Title:

Collective interactions, human mobility and viral evolution shaped the SARS-CoV-2 transmission in Mainland China.

Dong Wang[1,2], Yi Wang[1,2,3], Oswaldo Gressani[4], Jiaqi Chen[1,2], Yuxin Tao[1,2], Hao Wang[1,2], Shanshan Li[5], Dongxuan Chen[1,2], Eric H. Y. Lau[6], Yi Zhao[3], Peng Wu[1,2], Qingpeng Zhang[6,7], Benjamin J. Cowling[1,2,*], Sheikh Taslim Ali[1,2,*]

Affiliations:

[1] WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

[2] Laboratory of Data Discovery for Health, Hong Kong Science and Technology Park, Hong Kong Special Administrative Region, China

[3] School of Science, Harbin Institute of Technology, Shenzhen, China

[4] Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Data Science Institute, Hasselt University, Hasselt, Belgium

[5] Department of Neurosciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH 44195, USA.

[6] Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, China

[7] Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

*Corresponding authors

Email: bcowling@hku.hk (B.J.C) and alist15@hku.hk (S.T.A.)

Word count (abstract): 250

Abstract

Collective interaction of individuals in various settings is crucial for exposure to infections, encompassing complex viral interplay and amplifying infectious risk through phenomena such as social reinforcement, clustering and superspreading events, during the COVID-19 pandemic. However, standard epidemic models often inadequately capture such heterogeneity, overlooking the higher-order social structural. Spatiotemporal variation in transmission, an essential feature of the pandemic, remains poorly quantified at various scales, particularly in integrating high-resolution data streams and complex network approaches. We introduced a higher-order simplicial model that unifies human mobility data, genetic diversity and antigenic drift to systematically investigate the role of social reinforcement, spatiotemporal variation and genetic mutations in SARS-CoV-2 transmission. We found a median of 5.3% to 14.4% of infections across provinces were attributed to social reinforcement, while cluster heterogeneity contributed to a median of 17%–71% increase in susceptibility. Multiple viral interactions elevated transmissibility by 68%–70% across the periods of dominant variants. The reconstructed transmission networks underscored distinct spatiotemporal variation, with dynamic critical locations, varying mobility patterns, and evolving geographic cluster structures, by assessing complex networks. The influence of human mobility was found to be positive on transmission, effective distance was negatively associated with infection risks, while greater genetic diversity and antigenic drift were linked to higher susceptibility and transmissibility. Our proposed data-driven higher-order framework could help us to understand epidemics better by accounting the role of collective interactions, population mobility, and genetic mutation in transmission, which could inform the targeted interventions to mitigate SARS-CoV-2 and other respiratory pathogens.

The COVID-19 pandemic, caused by the SARS-CoV-2 virus [1,2], has posed unprecedented challenges on a global scale [3], underscoring the imperative for a comprehensive understanding of the dynamics and underlying transmission characteristics [4,5]. Several studies have highlighted the pivotal role of superspreading and clustering events in shaping the transmission dynamics of SARS-CoV-2 [6-8]. Social reinforcement [9-12], a phenomena that an individual behavior is influenced by collective interactions in groups such as households, workplaces, and restaurants, has emerged as a critical determinant of disease transmission, could occur not only after infection but also beforehand [13-15]. This effect is particularly salient for SARS-CoV-2, a respiratory pathogen transmissible via aerosols[16]. Recent evidence further indicated that when multiple strains co-circulate in a community, they may involve in interaction, mutual competition, and undergo significant mutations, potentially resulting in novel strains with more infectiousness or immune evading properties [17,18]. Incorporating the time-varying information on genetic mutation in epidemiological models is important to elucidating how viral evolution contributes to the transmission dynamics and outbreak trajectories [19-22,23]. Standard epidemic models typically assume a homogeneous, linear, pairwise relationship between the number of infectious contacts and the risk of infection for susceptible individuals, but often fail to capture the complex heterogeneity introduced by social reinforcement within collective interactions [9,10,13,24,25]. Accordingly, a higher-order modeling framework that integrates genetic information is warranted to elucidate the influence of social reinforcement and genetic dynamics embedded within collective interactions (human behaviors) on infectious disease transmission [23,26-29].

Moreover, pronounced spatiotemporal variation in transmission dynamics has been a defining features of the COVID-19 pandemic [2,30], particularly in large and diverse countries such as China, where social structures and mobility patterns differ substantially across provinces [31-33]. Transmission patterns have demonstrated significant temporal variability, underscoring the time-varying shift in community transmission [34]. Systematic analysis of spatiotemporal transmission variation facilitates the identification of high-risk regions at sub-national scale during specific periods and enables proactive implementation of targeted mitigation and control strategies [32,34,35]. However, attempts to characterize long-term, nationwide, spatiotemporal transmission dynamics, leveraging high-resolution data and network-based epidemiological modeling analysis remain limited, underscoring potential scopes of further investigation [32,34,36].

Along with spatiotemporal variation in transmission risk, heterogeneity in human social reinforcement is fundamental for understanding disease dynamics. These key factors account for the intricate interplay of social interactions, viral co-circulation and population mobility were barely assessed for their impact in shaping transmission dynamics [2,13,14,24,26,32,35]. To quantify the effects of social reinforcement on SARS-CoV-2 transmission and its spatiotemporal variation on outbreak patterns, we proposed a spatial higher-order mathematical modeling framework combined with complex network analysis, leveraging high-resolution surveillance data on epidemiology, human mobility and viral evolution across mainland China. Integrating social reinforcement within epidemiological models could provide critical insights into the challenges of controlling SARS-CoV-2 within collective settings. Furthermore,

spatiotemporal analysis of transmission networks could strengthen our understanding of the mechanisms driving pandemic spread, which could inform for effective public health interventions plan. Our general higher-order modeling framework underscored the necessity of incorporating social reinforcement, genetic signals in the co-circulating dynamics, and human mobility into epidemiological models, offering a practical perspective for the prevention and mitigation of the burden of SARS-CoV-2 and other respiratory pathogens under heterogeneous population settings.

Higher-order modelling with social reinforcement

We integrated the collective interaction with its impact as social reinforcement into our modeling framework to quantify its contribution to force of infection (FoI) for less or non-homogeneous population [9,10]. To illustrate this, we constructed the higher-order model by incorporating social reinforcement against the standard pairwise model as illustrated in Fig. 1a-b. In the standard model, the FoI was expressed as $\lambda(t)=\beta(t)I(t)$, where $\beta(t)$ denoted the time-varying transmission rates and $I(t)$ is the infection cases at time t. By incorporating nonlinear dynamics arising from collective interactions, the higher-order formulation extends the FoI to $\lambda(t)=\beta(t)I(t)+v(t)\beta(t)I(t)^2$, where $v(t)=v_{j0}(t)e^{\alpha_1 * div(t)}$ quantifies the transmissibility modulated by social reinforcement and genetic diversity. Here, $v_{j0}(t)$ represents the baseline impact of social reinforcement for dominant variant of concern j, $div(t)$ denotes genetic diversity at time t and α_1 is coefficient reflecting its impact. This modeling framework

transcends the limitations of traditional pairwise interaction paradigms by explicitly encoding group-level, collective transmission via social reinforcement [9,10]. Within this higher-order modeling paradigm, social structures were assimilated as simplicial complexes, with environments such as households, workplaces, restaurants, and social gatherings characterized as higher-dimensional simplices [24]. This formulation enabled an explicit characterization of collective interactions along with pairwise interactions, thereby overcoming the intrinsic limitations of traditional standard pairwise contact models.

We formulated the impact of social reinforcement within a simple spatial SEIRS modeling framework for COVID-19 transmission [37]. Our COVID-19 transmission model spitted the host population into $4\times n$ compartments according to their COVID-19 infection status and their located provinces: with S_i, E_i, I_i, and R_i representing susceptible, exposed, infectious, and recovered portions of individuals in province i, where $i\in\{1,\ldots,n\}$ and $n=31$ was the total number of provinces in mainland China. Compartments S_i was further subdivided into two subpopulations by considering individual status for social reinforcement in collective interactions. Specifically, S_{i1} and S_{i2} represented susceptible individuals without and with social reinforcement in interactions, where a proportion p_i was used to denote the subpopulation that has been impacted by social reinforcement for province i. To assess the impact of heterogeneity in collective interactions on susceptibility, we hypothesized that groups of susceptible individuals engaging collectively in effective contacts with multiple infectors elevated the population risk of infection. This was incorporated by defining the

relative FoI for sub-compartments as S_{i2}, by $\phi_i\lambda_{i2}(t)$, where ϕ_i denoted the hazard ratio which is modulated by the cluster heterogeneity for province i [38]. To quantify the impacts of social reinforcements on transmissibility, we posited that two infectors under social collective interactions may exhibit augmented transmissibility, and were, therefore $v(t)$ times more likely to transmit viruses. In addition to capturing collective interactions, our framework incorporates human mobility, viral evolution, and the strength of non-pharmaceutical interventions (NPIs). The model is schematically illustrated in Fig. 1c. Detailed mathematical formulations are provided in the Methods section, with parameter descriptions summarized in Table S1. We used Bayesian inference with the Hamiltonian Monte Carlo (HMC) method [39] to calibrate in predicting weekly COVID-19 incidence data in mainland China.

Epidemiological impact of social reinforcement

To quantify the epidemiological impact of social reinforcement on SARS-CoV-2 transmission, we parameterized the respective impacts (clustering, susceptibility and transmissibility impacts) in the model for predicting the reported incidence dataset (Figs. S1 and S2 in supplementary), thereby capturing the transmission dynamics reflective of spatial characteristics across provinces. Figs. 2a–c presented the proportion of infections attributable to social reinforcement (p_i), together with the resultant effects on susceptibility (ϕ_i) across 31 provinces, and the baseline transmissibility impacts (v_{j0}) for three periods, i.e., from 1 January 2020 to 6 March 2021 for ancestral strains; from 7 March 2021 to 31 December 2021 for Alpha/Delta variants;

and from 1 January 2022 to 28 February 2023 for the Omicron variant (Extended Data Fig 1a), respectively. This mechanistic framework is finally able to provide unequivocal insight into the role of social reinforcement in shaping transmission dynamics. Estimates for p_i revealed that 5.3% to 14.4% of individuals in susceptible population were affected by social reinforcement under collective settings. Furthermore, estimates for ϕ_i indicated that social heterogeneity was associated with a multiplicative effect in susceptibility risks ranging from 1.17 (95% CrI: 1.02, 3.66) to 1.71 (1.03, 4.40) times across provinces. Estimates for the variant-specific v_{j0} showed that transmissibility was augmented by 1.68 (1.03, 2.11) to 1.70 (1.34, 2.24) times comparing with pairwise interaction across the periods. The posterior distributions for the affected proportions and the increased susceptibility across provinces were presented in Extended Data Fig. 2.

Moreover, the estimate of coefficients α_1 indicated that genetic diversity (Extended Data Fig. 1b) was found to increase the transmissibility under collective interactions by 0.89 (95% CrI: 0.03, 7.76) units per unit change. Our estimate of α_2 suggested the antigenic drift (Extended Data Fig. 1c) would increase susceptibility by 0.07 (0.05, 0.10) units per unit change (Fig. 2d). In addition, the estimates of π_1 and π_1 indicated that per unit increase in NPI (Figs S3-S4) could reduce transmissibility by 2.78% (2.70%, 2.93%) units and increase ascertainment rate by 10.05% (9.58%, 10.52%) units respectively (Fig. 2e).

For model validation, we compared model performance between the higher-order and standard traditional models using identical datasets by assessing improvements in model likelihoods ($\Delta\ loglike$) (Fig. 2f). We observed a significant positive correlation (P-value = 0.007) between the magnitude of model improvement and the number of infections contributed from higher-order infection. Therefore, on average, higher-order model yielded better explanatory power in model-prediction by fitting observed case data (Fig. 2g) for recapitulating the timing and magnitude of distinct epidemic waves at the provincial level.

Spatiotemporal infection risk transmission

We reconstructed time-varying spatial transmission networks, determined by human mobility and population-level infection risk (details are in Methods section), to quantify the spatiotemporal variation of SARS-CoV-2 transmission, delineated by the distinctive characteristics of infection flows (IFs). As a proxy, we integrated the number of exposed infectious individuals $E_i(t)$ and unreported infectious population $(1\text{-}\rho_i(t))\ I_i(t)$, where $\rho_i(t)$ was the reporting rate for province i, mapped onto evolving mobility networks over time t. Fig. 3 illustrated the in- and out-bound IFs-based network *eigenvector centrality* during the pandemic at the spatiotemporal scale. In initial stages, Hubei province emerged as the primary *epicenter* of the outbreak for ancestral strain with an estimated total exported IFs of 1343015 (1266504, 1142249), by 6 March 2021 (Fig. 3a). As pandemic progressed, the principal locus of exported IFs shifted to Shanghai during the Alpha/Delta period (March–December 2021),

accounting for a cumulative exported IFs of 72756 (61267, 38193). During the subsequent Omicron wave (until February 2023), transmission hubs became increasingly diffuse, with Guangdong emerging as the dominant source, contributing an estimated 5109959 (4924523, 5306426) exported IFs during this large-scale outbreak.

Similarly, the earliest principal sink regions were those that absorbed the majority of imported infections: including Henan, Hunan, and Guangdong during initial period with estimated imported IFs of 227464 (213593, 241663), 219950 (208834, 231827), and 175293 (169758, 181992) respectively (Fig. 3b). Subsequently, the importation hubs moved to Jiangsu province during the Alpha/Delta period with an estimated imported IFs of 33932 (28901, 40070). With the onset of the Omicron wave, the scope of importation expanded, encompassing a wider array of provinces; notably, Hunan (1272795 (1226265, 1317602)) and Guangxi (1256078 (1215403, 1299642)) emerged as leading provinces in importing IFs. The spatiotemporal variation of IFs distributions across affected regions at bimonthly intervals were presented in Extended Data Figs. 3–4. The dynamic configurations of source and sink regions, derived from alternative centrality measures such as *degree*, *hub*, and *PageRank centralities*, exhibited comparable characteristics across three periods (Supplementary Figs. S5–S7). In addition, comprehensive visualizations of bimonthly spatial and temporal fluctuations in imported IFs were provided in Figs. S8–S13.

