The Growth of High-Speed Railway Network and its Effects on Labor Reallocation in China

Presented by Zijing Zhu

January 22, 2020

Introduction

Motivation:

- Expenditure in Transportation Infrastructure in Developing Countries:
  - 20% of World Bank lending goes for transportation infrastructure Baum-Snow et al.(2015)
- The ◂ Facts of HSR in China
- Decrease in Transportation Costs:(Cao et al(2013),Ahlfeldt et al(2010),Berhens and Pels(2011))
  - Transport Freights:Eaton and Kortum(2002),Donaldson and Hornbeck(2016),Grossman and Rossi-Hansberg(2008)
  - Transport Passengers:
    - HSR in China:Lin(2017),Lin(2019),Zheng and Kahn(2013),Zheng and Kahn(2018),Xu(2018)
    - HSR in Other Countries:Cheng et al(2014),Rus and Nombela (2007),Heuermann et al(2018),Bernard et al(2016)

Research Questions and Contributions

- Research Questions:
  How is labor reallocated across cities with the increase in connectivity to HSR in China?
- Contribution:
  - Estimate the increase of connectivity after connection:
    - Direct Connectivity ↑
    - Indirect Connectivity ↑: Spillover effects from nearby cities ✓
  - Estimate heterogeneous effects across different industries

Definitions and Data Sources

Variables	Definition	Data Source
Dependent Variables: Industries		
Em_i,t,k	Log value of employment for industry k for city i in year t	City Statistical Yearbook
Em_composition_i,t,k	Percentage of employment in industry k divided by total employment in for city i in year t	Author's Calculation
Independent Variables:		
Connectivity_i,t	Explained in Methodology	Author's Calculation
Other Variables:		
rgdp_i,t	Log value of real GDP for city i year t	City Statistical Yearbook
population_i,t	Log value of population for city i year t	City Statistical Yearbook
unemployment_i,t	Log value of unemployment for city i in year t	City Statistical Yearbook
rwage_i,t	Log value of average wage for city i year t	City Statistical Yearbook
rgov_exp_i,t	Log value of government expenditure for city i year t	City Statistical Yearbook
road usage_i,t	Log number of passengers who used road to transport in city i year t	City Statistical Yearbook
airplane usage_i,t	Log number of passengers who used airplane to transport in city i year t	City Statistical Yearbook

City i: excluding cities with missing values, in total there are 282 cities in prefecture level in the sample
Year t: range from 2003 to 2017

Methodology

Measure connectivity in different aspects:

- Degree: Important nodes have many connections
- Betweenness: Important nodes connect other nodes
- Center/Peripheral: Important nodes reach other nodes in the same network with shorter distance(not geographic distance)

Measure Connectivity

- Node: A B C D E
- Degree:number of direct connections to node v, $k_A = 1$
- Distance: Shortest Path
- Eccentricity:largest distance between node v and all other nodes: $\{A:3, B:2, C:2, D:3, E:2\}$
- Diameter: maximum distance between any pair of nodes: 3
- Radius: a network's minimun eccentricity: 2
- Center: Nodes with Eccentricity=Radius: $\{B, C, E\}$
- Periphery: Nodes with Eccentricity=Diameter:$\{A, D\}$

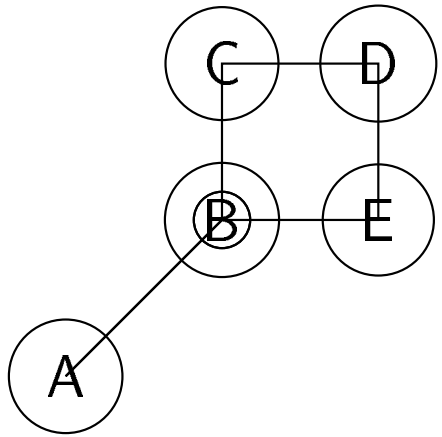


Figure: Example

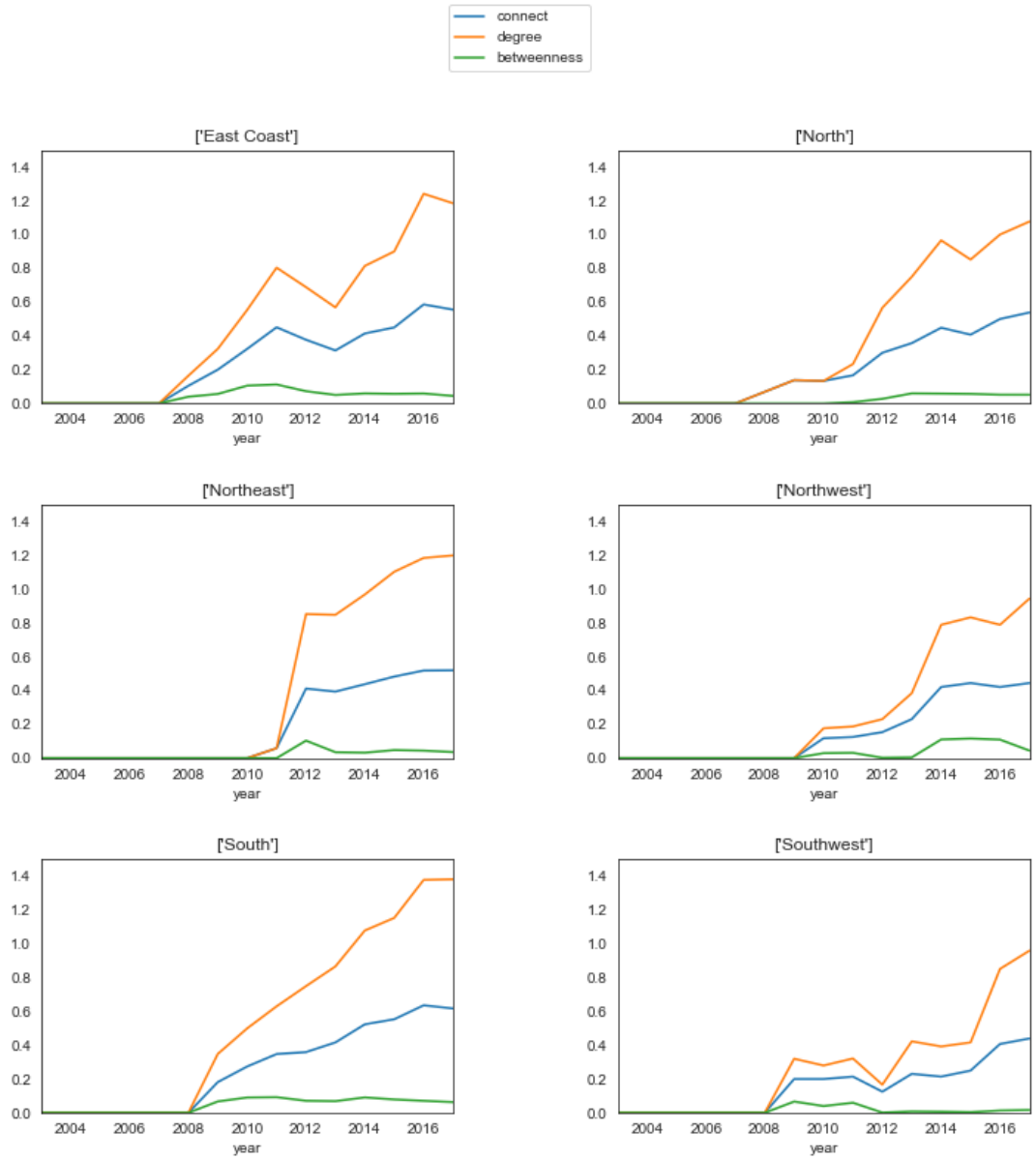
Measure Connectivity

Table: Different Measurements of Connectivity

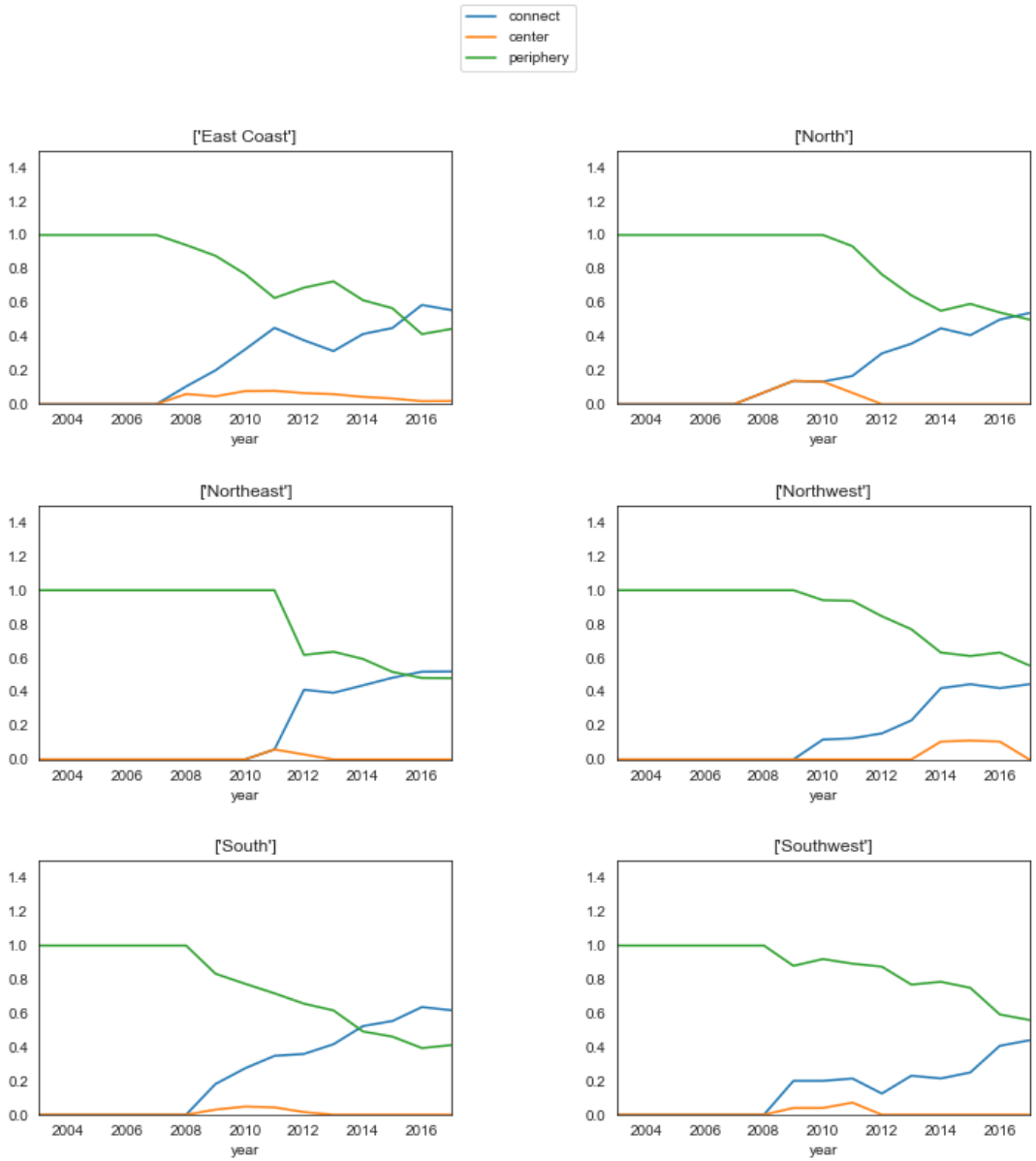
Variable	Definition	Calculation	Relation	FillNA
Dummy:				
Center	Nodes with eccentricity equal to radius	1 if center,0 if not	+	0
Periphery	Nodes with eccentricity equal to dismeter	1 if periphery, 0 if not	-	1
Continuous:				
Degree	# of neighbors	d(v)	+	0
Betweenness	# of shortest paths that pass node v	$\sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$	+	0

