



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Pathogen diversity in meta-population networks

Yanyi Nie^{a,b}, Xiaoni Zhong^a, Tao Lin^{b,*}, Wei Wang^{a,*}

^a School of Public Health, Chongqing Medical University, Chongqing, 400016, China

^b College of Computer Science, Sichuan University, Chengdu 610065, China

ARTICLE INFO

Keywords:

Pathogen mutation
Pathogen diversity
Meta-population networks

ABSTRACT

The pathogen diversity means that multiple strains coexist, and widely exist in the biology systems. The new mutation of SARS-CoV-2 leading to worldwide pathogen diversity is a typical example. What are the main factors of inducing the pathogen diversity? Previous studies indicated the pathogen mutation is the most important reason for inducing the pathogen diversity. The traffic network and gene network are crucial in shaping the dynamics of pathogen contagion, while their roles for the pathogen diversity still lacking a theoretical study. To this end, we propose a reaction–diffusion process of pathogens with mutations on meta-population networks, which includes population movement and strain mutation. We extend the Microscopic Markov Chain Approach (MMCA) to describe the model. Traffic networks make pathogen diversity more likely to occur in cities with lower infection densities. The likelihood of pathogen diversity is low in cities with short effective distances in the traffic network. Star-type gene network is more likely to lead to pathogen diversity than lattice-type and chain-type gene networks. When pathogen localization is present, infection is localized to strains that are at the endpoints of the gene network. Both the increased probability of movement and mutation promote pathogen diversity. The results also show that the population tends to move to cities with short effective distances, resulting in the infection density is high.

1. Introduction

Pathogenic mutations allow viral species to frequently undergo rapid evolution [1–3], leading to the phenomenon of pathogen diversity in which multiple strains coexist. COVID-19 is a classic example. Specifically, SARS-CoV-2 virus mutated into various strains such as Beta, Delta, and Omicron [4] during the pandemic. Influenza, ebolavirus, and HIV-1 have high mutation rates and are often described as having significant variability and unpredictable behavior [5–7]. Frequent mutations promote antigenic evolution of viruses, allowing them to evade recognition by the human immune system [8].

Population travel, and strains mutation make epidemics pose a great danger worldwide. Thus, it is important to predict, prevent and control the spread of pathogens. In order to accurately track and predict the spread of epidemics, existing studies attempted to model multi-strain with mutations [9]. In particular, the important feature of numerous viral species is cross-immunity, which is widely considered in pathogens mutation transmission models [10,11]. Related studies have found that cross-immunity can be approximated by the genetic distance of the pathogen [12,13]. In the sub-discipline of multi-strain disease modeling, existing models satisfy biological assumptions through adjusting modeling choices such as antigenic neighborhoods [14], age structure [15], and decision capture immune status [9]. The influence

of a potential genotypic network [16] was also explored in a multi-strain model. Genotype network is an efficient way to represent the genetic distances necessary for cross-immunity in multi-strain models [17]. The ability to control possible mutation paths between strains is the most important feature of the network [18]. Also, the impact of pathogen mutations on transmission is widely discussed. In two linked models, Girvan et al. [18] investigate how the interaction between the memory immune response and pathogen mutation impacts epidemic dynamics [19]. Their main result is that pathogens must mutate rapidly to remain viable in highly connected populations. To account for two important pathways, Williams et al. [20] construct a disease model with an underlying genetic network. They found that the four specific features defined in the model do not affect the classical epidemic threshold, but localize outbreaks around key strains and generate a second threshold for immune invasion. Schwarzendahl et al. [21] generalized the popular susceptible–infection–recovery model to account for mutations that cause recurrent new strains of bacteria. Their model predicts that mutations can cause a hyper-exponential increase in the number of infections at an early stage. Zhang et al. [22] present a comprehensive framework for epidemic spread under pathogen evolution that demonstrates how mutations can fundamentally affect the spread projection.

* Corresponding authors.

E-mail addresses: lintao@scu.edu.cn (T. Lin), wwzqbx@hotmail.com, wwzqbc@cqmu.edu.cn (W. Wang).

Population mobility (such as travel among cities) and distribution characteristics have important implications for the threshold and duration of epidemic outbreaks [23–25]. The meta-population model [26–29] is commonly used to describe the spatial distribution of populations and to discuss the effects of population movements on epidemic transmission. In this model, the population is divided into several sub-populations by geographic location. The different sub-populations are linked by population movement. Unlike the contact network model [30–33], the meta-population network is a coarse-grained description, where each node in the network represents a region (city). Individuals within the same node can react, for example, e.g., transmitting diseases, and they move to neighboring nodes according to certain rules. This reaction–diffusion [34,35] process portrays a real-world scenario of epidemic spread. Balcan et al. [36] provided a theoretical framework for analyzing contagion in a network of sites where people remember their residence. They discovered a phase transition between a regime where the infection impacts a significant portion of the system and one where just a tiny portion is impacted. Mata et al. [35] explore the effect of local population structure on meta-population networks in a contact process reaction–diffusion model. By analyzing several variations of the reaction–diffusion process, they concluded that the critical nature of the contact process model is independent of the population structure on the meta-population. Recently, the meta-population model has also been utilized in real-world virus transmission scenario construction [37,38], providing theoretical support for the development of non-pharmaceutical interventions, such as case isolation and the lockdown of entire populations.

The worldwide spread of pathogens with mutations has led to the existence of two types of regions. In some regions there is pathogen diversity, in others the infection is localized in few pathogens. We consider an area to have pathogen diversity [39–41] if there are infections caused by several different pathogens in that area. Existing studies extend mathematical models to analyze pathogen diversity based on host contact network structure [42], community structure [43], and cross-immune competition [44,45]. Specially, Abu-Raddad et al. [46] used a mathematical model to study the dynamics of antigenically diverse infection factors. They concluded that the location of the threshold is determined by the number of pathogens multiplying and the intensity of cross-immunization. In antigenically distinct pathogens, Gupta et al. [47] analyzed the impact of selection of host immune responses on transmission dynamics. The results indicate that strong selection can lead to stable isolation of pathogen populations into discrete strains with non-overlapping antigenic libraries. In the context of multi-pathogen transmission, Sridhar et al. [48] showed how their work on multi-strain transmission models with mutations answer some of the fundamental problems about COVID-19 transmission. They demonstrated the possible effect of an improbable mutation into a virulent strain using simulations. Based on COVID-19, the importance of incorporating mutations and evolutionary adaptations in epidemic models is also discussed [49].

In general, the effects of population movements (i.e. travel) and gene mutations on pathogen diversity are lacking in depth exploration in existing studies. Here, we therefore aim to introduce a more general model, taking into account the effect of population mobility and pathogen mutations on the localization of infection. Traffic network and gene network are utilized to model population movements and pathogen mutation pathways, respectively. The Microscopic Markov Chain Approach is also extended to mathematically model the reaction–diffusion process of multi-strain pathogens. The experimental results show that cities with lower infection rates are more likely to have pathogen diversity. In cities with short effective distances in traffic network, infection is also more likely to be localized on certain pathogens, leading to a lack of pathogen diversity. We also find that star-type gene networks are more prone to pathogen diversity. The gene network leads to a higher infection density of strains located at the endpoints. Restricting population movement and strain mutation can

also make it more likely that infections in cities is localized to one specific pathogen. Effective distance is used to measure the relationship between population mobility and the infection density. Individuals tend to move to cities with short effective distances, resulting in higher rates of infection in these cities.

The remainder of the paper is laid out as follows. First, we analyzed global data on SARS-CoV-2 variants in Section 2 and observed strain diversity. Then we show the main features of the dynamical model in Section 3. We handle the analytical derivation of the propagation dynamics process in Section 4 once the model has been established. The dynamic equations are also explained analytically in detail in Section 4. In Section 5, theoretical experiments were implemented to observe the phase transition phenomenon. Finally, in Section 6, we conclude the article by outlining the principal results of the work.

2. Empirical analysis

The SARS-CoV-2 virus has mutated into several different strains over the course of the pandemic. To get a clearer view of SARS-CoV-2 variants diversity phenomenon, we use the data collected from the GISAID [51] database to map the evolution of the percentage of SARS-CoV-2 variants from June 2020 to June 2022, as shown in Fig. 1. Strains diversity is clearly observed in most of the time periods. In a small period of time, the strains are relatively single, and the infection is localized on certain strains. To quantitatively measure the extent of infection localization, we introduce Effective Participation Ratio (EPR) [20]

$$\text{EPR} = \frac{M_i^{\text{eff}}}{M} = \frac{(\sum_x I_i^{x*})^2}{M \sum_x (I_i^{x*})^2}, i \in [1, N], \quad (1)$$

where I_i^{x*} is the final proportion of individuals infected with strain x in region i , and M represents the total number of strains. According to Eq. (1), the EPR values are available for global SARS-CoV-2 variants. As $\text{EPR} \rightarrow 1$, the number of infections caused by each strain is equal. At this time, the strains are diverse. As $\text{EPR} \rightarrow 0$, infections are caused by only one pathogen. We think that there is no pathogen diversity, infection is localized to certain strains. Accordingly, the pathogen diversity can be expressed mathematically as $1 - \text{EPR}$. Apparently, the EPR presents values converging to one after October 2021, indicating that localization is prominent during this time period. In contrast, in February 2021 and June 2021, the EPR values are larger, indicating that the global SARS-CoV-2 variants show strong diversity at this time.

