Supplementary Materials for
The effect of interurban movements on the spatial distribution of population

Jiachen Ye,[1] Qitong Hu,[1] Peng Ji,[1,*] Marc Barthelemy[2,3,*]

[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University,
Shanghai 200433, China
Shanghai 200433, China
[2]Institut de Physique Théorique, Université Paris Saclay, CEA, CNRS,
F-91191 Gif-sur-Yvette, France
[3]Centre d'Analyse et de Mathématique Sociales, (CNRS/EHESS) 54,
Boulevard Raspail, 75006 Paris, France

[*]E-mail: pengji@fudan.edu.cn, marc.barthelemy@ipht.fr.

Data for 2019

In order to evaluate the heterogeneity of flows of 2019 with comparison to that of 2020, we use the migration index during the same period of 2019 (re-scaled according to Chinese lunar calendar, from Jan. 12th to Feb. 23rd, 2019), and we would compute the distribution of all flows $N(i, j, d)$, for all cities i and j and all days d, though $N(i, j, d) = N_{\text{out}}(i, d) \times p(i, j, d)$. The Chinese Lunar New Year of 2019 is Feb. 5th. Here, $N_{\text{out}}(i, d)$ is migration index reflecting the size of the population moving into or out from a city/province, and $p(i, j, d)$ is migration ratio capturing the proportion of each origins and destination. However, the migration ratio is unavailable for 2019. We apply the data of $p(i, j, d)$ for 2020 to the computation of $N(i, j, d)$ for 2019, with results shown in Fig. S1. This result exhibits large heterogeneity of flows and displays a localized drop around LNY.

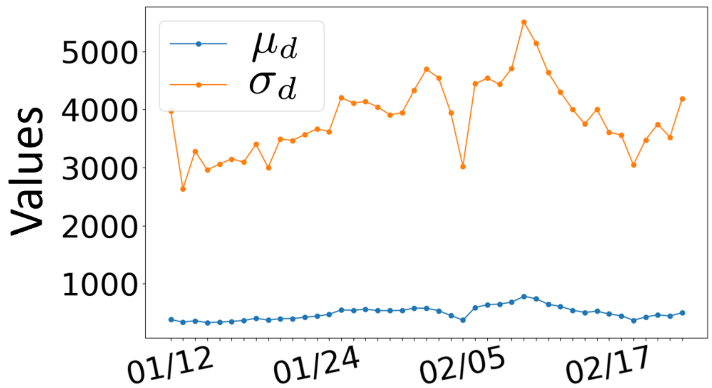

Figure S1: Temporal interurban flows for 2019. Average and standard deviation of N over traffic flows versus time for 2019 with the corresponding migration ratio for 2020.

Statistics of N_{in} and N_{out}

Versus population

We observe a power law relationship between the incoming/outgoing flows and city population in Fig. S2 which indicates that the larger a city and the more flows it carries. Compared to 2019, the differences between the scatter points for incoming flows corresponding to days before and after LNY are much larger in 2020. This result emphasizes again that travel ban causes indeed the sharp drop of standard deviation in Fig. 1 (b) of the main text rather than the low travel intention during the Spring Festival.

The power law fits that we obtain imply that $N_{\text{in}} \sim P_0^{\gamma_{\text{in}}}$ where $\gamma_{\text{in}} \approx 0.93$ before LNY and $\gamma_{\text{in}} \approx 0.88$ after LNY. Similarly for outgoing flows we obtain $\gamma_{\text{out}} \approx 0.85$ before and $\gamma_{\text{out}} \approx 0.93$ after LNY. We can interpret these results as a consequence of the conservation of the number of individuals traveling before and after LNY.

Distribution of $N_{\text{in,out}}$ for 2020

We show the distribution, average and standard deviation of incoming/outgoing flows in Fig. S3. We observe that these distributions are relatively broad, in particular outgoing flows (Fig. S3 (a) and (b)). We show the standard deviation of incoming flows and outgoing flows over cities for each day, and the corresponding average (the same for the incoming and outgoing flows) in Fig. S3 (c). Note that the standard deviation of N_{in} is smaller than that of N_{out} before Jan. 25th, the 2020 LNY, while the situation reverses after Jan. 25th. A reasonable explanation for this is that people go to a relative large number of hometowns from a relative small number of workplaces before the Spring Festival; Due to the travel ban, people do not come back after the Spring Festival.

