LETTER

Simulating the spread of epidemics in China on multi-layer transportation networks: Beyond COVID-19 in Wuhan

To cite this article: Tianyi Li 2020 *EPL* **130** 48002

A Letters Journal Exploring the Frontiers of Physics

May 2020

EPL, **130** (2020) 48002
doi: 10.1209/0295-5075/130/48002

www.epljournal.org

Simulating the spread of epidemics in China on multi-layer transportation networks: Beyond COVID-19 in Wuhan

Tianyi Li

System Dynamics Group, Sloan School of Management, Massachusetts Institute of Technology Cambridge, MA, USA

received 13 March 2020; accepted in final form 4 June 2020
published online 18 June 2020

PACS 89.75.-k – Complex systems
PACS 89.65.-s – Social and economic systems
PACS 95.75.-z – Observation and data reduction techniques; computer modeling and simulation

Abstract – A general-purpose simulator for the spread of epidemics in China is built. The Chinese public transportation system between over 340 prefectural-level cities is modeled as a *multi-layer* network having block structure, with layers representing different means of transportation (airlines, railways, sails, buses), and nodes representing two categories of cities (central, peripheral). At each city, an open-system SEIR model tracks the local spread of the disease, with population in- and out-flow exchanging with the overlying transportation network. The model accounts for 1) different transmissivities on different transportation media, 2) the transit of inbound flow, 3) cross-infection in public transportation due to path overlap, and that 4) infected population not entering public transportation, 5) recovered population not subject to repeated infections. The model is used to simulate the spread of COVID-19 in China; it shows that the framework is robust and reliable, and best-fit inversion results match public datasets to an extraordinary extent.

The COVID-19 [1–4] has incurred heavy casualties and tremendous economic loss in China. With the first case confirmed as early as December 8th, 2019 [5] and effective measures implemented in a national scale in China around 50 days later, the new Coronavirus has claimed over 80000 cases and more than 3200 deaths in China. The epidemic took place right before the Chinese New Year, and the massive population flow across the entire country aggravated its vicious spread. Facing the severe threat of the epidemic, since late January, the Chinese government has taken strong measures to quench the disease, mobilizing and coordinating available forces in a unprecedented scale, and thanks to the collective efforts of Chinese citizens, the epidemic has been largely put under control in most Chinese provinces by mid-February. Nevertheless, although the situation in China is picking up, the disease started to propagate worldwide, invading almost all countries in East Asia and soon reached Europe, America and Australia [6,7], and the epidemic will inevitably leave profound effects on the global economy. Right after the burst of COVID-19, studies set out to model the epidemic and use simulation results to nowcast and forecast its intensity [8–10]. COVID-19 is compared with the SARS in 2003, and people discovered that this time the virus is more contagious but less fatal. Arguably, the incubation period of COVID-19 is believed to be longer than that of SARS, and the early symptoms are less salient [11]. Reference [12] is the first formally published simulation model for COVID-19. Because this disease is originated only in Wuhan and is carried to other cities [12], assembles an SEIR model for Wuhan with population in- and out-flows, treating infections outside Wuhan as imported cases. The model is calibrated with the time series data of the confirmed cases in a list of global cities, and a posterior estimate of the basic reproduction number $R_0 = 2.68$ is carried out. Simulation models similar to the flavor of [12] are of great importance for policy decision-making, whose findings may significantly help guide the rescue and response for the epidemic.

Nevertheless, a few problems arise with the model in [12], hampering it to reach sufficient resolution. One big concern is that the model does not differentiate means of transportation during population flow. The public transportation system is composed of airlines, railways, sail routes and buses; especially in China where public transportation is exploited to a great extent, different means

of transportation leads to different contact rates and thus different transmissivities of the disease during travel, *e.g.*, the spread of epidemic is substantially easier on trains or buses than on airplanes. The differentiation of means of transportation is a must for high-resolution epidemic models regarding inter-city population flow since people spend non-trivial time on the route and may engage in various activities. Moreover, there exists cross-infection during travel due to path overlaps. Since railway, sails and bus travels are often not end-to-end, the spread of disease will likely take place along the way, among people taking the same vehicle yet having different destinations. Again, such an infection scenario is not negligible in China where the transportation system is crowded. Note that similar cross-infection concerns may arise in transportation service locations such as train stations and airports; yet concerns in these places are not as severe as during the travel where people are more densely located and have little choice of isolating from their neighbors who have unknown points of departure and destinations.