Eccentricity(node v): the largest distance between n and all other nodes in the same network
Radius: the minimum eccentricity of a network
Diameter: Maximum distance between any pair of nodes in the same network
Neighbors(node v): nodes that have direct connections with node v
FillNA: Fill the cities without HSR with this value

Heterogeneity in Connectivity Across Regions



Heterogeneity in Connectivity Across Regions



Methodology

Dealing with Endogenous issues:

- The Timing:
  - The start of construction: 12 out of 45 lines based on existing lines in 2005(Lin(2017)), Zheng and Kahn(2018)
  - The open date: depends on construction progress, which largely affected by engineering difficulties
  - Control for Location*Year fixed effect: Donaldson and Hornbeck(2016), Lin(2017), Giroud and Mueller(2015)
- The selection of cities:
  - Aim to connect all capital cities for each province(Ministry of Railway of China)
  - Cities connectivity measurements are calculated based their locations in HSR network
  - The Market Access Approach: Donaldson and Hornbeck(2016),Lin(2017)
  - ◂ This Paper

Benchmark Regression

$$Y_{i,t,k} = \sum_{k}^{4} \beta_k * X\_Spill_{i,t-1} * ind_k + \alpha_i + \gamma_k * \theta_t + \phi_l * \theta_t + \epsilon_{i,t}$$

- i stands for city, t stands for year, k stands for industry, l stands for region
- $Y_{i,t,k}$ is the dependent variable of interest, which includes log value and percentage composition of employment in different industries
- $X\_Spill_{i,t-1}$ is the independent varible that measures indirect connectivity
- standard errors are robust and clustered to city level

Table: Spillover Connectivity effects on Employment level

	Degree	Betweenness	Center	Periphery
Other_ns	0.07	0.48^{**}	0.22^{*}	-0.28^{***}
	(0.04)	(0.16)	(0.09)	(0.07)
Other_s	-0.07^{**}	-0.49^{***}	-0.23	0.59^{***}
	(0.03)	(0.15)	(0.15)	(0.07)
Skill	0.02	-0.47^{**}	-0.31^{**}	0.22^{***}
	(0.03)	(0.17)	(0.11)	(0.06)
Tourism	-0.06	-0.93^{**}	-0.57^{***}	0.22^{*}
	(0.05)	(0.30)	(0.13)	(0.09)
R^2	0.04	0.04	0.04	0.04
Adj. R^2	0.01	0.01	0.01	0.01
Num. obs.	12124	12124	12124	12124

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
standard errors are robust and cluster to city level

Table: Spillover Connectivity Effects on Industry Compositions

	Degree	Betweenness	Center	Periphery
Other_ns	2.93^{**}	12.33^{**}	6.23^{*}	-11.71^{***}
	(1.09)	(4.09)	(2.50)	(2.57)
Other_s	-2.87^{**}	-21.65^{***}	-10.06^{**}	23.82^{***}
	(0.91)	(4.48)	(3.70)	(2.49)
Skill	0.04	-14.87^{***}	-7.84^{**}	11.33^{***}
	(0.59)	(3.98)	(2.86)	(1.85)
Tourism	-0.10	-12.79^{***}	-7.03^{**}	11.68^{***}
	(0.51)	(3.68)	(2.43)	(1.82)
R^2	0.01	0.00	0.00	0.03
Adj. R^2	-0.02	-0.03	-0.03	-0.00
Num. obs.	12124	12124	12124	12124

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Additional Regression

$$Y_{i,t,k} = \sum_{k}^{4} \beta_{k,treated} * X_{i,t-1} * ind_k * D_{Treated_{i,t-1}}$$
$$+ \sum_{k}^{4} \beta_{k,other} * X_{i,t-1} * ind_k * D_{Other_{i,t-1}}$$
$$+ \alpha_i + \gamma_k * \theta_t + \phi_l * \theta_t + \epsilon_{i,t}$$

- i stands for city, t stands for year, k stands for industry, l stands for region
- $D_{Treated_{i,t-1}}$ and $D_{other_{i,t-1}}$ are dummy variables that distinguish different groups of cities
- standard errors are robust and clustered to city level

	Degree	Betweenness	Center	Periphery
Other_ns_t	0.06	−0.15	−0.05	−0.05
	(0.06)	(0.28)	(0.17)	(0.04)
Other_s_t	−0.19***	−0.63*	−0.52***	−0.22***
	(0.04)	(0.25)	(0.10)	(0.06)
Skill_t	0.05	0.13	−0.15	0.07
	(0.05)	(0.28)	(0.14)	(0.04)
Tourism_t	0.03	0.58*	0.54*	0.35***
	(0.06)	(0.26)	(0.21)	(0.07)
Other_ns_o	0.07*	0.69***	0.36**	0.06
	(0.04)	(0.21)	(0.12)	(0.04)
Other_s_o	−0.03	0.19	0.26	0.14**
	(0.02)	(0.16)	(0.17)	(0.05)
Skill_o	0.01	−0.03	−0.06	−0.08**
	(0.03)	(0.16)	(0.11)	(0.03)
Tourism_o	−0.09*	−0.80	−0.81***	−0.05
	(0.04)	(0.41)	(0.16)	(0.06)
R^2	0.04	0.04	0.04	0.05
Adj. R^2	0.01	0.01	0.01	0.03
Num. obs.	12124	12124	12124	12124

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

	Degree	Betweenness	Center	Periphery
Other_ns_t	4.45**	6.71	8.63*	2.23*
	(1.62)	(8.99)	(4.37)	(1.12)
Other_s_t	−5.72***	−14.24*	−11.56***	−4.80***
	(1.31)	(6.03)	(2.64)	(1.25)
Skill_t	1.24	6.79	2.20	2.13***
	(0.87)	(4.54)	(2.79)	(0.52)
Tourism_t	0.03	0.74	0.73	0.44
	(0.69)	(2.20)	(1.20)	(0.50)
Other_ns_o	2.43*	14.19*	4.98	−0.48
	(1.03)	(5.63)	(5.00)	(1.58)
Other_s_o	−1.93*	−7.69	0.20	2.60
	(0.85)	(4.89)	(5.04)	(1.66)
Skill_o	−0.36	−5.63**	−3.59***	−1.84***
	(0.52)	(2.17)	(1.02)	(0.49)
Tourism_o	−0.15	−0.87	−1.60*	−0.28
	(0.46)	(2.17)	(0.68)	(0.46)
R^2	0.01	0.00	0.00	0.01
Adj. R^2	-0.02	-0.03	-0.03	-0.02
Num. obs.	12124	12124	12124	12124

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Happy Chinese New Year!

The Spillover Effect

Spillover Effect from nearby cities:

$$X\_Spill_{i,t} = \sum_{c}^{5} Share_{c,i} * X_{c,t} \tag{1}$$

$$Share_{c,i} = \frac{log\_rgdp_{c,2002}}{\sum_{c \in N} log\_rgdp_{c,2002}} * \frac{log\_GeoDist_{i,c1}}{log\_GeoDist_{i,c}} \tag{2}$$

- Five neariest cities in terms geographic distance: ◂ Distance Distributions
- Share for each nearby city: ◂ Share Distributions

The Different Groups

The Three Groups: Following Giroud and Mueller(2015)

Treated Cities		Control Cities
Treated Details	Other	
Connected to HSR in year t-1, nearby cities can be connected or not Ex: A,B,C,D,E	Nearby cities connected to HSR in year t-1, not connected itself Ex: F	Not connected to HSR, no nearby cities get connected in year t-1 Ex: G

Note: In most of the cases(719 episodes), has HSR means near HSR, only in 15 episodes, the city has HSR but not near HSR