To identify key transmission pathways, we evaluated *link weight centrality* and *betweenness centrality* over time, enabling us to track the temporal evolution of network connectivity across three pandemic periods (Fig. 4) as well as at bimonthly intervals (Extended Data Figs. 5-6). During the ancestral strain period, the principal transmission pathways, quantified by *link weight* found to be managed from Hubei, extending predominantly to Henan (exported IFs: 209392 (195918, 222980)) and Hunan (exported IFs: 209528 (198511, 220923)) (Fig. 4a). These dominant pathways underwent a marked shift during the Alpha/Delta period, with the primary route connecting Shanghai and Jiangsu and a corresponding exported IFs 30288 (25510, 36182), substantially lower than that observed in the ancestral period. During Omicron period, transmission networks grew considerably more complex and spatially diffuse, giving rise to multiple provincially interlinked hubs as key locations of viral dissemination. Notably, the IFs from Guangdong to Fujian (exported IFs: 562761 (540650, 586860)), Hunan (1061837 (1015522, 1111964)), and Guangxi (1179367 (1137934, 1224628)) constituted major pathway of infections, significantly higher of those risks estimated during earlier periods, attributing the impact of the lifting of the "zero-COVID" policy in December 2022. Similarly, using network *betweenness centrality*, defined by the frequency with which a link appeared along shortest paths between all pairs of locations, we identified the Hubei–Hebei connection as particularly indispensable during the ancestral period (Fig. 4b). In the Alpha/Delta period, the central bridge pathway shifted to connect Shanghai and Beijing. The Omicron surge was characterized by further complexity and comparatively wide dispersal, with the most critical pathways emanating from Guangdong and Beijing toward Hainan.

Time-varying geographic cluster distributions and the relationship between infection outcomes and human mobility

Spatiotemporal analysis of inter-provincial transmission networks demonstrated that the geographic cluster distributions varied across three pandemic periods (Figs. 5a, 5e, 5i). Variation in geographic cluster distributions contributed to differences in infection risk and associated transmission across provinces. Notably, provinces within the same cluster were often geographically adjacent, highlighting the role of land-based transportation in shaping transmission dynamics. Although geographic cluster structure changed over time, the main clusters of infection risk flow consistently corresponded to the populous regions: northern, southern, and eastern provinces of China.

During the ancestral strain period, human mobility patterns associated with IFs centered on Hubei, which functioned as the primary source hub, with node and link sizes proportional to mobility volume (Fig. 5a). In the Alpha/Delta period, Shanghai emerged as a significant hub, though its influence and overall mobility volume were reduced (Fig. 5e). By the Omicron period, hub shifted to multiple provinces, notably Guangdong and Beijing, both exhibiting markedly increased IFs volumes (Fig. 5i). At bimonthly temporal resolution, an intriguing pattern emerged whereby, during the winter seasons (January–February) except 2021, Hainan Province in southeastern China clustered with provinces in Northeast China, likely reflecting

256 seasonal migration as populations from the Northeast moved to Hainan to escape winter
257 conditions (Extended Data Fig. 7).
258
259 Furthermore, we quantified the association between IFs outcomes (magnitudinal and temporal)
260 and human mobility under the predicted transmission networks, including the cumulative
261 number of infection cases (final size) and the timing required to reach specified case thresholds
262 across the periods (Fig. 5). During the ancestral period, Hubei acted as the main transmission
263 hub (Fig. 3a). To capture the mobility patterns related to this hub, we assessed how the number
264 of travelers from Hubei was linked to reported COVID-19 cases in other provinces. The
265 predicted provincial cumulative case counts strongly correlated (r = 0.7, P-value < 0.005) with
266 the volume of inbound travelers from Hubei (Fig. 5b). We evaluated *effective distance* [40], which
267 quantifies inter-regional connectivity based on travel flows, to better capture how transmission
268 relates to population mobility. The estimated final size for the period of ancestral strains found
269 to be negatively associated (r = -0.65, P-value < 0.005) with effective distance from Hubei
270 Province (Fig. 5c), while a positive association (r = 0.7, P-value < 0.005) was found between
271 the time to reach first 100 predicted cases following the initial outbreak in Hubei (1st January,
272 2020) and their effective distance (Fig. 5d). Comparable results found on the relationships for
273 observed final size on mobility and effective distance for ancestral period in China (Extended
274 Data Fig. 8). However, during Alpha/Delta and Omicron waves, these associations found to be
275 weaker (but same directions) and insignificant (Figs. 5f-h and 5j-l), suggesting the role of other
276 factors confounded as pandemic progressed.

Discussion

The higher-order modeling framework could assess and quantify the potential role of collective interactions within clustered social settings in shaping epidemic trajectories, assimilating social reinforcement under a simplicial complex [13,24,27]. In this framework, several settings, such as workplaces, restaurants, and households are included as higher-dimensional units (simplices) to account for the impact of transmission heterogeneity [9,24]. This formulation allowed for the explicit representation of collective interactions, rather than only pairwise interactions, and overcame the limitation of the common linear assumption of infectious contacts on infection risks. By capturing the higher-order topology of social contacts, typically overlooked in standard models, our framework enables rigorous identification of potential social determinants of epidemic transmission that are inaccessible to standard approaches (Fig. 1a-b) [9,10,24]. When coupled with empirical datasets on genetic diversity, population mobility, our model substantially enhances the granularity at which spatiotemporal transmission dynamics can be resolved at various scales (provincial, national, and global) (Fig. 1c) [17,23,26-29]. Crucially, the proposed higher-order modeling framework outperformed standard models in predicting pandemic trajectories [24], offering demonstrably superior explanatory and predictive power (Fig. 2f).

296 Social reinforcement revealed pronounced heterogeneity in both the proportion of infections
297 attributable to such reinforcement and the resultant increases in susceptibility across provinces
298 (Fig. 2a-b). Notably, while the proportion of infections influenced by social reinforcement
299 generally remains modest, ranging between 5.3% and 14.4% across mainland China, its impact
300 can be substantial, serving as a catalyst for critical phenomena such as superspreading events
301 [6-8,13,24,35]. Provinces exhibiting pronounced clustering manifested increased susceptibility up to
302 17% to 71% higher risk than those observed under pairwise mixing regimes. Within these
303 clustered settings [41], susceptibility demonstrably escalates with increasing population size, a
304 trend robustly captured by our higher-order modelling framework, which partitions the
305 susceptible population into discrete subgroups, thereby extending beyond the classical
306 assumption of pairwise interaction [9,10,13,24,29]. These findings contribute to a deeper
307 understanding of the underlying mechanisms driving clustering and superspreading
308 phenomena [9,29,42].
309
310 Furthermore, the higher-order contacts found to consistently escalate the transmissibility of
311 COVID-19 and consistent across the periods (Fig. 2c). One possible reason is that collective
312 interactions, which involve exposure to a greater number of infectious individuals, may result
313 in a higher cumulative dose of viral particles, surpassing the capacity of the human immune
314 system [9,13,29,43], thereby concurrently increasing the transmissibility. Crucially, transmissibility
315 driven by social reinforcement increases by more than half (68%–70%) compared to
316 transmission seen in pairwise contacts. Although it is rare for several infected people to be in

317 the same group when the overall infection rate is low, such events can spark major outbreaks
318 and superspreading [9,29,42]. These findings support the significance of non-pharmaceutical
319 interventions, such as social distancing and restrictions on large gatherings, in controlling the
320 spread of respiratory infectious diseases [30,32].
321
322 A nationally representative, population-based examination of the spatiotemporal variation and
323 unique characteristics of SARS-CoV-2 transmission would significantly deepen our
324 understanding of epidemic dynamics on both temporal and geographic scales [2]. By
325 reconstructing transmission networks derived from the spatial distribution of infectious
326 population and population mobility at temporal scale, our analysis reveals spatiotemporal
327 epidemiological characteristics, including critical locations (Fig. 3), key patterns of infection
328 flows (Fig. 4), and underlying geographic cluster structures (Fig. 5) across the provinces in
329 mainland China. We could identify several underline significant factors for the spatiotemporal
330 transmission pattern of COVID-19, which vary markedly across the provinces during three
331 predefined pandemic periods. In particular, the population mobility from epidemiological hubs
332 found to increase the infection rates in downstream (sink) regions by elevating the risk of
333 secondary infection (Fig. 5). The provinces with higher traveler volumes and shorter effective
334 distances from hub locations experienced significantly greater infection risks and earlier
335 epidemic onset. However, the strength of these associations declined during the Alpha/Delta
336 (Fig. 5f–h) and Omicron (Fig. 5j–l) periods, reflecting a diminished influence of centralized
337 transmission hubs [40,41,44]. This temporal shift indicates a gradual decentralization of

transmission pathways as the epidemic progressed, marked by the emergence of multiple, simultaneous outbreaks. Unveiling principal locations, pathways, geographic clusters, and the relationship between infection and population mobility would help us achieve effective and targeted control of epidemic transmission in large countries and at the global level [2,4,45].

Despite within-host viral interactions, competition, and mutation, the presence of multiple infectors in a group, each carrying viruses with distinct genetic backgrounds, amplifies genetic diversity and facilitates inter-viral interactions [23]. These interactions promote viral competition and mutation, collectively enhancing the potential for transmission dynamics by reshaping the associated susceptibility and hence transmissibility [17,23]. Our findings demonstrate that greater genetic diversity is associated with increased transmissibility under collective interactions (Fig. 2d). Furthermore, antigenic drift alters population immunity, a property we term immune efficacy with respect to susceptibility [46,47], and is associated with increased susceptibility as ongoing antigenic changes diminish the effectiveness of host immune responses (Fig. 2e).

The impact of NPIs stringency is evident in the attenuation of transmission rates, as demonstrated in Extended Data Figs. 9-10. We found that the increase in NPI stringency is associated with a reduction in the transmission rate, with the corresponding effects on the effective reproductive number, comparable to previous reports [32,33], illustrated in Figs. S14–S15. Besides, NPI stringency is positively correlated with improvements in case ascertainment,

as shown in Figs. S16–S17. Collectively, these results highlight the dual function of NPIs in both mitigating viral transmission and enhancing the rates of infection detection [48].

Our higher-order modelling framework that integrates viral genomic data, non-pharmaceutical interventions, and human mobility to quantify the effects of collective interactions and spatiotemporal heterogeneity within subpopulations, yields demonstrably improved predictive performance over conventional epidemic models. The flexibility of this paradigm highlights its broad applicability across diverse geographic settings and to other respiratory pathogens. Our study has limitations. First, we identify population mobility as a principal driver of the observed spatiotemporal variability in epidemiological dynamics. Although the use of mobility data from Baidu Qianxi substantially enhance inference [33], fully capturing the complexity of real-time, nationwide inter-regional mobility remains challenging. Second, while vaccination coverage is high, but its impact on disease infection rates appears to be mild to modest [49] and vaccination schemes were centralized with minimal variation in timing and coverage across the provinces in mainland China [50]. To focus on our primary objectives, we therefore opted out the vaccination impact in the model directly [2,18], however availability of such data could be tested for the model performance. Third, the limited availability of fine-grained cluster data precludes robust characterization of transmission dynamics within specific settings, such as workplaces or restaurants, across China at temporal scale. Therefore, agent-based analyses could be limited in such case; however, the individual-level investigations could be a better alternative[24,51]. Fourth, while our simplicial models with up to order 2 interactions offer an

important insight into the hidden mechanisms of infectious disease transmission. However, considering higher-order interactions (over order 3 or beyond), could be a pivotal to capture transmission dynamics better. Future research could account the contribution of such interactions of varying orders for transmission heterogeneity [9,10,24]. Finally, we considered the nationwide mobility datasets in mainland China [52], nevertheless the inclusion of international mobility data could provide better importation risk prediction over time. Although, international travel restrictions might reduce such risk of infection during the study period.

In conclusion, we could establish a data-driven, higher-order modelling framework that assimilates epidemiological, geographic, genomic data steams with complex network approaches to delineate the foundational role of social reinforcement in shaping collective transmission dynamics and the spatiotemporal evolution of transmission networks across regions and epidemic phases. Understanding the impacts of social reinforcement at various levels of transmission process and other potential factors would provide a perspective tool for mitigating the burden of COVID-19 and other respiratory pathogens and their emerging variants under heterogeneous population settings.

References

1 Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* **382**, 727–733 (2020). https://doi.org/10.1056/NEJMoa2001017

2 Pei, S., Yamana, T. K., Kandula, S., Galanti, M. & Shaman, J. Burden and characteristics of COVID-19 in the United States during 2020. *Nature* **601**, E6 (2022). https://doi.org/10.1038/s41586-021-04172-0

3 World Health, O. The impact of COVID-19 on health and care workers: a closer look at deaths. (World Health Organization, Geneva, 2021 https://iris.who.int/handle/10665/345300).

4 Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **395**, 931–934 (2020). https://doi.org/10.1016/S0140-6736(20)30567-5

5 Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* **20**, 553–558 (2020). https://doi.org/10.1016/S1473-3099(20)30144-4

6 Althouse, B. M. *et al.* Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *Plos Biol* **18**, e3000897 (2020). https://doi.org/10.1371/journal.pbio.3000897

7 Liu, Y., Eggo, R. M. & Kucharski, A. J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **395**, e47 (2020). https://doi.org/10.1016/S0140-6736(20)30462-1

8 Adam, D. C. *et al.* Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* **26**, 1714–1719 (2020). https://doi.org/10.1038/s41591-020-1092-0

9 Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat Commun* **10**, 2485 (2019). https://doi.org/10.1038/s41467-019-10431-6

10 Wang, D., Zhao, Y., Luo, J. & Leng, H. Simplicial SIRS epidemic models with nonlinear incidence rates. *Chaos* **31**, 053112 (2021). https://doi.org/10.1063/5.0040518

11 Hebert-Dufresne, L., Scarpino, S. V. & Young, J. G. Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement. *Nat Phys* **16**, 426–431 (2020). https://doi.org/10.1038/s41567-020-0791-2

12 Ye, Y. *et al.* Effect of heterogeneous risk perception on information diffusion, behavior change, and disease transmission. *Physical Review E* **102**, 042314 (2020). https://doi.org/10.1103/PhysRevE.102.042314

13 St-Onge, G., Sun, H., Allard, A., Hebert-Dufresne, L. & Bianconi, G. Universal nonlinear infection kernel from heterogeneous exposure on hgher-order networks. *Phys Rev Lett* **127**, 158301 (2021). https://doi.org/10.1103/PhysRevLett.127.158301

14 Bavel, J. J. V. *et al.* Using social and behavioural science to support COVID-19 pandemic response. *Nat Hum Behav* **4**, 460–471 (2020). https://doi.org/10.1038/s41562-020-0884-z

15 Leclerc, Q. J. *et al.* What settings have been linked to SARS-CoV-2 transmission clusters? *Wellcome Open Res* **5**, 83 (2020). https://doi.org/10.12688/wellcomeopenres.15889.2

16 Morawska, L. & Milton, D. K. It is time to address airborne transmission of coronavirus disease 2019 (COVID-19). *Clin Infect Dis* **71**, 2311–2313 (2020). https://doi.org/10.1093/cid/ciaa939

17 Jackson, B. *et al.* Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **184**, 5179–+ (2021). https://doi.org/10.1016/j.cell.2021.08.014

18 Carabelli, A. M. *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature Reviews Microbiology* **21**, 162–177 (2023). https://doi.org/10.1038/s41579-022-00841-7

19 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020). https://doi.org/10.1038/s41586-020-2180-5

20 Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020). https://doi.org/10.1038/s41586-020-2179-y