Second, ref. [12] initiates an SEIR compartment model only in Wuhan and does not capture the local evolution dynamics of the imposed diseases at other cities. This only holds true if the disease is not sufficiently contagious, or after the quarantine procedures are successfully implemented in all cities. Ideally, the same SEIR dynamics are supposed to be initiated in every city, contributing to the spread of the disease in a nationwide landscape. Third, the model neglects an important aspect of the population flow, the non-trivial occurrence of transit events in public transportation, and thus may possess a fundamental system error. Although highly developed, the Chinese transportation system could not (even get close to) realize cheap end-to-end travels between over 340 prefectural-level cities. Yet Chinese citizens are also much more likely than citizens of other countries to travel a long distance to a major city (*e.g.*, to make a living) due to an imbalanced development across the nation, which gives rise to the non-trivial role of transit during inter-city public transportation. Last, the model in [12] allows population out-flow of Wuhan from all four SEIR compartments. This may contain another system error as it is more appropriate to assume that the infected population are not traveling but instead stay local. Moreover, since the model did not trace the source of the population inflow, it has to put the recovered population into the susceptible compartment and essentially assumes repeated infections, which are often not the case for virus-triggered diseases.

The above problems could be resolved by assembling a network of the transportation system on top of the local evolution of epidemics and formulate open-system compartments, which will bring the model resolution to the next stage. On this network, nodes represent population districts (communities/cities/countries) and edges describe transportation availability between nodes. An identical compartment model is initiated at each node, which generates its local dynamics of the epidemic under the in- and out-flow of population that it exchanges from the overlying transportation network. Among various methodologies in epidemic research [13], this modeling approach has been widely taken by previous studies. Simulators are built to study the spread of epidemics and help design corresponding control policies after the burst of SARS [14], H5N1 [15], H1N1 [16], on a national [17–19] or global [14] scale. In these studies, the transportation system under concern is often considered as aggregated or of single-layer type [20,21] with few exceptions [17,22]; nevertheless, one notes that in a finer resolution, the public transportation system could be further modeled as a *multi-layer* network [23–25], with each layer representing a specific means of transportation [26–31]. On this network, nodes are maintained at different layers, in which case the network is sometimes termed as "multiplex" networks [32,33]. As in the single-layer representation, transportation takes place along network edges [34], captured by a set of flowmaps that record the flow between each pair of connected nodes on each layer. One expects that the multi-layer representation of the transportation system outstands the aggregated single-layer representation since by differentiating means of transportation, the model could account for different diffusion properties in the spread of epidemics; such a granular model might provide valuable insights for policy analysis such as the emergency response and policy measures regarding certain layers of transportation.

Consider the transportation network $\boldsymbol{G}$. At each node $i \in \boldsymbol{G}$, an SEIR model [35] is assembled, where the population P_i is divided into 4 compartments (Susceptible, Exposed, Infected, Recovered): $P_i = S_i + E_i + I_i + R_i$. Adopting the notations in [12], in the base model, the dynamics are governed by $\dot{S}_i = -\frac{S_i}{P_i}(\frac{R_0}{D_I} I_i + z)$, $\dot{E}_i = \frac{S_i}{P_i}(\frac{R_0}{D_I} I_i + z) - \frac{E_i}{D_E}$, $\dot{I}_i = \frac{E_i}{D_E} - \frac{I_i}{D_I}$, $\dot{R}_i = \frac{I_i}{D_I}$, where R_0, D_E, D_I are the basic reproduction number, the mean incubation period and the mean infection period, respectively; $z = z(t)$ is the zoonotic force, which is only nonzero for a certain period at the source node of the epidemics. Note that the venue from E_i directly leading to R_i is cut off, which exists in the general SEIR model. These dynamics work for a closed population which could not account for the spread of diseases due to the population flow between nodes. The flow is maintained on the overlying transportation network $\boldsymbol{G}$ which is cast as multi-layer, with each layer representing a specific means of transportation, including airline (A), railway (R), sail (S) and bus (B). We denote the set of layers as $Q = \{A, R, S, B\}$. These layers share the same set of nodes $V = \{i\}$ (cities), while having different sets of edges L_q for $q \in Q$ according to the availability of transportation infrastructures. Edges are undirected since the transportation between two cities is bi-way. Note that we do not have a separate layer for the (private) car travel and instead assume that cars share the routes in the bus layer (see below). International population flow are not included in the current model. Thus

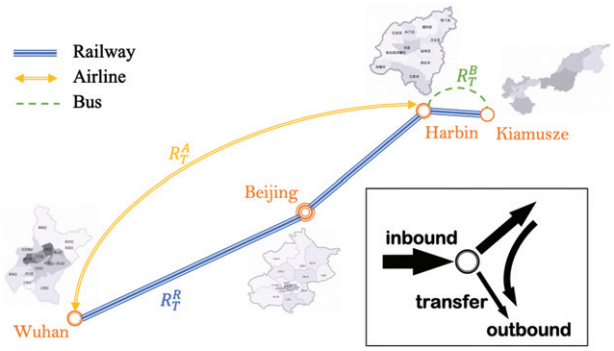

Fig. 1: Illustration of the multi-layer inter-city transportation network. People move between cities (nodes) by different means of transportation, possibly with multiple hops on the network across different layers. Cross-infection takes place during the travel, with different transportation media having different levels of transmissivity R_T. Each city i is associated with a specific transfer rate TR_i, the fraction of the inbound flow leaving for the next destination upon arrival.