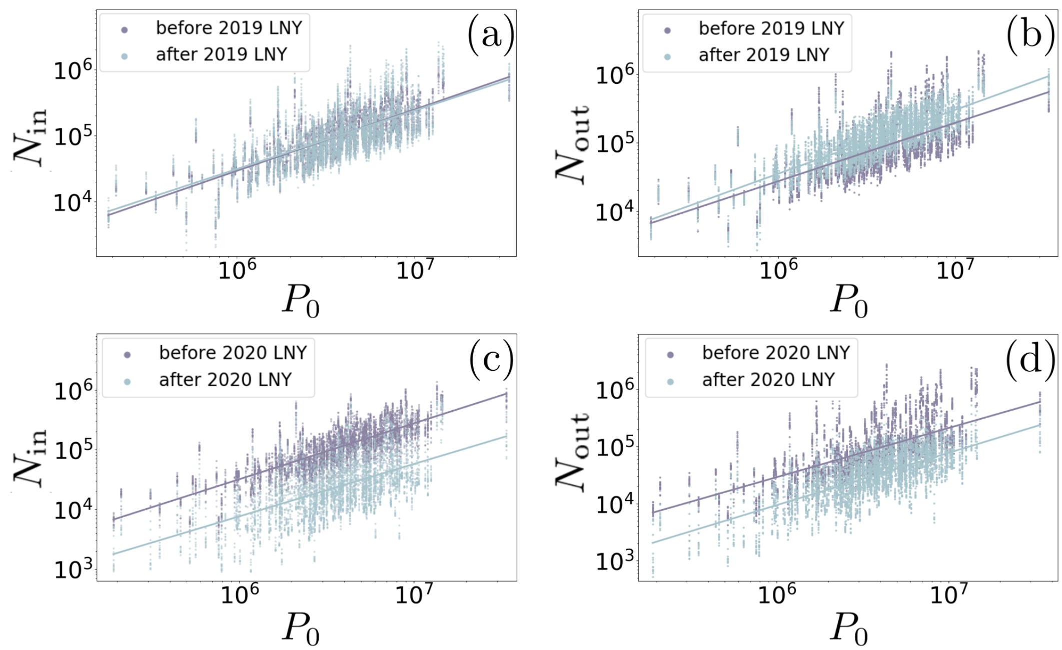

Figure S2: Relationship between incoming/outgoing flows and city population. Incoming flows (a) and outgoing flows (b) for all cities and days versus city population before and after the 2019 LNY in loglog. Incoming flows (c) and outgoing flows (d) for all cities and days versus city population before and after the 2020 LNY in loglog.

Distribution of $N_{\mathrm{in,out}}$ for 2019

We show the same quantities as above but for the year 2019. Here also, we observe broad distributions both for all incoming flows and all outgoing flows are shown in Fig. S4.

Fluctuations are larger around LNY where the total flow of individuals is larger, allowing for more heterogeneity. Before LNY individuals move from a large variety of cities to a relatively small number of hometowns explaining the large fluctuations of N_{out}. After LNY, individuals are returning from a small number of hometowns to a large variety of cities, inducing large fluctuations of N_{in}. The corresponding dispersion and relative dispersions Δ_d^{out}) and Δ_d^{in}) are shown in figures S4 (c) and (d). This also results in the heterogeneous distribution for the relative standard deviation of N_{in} and N_{out} averaged over cities for 2019 in Fig. S4 (e).

Gini indices for $N_{\mathrm{in,out}}$

We compute Gini indices for cities. Instead of showing results for all cities, we plot Gini indices versus the traffic flows to or from cities in set $\{i \in \mathcal{V} | \min_d \{\max_d \{\frac{N_{\mathrm{in}}(i,d)}{N_{\mathrm{out}}(i,d)}\}, \max_d \{\frac{N_{\mathrm{out}}(i,d)}{N_{\mathrm{in}}(i,d)}\}\} > 4.5\}$ and Wuhan in Fig. S5. In this case from these cities, many people go out to or come