21 Stockdale, J. E., Liu, P. & Colijn, C. The potential of genomics for infectious disease forecasting. *Nature Microbiology* **7**, 1736–1743 (2022). https://doi.org/10.1038/s41564-022-01233-6

22 Garg, M. *et al.* Disease prediction with multi-omics and biomarkers empowers case–control genetic discoveries in the UK Biobank. *Nature Genetics* **56**, 1821–1831 (2024). https://doi.org/10.1038/s41588-024-01898-1

23 Tran-Kiem, C. *et al.* Fine-scale patterns of SARS-CoV-2 spread from identical pathogen sequences. *Nature* **640** (2025). https://doi.org/10.1038/s41586-025-08637-4

24 Chen, Y. Z., Gel, Y. R., Marathe, M. & Poor, H. V. A simplicial epidemic model for COVID-19 spread analysis. *P Natl Acad Sci USA* **121** (2024). https://doi.org/10.1073/pnas.2313171120

25 Vasiliauskaite, V., Antulov-Fantulin, N. & Helbing, D. On some fundamental challenges in monitoring epidemics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **380**, 20210117 (2022). https://doi.org/doi:10.1098/rsta.2021.0117

26 Bauch, C. T. & Galvani, A. P. Social factors in epidemiology. *Science* **342**, 47–49 (2013). https://doi.org/10.1126/science.1244492

27 Battiston, F. *et al.* The physics of higher-order interactions in complex systems. *Nat Phys* **17**, 1093–1098 (2021). https://doi.org/10.1038/s41567-021-01371-4

28 Boccaletti, S. *et al.* The structure and dynamics of networks with higher order interactions. *Phys Rep* **1018**, 1–64 (2023). https://doi.org/10.1016/j.physrep.2023.04.002

29 Ferraz de Arruda, G., Aleta, A. & Moreno, Y. Contagion dynamics on higher-order networks. *Nature Reviews Physics* **6**, 468–482 (2024). https://doi.org/10.1038/s42254-024-00733-0

30 Gatto, M. *et al.* Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc Natl Acad Sci U S A* **117**, 10484–10491 (2020). https://doi.org/10.1073/pnas.2004978117

31 Jia, Q., Li, J., Lin, H., Tian, F. & Zhu, G. The spatiotemporal transmission dynamics of COVID-19 among multiple regions: a modeling study in Chinese provinces. *Nonlinear Dyn* **107**, 1313–1327 (2022). https://doi.org/10.1007/s11071-021-07001-1

32 Kraemer, M. U. G. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020). https://doi.org/10.1126/science.abb4218

33 Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020). https://doi.org/10.1126/science.abb6105

34 Sun, K. *et al.* Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, 254–+ (2021). https://doi.org/10.1126/science.abe2424

35 Rader, B. *et al.* Crowding and the shape of COVID-19 epidemics. *Nat Med* **26**, 1829–1834 (2020). https://doi.org/10.1038/s41591-020-1104-0

36 Xu, B. *et al.* Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* **7**, 106 (2020). https://doi.org/10.1038/s41597-020-0448-0

37 Pei, S., Yamana, T. K., Kandula, S., Galanti, M. & Shaman, J. Burden and characteristics of COVID-19 in the United States during 2020. *Nature* **598**, 338–341 (2021). https://doi.org/10.1038/s41586-021-03914-4

38 Shrestha, S. *et al.* Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Sci Transl Med* **5**, 191ra184 (2013). https://doi.org/10.1126/scitranslmed.3005982

39 Hoffman, M. D. & Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* **15**, 1593–1623 (2014).

40 Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013). https://doi.org/10.1126/science.1245200

41 Paredes, M. I. *et al.* Underdetected dispersal and extensive local transmission drove the 2022 mpox epidemic. *Cell* **187**, 1374–1386.e1313 (2024). https://doi.org/10.1016/j.cell.2024.02.003

42 Chen, D. *et al.* Investigating setting-specific superspreading potential and generation intervals of COVID-19 in Hong Kong. *Nat Commun* **16**, 5816 (2025). https://doi.org/10.1038/s41467-025-60591-x

43 Gama, J. A., Abby, S. S., Vieira-Silva, S., Dionisio, F. & Rocha, E. P. C. Immune subversion and quorum-sensing shape the variation in infectious dose among bacterial pathogens. *Plos Pathog* **8** (2012). https://doi.org/10.1371/journal.ppat.1002503

44 Tegally, H. *et al.* Dispersal patterns and influence of air travel during the global expansion of SARS-CoV-2 variants of concern. *Cell* **186**, 3277–3290.e3216 (2023). https://doi.org/10.1016/j.cell.2023.06.001

45 Lü, L. *et al.* Vital nodes identification in complex networks. *Physics Reports* **650**, 1–63 (2016). https://doi.org/10.1016/j.physrep.2016.06.007

46 Lavine, J. S., Bjornstad, O. N. & Antia, R. Immunological characteristics govern the transition of COVID-19 to endemicity. *Science* **371**, 741–745 (2021). https://doi.org/doi:10.1126/science.abe6522

47 Weisberg, S. P. *et al.* Distinct antibody responses to SARS-CoV-2 in children and adults across the COVID-19 clinical spectrum. *Nature Immunology* **22**, 25–31 (2021). https://doi.org/10.1038/s41590-020-00826-9

48 Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413 (2020). https://doi.org/10.1038/s41586-020-2293-x

49 Leung, K., Lau, E. H. Y., Wong, C. K. H., Leung, G. M. & Wu, J. T. Estimating the transmission dynamics of SARS-CoV-2 Omicron BF.7 in Beijing after adjustment of the zero-COVID policy in November–December 2022. *Nat Med* **29**, 579–582 (2023). https://doi.org/10.1038/s41591-023-02212-y

50 Zheng, W., Yan, X., Zhao, Z., Yang, J. & Yu, H. COVID-19 vaccination program in the mainland of China: a subnational descriptive analysis on target population size and current progress. *Infectious Diseases of Poverty* **10**, 124 (2021). https://doi.org/10.1186/s40249-021-00909-1

51 Du, Z. *et al.* Impact of RSVpreF vaccination on reducing the burden of respiratory syncytial virus in infants and older adults. *Nat Med* **31**, 647–652 (2025). https://doi.org/10.1038/s41591-024-03431-7

52 Zhu, K., Cheng, Z. & Wang, J. Measuring Chinese mobility behaviour during COVID-19 using geotagged social media data. *Humanities and Social Sciences Communications* **11**, 540 (2024). https://doi.org/10.1057/s41599-024-03050-0

Methods

Higher-order modelling with social reinforcements for collective interaction

Inspired by a simplicial higher-order model [9,10,24], we accounted for both standard homogeneous interactions and heterogeneous clustered group interactions to identify the impact of social reinforcements. We constructed the following deterministic skeleton of the model, incorporating collective interactions for social reinforcement and demographic profiles (Fig. 1). The spatiotemporal dynamics of the transmission process are subsequently described within this deterministic modeling framework.

$$\frac{\mathrm{d}S_i(t)}{\mathrm{d}t}=-(1-p_i)\mathrm{AD(t)}S_i(t)\lambda_{i1}(t)-p_i\phi_i\mathrm{AD(t)}S_i(t)\lambda_{i2}(t)+qR_i(t)$$

$$\frac{\mathrm{d}E_i(t)}{\mathrm{d}t}=\left(1-p_i\right)\mathrm{AD(t)}S_i\lambda_{i1}\ (t)+p_i\phi_i\mathrm{AD(t)}S_i(t)\lambda_{i2}\ (t)-\sigma E_i(t)$$

$$\frac{dI_i(t)}{dt}=\sigma E_i(t)-\gamma I_i(t)$$

$$\frac{\mathrm{d}R_i(t)}{\mathrm{d}t}=\gamma I_i(t)-qR_i(t)$$

$$N_i=S_i(t)+V_i(t)+E_i(t)+I_i(t)+R_i(t)$$

Where $S_i(t)$,$E_i(t)$,$I_i(t)$ and $R_i(t)$ denoted the proportion of susceptible individuals, exposed individuals, infectious individuals, and recovered individuals at time t for province i, respectively. $\mathrm{AD(t)}=e^{\alpha_2*\mathrm{ad(t)}}$ denoted the impact of antigenic drift on susceptibility, where $\mathrm{ad(t)}$ was the time-varying antigenic drift index and α_2 denoted its impact. An infected individual would become infectious after $1/\sigma$ days on average since infection (i.e., mean latent period), which was assumed to be 5 days [2,24,53]. The individual would be infectious for $1/\gamma$ days on average (i.e., mean infectious period), which was assumed to be 10 days [2,24,53]. The duration

564 of sterilizing immunity $1/q$ was assumed to be 610 days [46]. p_i represents the proportion of
565 infections influenced by social reinforcement under the higher-order collective interactions for
566 province i and ϕ_i denotes the impact of social reinforcements on susceptibility. Note that
567 when p_i=0, the high-order model will be degenerated into the standard homogeneous model.
568 The force of infection in standard and higher-order transmission settings with IFs can be
569 defined as follows:

570 $\lambda_{i1}(t)=\beta_i(t)(I_i(t)+\Delta_i(t))$

571 $\lambda_{i2}(t)=\beta_i(t)\left(I_i(t)+\Delta_i(t)+v(t)(I_i(t)+\Delta_i(t))^2\right)$

572 where $\beta_i(t)$ was the time-varying transmission rate for province i. $v(t)=v_{j0}(t)e^{\alpha_1*\mathrm{div}(t)}$
573 denoted the transmissibility impact under social reinforcements where more than one infector
574 are in the same cluster, and time t is segmented by the timing of the leading variants of
575 concern. Here, $v_{j0}(t)$ represents the baseline impact of social reinforcement for dominant
576 variant of concern j and $\mathrm{div}(t)$ denoted the genetic diversity data and α_1 was its impacts.
577 The impact of human mobility on provincial infectiousness is quantified by combining
578 mobility-driven transmission dynamics with spatiotemporal network properties, computed as
579 follows, $\Delta_i=\frac{1}{N_i}\sum_{k\neq i}\left(M_{ik}\left(1-\rho_k(t)\right)I_k-M_{ki}\left(1-\rho_i(t)\right)I_i\right)$, where M_{ik} is the population
580 flow from province k to province i, and $\rho_i(t)$ denotes the ascertainment rates (i.e., reporting
581 proportion) of province i. Here, we assumed that mobility was restricted for documented
582 infections.

583

We constructed the time-varying transmission rate $\beta_i(t)$ for province i at time t as a function of cubic *B-spline* and non-pharmaceutical interventions (NPIs), as $\beta_i(t)= exp(s_{i,t})$ * $\beta_{NPI,\,i}(t)$, where $s_{i,t}$ is a degree 3 spline function with 10 equally spaced knots based on the whole study period for the province i at time t. $\beta_{NPI,\,i}(t)=e^{-\pi_1 * NPI(i,t)}$ denotes the impact of the NPI and π_1 represents the level of impact and $NPI(i,t)$ represents the NPI stringency index for province i at time t. We assumed the time-varying ascertainment rates (i.e., reporting proportion) ρ_i for province i is often driven by the associated NPIs [54,55]. Therefore, assuming a positive association between the strength of NPI and ascertainment rates, we used a sigmoid function to describe time-varying ascertainment rates as $\rho_i(t)=\rho_0 e^{-\pi_2 * NPI(i,t)}$ where π_2 denotes the impact level and ρ_0 denotes the baseline ascertainment rate without the impact of NPI (i.e., $NPI(i,t)$=0) in the settings.

Model inference framework

To relate the predicted incidence to the empirically observed COVID-19 cases, let $C_i(t)$ be the cumulative observed cases at week t for province i, defined as $\frac{dC_i(t)}{dt}=\sigma E_i(t)$, accordingly, the observed cases $C_i^o(t)$ at time t for location i was derived as $C_i^o(t)=C_i(t)-C_i(t-1)$. Further, we described the observed incidence $y_i(w)$ at week w as a negative binomial process of the model induced cases as: $y_i(w)\sim NB\left(\sum_{t\in w} N_i\, C_i^o(t)\rho_i(t), \varphi_i\right)$, where $\sum_{t\in w} N_i\, C_i^o(t)\rho_i(t)$ denotes the mean of predicted cases at week w. N_i is the total population of province i and φ_i is a parameter denoting overdispersion.

Model fitting and estimation of parameters $\theta=\{s_i, p_i, \phi_i, v(t), \rho_0, a, b, c, d, \varphi_i\}$ were conducted within a Bayesian inference with Hamiltonian Monte Carlo (HMC) [39]. In this approach, prior distributions for certain model parameters were specified conditionally, reflecting the nature of the Bayesian paradigm. Uninformative priors were incorporated throughout the inference process (Supplementary 2). We employed the *RStan* package in R, leveraging the No-U-Turn Sampler (NUTS) algorithm [39]. Four independent Markov chains were initiated at different staring positions, each comprising 4,000 iterations with the first 2,000 iterations designated as burn-in. Convergence and thorough exploration of the posterior distribution were assessed by ensuring that the *R-hat* statistic for all parameters remained below 1.05, thereby indicating satisfactory mixing and convergence [56].

Reconstruction of the time-varying transmission network

In addition to characterizing the general population, we specifically investigated the mobility network of individuals at potential risk of infection. To reconstruct time-varying, virus-carrier-based transmission mobility networks, we integrated data on province-level human mobility with model-inferred, province-specific viral activity. The mobility of population at time t between provinces is denoted by a mobility matrix $M(t)$ with entries $M_{ji}(t)$ representing the number of individuals moving from province i to province j at time t. The infection dynamics in each province are captured by a potential IF matrix, represented as $E_i(t) + (1 - \rho_i(t))I_i(t)$, $E_i(t)$ represents the inferred exposed population, assumed to be unrestricted in their

mobility, $I_i(t)$ denotes the inferred infectious population at time t, where the mobility of such reported cases was prohibited, and thus we just considered the unreported infection cases, and $\rho_i(t)$ denotes the vector of ascertainment rates for all provinces at time t. The transmission network at time t can be computed as $T(t) = M(t) \cdot \mathrm{diag}(E(t) + (1 - \rho(t))I(t))$, where diag is used to denote a diagonal matrix, $E(t) = \{E_i(t)\}_{i=1\ldots31}$, $\rho(t) = \{\rho_i(t)\}_{i=1\ldots31}$ and $I(t)=\{I_i(t)\}_{i=1\ldots31}$. This integrative framework allows for the coupling of inter-provincial mobility patterns with local IF profiles, facilitating the characterization of spatiotemporal virus transmission across provinces.

