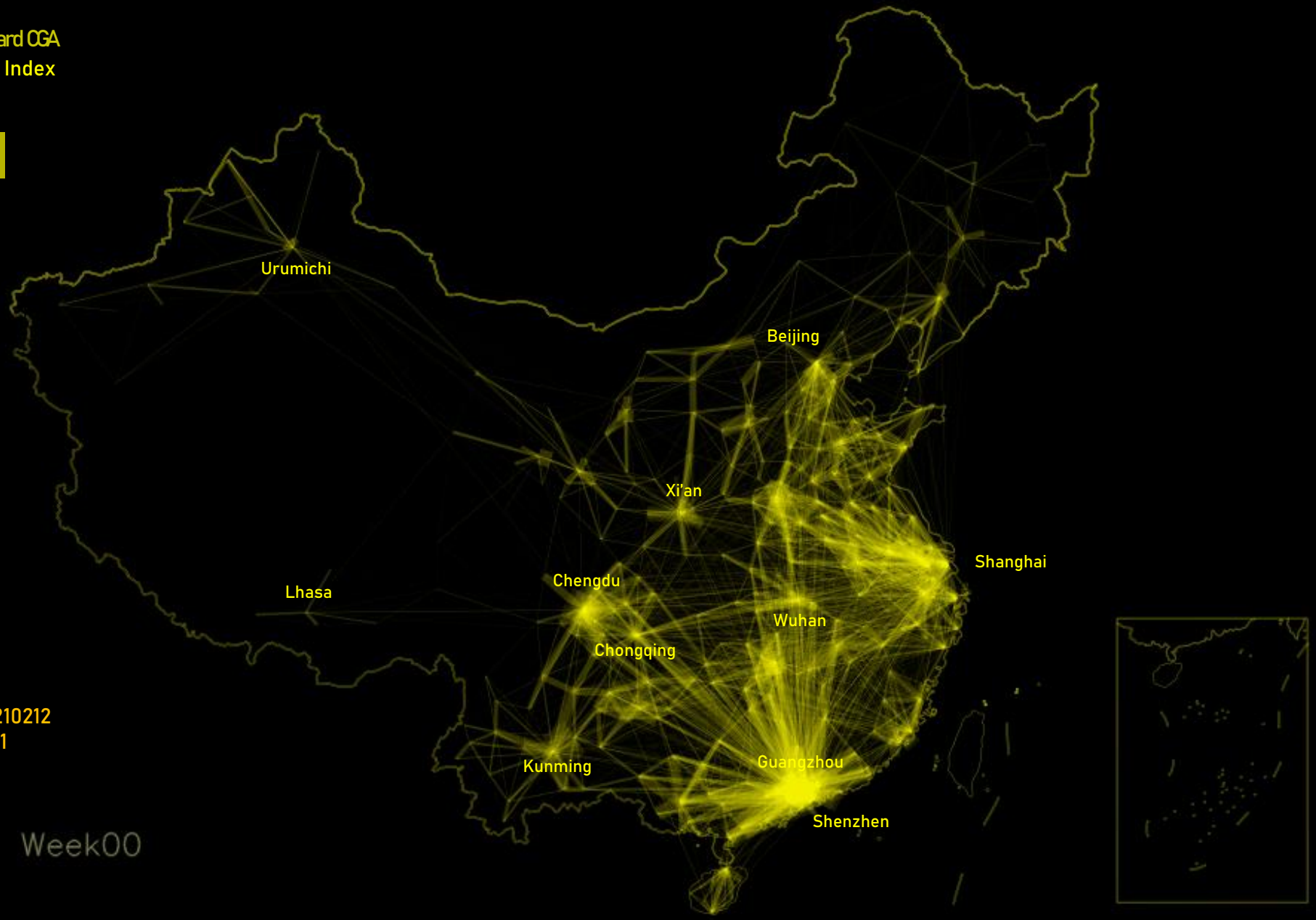
GUI

Harvard CGA

Global Urban Impulse Index

Weekly

CITY-LEVEL



* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

Week00

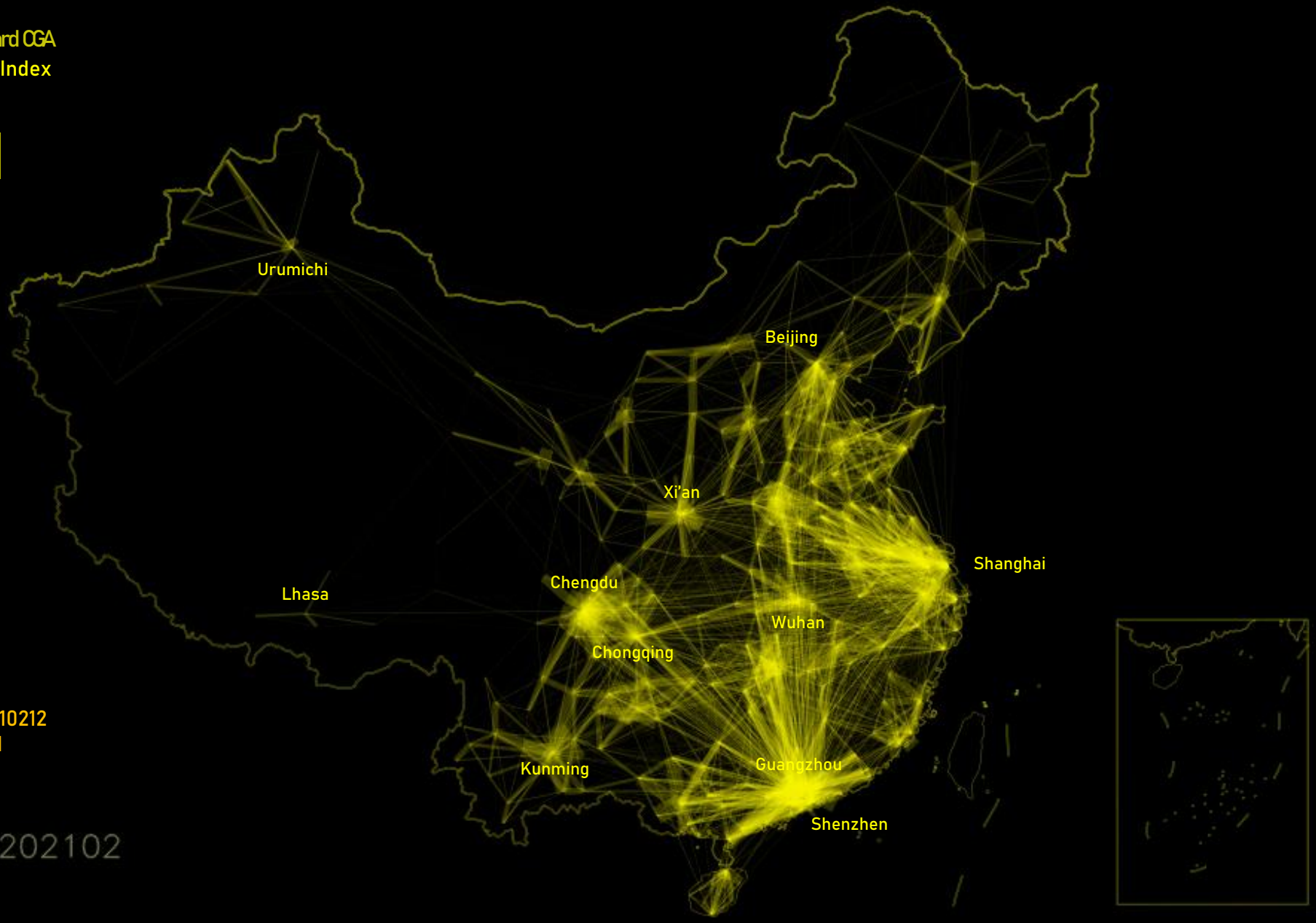
GUI

Harvard CGA

Global Urban Impulse Index

Monthly

CITY-LEVEL



* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

202102

GUI

Harvard CGA

Global Urban Impulse Index

Daily

PROVINCE-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

20210201

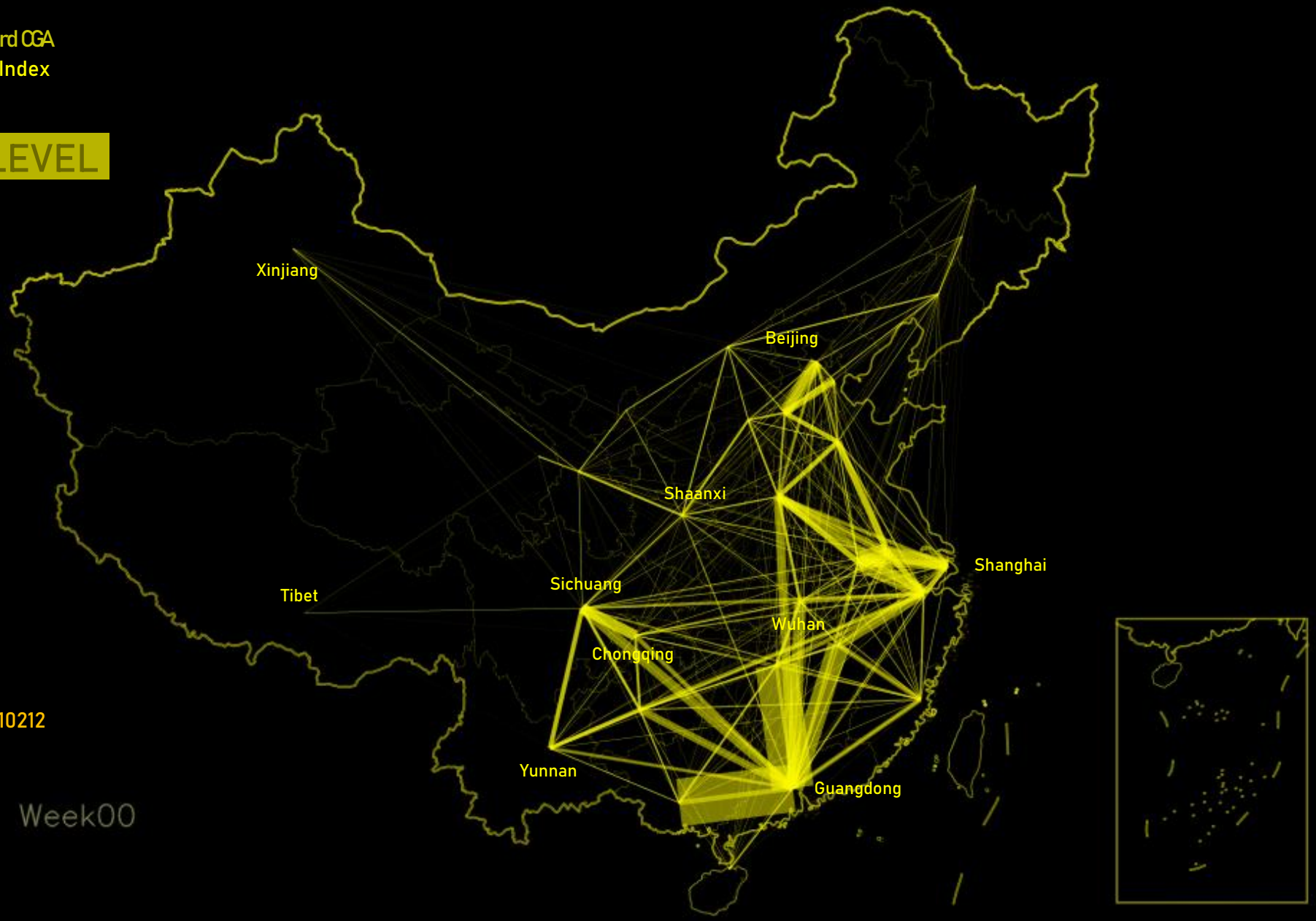
GUI

Harvard CGA

Global Urban Impulse Index

Weekly

PROVINCE-LEVEL



* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

Week00

GUI

Harvard CGA

Global Urban Impulse Index

Monthly

PROVINCE-LEVEL

* Spring Festival 20210212
* Labor Day 20210501

Baidu Data

20210201-20210630

202102

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

20210201-20210630

Comparison

Weibo Data vs Baidu Data

Aggregated Human Flow

Weekly Monthly SCALE

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Statistical summary of locational completeness of all OD matrixes.