Figure: Example

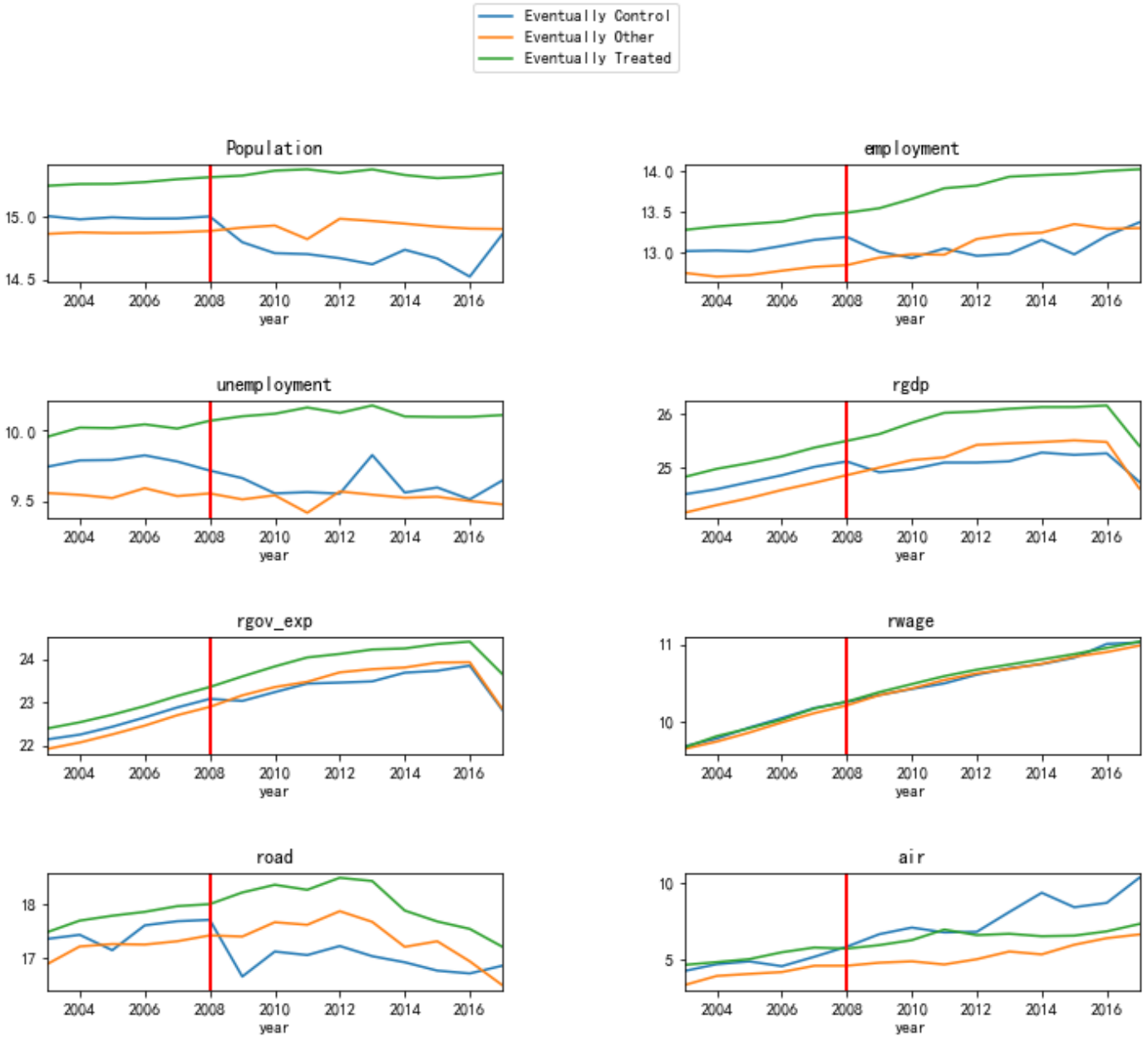
Summary Statistics
Back

Summary Statistics among Groups

Variables	Eventually control	Eventually Other	Eventually Treated
Count	435	1269	1752
Population	14.89	14.91	15.32
	0.81	0.62	0.58
employment	13.06	13.02	13.67
	0.69	0.58	0.86
unemployment	9.72	9.53	10.08
	0.74	0.66	0.79
rgdp	24.89	24.99	25.64
	0.91	0.82	1.02
rgov_exp	22.84	23.13	23.58
	0.93	0.81	1.04
rwage	10.19	10.4	10.44
	0.46	0.46	0.51
airplane_usage	5.84	5	6.1
	6.32	6.08	7.12
road_usage	17.3	17.31	17.93
	1.66	1.62	1.42

Summary Statistics among Groups

Back



Distance Distributions

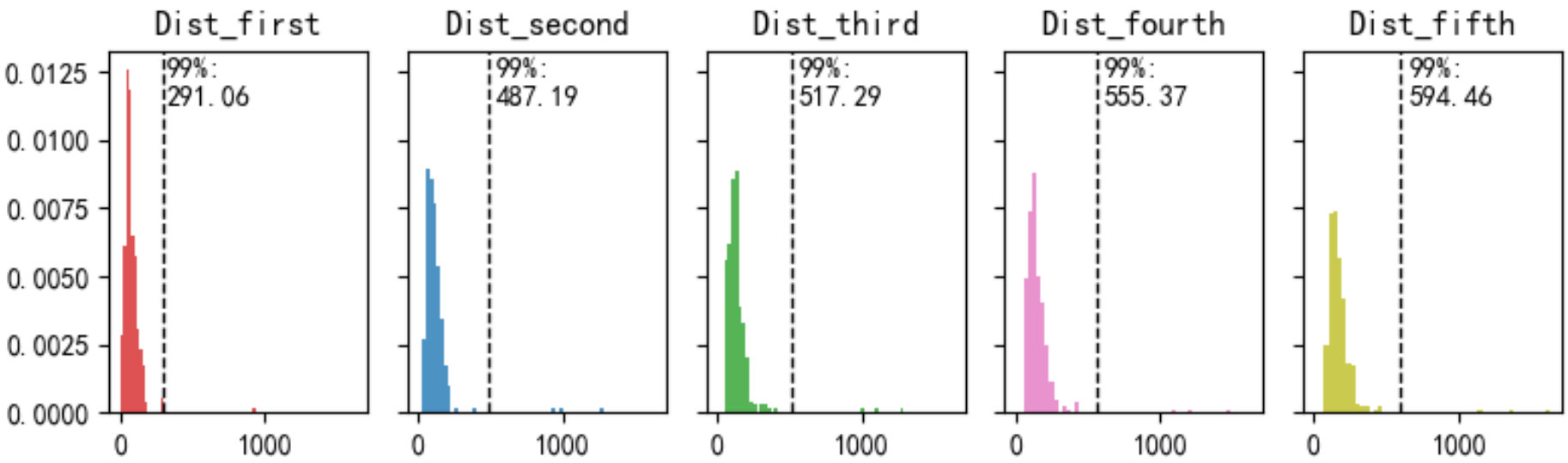
Back

Table: Distance(km) Summary Statistics

	Dist_first	Dist_second	Dist_third	Dist_fourth	Dist_fifth	total	aver_distance
count	282	282	282	282	282	282	282
mean	82.16	115.52	139.39	160.28	179.67	677.02	135.40
std	66.82	109.83	116.74	129.15	140.78	542.49	108.50

