An Investigation of Containment Measure Implementation and Public Responses to the COVID-19 Pandemic in Mainland China

Ji Liu†, Haoyi Xiong†, Xiakai Wang†, Jizhou Huang†,
Qiaojun Li†, Tongtong Huang†, Siyu Huang†, Haifeng Wang†, Dejing Dou†*

***Abstract*—While the COVID-19 epidemic expands among multiple countries, diverse measures have been exploited to halt the spread of COVID-19. In Mainland China, the containment measures can be categorized into two types, i.e., intra-city quarantine and isolation, and inter-city travel restriction. Both information acquisition and local economy play an important role while implementing the measures. In order to understand the relationship between the containment measures and pulic responses to the COVID-19 pandemic, we study the correlation among three factors, i.e., the information acquisition of containment measures, the public responses to the COVID-19 pandemic, and local economy of cities in Mainland China. We combine Markov Chain Monte Carlo (MCMC) and SIR-X to estimate the parameters related to the pandemic. Then, we investigate the correlations among multiple representative parameters including mobility, local economy, and information acquisition to understand the implementation of containment measures. We utilize the mobility data from Baidu Maps, the COVID-19 related search frequency data from Baidu Search Engine, and the data of Gross Domestic Product (GDP). From the analysis, we evidence that the the information acquisition is strongly correlated with the local economy. In addition, we find that the cities with stronger local economy have bigger inflows from Wuhan, while the citizens of the cities perform COVID-19-related searches more frequently and take the quarantine measure more strictly.**

***Index Terms*—COVID-19, P-value, SIR-X model, MCMC**

I. Introduction

Since the beginning of 2020, COVID-19 has quickly spread all around the world [1], [2], such as China [3], European countries [4], and United States [5]. In Mainland China, the total number of confirmed cases rapidly increased from about 500 to 84.5 thousand in the first five months of 2020. A bunch of containment measures have been carried by Chinese government in order to fight against the COVID-19 pandemic in Mainland China [6]–[8]. Some of the containment measures have also been implemented in other countries [9], [10].

In Mainland China, there are two types of containment measures, i.e., inter-city and intra-city. The intra-city measures refers to staying in hospitals or at home [11] when a citizen is suspected or confirmed to be infected by COVID-19 virus. This type of measure is denoted by the "quarantine" measure. Other measures are implemented, such as stay-at-home, closed schools [12]. Furthermore, strict travel restrictions were carried out in Wuhan, i.e., all transports were banned from the 23 January 2020. As shown in Figure 1, the outflows from Wuhan were significantly reduced. Inflows and outflows refer to the number of people entering and leaving a city. Figure 2 shows that inter-city travel was significantly reduced as well in order to reduce infection while the national spring vacation was prolonged.

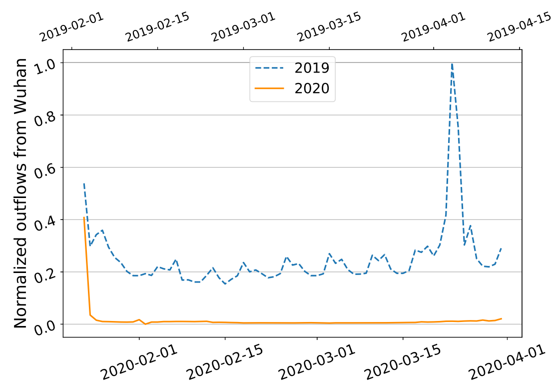


Fig. 1: The normalized outflows from Wuhan in 2019 and 2020, which is aligned based on the lunar calendar as the Spring Festival depends on the lunar calendar and has a big impact on the outflows, in Mainland China. The figure shows that the outflows are much reduced in 2020 compared with those in 2019.

Citizens generally utilize mobile applications, e.g., Baidu Migration [1] and search engines, e.g., Baidu [2], to get COVID-19 related information. Mobile applications are proven to be feasible for information acquisition [3], [13]–[19], while the search records correspond to the status of information acquisition and the statistical migration data reflect migration behavior. In order to understand the travel behavior of well-informed individuals, we study the correlation between the migration data and information acquisition. Moreover, we analyze the correlation between the information acquisition and the local economy to reveal the COVID-19 related search

*Corresponding author: doudejing@baidu.com
†Baidu Inc., Beijing, 100085, China.

[1] Baidu Migration - http://qianxi.baidu.com/
[2] Baidu - https://www.baidu.com/

978-1-6654-1685-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDH52753.2021.00046

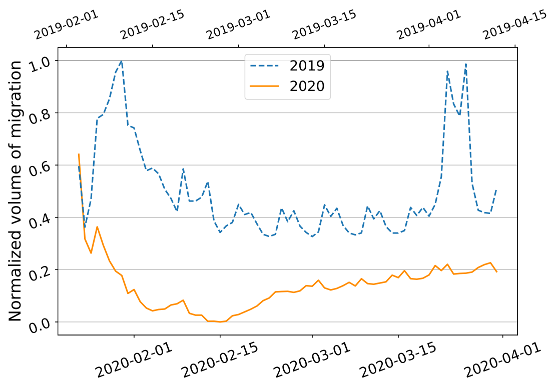


Fig. 2: The normalized volumes of migration in 2019 and 2020, which is aligned based on the lunar calendar as the Spring Festival depends on the lunar calendar and has a big impact on the volume of migration, in Mainland China. The figure shows that the volume of migration is much reduced in 2020 compared with those in 2019.

behavior in different economic situations.

In this paper, we combine a SIR-X model and Markov Chain Monte Carlo (MCMC) [20] methods to infer the representative parameters of the COVID-19 pandemic and the implementation of containment measures, in major cities of Mainland China. Afterward, in order to analyze the relationship among divers factors, we investigate the correlation among different parameters, i.e., local economic situation (GDP per capita), mobility, information acquisition status (COVID-19-related search frequency), and the parameters in the SIR-X model.

In this work, we focus on the correlation between local economy and the COVID-19-related representative parameters, e.g., search frequency, number of confirmed cases, inflows and outflows, which is different from other existing works [3], [6], [7], [11]. In addition, we avoid the influence of city scale by removing the city population size. Furthermore, we analyze the correlation by combining the SIR-X model and MCMC methods using the real data from Baidu Maps and Baidu Search Engine.