Data Source	Geographic Units	Daily 04/1,2021		Weekly 04/1-04/7,2021		Monthly 04/1/-04/30,2021		Total
		Records	%	Records	%	Records	%	
Weibo	Province	20	1.84	399	36.64	824	75.67	33x33=1089
	City	34	0.01	1599	1.18	6332	4.68	368x368=135424
Baidu	Province	975	89.53	1007	92.47	1012	92.93	33x33=1089
	City	35570	26.27	55275	40.82	72450	53.50	368x368=135424

3

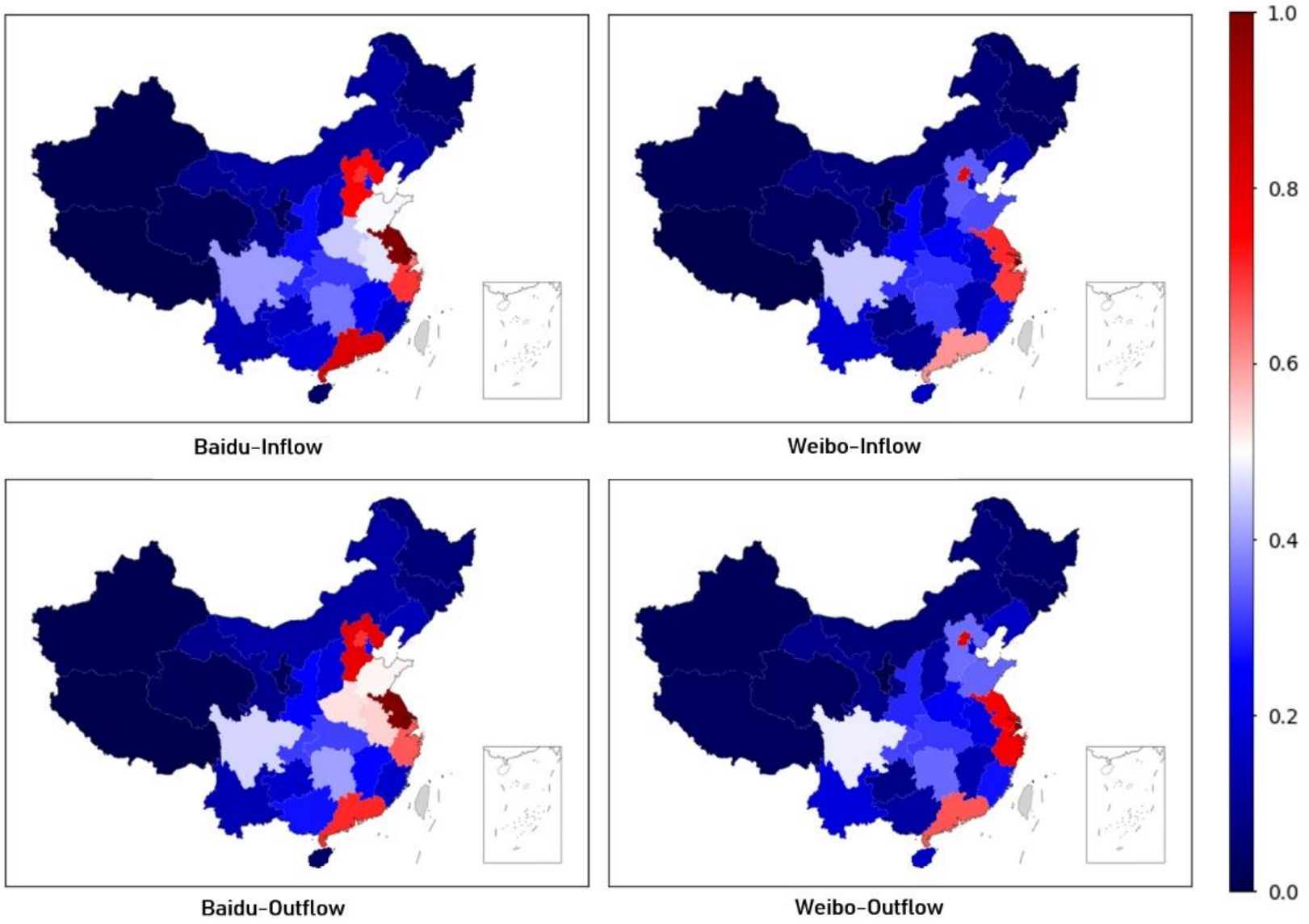
Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Normalized aggregated flow by Baidu Qianxi and Weibo tweets



3

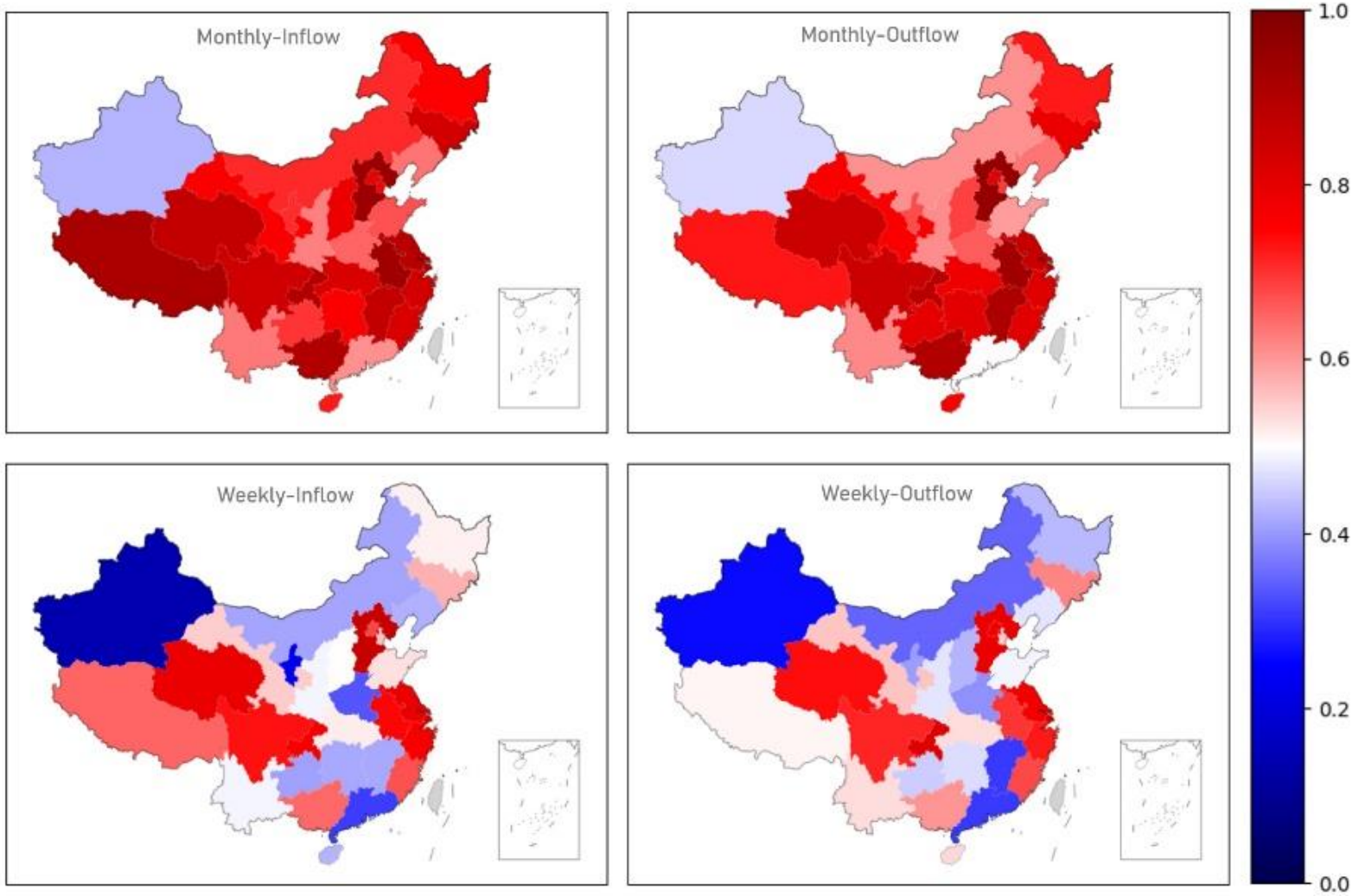
Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Correlation at province level based on flow count and Baidu index