3. Model descriptions

In the previous section, we described the phenomenon of diversity of pathogens in the real world. In this section, based on meta-population network, we propose a Susceptible–Infectious–Recovered–Susceptible (SIRS) infection model that allows for pathogenic mutations to occur. Both traffic and gene networks exist in our model, as shown in Fig. 2. The path of agents flowing between N subpopulations is defined by the traffic network. The mutation routes between M strains in the system are defined by the gene network. We describe the traffic network and gene network in detail in Section 3.1 and Section 3.2, respectively. The reaction–diffusion processes in the system involving all agents are described specifically in Section 3.3.

3.1. Traffic network

Inspired by the fact that the spread of epidemics with mutation is largely influenced by the structure of the traffic network [52,53], we use N subpopulations (nodes) constitute the traffic network in the system, as shown in Fig. 2a. For each node $i \in [1, N]$, there are n_i agents residing in the node, i.e., each agent has a corresponding residence i . The connectivity network between N nodes is described by the adjacency matrix \mathcal{W} , where \mathcal{W}_{ij} represents the connection weights

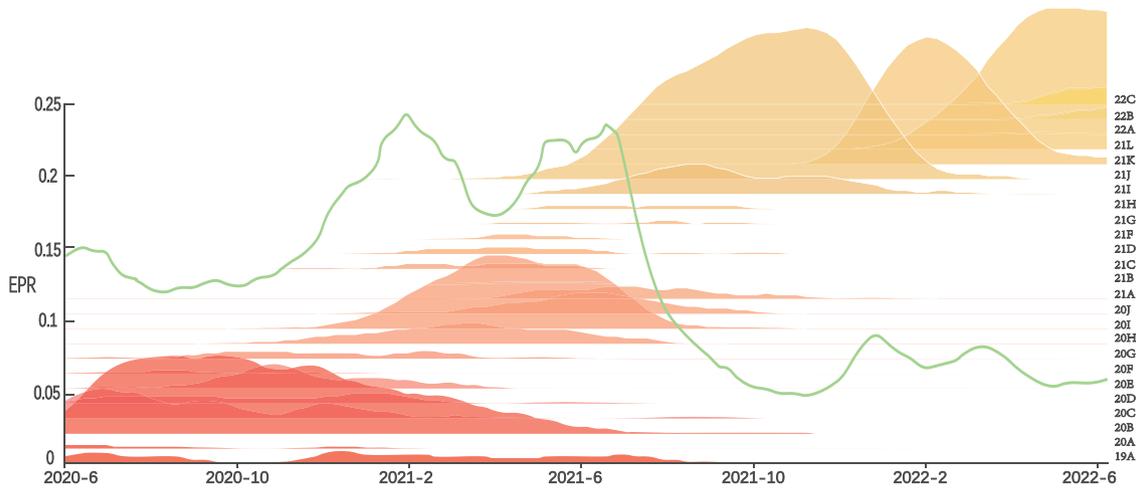


Fig. 1. Diversity of SARS-CoV-2 variants worldwide from June 2020 to June 2022. The graphic content can be divided into two parts. (1) The color block area. Each color row corresponds to a different variant strain, as shown in the legend on the right side. The right vertical coordinates, e.g. 22C, 22B, are the nomenclature of Nextstrain [50] for the SARS-CoV-2 mutant strain. The relative proportions of different strains are shown by the height of the color block area. (2) Green curve. The solid green line shows the EPR values of the strains at different times, corresponding to the vertical coordinates on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

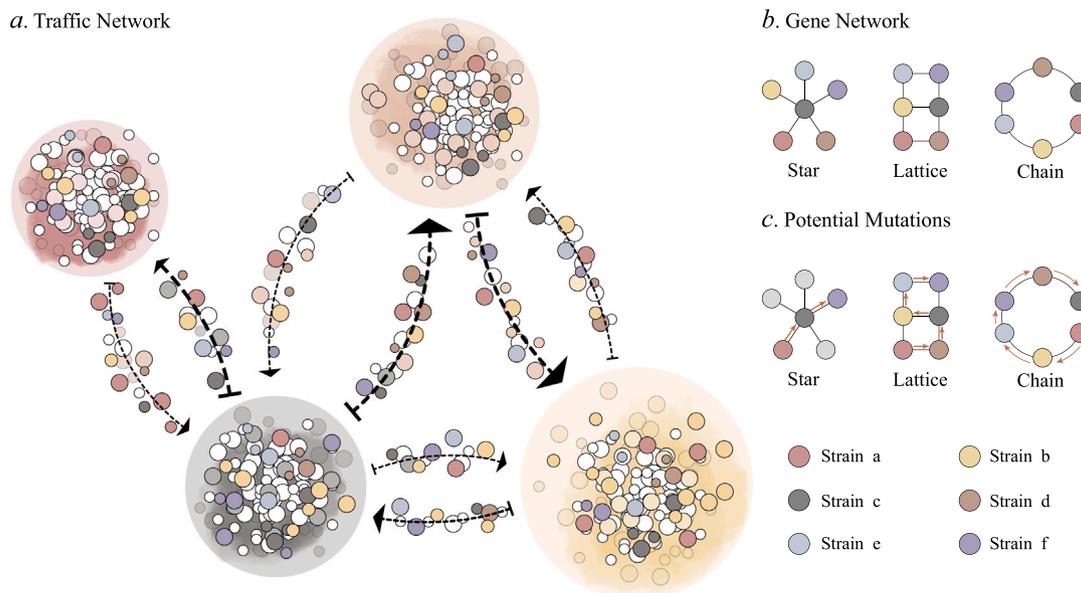


Fig. 2. Schematic diagram of traffic network and gene network. **a.** A traffic network with $N = 4$ subpopulations. The large circles are different subpopulations (nodes). Small circles of different colors in the subpopulations represent agents infected with different pathogens. Individuals move between different subpopulations, and the arrows are the movement paths, while the thickness of the arrows represents the different path weights. **b.** Three different types of gene networks, containing a total of $M = 6$ strains. **c.** One possible mutation pathway in each of the three gene networks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between nodes i and j . If $\mathcal{W}_{ij} \neq 0$, it means that the agents residing in the node have a certain probability of moving to node j ; otherwise, $\mathcal{W}_{ij} = 0$. In our model, the population travels according to the matrix \mathcal{W} , i.e., moves to a certain destination.

3.2. Gene network

The topology of mutation network can affect the mutation process [54]. In our model, the gene network allows us to specify potential mutational pathways between M strains. Fig. 2b shows three examples of genetic networks, consisting of the well-known lattice, chain and star models, respectively. Defined as an unweighted and undirected network, the gene network is described by the symmetric matrix \mathcal{A} , which is a collection of possible mutations. $\mathcal{A}_{xy} = 1$ means that there is a mutation path between strain $x \in [1, M]$ and strain $y \in [1, M]$, i.e., the

strain x can mutate to neighboring strain y , where $x \neq y$; otherwise $\mathcal{A}_{xy} = 0$. Each mutation will only occur between two pathogens that are directly adjacent to each other. As shown in Fig. 2c, taking mutations on a chain network as an example, pathogen a can only mutate to pathogen b, but not directly to pathogen e.

3.3. Reaction–diffusion process

There are three epidemic states for the agents residing in each node, susceptible S , infected (strain x) I_x , and recovered (strain x) R_x , where $x \in [1, M]$. At each time step t , one agent can only be in one of the three states mentioned above, and the reaction–diffusion process in the model is as follows.

(1) Firstly, agents move with probability p_d to a node different from their residence, i.e., travel to another city. Then each agent has probability p_d to move and probability $(1 - p_d)$ to stay at the residence.

(2) If the movement occurs, the target node of each movement is determined based on the network weight matrix \mathcal{W} . Specifically, $\frac{W_{ij}}{\sum_{l=1}^N W_{il}}$ represents the probability that an agent in node i moves to node j . After all movements are completed, the entire population showed a new spatial distribution at this time step.

(3) Based on the new spatial distribution of the meta-population, the contagion and recovery processes occur. Contagion is restricted to different nodes, i.e., the infection process of an agent is only relevant to the situation within the node it is currently in. We assume that the population within each node is well-mixed. The infection model of three epidemic states is as follows.

(i) For S -state agent. Agents in the S -state are not immune to any strains and can be infected by any of them. Then the S -state agent has a probability β of converting to the I_x -state at each contact with an I_x -state individual. We assume that this basic rate of transmission remains constant for all strains.