II. Methods: Combination of SIR-X and MCMC

In this section, we explain the SIR-X model to fit the number of confirmed cases of COVID-19. Then, we propose combining SIR-X and MCMC to estimate the parameters and to build the model. Afterwards, we compare the fitting number of accumulated confirmed cases and the official number.

Both Susceptible Infectious Recovered (SIR) model [21], [22] and Susceptible Exposed Infectious Recovered (SEIR) model [23]–[25] are widely used to estimate the outbreak of epidemics, while these methods do not consider the containment measures carried during the outbreak. SEIR model can be modified to consider the mobility of population [26] while the implementation of containment measures is still not considered. A Long-Short-Term-Memory (LSTM) model can be utilized to estimate the number of confirmed cases, which is still not able to take the containment measures into consideration [26] .

In this paper, we utilize the SIR-X model [11] to estimate the accumulated confirmed cases during the outbreak of the COVID-19 epidemic with the consideration of containment measures. The SIR-X model is a modified SIR model, which takes the containment measures into consideration. We utilize the same assumptions and representative parameters as those in [11]. We assume public containment measures, e.g., stay-at-home, reduced interaction among citizens, exist. These measures are denoted by 'containment' and represented by a variable κ_0. We assume that infected individuals are quarantined, which is denoted 'quarantine' and represented by a variable κ. We take α as the infection speed of an infected individual and β^{-1} as the average time that an infected individual remains infectious before recovery or removal. The SIR-X model is defined by the following differential equations:

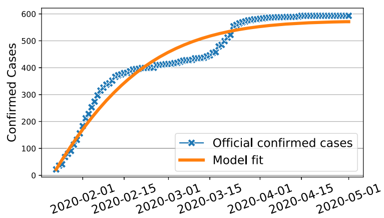
$$\begin{aligned}\partial_t S &= -\alpha SI - \kappa_0 S \\ \partial_t I &= \alpha SI - \beta I - \kappa_0 I - \kappa I \\ \partial_t R &= \beta I + \kappa_0 S \\ \partial_t X &= (\kappa + \kappa_0) I\end{aligned} \tag{1}$$

In the SIR-X model, I_0 refers to the number of initial infected individuals. R_0 is the basic reproduction number, which represents the average number of infections caused by an infected citizen before recovering or being removed [11]. We can calculate the reproduction number using $R_0 = \dfrac{\alpha}{\beta + \kappa + \kappa_0}$. We take $R_{0,free}$ to represent the reproduction number without containment or quarantine measures. S_0, R'_0, X_0 and I_0 represent the initial values of S, I, R and X at the beginning (January 23, 2020). S_0 is the population in the city, R'_0 is fixed as 0 and X_0 is the number of initial confirmed cases

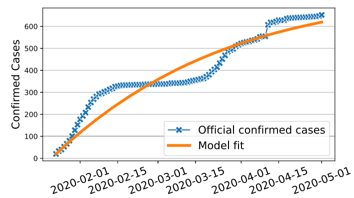
While high humidity and high temperature correspond to low transmission of COVID-19 [27], $R_{0,free}$ and β may differ according to different cities because of the divers local environments [28]. Thus, we exploit the MCMC [20] method to estimate the distribution of the parameters, i.e., α, β, κ, κ_0, I_0. However, we take the other parameters as fixed, i.e., S_0, R'_0, X_0 and I_0.

We utilize the uniform distribution as the prior distribution of parameters. We adopt the Sequential Monte Carlo sampler to infer the posterior distribution of the parameters with the consideration of the nonlinearity of SIR-X model. Finally, we take the expected value based on the posterior distribution of each parameter to construct the SIR-X model.

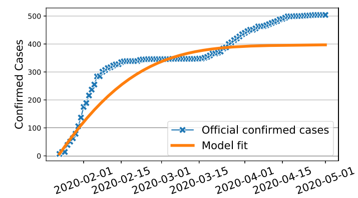
We exploit a priori conditions, i.e., $R_0 < R'_{0,free}$ and $\kappa_0 < \kappa$, to generate stable results while using the MCMC method. Without a priori conditions, the results of MCMC methods may be unstable, i.e., the results may be different for each execution. This may be caused by multiple possible solutions in the search space exist. Thus, we take advantage of a priori conditions to remove the unrealistic solutions. For instance, we assume $\kappa_0 < \kappa$, $R_0 < R'_{0,free}$, and the model fit number



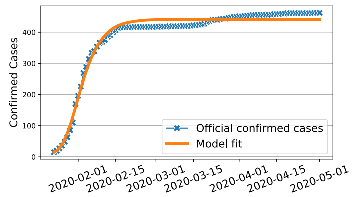
(a) Number of confirmed cases in Beijing.



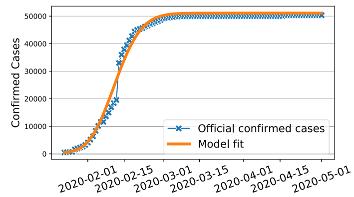
(b) Number of confirmed cases in Shanghai.



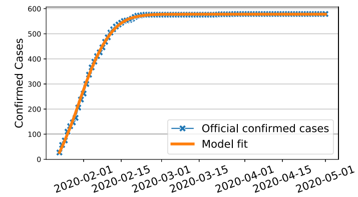
(c) Number of confirmed cases in Guangzhou.



(d) Number of confirmed cases in Shenzhen.



(e) Number of confirmed cases in Wuhan.



(f) Number of confirmed cases in Chongqing.

Fig. 3: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in Beijing, Shanghai, Guangzhou, Shenzhen, Wuhan and Chongqing.