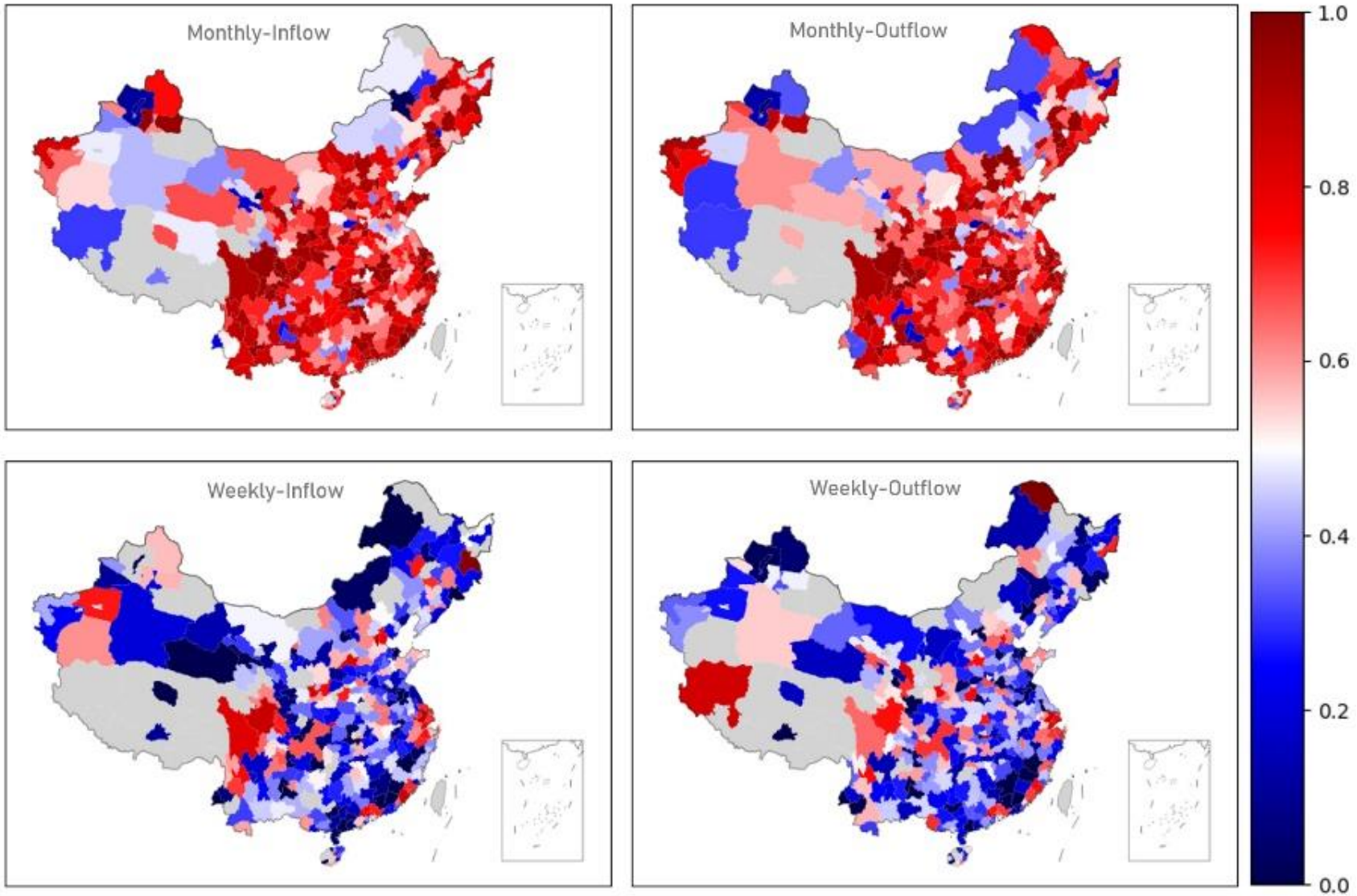
3 Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Correlation at city level based on flow count and Baidu index



3

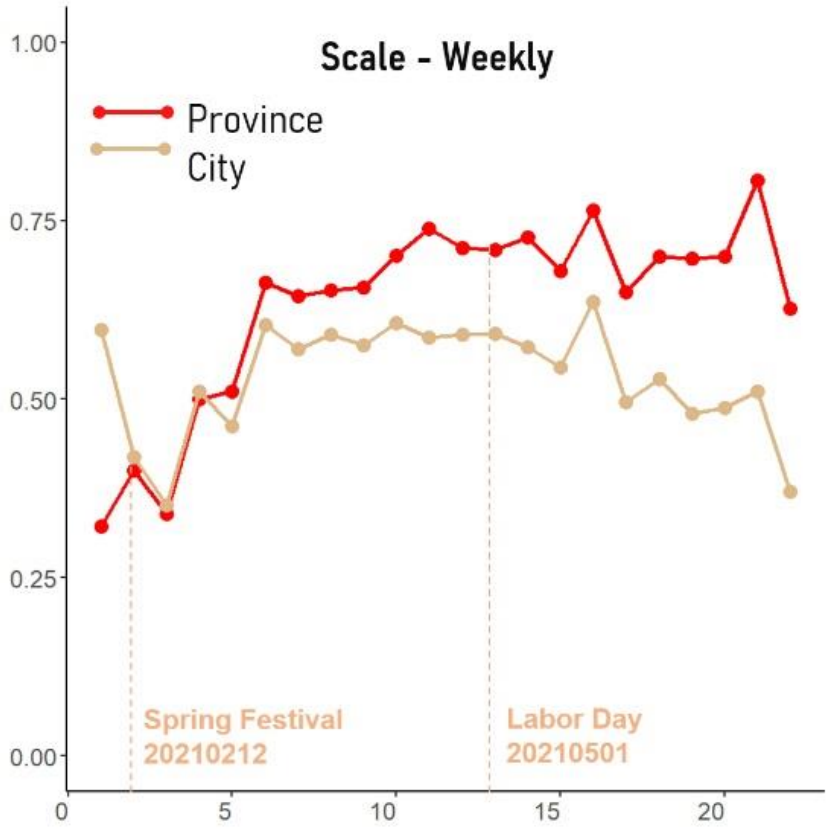
Comparison

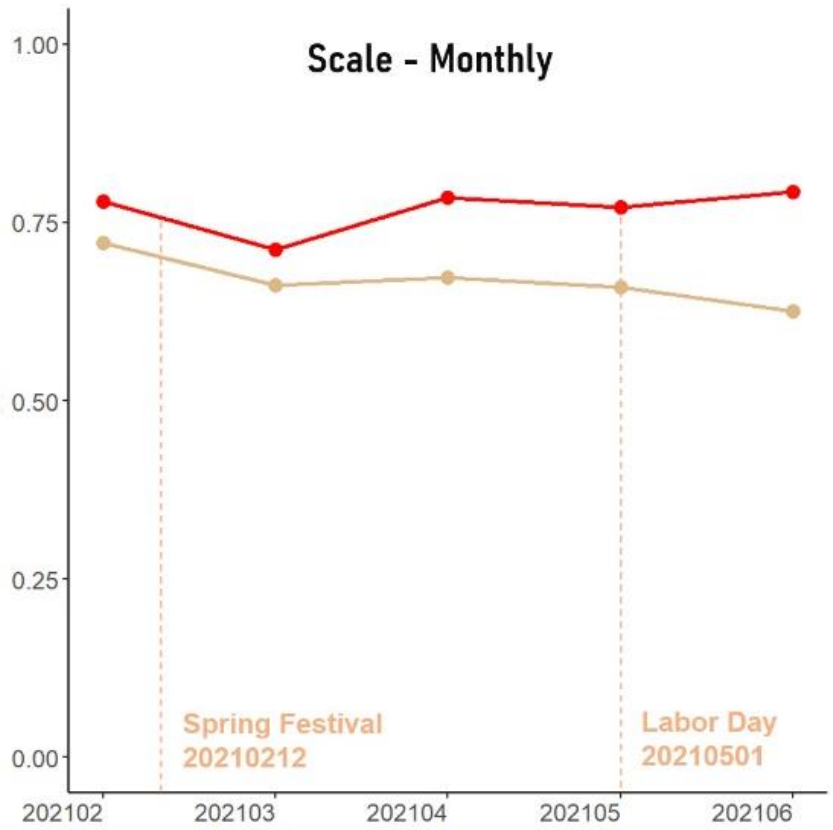
SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Human mobility data correlation between Baidu Qianxi and Weibo tweets over time





3

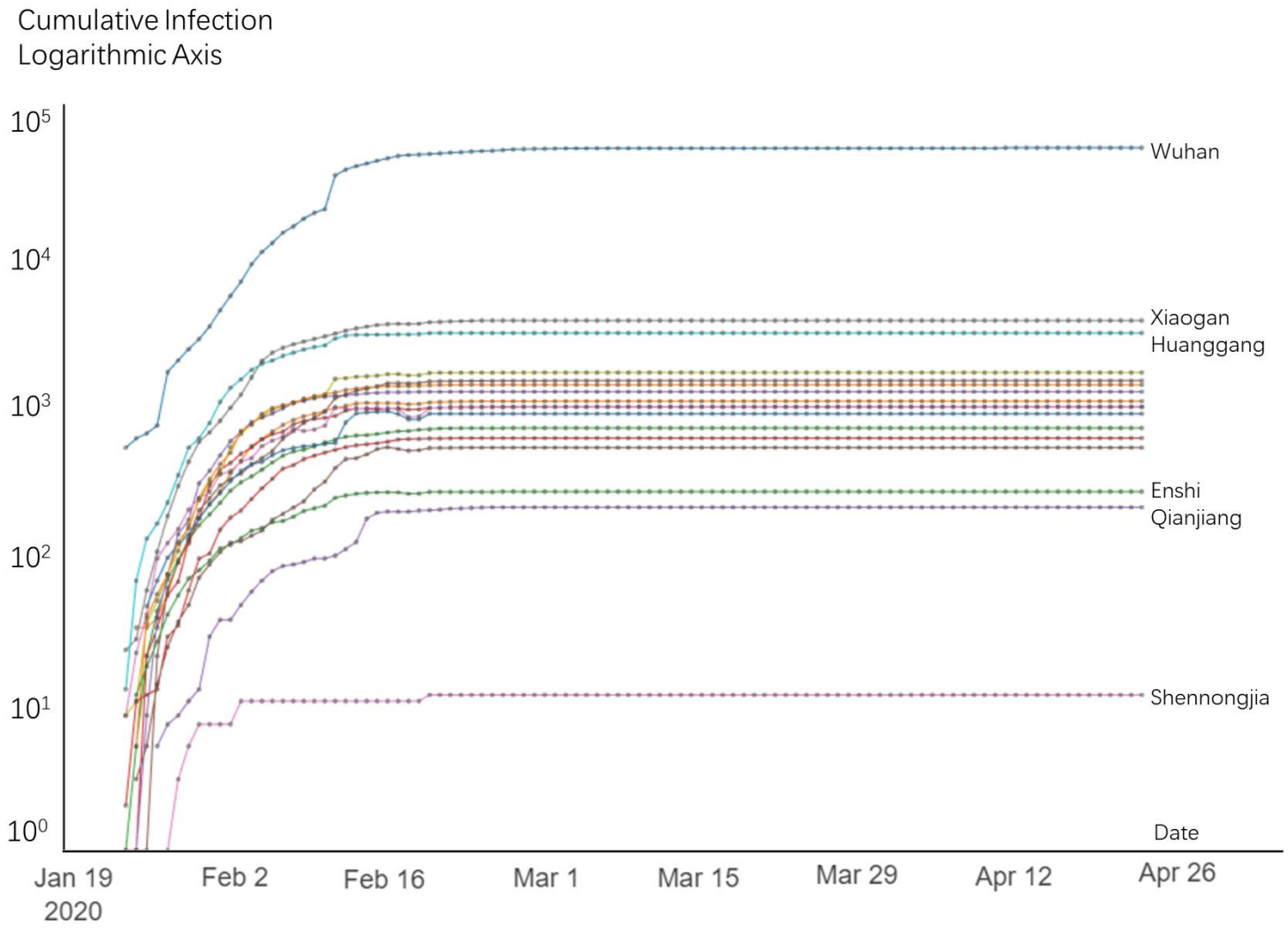
Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Application in COVID-19 Research



