Urban Epidemic Hazard Index for Chinese Cities: Why Did Small Cities Become Epidemic Hotspots?

Tianyi Li[*]

System Dynamics Group, Sloan School of Management, MIT

Jiawen Luo[†‡]

Institute of Geophysics, ETH Zurich

Cunrui Huang[§]

School of Public Health, Sun Yat-sen University

March 10, 2021

Abstract

Multiple small- to middle-scale cities, mostly located in northern China, became epidemic hotspots during the second wave of the spread of COVID-19 in early 2021. Despite qualitative discussions of potential social-economic causes, it remains unclear how this pattern could be accounted for from a quantitative approach. Through the development of an urban epidemic hazard index (EpiRank), we came up with a mathematical explanation for this phenomenon. The index is constructed from epidemic simulations on a multi-layer transportation network model on top of local SEIR transmission dynamics, which characterizes intra- and inter-city compartment population flow with a detailed mathematical description. Essentially, we argue that these highlighted cities possess greater epidemic hazards due to the combined effect of large regional population and small inter-city transportation. The proposed index, dynamic and applicable to different epidemic settings, could be a useful indicator for the risk assessment and response planning of urban epidemic hazards in China; the model framework is modularized and can be adapted for other nations without much difficulty.

Despite the nation-wide successful implementation of control measures against COVID-19 (*Kraemer et al.*, 2020; *Lai et al.*, 2020; *Tian et al.*, 2020), multiple small- to middle-scale cities in China became local epidemic hotspots during the new wave of the pandemic in early 2021; the list includes Tonghua, Songyuan, Suihua, Qiqihar, Heihe, and Xingtai etc. (*Yew*, 2021; *Steven*, 2021). Unlike their nearby capital cities (e.g., Shijiazhuang, Changchun, Harbin), these small cities are probably not known to many Chinese before they are enlisted as "high-risk regions" after the local bursts, and it is indeed an unexpected phenomenon that they are highlighted among the over 340 prefectural administrations in China. It is implied that the likelihood of epidemic hazards is high in these regions. Many social-economic factors may account for this fact (*Niu et al.*, 2020; *Qiu et al.*, 2020): for example, social scientists may observe that these cities are all located in the northeast part of China, where local economies are often underdeveloped, and local residents are often more behavioral active than they are supposed to be in face of the epidemic (e.g., *Shelach*, 2006); other conjectures may attend to the fact that since these are neither coastal cities nor metropolitans where imported cases are more common, local control measures and regulations are different and somewhat relaxed, which led

---

[*]tianyil@mit.edu

[†]jiawen.luo@erdw.ethz.ch

[‡]T. L and J. L contribute equally to this work.

[§]huangcr@mail.sysu.edu.cn

to heedlessness of early signals (e.g., *Ye et al.*, 2020; *Li et al.*, 2021), or that these northern regions have cold winters and also less residential housing space than the south, hence the hazard of severe infections was harbored (e.g., *Cao et al.*, 2020). Although these arguments are sound, it is desired that quantitative explanations could be addressed to decipher why such small cities in China stand out as epidemic hotspots.

This points to the significant necessity of the quantification of urban regions' epidemic hazards, desirably via a constructed index that assesses the extent of potential risk. Such indices of potential epidemic risk can be useful for effective decision analysis during epidemic response and planning (*Morse et al.*, 2012; *Hamele et al.*, 2018), critical to the mitigation of sudden and potentially catastrophic impacts of infectious diseases on society (*Shearer et al.*, 2020). Indeed, although the exit of COVID-19 is still on the fly, both methodologically and practically (*Thompson et al.*, 2020), it is nevertheless prudent to start getting prepared for the next pandemic (*Liu et al.*, 2018; *Neumann and Kawaoka*, 2019; *Simpson et al.*, 2020), both economically and ecologically (*Di Marco et al.*, 2020; *Dobson et al.*, 2020), through comprehensive pandemic risk management synthesis (*Studzinski and Pasteur*, 2020) and the upgraded implementation of digital technologies (*Budd et al.*, 2020)

Many successful attempts have been made to develop such epidemic risk indices from various angles. According to US CDC, the preparedness for influenza pandemics can be assessed with the Influenza Risk Assessment Tool (*Trock et al.*, 2012; *Cox et al.*, 2014; *Burke and Trock*, 2018) through the Pandemic Severity Assessment Framework (*Reed et al.*, 2013; *Holloway et al.*, 2014); for the same purpose, the Tool for Influenza Pandemic Risk Assessment (*WHO-TIPRA*, 2016) is recommended by WHO; and there are miscellaneous other tools developed for national-level pandemic planning, through either mathematical simulations (e.g., *Eichner et al.*, 2007) or scoring systems, in which case various social-economic factors are being considered (e.g., *McKay et al.*, 2019; *Oppenheim et al.*, 2019; *Grima et al.*, 2020).

However, it is suggested that currently a suitable public health evaluation framework for the assessment of epidemic risk and response is still not full-fledge in scale (*Warsame et al.*, 2020). Most proposed tools and frameworks are subject to a few shortcomings: (1) assessments are in most cases from the supply side (i.e., the preparedness) instead of the demand side (i.e., the actual risk); (2) assessment of pandemic potential is often virus-specific (i.e., pathological), not sufficiently general-purpose, as the risk potential is also determined by various societal factors (transportation, population (*Copiello and Grillenzoni*, 2020) etc.); (3) many indices rely on expert scoring systems that often depend largely on subjectivity, and the calling for mathematical models and algorithms for risk assessment and pandemic planning is compelling (*Wu and Cowling*, 2011); (4) finally, many models focus on nation-wide evaluation, and there is relatively little concentration on sub-nation (e.g., city) level analysis, except for a few interesting attempts (e.g., *Prieto et al.*, 2020; *Boyce and Katz*, 2002; *Zhu et al.*, 2020).

To deal with this problem, in this study we develop a novel epidemic hazard index for Chinese cities, which quantifies the potential risk of epidemic spread at the over 340 Chinese prefectural administrations. The index relies on a simulation model which integrates intra-city compartment dynamics and a detailed mathematical description of inter-city multi-channel transportation. Calculation of the hazard index is based on this dynamic system that simulates the domestic epidemic spread for user-specified diseases. In the model, intra-city evolution is governed by the SEIR dynamics assuming no epidemic response taken place, such that the constructed index serves as an early-warning indicator at early periods of epidemics, before the incidence of any structural change in the population flow upon policy intervention (e.g., *Schlosser et al.*, 2020; *Li*, 2021); inter-city transportation is modeled with a multi-layer bipartite network (*Li*, 2020), which make explicit considerations of various events during inter-city population flow, including transit, cross-infection due to path overlap, as well as the different transmissivities on different transportation media.