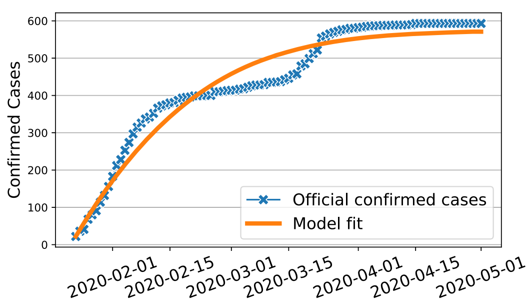


Fig. 4: Comparison of Number of confirmed cases between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in Beijing

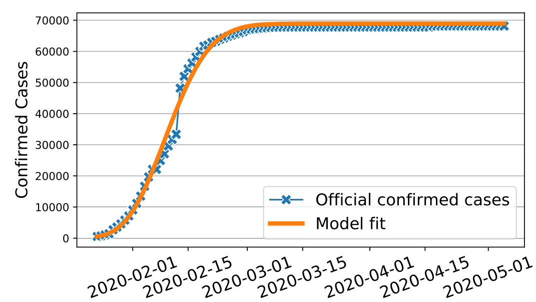


Fig. 5: Aggregated number of confirmed cases in Hubei. Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020

of accumulated confirmed cases should be equal or bigger than the official number of confirmed cases. $R'_{0,free}$ refers to the maximum value of R_0. We set $R'_{0,free}$ as 6.2, which is in accordance with the condition reported in [11], i.e. $1.4 < R_0 < 3.3$. If the a priori conditions are not satisfied, the fitting process will be repeated until reaching a limit, e.g., 20 times, in order to avoid infinite execution. Furthermore, we assume that the quarantine measure is applied more strictly on the infected individuals than other public citizens, i.e., $\kappa_0 < \kappa$.

We assume that few travelers or symptomatic infected individuals travel into or from a city while using the SIR-X model. As shown in Figure 1, we assume that few infected citizens travelled into other major cities than Wuhan after January 23, 2020 as almost all the transports were forbidden since January 23, 2020 and few people could go to other cities from Wuhan. In addition, we have another assumption: the number of infected cases in the inflows from other cities can be ignored compared to the number of infected individuals among the local citizens in a city. With these two assumptions, we can use the SIR-X and MCMC to estimate parameters for major cities in Mainland China based on the number of confirmed cases [1] from January 23, 2020 to May 1, 2020.

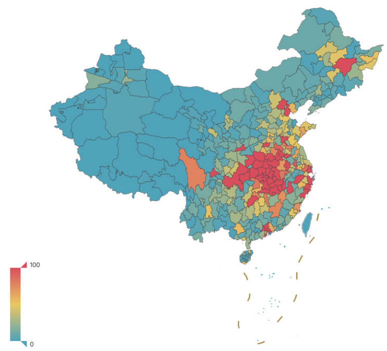
III. Results

In this section, we present the results of the combination of MCMC and SIR-X and the correlation results.

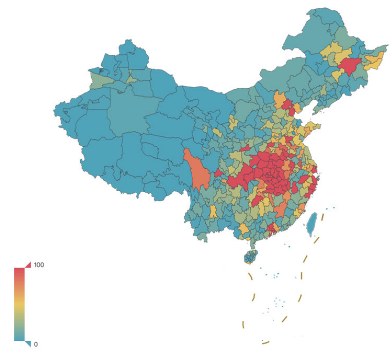
A. Results for Combination of MCMC and SIR-X

Figures 6a and 6b show that the combination of SIR-X model and MCMC captures well the number of confirmed cases at city scale. Figures 4 illustrates the confirmed cases in Beijing. The figures demonstrate that the model based on the combination of SIR-X and MCMC well fits to the number of confirmed cases in Beijing. However, there are some differences

[1]COVID-19 statistics - https://github.com/canghailan/Wuhan-2019-nCoV

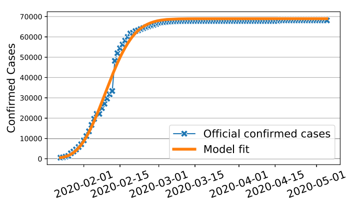


(a) The official number (May 1, 2020) of confirmed cases.

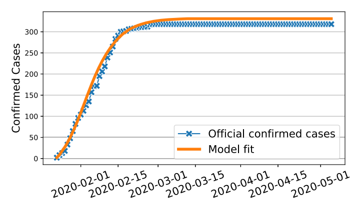


(b) The model fitting number (May 1, 2020) of confirmed cases.

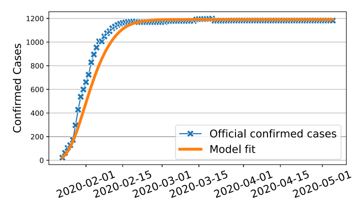
Fig. 6: The comparison between official number and model fitting number of confirmed cases in major cities of Mainland China.



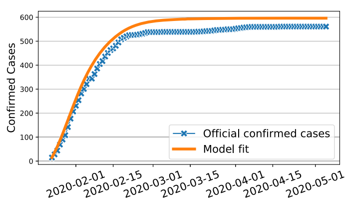
(a) Aggregated number of confirmed cases in Hubei.



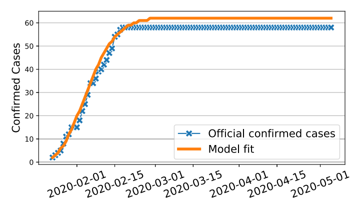
(b) Aggregated number of confirmed cases in Hebei.



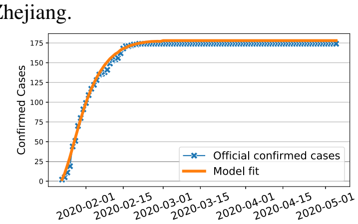
(c) Aggregated number of confirmed cases in Zhejiang.



(d) Aggregated number of confirmed cases in Sichuan.



(e) Aggregated number of confirmed cases in Xingjiang.



(f) Aggregated number of confirmed cases in Yunnan.

Fig. 7: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in Hubei, Hebei, Zhejiang, Sichuan, Xingjiang and Yunnan.

between the fitting number and the official number as there are many 174 confirmed cases from other countries. We find the same situation occurs in Shanghai (326 confirmed cases from other countries) and Guangzhou (127 [1] infected individuals from other countries) (please see details in Figures 3a - 3f for Shanghai, Shenzhen, Wuhan and Chongqing). We also calculate the confirmed cases at province scale by summing the number of confirmed cases of each affiliated city. Figures 5 shows the number of confirmed cases in Hubei. The combination of SIR-X and MCMC well captures the confirmed cases of provinces, as shown in Figures 7a - 7f for Hebei, Zhejiang, Sichuan, Xingjiang, Yunnan. In addition, we sum the number of confirmed cases of each province to calculate the number of confirmed cases in Mainland China as shown in Figure 9.