(ii) For I_x -state agent, two state transitions may occur.

a. I_x -state agents are transformed to R_x -state with recovery probability γ . We define an individual in the R_x -state to be completely immune to strain x and partially immune to strain y ($y \neq x$), i.e., the R_x -state agent may still be infected by strain y and transform into I_y -state.

b. Agents in the I_x -state mutate to the I_y -state with mutation probability μ , where $y \in \mathcal{H}_x$. \mathcal{H}_x is the set of direct neighbors of strain x in gene network \mathcal{A} .

(iii) For R_x -state agent, two state transitions are possible as well.

a. Agents in the R_x -state lose immunity with waning immunity rate α and shift to S -state.

b. R_x -state agents become I_y -state with a reduced rate T_{xy} , of being infected by strain y ($y \neq x$). T_{xy} is defined as an exponential decay function with respect to the genetic distance between strain x and strain y . Specially,

$$T_{xy} = 1 - e^{-L_{xy}/\Delta}.$$

where, $L_{xy} = L_{yx}$ is the genetic distance between strains x, y , which approximates the shortest path between strains x, y in the gene network. Δ ($0 < \Delta < \infty$) is the characteristic length of immunity transcendence [20], allowing us to investigate the immune characteristics of individuals in the system. As $\Delta \rightarrow 0$, then $T_{xy} \rightarrow 1$, immunity is pathogen specific. As $\Delta \rightarrow \infty$, then $T_{xy} \rightarrow 0$, immunity becomes more widespread, achieving universal protection against all pathogens.

(4) Finally, due to the commuting mobility patterns [28,55], all agents returns to the subpopulation where it resides. And another time step begins, $t = t + 1$.

4. Theoretical analysis

By extending the Microscopic Markov Chain Approach [31,33,56, 57], the dynamical process of our model can be described as the following time evolution equations.

Given N nodes and M strains, the variables $I_i^x(t)$, $R_i^x(t)$ and $S_i(t)$ ($i = 1, \dots, N; x = 1, \dots, M$) represent the proportion of I_x -state, R_x -state and S -state individuals residing at node i at time t , respectively. We first introduce the time evolution of $I_i^x(t)$, which has the evolution equation:

$$\begin{aligned} I_i^x(t+1) &= S_i(t)\Pi_i^x(t) \\ &+ I_i^x(t)(1 - \gamma - \mu \sum_{y=1}^M A_{xy}) \\ &+ \sum_{y=1}^M A_{yx} I_i^y(t)\mu \\ &+ \sum_{y=1}^M R_i^y T_{xy} \Pi_i^x(t). \end{aligned} \tag{2}$$

As shown on the right side of the equation, the four components constitute the evolution of $I_i^x(t)$. The first term, $S_i(t)\Pi_i^x(t)$ denotes the fraction of S -state agents residing at node i that infected by strain x and pass to I_x at time $t + 1$. $\Pi_i^x(t)$ is interpreted as the probability that an individual residing at node i is infected by strain x at time t . The second term accounts for the fraction of I_x -state agents residing at node i that do not recover and mutation at time $t + 1$. The third term, denoting the probability that the I_y -state ($y \neq x$) individual residing in node i , at time $t + 1$, transforms to I_x -state due to pathogen mutation. The fourth term is the probability that an individual in R_y -state ($y \neq x$), residing in node i , is infected by strain x and transforms to I_x -state at time $t + 1$. The probability $\Pi_i^x(t)$ reads:

$$\Pi_i^x(t) = (1 - p_d)P_i^x(t) + p_d \sum_{j=1}^N \frac{W_{ij}}{\sum_{l=1}^N W_{il}} P_j^x(t), \tag{3}$$

where W_{ij} is the weight of the link between nodes i and j , while p_d denotes the probability of moving. The first term on the right-hand side indicates the probability that the individual is infected by strain x when remaining at node i . The second one considers the probability that this agent contract the strain x when moving to any neighbor of i . The $P_i^x(t)$ represents the probability that the agent is infected by strain x in node i at time t . In particular, node i is not necessarily the residence of this agent. Under the well-mixed approximation for the subpopulations, this probability is:

$$P_i^x(t) = 1 - \prod_{j=1}^N [1 - \beta I_j^x(t)]^{n_{j \rightarrow i}}, \tag{4}$$

where $n_{j \rightarrow i}$ is the number of agents who move from node j to node i . The expression is:

$$n_{j \rightarrow i} = \delta_{ij}(1 - p_d)n_i + p_d \frac{W_{ji}}{\sum_{l=1}^N W_{jl}} n_j, \tag{5}$$

if $i = j$, $\delta = 1$; otherwise $\delta = 0$.

Then, the evolution process for R_x -state individual residing in node i is presented. Mathematically, the time evolution equation of $R_i^x(t)$ is written as:

$$R_i^x(t+1) = I_i^x(t)\gamma + R_i^x(t)[1 - \alpha - \sum_{y=1}^M T_{xy}\Pi_i^y(t)]. \tag{6}$$

The equation contains two parts. The first term on the right side of the equation represents the probability that an I_x -state individual residing at node i revert to R_x -state at time $t + 1$. The second term accounts for the probability that individuals in the R_x -state residing at node i do not transform to the S -state and infected by other pathogens, i.e., pathogens that are not x , at time $t + 1$.

Denoting $I_i(t)$ as the fraction of infected individuals residing at node i at time t . There are $I_i(t) = \sum_x I_i^x(t)$. Similarly, $R_i(t) = \sum_x R_i^x(t)$ is the proportion of agents residing in node i that are in the recovery state at time t . In our model, there are $S_i(t) + I_i(t) + R_i(t) = 1$.

Iterate the above equations until the dynamic achieves a stable state. The steady-state densities of I_x -state and R_x -state individuals are defined as asymptotic value $I_i^{x*} = \lim_{t \rightarrow \infty} I_i^x(t)$ and $R_i^{x*} = \lim_{t \rightarrow \infty} R_i^x(t)$, respectively. So there are $I_i^* = \sum_x I_i^{x*}$ and $R_i^* = \sum_x R_i^{x*}$ in the steady state. Based on the whole system perspective, defining I^* as the proportion of infected state individuals when system is stable, we have $I^* = 1/N \sum_i I_i^*$. Then the $R^* = 1/N \sum_i R_i^*$ is the proportion of individuals in the recovery state when the system is stable.

Analyzing the critical behavior of the dynamic system, we have

$$\begin{aligned} (\gamma + \mu \sum_{y=1}^M A_{xy}) I_i^{x*} &= [1 - \sum_{y=1}^M (I_i^{y*} + R_i^{y*})] \Pi_i^x \\ &+ \sum_{y=1}^M A_{yx} I_i^{y*} \mu \\ &+ \sum_{y=1}^M R_i^{y*} T_{xy} \Pi_i^x, \end{aligned} \tag{7}$$

and

$$(\alpha + \sum_{y=1}^M T_{xy} \Pi_i^y) R_i^{x*} = I_i^{x*} \gamma, \tag{8}$$

where

$$\begin{aligned} \Pi_i^x &= (1 - p_d) \left(1 - \prod_{j=1}^N (1 - \beta I_j^{x*})^{n_{j \rightarrow i}} \right) \\ &+ p_d \sum_{j=1}^N \frac{W_{ij}}{\sum_{l=1}^N W_{il}} \left(1 - \prod_{l=1}^N (1 - \beta I_l^{x*})^{n_{l \rightarrow j}} \right), \end{aligned} \tag{9}$$

and $n_{j \rightarrow i} = \delta_{ij}(1 - p_d)n_i + p_d \frac{W_{ji}}{\sum_{j=1}^N W_{ji}} n_j$. Injecting the solution into the first equation after isolating R_i^{x*} in the second equation, we have:

$$\begin{aligned} (\gamma + \mu \sum_{y=1}^M A_{xy}) I_i^{x*} &= \left[1 - \sum_{y=1}^M (I_i^{y*} + \frac{\gamma I_i^{y*}}{\alpha + \sum_{l=1}^M T_{yl} \Pi_i^l}) \right] \Pi_i^x \\ &+ \sum_{y=1}^M A_{yx} I_i^{y*} \mu \\ &+ \sum_{y=1}^M \frac{\gamma I_i^{y*}}{\alpha + \sum_{l=1}^M T_{yl} \Pi_i^l} T_{xy} \Pi_i^x. \end{aligned} \tag{10}$$

Approaching the critical point, there are $I_i^{x*} \ll 1$ for $\forall i$. Ignoring the higher order terms of I_i^{x*} in the above expression and linearize:

$$\begin{aligned} \Pi_i &\simeq (1 - p_d) \sum_{j=1}^N \beta I_j^{x*} n_{j \rightarrow i} \\ &+ p_d \sum_{j=1}^N \frac{W_{ij}}{\sum_{l=1}^N W_{il}} \sum_{l=1}^N \beta I_l^{x*} n_{l \rightarrow j}. \end{aligned} \tag{11}$$