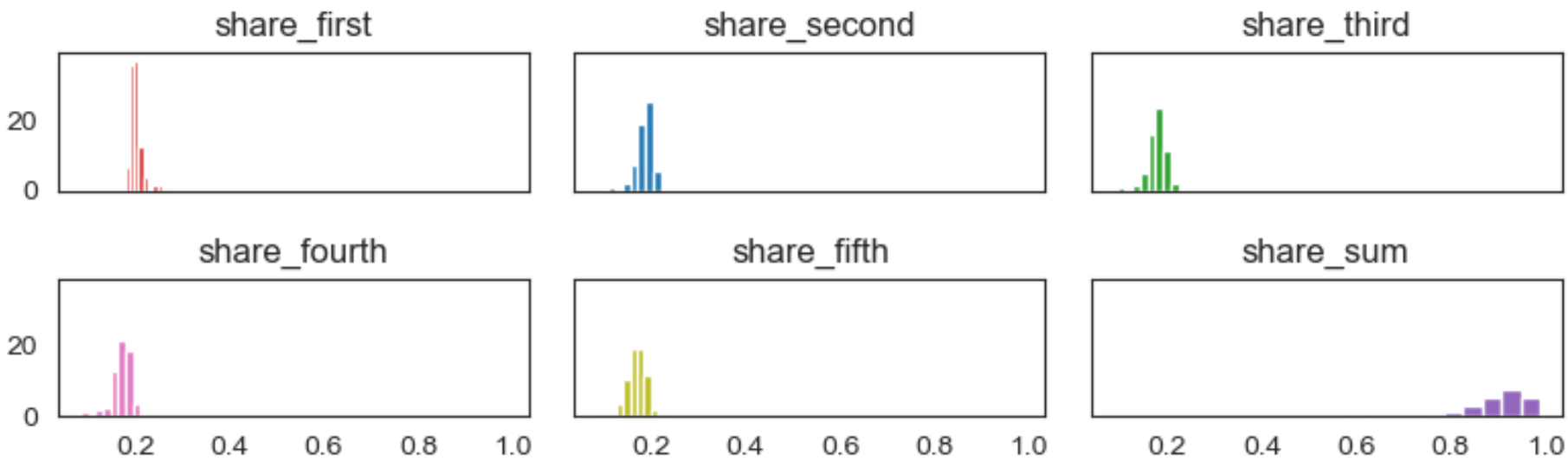
Note: 1 km = 0.62 miles

Figure: Frequency Histogram of Distance(km) in Order



Share Distributions

Figure: Distributions of Share



Back

The Treated Group

Figure: Example

The Treated Group

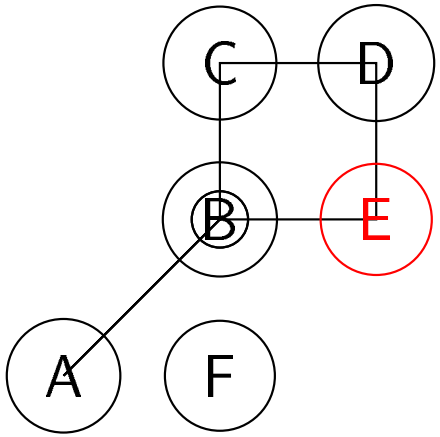


Figure: Example

The Treated Group

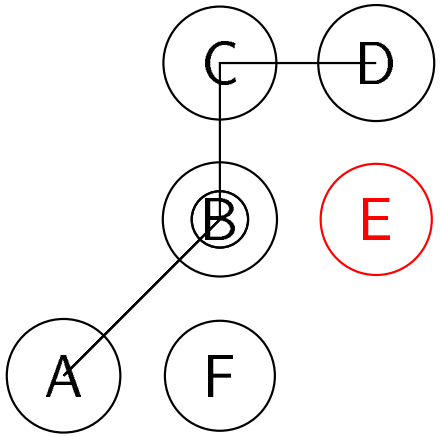
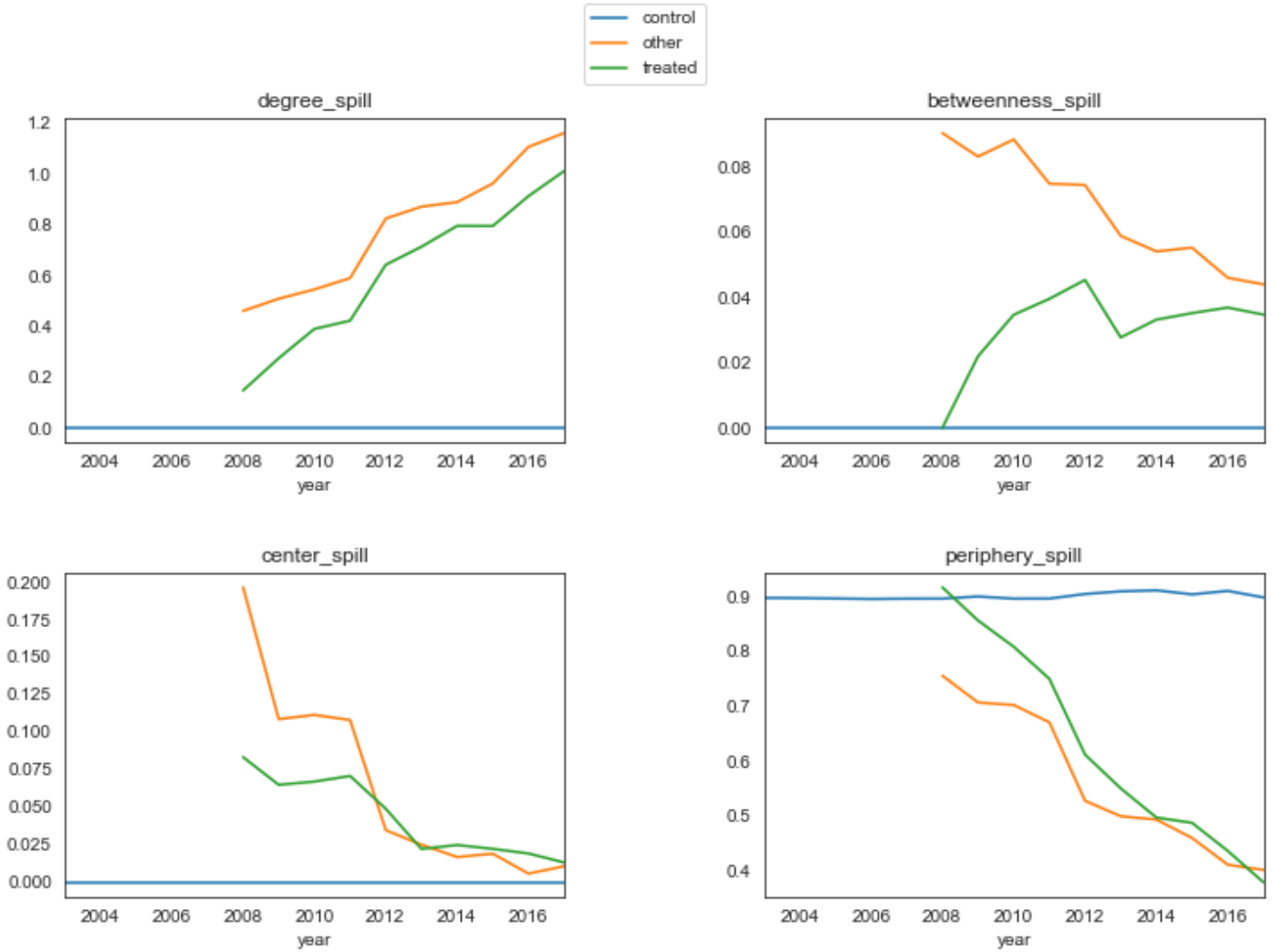


Figure: Example

Summary Statistics for Measuring Connectivity

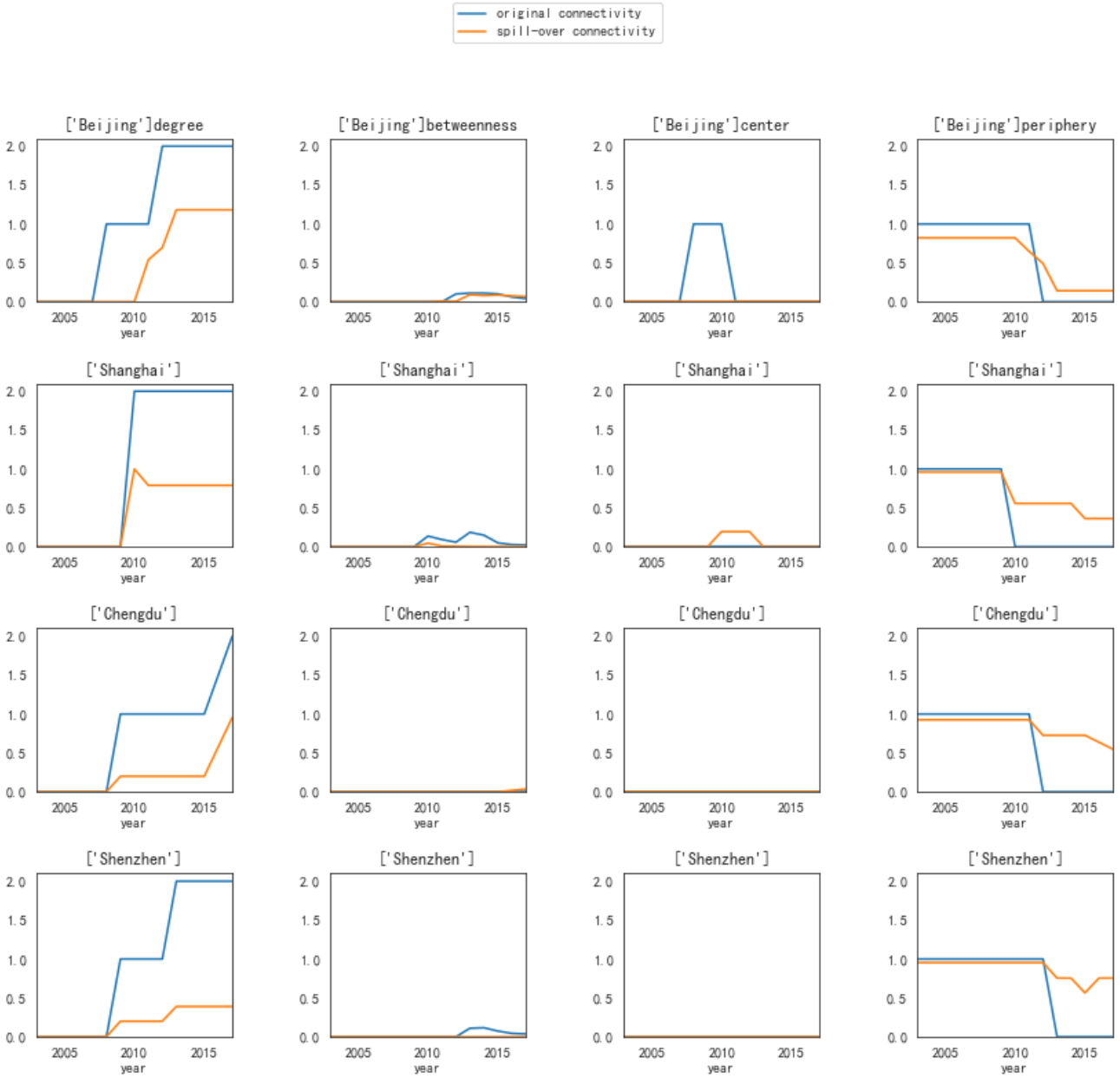
	degree_spill	degree	betweenness_spill	betweenness	center_spill	center	periphery_spill	periphery
count	3031	3031	3031	3031	3031	3031	3031	3031
mean	0.3	0.37	0.02	0.03	0.02	0.02	0.76	0.84
std	0.48	0.8	0.05	0.1	0.06	0.13	0.23	0.37
min	0	0	0	0	0	0	0	0
max	2.41	5	0.35	0.67	0.49	1	0.99	1

