Social Connections and the Spatial Spread of COVID-19

IIES, Stockholm University

May 31st, 2022

Research Question

How well can social connections be used to study the spread of infectious diseases?

- KNOWN: network data on human social interactions are more informative than geographic proximity
- UNCLEAR: which type of social interactions matters more for early stage spread and later-on local transmission

Setting: the Spread of COVID-19 in China in Early 2020

- Jan 23rd-March 23rd for all Chinese cities
- Combine social media network and travel network to measure social connections
- Three steps
  1. When did the first COVID-19 case show up?
  2. How wide spread was the subsequent local transmission?
  3. How does the interplay of travel and information driven by social connections matter for local transmission?

Data and Measurements

Social Connections (cross-sectional)

- Social Media Connection: 2013 snapshot of Weibo network created by Qin et al. (2021)
- Baidu Travel Connection: aggregated average value at city-to-city level from Baidu Migration during Jan 1st-Jan 23rd 2020
- Cellphone Travel from Wuhan:estimated population movement from Wuhan during Jan 1st-Jan 23rd 2020

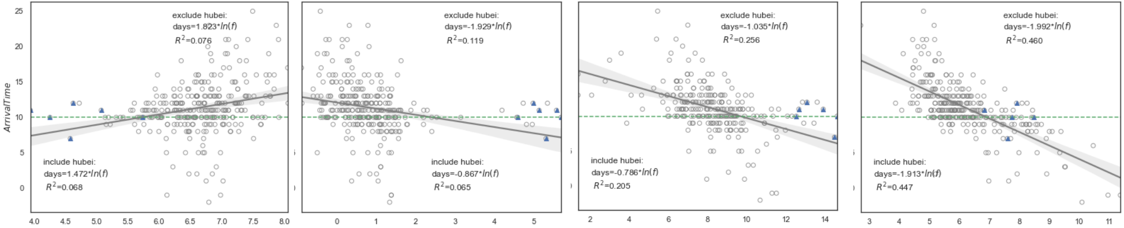
Infection and social distancing (city daily panel)

- Daily infections outside Wuhan (from DXY)
- Social Distance index computed using Baidu Within-City Traffic Index

▸ time line ▸ distribution of connection to wuhan by network ▸ PCA decomposition of connections

Predicting the Arrival Time

- Social media connections to Wuhan outperforms travel connections in predicting the arrival (higher R2)



(a) $\ln(Geographicdistance)$ (b) $BaiduTravelConnection$ (c) $CellphoneTravelConnection$ (d) $SocialMediaConnection$

- It also implies stronger transmission of COVID-19 (1.6 $\sim$ 2 days faster per SD) (bigger slope) table

Predicting the Accumulated Number of Infections

- Travel connection to Wuhan is a better predictor than social media connections (higher R2)

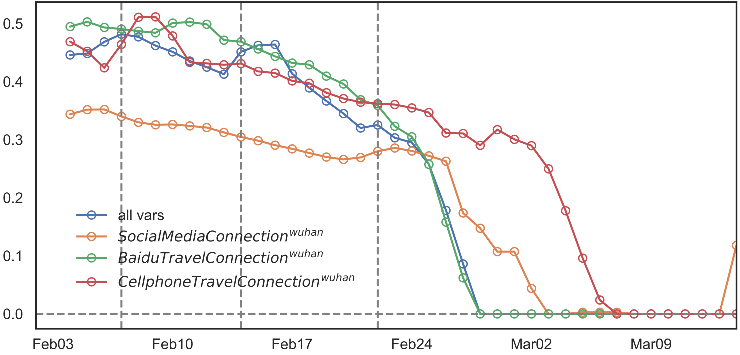


Figure: R-squared of prediction on number of infection using random forest regression

The Dual Effects of Infection Import Exposure (1/4)

To decompose by interacting the connection measures with travel restrictions

$$TravelExposure^k_{n,t} := \sum_{m \neq n} p^k_{nm}(1 - Lockdown_{m,t-1})Infections_{m,t-1} \tag{1}$$

$$CommunicationExposure^k_{n,t} := \sum_{m \neq n} p^k_{nm} Infections_{m,t-1} \tag{2}$$

$$\Delta y_{c,t} = \alpha Infections_{n,t-1} + \gamma^k TravelExposure^k_{n,t-1} + \theta^k CommunicationExposure^k_{n,t-1} + X_{n,t}\Phi + FEs + \varepsilon_{n,t}$$