3

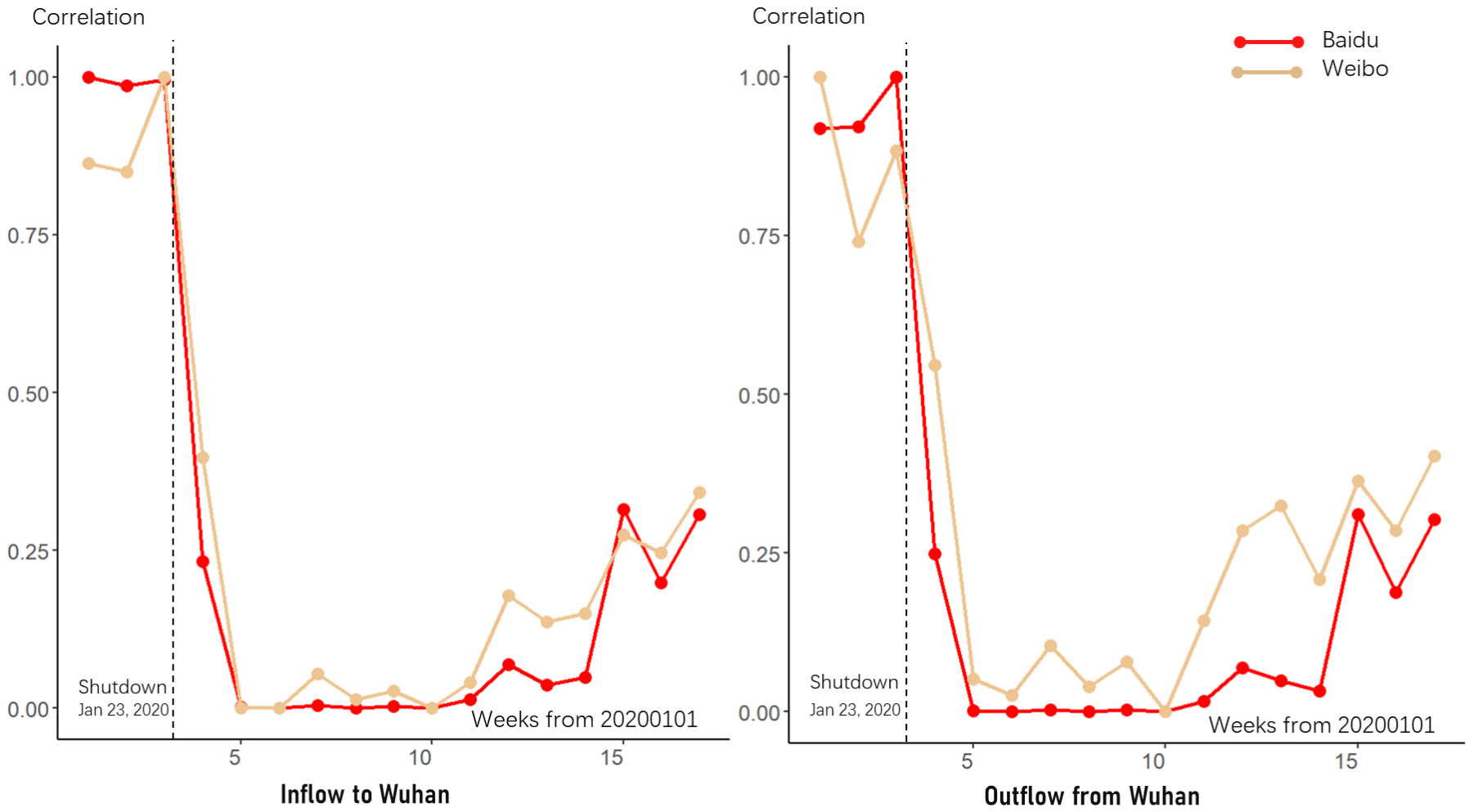
Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Total inflow and outflow of Wuhan from January 1 to April 30.



3

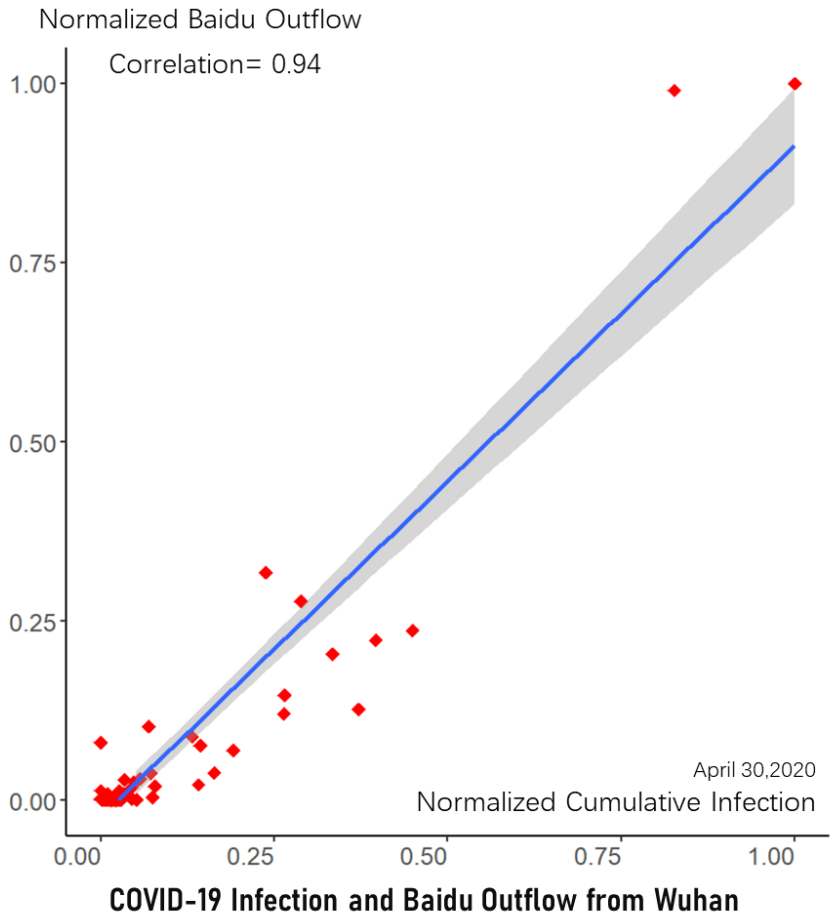
Comparison

SINA-Weibo Mobility Index

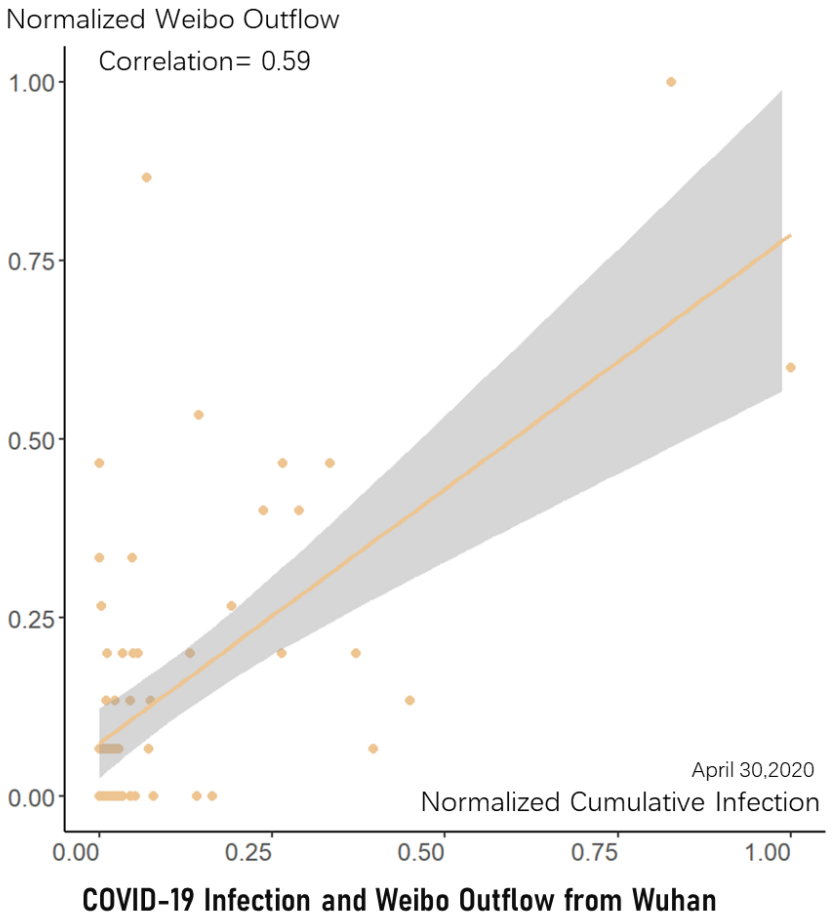
A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Comparison of COVID-19 and outflow from Wuhan to other cities in China



