REGULAR PAPER

Guojun Li · Baofeng Chang · Jian Zhao · Jiayang Wang · Fan He · Yongheng Wang · Ting Xu · Zhiguang Zhou⬤

VIVIAN: virtual simulation and visual analysis of epidemic spread data

Abstract Visual analysis of epidemic spread data is crucial in understanding the process of epidemic transmission, tracing the source of infection, evaluating the development of epidemic, and formulating reasonable policies. However, due to limited capabilities in collecting comprehensive data on the spread of the epidemic, it remains challenging to fully and visually comprehend the spatial and temporal changes in virus transmission, which, in turn, hampers efforts in exploration of macro-level pattern analysis and validation of micro-level facts. To fill this gap, in this paper, we propose a virtual simulation and visual analysis system, named VIVIAN. The user-friendly interaction design of the system enables rapid infection traceability and accurate investigation of close contacts. Furthermore, it establishes synchronous correlations between model parameters and epidemic prevention and control measures. The system supports users in simulating the formulation of epidemic prevention and control policies and evaluating their effectiveness. The system automatically generates the movement trajectories and contact situation based user-defined thresholds, which relaxes the data challenge and scalability of the system. The system is equipped with

G. Li · Z. Zhou (✉)
Intelligent Big Data Visualization Lab, Hangzhou Dianzi University, Hangzhou 310000, China
E-mail: zhgzhou@hdu.edu.cn

G. Li
E-mail: lgj_afeng@163.com

G. Li · J. Zhao · J. Wang
The School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou 310000, China
E-mail: zhaojian@zufe.edu.cn

J. Wang
E-mail: wjy4141@163.com

B. Chang
Zhejiang Airport Innovation Institute, Hangzhou 310000, China
E-mail: baofeng.chang@foxmail.com

F. He
Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou 310000, China
E-mail: fhe@cdc.zj.cn

Y. Wang
Research Center for Data Mining and Knowledge Discovery, Zhejiang Lab, Hangzhou 310000, China
E-mail: wangyh@zhejianglab.com

T. Xu
Department of Stomatology, First Affiliated Hospital, Zhejiang University, Hangzhou 310000, China
E-mail: xut@zju.edu.cn

multiple linked, intuitive, and interactive visualization charts for rapid infection traceability and accurate close contact investigation. In addition, the system supports users in formulating and evaluating of epidemic prevention and control policies. Case studies and expert interviews based on simulated data have demonstrated the effectiveness and practicality of the system, which make it, as a foundation, possible to be employed in certain scenes for epidemic prevention and controls.

Keywords Epidemic prevention and control · Visual analysis · Decision simulation

1 Introduction

Epidemic, caused by pathogenic microorganisms like severe acute respiratory syndrome (SARS) virus and COVID-19 virus, etc., can be transmitted between humans, animals, or a combination of both (Fang et al. 2009). The onset and transmission of such diseases are often abrupt and concealed, which allows for a possible rapid spread of the epidemic in its early stages. Especially for sudden and emerging epidemics, public understanding of the virus may be limited, leading to a low level of early epidemic prevention and control awareness and effort (Naseem et al. 2021). Furthermore, limitations in epidemic prevention and control policies in different regions can result in recurrent outbreaks of small-scale epidemics, posing a severe threat to public health, safety, and economic development. Examples of epidemics that have demonstrated rapid onset, transmission, and high fatality rates include the SARS virus, which emerged in 2003, and the COVID-19 virus, which emerged in 2019. Most infected patients have direct or indirect contact with the virus or reside in epidemic areas.

In recent years, especially under the background of the COVID-19 virus, there has been a growing focus on effectively preventing and controlling the spread of the epidemic, which has led to numerous studies conducted globally. For instance, some studies have investigated the impact of various measures, such as social distancing control and crowd movement restriction, on epidemic spread from the perspective of population movement trajectories (Silva et al. 2020; Lai et al. 2020). In the field of visualization, Chen et al. (2021) studied the spread pattern and transmission process of the novel coronavirus based on the distance attenuation law and spatial spread of epidemics. However, current epidemic analysis methods are often carried out from a macro-perspective (a fuzzy estimate of the trend of various epidemics in a large area) in a statistical manner, which makes it difficult to obtain precise contact data and may limit the ability to detect, understand, and control epidemics at a specific level (exploration of the epidemic spread characteristics in a small area) instead of a comprehensive and generalized analysis of the epidemic.

After extensive interviews and discussions with experts who are knowledgeable in the field of epidemic prevention and control, we have learned that the visualization technique is significant in many aspects such as mastering the specific dynamic transmission process of the epidemic, tracing the source of infection, identifying corresponding close contacts, and ultimately formulating and implementing effective epidemic prevention and control policies. However, there are three primary challenges in utilizing visualization technology for the analysis of outbreaks and emerging epidemics: Firstly, there are technical limitations in acquiring comprehensive and accurate spatial contact data of individuals, primarily due to factors such as constraints in crowd movement trajectory sensors (C1). Consequently, it becomes extremely challenging to collect and acquire such data. To address this issue, effective simulation methods are necessary to establish a reliable data foundation. Secondly, traditional epidemic analysis methods often focus on the macro-perspective, overlooking the temporal changes in population trajectories and spatial contact information (C2). It is essential for epidemic analysis to incorporate the abstract representation of rapid infection source tracing and precise identification of close contacts in an intuitive and visual manner. Thirdly, during the peak of an epidemic, multiple epidemic prevention and control measures are proposed. However, the lack of prevention and control simulation and evaluation systems makes it difficult for authorities to make prompt and accurate decisions in response to the sudden spread of the disease (C3).

To address the challenges, we propose VIVIAN, a virtual simulation and visual analysis system of epidemic spread data. First, considering the difficulty in acquiring real epidemic data, multiple virtual scenes are supported to be set up to meet the needs of users and capture the flow of people and spatial contact based on real-world scenes and crowd habits. Additionally, the system integrates the Susceptible–Exposed–Infected–Removed transmission dynamics (SEIR) model (He et al. 2020) to control the transmission process of epidemics dynamically and provide the necessary data for epidemic transmission simulation (C1). Second, the system employs visualization methods to present simulated human contact and virus

transmission events to enable experts to view the temporal and spatial situation of epidemic transmission. This approach facilitates tracing, verification, and decision-making analysis for epidemic prevention and control efforts, enabling experts to focus on key individuals in the virus transmission chain and effectively contain the spread of the epidemic (C2). Third, the system establishes a mapping correlation between SEIR model parameters and various epidemic prevention measures to simulate the impact of different preventive measures in different situations. By comparing the outcomes of preventive measures, the system evaluates and predicts the effectiveness of prevention and control efforts, providing managers with reference information to make informed decisions regarding epidemic prevention (C3).

The major contributions of this paper are summarized as follows:

- We propose a virtual simulation experiment, integrating a popular epidemic model, to allow users to generate epidemic data with user-defined requirements for subsequent epidemic analysis.
- We design a visual analysis prototype system integrating multiple interactive visualization charts to support users to analyze the epidemic data for data insights and decision-making.
- We present a complete and usable workflow for the comprehensive analysis and evaluation of most simulated epidemic data by integrating the population movement and epidemic parameters.

2 Related work

2.1 Epidemic data analysis

Effective control of epidemic depends on accurate analysis of disease surveillance data. This involves data collection, mining, and analysis of hidden information (Wang et al. 2023; Andrienko et al. 2021), which can infer the cause of epidemic deaths (Polonsky et al. 2019), establish monitoring systems (El-Khatib et al. 2019), trace the spread of disease, and adjust management strategies (Pak and Kasarskis 2015). Such data-driven analyses provide theoretical guidance to decision-makers for developing effective and timely prevention and control measures. Historically, experts and scholars at home and abroad have made significant contributions to epidemic prevention and control through data-based interventions. For example, A Corberán-Vallet and Lawson (2014) utilized real-time data to build a Bayesian hierarchical Poisson model for prospective analysis of epidemic data, which helped to predict and implement effective interventions. Pak and Kasarskis (2015) created a quantitative prediction model for epidemics using comprehensive clinical and immunoassay data sets, leading to accurate infection control and antimicrobial interventions. Merler et al. (2015) used the Markov chain Monte Carlo method to estimate the transmission parameters of the Ebola virus and designed a simulation model to explore the effectiveness of intervention measures based on the geographical and demographic data of Liberia. Drug intervention is the most direct and effective means of epidemic prevention and control. However, there is still a need for extensive research on drug coverage and efficacy (Mukandavire et al. 2020; Bartsch et al. 2020). In the early stages of a new epidemic outbreak, effective anti-epidemic drugs are not yet available, and non-drug interventions such as expanding the physical distance between people and restricting the flow of people become necessary to slow down or stop the spread of the epidemic (Silva et al. 2020; Tian et al. 2020; Chinazzi et al. 2020). Experts and scholars rely on real datasets to predict the development trends and inflection points of the epidemic using modeling and simulation methods (Ren et al. 2020; Kumar et al. 2021). The influence of factors such as weather and population density on the spread of the disease can be analyzed based on specific data to infer effective prevention and control measures and simulate the intervention effects (Wang et al. 2020; Rocklöv and Sjödin 2020). Finally, data visualization technologies can be used to evaluate the effectiveness of prevention and control measures (Sullivan et al. 2020), providing decision-makers with scientific and rational plans for protection, isolation, and control. However, during the early stages of a new epidemic outbreak, contact data is often scarce, which hinders effective epidemic analysis. Therefore, this paper utilizes virtual simulation technology to simulate contact data for the population and further analyzes epidemic transmission data based on the simulated data.