Such a highlight on transportation (i.e., spacial patterns) is core to the city-specific risk assessment of epidemics (*Smith et al.*, 2002) and natural hazards in general (*Greiving et al.*, 2006). Indeed, over the course of the still on-going pandemic, it is acknowledged that transportation, at both the global and the domestic level, plays a critical role during the spread of viruses and to a large extent may determine the severity of the disease at different geological divisions (*Christidis and Christodoulou*, 2020; *Liu et al.*, 2021). Essentially, compared to regression models (*Bo et al.*, 2014) or the machine-learning approach (*Feng et al.*, 2021), the highlight of transportation asks for epidemic risk analysis from a network perspective (e.g., *Chu et al.*, 2020;

Ge et al., 2020; *So et al.*, 2020), upon which the risk scores could then be computed from quantitative approaches (*So et al.*, 2020). An important precedent is the Global Epidemic and Mobility (GLEaM) model, which integrates sociodemographic and population mobility data in a spatially structured stochastic disease approach to simulate the spread of epidemics at the worldwide scale (*Balcan et al.*, 2010). GLEaM considers the commuting on the airport network on top of local disease transmission, where transportation is modeled via an effective operator; our model adopts a similar methodology yet constructs a more realistic multi-layer mathematical description of inter-city transportation, comparing also to various recent studies in the same line of research (*Jia et al.*, 2020; *Wang et al.*, 2020; *Zhang et al.*, 2020; *Chang et al.*, 2021).

Model

The model is developed in *Li* (2020) and is briefed summarized here. Assume a bi-partite graph with cities (nodes) classified as either central cities or peripheral cities. The network is multi-layer $G = (V, E_{A/B/R/S})$, specifying four means of inter-city transportation (i.e., different layers have different connectivites between nodes): Air (A), Bus (B), Rail (R) and Sail (S). At each node, the local urban population is divided into four compartments: Susceptible (S), Exposed (E), Infected (I), Recovered (R), and the intra-city epidemic spread follows the standard SEIR dynamics (e.g., *Sterman*, 2000; *Newman*, 2010). We track the in- and out-flow of the exposed, susceptible and recovered population at each node on the transportation network, which determines the open-system SEIR dynamics (for a specific city i):

$$
\begin{cases}
\dot{S}_i = -\dfrac{S_i}{P_i}\left(\dfrac{R_0}{D_I}I_i + z_i\right) + \Delta S_i^{in} - \Delta S_i^{out} \\[2mm]
\dot{E}_i = \dfrac{S_i}{P_i}\left(\dfrac{R_0}{D_I}I_i + z_i\right) - \dfrac{E_i}{D_E} + \Delta E_i^{in} - \Delta E_i^{out} \\[2mm]
\dot{I}_i = \dfrac{E_i}{D_E} - \dfrac{I_i}{D_I} \\[2mm]
\dot{R}_i = \dfrac{I_i}{D_I} + \Delta R_i^{in} - \Delta R_i^{out}.
\end{cases}
\tag{1}
$$

Epidemiological parameters R_0, D_E, D_I are the basic reproduction number, the incubation period, the infection period; $z_i = z_i(t)$ is the zoonotic force, i.e., the seed of the disease, with $z_i \neq 0$ only at node(s) that is the disease epicenter(s). The in- and out-flows of each city are determined via the flowmaps $F^q = \{f_{i,j}^q\}$ between pairs of cities (i, j) specified for each means of transportation q. During the inter-city flow, transit events are considered, where only a certain proportion of flow ($TR_{c/p}$ for central/peripheral cities, the same for each layer) enter the local population, and the rest are directed to other destinations. Moreover, cross-infections during inter-city travels are modeled, which take place between a susceptible person and an exposed person who share an overlapped travel path with the same destination. The strength of the cross-infection spillover effect R_T^q various on different transportation media (e.g., *Hu et al.*, 2020).

At each city i, the exposed inflow ($\Delta E_i^{in}(t)$) is given by

$$
\Delta E_i^{in}(t) = \sum_{q \in Q} \sum_{j \in V} \overline{f_{j,i}^q(t)\mu_j(t)}(1 - TR_i),
\tag{2}
$$

where

$$
\overline{f_{j,i}^q \mu_j} = f_{j,i}^q \mu_j + \sum_k^{p^q(i,k) \cap p^q(i,j) \neq 0} f_{k,i}^q \mu_k (R_T^q - 1) \frac{f_{j,i}^q(1 - \mu_j - \eta_j)\min(d_{j,i}^q, d_{k,i}^q)}{\sum_l^{p^q(i,k) \cap p^q(i,l) \neq 0} f_{l,i}^q(1 - \mu_l - \eta_l)\min(d_{l,i}^q, d_{k,i}^q)}
\tag{3}
$$

is the adjusted exposed flow from city j to i by means q, taking care of cross-infections ($d_{i,j}^q$ represents the shortest path distance between i and j on layer q), and

$$
\mu_i(t) = \frac{\Delta E_i^{out}(t) + \sum_{q \in Q} \sum_{j \in V} \overline{f_{j,i}^q(t-1)\mu_j(t)} TR_i^q}{\sum_{q \in Q} \sum_{j \in V} f_{i,j}^q(t)},
\tag{4}
$$

$$\eta_i(t) = \frac{\Delta R_i^{out}(t) + \sum\limits_{q \in Q}\sum\limits_{j \in V} f_{j,i}^q(t)\eta_j(t-1)TR_i^q}{\sum\limits_{q \in Q}\sum\limits_{j \in V} f_{i,j}^q(t)}, \tag{5}$$

are the time-stamped proportion of the exposed and recovered population among the total outflow population from city i; thus $(1 - \mu_i - \eta_i)$ is the proportion of the susceptible population among the total outflow from city i. The recovered inflow is tracking all the recovered people upon arrival (via η_j): $\Delta R_i^{in}(t) = \sum\limits_{q \in Q}\sum\limits_{j \in V} f_{j,i}^q(t)(1 - TR_i^q)\eta_j(t-1)$, and according to flow balance, the susceptible inflow is $\Delta S_i^{in}(t) = \sum\limits_{q \in Q}\sum\limits_{j \in V} f_{j,i}^q(t)(1 - TR_i^q) - \Delta E_i^{in}(t) - \Delta R_i^{in}(t)$. The outflow population from city i's population P_i is the total outbound flow minus the transferred inbound flow, contributed by the S, E, R compartments (with the proper assumption that I stay local, i.e., infected people do not participate in inter-city travels). Proportionally, the outflows are:

$$\Delta X_i^{out}(t) = X_i(t)\frac{\sum\limits_{q \in Q}\sum\limits_{j \in V} f_{i,j}^q(t) - \sum\limits_{q \in Q}\sum\limits_{j \in V} f_{j,i}^q(t)TR_i^q}{S_i(t) + E_i(t) + R_i(t)}. \tag{6}$$