[1]Confirmed cases in Guangzhou from National Health Commission - http://wjw.gz.gov.cn/ztzl/xxfyyqfk/yqtb/content/post_5815637.html

B. Results for Correlation

We collected four datasets, i.e., the number of confirmed cases (May 1, 2020), mobility (Baidu Maps), GDP [29], and COVID-19-related search frequency (Baidu Search) for 238 cities in Mainland China (excluding Wuhan). The GDP dataset represents local economy in 2019. The COVID-19-related search frequency refers to the ratio between COVID-19-related search volume from January to March 2020 and population in each city. We calculate the correlation among multiple representative parameters, i.e., local economy, mobility, search behaviors and the parameters estimated based on the method presented in Section II. Figure 8 summarized the results of

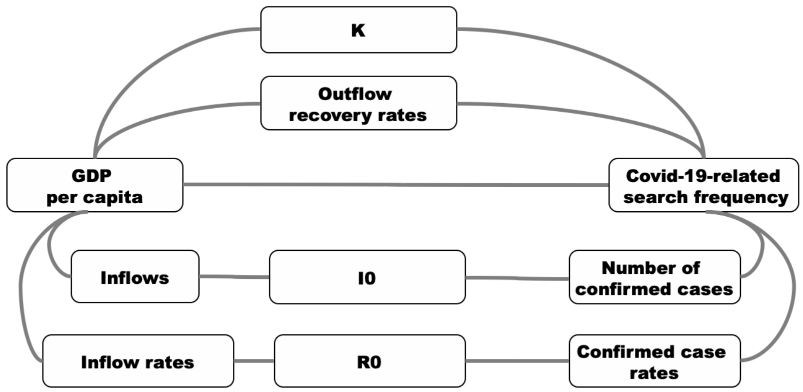


Fig. 8: Significant correlations among different factors.

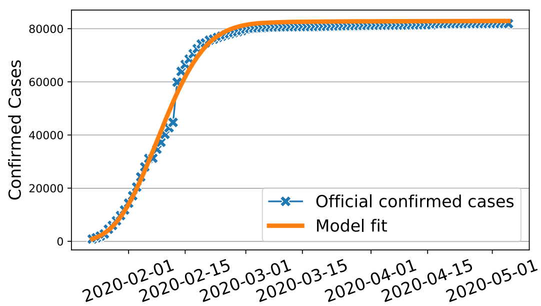


Fig. 9: Comparison between official number of confirmed cases and fitting number from the SIR-X model on May 1, 2020 in China.

our analysis. In this section, we show the correlation results obtained from the analysis.

1) Significant positive correlations have been evidenced between local economy and COVID-19-related search frequency: We have confirmed the significant positive correlations between COVID-19-related search frequency and local economy (Result I in Table I). We calculated the Pearson correlation coefficients [30] and conducted the Student's T-test (two tails) to verify the significance test, in order to analyze the correlation between two random variables (the same for the following analysis in the paper). The Pearson correlation between the local GDP per capita and the total COVID-19-related search volume (between January and March 2020) is $R^{***} = 90.4\%$ ($N = 238$ and p-value$= 5.92 \times 10^{-89} < 0.0001$) for each city.

We analyzed the significance of the correlations between COVID-19-related search frequency and GDP per capita, where we obtained the significant correlations as p-value$=2.33 \times 10^{-28} < 0.0001$. In addition, we carried out partial correlation analysis [31], [32] between the COVID-19-related search frequency and GDP per capita with the effects of the city population size (a controlling variable) removed. We obtained a strong correlation with significance as well, such that p-value$= 7.15 \times 10^{-22} < 0.0001$. Figure 10a visualize the correlations.

2) Correlation analysis for the spread of the COVID-19 pandemic: We have confirmed the significant positive correlation between inflows from Wuhan and GDP per capita (Result II in Table I). We conducted correlation between the inflows from Wuhan and the GDP per capita for every city in the study, where we got p-value$= 9.88 \times 10^{-12} < 0.0001$. The correlation reveals that cities with higher GDP per capita would attract larger inflows from Wuhan. In addition, in order to avoid the impact of the city scale, we analyze the correlation between the inflows rate from Wuhan, i.e., the ratio between inflows from Wuhan and the population, and GDP per capita. We obtained a strong positive correlation with significance as well, such that p-value$= 6.17 \times 10^{-7} < 0.0001$. Figures 10b and 10c visualize the correlations.

We have obtained the significant positive correlation between the inflows and I_0 (Result II in Table I). We have also obtained the significant positive correlation between the inflow rate and R_0 (Result III in Table I). We conducted correlation between the inflows from Wuhan and I_0, where we obtained a strong positive correlation with significance, such that p-value$= 8.11 \times 10^{-4} < 0.001$. This correlation indicates that cities with larger inflows from Wuhan have more initial infected cases, i.e., I_0 in the SIR-X model. Furthermore, we conducted the correlation between R_0 and the inflow rate, where we obtained a strong positive correlation with significance, such that p-value$= 3.27 \times 10^{-3} < 0.005$. Figures 10d and 11a show the visualization of the correlations.

We have obtained the significance of the positive correlation between the number of confirmed cases and the number of initial infected individuals (Result II in Table I). We also found the positive correlation between R_0 and the number of confirmed case rate (Result III in Table I). We analyzed correlation using the number of confirmed cases and I_0. We obtained a significant positive correlation, such that p-value$= 4.67 \times 10^{-4} < 0.001$, which confirms that the bigger number of initial infected individuals, the bigger number of confirmed cases. We investigate correlation between the confirmed case

TABLE I: Overall Results of Correlation Analysis (corresponding figures).