2.2 Epidemic models

An epidemic model is a mathematical tool used to investigate various aspects of disease transmission such as spread rate, spatial range, transmission pathways, and dynamic mechanisms. Dynamic models applied in
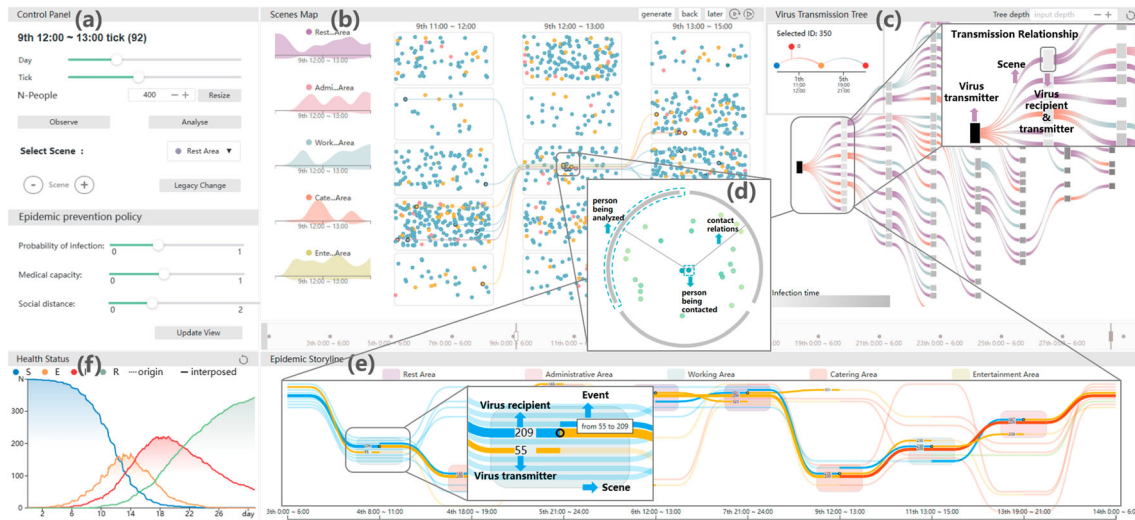
Fig. 1 A virtual simulation and visual analysis system of epidemic spread data. **a** The control panel provides the interface for user-defined setting the parameters necessary for the simulation of scenes and human movement patterns. **b** The scene map shows the distribution of people in the scenes, with the option to show the trajectory of people moving through the scenes, and provides an overview of the risks of the scenes. **c** The virus transmission tree shows the relationship between people spreading the virus and being spread and also shows the change in the person's health status when the person is selected. **d** The contact relationship shows the information about the contacts that the selected persons have with other persons during the specified period. **e** The epidemic storyline shows spatiotemporal information on the spread of the virus from the selected person to other persons. **f** The health status shows the change in health transitions over time for the four categories of Susceptible (*S*), Exposed (*E*), Infected (*I*), and Removed (*R*)

epidemic analysis can fully utilize the advantages of epidemic dynamic modeling to better understand and control virus transmission, providing valuable theoretical support for epidemic prevention and control. For instance, Angstmann et al. (2017) proposed a compartmental model based on a small, well-mixed population, while Brockmann and Helbing (2013) introduced an individual network epidemiological model based on complex contact relationships. Boudaoui et al. (2021) developed a fractional dynamic model that provides a deeper understanding of disease dynamics. Santos Gomes and Oliveira Serra (2021) proposed a machine learning model based on fuzzy clustering algorithms for adaptive tracking and real-time monitoring of the epidemic. Among various transmission dynamics models, the SEIR model fits well with many epidemics, making it widely used in epidemic transmission simulation. Many experts and scholars have made improvements and innovations based on this basic model to meet the needs of different situations. For example, Younsi et al. (2015) constructed a small-world network-based SEIR-SW model to investigate the influence of the susceptibility scale and topology structure of social networks on the spread of influenza epidemics, providing early warning displays and epidemic spread information to help decision-makers implement effective prevention and control measures. Kang et al. (2018) developed a delayed SEIR model with fixed latency based on scale-free networks to achieve more accurate simulations of epidemic transmission. Eduard et al. established a SEIR-D model using inverse parameter inference to predict the epidemic's development in a small range, accounting for the lack of simulation of epidemic transmission in a small range (Campillo-Funollet et al. 2021). Wu et al. proposed a SEIR model with variable infectivity and infinite time delay, considering the influence of infection age on epidemic transmission and analyzing the prediction of epidemics where the infection rate depends on infection age (Röst 2008). Additionally, to explore the impact of social media on the spread of the epidemic, Nie et al. (2021) constructed a SEIR model based on informatics to analyze the degree of social media influence and predict the epidemic's spread trend. However, epidemic models tend to be statistical, and it can be challenging to describe the specific transmission process in epidemic situations. This lack of detail can make it difficult for non-expert users to understand, and even expert users may miss essential characteristic information. Therefore, this paper proposes a visual analysis method that integrates model and data to facilitate a more intuitive perception of the epidemic transmission process. This approach establishes a connection between model parameters and specific virus transmission, making it easier to evaluate the effectiveness of epidemic prevention and control policies.

2.3 Spatiotemporal data visualization

Precision analysis is crucial for making epidemic contact data realistic and reflecting the actual world, and reasonable depiction of crowd movement trajectories is key. Time-series data analysis (Zhao et al. 2022; Peng et al. 2022; Zhao et al. 2023; Wang et al. 2020) can be a good representation of group trajectory information over time. Crowd trajectory analysis (Huang et al. 2019; Chen et al. 2017) is one of the current hotspots in the field of spatiotemporal data analysis, mainly used to extract the characteristics of crowd spatiotemporal trajectory (Afiq et al. 2019; Zhou et al. 2018). GPS positioning can be used for this type of analysis (Wang et al. 2014). For instance, in Nguyen et al. (2017), created an Android application that collects location data from registered users via GPS to study user movement prediction problems. Alternatively, WiFi network log data can be used to obtain time and space trajectory information. For example, Talavera et al. (2022) by analyzing the data obtained from WiFi infrastructure access points where users move, user mobility characteristics are extracted and a hierarchical user mobility model is generated to accurately describe user mobility behavior. Poucin et al. (2018) used WiFi network log data from Concordia University in Montreal, Canada, for analysis and clustering to identify the types of network users' activities in the building and to correlate the same type of activities of network users over multiple days. In the context of epidemic analysis, how to visualize the spatiotemporal movement data of infected individuals in an intuitive and reasonable manner remains a research focus. Baumgartl et al. (2020) proposed a novel visualization analysis method to support the analysis of transmission routes, patient contacts, epidemic progress, and patient schedules during hospitalization. Using this method, experts were able to extend the analysis of transmission routes to longer time intervals (years of data rather than days) and to more hospital rooms. Map-based scatter is capable of effectively expressing information about a person's geographic location (Zhou et al. 2020, 2023). However, when large-scale scattered points fill the map and interfere with our analysis, graph clustering (Zhou et al. 2022), graph sampling (Zhou et al. 2021) and dynamic networks analysis (Han et al. 2021) will be a good solution. Simpson et al. (2022) used data visualization techniques to construct dynamic maps to show the spatiotemporal variation of cholera infection and war conflict. Yang et al. (2022) designed an interactive visual analysis system called EpiMob (Epidemic Mobility) based on large human mobility data and urban POI data. The system interactively simulates changes in human mobility and infection status in response to the implementation of restrictive policies or a combination of policies (e.g., area lockdown, telecommuting, screening). To achieve this, a combination of simulation and visualization is used to fully utilize the advantages of visualization technology and develop spatiotemporal data visual analysis of epidemic diseases. This approach visually presents crowd movement regularities and the development trends of the epidemic situation in time and space. By using various visual mapping methods, it helps users perceive the epidemic development process and analyze the infection source.

3 Requirement analysis and pipeline

3.1 Requirement analysis

We discussed with two domain experts (E1 and E2), who are professionals in epidemic analysis, to obtain the requirements. E1 has four years of experience in public health security prevention and control and is well-versed in the analysis of epidemics. E2 is a campus educator and administrator with a keen interest in epidemic simulation and analysis, with the aim of providing reference value for campus epidemic prevention and control. Through discussions with experts and a literature review, we find that obtaining spatiotemporal data of the epidemic is crucial to the epidemic analysis. However, due to technology and privacy restrictions, real spatiotemporal epidemic data is hard to be obtained. Moreover, the factors that affect the simulation results should be editable by users to increase the generalizability of the simulation results. Additionally, experts pointed out that support is needed for epidemic analysis after the simulation of epidemic results, such as tracing the source of infection and screening close contacts, to aid epidemic prevention and control efforts. Overall, the requirements (R1-R4) can be summarized as follows:

3.1.1 R1. Simulation of epidemic spatiotemporal contact data

Acquiring population spatiotemporal contact data and virus transmission processes is challenging to the analysis of epidemics. To address this, we define customizable virtual scenes and migration preferences to

generate spatiotemporal contact simulation data. Furthermore, dynamic model is integrated to simulate the epidemic spread and generate epidemic simulation data, which establishes a data foundation for subsequent visual exploration of epidemic transmission laws.