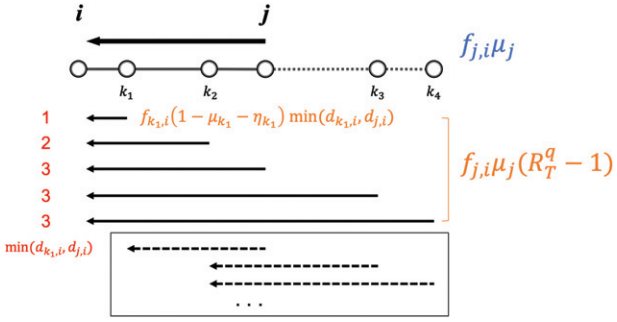

Fig. 2: Illustration of cross-infection during travel due to path overlap. If there are exposed cases in the population flow from city j to city i, then cross-infection takes place on all routes $k \longrightarrow i$ that shares a part with the route $j \longrightarrow i$. The spillover $f^q_{j,i}\mu_j(R^q_T - 1)$ of exposed cases is allocated proportionally to the susceptible population $f_{k,i}(1-\mu_k-\eta_k)$ among the outflow of city k, weighted by the shared length of the two route. Routes that share a finite part with $j \longrightarrow i$ but not terminate at i are not considered in the spillover.

we have (fig. 1):

$$\boldsymbol{G} = \begin{cases} G_A = (V, L_A), \\ G_R = (V, L_R), \\ G_S = (V, L_S), \\ G_B = (V, L_B). \end{cases} \tag{1}$$

Two properties differentiate means of transportation, besides layers' specific edge connectivities determined by the availability of infrastructures. First, for each q, there is a specific transfer rate $TR^q_i \in [0, 1]$ for every city i, which represents the proportion of inflow to city i that will be in transit and leave for the next destination, thus not entering the population stock P_i. Such transfer rates will be small for small cities and large for cities that are regional transportation centers. Ideally, the transfer rate will also be higher for certain transportation means such as railways and lower for others, say airlines (direct flights are primary among domestic air travels); yet for simplicity in the current model we assume that it is the same for all means of transportation at a certain city, *i.e.*, $TR^q_i = TR_i$. Second, on different transportation media the likelihood that an exposed person will infect others during travel is not the same, which is represented by an *in-travel* basic production number $R^q_T > 1$. Expectedly, such a number will be lower on airplanes than on trains or buses where people talk more often; it is safe to assume that R^q_T is homogeneous on different edges of layer q.

R_T characterizes the cross-infection of the disease during public transportation. Importantly, cross-infection has spillovers due to path overlaps on public transport media: a patient traveling from city j to city i by rail will likely spread the virus to everyone in the train during the travel, including those who might get on the train at a different city k. This effect is pervasive in the real world, especially in China where the public transportation is crowded, and could be exempted only for end-to-end travels like airlines and cars. Note that it is difficult to account for cross-infection in an aggregated network model, since transportation routes are not identical on different layers and thus layers cannot be combined. To model this path-dependent feature, we make two assumptions (fig. 2). First, it is assumed that only those people that get off the train at the same destination i as the patient will be infected. Implicitly, this is assuming that passengers sharing the same destination are more likely to stay close in travel, which could be viewed as valid in real practice as the closest contacts in travel often occur near the period of getting on and off the vehicle, the latter taking place right among people having the same destination; assigning tickets of the same destination to neighboring seats is also known to be the case for the ticket system in many situations. In a complete manner, the spillover will likely take place between any two routes that share a finite part; yet this will open a new dimension in the computation of the spillover and incur redundant complexity. Second, we assume that the strength of the spillover effect is proportional to the ratio of the (shortest path) distance between city k and i and the distance between j and i, for a node k that lies between j and i in the path; for a node k that lies beyond j in the path (*i.e.*, earlier in the path), the $k \to i$ distance is thresholded by the $j \to i$ distance.