COVID-19 Infection and Baidu Outflow from Wuhan



COVID-19 Infection and Weibo Outflow from Wuhan

3

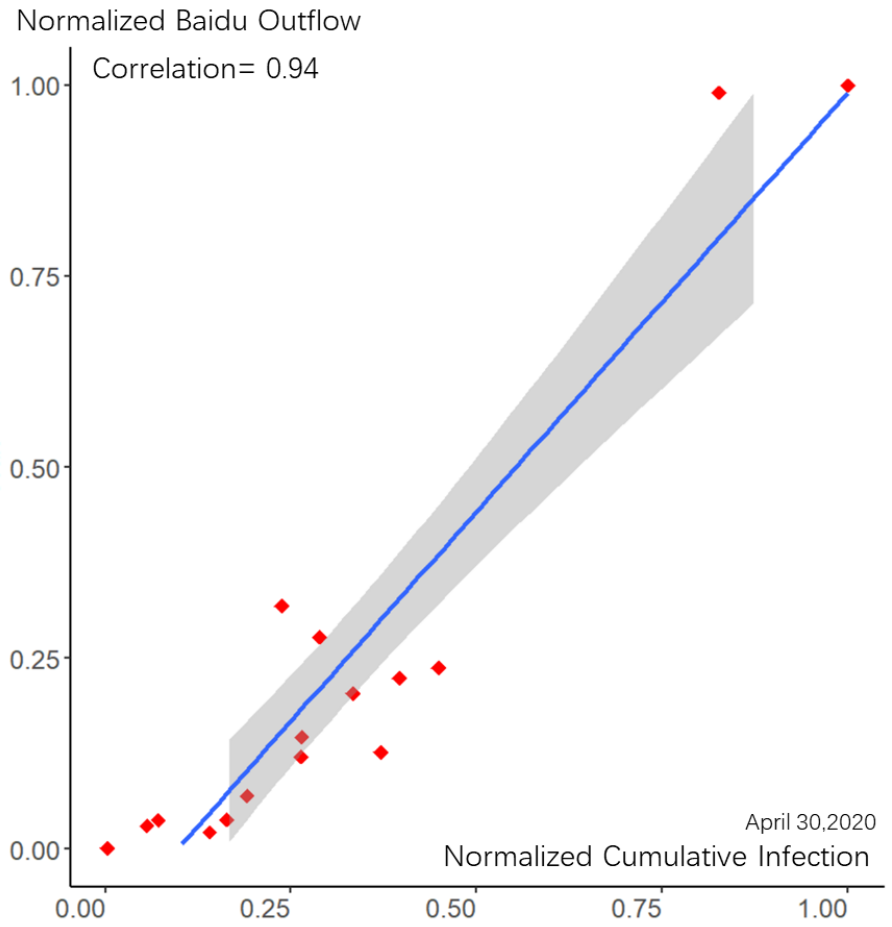
Comparison

SINA-Weibo Mobility Index

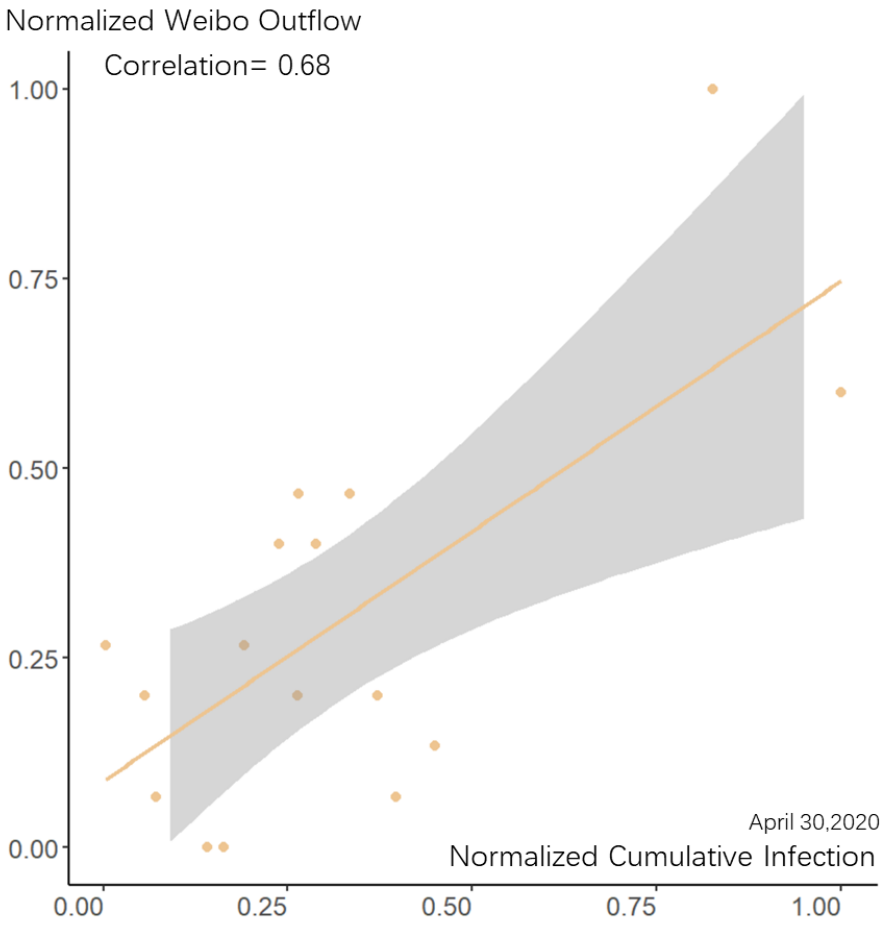
A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Comparison of COVID-19 and outflow from Wuhan to other cities in Hubei province



COVID-19 Infection and Baidu Outflow from Wuhan



COVID-19 Infection and Weibo Outflow from Wuhan

3

Comparison

Correlation coefficients between human mobility and COVID-19 infection

	Correaltionn for Flow Count			Correaltionn for Flow Rank		
	Baidu-COVID	Weibo-COVID	Baidu- Weibo	Baidu-COVID	Weibo-COVID	Baidu-Weibo
China	0.939	0.617	0.702	0.707	0.439	0.748
Hubei	0.941	0.679	0.812	0.868	0.417	0.586

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Yearly City level Weibo Index vs Geographic distance

Gravity Model $A_{ij}=kd_{ij}^{\beta}$

Weibo VS Distance

Linear Regressio...

File

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-1.2915	0.0521	-24.8111	0.0
Intercept	5.1603	0.1279	40.3359	0.0

Multiple R-Squared: 0.4166
Adjusted R-Squared: 0.4159

Baidu VS Distance

Linear Regressio...

File

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-2.335	0.0647	-36.0952	0.0
Intercept	9.6199	0.159	60.5054	0.0

Multiple R-Squared: 0.6018
Adjusted R-Squared: 0.6014

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Yearly Weibo Index vs Social economic factors

Gravity Model $A_{ij}=O_iD_jd_{ij}^{\beta}$, $O_i=\prod SE_i$, $D_j=\prod SE_j$

Weibo – distance -SES

Linear Regression Result View - 0:1060 - Linear Re...

File

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-1.2118	0.0419	-28.9219	0.0
GDP_O	1.5424	0.1522	10.1319	0.0
Population_O	-0.7786	0.1219	-6.3864	2.79E-10
FDI_O	0.2429	0.0588	4.1312	3.96E-5
Industry Water Usage_O	-0.1323	0.0443	-2.987	0.0029
Patent number_O	-0.4364	0.0774	-5.6414	2.30E-8
GDP_D	1.6575	0.1575	10.5223	0.0
Population_D	-0.9148	0.1266	-7.2268	1.10E-12
FDI_D	0.2947	0.0588	5.014	6.49E-7
Industry Water Usage_D	-0.1999	0.0435	-4.5976	4.92E-6
Patent number_D	-0.3646	0.0781	-4.67	3.50E-6
Intercept	-11.5838	0.8911	-13.0001	0.0

Multiple R-Squared: 0.6589
Adjusted R-Squared: 0.6544

Baidu – distance -SES

Linear Regression Result View - 0:1059 - Linear R...

File

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-2.1212	0.0529	-40.1197	0.0
GDP_O	0.8832	0.1921	4.5976	4.92E-6
Population_O	0.0357	0.1539	0.232	0.8166
FDI_O	0.0721	0.0742	0.9719	0.3314
Industry Water Usage_O	-0.2706	0.0559	-4.8407	1.54E-6
Patent number_O	-0.3041	0.0976	-3.1148	0.0019
GDP_D	1.0484	0.1988	5.2743	1.69E-7
Population_D	0.0131	0.1597	0.0817	0.9349
FDI_D	0.3862	0.0742	5.2077	2.40E-7
Industry Water Usage_D	-0.2637	0.0549	-4.8059	1.82E-6
Patent number_D	-0.3134	0.0985	-3.1812	0.0015
Intercept	-8.2999	1.1244	-7.3816	3.72E-13

Multiple R-Squared: 0.7462
Adjusted R-Squared: 0.7429

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Weibo Index vs Social economic factors at province level

Province Week

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.4232	0.0043	-97.6681	0.0
Intercept	0.9378	0.0026	360.288	0.0

Multiple R-Squared: 0.5445
Adjusted R-Squared: 0.5444

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.3969	0.01	-39.716	0.0
Intercept	0.3764	0.006	62.7091	0.0

Multiple R-Squared: 0.165
Adjusted R-Squared: 0.1649

Province Month

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.5167	0.0075	-68.7703	0.0
Intercept	0.9767	0.005	197.057	0.0

Multiple R-Squared: 0.5489
Adjusted R-Squared: 0.5488

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.6022	0.0158	-38.168	0.0
Intercept	0.6828	0.0104	65.6009	0.0

Multiple R-Squared: 0.2727
Adjusted R-Squared: 0.2725

Baidu

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-0.3933	0.0038	-104.2729	0.0
GDP_O	0.2648	0.0169	15.6929	0.0
Population_O	-0.017	0.0139	-1.2219	0.2218
FDI_O	0.0019	0.0098	0.1927	0.8472
Industry Water Usage_O	-0.0564	0.0048	-11.7745	0.0
Patent number_O	-0.0749	0.0178	-4.2063	2.62E-5
GDP_D	0.21	0.0171	12.2874	0.0
Population_D	0.0188	0.0141	1.3377	0.181
FDI_D	0.0073	0.01	0.7341	0.4629
Industry Water Usage_D	-0.051	0.0048	-10.7232	0.0
Patent number_D	-0.0589	0.0181	-3.2493	0.0012
Intercept	0.7341	0.0062	118.9362	0.0

Multiple R-Squared: 0.6831
Adjusted R-Squared: 0.6826

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-0.4393	0.0063	-69.6339	0.0
GDP_O	0.2883	0.0246	11.7154	0.0
Population_O	0.0321	0.02	1.5999	0.1097
FDI_O	0.0653	0.0148	4.4024	1.10E-5
Industry Water Usage_O	-0.0768	0.0084	-9.1123	0.0
Patent number_O	-0.1374	0.0258	-5.3292	1.04E-7
GDP_D	0.2853	0.0245	11.6383	0.0
Population_D	0.0071	0.0201	0.354	0.7233
FDI_D	0.0285	0.0149	1.9166	0.0554
Industry Water Usage_D	-0.0691	0.0084	-8.2238	2.22E-16
Patent number_D	-0.1121	0.0255	-4.3937	1.14E-5
Intercept	0.7094	0.0093	76.38	0.0

Multiple R-Squared: 0.7234
Adjusted R-Squared: 0.7227

Weibo

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-0.3963	0.0094	-42.2194	0.0
GDP_O	0.5758	0.042	13.7154	0.0
Population_O	-0.4836	0.0346	-13.9893	0.0
FDI_O	0.1975	0.0245	8.0636	8.88E-16
Industry Water Usage_O	-0.0564	0.0119	-4.7307	2.28E-6
Patent number_O	-0.165	0.0443	-3.723	0.0002
GDP_D	0.5705	0.0425	13.4105	0.0
Population_D	-0.4419	0.0351	-12.6071	0.0
FDI_D	0.1957	0.0249	7.8709	4.00E-15
Industry Water Usage_D	-0.0557	0.0118	-4.7092	2.53E-6
Patent number_D	-0.1771	0.0451	-3.9259	8.71E-5
Intercept	0.3072	0.0154	20.0023	0.0