3.1.2 R2. Visualization of epidemic spread process

The epidemic dynamics model, regarded as a black box, hinder users from exploring the real-time spatiotemporal evolution of epidemic transmission. To overcome this, classic visualizations such as scatter plots and tree diagrams are utilized to visually present the specific epidemic transmission process intuitively. Additionally, interactive exploration should be supported to help users understand and master the transmission characteristics and spatiotemporal evolution of epidemics.

3.1.3 R3. Analysis for traceability infection and investigating the close contact

Tracing the source of infection and analyzing close contacts are critical for interrupting the virus transmission chain and controlling the spread of the epidemic. However, traditional transmission dynamics models ignore to facilitate source tracing and close contact screening. Therefore, it is essential to provide users with visual tools to quickly trace the virus source, identify potential transmission pathways, and screen close contacts.

3.1.4 R4. Evaluation and visualization of intervention measures

Intervention measures such as crowd movement restrictions, social distancing control, and strong isolation strategies are commonly implemented to control epidemics (Tian et al. 2020). However, how to evaluate these measures is challenging to current works. These intervention measures should be customized and evaluated based on the epidemic data. This helps epidemic prevention and control staff to focus on prevention and control, and provides valuable references for the formulation of more effective human intervention strategies.
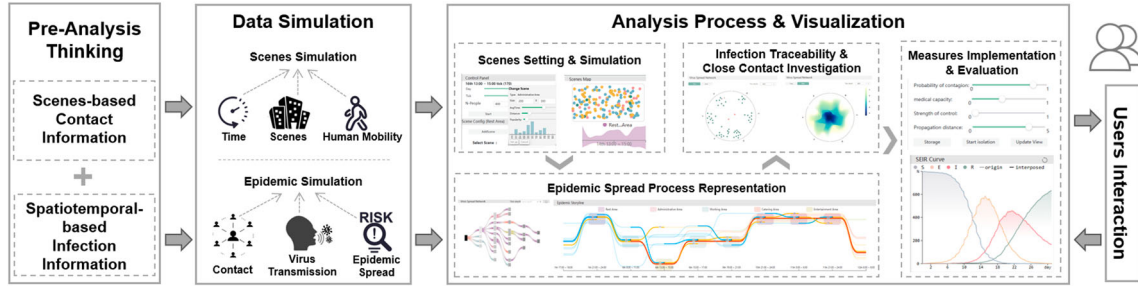
3.2 Pipeline

To address the aforementioned requirement, an interactive visualization and analysis system is designed to support the exploration of spatiotemporal patterns of epidemic spread, the evaluation of prevention measures, as well as the traceability of infection from epidemic exposure data. The pipeline of the visual analysis system is presented in Fig. 2, integrating three parts including pre-analysis and thinking, data simulation, and analysis process and visualization.

In this paper, to achieve effective epidemic simulation, we identify two key aspects for simulating epidemics, namely scene-based human contact information and spatiotemporal-based infection information, and consider the key factors for achieving simulation based on these two aspects. After research we identify the importance of time, scenes and human movement patterns in simulating the spatiotemporal trajectory of in scenes, and the importance of contact, virus transmission and epidemic spread in simulating spatiotemporal infection. To this end, we have designed a series of visualization interfaces incorporating these key factors to complete the epidemic analysis process including scenes setting and simulation, epidemic spread process representation, infection traceability and close contact investigation and epidemic decision evaluation. Multiple interactive interfaces facilitate a user-friendly epidemic simulation and analysis.

4 Methods

In this section, we provide a detailed description of the simulation process for the data base, which includes the scenes, human contact data, epidemic spread process, and the analysis of epidemic characteristics.

Fig. 2 The pipeline of VIVIAN includes pre-analysis thinking, data analysis, and an analysis workflow with a set of visualization and interaction for epidemic analysis process

4.1 Data basis

4.1.1 Crowd movement trajectory

As real data sets of epidemics are often difficult to obtain, the subsequent analysis of such data is challenging. Thus, this study relies on simulation to obtain the necessary data for analysis.

After several discussions with domain experts and a literature review, we find that fixed scene features are typically used when setting scenes in real environments. Thus, we have simplified the virtual scene into five basic functional scenes, namely *Rest Area*, *Entertainment Area*, *Catering Area*, *Working Area*, and *Administrative Area*. Users can expand the definition of scenes suitable for specific scenes based on these five basic scenes. This expansion can include adding or deleting scenes, defining scene attributes, and setting the number of people to simulate. Scene attributes include scene size, flow rate size between scenes based on the time axis, and the average contact distance between people in the scene. If the density of people in the scene exceeds the average contact distance density, people will move to the next scene based on the flow rate. Users can make necessary adjustments to express crowd movement rules under different scenes and states and thus obtain spatial contact simulation data.

4.1.2 Process of epidemic spread

After obtaining the spatial contact simulation data, the transmission process of epidemics was simulated in this study using the epidemic dynamics model. Over time, various epidemic dynamics modeling methods have been developed by scholars worldwide, including SIR, SIS, and SEIR (Kröger and Schlickeiser 2020; Dobie 2022; Heet al. 2020). Among these, the classical SEIR model is commonly used to describe epidemics that have an incubation period after infection, can be cured, and do not recur after treatment. The SEIR model classifies individuals into four categories: Susceptible (*S*), Exposed (*E*), Infected (*I*), and Recovered (*R*).

Susceptible (*S*) category represents the initial health state of the population. Exposed (*E*) category refers to the period after a susceptible person has been exposed to the virus, but has not yet fully developed the infection. Infected (*I*) category represents individuals who have developed the infection, and as time passes, individuals in the exposed state gradually progress to the infected state. Removed (*R*) category represents the final state of the population, where individuals have either recovered (Died individuals are not considered in this paper). The transition rate is shown in the following equation:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{r_t \beta S(I + \theta E)}{N}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{r_t \beta S(I + \theta E)}{N} - \sigma E$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \gamma_I I$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma_I I$$

The transmission process of epidemics was simulated through the epidemic dynamics model, where the probability of infection was denoted by β, the number of contacts at any time by r_t, θ represented the ratio of the transmission capacity of the exposed individuals (E) to that of the infected individuals (I), σ denoted the conversion rate of the exposed individuals (E) to the infected individuals (I), and γ_I indicated the recovery rate of the infected individuals (I).

By dividing the population into different categories in the space of whole domain, the process of disease transmission can be obtained. Users can simulate the spread of epidemics by setting basic parameters, such as the probability of virus infection, the propagation distance, the probability of infection based on space, the incubation period, and the recovery rate, to obtain spatiotemporal epidemic data and effectively depict the simulation process of epidemic spread.

4.2 Epidemic characteristics

4.2.1 Expression of activity trajectory of infected population

The scene setting and epidemic model are only capable of providing the overall activity rules of the population within the entire domain. However, during an epidemic, users are often interested in the activity tracks of specific population subsets, such as the selected individuals and the transmission paths of the infected.

The system in this paper displays the scene based on the time slice of the current time point and the time points before and after, as illustrated in Fig. 1b. Users can select the crowd they want to analyze by box selection in the scattered points of the map, and the location information of the marked crowd at the time points before and after will be connected by lines. The timeline can be dragged and dropped to visualize the movements of people at risk of infection.

Similarly, it is crucial to track the carriers of the virus and their transmission path to healthy individuals after an outbreak. As shown in Fig. 3a, we represent the transmission path of the virus in the population through storylines. The infected person being analyzed runs through the entire timeline, and the individual who has contact with the analyzed person and receives the virus at a specific time (reflected in the time axis) and place (reflected in the scene background) is shown in the storyline. The color of the line indicates the health status of the person, and when the color changes, it signifies a change in the health status of the individual. In addition, to avoid visual distractions, we only display the IDs of people within a given scene for the time in which the spread occurs, and we have designed the display with the people's health colors faded at the previous and later times to facilitate focus on the spread message.

4.2.2 Infection tracing and close contact investigation

In the context of unknown viruses, one of the major challenges in epidemic prevention and control is to interrupt the transmission chain and contain the source of infection. From the perspective of infection tracing: when tracing the origin of epidemics, the investigation usually starts with the contact history of the first detected patient in an attempt to identify the "patient zero". However, due to various uncontrollable factors such as privacy and the speed of recovery, it can be challenging to trace the precise source of the virus and identify the close contact with a high degree of certainty. From the point of view of close screening: close screening is equally important in containing the spread of the virus. Therefore, for patients who have already been identified, accurate screening of subsequent contact events and timely attention to contact personnel are crucial in blocking the transmission route.