Correlations	Coeff. (R)	p-value
Result I		
GDP per capita vs. COVID-19-related search volume	90.4%	< 0.0001
GDP per capita vs. COVID-19-related search frequency 10a	63.5%	< 0.0001
Result II		
GDP per capita vs. Inflows from Wuhan 10c	42.3%	< 0.0001
Inflows from Wuhan vs. I_0 10d	21.6%	< 0.001
I_0 vs. Number of confirmed cases 11b	22.5%	< 0.001
Number of confirmed cases vs. COVID-19-related search frequency 12b	41.5%	< 0.0001
Result III		
GDP per capita vs. Inflows from Wuhan / population 10b	31.6%	< 0.0001
Inflows from Wuhan / population vs. R_0 11a	19.0%	< 0.005
R_0 vs. Number of confirmed cases / population 11c	29.8%	< 0.0001
Number of confirmed cases / population vs. COVID-19-related search frequency 12c	21.4%	< 0.001
Result IV		
GDP per capita vs. κ 11d	17.3%	< 0.01
COVID-19-related search frequency vs. κ 12d	17.6%	< 0.01
Result V		
GDP per capita vs. Outflow recovery rates 12a	-46.5%	< 0.0001
COVID-19-related search frequency vs. Outflow recovery rates 13a	-51.5%	< 0.0001
Result VI		
Number of confirmed cases vs. Inflows from Wuhan 13b	82.7%	< 0.0001
Number of confirmed cases / population vs. Inflows from Wuhan 13c	36.1%	< 0.0001
Outflow recovery rates vs. Number of confirmed cases 13d	-31.6%	< 0.0001
Outflow recovery rates vs. Number of confirmed cases / population 14	-22.5%	< 0.001

rate and R_0, with p-value$= 2.82 \times 10^{-6} < 0.0001$. This correlation confirmed that cities with bigger R_0 have more confirmed case rate, i.e., the ratio between the confirmed cases and the population. Figures 11b and 11c show visualization of the correlations. In Figure 11b, we can categorize the cities into two groups, i.e., one with high I_0 and low number of cases (G1) and another with low I_0 and high number of cases (G2). G1 consists of 6 cities, i.e., Beijing, Shanghai, Guangzhou, Shenzheng, Chongqing, and Wenzhou, among which 5 cities are the most active cities in terms of economic development. The spreading speed of the COVID-19 pandemic in the cities of G1 is much faster than that in the cities of G2.

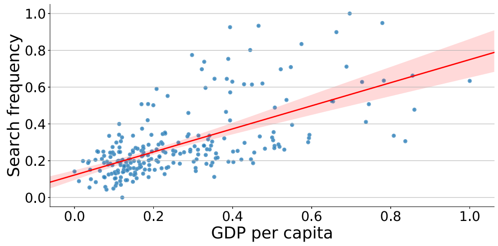
We evidenced a strong positive correlation with significance between the COVID-19-related search frequency and the number of confirmed cases, such that p-value$= 2.45 \times 10^{-11} < 0.0001$, as shown in Figures 12b . In addition, we evidenced a strong positive correlation with significance between the COVID-19-related search frequency and confirmed case rate, such that p-value$= 9.01 \times 10^{-4} < 0.001$, as shown in Figures 12c. Thus, we can confirm that when there are more confirmed cases, the citizens are more likely to perform COVID-19 related search.

3) Correlation analysis for the interaction between local economy and the implementation of containment measures: In order to analyze the interaction between local economy and the implementation of containment measures, we analyze the correlation among three factors, i.e., information acquisition, local economy, and containment measures. As the quarantine measure is critical to the number of confirmed cases, we study the correlation between κ and other factors, i.e., information acquisition and local economy. We also analyze the realization of containment measures, i.e., the outflow recovery rate.

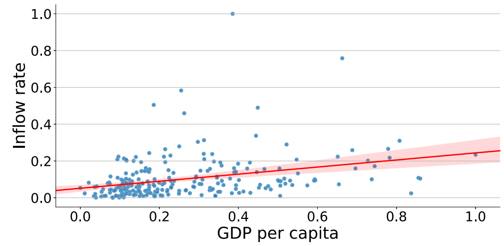
We have confirmed the significance of the positive correlation between GDP per capita and the implementation of the quarantine measure for major Chinese cities in the study (Result IV in Table I). We correlated κ and the GDP per capita, while p-value$= 7.46 \times 10^{-3} < 0.01$). In addition, we correlated the outflow recovery rate and the GDP per capita, where p-value$= 3.82 \times 10^{-14} < 0.0001$). Figures 11d and 12a show the correlations. Note that small outflow recovery represents that the quarantine measures are well performed. Thus, the significant negative correlation between GDP per capita and the outflow recovery rate shown in Figure 12a confirms that people with high GDP per capita would try harder to realize the quarantine measures and to reduce travelling.

We have found the significant positive correlations between the COVID-19-related search frequency and the realization of quarantine measure. We analyzed the correlation using κ and the COVID-19-related search frequency. We found a significant positive correlation, such that p-value$= 6.35 \times 10^{-3} < 0.01$. Figures 12d shows the correlations. We evidenced that the negative correlations between the the outflow recovery rate and COVID-19-related search frequency with p-value$= 1.51 \times 10^{-17} < 0.0001$). The correlation analysis result indicates that citizens with higher per capita COVID-19-related search frequency apply the containment measures more strictly. Figure 13a shows the correlations.

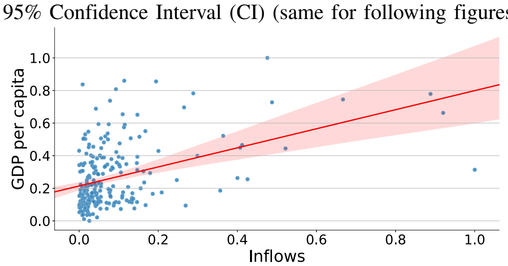
4) Correlation analysis for the interaction between the population mobility and the COVID-19 pandemic: In this section, we analyze the correlation among inflows to Wuhan, outflows from Wuhan and the number of confirmed cases. We first focus on the correlation between the number of confirmed



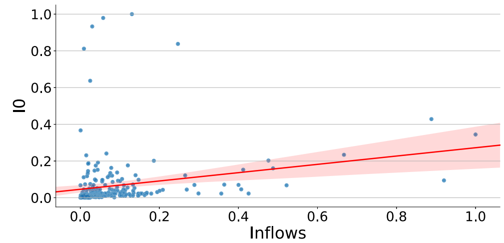
(a) Significant positive correlation between COVID-19-related search frequency and GDP per capita. Shaded area represents the 95% Confidence Interval (CI) (same for following figures).