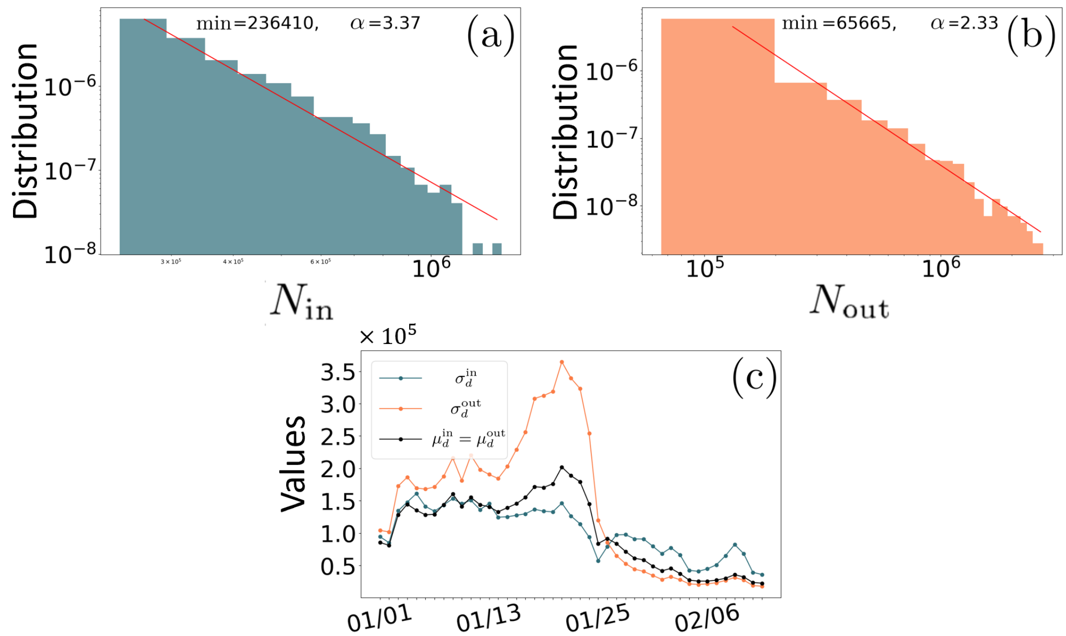

Figure S3: Distribution and variations of incoming and outgoing flows for 2020. (a) Distribution of all incoming flows in loglog with parameters of power law fitting on the top middle. (b) Distribution of all outgoing flows in loglog with parameters of power law fitting on the top middle. (c) Average and standard deviation of $N_{\rm in}$ and $N_{\rm out}$ over cities versus time.

in from many different cities. These cities are critical and include Shanghai, Beijing, and so on. Due to travel bans, Wuhan exhibits specific features of scatter points with clear separation before and after LNY (see the scatter points in blue at the upper left corner corresponding to Wuhan in Fig. S5). We also observe here in both cases a decreasing behavior on average. This is more salient for $N_{\rm out}$ where the trend is clearly visible. This indicates that for larger outgoing flows, the Gini is smaller with no clearly dominant flow.

Statistical structure of the national population

We first show the population distribution in Fig. S6 (a). We observe a broad distribution and a power law fit gives the exponent $\alpha \approx 5$. We also show the number of important cities quantified by the integer part of $n[1 - G(d)]$, in Fig. S6 (b).

Pendular ratio

In order to test the dependence of the pendular ratios on the criteria for defining classes, we change the criteria as follows: here if the value of $R(i, 1)$ is larger than 1.2, we classify