- $\theta^k CommunicationExposure^k_{n,t-1}$: effect under lock-down (no travel outflow from export cities)
- $\gamma^k TravelExposure^k_{n,t-1}$: additional effect when there is no travel restrictions example

Results

- Travel speeds up spread while communication slows down spread

effects on infection
effects on social distancing

Conclusion and Take away

To conclude

- Construct a Social Media Connection index between Chinese cities using Weibo data
- Nott only a measure of social contact proximity but also a conduit of information
  - outperform travel connections in predicting when and where
  - can both speed up (travel) and slow down (communication) the spread

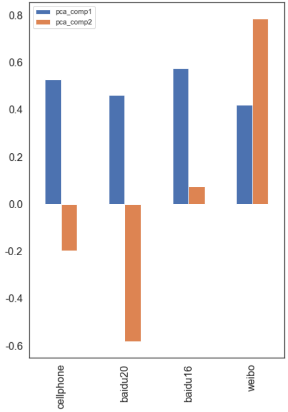
Take Away

- Quick identification of high-risk regions upon initial outbreak

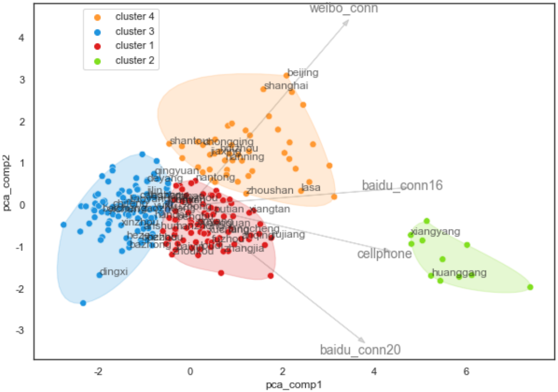
Appendix

PCA Decomposition

back

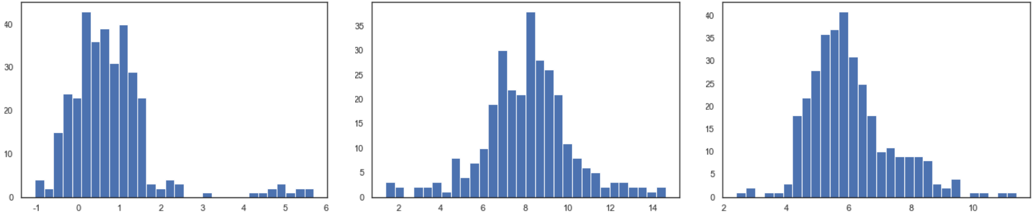


(a) factor loading



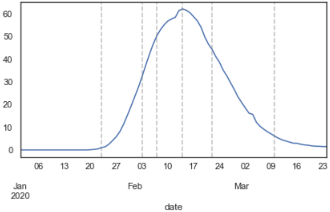
(b) K-means clustering on PCA components

Distribution of Connections to Wuhan by Network ▸ back

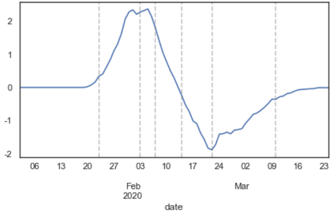


(a) $BaiduTravelConnection^{wuhan}$ (b) $CellphoneTravelConnection^{wuhan}$ (c) $SocialMediaConnection^{wuhan}$