Transmission network assessment and spatiotemporal characteristics

The properties of complex transmission networks were primarily characterized by (1) the identification of potential infection hubs—critical nodes or provinces implicated in disease transmission [33,45,57]; (2) the delineation of high-risk transmission pathways, key mobility links facilitating disease dissemination; and (3) the recognition of transmission geographic clusters—groups of provinces tightly interconnected through frequent exchanges [58]. The elucidation of these structural elements within transmission networks is essential for informing the design of targeted and effective intervention strategies. To identify putative infection hubs or key nodes, we employed several centrality metrics: (a) *Degree centrality*, quantifying the number of direct connections per node, where provinces with high degree centrality are positioned to facilitate rapid disease spread owing to numerous immediate contacts; (b) *Hub centrality*, as defined by Kleinberg's algorithm and calculated by *hub score*, providing a

measure of a node's prominence as a key transmitter within the network, reflecting its overall influence on epidemic dynamics; (c) *Eigenvector centrality*, which captures not only the direct connectivity of a node but also considers the influence exerted by its neighbors—provinces with elevated eigenvector centrality are thus both extensively connected and linked to other influential provinces; and (d) *PageRank centrality*, ranking nodes based on both the quantity and quality of incoming connections, thereby highlighting those provinces of greatest strategic importance within the transmission network [45]. Considering vital locations, we assess key nodes in the transmission network from two perspectives: (1) sources, defined as principal outflow provinces from which SARS-CoV-2 disseminates outward, and (2) sinks, or primary inflow provinces to which infection is introduced.

A suite of well-established centrality metrics was employed to elucidate key links within the transmission network across three distinct periods: (a) *Weight centrality*, which quantifies the strength of interconnections in a weighted network, wherein link weights correspond to the probability or intensity of disease spread—heavier links thereby reflect a greater potential for pathogen transmission [58]. (b) *Link betweenness centrality*, which assesses the frequency with which a given link appears on the shortest paths between node pairs. Links with high betweenness centrality serve as critical "bridges" within the network, whose removal is likely to fragment connectivity and substantially impede the progression of outbreaks—such links thus represent pivotal routes for transmission [59].

667 To further delineate the structure of transmission, we implemented the walk trap algorithm to
668 detect geographic transmission clusters in time-varying, weighted, directed networks by
669 interrogating node and link interrelationships alongside their contributions to epidemic
670 trajectories [60,61]. Such clusters typically emerge when groups of infected individuals in one
671 province establish strong connections with those in other provinces, revealing localized
672 outbreaks that share common infection sources. These clusters may function as amplification
673 zones that not only sustain intra-regional spread but also potentiate the exportation of infections
674 to new geographic areas.
675

676 **The impact of human mobility on transmission dynamics**

677 To assess the influence of human mobility on the transmission dynamics of SARS-CoV-2, we
678 examined the relationship between human mobility patterns and key epidemiological
679 characteristics of SARS-CoV-2 [32,40], including the relationship between cumulative human
680 mobility from the epidemic hub and the number of reported cases, considering both
681 accumulative migration metrics and effective distance cumulative number of infection cases
682 and the timing required to reach 100 reported case thresholds [40]. Here, the effective distance
683 from province j to i is defined as [40] d_{ij}=1-log(P_{ij}), where P_{ij} is the fraction of flux leaving
684 province j that is arriving at province i, and therefore can also be written as $P_{ij}=M_{ij}/M_j$,
685 where $M_j=\sum_m M_{mj}$ is the total number of population leaving province j.
686

Ethics statement

This study was conducted in strict accordance with all relevant ethical regulations and guidelines. Ethical approval was obtained from the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB; reference number: UW 22-761). All methods were performed according to relevant guidelines and regulations. Informed consent was obtained from all participants and/or their legal guardians, as appropriate.

Data availability

We obtained province-level, daily reported COVID-19 surveillance data for mainland China spanning 1 January 2020 to 28 February 2023 from the Johns Hopkins University Coronavirus Resource Center (JHU CRC) (https://github.com/CSSEGISandData/COVID-19) [2,62,63]. Human mobility patterns over the study period were characterized using data provided by Baidu Corporation (https://qianxi.baidu.com), which represents one of the most widely utilized open-access nationwide mobility datasets in mainland China [52]. This dataset comprises both the migration scale index and the migration proportion for each province across time. Data on the stringency of non-pharmaceutical interventions (NPIs) were sourced from the Oxford COVID-19 Government Response Tracker (OxCGRT) (https://github.com/OxCGRT/covid-policy-tracker) [64], and the stringency index was rescaled to the unit interval [0, 1] for incorporation into the model. Information on SARS-CoV-2 lineages circulating in mainland China during

the study period was obtained from publicly available submissions to GISAID (https://gisaid.org/). All genome sequences and associated metadata in this dataset are published in GISAID's EpiCoV database. Detailed information for each sequence, including accession number, virus name, collection date, originating and submitting laboratories, and authors, can be found under EPI_SET_250805xc. This dataset includes 27,605 genome sequences collected between 2020-01-01 and 2023-02-28. These datasets were utilized to ascertain the temporal emergence in predominance of variants of concern genetic diversity and antigenic drift. Provincial population sizes for 2021 were obtained from the China City Statistical Yearbook (http://olap.epsnet.com.cn/). The geographic location of each provincial unit was determined by the latitude and longitude coordinates of its central point.

Code availability

All data and code for the study analysis and simulations in this paper will be publicly available from https://github.com/WangDongHKU/high_order_modelling.

Acknowledgments

The authors thank Julie Au for technical assistance. We acknowledge the research computing facilities and advisory services (HPC) offered by Information Technology Services, The University of Hong Kong and Laboratory of Data Discovery for Health ($D^2$4H).

Funding

This project was supported by the Health and Medical Research Fund (22210582 and 24230712, D. W. and S.T.A.), Hong Kong Special Administrative Region, Research Grants Council of the Hong Kong Special Administrative Region (17100225, S.T.A) and the AIR@innoHK program of the Innovation and Technology Commission of the Hong Kong SAR Government (B.J.C). The funding bodies had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish.

Author contributions: All authors meet the ICMJE criteria for authorship. Conceptualization: S.T.A., D.W. and B.J.C.; Data curation: Y.W.,Y.T., D.W. and S.T.A.; Methodology: D.W. and S.T.A.; Formal analysis: D.W. and S.T.A.; Investigation: D.W. and S.T.A.; Visualization: D.W. and S.T.A.; Software: D.W.; Validation: D.W., O.G. and S.T.A.; Resources: S.T.A. and B.J.C.; Funding acquisition: D.W., S.T.A. and B.J.C.; Supervision: S.T.A.; Validation: D.W., and S.T.A.; Writing-original draft: D.W. and S.T.A. ; Writing-review and editing: Y.W., J.C., Y.T., H. W., D.C., O.G., E.H.Y.L, Y.Z., P.W. Q.Z. and B.J.C. All authors discussed and approved the manuscript.

Competing interests: BJC consults for AstraZeneca, Fosun Pharma, GSK, Haleon, Moderna, Novavax, Pfizer, Roche and Sanofi Pasteur. The authors report no other potential conflicts of interest.

Figures

Figure 1. Epidemic modelling with and without social reinforcement, and higher-order modelling schematic representation under SEIR framework. (**a)** Schematic of the standard epidemiological model without social reinforcement, illustrating baseline infection risk. (**b)** Higher-order model incorporating social reinforcement, which modifies infection risk within cluster settings. The infection risk includes both the standard one $\beta(t)I(t)$ and another one $\beta(t)I(t)^2$ from social reinforcements, where the formula $I(t)^2$ in population, the level is equivalent to two infections in a cluster at the individual level. Susceptible and infected individuals are represented in black and red, respectively. **(c)** Schematic representation of SARS-CoV-2 transmission model under a simple SEIR framework, where individuals progress along $S_i \rightarrow E_i \rightarrow I_i \rightarrow R_i \rightarrow S_i$ at per capita rates λ, σ, and γ for province i, respectively. The progression of individuals infected not under social reinforcement is tracked via classes S_{i1} and S_{i2}, where $p(t)$ is the proportion of infected individuals under social reinforcements. $\phi_i > 1$ captures the enhanced susceptibility arising from collective interactions, as determined by cluster heterogeneity. $v(t) > 0$ represents the increased transmissibility conferred by the concurrent presence of two infectors within a cluster, in contrast to standard pairwise transmission dynamics.

Figure 2. Effects of social reinforcement and higher-order model predictions across 31 provinces during the COVID-19 epidemic. **(a)** Proportion of infection cases influenced by social reinforcement mechanisms across provinces. **(b)** Provincial variation in the impact of

769 cluster heterogeneity on susceptibility. **(c)** Effects of collective interactions on SARS-CoV-2
770 transmissibility across distinct epidemic periods characterized by the predominance of different
771 SARS-CoV-2 variants of concern. **(d)** The impact of viral genetic diversity on transmissibility
772 and the impact of viral antigenic drift on susceptibility. **(e)** The impact of non-pharmaceutical
773 interventions on transmissibility and ascertainment rates. **(f)** Improvement in predictive
774 performance of the higher-order model relative to the standard model as a function of higher-
775 order transmission, where the red line denoted the predicted mean with 95% confidence
776 intervals (CIs) in orange shade and 95% prediction intervals (PIs) in gray shade. **(g)** Model fit
777 to weekly reported case numbers (black dots) for 31 provinces in mainland China; red solid
778 lines represent mean model estimates and shaded regions indicate 95% PIs.
779
780 **Figure 3**. Identification of key transmission locations for SARS-CoV-2 in mainland China,
781 from January 2020–to February 2023. **(a)** Spatial distribution of origin locations, as determined
782 by eigenvector centrality within the outgoing transmission network, highlighting provinces
783 with the highest outflow of SARS-CoV-2 for the three major variants of concern. **(b)** Spatial
784 distribution of key destination locations, based on eigenvector centrality for incoming
785 transmission, indicating provinces with the highest inflow of SARS-CoV-2 associated with the
786 three variants of concern. **(c)** Temporal evolution of dominant source (outflow) provinces
787 identified as vital transmission origins. **(d)** Temporal evolution of major sink (inflow)
788 provinces representing crucial destinations for viral spread. Greater intensity of red in the index

corresponds to increasingly significant locations, while blue specifically denotes the most critical transmission hubs at each time point.

Figure 4. Key link characteristics of SARS-CoV-2 transmission networks across three epidemic periods in mainland China, from January 2020 to February 2023. Link importance within the transmission networks is quantified using link weight centrality **(a)** and betweenness centrality **(b)** for three distinct epidemic periods. Higher values denote greater structural and epidemiological importance of transmission pathways.

Figure 5. Time varying geographic clusters of transmission networks and associations between human mobility and SARS-CoV-2 transmission dynamics across three pandemic periods in mainland China, from January 2020 to February 2023. **(a)** Geographic clustering of provinces within the transmission network during the ancestral strain period; provinces sharing the same color belong to the same cluster, and node size reflects the magnitude of infection risk outflow. **(b–d)** Relationships during the ancestral strain period: **(b)** reported cases versus inter-provincial mobility, **(c)** reported cases versus effective distance, and **(d)** time to reach 100 cases versus effective distance in the early epidemic period. **(e–h)** and **(i–l)** present analogous analyses for the Alpha/Delta and Omicron variant periods, respectively, illustrating changes in cluster structure and associations between mobility patterns and transmission characteristics for each period.

Extended Data Fig. 1 The virus genetic drift data for SARS-CoV-2 from December 2019 to February 2023. **(a)** the proportion of variants of concern, **(b)** genetic viral diversity and **(c)** antigenic drift of SARS-CoV-2 over time. The data sets are collected from GISAID (https://gisaid.org). Originating Lab and Submitting Lab and the list of Authors, visit https://doi.org/10.55876/gis8.250805xc.

Extended Data Fig. 2 The posterior distributions of transmission parameters across provinces: **(a)** the proportion of infection cases influenced by social reinforcement mechanisms across provinces, **(b)** the impact of collective interaction on susceptibility across provinces.

Extended Data Fig. 3. The spatiotemporal distribution of eigenvector-based key source locations in mainland China from January 2020 to February 2023, presented at a bimonthly resolution. Color gradients ranging from light to dark denote increasing importance of each location as a source, with corresponding values of the importance index indicated in the legend.

Extended Data Fig. 4. The spatiotemporal distribution of eigenvector-based vital sink locations in mainland China from January 2020 to February 2023 on a bimonthly scale. The color from light to dark indicates the importance of location from weak to strong, and the importance index in specific regions is shown in the legend.

Extended Data Fig. 5. Key link properties of SARS-CoV-2 transmission networks in mainland China, quantified by link weight centrality at bimonthly intervals from January–February 2020 to January–February 2023. Higher centrality values denote greater significance within the network.

Extended Data Fig. 6. Key link properties of SARS-CoV-2 transmission networks in mainland China, quantified by betweenness centrality at bimonthly intervals from January–February 2020 to January–February 2023. Higher centrality values denote greater significance within the network.

Extended Data Fig. 7. The geographic cluster distribution of the transmission network at a bimonthly scale from January 2020 to February 2023. The same colour in a subfigure denotes the same cluster of provinces. The node size represents the outflows of unreported infectious cases and the exposed population in that province. For each period, we showed the top 100 directed risk mobility.

Extended Data Fig. 8. The associations between human mobility and SARS-CoV-2 transmission dynamics across three pandemic periods in mainland China from January 2020 to February 2023. (a, d) Relationships during the ancestral strain period: (a) reported cases versus inter-provincial mobility, (d) reported cases versus effective distance (b, e) and (c, f) present analogous analyses for the Alpha/Delta and Omicron variant periods, respectively.

Extended Data Fig. 9. Temporal trends in the regional impacts of non-pharmaceutical interventions (NPIs) on COVID-19 transmission rate across mainland China from January 2020 to February 2023. Each section of the figure represents a different province.

Extended Data Fig. 10. The spatiotemporal distribution of the impact of NPI on COVID-19 transmission rate in mainland China from January 2020 to February 2023 on a bimonthly scale. Color intensity, progressing from light to dark, corresponds to increasing impact levels, with quantitative values for specific regions detailed in the accompanying legend.