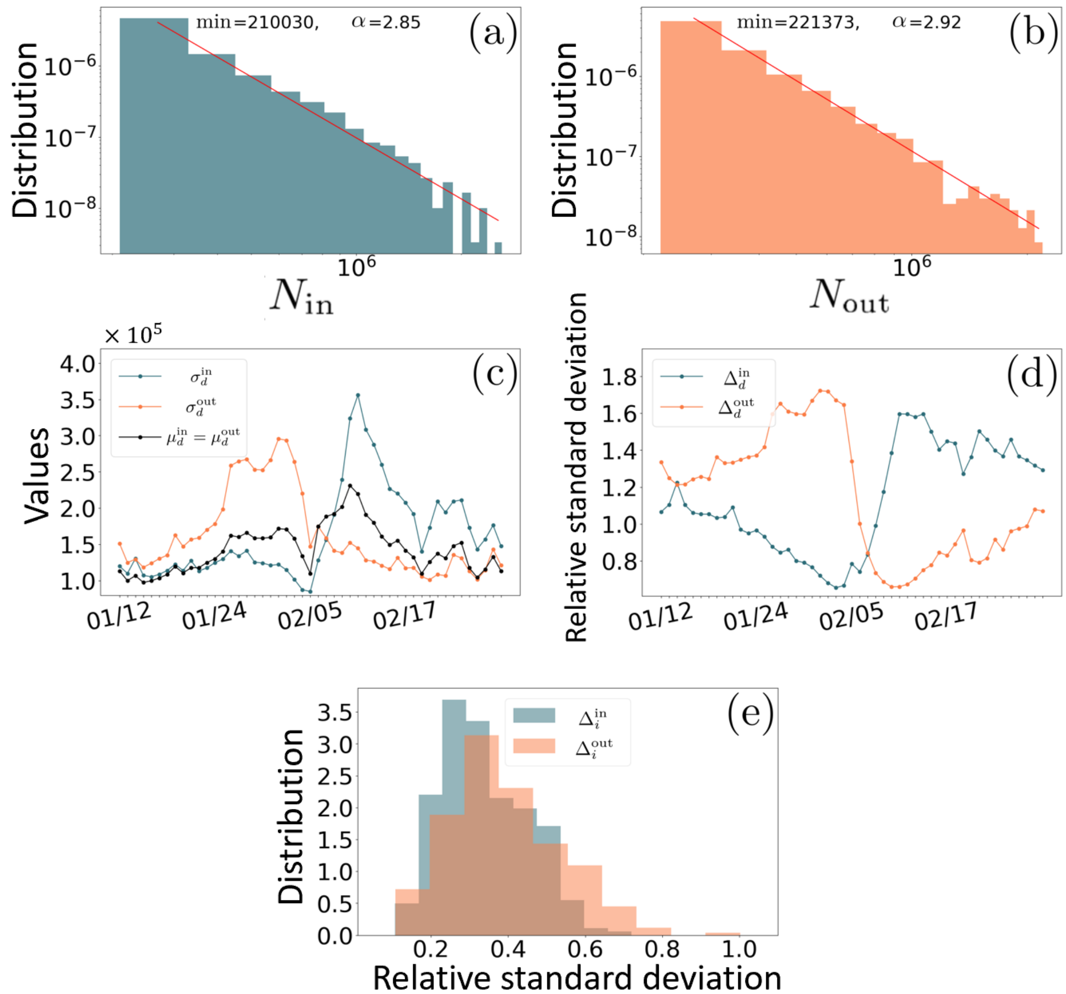

Figure S4: Distribution and variations of incoming and outgoing flows for 2019. (a) Distribution of all incoming flows in loglog with parameters of power law fitting on the top middle. (b) Distribution of all outgoing flows in loglog with parameters of power law fitting on the top middle. (c) Average and standard deviation of N_{in} and N_{out} over cities versus time. (d) Relative standard deviation of N_{in} and N_{out} over cities versus time. (e) Distribution of the relative standard deviation of N_{in} and N_{out} over time.

city i as a 'receiver' city. If the value is less than 0.8, we classify city i as an 'emitter' city. Finally, if the value is between 0.8 and 1.2, we classify city i as a 'transit' city. We show the location of three categories of cites on the map of China for 2019 and 2020 in Fig. S7.

Compared to the criteria in the main text, the number of transit cities decreases, while the number of emitter and receiver cities increases. However, as shown in Fig. S8, the patterns of the average value of pendular ratio corresponding to three categories of cites

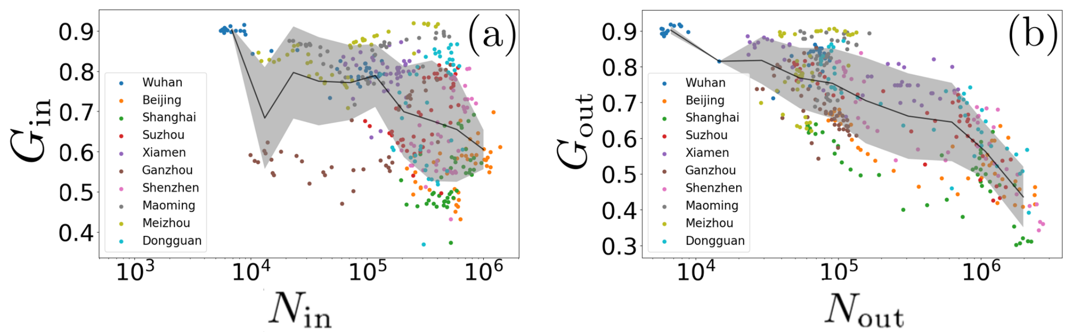

Figure S5: Relationship between Gini index and incoming/outgoing flows. (a) $G_{\rm in}$ versus $N_{\rm in}$ for some cities and all days. (b) $G_{\rm out}$ versus $N_{\rm out}$ for some cities and all days.

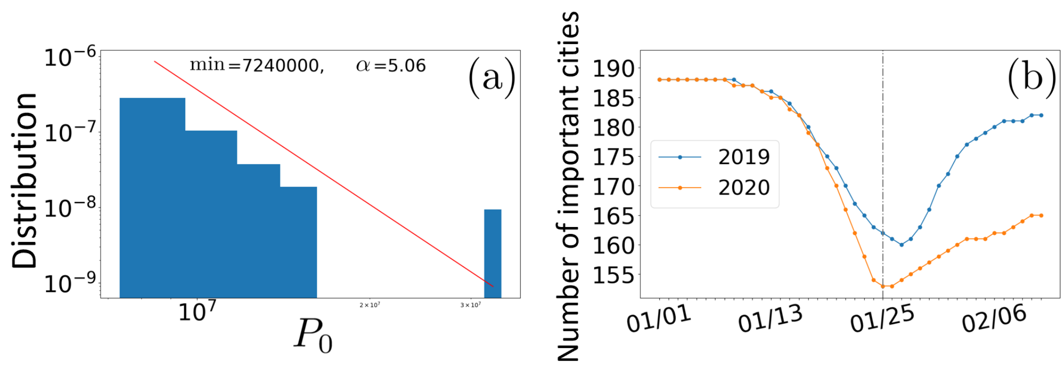

Figure S6: City population distribution and important cities.(a) Distribution of populations of all cities in loglog. (b) The number of important cities versus time corresponding to 2019 and 2020. We match the time scale for 2019 and 2020 according to LNY (for the sake of clarity, we show on the x-axis the dates for 2020 only). The vertical line highlights the LNY.

remain unchanged.