Multiple R-Squared: 0.3233
Adjusted R-Squared: 0.3224

Variable	Coeff.	Std. Err.	t-value	P>|t|
Distance	-0.509	0.0129	-39.4983	0.0
GDP_O	0.9456	0.0503	18.81	0.0
Population_O	-0.5221	0.0409	-12.7548	0.0
FDI_O	0.4192	0.0303	13.8378	0.0
Industry Water Usage_O	-0.1384	0.0172	-8.0367	1.11E-15
Patent number_O	-0.5187	0.0526	-9.8518	0.0
GDP_D	0.8649	0.0501	17.2718	0.0
Population_D	-0.43	0.041	-10.4766	0.0
FDI_D	0.3966	0.0303	13.0732	0.0
Industry Water Usage_D	-0.1127	0.0172	-6.5609	6.05E-11
Patent number_D	-0.5256	0.0521	-10.0869	0.0
Intercept	0.4377	0.019	23.0708	0.0

Multiple R-Squared: 0.5781
Adjusted R-Squared: 0.5769

3

Comparison

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Weibo Index vs Social economic factors at province level

City Week

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.5977	0.0041	-146.6751	0.0
Intercept	1.0143	0.0025	412.126	0.0

Multiple R-Squared: 0.477
Adjusted R-Squared: 0.4769

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.2207	0.0052	-42.4147	0.0
Intercept	0.204	0.0032	63.7055	0.0

Multiple R-Squared: 0.0675
Adjusted R-Squared: 0.0674

City Month

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.6501	0.0044	-149.0738	0.0
Intercept	1.0343	0.0028	367.8813	0.0

Multiple R-Squared: 0.475
Adjusted R-Squared: 0.475

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.345	0.0063	-55.1408	0.0
Intercept	0.3319	0.0041	80.6335	0.0

Multiple R-Squared: 0.1023
Adjusted R-Squared: 0.1022

Baidu

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.7307	0.0038	-194.0479	0.0
Population	0.1432	0.0144	9.9141	0.0
GDP	0.4976	0.0151	32.8618	0.0
FDI	-0.0673	0.0065	-10.3408	0.0
Patent	-0.1436	0.0112	-12.8776	0.0
Population (#1)	0.1532	0.0144	10.6736	0.0
GDP (#1)	0.4243	0.0148	28.7122	0.0
FDI (#1)	-0.0302	0.0063	-4.8245	1.41E-6
Patent (#1)	-0.0939	0.0109	-8.6297	0.0
Intercept	0.5533	0.0046	119.0216	0.0

Multiple R-Squared: 0.6675
Adjusted R-Squared: 0.6673

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.7778	0.0041	-190.7664	0.0
Population	0.1045	0.0137	7.6271	2.51E-14
GDP	0.4954	0.0144	34.3164	0.0
FDI	-0.0609	0.006	-10.1356	0.0
Patent	-0.0989	0.0104	-9.4652	0.0
Population (#1)	0.075	0.0136	5.5107	3.62E-8
GDP (#1)	0.4154	0.0137	30.2364	0.0
FDI (#1)	-0.0471	0.0056	-8.4372	0.0
Patent (#1)	-0.0597	0.0099	-6.03	1.67E-9
Intercept	0.5822	0.0048	122.1978	0.0

Multiple R-Squared: 0.663
Adjusted R-Squared: 0.6629

Weibo

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.3127	0.0057	-54.5527	0.0
Population	-0.2205	0.0227	-9.7219	0.0
GDP	0.412	0.0233	17.6507	0.0
FDI	-0.0202	0.0099	-2.0358	0.0418
Patent	-0.042	0.0172	-2.4482	0.0144
Population (#1)	-0.1776	0.0226	-7.8541	4.22E-15
GDP (#1)	0.4147	0.0233	17.7614	0.0
FDI (#1)	-0.0244	0.0099	-2.4565	0.014
Patent (#1)	-0.0503	0.0172	-2.9288	0.0034
Intercept	-0.1532	0.0074	-20.825	0.0

Multiple R-Squared: 0.1958
Adjusted R-Squared: 0.1955

Variable	Coeff.	Std. Err.	t-value	P>|t|
distance	-0.5144	0.0067	-77.1962	0.0
Population	-0.3534	0.023	-15.35	0.0
GDP	0.6948	0.0235	29.5216	0.0
FDI	-0.0395	0.0096	-4.094	4.25E-5
Patent	-0.0868	0.0169	-5.1405	2.76E-7
Population (#1)	-0.3527	0.0228	-15.4482	0.0
GDP (#1)	0.6487	0.0232	27.9432	0.0
FDI (#1)	-0.0395	0.0095	-4.1516	3.31E-5
Patent (#1)	-0.0679	0.0167	-4.0579	4.97E-5
Intercept	-0.1816	0.0081	-22.4508	0.0

Multiple R-Squared: 0.314
Adjusted R-Squared: 0.3137

4

Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

File List

For Each Year (2018-2020)
Tables for provinces (n):
Weekly tables (one table for all weeks):
- Weekly out-flows between provinces by counts (n x number of weeks x n)
- Weekly in--flows between provinces by counts (n x number of weeks x n)

Monthly tables (one table for all 12 months):
- Monthly out-flows between provinces by counts (n x number of months x n)
- Monthly in--flows between provinces by counts (n x number of months x n)

Yearly tables (one table for each year)::
- Yearly out-flows between provinces by counts (nxn)
- Yearly in--flows between provinces by counts (nxn)

Mobility scales by province
- Weekly out-flow total scales by province (n x number of weeks)
- Weekly in-flow total scales by province (n x number of weeks)
- Monthly out-flow total scales by province (n x number of months)
- Monthly in-flow total scales by province (n x number of months)
- Yearly out-flow total scales by province (n x 1)
- Yearly in-flow total scales by province (n x 1)

Tables for cities (m):
Weekly tables (one table for each week):
- Weekly out-flows between cities by counts (m x number of weeks x m)
- Weekly in--flows between cities by counts ratio (m x number of weeks x m)

Monthly tables (one table for each month)::
- Monthly out-flows between cities by counts ratio (m x number of months x m)
- Monthly in--flows between cities by counts ratio (m x number of months x m)

Yearly tables (one table for each year)::
- Yearly out-flows between cities by counts ratio (mxm)
- Yearly in--flows between cities by counts ratio (mxm)

Mobility scales by city
- Weekly out-flow scales by city (m x number of weeks)
- Weekly in-flow scales by city (m x number of weeks)
- Monthly out-flow scales by city (m x number of months)
- Monthly in-flow scales by city (m x number of months)
- Yearly out-flow scales by city (m x1)

Yearly in-flow scales by city (m x 1)

4 Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Illustration of Inflow/Outflow—Counts Index

Example: 2018_City_Monthly_Inflow_Counts.csv

1st Column
D_Code, Code for **Destination City** (or Prov)

2nd Column
Month (or First day of Weeks)
For Yearly data, there is no **Time** column

3rd Column~ Nth Column
O_Code, Code for Origin **City** (or Province)

D_Code	Time	O_1100	O_1200	O_1301	O_1302	O_1303	O_1304	O_1305	O_1306
D_1100	2018-01	4928	61	21	6	7	4	2	21
D_1100	2018-02	3499	49	19	8	3	8	4	17
D_1100	2018-03	5199	71	18	10	10	2	1	25
D_1100	2018-04	5235	102	28	10	11	5	2	21
D_1100	2018-05	4324	76	19	10	12	3	1	15
D_1100	2018-06	4886	75	24	14	19	6	2	13
D_1100	2018-07	4311	53	17	6	27	2	2	12
D_1100	2018-08	4633	59	20	11	27	4	3	21
D_1100	2018-09	4776	65	18	10	21	2	3	21
D_1100	2018-10	4534	101	36	13	27	7	4	42
D_1100	2018-11	4981	56	23	7	8	6	1	13
D_1100	2018-12	5203	61	16	9	5	2	2	9
D_1200	2018-01	59	931	1	4	3	0	0	4
D_1200	2018-02	62	760	0	2	1	1	1	3
D_1200	2018-03	82	964	4	5	3	2	0	4
D_1200	2018-04	82	1026	3	4	3	0	1	6
D_1200	2018-05	82	885	5	4	3	1	0	2
D_1200	2018-06	86	994	3	5	7	2	0	5
D_1200	2018-07	72	772	4	6	17	1	1	6

If **D_code**= **O_code**, the values means the total amount of Weibo users who posted Weibo tweets in the same city (or Province)

The inflow scale from **Origin city**/Province to **Destination city**/Province (**O_Code**)

4

Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Illustration of Inflow/Outflow—TotalScale Index

Example: 2018_City_Monthly_Inflow_TotalScale.csv

1st Column
D_Code, Code for **Destination City** (or Prov)

2nd Column ~Nth Column
Month (or First day of Weeks)

D_Code	2018-01	2018-02	2018-03	2018-04	2018-05	2018-06	2018-07	2018-08	2018-09
D_1100	5644	4137	5975	6189	5012	5694	5052	5569	5522
D_1200	1090	939	1170	1299	1057	1217	993	1061	1176
D_1301	504	500	547	556	515	563	486	492	524
D_1302	219	224	189	220	176	233	190	213	210
D_1303	131	156	138	185	150	259	285	287	228
D_1304	167	228	156	157	139	170	148	159	152
D_1305	98	145	96	94	97	102	83	90	96
D_1306	264	309	240	274	249	282	221	227	263

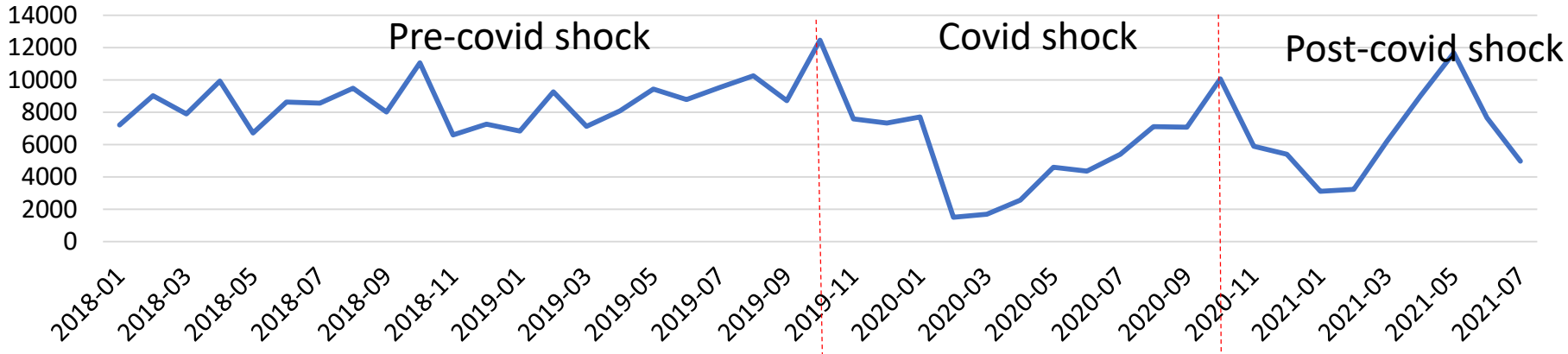
The Total inflow scale from all other **cities**/Provinces to **Destination city** /Province (**O_Code**)