Substitute the expression of $n_{j \rightarrow i}$ into equation Eq. (11) and ignore the higher-order terms of I_i^{x*} :

$$\begin{aligned} \Pi_i &\simeq \beta \sum_{j=1}^N [(1 - p_d)^2 \delta_{ij} n_i + p_d(1 - p_d) n_j (F_{ij} + F_{ji})] \\ &+ p_d^2 \sum_{l=1}^N n_j F_{il} F_{jl} I_j^{x*}. \end{aligned} \tag{12}$$

Defining the row-stochastic matrix \mathbf{F} as

$$F_{ij} = \frac{W_{ij}}{\sum_{l=1}^N W_{il}}. \tag{13}$$

Then, substitute Eq. (12) into Eq. (10), and keeping up to first order in I_i^{x*} , we obtain:

$$\begin{aligned} (\gamma + \mu \sum_{y=1}^M A_{xy}) I_i^{x*} &= \beta \sum_{j=1}^N [(1 - p_d)^2 \delta_{ij} n_i + p_d(1 - p_d) n_j (F_{ij} + F_{ji})] \\ &+ p_d^2 \sum_{l=1}^N n_j F_{il} F_{jl} I_j^{x*} + \sum_{y=1}^M A_{yx} I_i^{y*} \mu. \end{aligned} \tag{14}$$

To simplify the solution, assume that $I_i^{x*} = I_i^{y*}$, where $x, y = 1, \dots, M$ and $x \neq y$. Therefore Eq. (14) can be rewritten as:

$$\begin{aligned} \gamma I_i^{x*} &= \beta \sum_{j=1}^N [(1 - p_d)^2 \delta_{ij} n_i + p_d(1 - p_d) n_j (F_{ij} + F_{ji})] \\ &+ p_d^2 \sum_{l=1}^N n_j F_{il} F_{jl} I_j^{x*} \\ &= \beta \sum_{j=1}^N M_{ij} I_j^{x*}. \end{aligned} \tag{15}$$

Therefore, Eq. (15) can be written as

$$\frac{\gamma}{\beta} I_i^{x*} = (\mathbf{M} I_i^{x*}), \tag{16}$$

where,

$$\begin{aligned} M_{ij} &= (1 - p_d)^2 \delta_{ij} n_j + p_d(1 - p_d) n_j (F + F^T)_{ij} \\ &+ p_d^2 n_j (F F^T)_{ij}. \end{aligned} \tag{17}$$

An epidemic outbreak only occurs when $\frac{\gamma}{\beta}$ is an eigenvalue of the matrix \mathbf{M} . Thus, the critical point is

$$\beta_c = \frac{\gamma}{\Lambda_{\max}(\mathbf{M})}, \tag{18}$$

where $\Lambda_{\max}(\mathbf{M})$ is the largest eigenvalue of matrix \mathbf{M} . And the critical point β_c is approximately equal to the reciprocal of the Basic Reproduction Number (R_0) [58].

5. Numerical validation

In this section, we present experimental results by iteratively computing the equations of Section 4 to investigate the impact of traffic and gene networks on pathogen diversity. To understand the effects of city location and population movement on mutant pathogen diversity in the real world, we refer to the population distribution data of the city of Cali, Colombia, for traffic network construction. The network data obtained from A. Arenas team and presented in Ref. [27]. The city has 2.4×10^6 inhabitants, who are formally divided into $N = 22$ districts according to their residential location. We construct the matrix \mathcal{W} based on the movement patterns between these $N = 22$ subpopulations. In the traffic network constructed from the population distribution data of Cali, except for cities 3 and 14 corresponding to Fig. 3, the other cities are not completely connected, that is, only individuals in cities 3 and 4 may move to any other cities, and individuals in other cities may also move to cities 3 and 14. In addition, paying attention to the connection weights between 22 cities, we can see that the sum of the connection weights pointing to cities 3 and 5 is the largest, indicating that individuals are more likely to move to these two cities. In addition, in order to learn as much as possible about how gene networks affect pathogen diversity, we worked to keep the gene networks themselves simple, using the well-known graphical toy model consisting of lattices, chains, and stars to construct the gene networks. All three gene networks contain $M = 25$ strains and define mutational pathways between them.

To fully and clearly explore the impact of the traffic and the gene network, the specific pathogen infections in each subpopulation are included in the study. We first observe the results of infection in different cities. As shown in Fig. 3, in the results based on either gene network, we find a higher density of infection in city 6, city 13, and city 14. Afterwards, we observe the infection of different strains in each city under three gene networks. When the gene network is star-type, we find that the proportion of arbitrary pathogens is almost uniform in the same city. This phenomenon is due to the fact that the central node of the star network balances the mutations of various pathogens very well. When looking at the lattice-type gene network, as in Fig. 3(e), we find that the infection is almost localized around strains a, e, f, j, k, p, u, y , i.e. the infection rate is relatively greater in these strains. These strains are just at the edge of the gene network. The phenomenon indicates that the gene network structure leads to outbreak localization on certain viruses to some extent. Also, we observe that pathogen localization is particularly evident in city 6, city 13, and city 14, where infection densities are high. Pathogen diversity is more pronounced in the remaining areas with infection. This suggests that the traffic network and genetic network together influence pathogens transmission and variation. With enhanced movement probabilities p_d , as shown in Fig. 4(g) and (h), the likelihood of individuals going outside increases, promoting the occurrence of pathogen diversity. At this point, infection rates between cities become close, and traffic networks play a smaller role. The gene

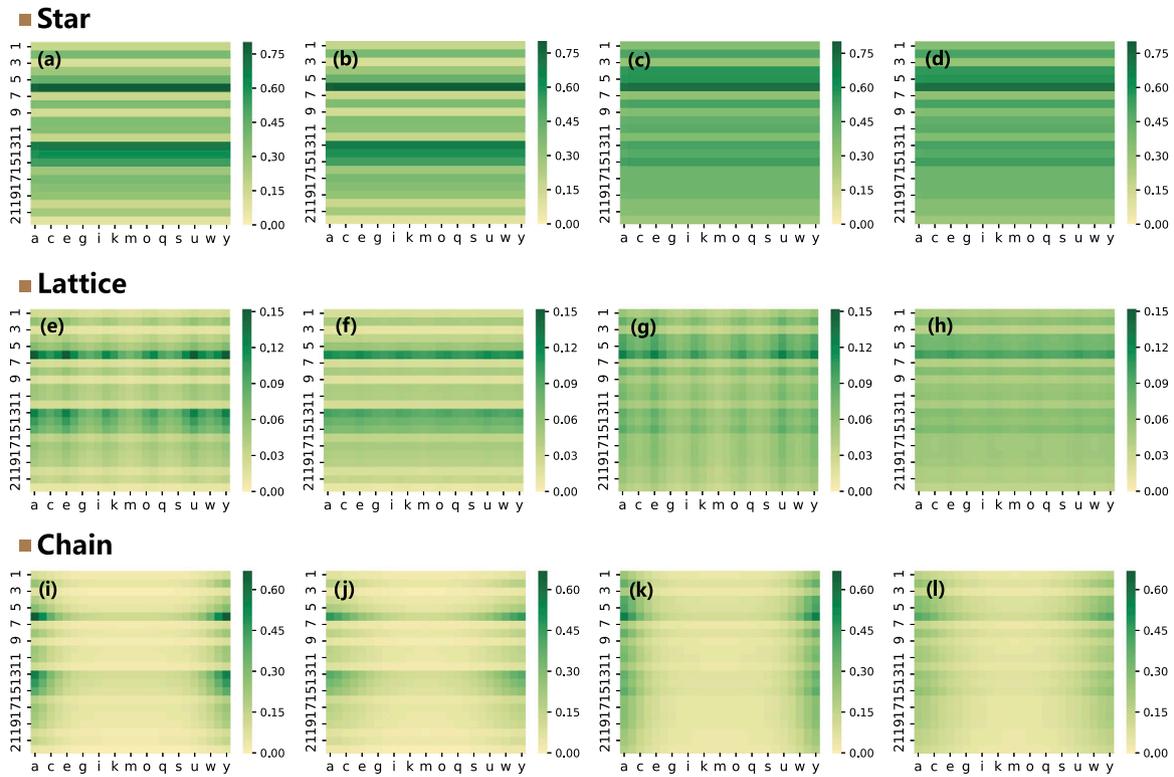


Fig. 3. Infection by different $M = 25$ strains in all subpopulations. The horizontal coordinates of the graph indicate the different pathogens and the vertical coordinates are all cities. The color represents the infection rate of a particular pathogen in a certain area. Figs (a)–(d), Figs (e)–(h), and Figs (i)–(l) show the results based on the star-type, lattice-type and chain-type gene networks, respectively. We set movement rate $p_d = 0.1$, mutation rate $\mu = 0.01$ in Figs (a)(e)(i); $p_d = 0.1$, $\mu = 0.05$ in Figs (b)(f)(j); $p_d = 0.4$, $\mu = 0.01$ in Figs (c)(g)(k); and $p_d = 0.4$, $\mu = 0.05$ in Figs (d)(h)(l). The remaining parameters are $\beta = 0.0001$, $\Delta = 10$, $\gamma = 0.08$, and $\alpha = 0.02$. For ease of observation, the values of infections on the star-type gene network (first row) are magnified by a factor of 100. The results on the lattice as well as on the chain gene network are magnified by a factor of 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

network still allows for a high infection density of endpoint strains a, e, f, j, k, p, u, y . In addition, the increases of mutation probability μ also promotes pathogen diversity, as shown in Fig. 3(f). In this case, the gene network effect is reduced and the infection rate tends to be consistent between strains. The traffic network still plays a role, allowing a high infection density in city 6, city 13, and city 14. Finally, when looking at the infection results on the chain-type gene network, as shown in Fig. 3(i), in the results for city 6, city 13, and city 14 we can clearly observe that the infection is localized on pathogen a and strain y . These two pathogens are at the both ends of the gene network. The phenomenon again shows that the gene network and the traffic network have a significant effect on the localization of pathogens.