Statistics at the inter-province level

Statistics of flows

In what follows, we show some corresponding results based on province-level data instead of city-level data. The distribution of traffic flows between provinces exhibits large heterogeneity with the exponent of power law fit α around 2.24, as shown in Fig. S9 (a). A sharp drop of the standard deviation after Jan. 25th is observed in Fig. S9 (b).

We show the relative standard deviation of N over flows versus time with an order around 2.96 in Fig. S10 (a) and the distribution of the relative standard deviation of N over time concentrating around 0.64 in Fig. S10 (b). Large heterogeneity of traffic flows

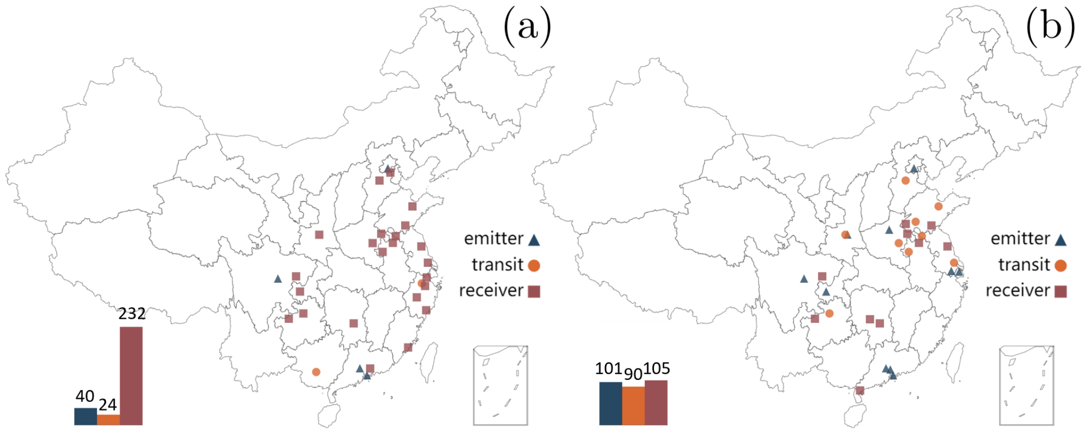

Figure S7: Spatial distribution of different categories of cites. (a) Emitter, receiver and transit cities according to the value of $R(i, 1)$ for 2019 with the number of three categories of provinces at the lower left corner. (b) Emitter, receiver and transit cities according to the value of $R(i, 1)$ for 2020 with the number of three categories of provinces at the lower left corner.

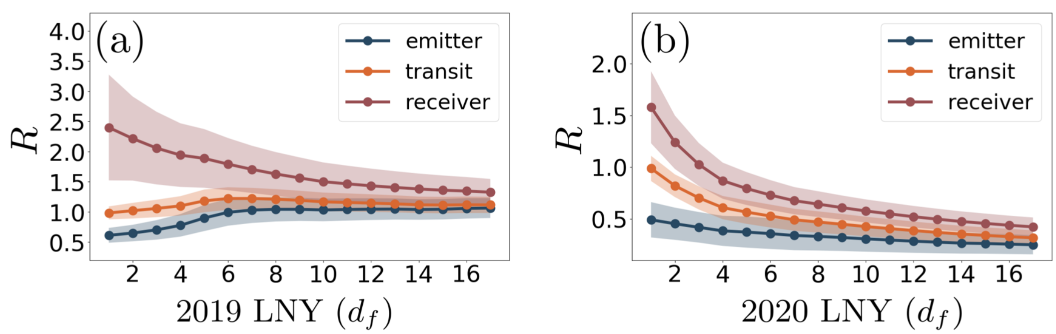

Figure S8: Pendular ratio comparison. (a) Mean value of the pendular ratios over cities according to the classification (receiver, emitter or transit cities) with different criteria versus days from LNY in 2019. The colored areas correspond to one standard deviation. (b) Mean value of the pendular ratios over cities according to the classification with different criteria versus days from LNY in 2020 with shaded areas representing standard deviation.

between provinces confirms the difficulty of modeling these flows. The relative standard deviations corresponding to incoming and outgoing flows with smaller relative dispersions are shown in figures S10 (c) and (d).

We compare the incoming flows and outgoing flows versus city population in 2019 and 2020 at province-level with days before and after LNY highlighted by different colors in Fig. S11. Compared to 2019 (figures S11 (a) and (b)), the differences between days before and after LNY are much larger in 2020 (figures S11 (c) and (d)).