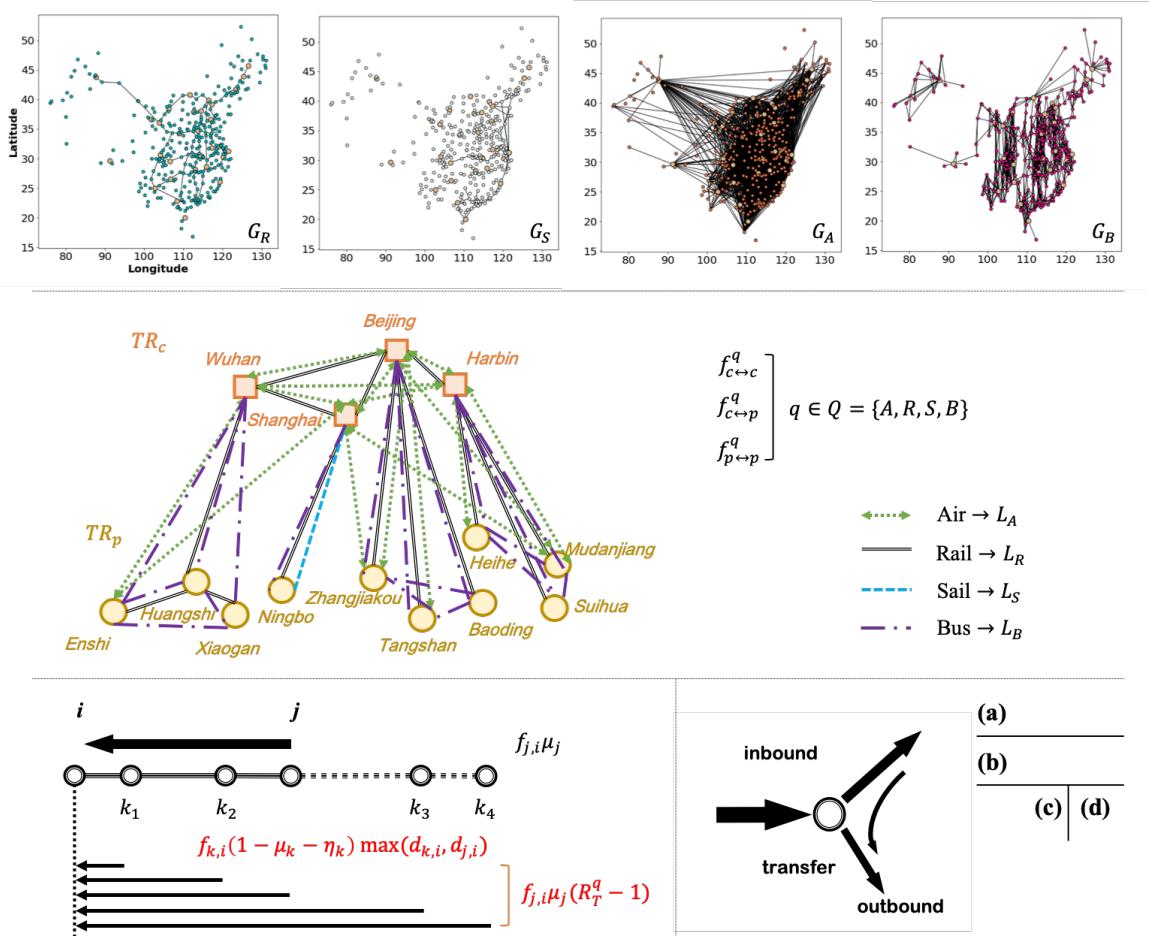with X being S, E or R.

Figure 1: Multi-channel transportation network model with detailed flow description. (a) Four layers of inter-city transportation (Air, Bus, Sail, Rail) between Chinese prefectural administrations. (b) Bi-partite categorization of nodes (central cities vs. peripheral cities). (c) Cross-infection during travel due to path overlaps. (d) Inbound/outbound flow upon city arrival.

Overall, the multi-layer network model is summarized in Figure 1 (see more details in (Li, 2020)).

EpiRank

With the constructed model framework, essentially a simulator of the spread of epidemics, we are able to simulate imaginary diseases of arbitrary epidemiological features originated from arbitrary epicenters. Suppose an epidemic initiated at node i, with a certain set of epidemiological parameters R_0, D_E, D_I. It is important to quantify the intensity of this epidemic, i.e., the extent of its spread on the domestic scale. This points to a centrality measure of the node that characterizes nodes' ability in spreading an epidemic. Borrowing from the idea of PageRank (*Page et al.*, 1999; *Langville and Meyer*, 2011), we construct this new centrality score in a way similar to the eigenvalue centrality and term it as the *EpiRank*.

Start the simulator with the disease seeded at node i upon a given zoonotic force, specified with $z(t)$ during period t_z^s to t_z^e. Consider a constant force z_0 over such a period (whose length is Δt_z), and the overall zoonotic force Z is:

$$Z = \int_{t_z^s}^{t_z^e} z(t)dt \sim \sum_{t_z^s}^{t_z^e} z(t) = (t_z^e - t_z^s)z_0 = \Delta t_z z_0. \tag{7}$$

The simulation proceeds, during which the local disease spreads to the entire nation via the transportation network. After τ time steps, we obtain the number of infected cases at city j, denoted as I_{ij}^τ, with the first subscript indicating the epicenter. Similarly we obtain R_{ij}^τ, E_{ij}^τ etc. We define the normalized total infection at city j as $U_{ij}^\tau = (I_{ij}^\tau + R_{ij}^\tau)/Z$. U_{ij}^τ is then used to compute the epidemic score EpiRank for node i:

$$h_i^\tau = \alpha \sum_{j \in G, j \neq i} \frac{f(U_{ij}^\tau)}{\sum_{j \in G} f(U_{ij}^\tau)} h_j^\tau + (1 - \alpha)U_{ii}^\tau. \tag{8}$$

$f(U_{ij}^\tau)$ represents a specific function of U_{ij}^τ to indicate the relative weights of each city j contributing to the score for city i. Here we simply allow $f(U_{ij}^\tau) = U_{ij}^\tau$ but further considerations could be made, for example, applying a cutoff $f(U_{ij}^\tau) = max(U_{ij}^\tau - U_0, 0)$.