Next, the effects of traffic network and gene network on pathogens diversity are further explored. For a more visual view, we first plot the infection results in the map for $N = 22$ cities. As shown in Figs. 4(a)–(c), we find that the infection density is higher in all regions when the gene network is based on chain-type. Obviously, city 6, city 13, and city 14 are always the regions with the highest infection density. By observing Figs. 4(d)–(f), we find that in the star-type gene networks, the area of various color sectors in each subpopulation is almost the same, which indicates that the infection rate of various pathogens does not differ much. In the lattice-type gene network, there are tiny differences in the area of every sector, and they are not very obvious. Pathogen diversity is present in all cities in the star-type gene network. And in the lattice-type gene network, it can be find that the proportion of pathogen a and strain y is significantly higher than that of other pathogens. This phenomenon suggests that lattice-type and chain-type gene networks are more likely to lead to pathogen localization (lower diversity) for the same population movements on the traffic network.

We introduce Effective Distance (ED) to discuss the reasons for the appearance of pathogen diversity. The effective distance from a node i

to a connected node j is defined as [59] :

$$d_{ij} = (1 - \log R_{ij}), \tag{19}$$

where $R_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$. In generally, $d_{ij} \neq d_{ji}$. We assume that there are $L = \{P_1, \dots, P_L\}$ paths between arbitrary two nodes i and j in the network. The effective length $N_l, l \in [1, L]$ of path $P_l, l \in [1, L]$ is the sum of effective distances along the legs of the path. Mathematically, $N_l = d_{ik} + d_{kj}$, when path P_l covers legs $i \rightarrow k$ and $k \rightarrow j$. For the paths of two nodes i and j , $\Gamma = \{N_1, \dots, N_L\}$ is the set of effective length. Then the effective distance D_{ij} from an arbitrary node i to another node j in the network is defined as:

$$D_{ij} = \min(\Gamma). \tag{20}$$

So, in our model, the effective distance ED_i for node i is defined as:

$$ED_i = \sum_j^N D_{ji}, i = [1, N]. \tag{21}$$

This indicator is a measure of the importance of a city in the traffic network to a certain extent. That is, a smaller ED_i indicates a greater tendency for the population to move to city i .

The relationship between the effective distance (ED) and the infection density (I^*) of each city are discussed. As shown in Fig. 5, the relationship between infection density and effective distance can be fitted by a straight line with negative slope. This result indicates that the greater the effective distance, the lower the infection density of the city. Specifically, the long effective distance of the city makes the inflowing population relatively small, which reduces the risk of population infection to a certain extent, and eventually leads to a low infection density. However, most individuals flow into cities with short effective distances, which increases the possibility of inflow of infected

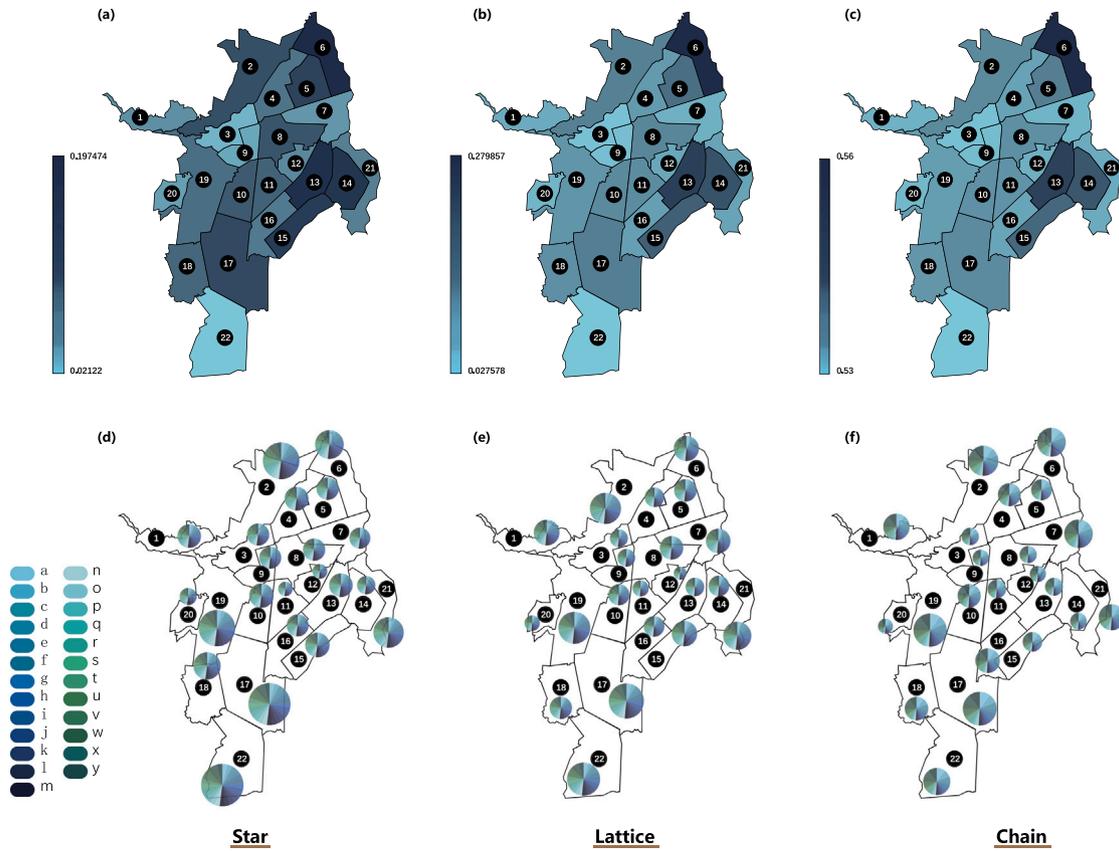


Fig. 4. Infection of 22 cities in Cali. Figs (a)–(c) show the results of infection in every city when the gene network is based on star, lattice, and chain, respectively. The colors in the graph represent the prevalence of infection in a particular region. Figs (d)–(f) show the percentage of different pathogens infections in each city when the gene network is star, lattice, and chain, respectively. The different colors in the pie chart represent different pathogens. The parameters of the experiment are $\beta = 0.001$, $p_d = 0.1$, $\mu = 0.01$, $\Delta = 10$, $\gamma = 0.08$, and $\alpha = 0.02$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

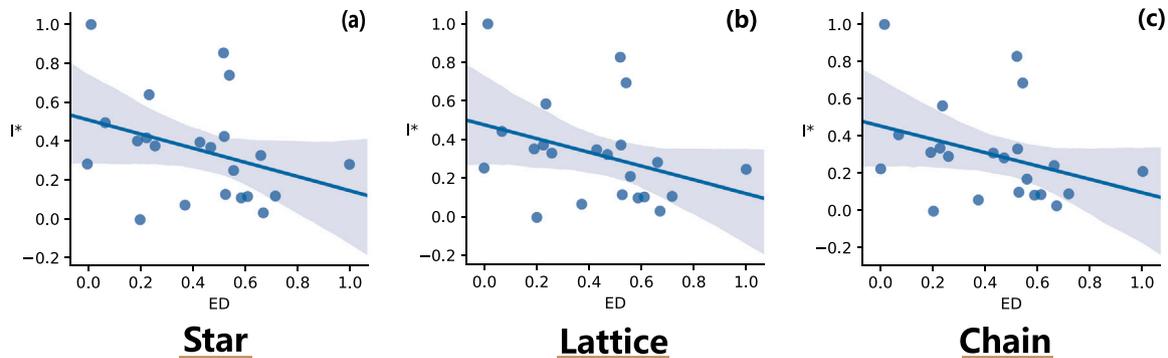


Fig. 5. The correlation between I^* and ED for the 22 cities. (a)–(c) show the results on the star-type, lattice-type and chain-type gene networks, respectively. The data in the figure are normalized for ease of side observation. Each small blue dot in the diagram represents a city. The blue line is the linear fit of the scattered points. We set the confidence interval to 95%, as shown in the blue shaded interview. The experiment parameters are consistent with Fig. 4.

individuals, resulting in higher infection density in cities with short effective distances. So, the greater infection density in cities 6, 13, 14 in Fig. 4 can be explained by their short effective distance in the traffic network. The above phenomenon is found in all three different gene networks.