In this paper, based on an exploration of the transmission characteristics and evolution rules of epidemics, we developed a contact correlation view to trace the source of infection and close contact investigation, as depicted in Fig. 3b. The circles represent infected individuals, while the concentric circles represent individuals who have been in contact with the infected individuals. The distribution of individuals within the circles is determined by the number of people who have been in contact with the infected individuals and the duration of contact. The closer the distribution is to the center, the more people have been in contact with the infected individuals. Additionally, the impact of the number of people in contact is greater than that of the duration of contact. The closer an individual is to the center, the higher the likelihood that they will become infected, indicating that they need more to be noticed.

After identifying individuals at risk of infection, users can perform traceability analysis by moving forward in time. The closer people are to the center of the circle, the more likely they are to have had contact
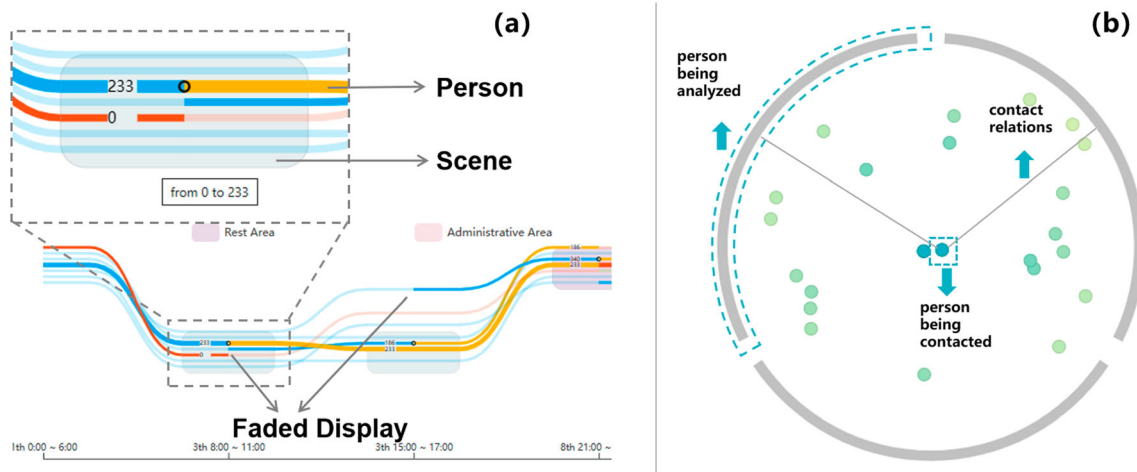
Fig. 3 The visual encoding of the epidemic storyline (**a**) and the infection tracing (**b**)

with multiple infected individuals, and this contact would have occurred some time before the analysis time. Based on this analysis, individuals of high concern are most likely to be the source of the infection that has spread to the individuals being analyzed. Users can also perform close contact screening by moving backward in time. The closer people are to the center, the more likely they are to have had contact with multiple infected individuals, and their risk of infection would be much higher than that of other individual contacts. Thus, these individuals should also be given high priority in the close investigation.

5 Visual design and interaction

5.1 Visual design

We provide an overview of our system interface in Fig. 1, followed by a detailed discussion of the visual coding and design of each view.

5.1.1 Scenes map view

As shown in Fig. 1b, in the map view, the time scene slices are displayed based on the predetermined scene in the current analysis period and the preceding and succeeding time periods. The scenes are differentiated with various colors, with personnel represented as scattered points. The scattered colors indicate the health status of four categories of personnel (S-blue, E-orange, I-red, and R-green). As the analysis time progresses, the personnel's location in the scene and their health state undergo changes. The left side of the map view contains a curve representing the scene infection risk value. Infection risk value is the quantitative value of infection risk faced by ordinary people present at the scene, which is determined by the corresponding time, the number of virus carriers present and the number of contact events. The top half of the curve represents the relative value at risk (VAR) for each scene based on current time, the bottom half represents the absolute VAR for each scene based on global time.

5.1.2 Virus transmission tree view

As shown in Fig. 1c, the virus transmission view allows for visualization of virus transmission among individuals in the global simulation scene from an observational standpoint. Starting from left to right, dissemination and propagation nodes can be observed. The node colors are a gradient of gray, indicating the backward order of the propagation time. The darker the color, the later the propagation time. The lines connecting the nodes represent the propagation relationship while the colors at both ends of the nodes map the scene during the corresponding person's propagation. Moreover, by clicking on a personnel node, it is selected as the starting node, and its infected successors' nodes are expanded. Additionally, the health

transformation diagram of the personnel is displayed at the top of the tree. The scene is represented by the color of the arc, indicating when and where the personnel's status changes.

5.1.3 Contact relationship view

As shown in Fig. 1d and described in detail in section 4.2.2, the curved area on the ring corresponds to each person being analyzed, with the person in the ring representing those in contact with them. The closer the distribution is to the center, the higher the number of people contacted, and the longer the contact time. Nodes on the ring are mapped using gradient shading, with darker colors indicating greater attention paid to the node. Clicking on a node on the ring also displays information about their connections with other individuals on the ring. Furthermore, we have combined a global-based timeline below the scene view and relational view, which includes a time flow pointer indicating the current analysis time and an interactive time selection pointer. The selection range between the two pointers can be utilized for subsequent infection tracing and close contact investigation during the specified period.

5.1.4 Epidemic storyline view

As shown in Fig. 1e, we have designed a storyline to illustrate the temporal and spatial processes of how an infected source transmits the virus to other patients. In the storyline, we represent people as lines, with the area background representing the scene (the scene color mapping is shown above the view and follows the same representation as the Scenes Map view). People flow between different scenes, and we distinguish between infected and uninfected individuals using the thickness of the lines. Thick lines represent infected individuals, while thin lines represent uninfected individuals. When a virus transmission event occurs, the transmission relationship between infected and uninfected individuals is marked, and the health status of the infected person changes from susceptible to exposed status (blue to orange). This storyline serves to provide a detailed illustration of the virus transmission process and related events.

5.1.5 Health status view

As shown in Fig. 1f. As depicted in the diagram, this view showcases the health changes of the population in a scene through the increase and decrease of curves. The health status is mapped using colors in the same way as in the Scenes Map view. Through this view, macro-information such as the timing of peak infection and the highest risk of infection can be understood within the epidemic simulation scene. Additionally, we can use dotted and solid lines to compare the original population health status with that following the implementation of epidemic prevention and control measures. This enables an evaluation of the measures effectiveness in improving overall health status (dotted line representing the original and solid line representing the status after measures have been implemented).

5.2 Interaction

In this section, we will begin with the control panel. As depicted in Fig. 1a, the control panel offers parameter setting options for epidemic simulation. For instance, by entering the value of "N-people" and clicking the "resize" button, the number of people in the simulation can be adjusted. The "observe" button provides enables the user to observe the entire epidemic spread process, displaying information such as the health status and transfer of people in the scene, details of the story line, and the spread of the epidemic. Likewise, the "analyse" button allows for analysis of unknown epidemic outbreaks from the simulation point of view, including the analysis of traceability and close connection investigation. The "add scene" and "delete scene" buttons provide scene adding and deleting operations, while the "select scene" option enables selection of scene attributes to be edited via a drop-down box. Scene attributes that can be modified include scene size, average personnel contact distance, scene density, and scene flow rate setting based on the time axis.

   In particular, when the personnel density in the scene exceeds a threshold, personnel will no longer continue to flow into that scene. They will instead be directed to the next-high flow scene. Such property setting actions are also provided when adding scenes. Furthermore, virus legacy risk settings are provided for different virus capabilities. The "Legacies Change" button permits user-defined size of legacy risk size to be set when considering the legacy risk of virus.
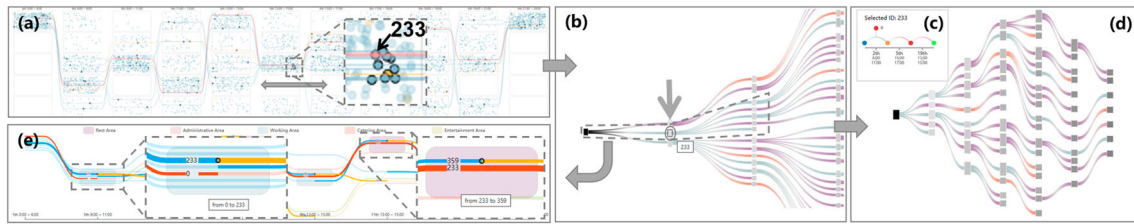
Fig. 4 Analysis of the epidemic spread process based on a person of interest. **a** Shows the movement of the person and those around him within each scene on day 6. **b** Provides an indication of the infection relationship of the person based on the global virus transmission tree. **c** Shows the health transitions of the person. **d** Shows the spread of the virus caused by the person. **e** Shows spatiotemporal information on the infection of other individuals by the person
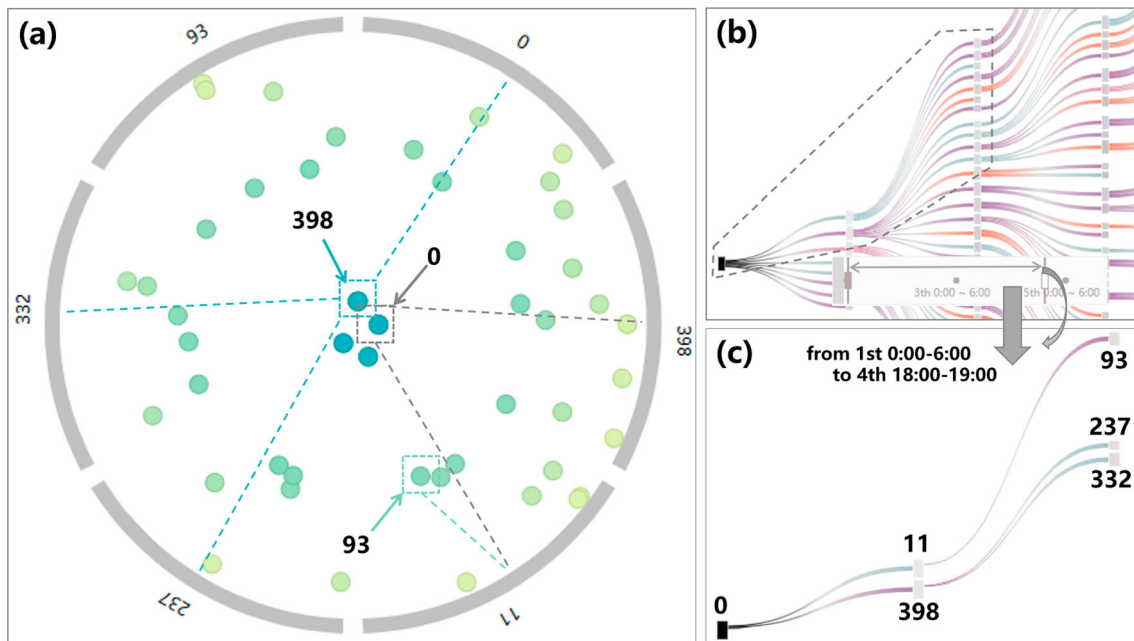
Fig. 5 Traceability of infection based on 6 analyzed persons(*Person 0, 398, 11, 237, 332 and 93*). **a** shows the contacts of the analyzed person with other people during the analyzed time period, and detailed virus transmission is verified by adjusting the time (**b**, **c**)