53 Hatter, L., Eathorne, A., Hills, T., Bruce, P. & Beasley, R. Respiratory syncytial virus: paying the immunity debt with interest. *The Lancet Child & Adolescent Health* **5**, e44–e45 (2021). https://doi.org/10.1016/S2352-4642(21)00333-3

54 Li, Y. *et al.* The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis* **21**, 193–202 (2021). https://doi.org/10.1016/S1473-3099(20)30785-4

55 Haug, N. *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav* **4**, 1303–1312 (2020). https://doi.org/10.1038/s41562-020-01009-0

56 Dehning, J. *et al.* Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, 160–+ (2020). https://doi.org/10.1126/science.abb9789

57 Alrasheed, H., Alballa, N., Al-Turaiki, I., Almutlaq, F. & Alabduljabbar, R. City Transmission Networks: Unraveling Disease Spread Dynamics. *ISPRS International Journal of Geo-Information* **13**, 283 (2024). https://doi.org/10.3390/ijgi13080283

58 Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Phys Rep* **424**, 175–308 (2006). https://doi.org/10.1016/j.physrep.2005.10.009

59 Brandes, U. A faster algorithm for betweenness centrality. *J Math Sociol* **25**, 163–177 (2001). https://doi.org/Doi 10.1080/0022250x.2001.9990249

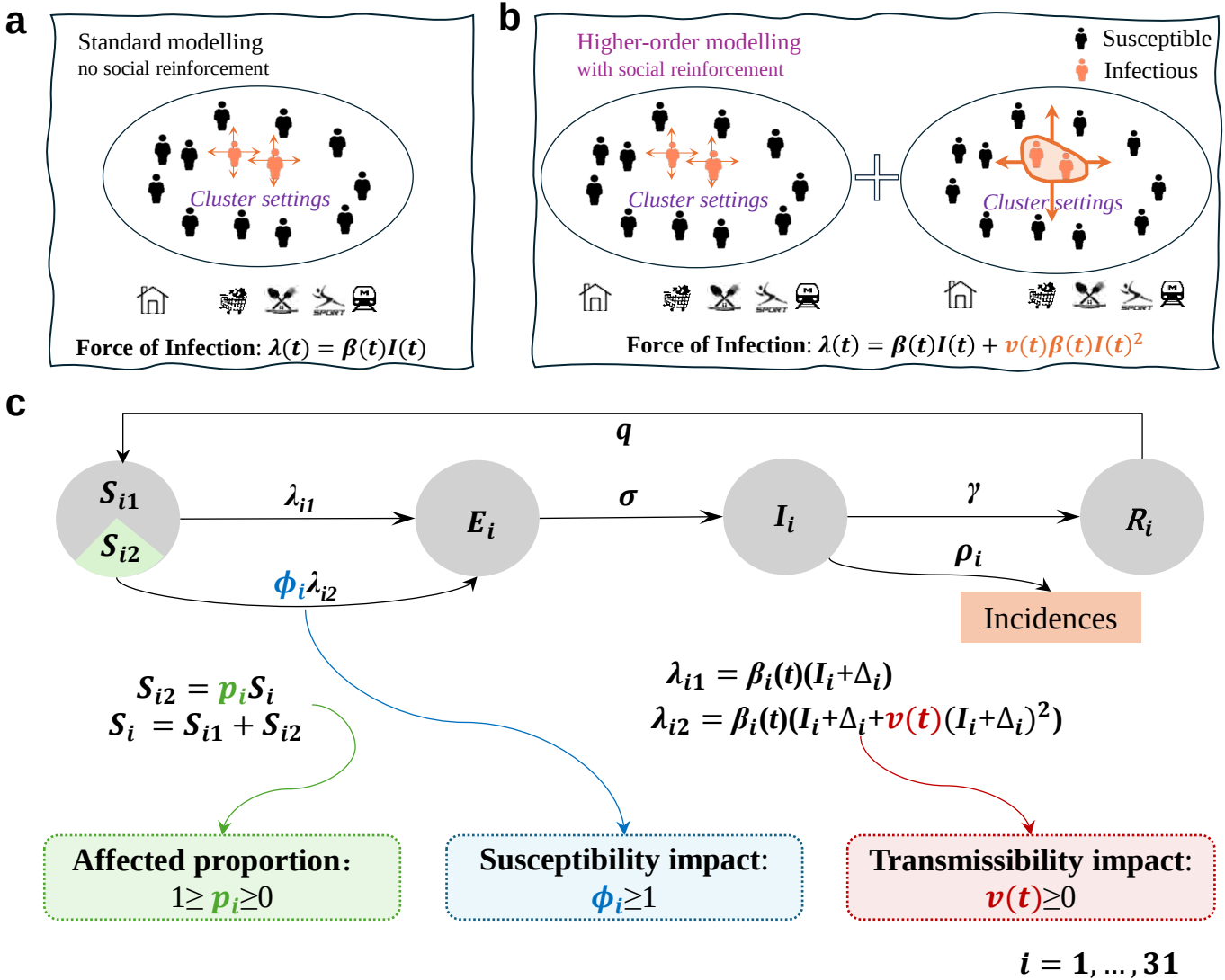
60 Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A* **105**, 1118–1123 (2008). https://doi.org/10.1073/pnas.0706851105

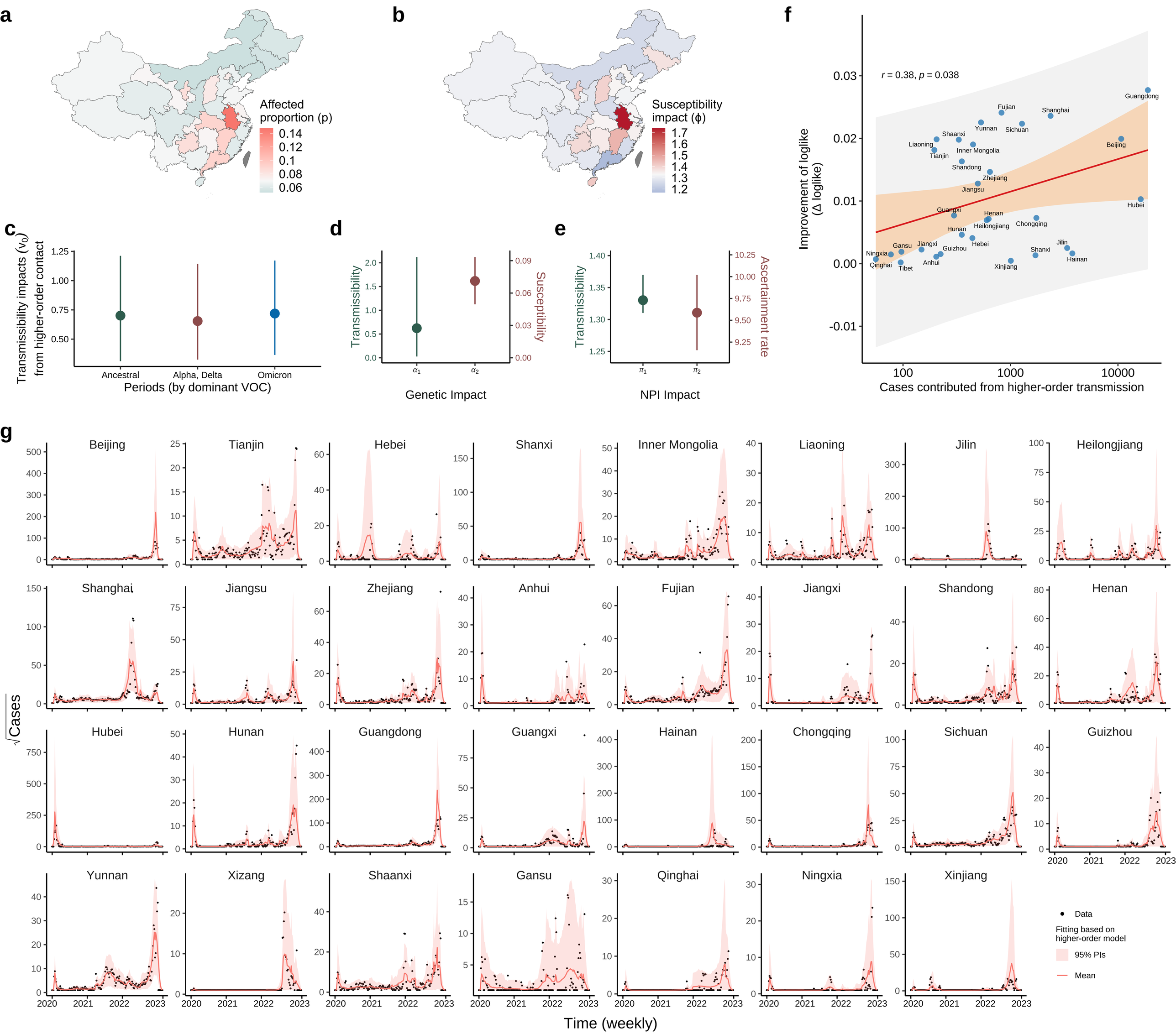
61 Brandes, U. *et al.* On modularity clustering. *Ieee T Knowl Data En* **20**, 172–188 (2008). https://doi.org/10.1109/Tkde.2007.190689

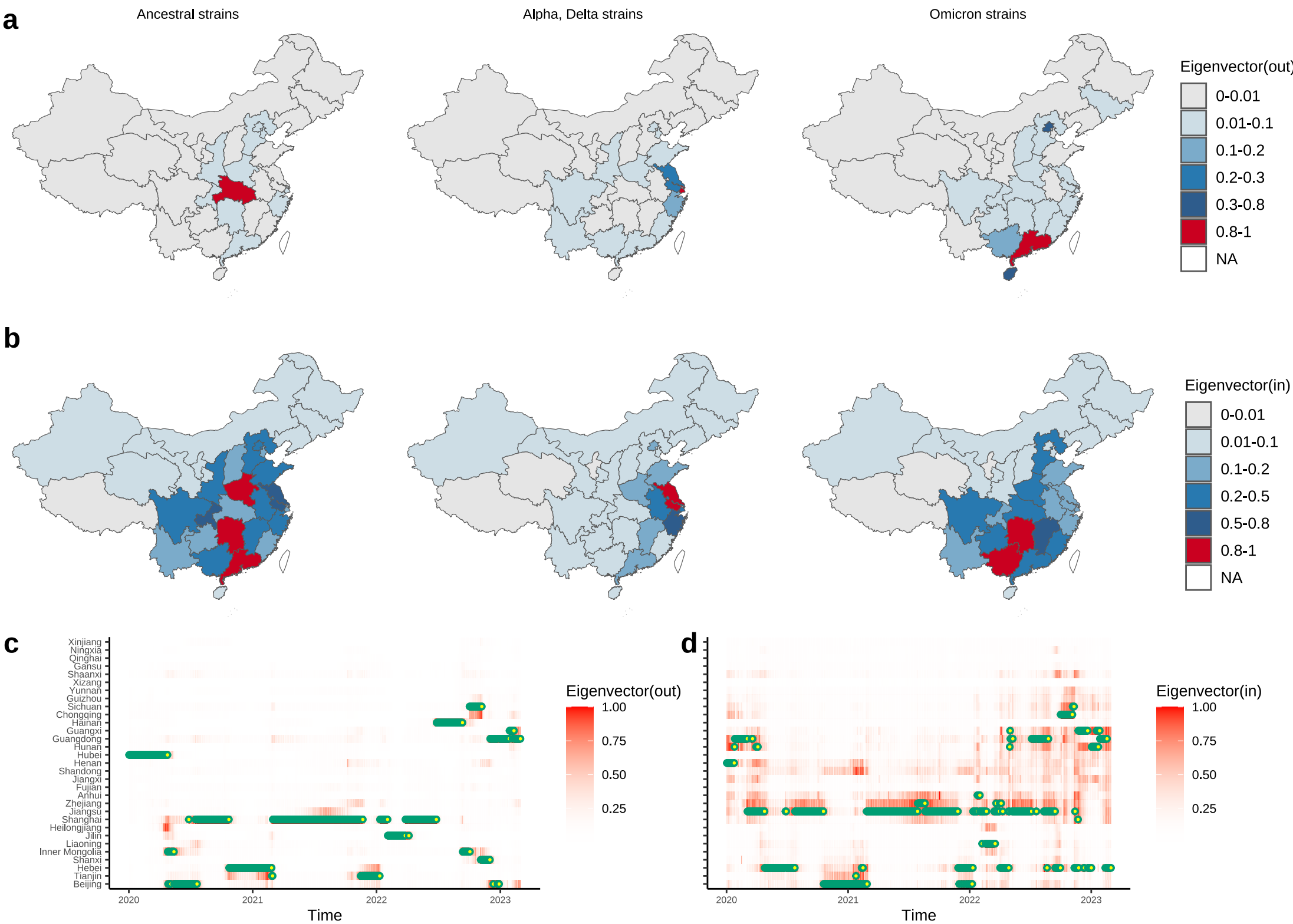
62 Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* **20**, 533–534 (2020). https://doi.org/10.1016/S1473-3099(20)30120-1

63 Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020). https://doi.org/doi:10.1126/science.abb3221

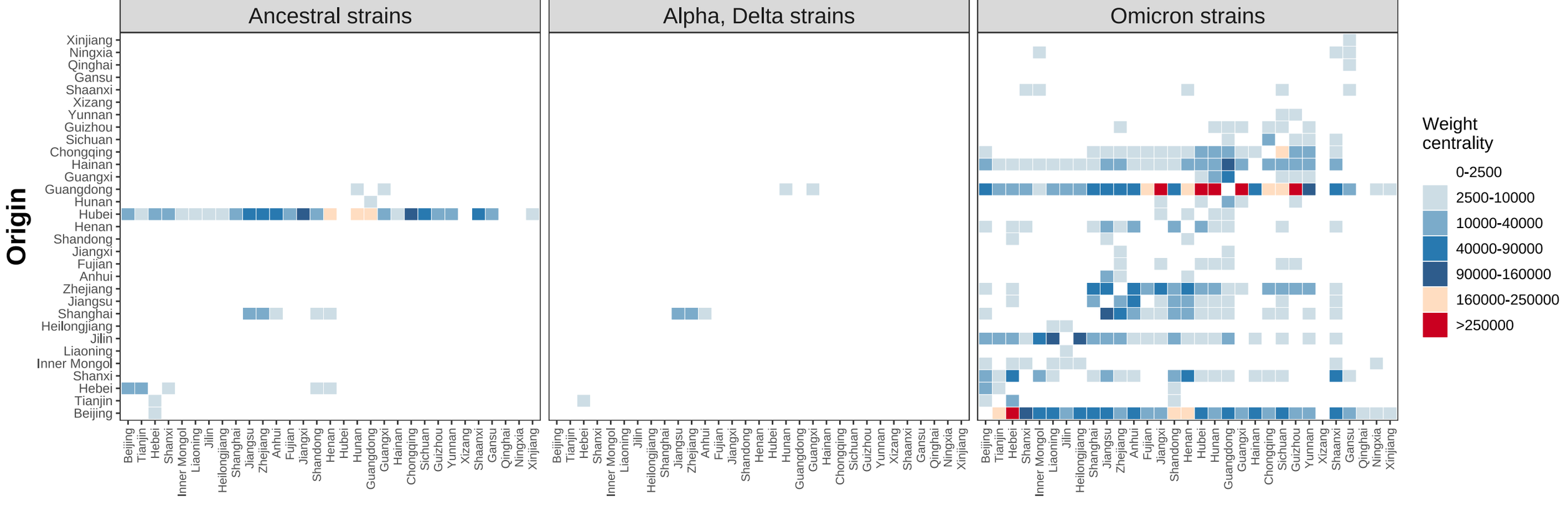
64 Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* **5**, 529–538 (2021). https://doi.org/10.1038/s41562-021-01079-8


a
Standard modelling
no social reinforcement
Cluster settings
Force of Infection: $\lambda(t) = \beta(t)I(t)$
b
Higher-order modelling
with social reinforcement
Susceptible
Infectious
Cluster settings
Cluster settings
Force of Infection: $\lambda(t) = \beta(t)I(t) + v(t)\beta(t)I(t)^2$
c
q
S_{i1}
S_{i2}
λ_{i1}
E_i
σ
I_i
γ
R_i
ρ_i
Incidences
$\phi_i\lambda_{i2}$
$S_{i2} = p_i S_i$
$S_i = S_{i1} + S_{i2}$
$\lambda_{i1} = \beta_i(t)(I_i+\Delta_i)$
$\lambda_{i2} = \beta_i(t)(I_i+\Delta_i+v(t)(I_i+\Delta_i)^2)$
Affected proportion:
$1 \geq p_i \geq 0$
Susceptibility impact:
$\phi_i \geq 1$
Transmissibility impact:
$v(t) \geq 0$
$i = 1, \ldots, 31$