On the transportation network, we track the in- and out-flow of the E, S, R population of each city. For each q, the flowmap is characterized by a matrix $F^q = \{f^q_{j,i}\}$, which is not guaranteed to be symmetric since the flow $f^q_{i,j}$ from city i to j is not equal to the flow $f^q_{j,i}$. Upon all layers, at time t, the inflow of the exposed population to city i summarizes over all cities and all means of transportation:

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

with the fraction of people in transit deducted. Here $\overline{f^q_{j,i}\mu_j}$ is the adjusted exposed population flow from city

j to i by means q, taking care of the spillover effect in cross-infection, given by ($d^q_{i,j}$ is the shortest path distance between i and j on layer q):

$$\overline{f^q_{j,i}\mu_j} = f^q_{j,i}\mu_j + \sum_{k}^{p^q(i,k)\cap p^q(i,j)\neq 0} f^q_{k,i}\mu_k(R^q_T - 1) \times \frac{f^q_{j,i}(1-\mu_j-\eta_j)\min(d^q_{j,i}, d^q_{k,i})}{\sum_{l}^{p^q(i,k)\cap p^q(i,l)\neq 0} f^q_{l,i}(1-\mu_l-\eta_l)\min(d^q_{l,i}, d^q_{k,i})}. \quad (3)$$

And we have

$$\mu_i(t) = \frac{\Delta E^{out}_i(t) + \sum_{q\in Q}\sum_{j\in V}\overline{f^q_{j,i}(t-1)\mu_j(t)}TR^q_i}{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t)} \quad (4)$$

and

$$\eta_i(t) = \frac{\Delta R^{out}_i(t) + \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)\eta_j(t-1)TR^q_i}{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t)}, \quad (5)$$

which are the (time-stamped) proportion of the E and R population among the total outflow population of city i, respectively. Therefore, $(1-\mu_i-\eta_i)$ is the proportion of the S population. The exposed population among the total outbound is contributed by both the exposed originated from city i (ΔE^{out}_i, see below) and the exposed among the transit of inbound (second term in the numerator of μ_i), and is dynamically updated at each step; we denote the proportion by $\mu_i(t)$. Similarly, we track the stock (via proportion $\eta_i(t)$) of the recovered population among the total outflow of a city. This tracking is an important caveat in this model to make sure that a recovered person will unlikely to be infected again, and only the susceptible population $(1-\mu_i-\eta_i)$ among the total outflow along each edge will be subject to cross-infection.

The outflow population from city i's population P_i is the total outbound flow minus the transferred inbound flow, which is contributed to by the S, E, R compartments (not I), different from the case in [12]. Proportionally, the outflow of the exposed is

$$\Delta E^{out}_i(t) = E_i(t)\frac{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t) - \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)TR^q_i}{S_i(t)+E_i(t)+R_i(t)}. \quad (6)$$

Note that the flowmaps and the transfer rates should ensure that at each city, the inbound transfer flow should always be smaller than the outbound flow, *i.e.*,

$$\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t) - \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)TR^q_i > 0 \implies TR^q_i < \frac{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t)}{\sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)}, \quad \forall\, i\in V, \quad q\in Q. \quad (7)$$

For the recovered population, the in-flow is tracking all the recovered people (through η_j) upon arrival:

$$\Delta R^{in}_i(t) = \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)(1-TR^q_i)\eta_j(t-1), \quad (8)$$

and the out-flow is proportional to ΔE^{out}_i:

$$\Delta R^{out}_i(t) = R_i(t)\frac{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t) - \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)TR^q_i}{S_i(t)+E_i(t)+R_i(t)}. \quad (9)$$

For the susceptible population, according to the flow balance, ΔS^{in}_i is the total un-transferred inflow subtracting the recovered and the exposed in-flow:

$$\Delta S^{in}_i(t) = \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)(1-TR^q_i) - \Delta E^{in}_i(t) - \Delta R^{in}_i(t) \quad (10)$$

and

$$\Delta S^{out}_i(t) = S_i(t)\frac{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t) - \sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)TR^q_i}{S_i(t)+E_i(t)+R_i(t)}. \quad (11)$$

In the end, we arrive at the open-system SEIR model accounting for the transportation flow:

$$\begin{cases} \dot{S}_i = -\frac{S_i}{P_i}\left(\frac{R_0}{D_I}I_i + z\right) + \Delta S^{in}_i - \Delta S^{out}_i, \\ \dot{E}_i = \frac{S_i}{P_i}\left(\frac{R_0}{D_I}I_i + z\right) - \frac{E_i}{D_E} + \Delta E^{in}_i - \Delta E^{out}_i, \\ \dot{I}_i = \frac{E_i}{D_E} - \frac{I_i}{D_I}, \\ \dot{R}_i = \frac{I_i}{D_I} + \Delta R^{in}_i - \Delta R^{out}_i, \end{cases} \quad (12)$$

with μ_i and η_i updated according to (4) and (5). Note that in the model we do not assume a balance of the inbound and outbound flow at each city; there is no such equilibrium for the population flow in real practice, and each city's transient population is in constant dynamics. However, (7) should always hold at each city.