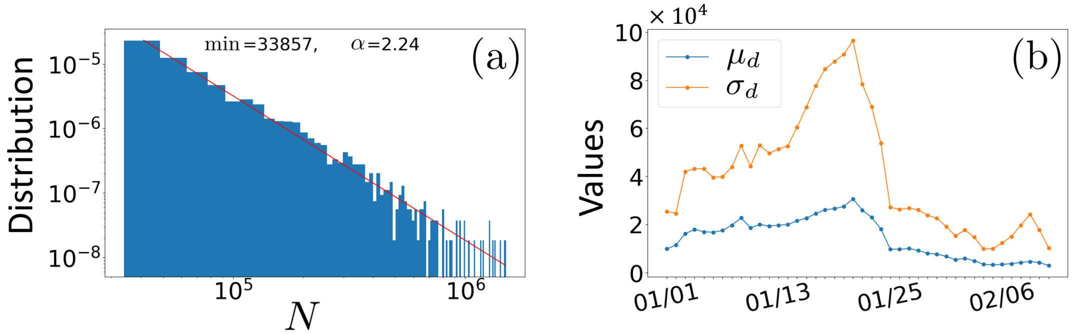

Figure S9: Statistics of inter-province flows. (a) Distribution of all traffic flows $N(i, j, d)$ in loglog. The line is a power law fit of the form $N^{-\alpha}$ with exponent $\alpha = 2.27$. (b) Average and standard deviation of N over traffic flows versus time.

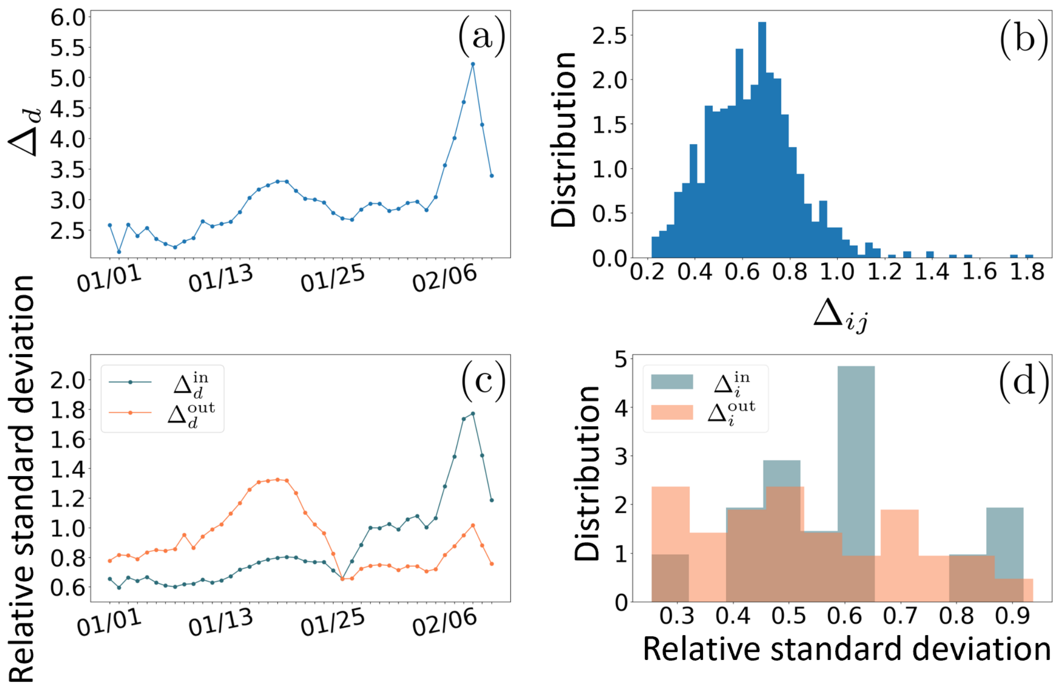

Figure S10: Incoming and outgoing flows at the inter-province level. (a) Relative standard deviation of N over traffic flows versus time. (b) Distribution of the relative standard deviation of N over time. (c) Relative standard deviation of N_{in} and N_{out} over provinces versus time. (d) Distribution of the relative standard deviation of N_{in} and N_{out} over time.

Statistical structure of the national population

The trends of the population Gini index for 2019 and 2020 are shown in Fig. S12. The Gini index reaches its maximum around LNY and returns to normal state gradually.