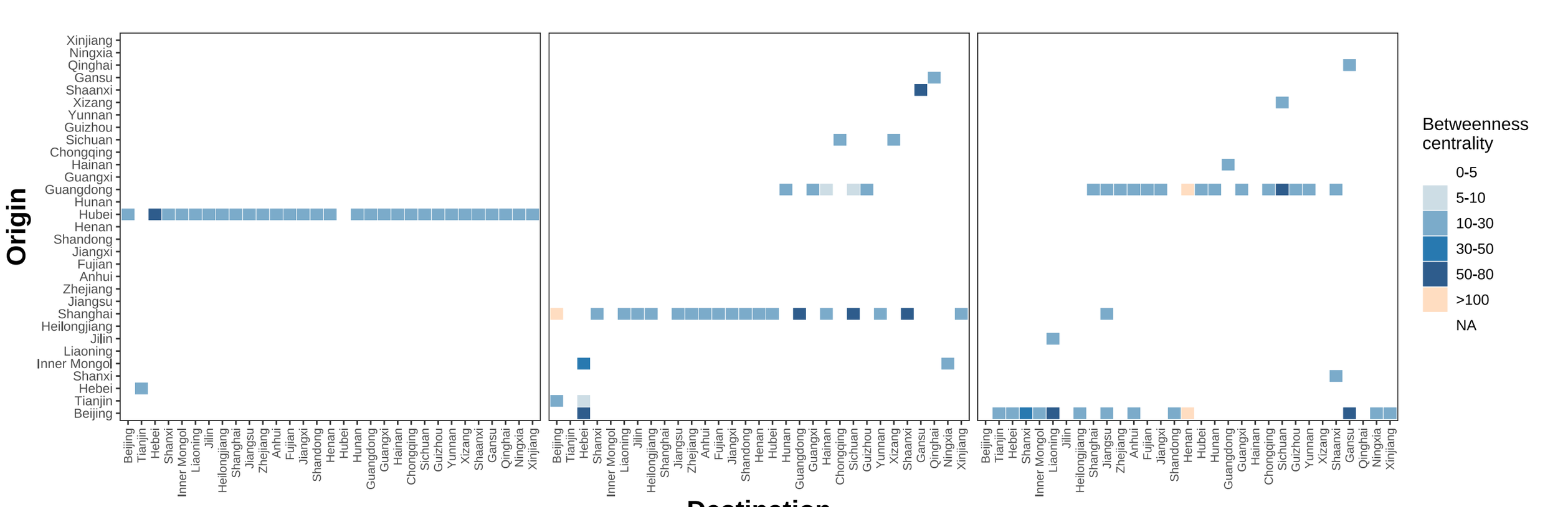

a
Affected proportion (p)
0.14
0.12
0.1
0.08
0.06
b
Susceptibility impact (ϕ)
1.7
1.6
1.5
1.4
1.3
1.2
c
Transmissibility impacts (v_0) from higher-order contact
1.25
1.00
0.75
0.50
Ancestral
Alpha, Delta
Omicron
Periods (by dominant VOC)
d
Transmissibility
Susceptibility
2.0
1.5
1.0
0.5
0.0
0.09
0.06
0.03
0.00
α_1
α_2
Genetic Impact
e
Transmissibility
Ascertainment rate
1.40
1.35
1.30
1.25
10.25
10.00
9.75
9.50
9.25
π_1
π_2
NPI Impact
f
r = 0.38, p = 0.038
Improvement of loglike (Δ loglike)
0.03
0.02
0.01
0.00
-0.01
Guangdong
Fujian
Shanghai
Yunnan
Sichuan
Shaanxi
Liaoning
Inner Mongolia
Beijing
Tianjin
Shandong
Zhejiang
Jiangsu
Hubei
Guangxi
Henan
Heilongjiang
Chongqing
Hunan
Hebei
Gansu
Jiangxi
Guizhou
Jilin
Shanxi
Ningxia
Qinghai
Tibet
Anhui
Hainan
Xinjiang
100
1000
10000
Cases contributed from higher-order transmission
g
Beijing
Tianjin
Hebei
Shanxi
Inner Mongolia
Liaoning
Jilin
Heilongjiang
Shanghai
Jiangsu
Zhejiang
Anhui
Fujian
Jiangxi
Shandong
Henan
Hubei
Hunan
Guangdong
Guangxi
Hainan
Chongqing
Sichuan
Guizhou
Yunnan
Xizang
Shaanxi
Gansu
Qinghai
Ningxia
Xinjiang
√Cases
2020
2021
2022
2023
Time (weekly)
Data
Fitting based on higher-order model
95% PIs
Mean


a
Ancestral strains
Alpha, Delta strains
Omicron strains
Eigenvector(out)
0-0.01
0.01-0.1
0.1-0.2
0.2-0.3
0.3-0.8
0.8-1
NA
b
Eigenvector(in)
0-0.01
0.01-0.1
0.1-0.2
0.2-0.5
0.5-0.8
0.8-1
NA
c
Xinjiang
Ningxia
Qinghai
Gansu
Shaanxi
Xizang
Yunnan
Guizhou
Sichuan
Chongqing
Hainan
Guangxi
Guangdong
Hunan
Hubei
Henan
Shandong
Jiangxi
Fujian
Anhui
Zhejiang
Jiangsu
Shanghai
Heilongjiang
Jilin
Liaoning
Inner Mongolia
Shanxi
Hebei
Tianjin
Beijing
2020
2021
2022
2023
Time
Eigenvector(out)
1.00
0.75
0.50
0.25
d
2020
2021
2022
2023
Time
Eigenvector(in)
1.00
0.75
0.50
0.25

a


Ancestral strains
Alpha, Delta strains
Omicron strains
Origin
Xinjiang
Ningxia
Qinghai
Gansu
Shaanxi
Xizang
Yunnan
Guizhou
Sichuan
Chongqing
Hainan
Guangxi
Guangdong
Hunan
Hubei
Henan
Shandong
Jiangxi
Fujian
Anhui
Zhejiang
Jiangsu
Shanghai
Heilongjiang
Jilin
Liaoning
Inner Mongol
Shanxi
Hebei
Tianjin
Beijing
Weight centrality
0-2500
2500-10000
10000-40000
40000-90000
90000-160000
160000-250000
>250000