The pathogen localization phenomenon, i.e. lack of pathogen diversity, is also discussed in depth. We calculate the EPR for 22 subpopulations and the results are displayed in Fig. 6. We find that on the star-type network, there are $EPR \rightarrow 1$ in all cities. The phenomenon verifies the conclusion in Fig. 3, indicating that star-type gene network lead to pathogen diversity without infection localization. EPR values

are smaller in cities 6,13,14,15 when the gene network is lattice-type and chain-type, suggesting the presence of pathogen localization, i.e., lack of pathogen diversity, in these cities.

The effective distance is also used to further explore the reason why the pathogen localization is only evident in certain cities. The correlation between Effective Participation Ratio (EPR) and effective distance (ED) for the 22 cities is plotted in Fig. 7. It can be found that the EPR values of pathogens in cities are essentially distributed in a positive correlation with the effective distance, indicating that the smaller the effective distance from the city, the greater the probability of pathogen localization happening, i.e., no pathogen diversity. This phenomenon is due to the fact that the population is more likely to move to a city with a short effective distance, which makes the

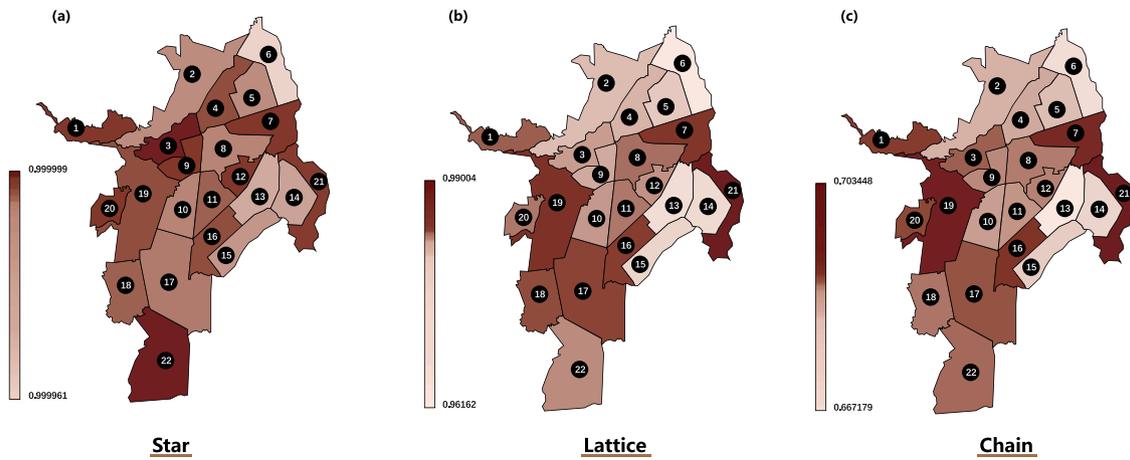


Fig. 6. The EPR of 22 cities. (a)–(c) show the calculated results of EPR in each city when the gene network is star-type, lattice-type, and chain-type, respectively. The color in the graph is the EPR value for a region, representing the degree of localization of the infection. The parameters in this experiment are fixed as $\beta = 0.001$, $p_d = 0.1$, $\mu = 0.01$, $\Delta = 10$, $\gamma = 0.08$, and $\alpha = 0.02$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

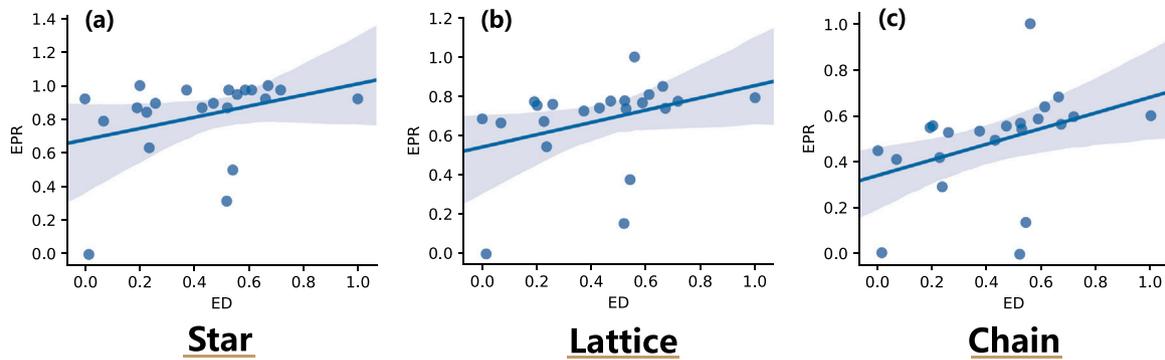


Fig. 7. The correlation between EPR and ED for the 22 cities. The effective participation ratio is positively correlated with the effective distance. The legend meanings and experimental parameters are consistent with Fig. 6.

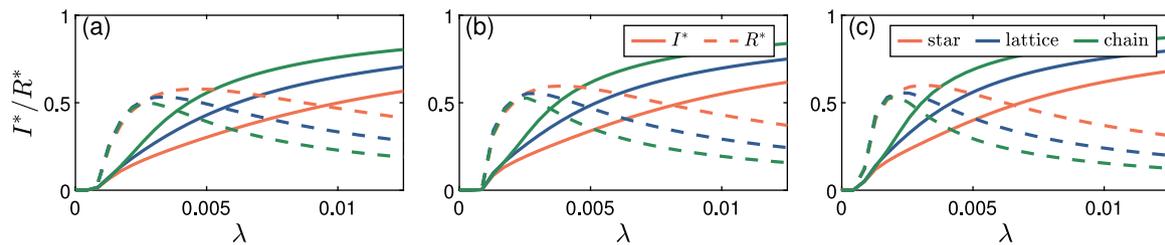


Fig. 8. Theoretical calculation results for I^* and R^* on the three gene networks. We fix the recovery rates $\gamma = 0.08$, the mutation rate $\mu = 0.01$, waning immunity rate $\alpha = 0.02$, transcending immunity $\Delta = 10$ and vary the infection rate λ under three values of movement rates: $p_d = 0.1$ (Left), $p_d = 0.4$ (Center), and $p_d = 0.7$ (Right). In this experiment, the value of β is scaled by the equation $\lambda = \beta/\mu$.

population in the city more likely to be infected, the infection density is high, and the pathogen localization is more obvious. However, few individuals flow to cities with long effective distances, so that few individuals are infected by any strain in these cities, resulting in less obvious localization of pathogens.

This phenomenon is due to the fact that the population prefers to move to cities with short effective distances, which leads to a high density of infection and more pronounced pathogen localization in that city. When the gene network is chain-type, this phenomenon is particularly evident.

We finally wonder the impact of pathogen mutation profiles under different genetic networks on the infection of the whole system. The final fraction of individuals in the infected state I^* and the recovered state R^* versus the infection rate λ under three different gene networks are shown in Fig. 8. An important feature of the meta-population

model is population mobility, which can have a significant impact on epidemic transmission. Fig. 8 also simultaneously explores infections and recovery at different rates of population movement p_d at the same time. We find that the infection range of gene network based on chain is wider than that of gene network based on lattice. And the infection range is smallest when gene network based on star structure. This phenomenon demonstrates the influence of genetic networks on the outcome of infection. The percentage of recovery state population is just the opposite of the above phenomenon. We detect a maximum in R^* (solid line) around the inflection point of every I^* curve (dashed line). Then as the infection rate λ increases, the fraction of R^* decreases. This is the result of a combination of infection rates λ and transcending immunity Δ . When λ increases to a sufficiently large value, it counteracts the immunity brought by Δ , causing the recovery state individuals to shift to the infection state. When focusing

on the effect of the movement probability p_d on the infection, we find a significant increase in the proportion of I^* as p_d increased. This suggests that population movements can cause an increase in infections.

6. Conclusion

To explore the factors of pathogen diversity, in that work, we introduce pathogenic mutations in the meta-population model. Traffic network is used to model the population movement in the meta-population. And the potential mutation pathways of pathogens are defined by the gene network. To explore the effect of traffic network and gene network on the pathogen diversity under strains mutation scenarios, a Markov chain theoretical framework based on the SIRS infection model is proposed. Real population distribution data of Cali city are used to construct the traffic network, and construct three toy gene networks—the star, lattice, chain.