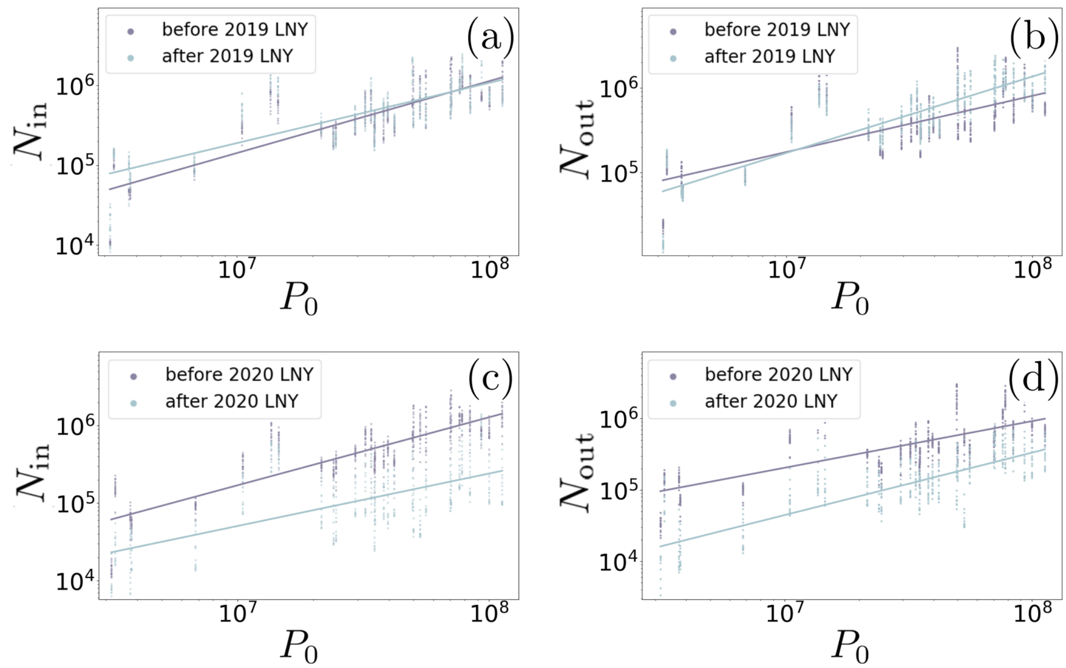

Figure S11: Relationship between incoming/outgoing flows and province population. Incoming flows (a) and outgoing flows (b) for all provinces and days versus province population before and after the 2019 LNY in loglog. Incoming flows (c) and outgoing flows (d) for all provinces and days versus province population before and after the 2020 LNY in loglog.

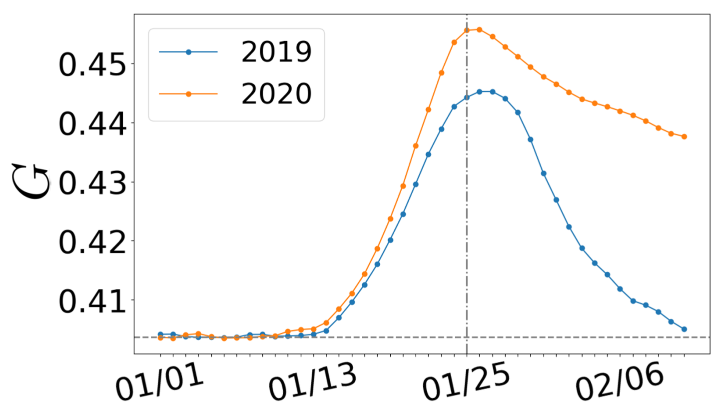

Figure S12: Variations of population Gini index at the inter-province level. Temporal variations during the Spring Festical of the population Gini index for 2019 and 2020. The dotted line represents the value 'at rest'.

Compared to 2019, the Gini index corresponding to 2020 has a higher peak and decreases with a slower speed.

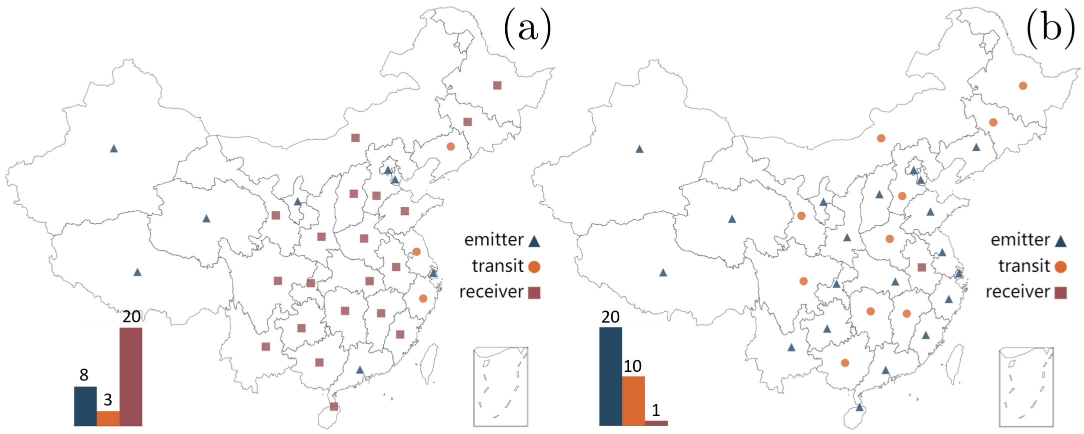

Figure S13: Spatial distribution of different categories of provinces. (a) Emitter, receiver and transit provinces according to the value of $R(i,1)$ for 2019 with the number of three categories of provinces at lower left corner. (b) Emitter, receiver and transit provinces according to the value of $R(i,1)$ for 2020 with the number of three categories of provinces at lower left corner.