Change of Connectivity among Groups



Examples from the Big Cities

Back

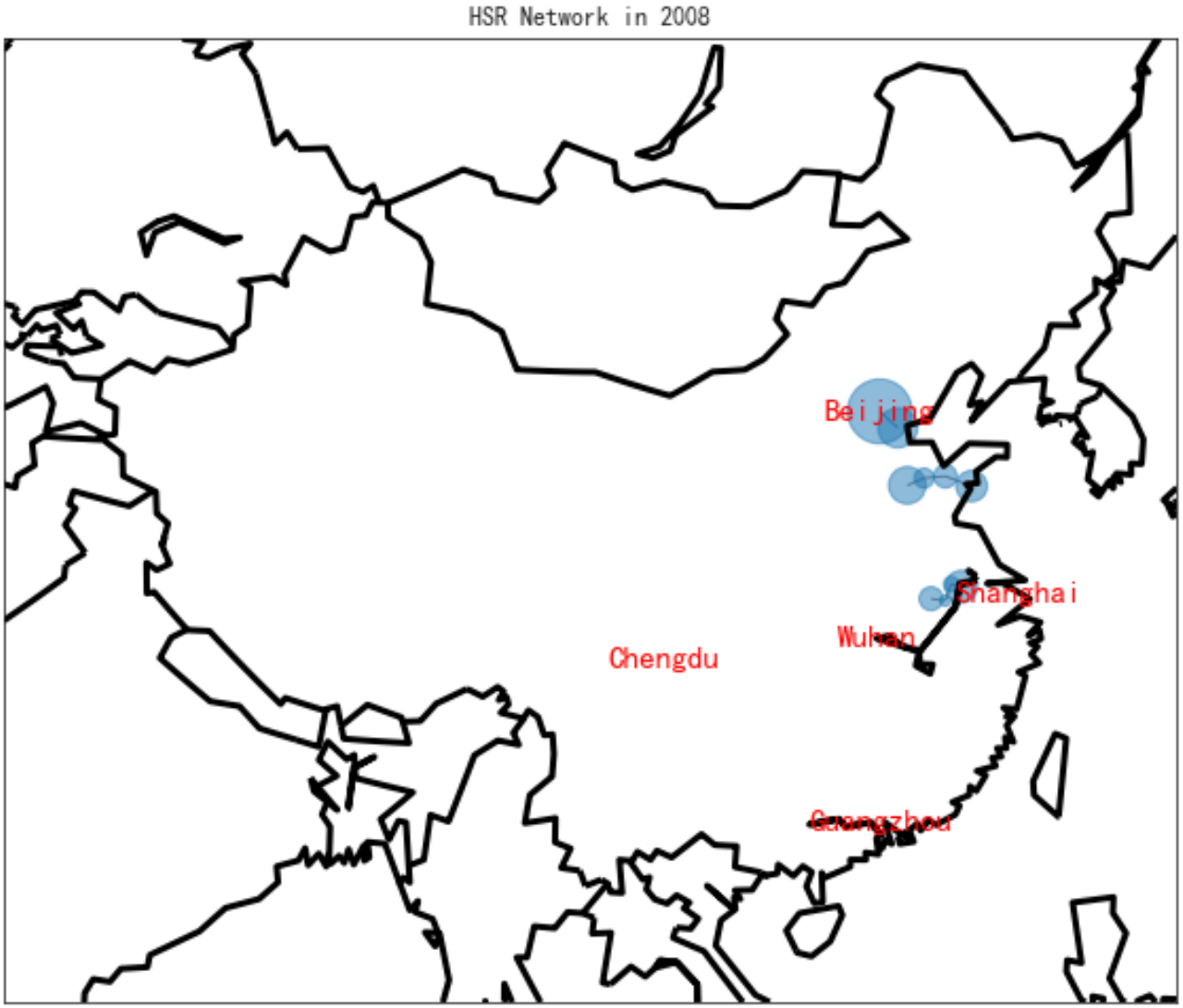


Facts

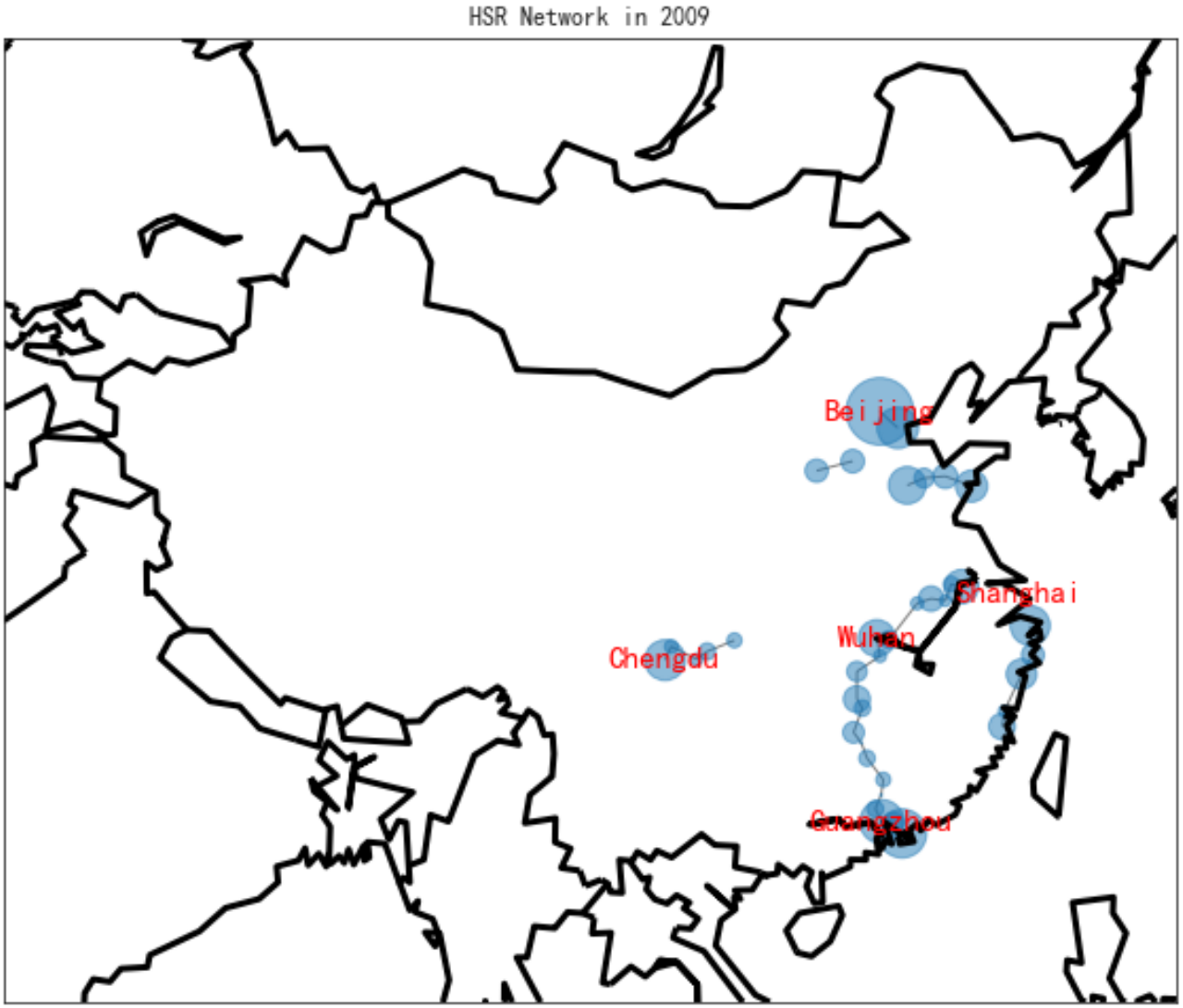
- The Project: (www.gaotie.cn)
  - Largest network in high-speed rail, almost covers two-thirds of the world's commercial high-speed rail tracks
  - Speed: 200 km/h to 380 km/h (155 MPH to 236 MPH)
  - Four Verticals and Four Horizontals to Eight Verticals and Eight Horizontals: ◂ The Network Growth
- The Usage:
  - The advantages over other methods
  - HSR has been replacing bus and cars in shorter distance trips and airlines and conventional railways for longer distance trips.(Lin(2014))
- The Passengers: (Jianbin(2011),Wu et al.(2013),Olivier et al.(2014))
  - Average monthly Income ranges from 4300 to 6700 yuan. (High income group).
  - A large proportion of the passenger travels for business purposes. (Shorter lines: 25% to 40%; longer lines: 40% to 60%)

