

## Multimodal Epidemic Visual Analytics and Modeling

Seokyeon Kim  
Sejong University  
[ksy0586@sju.ac.kr](mailto:ksy0586@sju.ac.kr)

Yun Jang  
Sejong University  
[jangy@sejong.edu](mailto:jangy@sejong.edu)

### Abstract

*The risk of infectious disease increases due to various factors, including the dense population, development of various transportations, urbanization, and abnormal weather conditions. Since the speed of epidemic spread is fast, it is necessary to respond quickly to prevent the high fatality rate. Therefore, a fast search for the highly accurate spreading model has to be focused on the proper analysis of disease spreading. There have been many studies to understand the disease spreading, and the epidemic model is often used to analyze and predict the spread of infectious disease. However, it is limited to apply the epidemic model for the spread analysis because it captures spreading changes only within the defined area. In this paper, we propose a framework for the disease spreading simulation with multimodal factors in the epidemic model and networks of possible spread routes. Our system provides an interactive simulation environment with the interregional disease spreading according to various spread parameters. Moreover, to understand the spreading directions, we extract vector fields over time and visualize the vector fields with the fatality of the disease. Therefore, users can understand the disease spreading phenomena and obtain appropriate models through our framework.*

### 1. Introduction

The human lifespan has been increased as medicine develops. However, the risk of disease also increases at the same time since humans are more exposed to many different types of infectious diseases. The pandemic is transmitted between humans, and as the longevity and population density increase, the risk of the epidemic spared has increased. Moreover, the new extraordinary deleterious disease has been discovered and their influence on humans is considerably threatening.

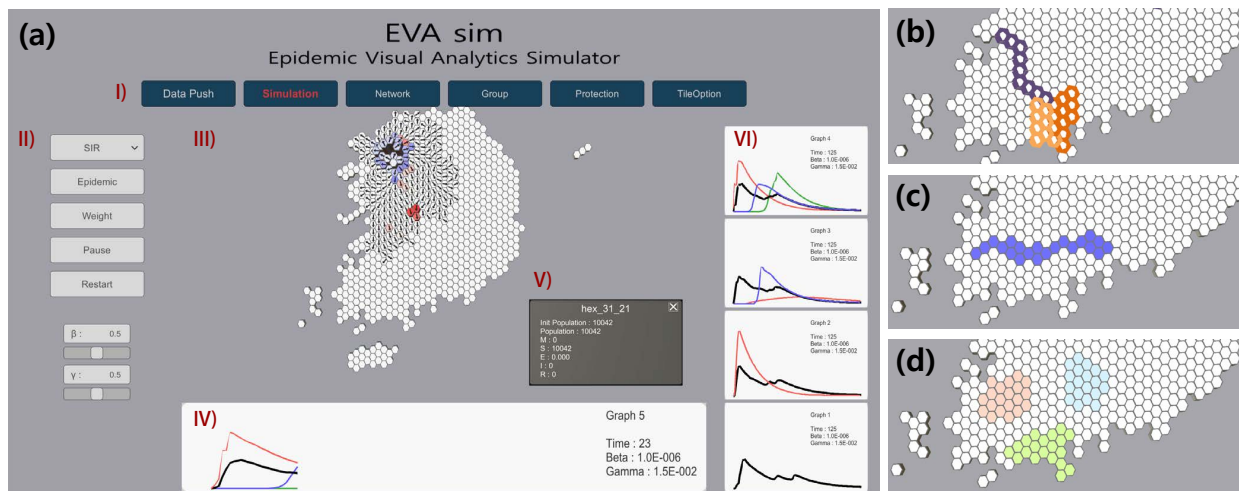
The WHO manages six pandemic phases as follows [1]. In phase 1, only animal-to-animal infections

are reported and do not cause human infection. In phase 2 humans are infected by wild animals, which is considered as a potential pandemic threat. In phase 3, animal-to-human transmission occurs, creating small clusters of infected areas. In phase 4, human-to-human transmission occurs, which causes community-level outbreaks. Phase 5 refers to the imminent situation of a global pandemic, spreading to at least two of the WHO regions divided into six regions. Phase 6 is a situation where another region is infected, indicating that an pandemic is ongoing.

Animal-to-animal transmission or human-to-animal infection only spreads to a small extent. However, in the pandemic phase 4 and above, human-to-human transmission occurs and the radius of spread increases with the size of the community. To understand the spreading patterns of newly found disease, there have been many studies on various factors that cause the disease spreading. Since the speed of epidemic spread is fast, it is necessary to react promptly to avoid the uncontrollable disease spread. In previous years, most of the epidemic models were simulated under certain conditions within limited areas.