The impact of traffic network on pathogen diversity is discussed first. We find that pathogen localization is more likely to occur in cities with higher infection densities. In contrast, pathogen diversity is present in cities with lower infection densities. The underlying structure of the gene network also has an impact on pathogen diversity. We find that compared with star-type gene network, lattice-type and chain-type gene network are more likely to cause pathogen localization (i.e., lack of diversity). The mutational pathways defined by the gene network give a high probability that the localization at the endpoint position.

Both the increase of individual movement probability and strain mutation rate are found to contribute to pathogen diversity. As the probability of movement increases, the role of traffic network decreases and the infection density tends to be consistent between cities. The gene network still causes slightly greater infection densities in endpoint strains. An increase in mutation probability leads to a decrease in the role of gene networks for localization. That is, the infection rate becomes the same across strains and there is no pathogen localization.

Effective distance is also introduced to measure the relationship between localization and population mobility. We find that effective distance is negatively associated with infection density. A short effective distance makes the population tend to move towards the city, which results in the higher infection density in the city 6, city 13, and city 14. In addition, the positive correlation between effective participation ratio and effective distance suggests that cities with short effective distances are more likely to have pathogen localization.

Overall, our study provides some theoretical basis for exploring the effects of population movement and variation on pathogen diversity. In the future, this research could be extended to analyze the transmission evolution of real-world viruses with mutations, such as COVID-19. Considering the fact that SARS-CoV-2 mutation and population travel make infectious diseases prevalent globally, based on the existing studies with mutations, this study includes the impact of population movement on epidemic transmission into the discussion, and further completes the epidemic spreading model. Our results may provide some insight into proposing effective epidemic contain strategies. On the other hand, unlike the classical setting for studying mutation and evolution, our proposed model assumes equal infection rate and recovery rate for each mutant strain, which weakens the difference in infection between strains to a certain extent, resulting in a distance from the real result of transmission in reality. Also, the model we propose is a theoretical model, lacking real data support, and a more realistic data-driven epidemic model can be constructed in the future.

CRedit authorship contribution statement

Yanyi Nie: Devised the research project, Numerical simulations, Analyzed the results, Writing – original draft. **Xiaoni Zhong:** Analyzed the results. **Tao Lin:** Analyzed the results, Writing – original draft. **Wei Wang:** Devised the research project, Numerical simulations, Analyzed the results, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was partially supported by the Social Science Foundation of Chongqing, No. 2021PY53, Natural Science Foundation of Chongqing, China, No. cstc2021jcyj-msxmX0132, Natural Science Foundation of Yuzhong District, Chongqing, No. 20210117, National Natural Science Foundation of China under Grants No. 61903266, and Science and Technology Research Program of Chongqing Municipal Education Commission, No. KJQN202200429.

References

- [1] de Jong JC, Smith DJ, Lapedes AS, Donatelli I, Campitelli L, Barigazzi G, Van Reeth K, Jones TC, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. Antigenic and genetic evolution of swine influenza A (H3N2) viruses in Europe. *J Virol* 2007;81(8):4315–22. <http://dx.doi.org/10.1128/JVI.02458-06>.
- [2] Tsatsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in Chikungunya Virus affects vector specificity and epidemic potential. *PLoS Pathog* 2007;3(12):e201. <http://dx.doi.org/10.1371/journal.ppat.0030201>.
- [3] Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, Weinert LA, Wang Z, Guo Z, Xu L, Zhang Y, Zheng H, Qin N, Xiao X, Wu M, Wang X, Zhou D, Qi Z, Du Z, Wu H, Yang X, Cao H, Wang H, Wang J, Yao S, Rakin A, Li Y, Falush D, Balloux F, Achtman M, Song Y, Wang J, Yang R. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci* 2013;110(2):577–82. <http://dx.doi.org/10.1073/pnas.1205750110>.
- [4] Koelle K, Martin MA, Antia R, Lopman B, Dean NE. The changing epidemiology of SARS-CoV-2. *Science* 2022;375(6585):1116–21. <http://dx.doi.org/10.1126/science.abm4915>.
- [5] Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *J Virol* 2010;84(19):9733–48. <http://dx.doi.org/10.1128/JVI.00694-10>.
- [6] Lyons DM, Lauring AS. Mutation and Epistasis in Influenza Virus Evolution. *Viruses* 2018;10(8):407. <http://dx.doi.org/10.3390/v10080407>.
- [7] Shao W, Li X, Goraya MU, Wang S, Chen J-L. Evolution of Influenza A Virus by mutation and re-assortment. *Int J Mol Sci* 2017;18(8):1650. <http://dx.doi.org/10.3390/ijms18081650>.
- [8] Shan C, Xia H, Haller SL, Azar SR, Liu Y, Liu J, Muruato AE, Chen R, Rossi SL, Wakamiya M, Vasilakis N, Pei R, Fontes-Garfias CR, Singh SK, Xie X, Weaver SC, Shi P-Y. A Zika virus envelope mutation preceding the 2015 epidemic enhances virulence and fitness for transmission. *Proc Natl Acad Sci* 2020;117(33):20190–7. <http://dx.doi.org/10.1073/pnas.2005722117>.
- [9] Kucharski AJ, Andreasen V, Gog JR. Capturing the dynamics of pathogens with many strains. *J Math Biol* 2016;72(1–2):1–24. <http://dx.doi.org/10.1007/s00285-015-0873-4>.
- [10] Gilchuk I, Gilchuk P, Sapparapu G, Lampley R, Singh V, Kose N, Blum DL, Hughes LJ, Satheskumar PS, Townsend MB, Kondas AV, Reed Z, Weiner Z, Olson VA, Hammarlund E, Raue H-P, Slifka MK, Slaughter JC, Graham BS, Edwards KM, Eisenberg RJ, Cohen GH, Joyce S, Crowe JE. Cross-neutralizing and protective human antibody specificities to Poxvirus Infections. *Cell* 2016;167(3):684–94.e9. <http://dx.doi.org/10.1016/j.cell.2016.09.049>.
- [11] Peeters B, Reemers S, Dortmans J, de Vries E, de Jong M, van de Zande S, Rottier PJ, de Haan CA. Genetic versus antigenic differences among highly pathogenic H5N1 avian influenza A viruses: consequences for vaccine strain selection. *Virology* 2017;503:83–93. <http://dx.doi.org/10.1016/j.virol.2017.01.012>.
- [12] Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordon A, Balmaseda A, Harris E. Antibody-dependent enhancement of severe dengue disease in humans. *Science* 2017;358(6365):929–32. <http://dx.doi.org/10.1126/science.aan6836>.
- [13] Halstead SB. Neutralization and antibody-dependent enhancement of Dengue Viruses. In: *Advances in virus research*, Vol. 60. Elsevier; 2003, p. 421–67.
- [14] Ferguson N, Andreasen V. The influence of different forms of cross-protective immunity on the population dynamics of antigenically diverse pathogens. In: Castillo-Chavez C, Blower S, van den Driessche P, Kirschner D, Yakubu A-A, editors. *Mathematical approaches for emerging and reemerging infectious diseases: models, methods, and theory, the ima volumes in mathematics and its applications*. New York, NY: Springer; 2002, p. 157–69. http://dx.doi.org/10.1007/978-1-4613-0065-6_9.