◂ Back

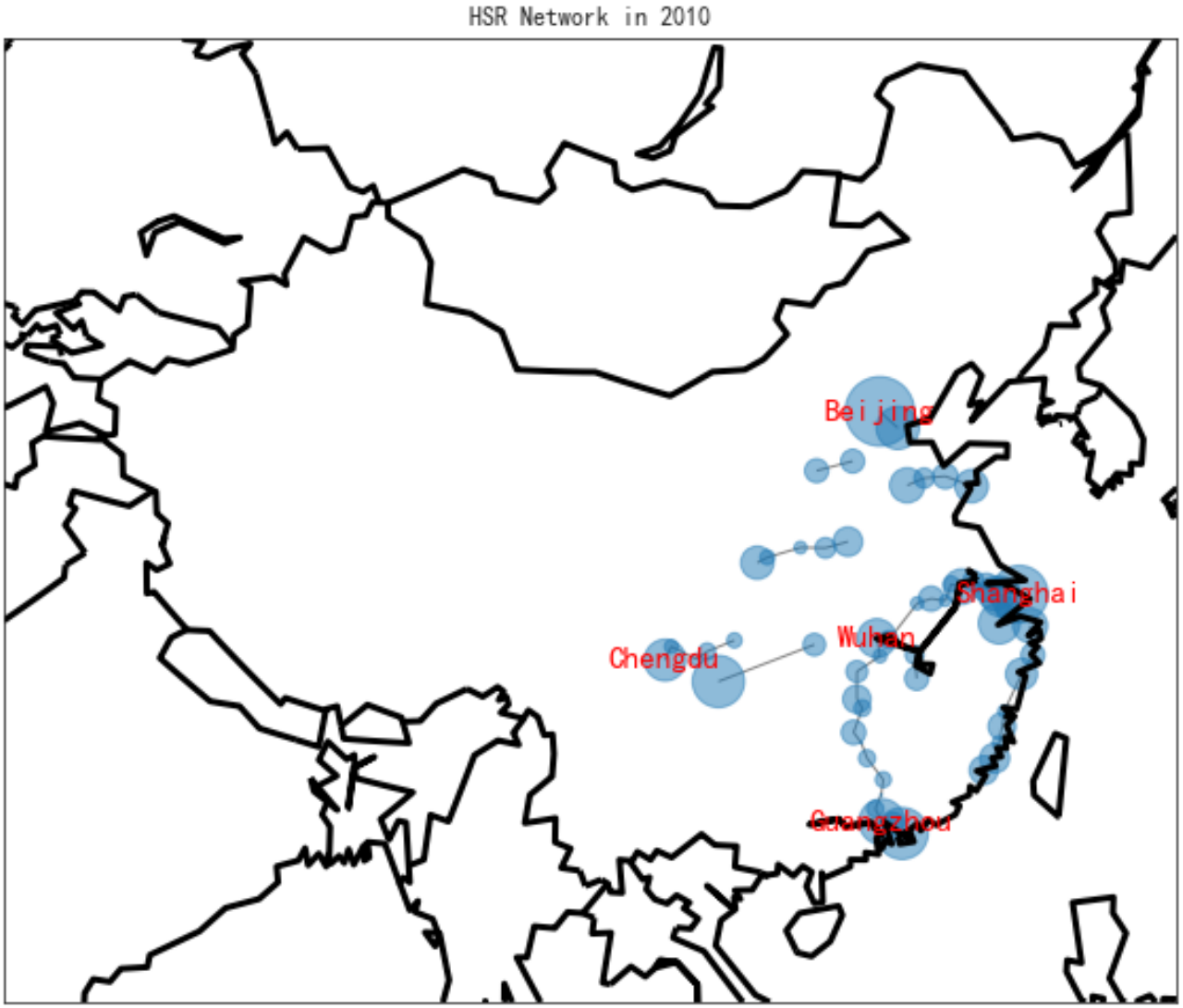
The Network


HSR Network in 2008
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

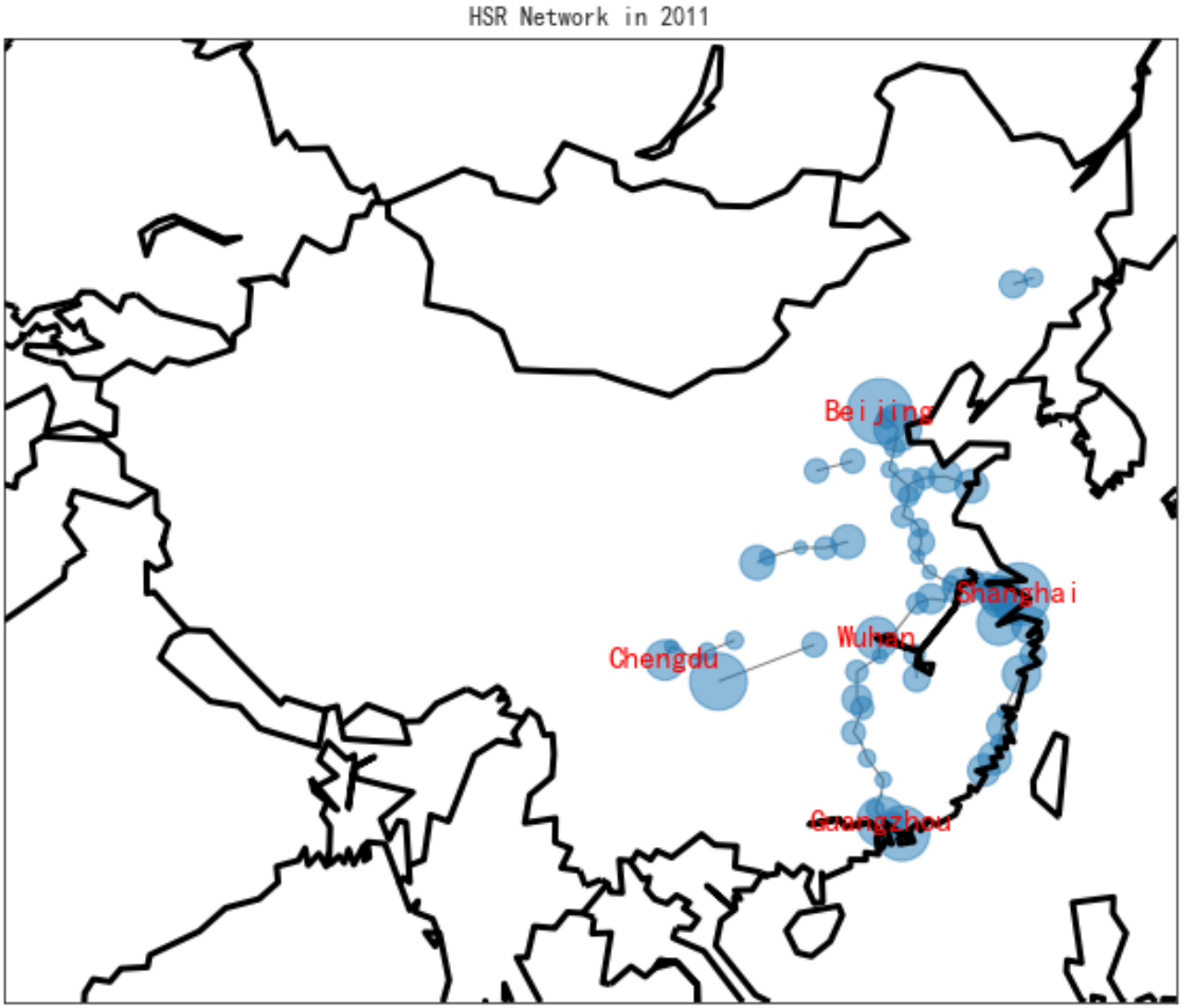
The Network


HSR Network in 2009
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

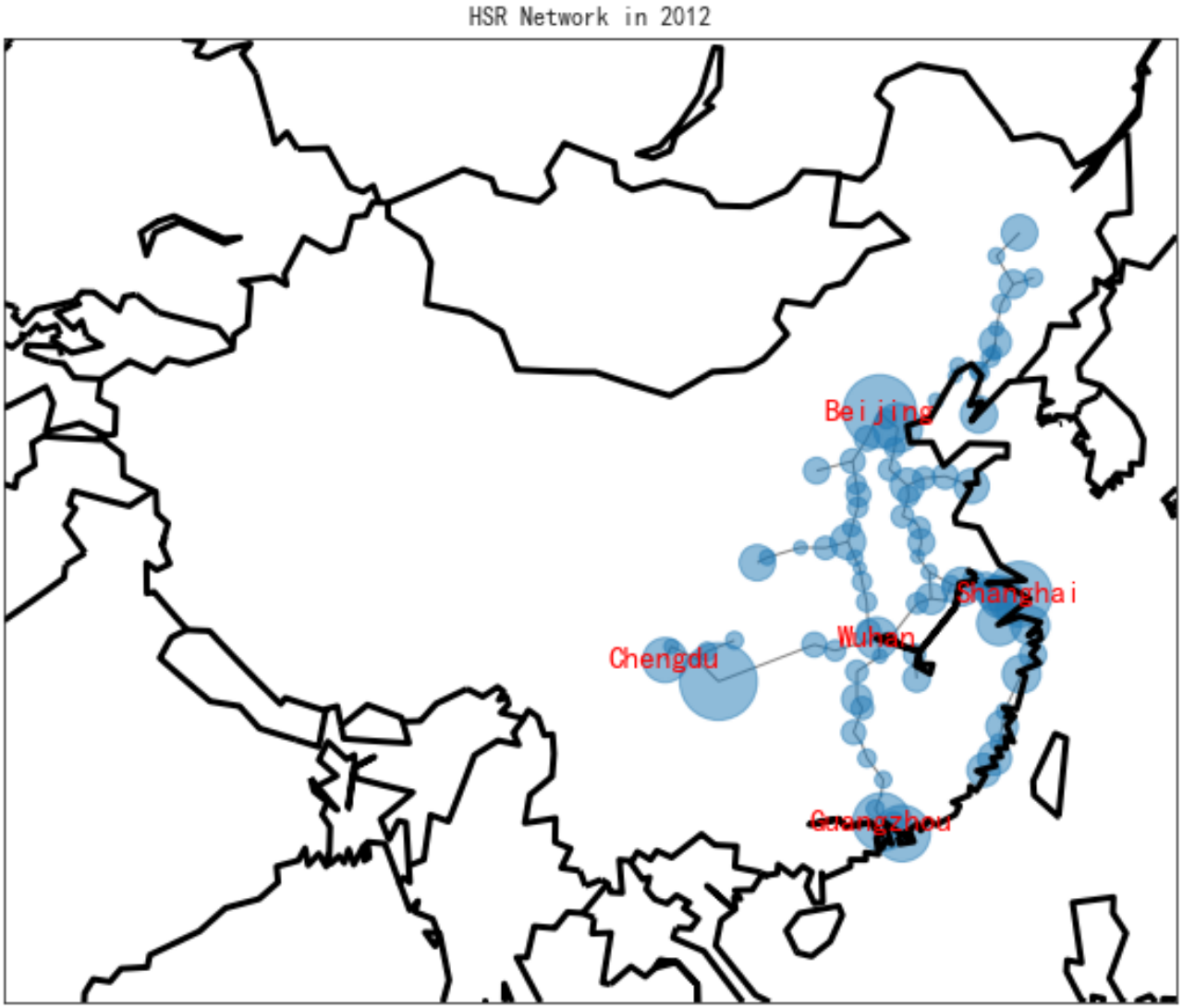
The Network


HSR Network in 2010
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

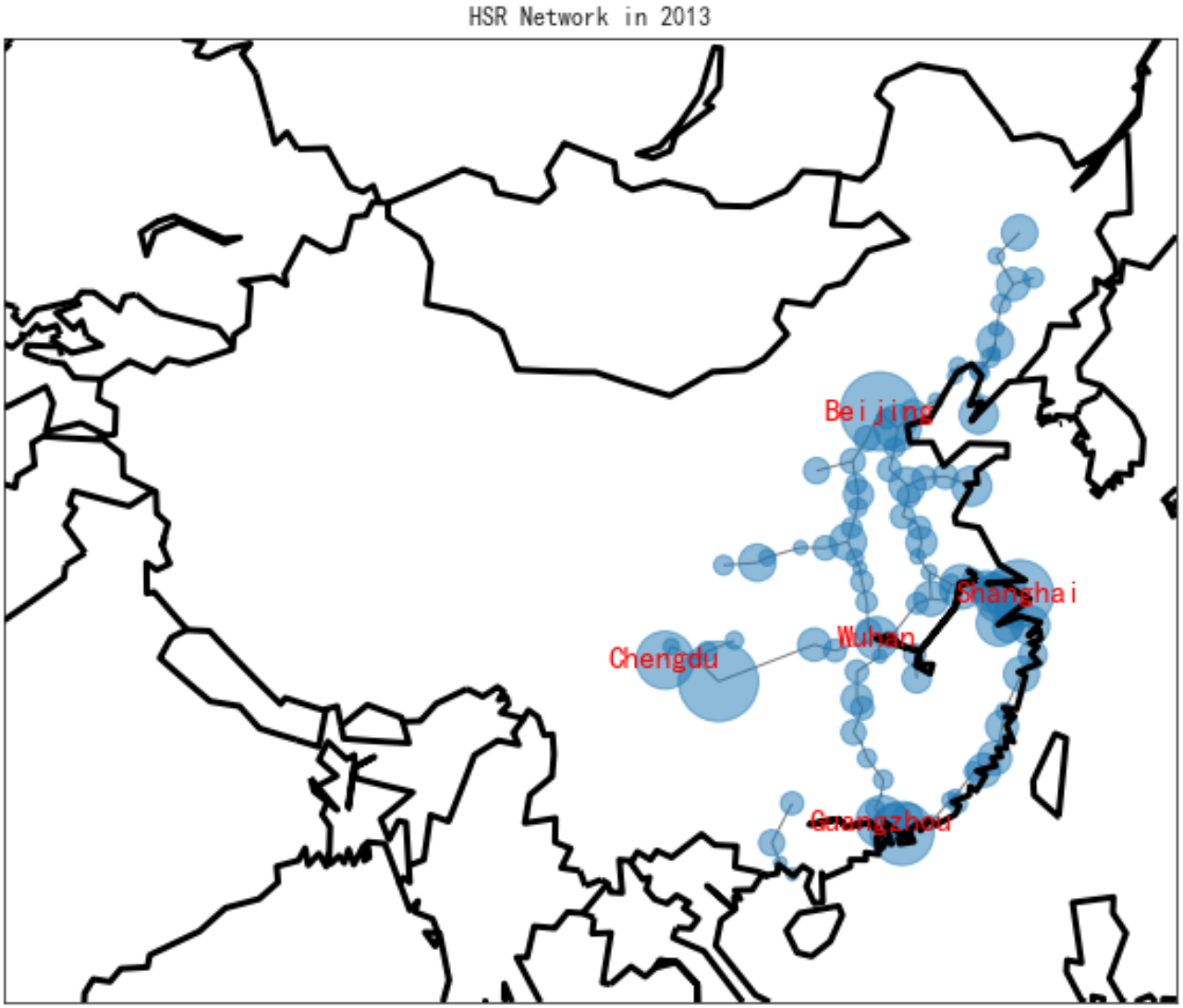
The Network


HSR Network in 2011
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

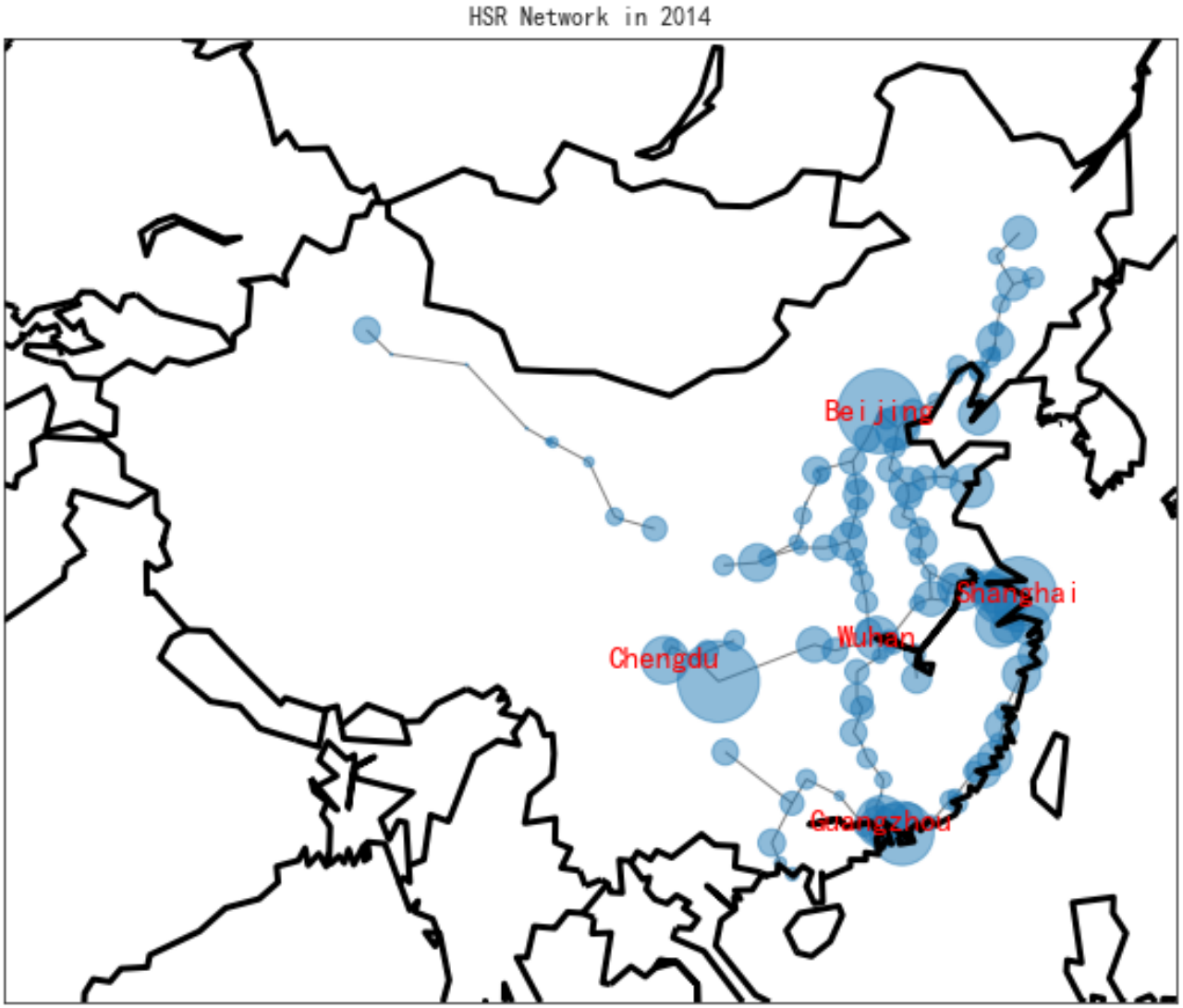
The Network


HSR Network in 2012
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

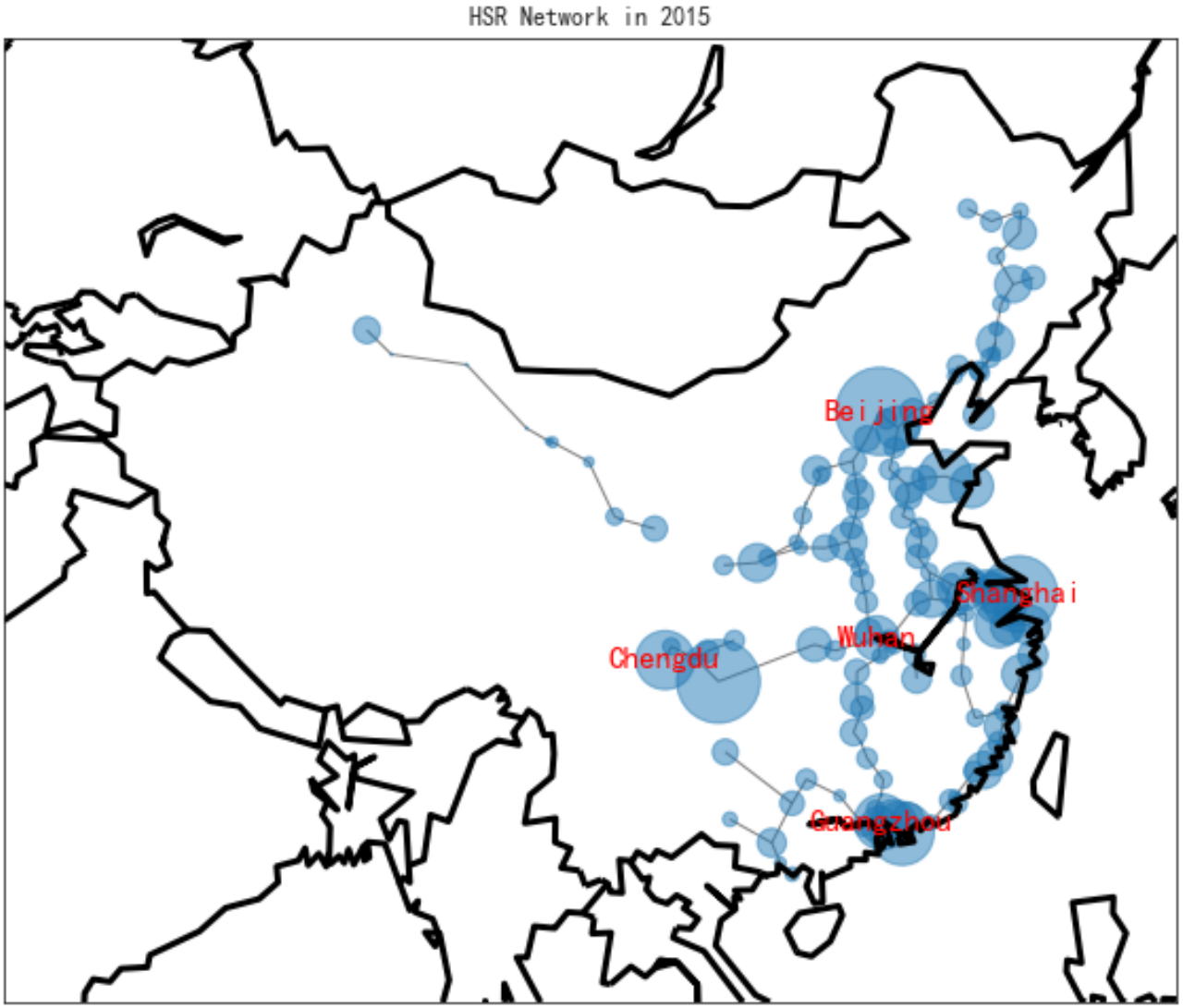
The Network


HSR Network in 2013
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

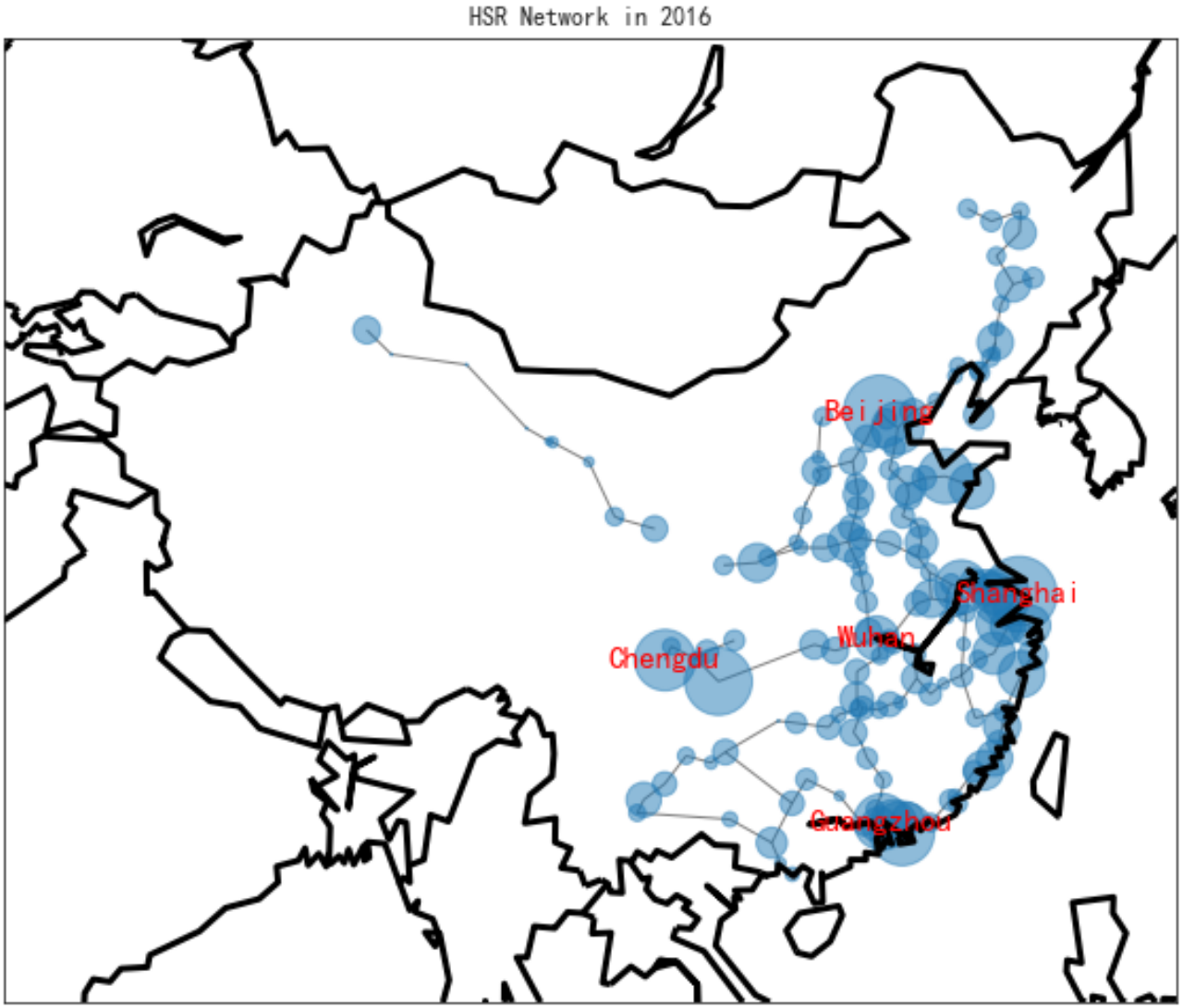
The Network


HSR Network in 2014