As technologies for the data collection and storage advance, lots of data are available for the research purpose. Especially, the information related to disease, including patients and hospitals, is available for the use under privacy policies. Therefore, it is possible to model and simulate the infectious disease spreading based on the actual data. On the other hand, many epidemic models are representing actual disease spreading. However, it is not easy to find a proper model with pertinent parameters and predict spreading patterns for emergent cases. Although experienced experts investigate an epidemic outbreak, it is still difficult to estimate the disease spreading without any professional support due to a large amount of data. Thus, it is necessary to develop a suitable system simulating disease spreading and discovering appropriate epidemic models for immediate emergency responses.

In this paper, we present an interactive multimodal



**Figure 1. Epidemic VA system. (a) presents our system; (I)-(VI) represent the control panels, simulation map view, graph view, detail information view, and history view, respectively. (b)-(d) visualize settings of the protection model, additional roads networks, and area grouping, respectively.**

epidemic visual analytics and modeling system to analyze the epidemic situations. Our system consists of an epidemic simulation and epidemic modeling. The epidemic simulation utilizes actual environmental factors, including population, transportation, weather condition, transmission rate, and antiviral supplies, to simulate disease spreading patterns. On the other hand, epidemic modeling provides an optimal disease spreading model based on the patient and hospital data to predict the risk and disease spreading patterns.

For the epidemic modeling, our system loads temporal infectious rates and performs optimizations to estimate the parameters of epidemic models. Besides, analysts can add environment conditions to the optimization as well as the simulation of the future epidemic patterns. Our system enables analysts to simulate possible cases with different conditions interactively and to model current epidemic patterns.

The major contributions of our work are as follows:

- We have developed an interactive epidemic simulation framework to analyze the infectious disease spreading.
- Interactive condition setting is designed intuitively for the better usability of the epidemic simulation.
- Major epidemic flow patterns are extracted and visualized to estimate the spreading patterns.
- Epidemic model is approximated based on the actual disease and infectious information to characterize disease for the automated prediction.

## 2. Related Work

The epidemic model has been developed as a simplified model to understand disease spread. Traditional epidemic models [2, 3] consider population and interactions between individuals. Kermack-McKendrick model [4] is the first model that provides a mathematical description of the kinetic transmission of an epidemic in an unstructured population. Webb [5] proposes and analyzes a model in a one-dimensional bounded space. The random diffusion governs spatial mobility with coefficients  $k_1$  and  $k_2$  for the susceptible and infected classes. May et al. [6] introduce SIR model and a fixed population is applied with three different variables;  $S(t)$ :susceptible,  $I(t)$ :infected, and  $R(t)$ :recovered. Various models are derived from the SIR model, such as SI and SIS [7], SEIR [8], modified SIR [9, 10], MSIR [11], and MSEIR [12]. However, these epidemic models are designed to estimate parameters for the diffusion patterns after disease outbreaks without considering the spatial analysis of disease spreading.

In addition to the epidemic models, various disease simulation research has been proposed. FluAid [13] is provided by the United States Department of Health and Human Services to assist state and local level planners in preparing for the next influenza pandemic. Carley et al. [14] propose a BioWar system with an agent-based model to present disease diffusion that encompasses certain exogenous factors, such as media information, geographic information, and weather information with social contact network and diffusion information. Chao et al. [15] present FluTE also with an agent-based model to simulate the disease spread across large populations.

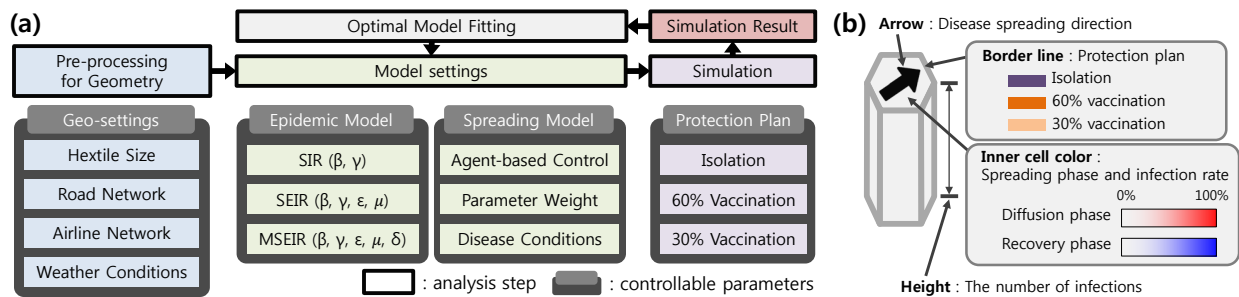


Figure 2. System overview. (a) presents system flow and controllable parameters. (b) is our visual encoding.