When real datasets could be acquired, ideally, the flow $f^q_{j,i}$ from node j to node i by means q, is supposed to be cast as a time series that incorporates seasonal features but approximately remains invariant on the annual basis. Then for a specific starting time of an epidemic, a cursor is placed on the annual curve and a window of the flowmap is then cut out from this point on for the usage in simulation. In the cases where only the aggregated flow of all transportation layers between two cities $F_{i,j} = \sum_q \{f^q_{j,i}\}$ is available, one may apply a multinomial logit model to determine the flow for each layer, as in [36]. For the transfer rate, if the city's (average/annual) transient population P^T_i is available, then TR could be calculated by comparing the aggregated out-flow and the transient population:

$$TR_i = \frac{\sum_{q\in Q}\sum_{j\in V} f^q_{i,j}(t) - P^T_i}{\sum_{q\in Q}\sum_{j\in V} f^q_{j,i}(t)}. \quad (13)$$

When datasets are not available, based on empirical observations, the multi-layer network could be further

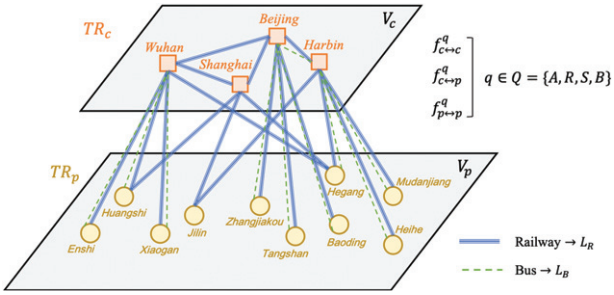

Fig. 3: Illustration of the two-block structure of the transportation network. Cities are categorized as *central cities* V_c and *peripheral cities* V_p. For demonstration, only two layers L_R and L_B are shown. The block structure facilitates the determination of the transfer rate TR and the flowmaps F.

represented by a two-block structure, with nodes V divided into two categories: *central cities* V_c and *peripheral cities* V_p (fig. 3). Upon this division, we assume different flow strength on edges between $V_c \leftrightarrow V_p$, $V_c \leftrightarrow V_c$ and $V_p \leftrightarrow V_p$ when constructing the flowmaps F. We are also able to assign a homogeneous value of TR to each category of nodes. These treatments bring down the model's parameter space to a feasible region (2 for TR, $4\times3 = 12$ for F). This division of cities' roles is valid in China, since Chinese cities are often categorized on a level base [37]. In the current study we apply a two-bin partition, regarding all provincial capital cities plus level 0–2 cities as central cities V_c (54 counts, table S1 in the Supplementary material `Supplementarymaterial.pdf` (SM)) with the rest 293 cities belong to V_p (centrality features on the multi-layer graph support this division, fig. S2 in the SM); expectedly, in a finer scale, cities could be categorized into more bins, and the flowmap will have a multi-block structure.

We use the model to simulate the spread of the COVID-19 in Chinese cities, with parameters determined from empirical considerations. Edge connectivities are different for each layer [38] as different layers of the public transportation system play different roles in serving the demand of population flow (fig. S1 in the SM). For airlines, we obtain public data on the airline schedules for 90 cities; for railways, we use the 11 primary railroads in China which occupy the major railway passenger flow, connecting ∼120 cities; for sails, there are roughly 19 passenger lines in China yet they take a very small part (<2%) in Chinese public transportation; for buses, as schedules are difficult to collect, according to empirical considerations, we construct edges between cities within a mid-range geological distance (150 km), and between central cities to other cities in the province. For epidemiological parameters, we let $R_0 = 2.68$ and $D_E = 6$ days [12], and use both a short and a long D_I: 2.4 days [12], 10 days [39]; we starts the simulation at December 8th, 2019 with a nonzero zoonotic force $z = 27$ cases/day (the total number of confirmed cases in December) before January 1st the date of the closure of the Wuhan seafood market. We tested the values suggested by [12] ($z = 86$ extending 30 days) and results show a misfit three times larger than using our values for z; therefore the values in [12] are not used. For parameters of the transportation network, R_T is assumed to be: $\{R_T^A, R_T^R, R_T^S, R_T^B\} = \{1.2, 1.5, 1.5, 3\}$, smallest for airlines, largest for buses. Flowmaps are determined empirically (all units are people/day). Airlines are the major transportation means between central cities, and we determine $f^A_{cc/cp/pp} = 1000/500/0$, assuming no airline transport between peripheral cities, four flights between central cities per day, and half the amount between a central city and a peripheral city. For railways the flow is determined as $\hat{f}^R_{cc/cp/pp} = 2000/200/500$, and $f_{i,j} = \hat{f}_{i,j}/d_{i,j}$, *i.e.*, passenger flows are inversely proportional to the shortest path distance between cities, under the carriage capacity 2000 people/train. For sail routes, $f^S_{cc/cp/pp} = 100/100/100$, which occupy a small fraction of the total flow. For buses, we let $f^B_{cc/cp/pp} = 0/3000/1000$, assuming no bus travel between central cities; $f^A_{pp} = 0$ and $f^B_{cc} = 0$ reinforce the two-block structure of the model. Buses are heavily used in provincial transport, for travelers in peripheral cities to go to either the local central city or nearby peripheral cities. We let the numbers (3000 and 1000) be three/two times the normal flow (50 people/bus, 20 or 10 buses between central-peripheral and peripheral-peripheral) to account for the private transportation by cars that nevertheless contributes to the entire population flow, which is mostly provincial. Implicitly, as car travels are aggregated into bus travels, R_T^B represents the effective value watered down from a real transmissivity on buses. Note the aggregation of bus and car layers are valid in the sense of cross-infection as they are both considered as end-to-end, in which case the two layers have the same routes and thus could be aggregated. The assumed constant flowmaps do not consider seasonal effects, which might be significant in certain cases, exactly like for COVID-19 which took place right before the Chinese New Year. Although not obtained from a full-parameter inversion, the values for R_T and F that we use are nevertheless well tested by extensive forward simulation runs and are confirmed as quasi-locally-optimal values.