(b) Significant positive correlations between inflow rate from Wuhan and GDP per capita.



(c) Significant positive correlations between inflows from Wuhan and GDP per capita.



(d) Significant positive correlations between I_0 and the inflows from Wuhan.

Fig. 10: Correlation analysis.

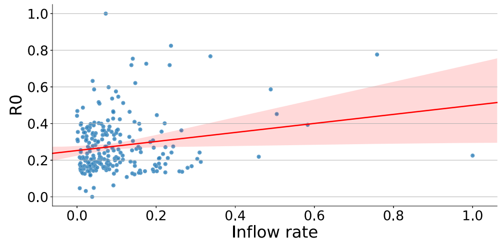
cases (or confirmed case rates) and inflows from Wuhan. Then, we analyze the correlation between the outflow recovery rates and the number of confirmed cases (or confirmed case rates).

We have evidenced the significance of the correlations between the population inflows from Wuhan and local infections (both the number of confirmed cases and the infection rate) for major Chinese cities in the study. The Pearson correlation between the number of confirmed cases of every city and the number of Baidu Maps users migrated to the city from Wuhan (during 1 to 23 Jan 2020) is $R^{***} = 82.7\%$ ($N = 238$ and p-value$= 6.31 \times 10^{-61} < 0.0001$). Similar correlation results have been found in [18], [19]. However, we considered that the scale of the city would be a threat to the validity of this observation, as a larger city would attract more population inflows, while a larger city with greater population would experience more infections. We therefore tested the significance of correlations between the local infection rate (number of confirmed cases/population of the city) and the population inflows from Wuhan, where we evidenced the significance in the correlations as $R^{***} = 36.1\%$ ($N = 238$ and p-value$=9.76 \times 10^{-9} < 0.0001$). Furthermore, we carried out partial correlation analysis between the number of confirmed cases and the population inflows from Wuhan with the effects of the city population size (a controlling variable) removed. The partial correlation analysis gave a strong correlation with significance, such that $R^{***} = 56.6\%$ ($N = 238$ and p-value$= 1.41 \times 10^{-21} < 0.0001$). Please see Figures 13b and 13c for the correlations in details.

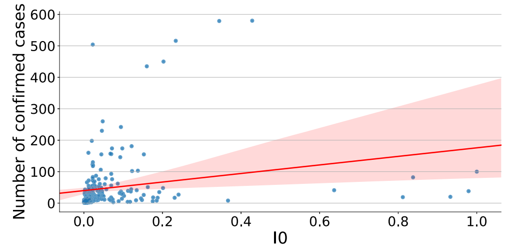
Negative correlations have been evidenced between the population outflows and local infections for major Chinese cities We have evidenced the significance of the negative correlation between the outflow populations and local infections (both the number of confirmed cases and the infection rate) for major Chinese cities in the study. Among all 238 cities in the correlation study, we found the outflow recovery rates range between 17.9% to 66.8% while more than 90% of cities had outflow recovery rates lower than 50%. We hypothesized that people would try harder to escape the cities with more infections. Therefore, for every city in the study, we correlated the outflow recovery rate with the number of confirmed cases, as well as the outflow recovery rate with the local infection rate, where we obtained Pearson correlation coefficients of $R^{***} = -31.6\%$ ($N = 238$ and p-value$= 6.41 \times 10^{-7} < 0.0001$) and $R^{**} = -22.5\%$ ($N = 238$ and p-value$= 0.000459 < 0.0005$). The correlation analysis result suggests that people under more critical situations are more likely to refrain from travelling out (fewer travels under higher infections), as the correlation is significantly negative between the outflow recovery rates and infections. Please see also Figures 13d and 14 for the visualization of the correlations.

IV. Discussion

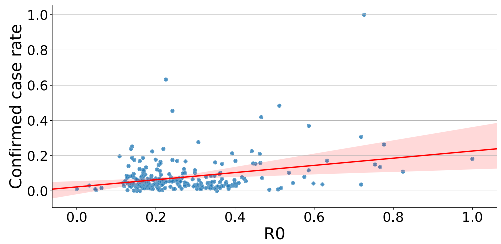
From the analysis, we find that the cities of strong local economy may receive many infected migrations and the speeds of spreading epidemics are also much faster than that in other cities. Thus, when there is COVID-19 or similar pandemic, it is of much importance to take strict measure in the cities of strong local economy, e.g., wearing masks, reducing mass gatherings etc. In addition, the information acquisition is also critical to the execution of measures as well-informed



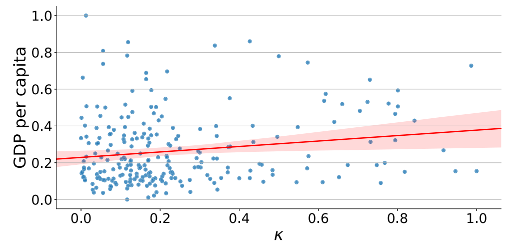
(a) Significant positive correlations between R_0 and the inflow rate from Wuhan.



(b) Significant positive correlations between the number of confirmed cases and I_0



(c) Significant positive correlations between confirmed case rate and R_0.



(d) Significant positive correlations between κ and GDP per capita.

Fig. 11: Correlation analysis.

citizens prefer strictly performing the quarantine measurements. Furthermore, the results of this work may potentially be used in two major applications. First, the model can be used to predict the confirmed cases. Second, the correlation study can be used to take measures in high-risk cities to prevent the outbreak of pandemic.