In the Scenes Map View, hovering the mouse over a scatter point allows us to view the ID information of the corresponding person, used to identify the person. We can then proceed to analyze the data based on whether we want to single-select, box-select, or multiple-select people. The selected individuals will have their scene activity tracks displayed in the upper and lower time frames, through the way of connection. Additionally, clicking on a single person will display the virus infection situation spread by them in the "virus spread network" and their health status change information. Moreover, a storyline will provide information on the spatiotemporal transmission of the virus from the selected carriers to other health personnel. In the upper right corner of the map view, we have provided the functionality of time-based play, pause, and replay. By selecting "back" or "later", we can play forward or backward in time to simulate the development of the epidemic in a scene, with corresponding changes in track information of the selected people (single, box, multiple).

In the Virus Transmission Tree View, in addition to selecting an analyst in the map to view the spread of the virus, we can also click on a person in the global spread tree for analysis. The health status change view and story line view of the person are also changed to show the information of the selected person. When the mouse hover node can also view the corresponding personnel ID information. By rolling the mouse wheel, the propagation tree can be scaled, so that we can view detailed information. At the top right of the tree, we provide a "tree depth" feature that allows you to input numbers to see how many generations the virus has
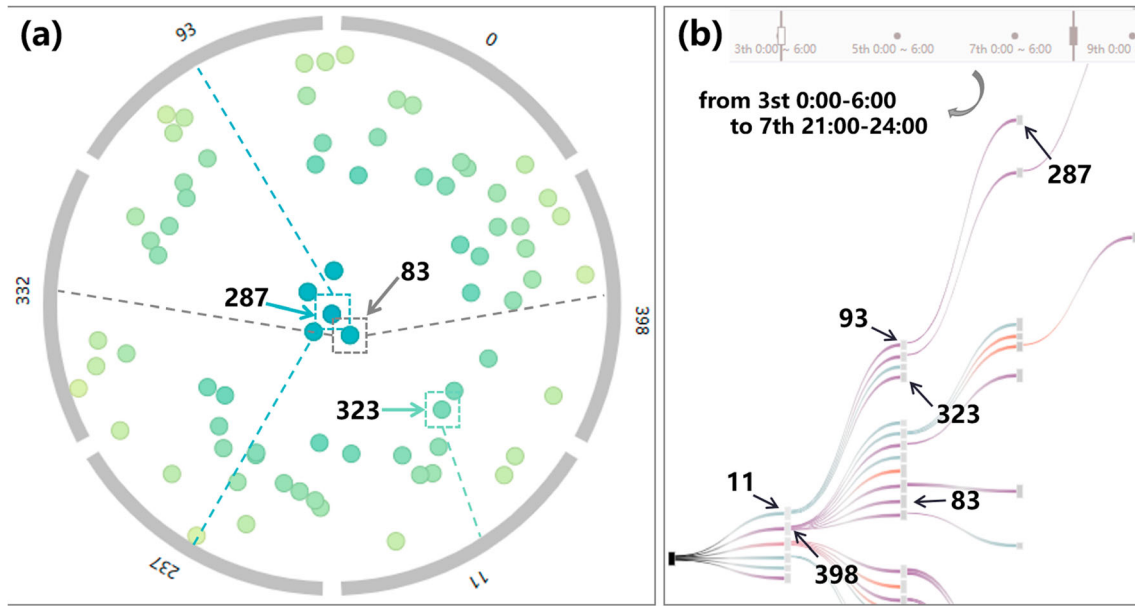
Fig. 6 Close contact investigation based on 6 analyzed persons(*Person 0, 398, 11, 237, 332 and 93*). **a** shows the contacts of the analyzed person with other people during the analyzed time period, and detailed virus transmission is verified by adjusting the time (**b**)

spread into. When the propagation tree changes to that of a specific analyst, you can click the back button to return to the global virus propagation tree.

In the Epidemic Storyline View, hovering over a given scene displays the time period during which virus transmission event occurred (e.g., "from 123 to 321"). In addition, users can zoom in and out of the storyline by scrolling the mouse wheel, enabling to view details.

For simulating the implementation of epidemic prevention measures, users may adjust the respective slider on the control panel to analyze the desired value. Once the parameters are set, clicking the "Implementation" button allows for the comparison of before and after implementation of the measures in the "health status" view.

6 Evaluation

In this section, we provide additional validation for the effectiveness of our system in analyzing the spatiotemporal dynamics of the epidemic as described in this paper. However, considering the pressure of computing power on the equipment and the decrease of analysis efficiency caused by the number of people and the scale of the scenes, we finally chose a more suitable simulation scene for our validation and analysis work. After several discussions with expert E2, we finally settle on implementing a data simulation based on a campus scene. We have set up the scenes as a Resting area (dormitory), an Administrative area (office building), a Catering area (restaurant), a Working area (school building) and a Entertainment area (playground). The size of the simulation was set at 400 people, considering that too many people could easily cause visual disruption to the user analysis. The simulation is carried out over a period of one month (31 days). Following the students' weekday habits of getting up in the morning and having breakfast, attending classes in the morning, eating lunch in the canteen, attending classes in the afternoon, eating dinner in the canteen and returning to the dormitory for a break after classes in the evening. In view of the short lunch and dinner periods and the staggered eating behavior, we divide the lunch and dinner periods into two parts. For this purpose each day is divided into 11 periods, namely the early morning break (0:00–6:00), the breakfast period (6:00–8:00), the morning courses time (8:00–11:00), the first half of the lunch period (11:00–12:00), the second half of the lunch period (12:00–13:00), the first half of the afternoon courses time (13:00–15:00), the second half of the afternoon courses time (15:00–17:00), first half of the dinner period (17:00–18:00), second half of the dinner period (18:00–19:00), evening courses time (19:00–21:00), and evening break
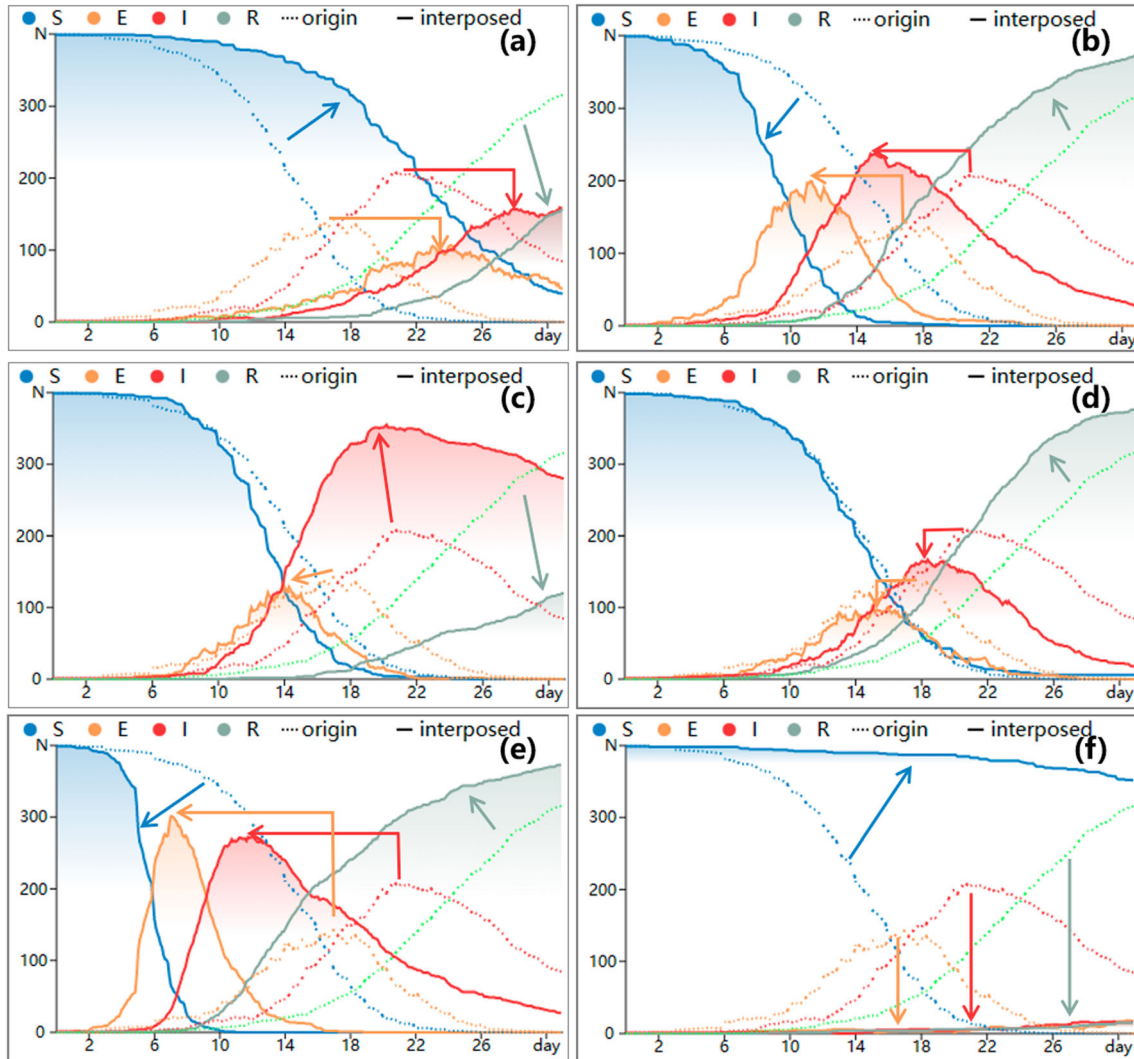
Fig. 7 Evaluation of the epidemic prevention decisions effectiveness. **a**, **b** show the adjustment of the values of the "probability of infection" to 0.20 and 0.50. **c**, **d** show the adjustment of the values of the "medical capacity" to 0.20 and 0.70. **e**, **f** show the adjustment of the values of the "social distance" to 0.30 and 1.00