This score h_i thus indicates the spreading intensity of epidemics at any node i, which is contributed by (1) the city's local severity of the epidemic, and (2) its ability of spreading the disease to other cities, with the intensity scores of other cities contributing to its own score at particular weights. $\alpha \in [0, 1]$ is the modulating parameter weighing over these two effects: a small α concerns more on the city-wide local spread of the epidemic ($\alpha = 0$ corresponds to a complete local index), while a large α puts more weight on the city's capability of spreading the disease out.

Write $\boldsymbol{W} = \{W_{ij}\} = \{U_{ij}^\tau / \sum_{j \in G} U_{ij}^\tau\}$, then $\boldsymbol{h} = \{h_i^\tau\}$ could be calculated with:

$$\begin{cases} h_i^\tau = \alpha \displaystyle\sum_{j \in G, j \neq i} \frac{U_{ij}^\tau}{\sum_{j \in G} U_{ij}^\tau} h_j^\tau + (1 - \alpha)U_{ii}^\tau \\[2ex] \Longrightarrow (1 + \alpha \frac{U_{ii}^\tau}{\sum_{j \in G} U_{ij}^\tau})h_i^\tau = \alpha \displaystyle\sum_{j \in G} \frac{U_{ij}^\tau}{\sum_{j \in G} U_{ij}^\tau} h_j^\tau + (1 - \alpha)U_{ii}^\tau \\[2ex] \Longrightarrow (1 + \alpha W_{ii})h_i^\tau = \alpha \displaystyle\sum_{j \in G} W_{ij} h_j^\tau + (1 - \alpha)U_{ii}^\tau \\[2ex] \Longrightarrow [\boldsymbol{I} - \alpha(\boldsymbol{W} - \text{diag}(\boldsymbol{W}))]\boldsymbol{h} = (1 - \alpha)\text{diag}(\boldsymbol{U})\mathbb{1} \\[2ex] \Longrightarrow \boldsymbol{h} = (1 - \alpha)[\boldsymbol{I} - \alpha(\boldsymbol{W} - \text{diag}(\boldsymbol{W}))]^{-1}\text{diag}(\boldsymbol{U})\mathbb{1}. \end{cases} \tag{9}$$

When $\alpha = 0$, $\boldsymbol{h} = \text{diag}(\boldsymbol{U})\mathbb{1}$. When $\alpha = 1$, $\boldsymbol{h}$ is the eigenvector of $[\boldsymbol{W} - \text{diag}(\boldsymbol{W})]$ of eigenvalue 1, in which case we may impose a value for $\max(\boldsymbol{h})$; wlog, we consider $\alpha \neq 1$. Note $\boldsymbol{h} = \boldsymbol{h}(\Delta t_z, z_0)$ but not $\boldsymbol{h} = \boldsymbol{h}(Z)$, since $\boldsymbol{U} = \boldsymbol{U}(\Delta t_z, z_0) \neq \boldsymbol{U}(Z)$, i.e., the spread patterns are different under different distributions of the same overall zoonotic force.

Results

Connectivities of each layer (i.e., transportation routes) are determined from public datasets and empirical considerations; city information (population, GDP etc.) is obtained from public datasets; transportation parameters (flowmap, transfer rate, cross-infection strength) are determined through fitting the early spread of COVID-19 in China in January-February 2020, where the multi-parameter inversion is conducted via a smart gradient method (see (*Li*, 2020) and Supplemental Materials). Transfer rate at central/peripheral cities are $TR_c = 0.4$ and $TR_p = 0.05$; the cross-infection strength are $\{R_T^A, R_T^R, R_T^S, R_T^B\} = \{1.2, 1.5, 1.5, 3\}$ on the four transportation media; inter-city flows are different for different types of city pairs (central-central, central-peripheral, peripheral-peripheral), and are determined at $f_{cc/cp/pp}^A = 1000/500/0$, $\hat{f}_{cc/cp/pp}^R = 2000/200/500$, $f_{cc/cp/pp}^S = 100/100/100$, $f_{cc/cp/pp}^B = 0/3000/1000$. These transportation parameters well fit the early spread of COVID-19 in China; they are independent of epidemiological concerns and are fixed throughout the simulations.

For epidemiological parameters, the zoonotic force z is assumed to be 5 persons/day at Day 1 and zero afterwards, at a single epicenter (the simulator nevertheless allows for simultaneous bursts at multiple epicenters). The base-case disease is fixed at $R_0 = 2.5$, $D_E = 6\ days$ and $D_I = 3\ days$, i.e., a mild reproduction of virus and a medium-range infection duration, close to the clinical parameters of COVID or SARS (e.g., *Wu et al.*, 2020). Unintervened spread of this seeded disease to all Chinese prefectural districts is simulated for 30 days, after which the ever-infected population $(I_i + R_i)$ in each city i are recorded and are used to compute the hazard index h.