Given the large parameter space even for the simplified model (2 for TR, 12 for F, 4 for R_T and 5 epidemiological parameters), a full-parameter inversion is currently not feasible on personal computing devices. With other parameters determined with empirical considerations, TR_c and TR_p are set as open parameters and we initiate a partial inversion using the Powell optimizer [40]. At each run, the model extends 53 time steps (days), from December 8 (date of the first case minus an average incubation period), to January 23 (Chinese New Year's Eve). This stopping time is reasonable since in China most inter-city winter travels took place before this date, right after which the government took urgent measures and effectively shut down population flow. After January 23, confirmed cases of the epidemic are collected from each prefectural-level city and are reported to the public; we use this dataset for inversion [41]. We observe that

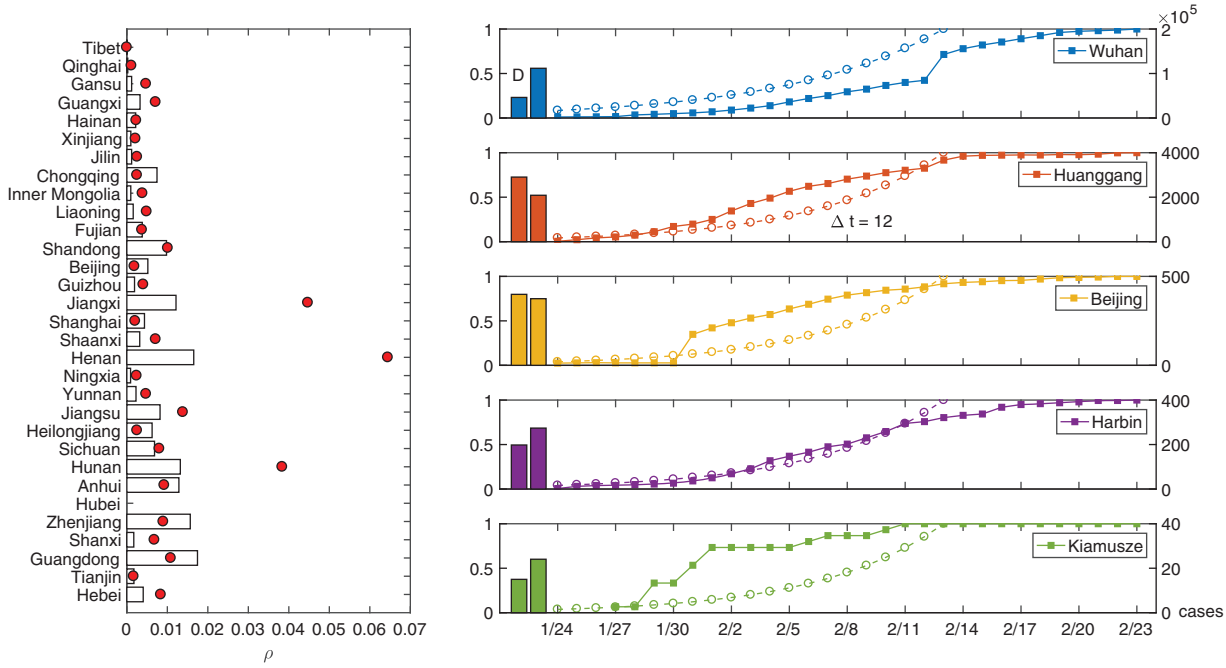

Fig. 4: Inversion results for $D_I = 2.4$ days. Left: fraction of confirmed cases in Chinese provincial districts except Hubei (Hong Kong, Macau and Taiwan not shown). Data: bars. Best-fit results: red dots. Right: normalized time series of the epidemic (left y-axis) and the absolute number of confirmed cases (bar, right y-axis) in Wuhan (origin), Beijing (capital), Huanggang (peripheral city in Hubei), Harbin (central city outside Wubei) and Kiamusze (peripheral city outside Wubei). Data: solid lines, first bar on the left. Best-fit results: dashed lines, second bar on the left. Simulation time series are shift to the left by $\Delta t = 12$.

the fraction of confirmed cases in each city ρ among national headcounts remains roughly invariant after around February 15th (fig. S3 in the SM) due to the reduction of daily new headcounts as the effect of nationwide measures, with Wuhan having ∼60% of all confirmed cases and all cities in the Hubei Province having >80% cases. Hence instead of matching the entire data time series, for the inversion we minimize the misfit between the simulated and realistic ρ obtained at the end of the two time series (summing over all cities except Wuhan in prevention of double-counting as $\sum_i \rho_i = 1$). This choice of misfit is also consistent with the fact that there is not a reliable estimate of the zoonotic force (which may not be a significant force for this disease, as informed by an anonymous reviewer of this paper) and hence our model focuses more on the relative strength of the epidemic in each city. Further, this treatment also obviates the potential system error of insufficient reporting, in which case the fraction rather than the absolute number of cases is certainly more reasonable to be used for inversion. Since the simulation stops at January 23 but confirmed cases are reported gradually afterwards, when calculating the ρ of the simulation result we regard the exposed as additional infected cases as they will eventually enter the stock, a valid treatment given that the incubation period is unlikely to be as long as 30 days (January 23 to February 23).