(21:00–24:00). We simulate a total of five scenes (*Rest Area*, *Entertainment Area*, *Catering Area*, *Working Area*, and *Administrative Area*), each with different circulation rates of the population over time.

6.1 Case study

6.1.1 Epidemic simulation

During an epidemic, experts need to track the movements of the person of concern, the timing of changes in their health status, and the transmission of the virus caused by them over a specific period.

Using an observational approach, we identify an infected person on day 6 during the 13:00–15:00 phase, and select their neighbors by boxing in nodes to observe their movements throughout the day by adjusting the time, as illustrated in Fig. 4a. Our investigation reveals that the infected person has an *ID of 233* and is a descendant of the "Patient Zero" in the global virus transmission relationship. As depicted in Fig. 4b, selecting the *Person 233* node in the tree diagram regenerates the tree starting from *Person 233* to display the infection's spread, as shown in Fig. 4d. We also observe that *Person 233* contracts the virus on day 2, the same day as Patient Zero, is healthy on day 5, and does not recover until day 19, as shown in Fig. 4c.

Based on the above observations, experts also want to view spatiotemporal information regarding the transmission of the virus from *Patient 233* to other individuals. Figure 4e shows that by selecting the *Person 233* node in the tree diagram, the storyline changes accordingly to reveal who transmits the virus to *Person 233* (8:00–11:00 on day 3, Patient Zero in the *Working Area* scene transmits the virus to *Person 233*), and who in turn transmits the virus to others (subsequently at different times in the *Working Area* and *Rest Area*, respectively, e.g., 13:00–15:00 on day 11 spreading the virus to *Person 359*).

6.1.2 Infection tracing and close contact investigation

During the investigation, an infected person (identified as *Person 398*) is discovered during the 17:00–18:00 period on day 4. Following this discovery, a thorough epidemic investigation is conducted at the next time period (18:00–19:00 on day 4) and two other infected persons (*Person 0, 11*), as well as three individuals who are found to be carriers of the virus (*Person 93, 237, 332*) are identified. By selecting these individuals and considering the time period between 0:00–6:00 on day 1 and, 18:00–19:00 on day 4, the exposure of the five analyzed individuals to others during this time is determined, as shown in Fig. 5a. Based on the analysis, it is observed that both *Person 11* and *Person 398* have contact with other individuals who are infected, implying that the other individuals who are in contact with them are potentially infected by these two individuals. Further investigation reveals a *Person 0* who had come into contact with *Person 11* and *Person 398* at an earlier time, thus suggesting that *Person 0* also has a higher suspicion of infection. In order to confirm the validity of the retrospective analysis, a virus transmission tree is analyzed based on an observational perspective. After adjusting for the same time and angle of analysis (Fig. 5b–c), confirming that the virus is indeed transmitted by *Person 0*, who subsequently transmits the virus to *Persons 11* and *398*, who then transmit the virus to *Patients 93,237* and *332*.

In addition, to further prevent the spread of the epidemic, we need to conduct contact tracing among confirmed cases' close contacts over a 5-day period (0:00–6:00 on day 3 to 21:00–24:00 on day 7). According to Fig. 6, numerous individuals (*Person 287, 83*) are in contact with multiple confirmed cases, putting them at high risk for infection as shown in the diagram of virus spread. Furthermore, individuals who have had contact with only one confirmed case are not necessarily safe from infection, for example, *Person 323* in Fig. 6.

Therefore, it is important to precisely identify their close contacts and implement control measures for all identified at-risk individuals in order to effectively stop the further spread of the epidemic.

6.1.3 Evaluation of epidemic prevention measures

The ultimate goal of epidemic analysis is to aid prevention management in formulating appropriate epidemic prevention policies that quickly and effectively control the spread of the virus in the population. To achieve this, we simulate the natural state and utilize different means of epidemic prevention interventions by simultaneously correlating them with model parameters, and combine SEIR time-varying curve views to demonstrate the effects of varying epidemic prevention measures.

Probability of infection   The default value of "probability of infection" is 0.35. Figure 7a–b illustrates that the "probability of infection" values are adjusted to 0.20 and 0.50, respectively. A comparison of pre-intervention and post-intervention SEIR curves indicates that lowering the transmission probability results in a slower rate of decline in the susceptible population curve, a delay in its inflection point, while the rate of latent and infected curves decreases. Conversely, an increase in transmission probability leads to a faster decline in the susceptible population curve and a corresponding forward shift in its inflection point, whereas the latent and infected curves increase more rapidly and reach their peak numbers earlier. This suggests that the implementation of measures to control virus transmission has effectively reduced infection rates in susceptible populations.

Medical capacity   The default value of "medical capacity" is 0.35. As shown in Fig. 7c, d, the "medical capacity" parameter was changed to 0.20 and 0.70, respectively. By comparing the SEIR curves before and after changing the parameter value, it is clear that an effective vaccine can significantly increase the cure rate, though the effect on susceptible and latent individuals is insignificant. However, the effect on infected and recovered individuals is notable. As the cure rate increases, the peak of the infected individuals curve decreases significantly, and the curve for recovered individuals increases notably. On the other hand, if an

effective vaccine is not developed in time for the virus, or if the virus mutates so that the original vaccine is no longer suitable, the peak of the infected individuals curve will be much higher.

Social distance The default value for "social distance" is 0.5, which users can adjust to examine the impact of social distancing measures on virus prevention and control. Figure 7e, f shows that setting "social distance" values to 0.3 and 1.0 results in a sharp decline in the number of susceptible individuals during high-density crowd activities when social distance is reduced. Meanwhile, the latent and infected curves exhibit significantly accelerated growth with a higher peak, exceeding that of the normal curve. Conversely, if the physical distance between individuals increases, the number of latent and infected individuals decreases markedly. By increasing the distance between individuals to 1.0, we observe that only 23 newly infected individuals arise in our simulations. Thus, controlling physical distance among individuals is an effective intervention in curtailing the spread of infection among the population.

This demonstrates the effectiveness of prevention methods in curbing the spread of the virus. Officials should encourage the public to gather in smaller groups, practice proper hand hygiene and wear face masks. Moreover, they should devote more resources to developing vaccines to safeguard public health and safety.

6.2 Expert interview

To assess the effectiveness of VIVIAN, we conducted interviews with five epidemiologists consisting of two domain experts (E1 and E2, who have been introduced in Section 3.1) and three PhDs (E3–E5). Three epidemiologists, E1–E3, used VIVIAN for the first time while researching epidemic prevention.