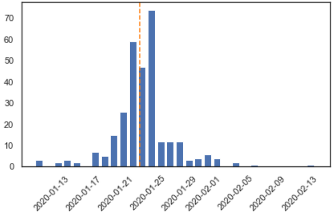
Timeline ▸ back



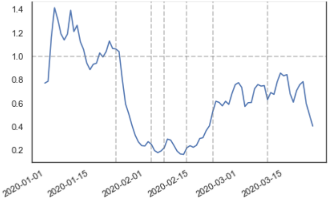
(a) avg current daily infection



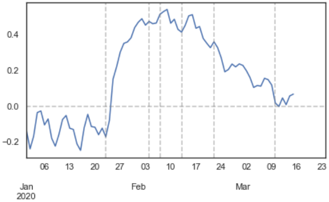
(b) avg change in daily infection



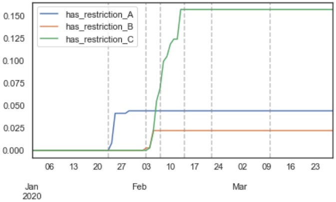
(c) Distribution of arrival date



(d) outflow travel index



(e) social distancing index



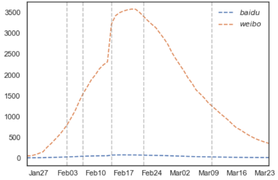
(f) social distancing index

Predicting the Arrival Time ▸ back

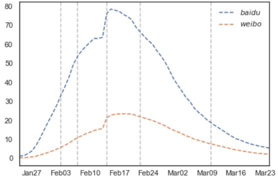
VARIABLES	(1)	(2)	(3)	(4)	(5)
$SocialMediaConnection^{wuhan}$	-2.6039***			-2.1878***	-2.0373***
	(0.187)			(0.238)	(0.265)
$BaiduTravelConnection^{wuhan}$		-1.2324***		-0.0738	-0.2329
		(0.234)		(0.257)	(0.284)
$CellphoneTravelConnection^{wuhan}$			-2.0960***	-0.6656**	-0.3316
			(0.219)	(0.321)	(0.414)
Observations	244	243	244	243	243
R-squared	0.446	0.103	0.275	0.468	0.473
Mean	11.209	11.210	11.209	11.210	11.210
Control	NO	NO	NO	NO	YES

Two Examples on Infection Import Exposure

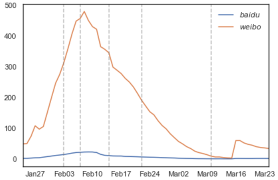
back



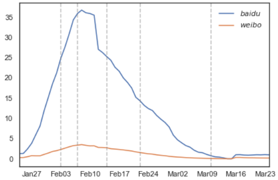
(a) $CommunicationExposure$, Guangzhou



(b) $CommunicationExposure$, Yongzhou



(c) $TravelExposure$, Guangzhou



(d) $TraveExposure$, Yongzhou

The Dual Effects of Infection Import Exposure on Newly Infections

▸ back

VARIABLES	(1) early	(2) early	(3) early	(4) later	(5) later	(6) later
$CommunicationExposure^{weibo}$	-1.4404***		-1.4032***	-0.2937***		-0.2479***
	(0.447)		(0.470)	(0.062)		(0.053)
$TravelExposure^{weibo}$	1.3883***		1.2727***	0.3376***		0.2914***
	(0.366)		(0.375)	(0.088)		(0.085)
$CommunicationExposure^{baidu}$		-0.7115***	-0.5829***		-0.2593***	-0.2261***
		(0.170)	(0.127)		(0.050)	(0.045)
$TravelExposure^{baidu}$		0.5336***	0.4284***		0.2892***	0.2281***
		(0.130)	(0.120)		(0.073)	(0.067)
Observations	6,224	6,224	6,224	8,475	8,475	8,475
R-squared	0.340	0.299	0.351	0.195	0.194	0.205
Mean	-0.029	-0.029	-0.029	-0.099	-0.099	-0.099
Controls	X	X	X	X	X	X
dateFE	X	X	X	X	X	X
SEcluster	City+date	City+date	City+date	City+date	City+date	City+date

The Dual Effects of Infection Import Exposure on Social Distancing

back

VARIABLES	(1) early	(2) early	(3) early	(4) later	(5) later	(6) later
$CommunicationExposure^{weibo}$	0.0897***		0.0900***	0.0095		0.0307
	(0.030)		(0.030)	(0.030)		(0.028)
$TravelExposure^{weibo}$	0.1129***		0.1157***	0.0055		-0.0137
	(0.030)		(0.031)	(0.028)		(0.028)
$CommunicationExposure^{baidu}$		-0.0267	-0.0360		-0.0872**	-0.0925**
		(0.035)	(0.033)		(0.038)	(0.038)
$TravelExposure^{baidu}$		-0.0310	-0.0334		0.0852***	0.0932***
		(0.033)	(0.030)		(0.030)	(0.032)
Observations	5,669	5,669	5,669	7,633	7,633	7,633
R-squared	0.887	0.882	0.888	0.912	0.913	0.913
Mean	0.006	0.006	0.006	0.002	0.002	0.002
Controls	X	X	X	X	X	X
CityFE	X	X	X	X	X	X
dateFE	X	X	X	X	X	X
SEcluster	City+date	City+date	City+date	City+date	City+date	City+date