Although real datasets for flowmaps are not available at the current stage and a full-parameter inversion is not conducted, quite surprisingly, results of the coarse partial inversion nevertheless demonstrated great fitting performance of our model. The best-fit $TR_c = \mathbf{37.86\%}$, $TR_p = \mathbf{4.29\%}$ for $D_I = 2.4$ days, and $\mathbf{19.20\%}$ and $\mathbf{5.15\%}$ for $D_I = 10$ days, *i.e.*, on average around 20–40% inbound flow at central cities are in transit, and the number is roughly 5% for peripheral cities. These results are not contradictory to empirical observations, although it is difficult to further match them with real data (see the SM for discussion). We show the best-fit ρ of Chinese provincial districts (except Hubei; Hong Kong, Macau and Taiwan not shown) aggregated from prefectural-level cities for $D_I = 2.4$ days (results do not change much for $D_I = 10$ days; see fig. S4 in the SM). It suggests that our results recover the data to a satisfying extent (fig. 4, left): the severe situation of the disease in Hehan, Hunan, Shandong and Jiangxi is correctly revealed, while the spread in provinces such as Chongqing, Anhui, Zhengjiang, Guangdong and Heilongjiang is underestimated. This result might be consistent with public news that after the burst of the epidemic there are uncommonly large population flow directed into these provinces from Wuhan, a situation not captured by our simplified flowmaps. Entire best-fit time series are also shown (fig. 4, right) on five example cities: Wuhan, Beijing (capital), Huanggang (peripheral city in Hubei), Harbin (central city outside Wubei) and Kiamusze (peripheral city outside Wubei). The data (solid lines) are S-shape as the epidemic is gradually under control in February, while the simulation time series (dashed lines) are demonstrating exponential growth, as one expects

from the SEIR model. Nevertheless, applying a uniform time-shift ($\Delta t = 12$) assuming implicitly that all cities took measures at the same time, which was essentially the case in China, simulation results match the initial part of the real S-curve by a large margin. However, the matching of the two curves should not be over-interpreted as the model is always able to generate exponential growth and the value of the time-shift is not warranted. The absolute values of the confirmed cases, instead of the fraction ρ, are also fit reasonably well (bars in fig. 4, right), suggesting that our choice of model parameters makes certain sense (see the SM for other directions of future work). Overall, given the coarse treatment during the parameter determination and data-processing process, these partial inversion results are believed to be acceptable, suggesting that the model is promising in generating reference dynamics for the spread of epidemics in China.