6.2.1 Procedure

We completed the following steps to collect their comments. Training (20 min). We introduced our motivation, related definitions, and visual designs following the visualization introductory description guidelines. Full epidemic analysis process (20 min). We let epidemiologists complete a full process for the epidemic simulation and the exploration processes were recorded. Interview (20 min). We asked them to evaluate our system from three aspects, i.e., universal scene setting, the effectiveness of the interaction, and the intuitiveness of the visual cues.

6.2.2 Universal scene setting

By setting up scene attributes and human movement patterns, the epidemiologists pointed out that our system VIVIAN can effectively cater for scene simulation in different situations. H2 mentioned that as a campus manager, the simulation of an epidemic for a campus scene is different from other scenes, and people move differently. The user-defined settings meet the needs of scene-specific simulations. All epidemiologists affirmed the significance of our work and believed that VIVIAN is universal.

6.2.3 Effectiveness of the interaction

All epidemiologists were very impressed by our system VIVIAN, especially the time-based projection of the spread of the epidemic, which demonstrated the full course of the epidemiological situation and provided information on the trajectory of the analyzed persons. H1 and H5 still pointed out that the time selection operations allow for targeted infection tracing and close contact identification, enabling users to analyze infection contact information for a specific time period. Finally, experts praised the excellent performance of our system in supporting epidemic spread data analysis.

6.2.4 Intuitiveness of the visual cues

Epidemiologists have shown that epidemic spread data contain complex spatiotemporal content. Based on the visual cues our visual system provide, confusion of time, space, people and events in the simulated scene is well avoided, allowing intuitive and orderly epidemic analysis process. Meanwhile, epidemiologists were pleasantly surprised by the visual attention-based analysis of infection traceability and close contact investigation, which effectively guided personnel at high risk of infection.

7 Discussion

Efficient epidemic analysis enables experts to quickly grasp epidemic development, accurately identify key players in the transmission chain, and make prompt and accurate epidemic prevention decisions to maximize safety. Access to comprehensive spatial and temporal epidemic data is crucial to achieve these goals. Accordingly, we developed VIVIAN, a virtual simulation and visual analysis system of epidemic spread data for analyzing epidemic processes and evaluating epidemic prevention decisions more effectively. However, some problems remain unsolved in this research, which will be addressed in future work.

7.1 Application of real data sets

Analysis of epidemics based on realistic datasets provides an accurate representation of its spread. Nonetheless, obtaining useful spatiotemporal information for the development of an epidemic in small-scale scenarios remains challenging. In this paper, we obtained key spatiotemporal contact data through simulation, which is universal enough to accommodate a variety of scenarios but lacks realism. In future work, we will continue to seek realistic datasets on matched spatiotemporal base data and virus transmission data to complete analyses of epidemic spread based on small-scale scenes.

7.2 Reflections on the evaluation of isolation measures

Epidemic decisions in our simulation include reducing infection probability, improving medical capacity, and increasing social distance. However, in the case of a real epidemic, isolation is the most effective way to isolate infected individuals from the overall scenes. Often, this requires distinct scene settings, though discussions with experts show that using isolation scenarios does not always fit simulation analysis needs. In future work, we will integrate analysis requirements that account for the isolation strategy.

7.3 Refinement of epidemic characteristics

Each pandemic exhibits unique characteristics, such as a high infection rate or low cure rate, that can result in varying patterns of epidemic spread. However, in this paper, we chose not to account for numerous epidemic characteristics in our epidemic simulations, making them more macro-oriented. Thus, future work will involve incorporating these characteristics from previous epidemics to improve the epidemic simulation database.

8 Conclusion

With a focus on essential elements of analyzing epidemics, we designed and built VIVIAN—a virtual simulation and visual analysis system of epidemic spread data, which enables user-defined scene attribute definition and epidemic spread control simulation. In addition, we integrate a set of visual cues and interactions to assist users with epidemic process analysis, including scenes setting & simulation, epidemic spread process representation, infection traceability & close contact investigation and epidemic decision evaluation. Case studies and expert interviews demonstrated the usefulness and effectiveness of the system.

References

Afiq A, Zakariya M, Saad M, Nurfarzana A, Khir MHM, Fadzil A, Jale A, Gunawan W, Izuddin Z, Faizari M (2019) A review on classifying abnormal behavior in crowd scene. J Vis Commun Image Represent 58:285–303. https://doi.org/10.1016/j.jvcir.2018.11.035

Andrienko N, Andrienko G, Miksch S, Schumann H, Wrobel S (2021) A theoretical model for pattern discovery in visual analytics. Vis Inform 5(1):23–42. https://doi.org/10.1016/j.visinf.2020.12.002

Angstmann CN, Erickson AM, Henry BI, McGann AV, Murray JM, Nichols JA (2017) Fractional order compartment models. SIAM J Appl Math 77(2):430–446. https://doi.org/10.1137/16M1069249

Bartsch SM, O'Shea KJ, Ferguson MC, Bottazzi ME, Wedlock PT, Strych U, McKinnell JA, Siegmund SS, Cox SN, Hotez PJ (2020) Vaccine efficacy needed for a covid-19 coronavirus vaccine to prevent or stop an epidemic as the sole intervention. Am J Prev Med 59(4):493–503. https://doi.org/10.1016/j.amepre.2020.06.011

Baumgartl T, Petzold M, Wunderlich M, Hohn M, Archambault D, Lieser M, Dalpke A, Scheithauer S, Marschollek M, Eichel V (2020) In search of patient zero: visual analytics of pathogen transmission pathways in hospitals. IEEE Trans Vis Comput Graph 27(2):711–721. https://doi.org/10.1109/TVCG.2020.3030437

Boudaoui A, Moussa Y, Hammouch Z, Ullah S (2021) A fractional-order model describing the dynamics of the novel coronavirus (covid-19) with nonsingular kernel. Chaos Solitons Fractals 146:110859. https://doi.org/10.1016/j.chaos.2021.110859

Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. Science 342(6164):1337–1342. https://doi.org/10.1126/science.1245200

Campillo-Funollet E, Van Yperen J, Allman P, Bell M, Beresford W, Clay J, Dorey M, Evans G, Gilchrist K, Memon A (2021) Predicting and forecasting the impact of local outbreaks of covid-19: use of SEIR-D quantitative epidemiological modelling for healthcare demand and capacity. Int J Epidemiol 50(4):1103–1113. https://doi.org/10.1093/ije/dyab106

Chen W, Huang Z, Wu F, Zhu M, Guan H, Maciejewski R (2017) Vaud: a visual analysis approach for exploring spatio-temporal urban data. IEEE Trans Vis Comput Graph 24(9):2636–2648. https://doi.org/10.1109/TVCG.2017.2758362

Chen Y, Li Y, Feng S, Man X, Long Y (2021) Gravitational scaling analysis on spatial diffusion of covid-19 in Hubei Province, China. PLoS ONE 16(6):0252889. https://doi.org/10.1371/journal.pone.0252889

Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Piontti A, Mu K, Rossi L, Sun K (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science 368(6489):395–400. https://doi.org/10.1126/science.aba9757

Corberán-Vallet A, Lawson AB (2014) Prospective analysis of infectious disease surveillance data using syndromic information. Stat Methods Med Res 23(6):572–590. https://doi.org/10.1177/0962280214527385

Dobie AP (2022) Susceptible-infectious-susceptible (SIS) model with virus mutation in a variable population size. Ecol Complex 50:101004. https://doi.org/10.1016/j.ecocom.2022.101004

El-Khatib Z, Taus K, Richter L, Allerberger F, Schmid D (2019) A syndrome-based surveillance system for infectious diseases among asylum seekers in Austrian reception centers, 2015–2018: analysis of reported data. JMIR Public Health Surveill 5(1):11465. https://doi.org/10.2196/11465

Fang L-Q, De Vlas SJ, Feng D, Liang S, Xu Y-F, Zhou J-P, Richardus JH, Cao W-C (2009) Geographical spread of SARS in mainland China. Trop Med Int Health 14:14–20. https://doi.org/10.1111/j.1365-3156.2008.02189.x

Han D, Pan J, Pan R, Zhou D, Cao N, He J, Xu M, Chen W (2021) iNet: visual analysis of irregular transition in multivariate dynamic networks. Front Comput Sci 16:1–16. https://doi.org/10.1007/s11704-020-0013-1

He S, Peng Y, Sun K (2020) SEIR modeling of the covid-19 and its dynamics. Nonlinear Dyn 101:1667–1680. https://doi.org/10.1007/s11071-020-05743-y

Huang Z, Zhao Y, Chen W, Gao S, Yu K, Xu W, Tang M, Zhu M, Xu M (2019) A natural-language-based visual query approach of uncertain human trajectories. IEEE Trans Vis Comput Graph 26(1):1256–1266. https://doi.org/10.1109/TVCG.2019.2934671

Kang H, Sun M, Yu Y, Fu X, Bao B (2018) Spreading dynamics of an SEIR model with delay on scale-free networks. IEEE Trans Netw Sci Eng 7(1):489–496. https://doi.org/10.1109/TNSE.2018.2860988

Kröger M, Schlickeiser R (2020) Analytical solution of the sir-model for the temporal evolution of epidemics. Part A: time-independent reproduction factor. J Phys A Math Theor 53(50):505–601. https://doi.org/10.1088/1751-8121/abc65d