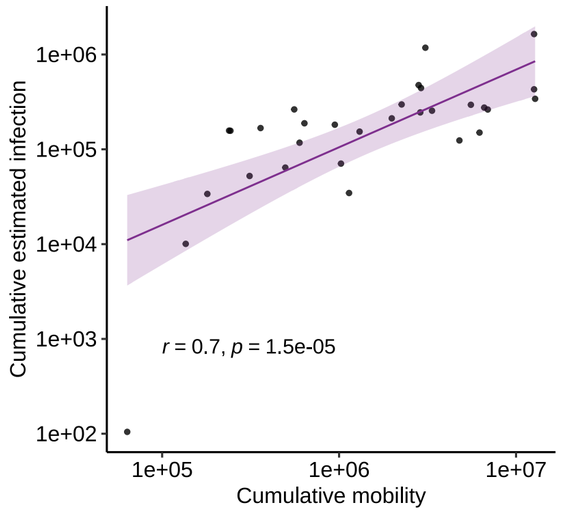
b


Origin
Destination
Betweenness centrality
0-5
5-10
10-30
30-50
50-80
>100
NA

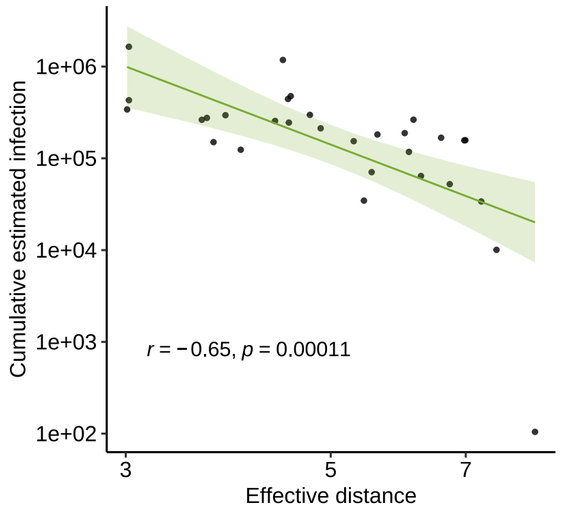
a

Ancestral strains

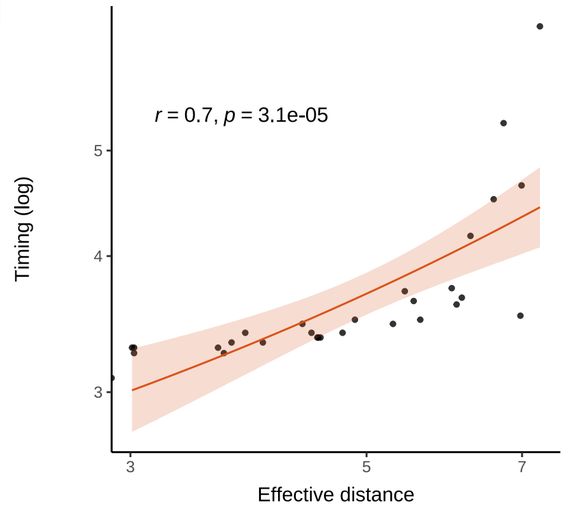
b



c

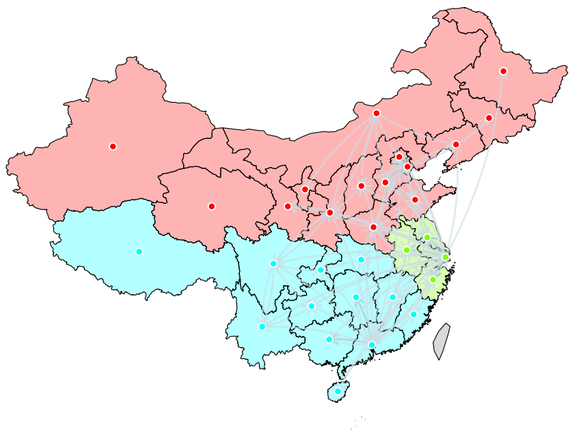


d

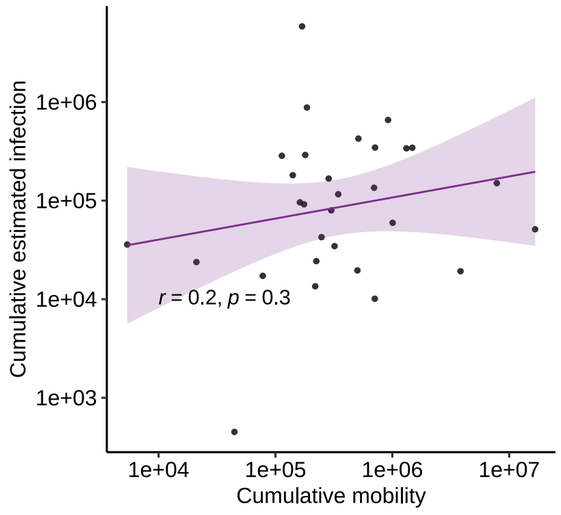


e

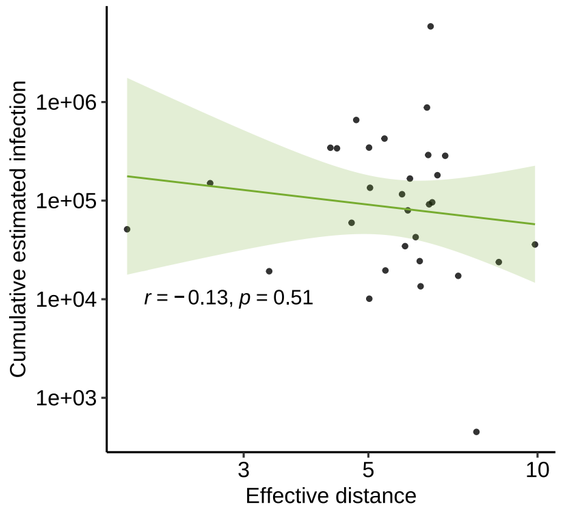
Alpha, Delta strains



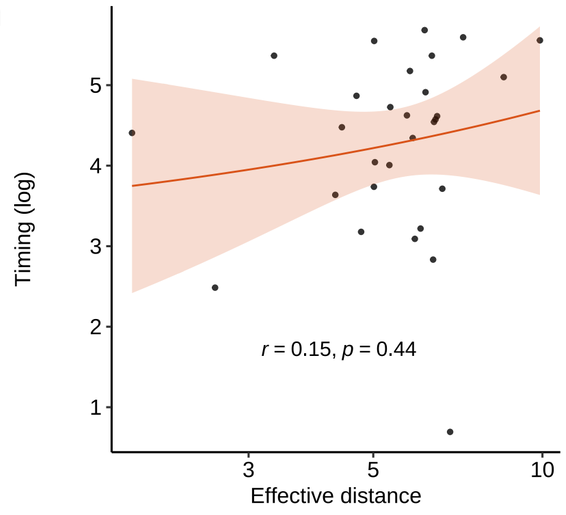
f



g

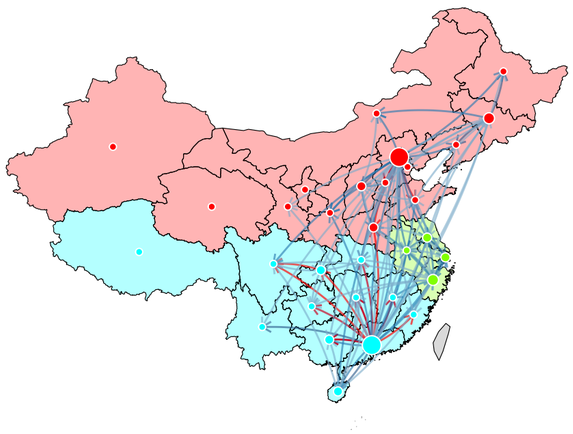


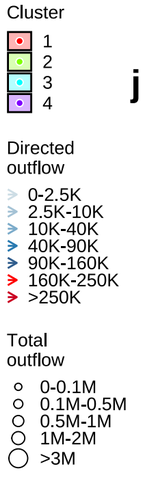
h



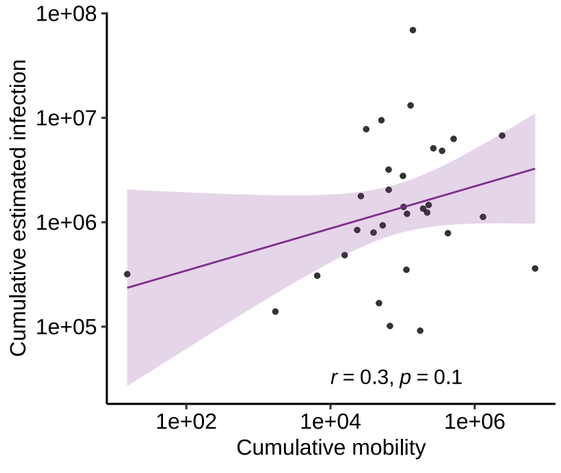
i

Omicron strains

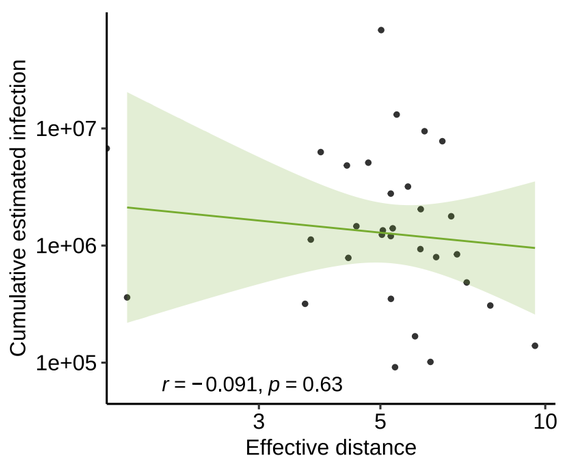




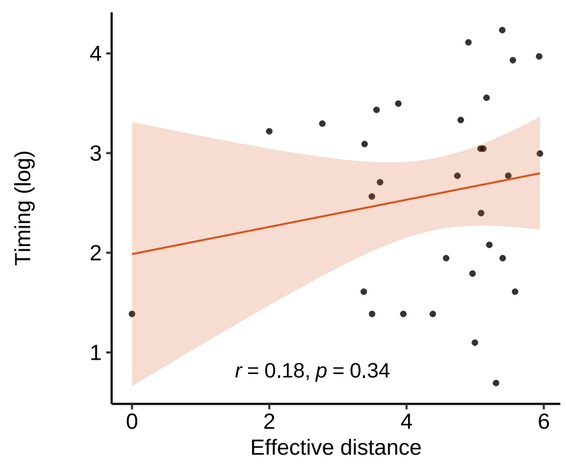
j

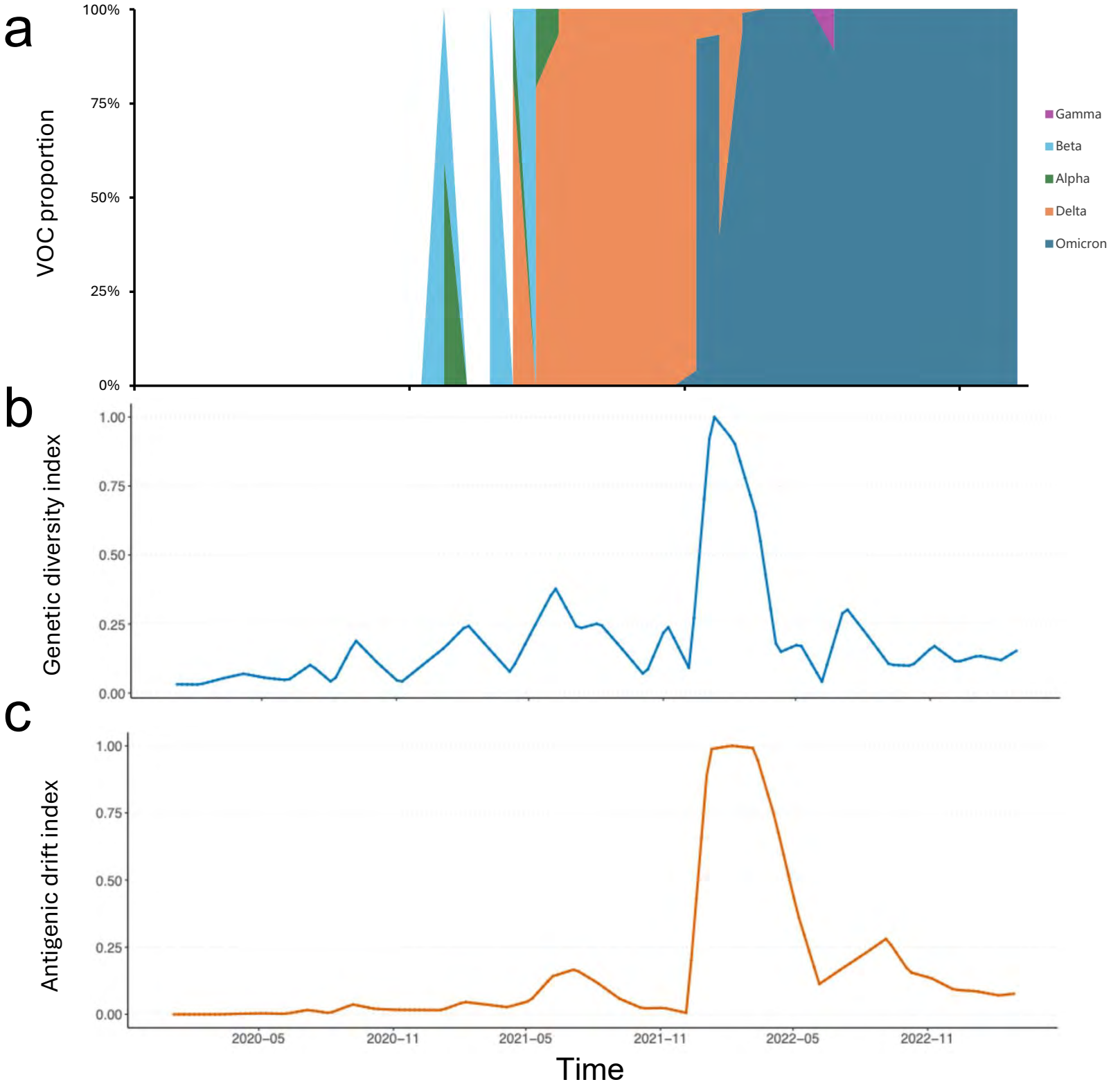


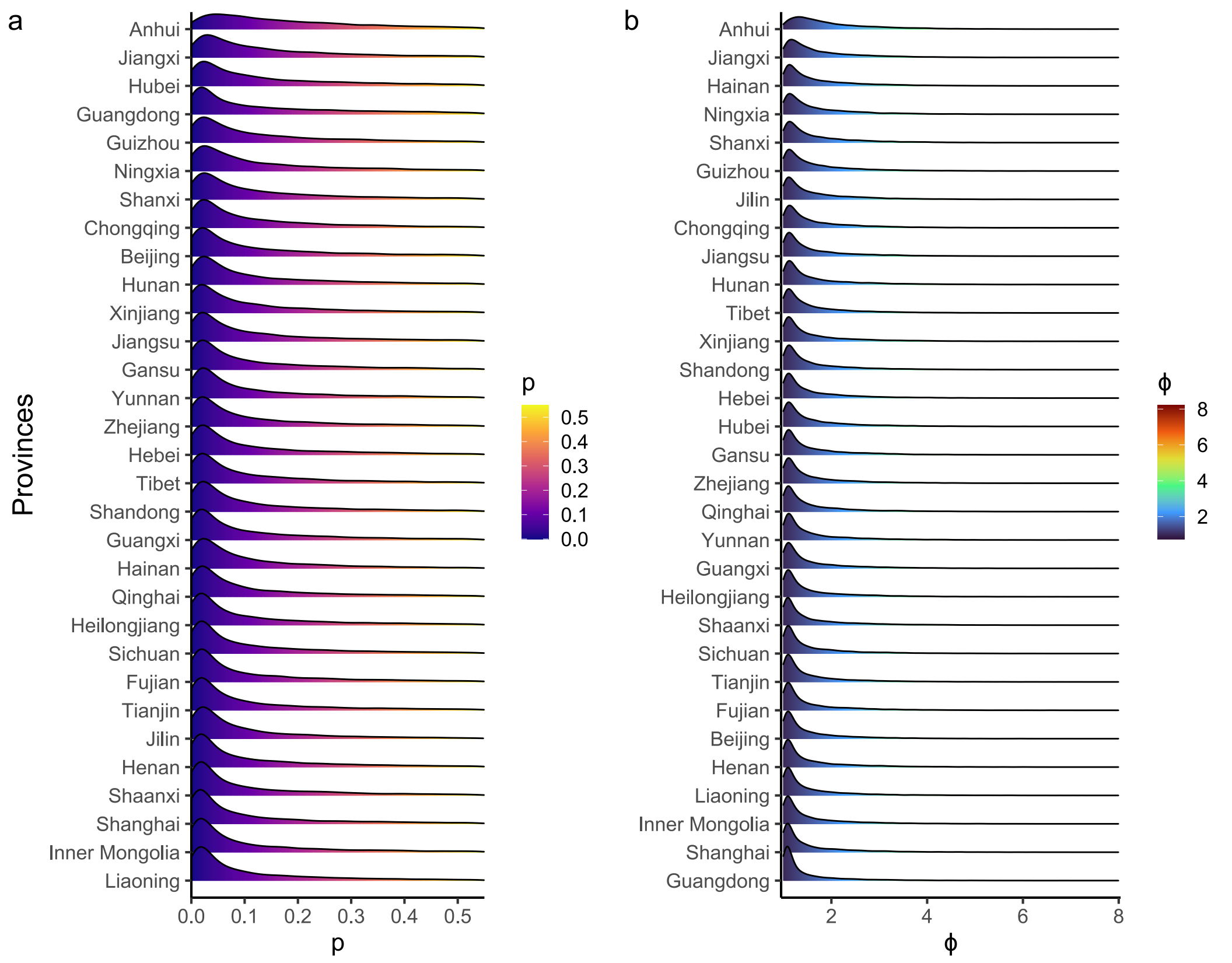
k



l




a
VOC proportion
100%
75%
50%
25%
0%
Gamma
Beta
Alpha
Delta
Omicron
b
Genetic diversity index
1.00
0.75
0.50
0.25
0.00
c
Antigenic drift index
1.00
0.75
0.50
0.25
0.00
2020–05
2020–11
2021–05
2021–11
2022–05
2022–11
Time


a
b
Provinces
Anhui
Jiangxi
Hubei
Guangdong
Guizhou
Ningxia
Shanxi
Chongqing
Beijing
Hunan
Xinjiang
Jiangsu
Gansu
Yunnan
Zhejiang
Hebei
Tibet
Shandong
Guangxi
Hainan
Qinghai
Heilongjiang
Sichuan
Fujian
Tianjin
Jilin
Henan
Shaanxi
Shanghai
Inner Mongolia
Liaoning
0.0
0.1
0.2
0.3
0.4
0.5
p
p
Anhui
Jiangxi
Hainan
Ningxia
Shanxi
Guizhou
Jilin
Chongqing
Jiangsu
Hunan
Tibet
Xinjiang
Shandong
Hebei
Hubei
Gansu
Zhejiang
Qinghai
Yunnan
Guangxi
Heilongjiang
Shaanxi
Sichuan
Tianjin
Fujian
Beijing
Henan
Liaoning
Inner Mongolia
Shanghai
Guangdong
2
4
6
8
ϕ
ϕ

a. Jan-Feb 2020

b. Mar-Apr 2020

c. May-Jun 2020

d. Jul-Aug 2020

e. Sep-Oct 2020

f. Nov-Dec 2020

g. Jan-Feb 2021

h. Mar-Apr 2021

i. May-Jun 2021

j. Jul-Aug 2021

k. Sep-Oct 2021

l. Nov-Dec 2021

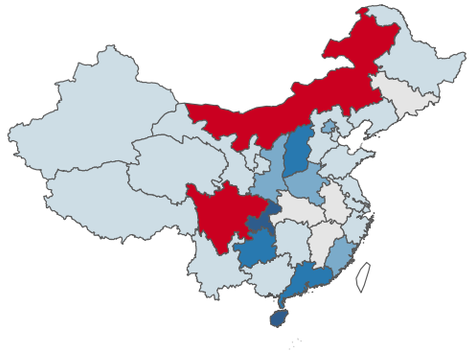
m. Jan-Feb 2022

n. Mar-Apr 2022

o. May-Jun 2022

p. Jul-Aug 2022

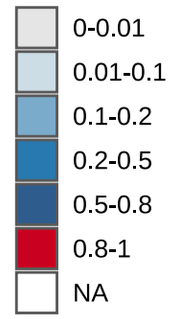
q. Sep-Oct 2022



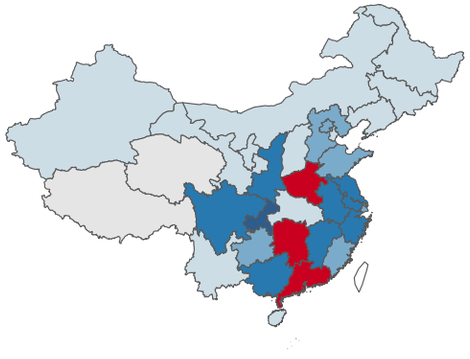
r. Nov-Dec 2022

s. Jan-Feb 2023

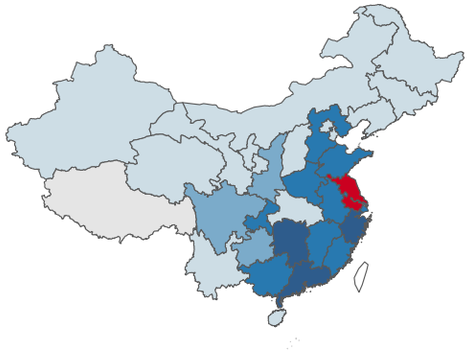
Eigenvector(out)