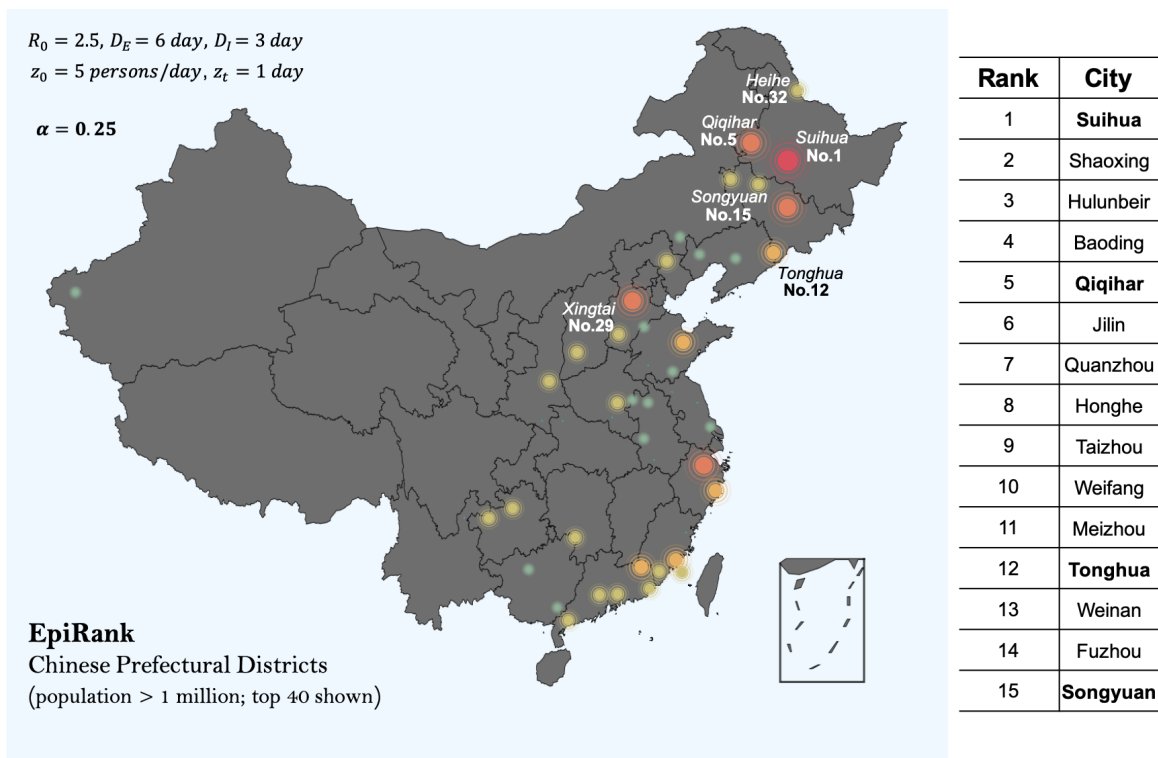| Rank	City
1	**Suihua**
2	Shaoxing
3	Hulunbeir
4	Baoding
5	**Qiqihar**
6	Jilin
7	Quanzhou
8	Honghe
9	Taizhou
10	Weifang
11	Meizhou
12	**Tonghua**
13	Weinan
14	Fuzhou
15	**Songyuan**

Figure 2: Urban Epidemic Hazard Index (EpiRank) of Chinese Cities (with population greater than 1 million). Assuming a simulated disease with $R_0 = 2.5$, $D_E = 6\ days$, $D_I = 3\ days$, seeded by a 1-day zoonotic force at the strength of 5 *persons/day*; $\alpha = 0.25$. Table: Top 15 ranks. Among 300 cities, the six ground-true cities are successfully indicated at high rankings (four in top 5% and all six in top 10%).

Cities having population larger than 1 million (300 out of 347) participate in the calculation of $\boldsymbol{h}$ under equation (9). For a value of $\alpha = 0.25$, the determined rankings of cities' epidemic hazards are shown in Figure 2 (top 40 in the graph and top 15 in the table). One sees that, quite strikingly, the six small cities

where the new bursts of COVID took place (Tonghua, Songyuan, Suihua, Qiqihar, Heihe, Xingtai) are successfully highlighted by the computed hazard index. All six cities rank within or near top 10% in the list, including four cities ranking within the top 15. Tests suggest that the result is robust; the high ranks of the six denoted ground-true cities are largely invariant to fluctuations in both transportation parameters and epidemiological parameters.

The hazard rankings are computed at different α, and correlation of the ranks is demonstrated via the Spearman's correlation coefficient (Figure 3; comparing top 30 entries of each rank). A stable ranking at small values of α is identified, along with a second invariance at the larger end (Figure 3 center). Indeed, the ranking is almost completely different at, for example, $\alpha = 0.1$ vs. $\alpha = 0.8$, with the latter having a new set of cities ranked top in the list which are mostly located in the middle of China. This is consistent with our theory, as a small/large α points to either of the two end-members of epidemic hazards: the city-wide local spread, or the capability of spreading the disease to other locations. Therefore, cities ranking high at small α are regions with mass population but relatively small transportation means, in which case local epidemic bursts are severely harbored (*Iacus et al.*, 2020) but not much spillovered to other cities. On the opposite, cities ranking high at large α are regions where inter-city transportation is sufficiently viable with respect to the humble local population; in this case, when seeded a virus, the city is less likely to become a closed epidemic cluster than to enormously spread the disease out to other regions. The epidemic hazard index thus also implies the condition of a city's inter-city transportation infrastructures.
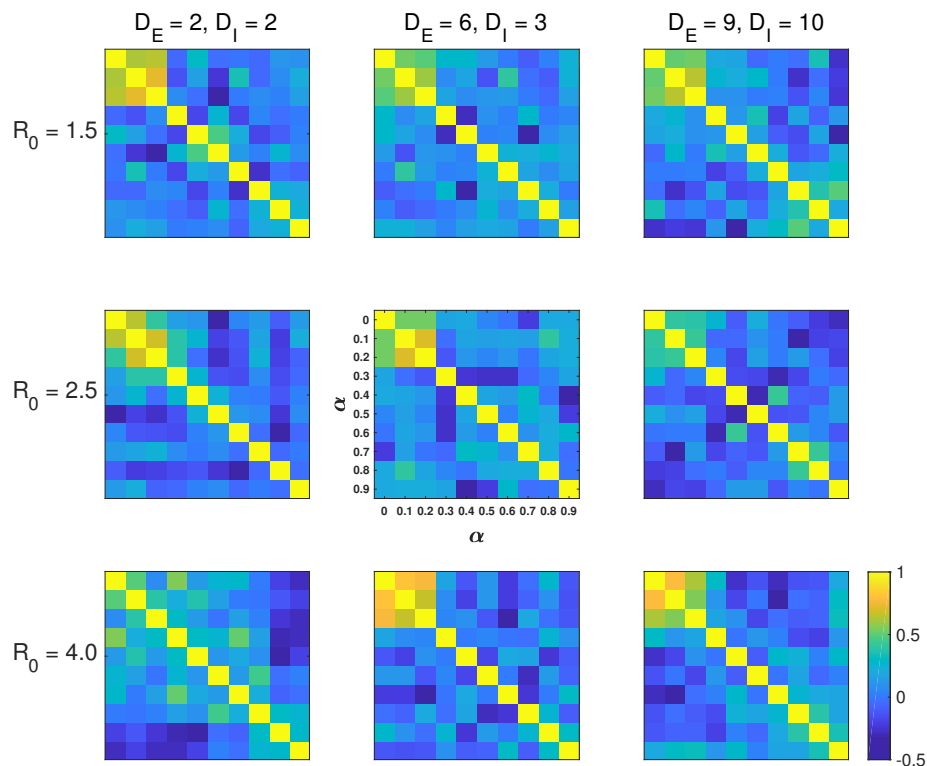

Figure 3: Spearman's correlation of hazard rankings (top 30 in the list) at different α (each sub-figure), for different sets of epidemiological parameters: reproduction number $R_0 = 1.5/2.5/4$ and infection duration $(D_E, D_I) = (2, 2)/(6, 3)/(9, 10)$ days. Base case (center sub-figure) is $R_0 = 2.5$ and $(D_E, D_I) = (6, 3)$.