- [15] Kucharski AJ, Gog JR. Age profile of immunity to influenza: Effect of original antigenic sin. *Theor Popul Biol* 2012;81(2):102–12. <http://dx.doi.org/10.1016/j.tpb.2011.12.006>.
- [16] Wagner A. A genotype network reveals homoplastic cycles of convergent evolution in influenza A (H3N2) haemagglutinin. *Proc R Soc B* 2014;281(1786):20132763. <http://dx.doi.org/10.1098/rspb.2013.2763>.
- [17] Uekermann F, Sneppen K. Spreading of multiple epidemics with cross immunization. *Phys Rev E* 2012;86(3):036108. <http://dx.doi.org/10.1103/PhysRevE.86.036108>.
- [18] Girvan M, Callaway DS, Newman MEJ, Strogatz SH. Simple model of epidemics with pathogen mutation. *Phys Rev E* 2002;65(3):031915. <http://dx.doi.org/10.1103/PhysRevE.65.031915>.
- [19] Li H-J, Xu W, Song S, Wang W-X, Perc M. The dynamics of epidemic spreading on signed networks. *Chaos Solitons Fractals* 2021;151:111294. <http://dx.doi.org/10.1016/j.chaos.2021.111294>.
- [20] Williams BJM, St-Onge G, Hébert-Dufresne L. Localization, epidemic transitions, and unpredictability of multistrain epidemics with an underlying genotype network. *PLoS Comput Biol* 2021;17(2):e1008606. <http://dx.doi.org/10.1371/journal.pcbi.1008606>.
- [21] Schwarzendahl FJ, Grauer J, Liebchen B, Löwen H. Mutation induced infection waves in diseases like COVID-19. *Tech. rep.*, 2021.
- [22] Zhang X, Zhongyuan R, Zheng M, Zhou J, Stefano B, Barzel B. Epidemic spreading under pathogen evolution. 2021, Preprint, In Review.
- [23] Feng L, Zhao Q, Zhou C. Epidemic spreading in heterogeneous networks with recurrent mobility patterns. *Phys Rev E* 2020;102(2):022306. <http://dx.doi.org/10.1103/PhysRevE.102.022306>.
- [24] Dalziel BD, Pourbohloul B, Ellner SP. Human mobility patterns predict divergent epidemic dynamics among cities. *Proc R Soc B* 2013;280(1766):20130763. <http://dx.doi.org/10.1098/rspb.2013.0763>.
- [25] Li W-J, Chen Z, Jin K-Z, Li L, Yuan L, Jiang L-L, Perc M, Kurths J. Eliminating poverty through social mobility promotes cooperation in social dilemmas. *Chaos Solitons Fractals* 2022;156:111845. <http://dx.doi.org/10.1016/j.chaos.2022.111845>.
- [26] Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. *Rev Modern Phys* 2015;87(3):925–79. <http://dx.doi.org/10.1103/RevModPhys.87.925>.
- [27] Gómez-Gardeñes J, Soriano-Paños D, Arenas A. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nat Phys* 2018;14(4):391–5. <http://dx.doi.org/10.1038/s41567-017-0022-7>.
- [28] Colizza V, Vespignani A. Invasion threshold in heterogeneous metapopulation networks. *Phys Rev Lett* 2007;99(14):148701. <http://dx.doi.org/10.1103/PhysRevLett.99.148701>.
- [29] Wang Z, Moreno Y, Boccaletti S, Perc M. Vaccination and epidemics in networked populations—an introduction. *Chaos Solitons Fractals* 2017;103:177–83. <http://dx.doi.org/10.1016/j.chaos.2017.06.004>.
- [30] Newman MEJ, Ferrario CR. Interacting epidemics and coinfection on contact networks. *PLOS ONE* 2013;8(8):e71321. <http://dx.doi.org/10.1371/journal.pone.0071321>.
- [31] Nie Y, Li W, Pan L, Lin T, Wang W. Markovian approach to tackle competing pathogens in simplicial complex. *Appl Math Comput* 2022;417:126773. <http://dx.doi.org/10.1016/j.amc.2021.126773>.
- [32] Wang W, Li W, Lin T, Wu T, Pan L, Liu Y. Generalized k-core percolation on higher-order dependent networks. *Appl Math Comput* 2022;420:126793. <http://dx.doi.org/10.1016/j.amc.2021.126793>.
- [33] Nie Y, Zhong X, Wu T, Liu Y, Lin T, Wang W. Effects of network temporality on coevolution spread epidemics in higher-order network. *J King Saud Univ - Comput Inf Sci* 2022;34(6, Part A):2871–82. <http://dx.doi.org/10.1016/j.jksuci.2022.04.004>.
- [34] Colizza V, Pastor-Satorras R, Vespignani A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat Phys* 2007;3(4):276–82. <http://dx.doi.org/10.1038/nphys560>.
- [35] Mata AS, Ferreira SC, Pastor-Satorras R. Effects of local population structure in a reaction–diffusion model of a contact process on metapopulation networks. *Phys Rev E* 2013;88(4):042820. <http://dx.doi.org/10.1103/PhysRevE.88.042820>.
- [36] Balcan D, Vespignani A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat Phys* 2011;7(7):581–6. <http://dx.doi.org/10.1038/nphys1944>.
- [37] Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini Jr. IM, Merler S, Pentland A, Vespignani A, Moro E, Moreno Y. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav* 2020;4(9):964–71. <http://dx.doi.org/10.1038/s41562-020-0931-9>.
- [38] Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, Soriano-Paños D, Steinegger B. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys Rev X* 2020;10(4):041055. <http://dx.doi.org/10.1103/PhysRevX.10.041055>.
- [39] Giangaspero M, Decaro N, Turno P, Apicella C, Gargano P, Buonavoglia C. Pathogen spread and globalization: The case of Pestivirus heterogeneity in southern Italy. *Res Vet Sci* 2019;125:100–12. <http://dx.doi.org/10.1016/j.rvsc.2019.05.017>.
- [40] Magalhães LMD, Gollob KJ, Zingales B, Dutra WO. Pathogen diversity, immunity, and the fate of infections: Lessons learned from *Trypanosoma cruzi* human–host interactions. *Lancet Microbe* [http://dx.doi.org/10.1016/S2666-5247\(21\)00265-2](http://dx.doi.org/10.1016/S2666-5247(21)00265-2).
- [41] Gupta S, Maiden MCJ. Exploring the evolution of diversity in pathogen populations. *TIM* 2001;9(4):181–5. [http://dx.doi.org/10.1016/S0966-842X\(01\)01986-2](http://dx.doi.org/10.1016/S0966-842X(01)01986-2).
- [42] Buckee CO, Koelle K, Mustard MJ, Gupta S. The effects of host contact network structure on pathogen diversity and strain structure. *Proc Natl Acad Sci* 2004;101(29):10839–44. <http://dx.doi.org/10.1073/pnas.0402000101>.
- [43] Buckee C, Danon L, Gupta S. Host community structure and the maintenance of pathogen diversity. *Proc R Soc B* <http://dx.doi.org/10.1098/rspb.2007.0415>.
- [44] Gomes MGM, Medley GF, Nokes DJ. On the determinants of population structure in antigenically diverse pathogens. *Proc R Soc B* 2002;269(1488):227–33. <http://dx.doi.org/10.1098/rspb.2001.1869>.
- [45] Cobey S, Lipsitch M. Pathogen diversity and hidden regimes of apparent competition. *Amer Nat* 2013;181(1):12–24. <http://dx.doi.org/10.1086/668598>.
- [46] Abu-Raddad LJ, Ferguson NM, cross immunity Theimpactof. Mutation and stochastic extinction on pathogen diversity. *Proc R Soc B* 2004;271(1556):2431–8. <http://dx.doi.org/10.1098/rspb.2004.2877>.
- [47] Gupta S, Ferguson N, Anderson R. Chaos, Persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 1998;280(5365):912–5. <http://dx.doi.org/10.1126/science.280.5365.912>.
- [48] Sridhar A, Yağan O, Eletreby R, Levin SA, Plotkin JB, Poor HV. Leveraging a multiple-strain model with mutations in analyzing the spread of Covid-19. In: ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2021, p. 8163–7. <http://dx.doi.org/10.1109/ICASSP39728.2021.9414595>.
- [49] Yagan O, Sridhar A, Eletreby R, Levin S, Plotkin JB, Poor HV. Modeling and analysis of the spread of COVID-19 under a multiple-strain model with mutations. *Harvard Data Sci Rev (Special Issue 1)* <http://dx.doi.org/10.1162/99608f92.a11bf693>.
- [50] Nextstrain, <https://nextstrain.org/ncov/gisaid/global/all-time>.
- [51] Gisaid, <https://www.gisaid.org/>.
- [52] Lobinska G, Pautzner A, Traulsen A, Pilpel Y, Nowak MA. Evolution of resistance to COVID-19 vaccination with dynamic social distancing. *Nat Hum Behav* 2022;6(2):193–206. <http://dx.doi.org/10.1038/s41562-021-01281-8>.
- [53] Zhang X, Lobinska G, Feldman N, Dekel E, Nowak MA, Pilpel Y, Pautzner Y, Barzel B, Pautzner A. A spatial vaccination strategy to reduce the risk of vaccine-resistant variants. *PLoS Comput Biol* 2022;18(8):e1010391. <http://dx.doi.org/10.1371/journal.pcbi.1010391>.
- [54] Lieberman E, Hauert C, Nowak MA. Evolutionary dynamics on graphs. *Nature* 2005;433(7023):312–6. <http://dx.doi.org/10.1038/nature03204>.
- [55] Belik V, Geisel T, Brockmann D. Natural human mobility patterns and spatial spread of infectious diseases. *Phys Rev X* 2011;1(1):011001. <http://dx.doi.org/10.1103/PhysRevX.1.011001>.
- [56] Gómez S, Arenas A, Borge-Holthoefer J, Meloni S, Moreno Y. Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *Europhys Lett* 2010;89(3):38009. <http://dx.doi.org/10.1209/0295-5075/89/38009>.
- [57] Liu C, Zhou N, Zhan X-X, Sun G-Q, Zhang Z-K. Markov-based solution for information diffusion on adaptive social networks. *Appl Math Comput* 2020;380:125286. <http://dx.doi.org/10.1016/j.amc.2020.125286>.
- [58] Hiraoka T, Rizi AK, Kivelä M, Saramäki J. Herd immunity and epidemic size in networks with vaccination homophily. *Phys Rev E* 2022;105(5):L052301. <http://dx.doi.org/10.1103/PhysRevE.105.L052301>.
- [59] Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013;342(6164):1337–42. <http://dx.doi.org/10.1126/science.1245200>.