C

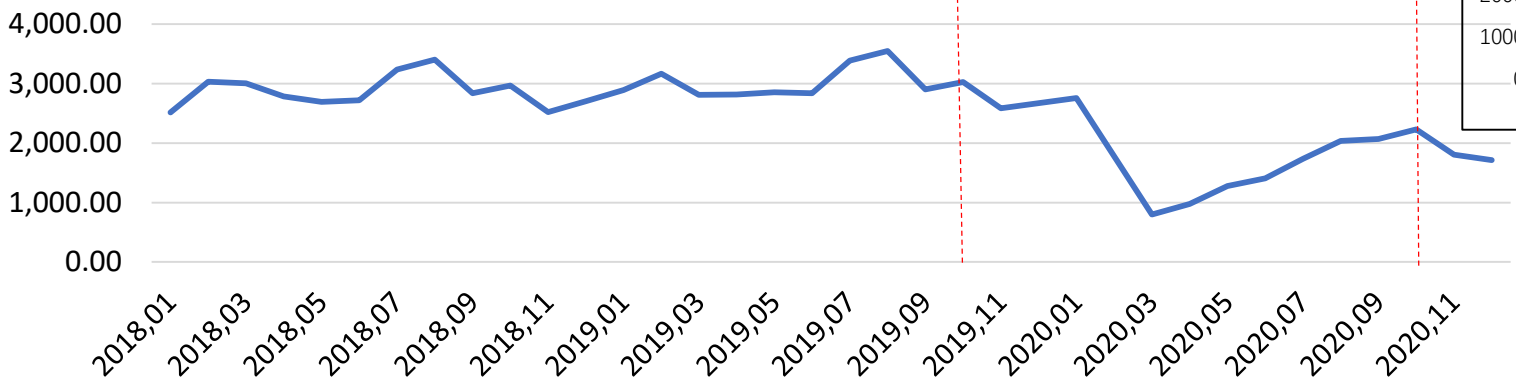
Take a Pulse of China with Weibo Mobility Index Data

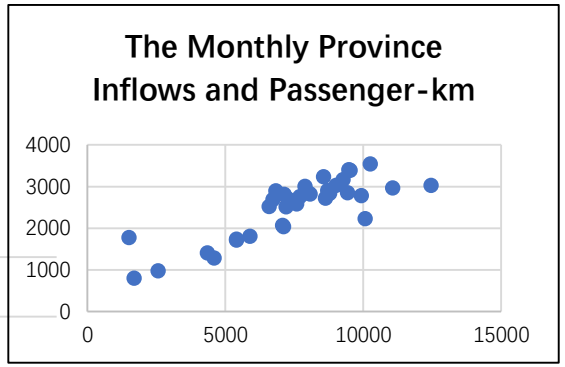
4 Discussion

The Monthly Mobility Scales of All Provinces from 2021 to 2021



The Monthly Passenger-km (in 100 million)





Data source: http://china-data-online.com

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

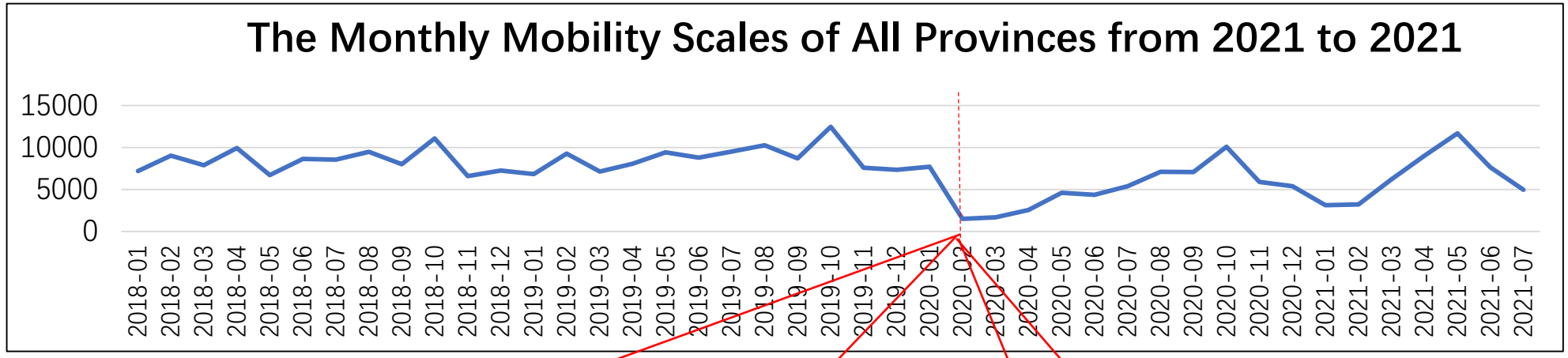
Center for Geographic Analysis
Harvard University

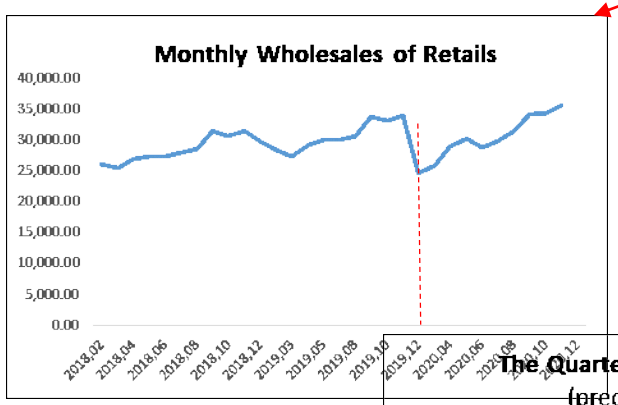
4 Discussion

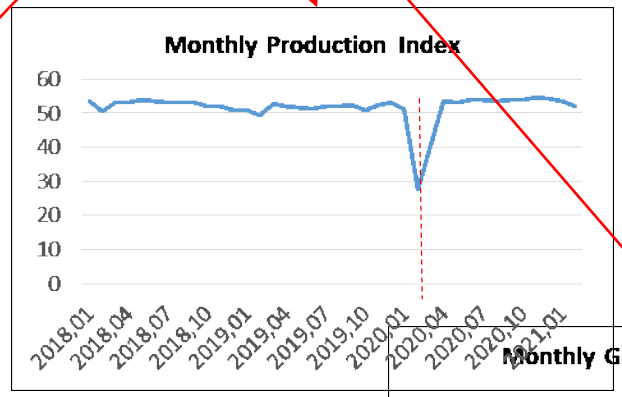
SINA-Weibo Mobility Index

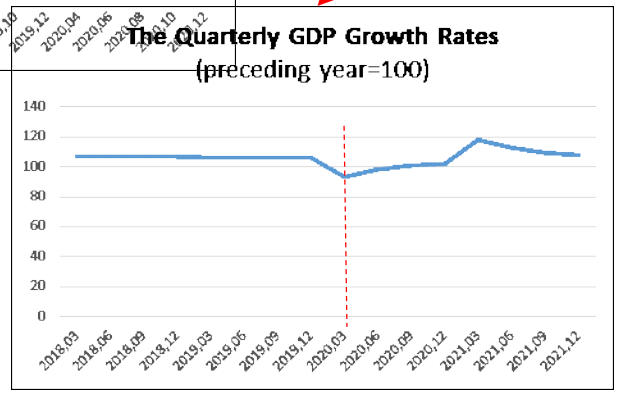
A Initiative Project of Global Urban Impulse

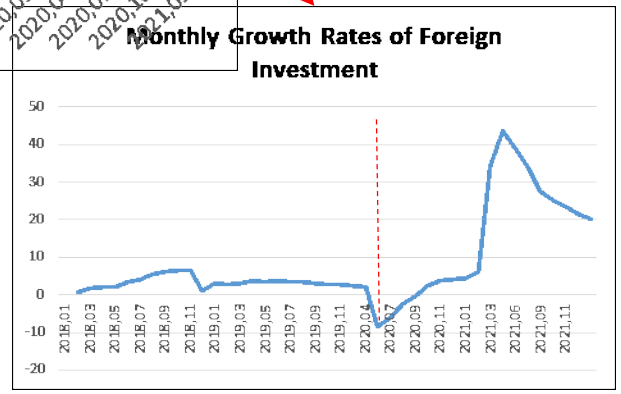
Center for Geographic Analysis
Harvard University











Data source: http://china-data-online.com

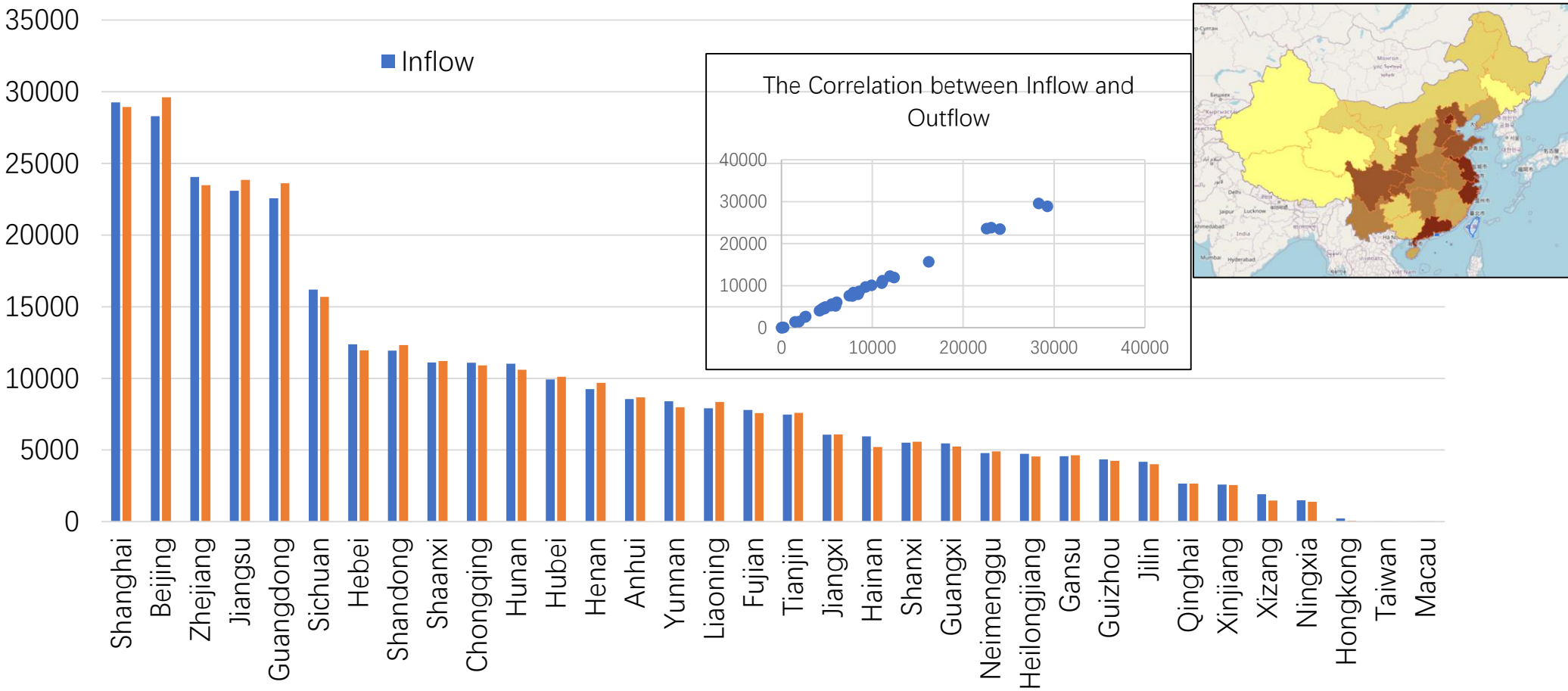
4 Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