We initiated simulations for different sets of epidemiological parameters of the assumed disease (Figure 3), with the combination of low/medium/high infectivity ($R_0 = 1.5/2.5/4$) and short/medium/long infection

duration $((D_E, D_I) = (2,2)/(6,3)/(9,10)$ days). The invariance at small α is largely maintained across the experiments, expect for a very severe virus with high infectivity and short duration $(R_0 = 4.0, (D_E, D_I) = (2,2))$. In this case, 30 days is sufficient for most population in most cities to get infected, and thus top rankings lean instead on densely populated cities. The second invariance around high α is also identifiable, although not as clear as the first one. In some cases there is a third cluster at intermediate values of α, but its significance is not as high as the first two which have well-grounded interpretations. Overall, one is able to conclude that the two end-options of the epidemic hazard index, using small or large α, hold meaningful across different scenarios of epidemic onset.
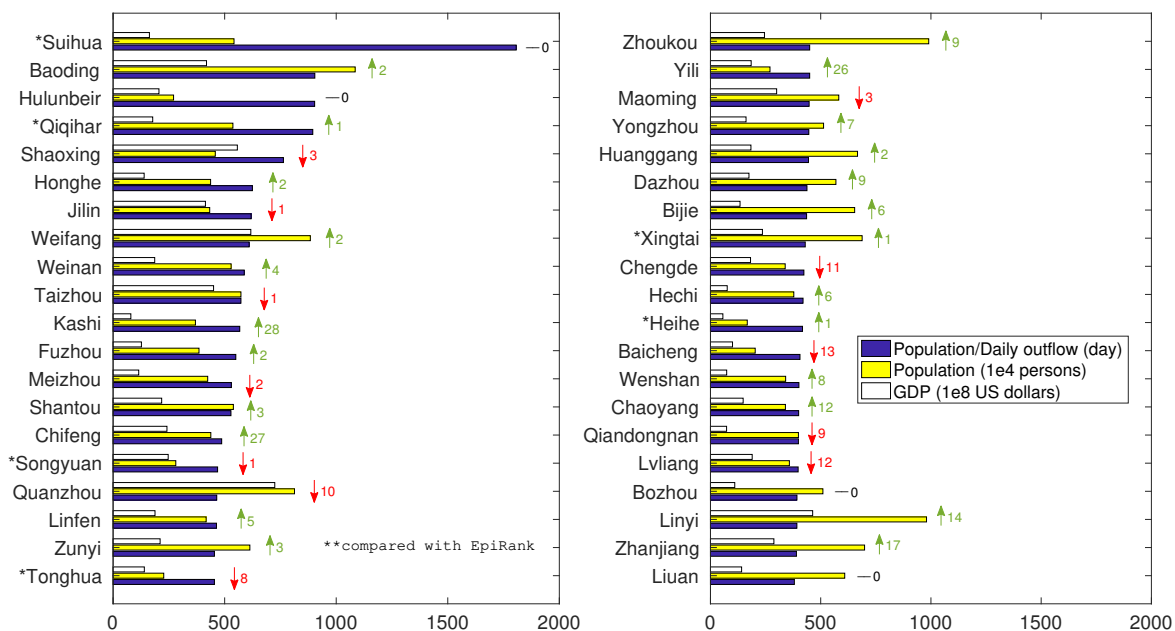

Figure 4: Population/daily outflow as a simplified epidemic hazard indicator (showing top 40 cities). Ranking compared to EpiRank (results in Figure 2); differences marked with green, red or black arrows and numbers. The ratio is effective but less accurate than EpiRank in highlighting the six epicenters (with star marks). Shear population or city GDP does not provide the same indication.

Discussion

From the model and results one deduces that, the high epidemic hazard of these small-scale cities computed at low α, in which case h_i draws heavily on a city's own infection U_{ii}, derives from the combined effect of two factors: a relatively large local population, and a small inter-city transportation flow. Intuitively, a serious epidemic cluster at the regional scale is going to develop, when the region is sufficiently populated, and not much inter-city outflow is dispersing the infection out of the epicenter. This inspires the idea that alternatively, we could compute the population/outflow ratio of each city and use this quantity to indicate cities' epidemic hazard. Results show that (Figure 4) similar to EpiRank, this ratio does serve as a good hazard indicator, under which the six denoted cities are listed with high ranks; furthermore, by contrast, shear population or city GDP, arguably two most considered social-economic indicators of urban regions (*Niu et al.*, 2020), are not valid to reflect the ground-true ranking. Conceptually, analysis on EpiRank help us pin down these two quantities among various social-economic factors in establishing a promising mathematical explanation of the observed phenomenon. For robust tests, we proportionally increased and decreased values on the flowmaps; results suggest that the effectiveness of this ratio (and certainly the

effectiveness of EpiRank) in highlighting the six epicenters is largely invariant to changes in absolute flow strength.

Nevertheless, it is seen that the simple population/outflow ratio, although still effective and easy to compute, is not as accurate and informative as EpiRank. This time the six ground-true cities are overall lower ranked, with only 2 out 6 in top 5%. This is because this ratio only considers a city's own population and transportation condition, whereas EpiRank takes a full account of the regional and then the entire national picture, under the networked dynamics approach. Indeed, it is not empirically inconsistent to argue that the six high-epidemic-risk cities are all located in the north, not only because they themselves have large population and small inter-city outflow, but also because it is exactly that cities in northern China, with which the six cities exchange most outflow population, all tend to have such features and therefore the effect of local clusters is further locked in. The advantage of EpiRank is implied; certainly, the simple ratio is also not able to reveal cities' ability of spreading the disease out, as EpiRank can shed light on with high α.