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

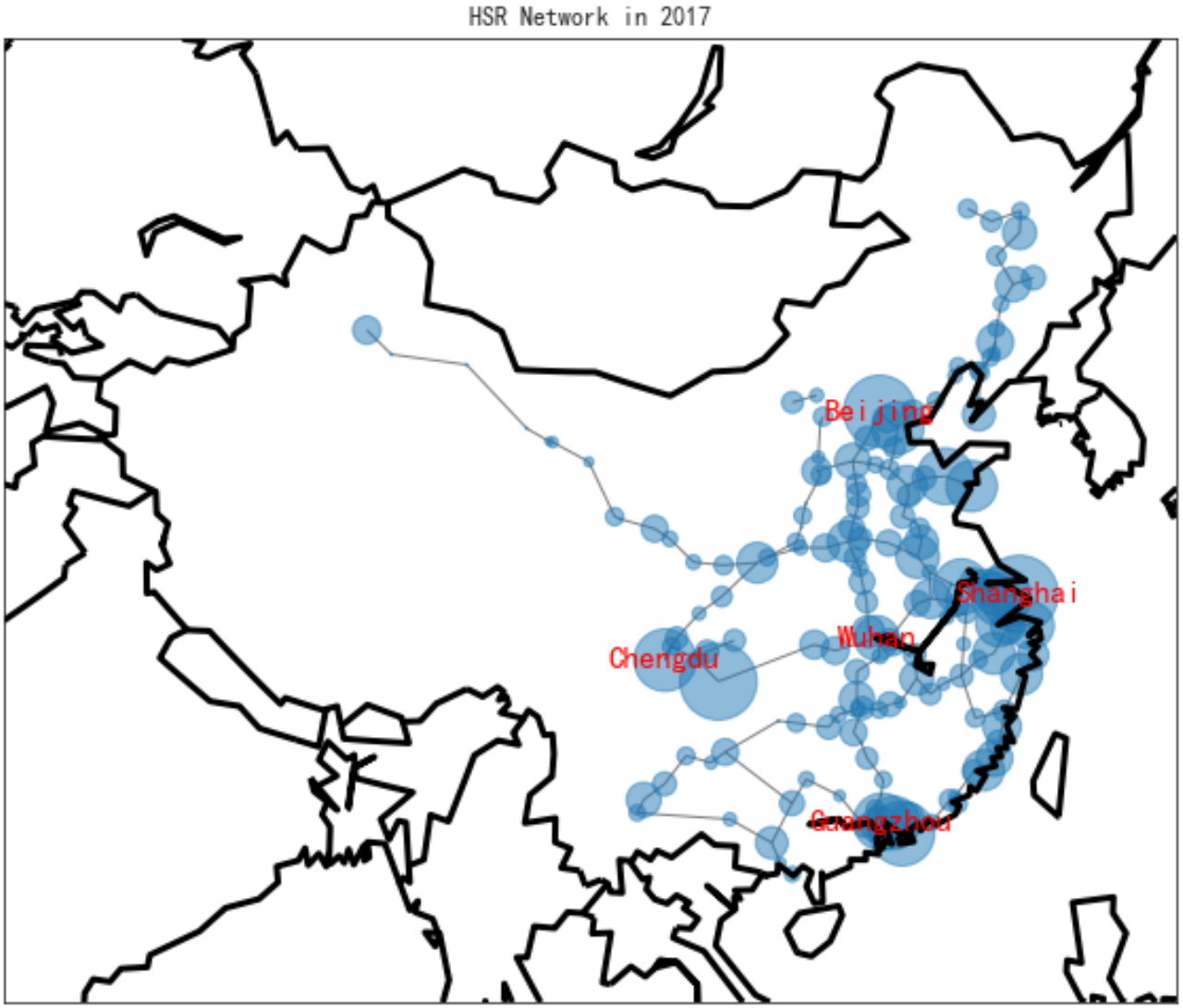
The Network


HSR Network in 2015
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

The Network


HSR Network in 2016
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

The Network


HSR Network in 2017
Beijing
Shanghai
Wuhan
Chengdu
Guangzhou

The Network

Back

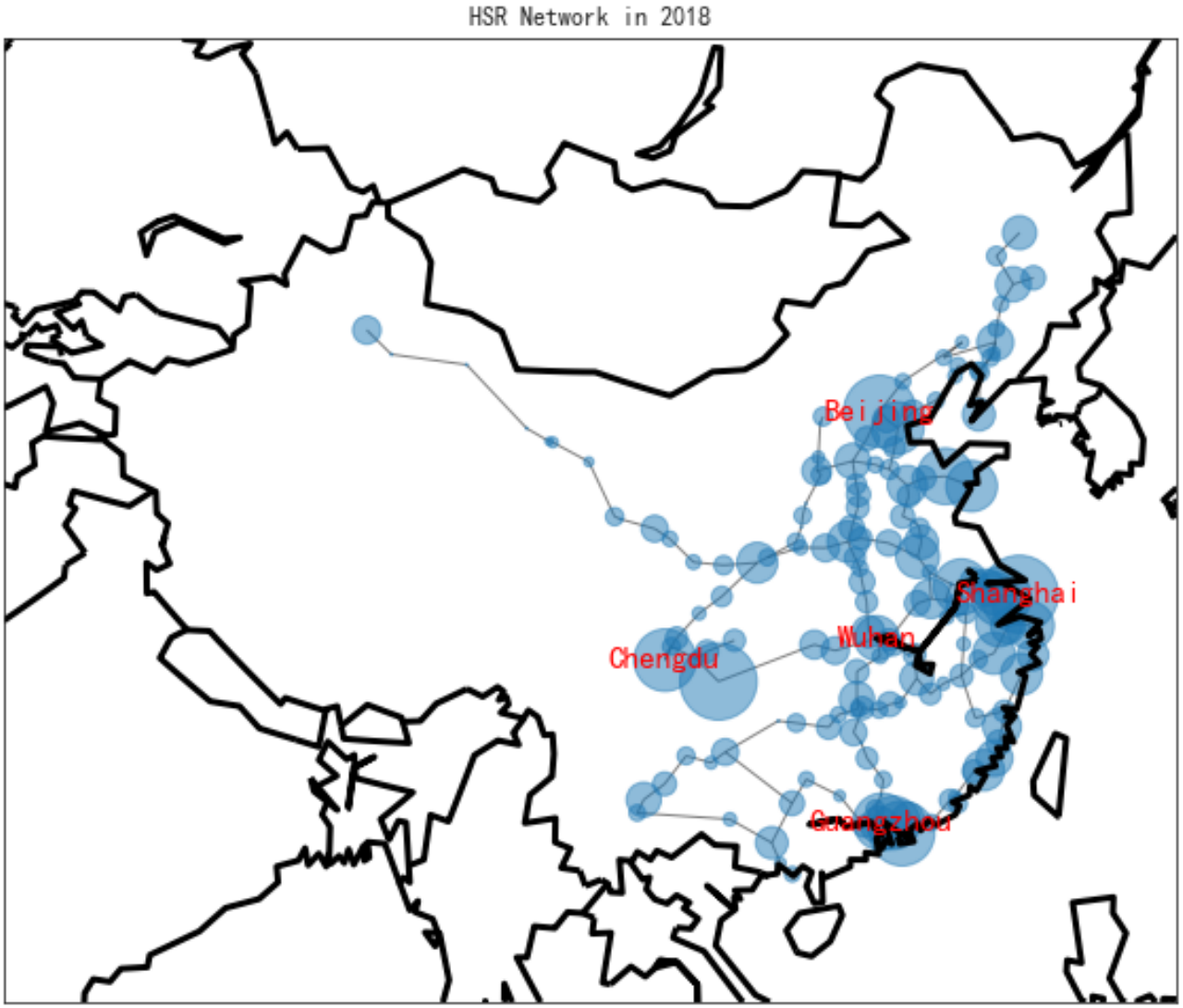


Table: 19 Industries in China

Back

Chinese Industries	US industries	NAICS
Skilled Employment:		
Finance and insurance	Finance and insurance	52
Real estate	Real estate and rental and leasing	53
Information transmission, computer service and software	Information	51
Leasing and Business service	Professional,scientific and technical services	54
Scientific research,technical service	Professional,scientific and technical services	54
Culture,sports and entertainment	Arts,entertainment and recreation	71
Tourism-related employment:		
Hotels and catering service	Accommodation and food service	72
Other service employment:		
Transportation,warehousing and post	Transportation,warehousing	48
Management of water conservancy,environment and public facilities		
Household services, repair and other service		
Education	Educational Service	61
Health, social work	Health care and social assistance	62
Public Management and Social 0rganization		
Wholesale and retail trade	wholesale and retail trade	42,43
Other non-service employment:		
Agricultur, forestry,animal productionand hunting,fishing	Forestry,fishing,hunting and agriculture support	11
Mining and quarrying	Mining	21
Manufacturing	Manufacturing	31
Production and Distribution of Electricity, Gas and Water	Utilities	22
Construction	Construction	23