V. Conclusion

In this paper, we combine the SIR-X model and MCMC methods to infer the representative parameters related to the COVID-19 pandemic in major cities of Mainland China. Afterward, we analyzed the correlation among the spread of COVID-19 pandemic, the execution of containment measures, and the local economy. We investigated the correlation using the mobility data and search data from Baidu Maps and Baidu Search Engine. We confirmed strong correlations among divers factors. The inflows from Wuhan are attracted by the cities with higher GDP per capita, where the number confirmed cases is bigger. In these cities, the citizens perform COVID-19 related search more frequently while applying the containment measures more strictly. These correlations reveal that the better the local economy is and the timelier information is acquired by citizens, the better the containment measures are carried out, which help fight against the COVID-19 pandemic.

References

[1] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China," *China CDC Weekly*, vol. 41, no. 2, p. 145, 2020.

[2] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76, pp. 71–76, 2020.

[3] M. Ahinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, 2020.

[4] A. Deslandes, V. Berti, Y. Tandjaoui-Lambotte, C. Alloui, E. Carbonnelle, J. Zahar, S. Brichler, and Y. Cohen, "SARS-CoV-2 was already spreading in france in late december 2019," 2020.

[5] J. H. University, "Johns Hopkins University Coronavirus Resource Center," http://https://coronavirus.jhu.edu/, 2020, [Online; accessed 18-May-2020].

[6] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage *et al.*, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, pp. 493–497, 2020.

[7] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang *et al.*, "An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China," 2020.

[8] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying SARS-COV-2 transmission suggests epidemic control with digital contact tracing," 2020.

[9] S. Engle, J. Stromme, and A. Zhou, "Staying at Home: Mobility Effects of COVID-19," 2020.

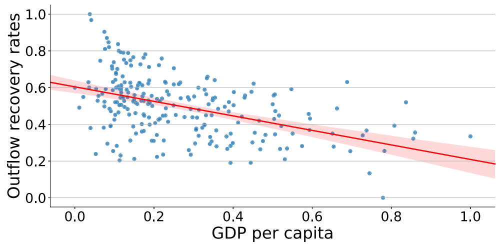
[10] A. Remuzzi and G. Remuzzi, "Covid-19 and italy: what next?" *The Lancet*, vol. 395, no. 10231, pp. 1225 – 1228, 2020.

[11] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in recent confirmed covid-19 cases in china," *Science*, vol. 368, no. 6492, pp. 742–746, 2020.

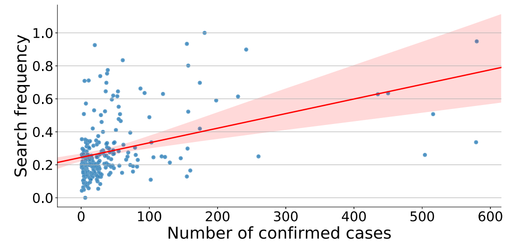
[12] S. Chen, J. Yang, W. Yang, C. Wang, and T. Bärnighausen, "COVID-19 control in china during mass population movements at new year," *Lancet*, vol. 395, no. 10226, pp. 764–766, 2020.

[13] S.-Y. Ren, R.-D. Gao, and Y.-L. Chen, "Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic," *World Journal of Clinical Cases*, vol. 8, no. 4, p. 652, 2020.

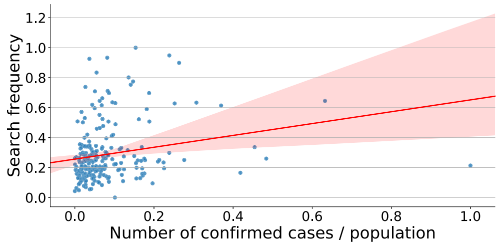
[14] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai *et al.*, "Modified SEIR and AI prediction of the



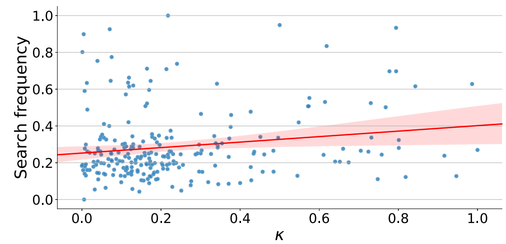
(a) Significant negative correlations between the outflow recovery rate and GDP per capita.



(b) Significant positive correlations between COVID-19-related search frequency and the number of confirmed cases.



(c) Significant positive correlations between per capita COVID-19-related search frequency and confirmed case rate.



(d) Significant positive correlations between κ and the COVID-19-related search frequency.

Fig. 12: Correlation analysis.

epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, p. 165, 2020.

[15] H. Li, X. Chen, and H. Huang, "The novel coronavirus outbreak: what can be learned from China in public reporting?" *Global Health Research and Policy*, vol. 5, no. 1, pp. 1–3, 2020.

[16] S. Chen, J. Yang, W. Yang, C. Wang, and T. Bärnighausen, "COVID-19 control in China during mass population movements at new year," *The Lancet*, vol. 395, no. 10226, pp. 764–766, 2020.

[17] J. Zhang, G. Lin, J. Zeng, J. Lin, J. Tian, and G. Li, "Challenges of SARS-COV-2 and lessons learnt from SARS in guangdong province, China," *Journal of Clinical Virology*, vol. 126, p. 104341, 2020.

[18] H. Chen, Y. Chen, B. Sun, P. Wang, L. Wen, Z. Lian, Y. Lu, Y. Qi, S. Zhao, L. Zhang *et al.*, "Correlation between the migration scale index and the number of new confirmed Novel Coronavirus Pneumonia cases in China," *Virus*, vol. 11, p. 3, 2020.

[19] P. Zhong, S. Guo, and T. Chen, "Correlation between travellers departing from Wuhan before the Spring Festival and subsequent spread of COVID-19 to all provinces in China," *Journal of Travel Medicine*, vol. 27, no. 3, pp. 1–4, 2020.

[20] W. R. Gilks, "Markov chain monte carlo," 2005.

[21] A. A. Toda, "Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact," 2020.

[22] A. Simha, R. V. Prasad, and S. Narayana, "A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe," 2020.

[23] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," 2020.

[24] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, J. Zhang, G. Xu, and E. Jia, "The effectiveness of quarantine of wuhan city against the corona virus disease 2019 (COVID-19): A well-mixed seir model analysis," 2020.