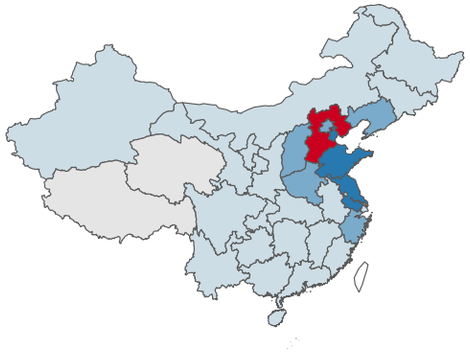
a. Jan-Feb 2020



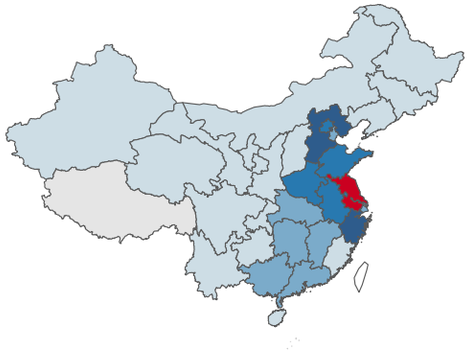
b. Mar-Apr 2020



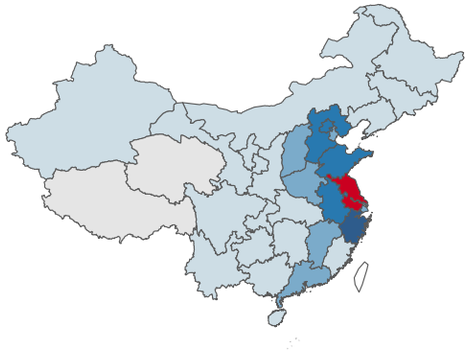
c. May-Jun 2020



d. Jul-Aug 2020



e. Sep-Oct 2020



f. Nov-Dec 2020

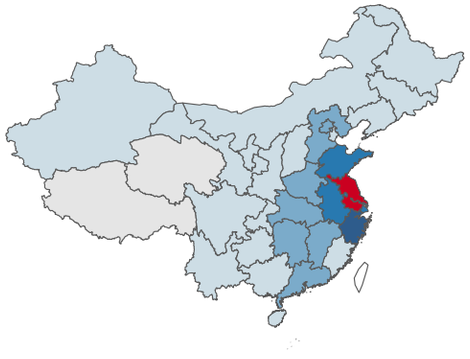
g. Jan-Feb 2021

h. Mar-Apr 2021

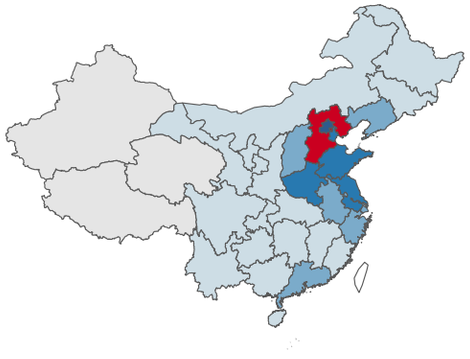
i. May-Jun 2021

j. Jul-Aug 2021

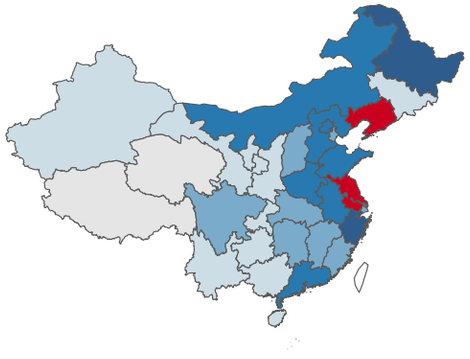
k. Sep-Oct 2021



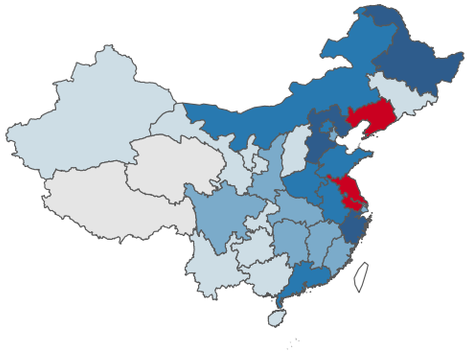
l. Nov-Dec 2021



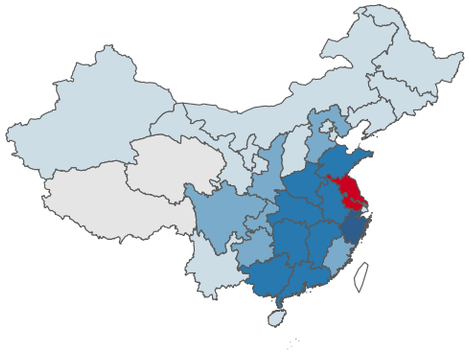
m. Jan-Feb 2022



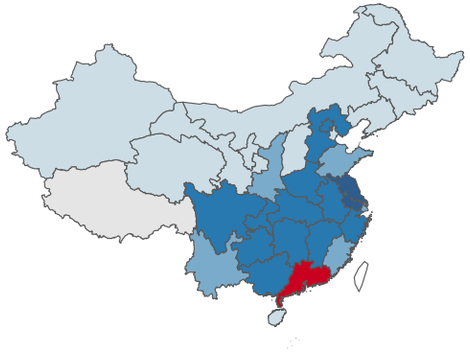
n. Mar-Apr 2022



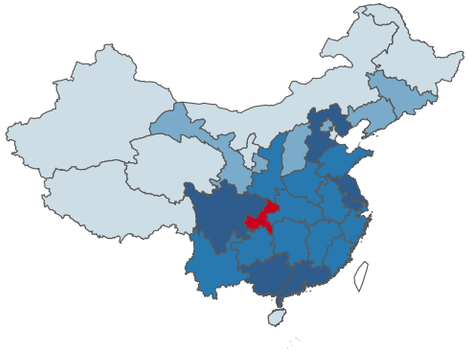
o. May-Jun 2022



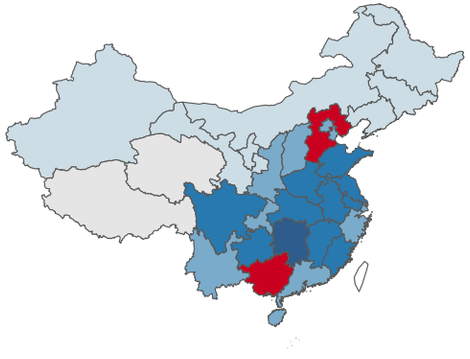
p. Jul-Aug 2022



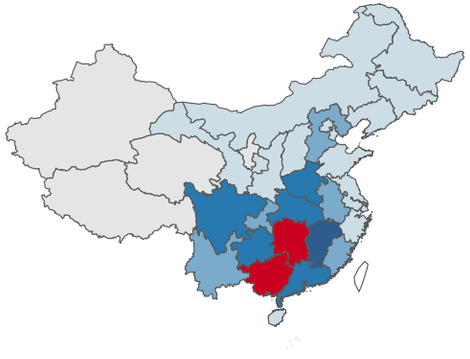
q. Sep-Oct 2022



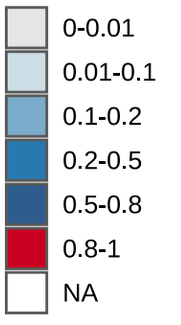
r. Nov-Dec 2022



s. Jan-Feb 2023



Eigenvector(in)



a. Jan-Feb 2020
b. Mar-Apr 2020
c. May-Jun 2020
d. Jul-Aug 2020
e. Sep-Oct 2020
f. Nov-Dec 2020
g. Jan-Feb 2021
h. Mar-Apr 2021
i. May-Jun 2021
j. Jul-Aug 2021
k. Sep-Oct 2021
l. Nov-Dec 2021
m. Jan-Feb 2022
n. Mar-Apr 2022
o. May-Jun 2022
p. Jul-Aug 2022
q. Sep-Oct 2022
r. Nov-Dec 2022
s. Jan-Feb 2023

Origin

Destination

Xinjiang, Ningxia, Qinghai, Gansu, Shaanxi, Xizang, Yunnan, Guizhou, Sichuan, Chongqing, Hainan, Guangxi, Guangdong, Hunan, Hubei, Henan, Shandong, Jiangxi, Fujian, Anhui, Zhejiang, Jiangsu, Shanghai, Heilongjiang, Jilin, Liaoning, Inner Mongol, Shanxi, Hebei, Tianjin, Beijing

Weight centrality
0-100
100-2500
2500-10000
10000-40000
40000-90000
90000-160000
>160000

a. Jan-Feb 2020
b. Mar-Apr 2020
c. May-Jun 2020
d. Jul-Aug 2020
e. Sep-Oct 2020
f. Nov-Dec 2020
g. Jan-Feb 2021
h. Mar-Apr 2021
i. May-Jun 2021
j. Jul-Aug 2021
k. Sep-Oct 2021
l. Nov-Dec 2021
m. Jan-Feb 2022
n. Mar-Apr 2022
o. May-Jun 2022
p. Jul-Aug 2022
q. Sep-Oct 2022
r. Nov-Dec 2022
s. Jan-Feb 2023

Origin

Destination

Betweenness centrality
0-5
5-10
10-30
30-50
50-80
80-100
>100
NA

a. Jan-Feb 2020

b. Mar-Apr 2020

c. May-Jun 2020

d. Jul-Aug 2020

e. Sep-Oct 2020

f. Nov-Dec 2020

g. Jan-Feb 2021

h. Mar-Apr 2021

i. May-Jun 2021

j. Jul-Aug 2021

k. Sep-Oct 2021

l. Nov-Dec 2021

m. Jan-Feb 2022

n. Mar-Apr 2022

o. May-Jun 2022

p. Jul-Aug 2022

q. Sep-Oct 2022

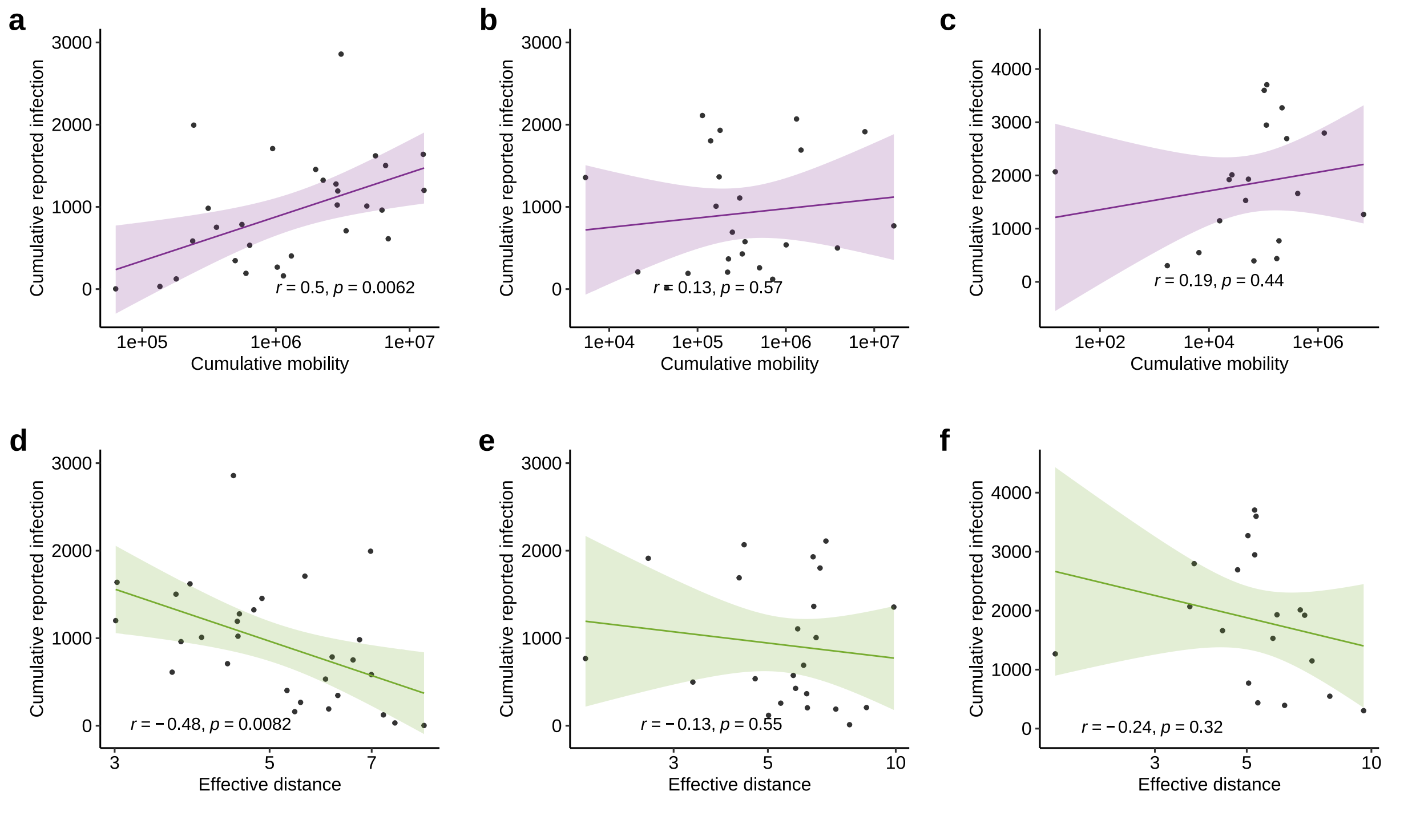
r. Nov-Dec 2022

s. Jan-Feb 2023

Cluster
1
2
3
4

Directed outflow
0-2.5K
2.5K-10K
10K-40K
40K-90K
90K-160K
160K-250K
>250K

Total outflow
0-0.1M
0.1M-0.5M
1M-2M
2M-3M


a
Cumulative reported infection
3000
2000
1000
0
1e+05
1e+06
1e+07
Cumulative mobility
r = 0.5, p = 0.0062
b
Cumulative reported infection
3000
2000
1000
0
1e+04
1e+05
1e+06
1e+07
Cumulative mobility
r = 0.13, p = 0.57
c
Cumulative reported infection
4000
3000
2000
1000
0
1e+02
1e+04
1e+06
Cumulative mobility
r = 0.19, p = 0.44
d
Cumulative reported infection
3000
2000
1000
0
3
5
7
Effective distance
r = −0.48, p = 0.0082
e
Cumulative reported infection
3000
2000
1000
0
3
5
10
Effective distance
r = −0.13, p = 0.55
f
Cumulative reported infection
4000
3000
2000
1000
0
3
5
10
Effective distance
r = −0.24, p = 0.32

Beijing
Tianjin
Hebei
Shanxi
Inner Mongol
Liaoning
Jilin
Heilongjiang
Shanghai
Jiangsu
Zhejiang
Anhui
Fujian
Jiangxi
Shandong
Henan
Hubei
Hunan
Guangdong
Guangxi
Hainan
Chongqing
Sichuan
Guizhou
Yunnan
Xizang
Shaanxi
Gansu
Qinghai
Ningxia
Xinjiang
NPI impact
β_{NPI}
Time
NPI impact
β_{NPI}
Mean
95% CrI

a. Jan-Feb 2020

b. Mar-Apr 2020

c. May-Jun 2020

d. Jul-Aug 2020

e. Sep-Oct 2020

f. Nov-Dec 2020

g. Jan-Feb 2021

h. Mar-Apr 2021

i. May-Jun 2021

j. Jul-Aug 2021

k. Sep-Oct 2021

l. Nov-Dec 2021

m. Jan-Feb 2022

n. Mar-Apr 2022

o. May-Jun 2022

p. Jul-Aug 2022

q. Sep-Oct 2022

r. Nov-Dec 2022

s. Jan-Feb 2023

NPI impact

β_{NPI}

0.6

0.5

0.4