We apply the criteria of three categories of provinces similar to that for city: If the value of $R(i,1)$ is larger than 1.2, we classify province i as a 'receiver' province. If the value is less than 0.8, we classify province i as an 'emitter' province. Finally, if the value is between 0.8 and 1.2, we classify province i as a 'transit' province. We show the location of three categories of cites on map of China for 2019 and 2020 in Fig. S13 and observe that receiver provinces are the majority in 2019 while emitter provinces are the majority in 2020. This results seem to make sense since most people defer the return time due to the travel ban, so that most provinces are 'emitters' in 2020.

We show in figures S14 (a) and (b) the pendular ratio for 2019 and 2020 for all provinces and highlight 5 provinces: Hubei, Beijing, Tianjin, Chongqing and Shanghai, corresponding to the origin province of COVID-19 and four province-level municipalities. We note that the curve corresponding to Hubei is in the bottom from all provinces in Fig. S14 (b), indicating that except Wuhan, the origin city of COVID-19, people also avoid going to cities of Hubei. We observe that, in 2019, the pendular ratios of all the three types of cities return to 1, meaning that the majority of individuals who went away for the holidays came back, as shown in Fig. S14 (c). The situation for 2020 is very different with a pendular ratio for all types of cities that converges to a value less than 1, indicating that the majority of people who went away for the holidays did not back yet, as shown in Fig. S14 (d).

To sum up, we show that the traffic flows between provinces are very heterogeneous, and display both large temporal and spatial fluctuations, and so on. These results for province-level are in good agreement with for city-level, indicating that our methods are applicable to both scales. Despite the detailed characters of traffic flows revealed by

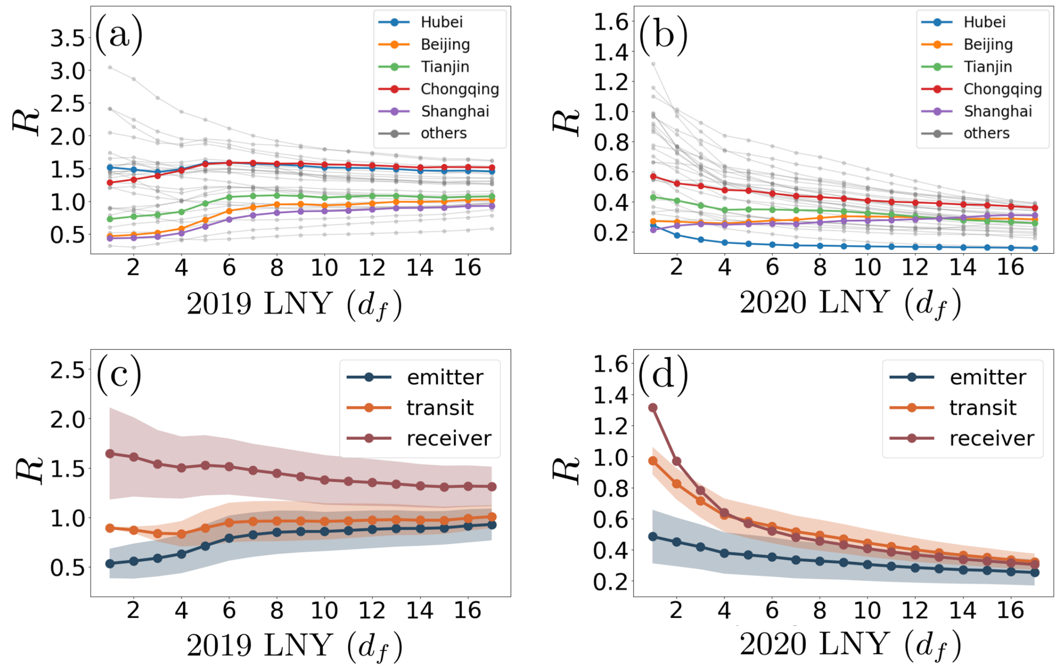

Figure S14: Pendular ratio comparison. (a) Pendular ratio for all provinces versus days d_f from LNY in 2019. We highlighted five important provinces. (b) Pendular ratio for all provinces versus days d_f from LNY in 2020 with highlight of 5 provinces. (c) Mean values of the pendular ratio over provinces according to the classification (receiver, emitter or transit provinces) versus days from LNY in 2019. The colored areas correspond to one standard deviation. (d) Mean value of the pendular ratio over provinces according to the classification versus days from LNY in 2020 with colorbar representing standard deviation.

results for city-level and province-level, a global view of a higher level is also necessary. The statistical properties of the interurban mobility help us to understand the effect of travel restrictions, their impact on and the control of epidemic spread.