[25] S. Tang, B. Tang, N. L. Bragazzi, F. Xia, T. Li, S. He, P. Ren, X. Wang, C. Xiang, Z. Peng, J. Wu, and Y. Xiao, "Analysis of covid-19 epidemic traced data and stochastic discrete transmission dynamic model," 2020.

[26] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He, "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, pp. 165–174, 2020.

[27] J. Wang, K. Tang, K. Feng, and W. Lv, "High Temperature and High Humidity Reduce the Transmission of COVID-19," *Social Science Research Network (SSRN)*, 2020.

[28] S. Huang, J. Liu, H. Xiong, J. Huang, H. An, and D. Dou, "The weather impacts the outbreak of covid-19 in mainland china," *arXiv preprint arXiv:2006.10376*, 2020.

[29] N. B. of Statistics of China, "National Bureau of Statistics of China," http://www.stats.gov.cn/, 2019, [Online; accessed 21-May-2020].

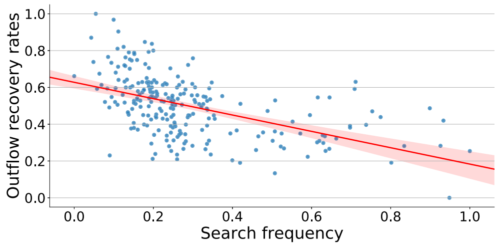
[30] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise Reduction in Speech Processing*, vol. 2, pp. 1–4, 2009.

[31] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian & New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.

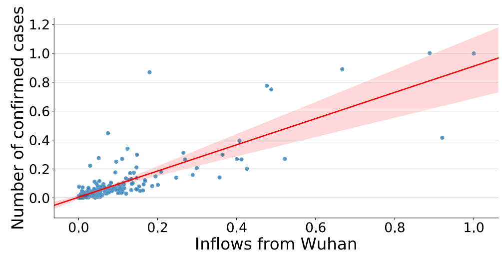
[32] D. Y. Kenett, X. Huang, I. Vodenska, S. Havlin, and H. E. Stanley, "Partial correlation analysis: applications for financial markets," *Quantitative Finance*, vol. 15, no. 4, pp. 569–578, 2015.

APPENDIX

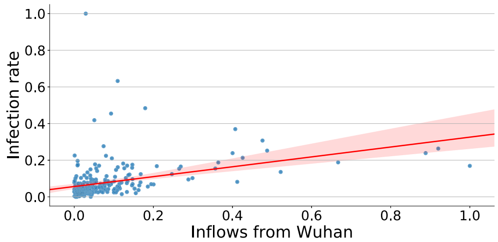
Figure 1 shows the volumes of population inflows from Wuhan evolving over time and Figure 2 presents the volumes of COVID-19-related searches evolving over time.



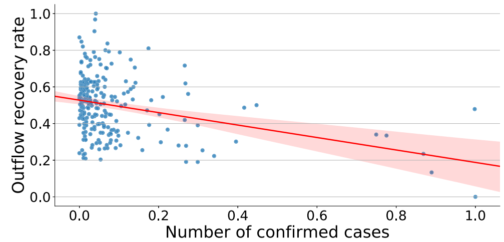
(a) Significant negative correlations between the outflow recovery rate and the per capita COVID-19-related search activities.



(b) Significant positive correlations between the number of confirmed cases and the inflows from Wuhan.



(c) Significant positive correlations between the infection rate and the inflows from Wuhan.



(d) Significant negative correlations between the outflow recovery rate and the number of confirmed cases.

Fig. 13: Correlation analysis.

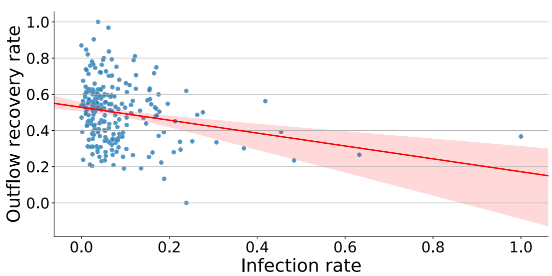


Fig. 14: Significant negative correlations between the outflow recovery rate and the infection rate.

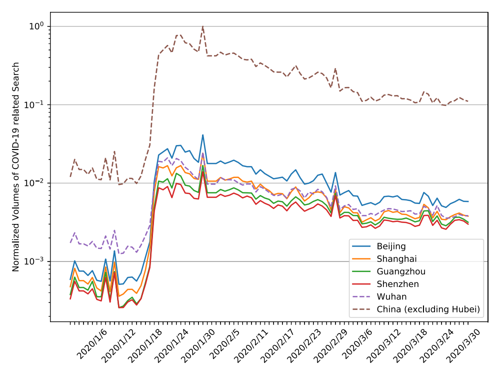


Fig. 2: The normalized volumes of COVID-19-related searches over time (in logarithmic scale). We normalized the plots proportionally using the maximum volume of the search (the national peak arrived on 31 Jan 2020 — the date that the first patient cured of novel coronavirus in Wuhan was discharged from the hospital). Wuhan, which has a smaller population than the other four cities, contributed even higher volumes of related searches in the early of January 2020. It might reflect the local infections in the early stage, as well as the collective responses of populations. Nationwide, compared to the peak time, the overall volumes of COVID-19-related searches have dropped to 10% by the end of March 2020. It is possibly due to the collective responses of populations to the relative local containment of the pandemic.

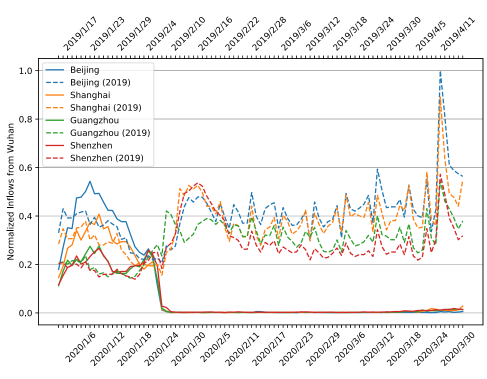


Fig. 1: The comparisons of normalized volumes of population inflows from Wuhan between 2020 and 2019.