Although a promising quantitative explanation for the researched phenomenon is developed, it is yet indiscreet to conclude that EpiRank is by any means a sufficiently accurate index of urban epidemic hazards. The current dynamic network model draws little besides the two aspects, urban population and inter-city transportation, and too many real-world factors are left out. Validation of EpiRank results is also difficult to be conducted in a systematic and rigorous way, besides using the six epicenters as the ground truth. Despite a mathematically consistent and empirically effective approach, the proposed simulation framework and the constructed EpiRank index needs further analysis and extensive tests in various settings (e.g., to investigate the situation in the US (e.g., *Chang et al.*, 2021)), before their powers and shortcomings could be substantially uncovered; this study only serves as a first attempt.

References

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., & Wu, J. (2020), Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, *Journal of Clinical Medicine*, 9(2), 462.

Xiao, Y., Tang, S., & Wu, J. (2015), Media impact switching surface during an infectious disease outbreak, *Scientific reports*, 5, 7838.

Zhang, J., Lou, J., Ma, Z., & Wu, J. (2005), A compartmental model for the analysis of SARS transmission patterns and outbreak control measures in China, *Applied Mathematics and Computation*, 162(2), 909-924.

Li, T. (2020), Simulating the spread of epidemics in China on multi-layer transportation networks: Beyond COVID-19 in Wuhan, *EPL*, 130(4), 48002.

van den Driessche, P. , & Watmough, J. . (2002), Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Mathematical Biosciences*, 180(1-2), 29-48.

Oppenheim, B. , Gallivan, M. , Madhav, N. K. , Brown, N. , & Ayscue, P. . (2019), Assessing global preparedness for the next pandemic: development and application of an epidemic preparedness index, *British Medical Journal Global Health*, 4(1), e001157.

Boyce, M. R. , & Katz, R. . (2020), Rapid urban health security assessment tool: a new resource for evaluating local-level public health preparedness, *BMJ Global Health*, 5(6).

Warsame, A., Blanchet, K., & Checchi, F. (2020), Towards systematic evaluation of epidemic responses during humanitarian crises: a scoping review of existing public health evaluation frameworks, *BMJ global health*, 5(1).

Shearer, F. M., Moss, R., McVernon, J., Ross, J. V., & McCaw, J. M. (2020), Infectious disease pandemic planning and response: Incorporating decision analysis, *PLoS medicine*, 17(1), e1003018.

Di Marco, M., Baker, M. L., Daszak, P., De Barro, P., Eskew, E. A., Godde, C. M., ... & Ferrier, S. (2020), Opinion: Sustainable development must account for pandemic risk, *Proceedings of the National Academy of Sciences*, 117(8), 3888-3892.

Hamele, M., Neumayer, K., Sweney, J., & Poss, W. B. (2018), Always ready, always prepared?preparing for the next pandemic, *Translational Pediatrics*, 7(4), 344.

Simpson, S., Kaufmann, M. C., Glozman, V., & Chakrabarti, A. (2020), Disease X: accelerating the development of medical countermeasures for the next pandemic, *The Lancet Infectious Diseases*.

Liu, W. J., Bi, Y., Wang, D., & Gao, G. F. (2018), On the centenary of the Spanish flu: being prepared for the next pandemic, *Virologica Sinica*, 33(6), 463-466.

Thompson, R. N., Hollingsworth, ... & Restif, O. (2020), Key questions for modeling COVID-19 exit strategies, *Proceedings of the Royal Society B*, 287(1932), 20201405.

Neumann, G., & Kawaoka, Y. (2019), Predicting the next influenza pandemics, *The Journal of infectious diseases*, 219, S14-S20.

Morse, S. S., Mazet, J. A., ... & Daszak, P. (2012), Prediction and prevention of the next pandemic zoonosis, *The Lancet*, 380(9857), 1956-1965.

Reed, C., Biggerstaff, M., ... & Jernigan, D. B. (2013), Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics, *Emerging infectious diseases*, 19(1), 85.

Trock, S. C., Burke, S. A., & Cox, N. J. (2012), Development of an influenza virologic risk assessment tool, *Avian Diseases*, 56(4s1), 1058-1061.

Cox, N. J., Trock, S. C., & Burke, S. A. (2014), Pandemic preparedness and the influenza risk assessment tool (IRAT), In *Influenza pathogenesis and control-Volume I* (pp. 119-136). Springer, Cham.

Burke, S. A., & Trock, S. C. (2018), Use of influenza risk assessment tool for prepandemic preparedness, *Emerging infectious diseases*, 24(3), 471.

Holloway, R., Rasmussen, S. A., Zaza, S., Cox, N. J., Jernigan, D. B., & Influenza Pandemic Framework Workgroup. (2014), Updated preparedness and response framework for influenza pandemics, *Morbidity and Mortality Weekly Report: Recommendations and Reports*, 63(6), 1-18.

World Health Organization, (2016), Tool for influenza pandemic risk assessment (TIPRA) (No. WHO/OHE/PED/GIP/2016.2), *World Health Organization*.

Dobson, A. P., Pimm, S. L., ... & Vale, M. M. (2020), Ecology and economics for pandemic prevention, *Science*, 369(6502), 379-381.

Wu, J. T., & Cowling, B. J. (2011), The use of mathematical models to inform influenza pandemic preparedness and response, *Experimental Biology and Medicine*, 236(8), 955-961.

Bo, Y. C., Song, C., Wang, J. F., & Li, X. W. (2014), Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in Mainland China, *BMC Public Health*, 14(1), 1-13.

Greiving, S., Fleischhauer, M., & Lückenkötter, J. (2006), A methodology for an integrated risk assessment of spatially relevant hazards, *Journal of environmental planning and management*, 49(1), 1-19.

Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E., & Real, L. A. (2002), Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes, *Proceedings of the National Academy of Sciences*, 99(6), 3668-3672.

Balcan, D., Gonçalves, B., Hu, H., Ramasco, J. J., Colizza, V., & Vespignani, A. (2010), Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model, *Journal of computational science*, 1(3), 132-145.

Studzinski, N. G., & Pasteur, L. (2020), Comprehensive pandemic risk management: A systems approach, Visiting International Research Fellow Policy Institute, King?s College, London.

McKay, S., Boyce, M., Chu-Shin, S., Tsai, F. J., & Katz, R. (2019), An Evaluation Tool for National?Level Pandemic Influenza Planning, *World Medical & Health Policy*, 11(2), 127-133.

Eichner, M., Schwehm, M., Duerr, H. P., & Brockmann, S. O. (2007), The influenza pandemic preparedness planning tool InfluSim, *BMC infectious diseases*, 7(1), 17.