The Inflow and Outflow Index by Provinces of China (2018-2021)

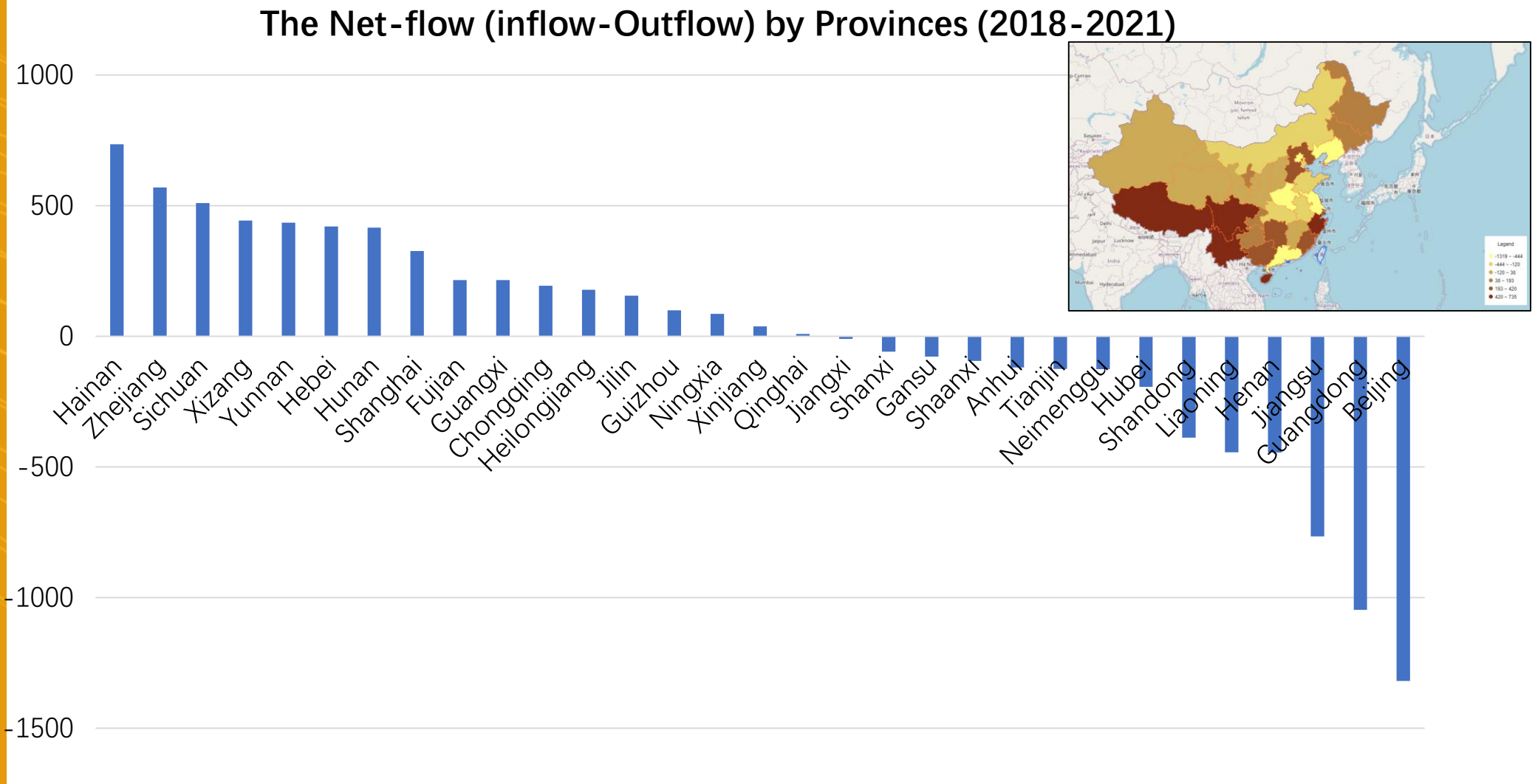


4 Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University



4

Discussion

SINA-Weibo Mobility Index

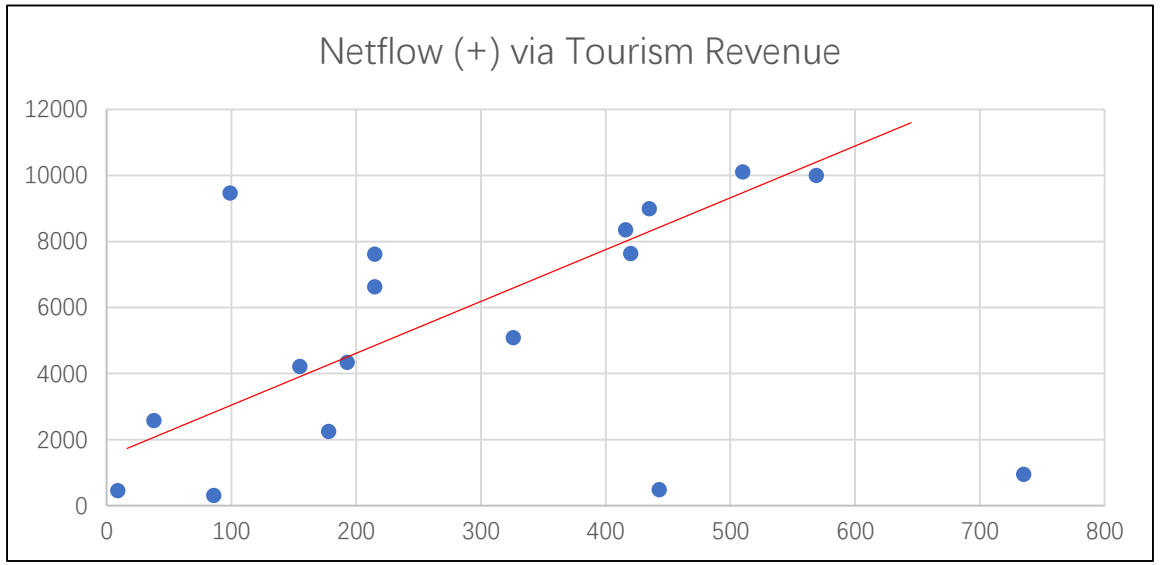
A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Top 10 Provinces with Netflows (+)

- Hainan
- Zhejiang
- Sichuan
- Xizang
- Yunnan
- Hebei
- Hunan
- Shanghai
- Fujian
- Guangxi

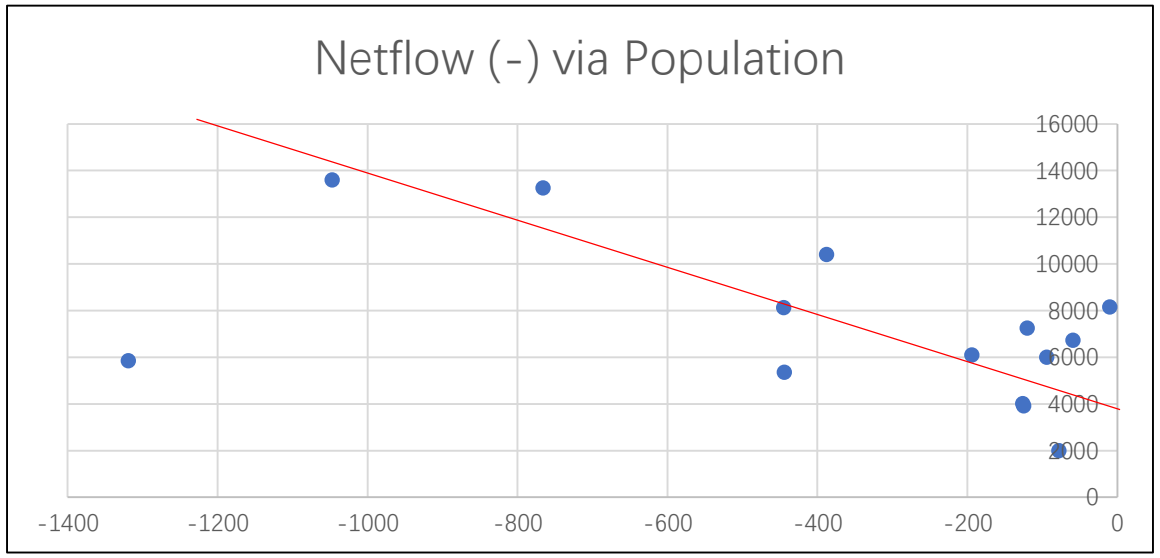
Provinces with (+) Netflows: Tourist attraction



Top 10 Provinces with Netflows (-)

- Beijing
- Guangdong
- Jiangsu
- Henan
- Liaoning
- Shandong
- Hubei
- Neimenggu
- Tianjin
- Anhui

Provinces with (-) Netflows: Population



4

Discussion

SINA-Weibo Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Prospective Applications

- ☐ Spatiotemporal changes in regional connectivity pre and post pandemic
- ☐ Predictability and detection of emerging events and shocks
- ☐ Assessment of shock impacts on population, business and economy
- ☐ Understanding the motivation of mobility and spatiotemporal difference
- ☐ Identification of long-term/short-term and intra/inter-region mobility
- ☐ Urban / Rural vitality index

Data Access:

China Data Lab, 2022, "Weibo Mobility Index", https://dataverse.harvard.edu/dataverse/weibomobilityindex, Harvard Dataverse.

Reference:

Liu, L.; Wang, R.; Guan, W.W.; Bao, S.; Yu, H.; Fu, X.; Liu, H. 2022. **Assessing Reliability of Chinese Geotagged Social Media Data for Spatiotemporal Representation of Human Mobility**. ISPRS Int. J. Geo-Inf. 2022, 11, 145. https://doi.org/10.3390/ijgi11020145.

Acknowledgement:

- MIT Sustainable Urbanization Lab, MIT
- NSF award number 1841403
- Center for Geographic Analysis, Harvard University
- Geocomputation Center for Social Sciences, Wuhan University
- China Data Institute
- Future Data Lab

MIT | Sustainable Urbanization Lab

SINA-Weibo
Mobility Index

A Initiative Project of Global Urban Impulse

Center for Geographic Analysis
Harvard University

Spatial Data Lab:

http://spatialdatalab.org

Contact:

spatialdatalab@lists.fas.harvard.edu
office@chinadatacenter.net