In this study, an SEIR model is used as the baseline epidemic model; thus this simulator is only reliable to generate dynamics for the period where no effective government intervention has been implemented. After actions are taken, the SEIR compartments will be invalidated, and elaborated compartment models could be adopted to account for government's measures such as quarantine and the reportage of suspected cases [42], as well as the shutdown of transportation. Another extension to the model is to relax the assumption in the spillover of cross-infection, and instead allow that cross-infection occurs between all routes sharing a finite part in the path. This will increase one search depth in the computation and will be feasible on massive-scale clusters; a super computing device will also facilitate a full-parameter inversion for COVID-19, ideally with real datasets assembled for flowmaps, whose results may uncover further information for the study of this urgent on-going epidemic. Overall, constructed on the multi-layer network flow model, this general-purpose simulator for the city-level spread of epidemics in China has an adaptive nature that could be tuned for specific usage, which might be helpful for policy analysis in emergence response, and early warnings of future events.

REFERENCES

[1] Gralinski L. E. *et al.*, *Viruses*, **12** (2020) 135.
[2] Wang C. *et al.*, *The Lancet*, **395** (2020) 470.
[3] Zhou P. *et al.*, *Nature*, **579** (2020) 270.
[4] Zhu N. *et al.*, *New Engl. J. Med.*, **382** (2020) 727.
[5] Li Q. *et al.*, *New Engl. J. Med.*, **382** (2020) 1199.
[6] Holshue M. L. *et al.*, *New Engl. J. Med.*, **382** (2020) 929.
[7] Phan L. T. *et al.*, *New Engl. J. Med.*, **382** (2020) 872.
[8] Liu T. *et al.*, *Transmission dynamics of 2019 novel coronavirus (2019-nCoV)* (2020), available on SSRN.
[9] Majumder M. and Mandl K. D., *Early transmissibility assessment of a novel coronavirus in Wuhan, China*, (January 23, 2020), available on SSRN.
[10] Tang B. *et al.*, *Infectious Disease Model.*, **5** (2020) 248.
[11] Huang C. *et al.*, *The Lancet*, **395** (2020) 497.
[12] Wu J. T., Leung K. and Leung G. M., *The Lancet*, **395** (2020) 689.
[13] Riley S., *Science*, **316** (2007) 1298.
[14] Hufnagel L., Brockmann D. and Geisel T., *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004) 15124.
[15] Germann T. C. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 5935.
[16] Brockmann D. and Helbing D., *Science*, **342** (2013) 1337.
[17] Balcan D. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 21484.
[18] Ferguson N. M. *et al.*, *Nature*, **437** (2005) 209.
[19] Ferguson N. M. *et al.*, *Nature*, **442** (2006) 448.
[20] Colizza V. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 2015.
[21] Ruan Z. *et al.*, *Sci. Rep.*, **5** (2015) 11401.
[22] Saumell-Mendiola A. *et al.*, *Phys. Rev. E*, **86** (2012) 026106.
[23] Boccaletti S. *et al.*, *Phys. Rep.*, **544** (2014) 1.
[24] De Domenico M. *et al.*, *Nat. Phys.*, **12** (2016) 901.
[25] De Domenico M. *et al.*, *Phys. Rev. X*, **3** (2013) 041022.
[26] Aleta A. *et al.*, *Sci. Rep.*, **7** (2017) 44359.
[27] Cardillo A. *et al.*, *Eur. Phys. J. ST*, **215** (2013) 23.
[28] Chan W. K. V. and Hsu C., *IIE Trans.*, **47** (2015) 929.
[29] Gallotti R. and Barthelemy M., *Sci. Rep.*, **4** (2014) 6911.
[30] Kurant M. and Thiran P., *Phys. Rev. Lett.*, **96** (2006) 138701.
[31] Tsiotas D. and Polyzos S., *J. Comput. Netw.*, **3** (2015) 642.
[32] Nicosia V. *et al.*, *Phys. Rev. Lett.*, **111** (2013) 058701.
[33] Solé-Ribalta A., Gomez S. and Arenas A., *Phys. Rev. Lett.*, **116** (2016) 108701.
[34] Morris R. G. and Barthelemy M., *Phys. Rev. Lett.*, **109** (2012) 128703.
[35] Newman M., *Networks* (Oxford University Press) 2018.
[36] Zhang L. *et al.*, *Trans. Res. Part C*, **49** (2014) 73.
[37] Chinese cities are effectively ranked on a level base; see `https://baike.baidu.com`.
[38] Edge connectivities of the first three layers (airline, railway, sail) are collected from public datasets and available at `https://github.com/TimothyLi0123/WH`.
[39] He X. *et al.*, *Nat. Med.*, **26** (2020) 672.
[40] Powell M. J. D., *Comput. J.*, **7** (1964) 155.
[41] Datasets obtained from a public repository at `https://github.com/BlankerL/DXY-COVID-19-Data`.
[42] Zhang J. *et al.*, *Appl. Math. Comput.*, **162** (2005) 909.