Grima, S., Kizilkaya, M., Rupeika-Apoga, R., Romānova, I., Gonzi, R. D., & Jakovljevic, M. (2020), A Country Pandemic Risk Exposure Measurement Model, *Risk Management and Healthcare Policy*, 13, 2067.

Christidis, P., & Christodoulou, A. (2020), The Predictive Capacity of Air Travel Patterns during the Global Spread of the COVID-19 Pandemic: Risk, Uncertainty and Randomness, *International Journal of Environmental Research and Public Health*, 17(10), 3356.

Chu, A. M., Tiwari, A., & So, M. K. (2020), Detecting early signals of COVID-19 global pandemic from network density, *Journal of travel medicine*, 27(5), taaa084.

So, M. K., Tiwari, A., Chu, A. M., Tsang, J. T., & Chan, J. N. (2020), Visualising COVID-19 pandemic risk through network connectedness, *International Journal of Infectious Diseases*.

So, M. K., Chu, A. M., Tiwari, A., & Chan, J. N. (2020), On Topological Properties of COVID-19: Predicting and Controling Pandemic Risk with Network Statistics, *medRxiv*.

Prieto, J., Malagón, R., Gomez, J., & León, E. (2020), Urban Vulnerability Assessment for Pandemic surveillance, *medRxiv*.

Sterman, J. D. (2000), Business dynamics: systems thinking and modeling for a complex world (No. HD30. 2 S7835 2000).

Newman, M. (2010), Networks: an introduction, *Oxford university press*.

Wu, J. T., Leung, K., & Leung, G. M. (2020), Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *The Lancet*.

Shelach, G. (2006), Leadership strategies, economic activity, and interregional interaction: social complexity in northeast China, *Springer Science & Business Media*.

Yew Lun Tian (2021), As China COVID-19 cases rise, millions more placed under lockdown, *Reuters*, Jan. 13.

Steven Lee Myers (2021), Facing New Outbreaks, China Places Over 22 Million on Lockdown, *New York Times*, Jan. 15.

Qiu, Y., Chen, X., & Shi, W. (2020), Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China, *Journal of Population Economics*, 33, 1127-1172.

Cao, Y., Liu, R., Qi, W., & Wen, J. (2020), Spatial heterogeneity of housing space consumption in urban China: Locals vs. inter-and intra-provincial migrants, *Sustainability*, 12(12), 5206.

Ye, Y., Xu, X., Wang, S., Wang, S., Xu, X., Yuan, C., ... & Wu, X. (2020), Evaluating the Control Strategies and Measures for COVID-19 Epidemic in Mainland China: A City-level Observational Study.

Li, A., Liu, Z., Luo, M., & Wang, Y. (2021), Human mobility restrictions and inter-provincial migration during the COVID-19 crisis in China, *Chinese Sociological Review*, 53(1), 87-113.

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... & Scarpino, S. V. (2020), The effect of human mobility and control measures on the COVID-19 epidemic in China, *Science*, 368(6490), 493-497.

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., ... & Tatem, A. J. (2020), Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature*, 585(7825), 410-413.

Tian, H., Liu, Y., Li, Y., Wu, C. H., Chen, B., Kraemer, M. U., ... & Dye, C. (2020), An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China, *Science*, 368(6491), 638-642.

Ge, J., He, D., Lin, Z., Zhu, H., & Zhuang, Z. (2020), Four-tier response system and spatial propagation of COVID-19 in China by a network model, *Mathematical Biosciences*, 330, 108484.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999), The PageRank citation ranking: Bringing order to the web, *Stanford InfoLab*.

Langville, A. N., & Meyer, C. D. (2011), Google's PageRank and beyond: The science of search engine rankings, *Princeton university press*.

Hu, M., ... & Lai, S. (2020), Transmission risk of SARS-CoV-2 on airplanes and high-speed trains, *medRxiv*.

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021), Mobility network models of COVID-19 explain inequities and inform reopening, *Nature*, 589(7840), 82-87.

Copiello, S., & Grillenzoni, C. (2020), The spread of 2019-nCoV in China was primarily driven by population density, Comment on "Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China" by Zhu et al, *Science of The Total Environment*, 744, 141028.

Zhu, S., Bukharin, A., Xie, L., Santillana, M., Yang, S., & Xie, Y. (2020). High-resolution Spatio-temporal Model for County-level COVID-19 Activity in the US, *arXiv*:2009.07356.

Liu, J., Hao, J., Sun, Y., & Shi, Z. (2021), Network Analysis of Population Flow among Major Cities and Its Influence on COVID-19 Transmission in China, *Cities*, 103138.

Niu, X., Yue, Y., Zhou, X., & Zhang, X. (2020), How Urban Factors Affect the Spatiotemporal Distribution of Infectious Diseases in Addition to Intercity Population Movement in China, *ISPRS International Journal of Geo-Information*, 9(11), 615.

Feng, S., Feng, Z., Ling, C., Chang, C., & Feng, Z. (2021), Prediction of the COVID-19 epidemic trends based on SEIR and AI models, *PloS one*, 16(1), e0245101.

Iacus, S. M., Santamaria, C., Sermi, F., Spyratos, S., Tarchi, D., & Vespe, M. (2020), Human mobility and COVID-19 initial dynamics, *Nonlinear Dynamics*, 101(3), 1901-1919.

Budd, J., Miller, B. S., ... & McKendry, R. A. (2020), Digital technologies in the public-health response to COVID-19, *Nature medicine*, 26(8), 1183-1192.

Li, R. (2021), Mobility restrictions are more than transient reduction of travel activities, *PNAS*, 118(1).

Schlosser, F., ... & Brockmann, D. (2020), COVID-19 lockdown induces disease-mitigating structural changes in mobility networks, *PNAS*, 117(52), 32883-32890.

Jia, J. S. , Lu, X. , Yuan, Y. , Xu, G. , & Christakis, N. A. (2020), Population flow drives spatio-temporal distribution of covid-19 in china, *Nature*, 582(7812), 1-11.

Zhang, J., Dong, L., Zhang, Y., Chen, X., Yao, G., & Han, Z. (2020), Investigating time, strength, and duration of measures in controlling the spread of COVID-19 using a networked meta-population model, *Nonlinear Dynamics*, 101(3), 1789-1800.

Wang, B., Gou, M., Guo, Y., Tanaka, G., & Han, Y. (2020), Network structure-based interventions on spatial spread of epidemics in metapopulation networks, *Physical Review E*, 102(6), 062306.