Kumar A, Choi T-M, Wamba SF, Gupta S, Tan KH (2021) Infection vulnerability stratification risk modelling of covid-19 data: a deterministic SEIR epidemic model analysis. Ann Oper Res. https://doi.org/10.1007/s10479-021-04091-3

Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, Wesolowski A, Santillana M, Zhang C, Du X (2020) Effect of non-pharmaceutical interventions to contain covid-19 in China. Nature 585(7825):410–413. https://doi.org/10.1038/s41586-020-2293-x

Merler S, Ajelli M, Fumanelli L, Gomes MF, Piontti AP, Rossi L, Chao DL, Longini IM, Halloran ME, Vespignani A (2015) Spatiotemporal spread of the 2014 outbreak of ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. Lancet Infect Dis 15(2):204–211. https://doi.org/10.1016/S1473-3099(14)71074-6

Mukandavire Z, Nyabadza F, Malunguza NJ, Cuadros DF, Shiri T, Musuka G (2020) Quantifying early covid-19 outbreak transmission in south Africa and exploring vaccine efficacy scenarios. PloS ONE 15(7):0236003. https://doi.org/10.1371/journal.pone.0236003

Naseem U, Razzak I, Khushi M, Eklund PW, Kim J (2021) Covidsenti: a large-scale benchmark twitter data set for covid-19 sentiment analysis. IEEE Trans Comput Soc Syst 8(4):1003–1015. https://doi.org/10.1109/TCSS.2021.3051189

Nguyen BT, Nguyen NV, Nguyen NT, Tran MHT (2017) A potential approach for mobility prediction using GPS data. In: 2017 seventh international conference on information science and technology (ICIST), pp 45–50. IEEE, Da Nang, Vietnam. https://doi.org/10.1109/ICIST.2017.7926813

Nie Q, Liu Y, Zhang D, Jiang H (2021) Dynamical SEIR model with information entropy using covid-19 as a case study. IEEE Trans Comput Soc Syst 8(4):946–954. https://doi.org/10.1109/TCSS.2020.3046712

Pak TR, Kasarskis A (2015) How next-generation sequencing and multiscale data analysis will transform infectious disease management. Clin Infect Dis 61(11):1695–1702. https://doi.org/10.1093/cid/civ670

Peng Y, Fan X, Chen R, Yu Z, Liu S, Chen Y, Zhao Y, Zhou F (2022) Visual abstraction of dynamic network via improved multi-class blue noise sampling. Front Comput Sci 17(1):171701. https://doi.org/10.1007/s11704-021-0609-0

Polonsky JA, Baidjoe A, Kamvar ZN, Cori A, Durski K, Edmunds WJ, Eggo RM, Funk S, Kaiser L, Keating P (2019) Outbreak analytics: a developing data science for informing the response to emerging pathogens. Philos Trans R Soc B 374(1776):20180276. https://doi.org/10.1098/rstb.2018.0276

Poucin G, Farooq B, Patterson Z (2018) Activity patterns mining in WI-FI access point logs. Comput Environ Urban Syst 67:55–67. https://doi.org/10.1016/j.compenvurbsys.2017.09.004

Ren J, Yan Y, Zhao H, Ma P, Zabalza J, Hussain Z, Luo S, Dai Q, Zhao S, Sheikh A (2020) A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for covid-19. IEEE J Biomed Health Inform 24(12):3551–3563. https://doi.org/10.1109/JBHI.2020.3027987

Rocklöv J, Sjödin H (2020) High population densities catalyse the spread of covid-19. J Travel Med 27(3):038. https://doi.org/10.1093/jtm/taaa038

Röst G (2008) SEIR epidemiological model with varying infectivity and infinite delay. Math Biosci Eng 5(2):389–402

Santos Gomes DC, Oliveira Serra GL (2021) Machine learning model for computational tracking and forecasting the covid-19 dynamic propagation. IEEE J Biomed Health Inform 25(3):615–622. https://doi.org/10.1109/JBHI.2021.3052134

Silva PC, Batista PV, Lima HS, Alves MA, Guimarães FG, Silva RC (2020) Covid-abs: an agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos Solitons Fractals 139:110088. https://doi.org/10.1016/j.chaos.2020.110088

Simpson RB, Babool S, Tarnas MC, Kaminski PM, Hartwick MA, Naumova EN (2022) Dynamic mapping of cholera outbreak during the Vemeni civil war, 2016–2019. J Public Health Policy 43(2):185–202. https://doi.org/10.1057/s41271-022-00345-x

Sullivan PS, Woodyatt C, Koski C, Pembleton E, McGuinness P, Taussig J, Ricca A, Luisi N, Mokotoff E, Benbow N (2020) A data visualization and dissemination resource to support HIV prevention and care at the local level: analysis and uses of the AIDSVU public data resource. J Med Internet Res 22(10):23173. https://doi.org/10.2196/23173

Talavera F, Lera I, Guerrero C (2022) Analysis of wireless network access logs for a hierarchical characterization of user mobility. J King Saud Univ-Comput Inf Sci 34(6):2471–2487. https://doi.org/10.1016/j.jksuci.2022.03.014

Tian H, Liu Y, Li Y, Wu C-H, Chen B, Kraemer MU, Li B, Cai J, Xu B, Yang Q (2020) An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in China. Science 368(6491):638–642. https://doi.org/10.1126/science.abb6105

Wang X, Wu Z, Huang W, Wei Y, Huang Z, Xu M, Chen W (2023) VIS+ AI: integrating visualization with artificial intelligence for efficient data analysis. Front Comput Sci 17(6):1–12. https://doi.org/10.1007/s11704-023-2691-y

Wang F, Chen W, Wu F, Zhao Y, Hong H, Gu T, Wang L, Liang R, Bao H (2014) A visual reasoning approach for data-driven transport assessment on urban roads. In: 2014 IEEE conference on visual analytics science and technology (VAST), pp 103–112. IEEE, Paris, France. https://doi.org/10.1109/VAST.2014.7042486

Wang X, Chen W, Xia J, Chen Z, Xu D, Wu X, Xu M, Schreck T (2020) Conceptexplorer: visual analysis of concept drifts in multi-source time-series data. In: IEEE conference on visual analytics science and technology (VAST), pp 1–11. IEEE, Salt Lake City, UT, USA. https://doi.org/10.1109/VAST50239.2020.00006

Wang J, Tang K, Feng K, Lv W (2020) High temperature and high humidity reduce the transmission of covid-19. Available at SSRN 3551767, 2020 https://doi.org/10.2139/ssrn.3551767

Yang C, Zhang Z, Fan Z, Jiang R, Chen Q, Song X, Shibasaki R (2022) Epimob: interactive visual analytics of citywide human mobility restrictions for epidemic control. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2022.3165385

Younsi F-Z, Bounnekar A, Hamdadou D, Boussaid O (2015) SEIR-SW, simulation model of influenza spread based on the small world network. Tsinghua Sci Technol 20(5):460–473. https://doi.org/10.1109/TST.2015.7297745

Zhao Y, Ge L, Xie H, Bai G, Zhang Z, Wei Q, Lin Y, Liu Y, Zhou F (2022) ASTF: visual abstractions of time-varying patterns in radio signals. IEEE Trans Vis Comput Graph 29(1):214–224. https://doi.org/10.1109/TVCG.2022.3209469

Zhao Y, Lv S, Long W, Fan Y, Yuan J, Jiang H, Zhou F (2023) Malicious webshell family dataset for webshell multi-classification research. Vis Inform. https://doi.org/10.1016/j.visinf.2023.06.008

Zhou Z, Meng L, Tang C, Zhao Y, Guo Z, Hu M, Chen W (2018) Visual abstraction of large scale geospatial origin-destination movement data. IEEE Trans Vis Comput Graph 25(1):43–53. https://doi.org/10.1109/TVCG.2018.2864503

Zhou Z, Ma Y, Zhang Y, Liu Y, Liu Y, Zhang L, Deng S (2021) Context-aware visual abstraction of crowded parallel coordinates. Neurocomputing 459:23–34. https://doi.org/10.1016/j.neucom.2021.05.005

Zhou Z, Sun L, Wang H, Yu W, Liu Y, Zhang X, Wang Y, Chen W (2022) iMGC: Interactive multiple graph clustering with constrained Laplacian rank. IEEE Trans Hum–Mach Syst 53(2):427–437. https://doi.org/10.1109/THMS.2022.3227181

Zhou Z, Zheng F, Wen J, Chen Y, Li X, Liu Y, Wang Y, Chen W (2023) A user-driven sampling model for large-scale geographical point data visualization via convolutional neural networks. IEEE Trans Hum–Mach Syst 53(5):885–894. https://doi.org/10.1109/THMS.2023.3296692

Zhou Z, Zhang X, Yang Z, Chen Y, Liu Y, Wen J, Chen B, Zhao Y, Chen W (2020) Visual abstraction of geographical point data with spatial autocorrelations. In: IEEE conference on visual analytics science and technology (VAST), pp 60–71. IEEE, Salt Lake City, UT, USA. https://doi.org/10.1109/VAST50239.2020.00011