Mniszewski et al. [16] have developed EpiSims, which applies the regional population by an agent-based model and constructs social contact network from daily social activities. Balcan et al. [17] present an effect of mobility networks for the spatial spreading of infectious disease, but the mobility networks are connections between two destinations, not actual geographical networks. Vaccination [18] and school closure [19] are applied in the H1N1 influenza simulation to confirm their effectiveness in the mitigation of an influenza epidemic.

There has been lots of research on visual analytics for various data analysis. The primary purpose of the work is to synthesize information and derive insight from data [20]. Visual analytics enables users to analyze data efficiently. In the disease simulation domain, Guo [21] proposes a visual analytics approach to discover spatial patterns in pandemic data. In order to design and plan more effective containment strategies before pandemic diseases happen, Guo presents a new graph partitioning method that consists of a graph clustering, linear ordering, and matrix-based visualization. Ramanathan et al. [22] propose a verification method for the epidemiological model with a visual analytics approach. Afzal et al. [23] propose a visual analytics decision support environment for epidemic modeling and response evaluation.

AnyLogic [24] is a simulation tool that supports all of the most common simulation methodologies, such as System Dynamics, Process-centric, and Agent-Based modeling. STEM [25] is a framework and development tool designed to help scientists create and use spatial and temporal models of infectious disease. FRED [26] has been designed as a flexible, modular, open-source framework for epidemic modeling. GLEAMviz [27] has been developed to explore a more realistic epidemic spreading at a global scale. The system utilizes the global transportation network to simulate the disease spreading globally. Bryan et al. [28] present predictive analytics for spatiotemporal epidemic simulations with an emulator workflow and visual analytics interfaces.

### 3. Design Goals

There are a few systems for the disease spreading analysis, but they support only limited functionalities and visual representations. We have designed our system based on the following design goals for the system capabilities and visualizations:

- *The system should be able to express various information at once.* Figure 1 (a)-(III) shows the visualization map view for simulation. (IV) presents the graphs of infections, and (V) presents a detailed information for the selected hextile. Figure 2 (b) shows our visual encoding on the hextile. The total number of infections is encoded in the height of the cell. The spreading phase and infection rate are encoded in the inner cell color. The red color presents the infection diffusion phase, whereas the blue color indicates the recovery phase. In order to reveal the spreading directions, we represent the spreading directions.
- *The system should be able to compare different simulation outputs.* We have designed a history view in order to provide multiple simulation results in Figure 1 (a)-(VI). Once the simulation is completed, the infection rate graphs are moved to the history view to compare the multiple simulation results quickly.
- *An appropriate action should be taken based on the disease spread simulation.* A protection plan has been designed to provide suitable actions, including an isolation plan and a vaccination plan. Our system supports the protection plan simulation on the fly.
- *The system should be able to show the infection rates in specific regions.* Disease spreading patterns have geographical characteristics. Our system allows users to group interesting regions, and the system shows the infection rate changes for each group. The black graph in Figure 1 (IV) represents the total infection rate in the entire region, and color-graphs represent the infection rate of each group that classified by analyst.
- *The system should be capable of configuring various networks.* Our system uses the actual road networks for more accurate predictive disease spreading, and we obtain the actual network from the online map service.

**Table 1. Comparison of epidemic analytics systems among ① GLEAMviz, ② EpiSIM, ③ FRED, ④ STEM, and ⑤ AnyLogic. - O: Fully supported,  $\Delta$ : Partially supported, X: Not supported**

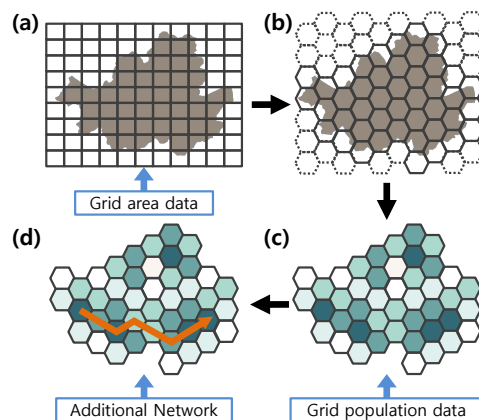
	①	②	③	④	⑤	Ours
Epidemic model	O	O	O	O	O	O
Real road network	O	X	X	X	X	O
Airline network	$\Delta$	X	X	$\Delta$	X	O
Disease direction	X	X	X	X	X	O
Protection plan	X	$\Delta$	X	$\Delta$	X	O

## 4. System Overview

Our system provides an interactive multimodal epidemic visual analytics and modeling framework allowing users to simulate and model disease spreading patterns. The system flow is represented in Figure 2 (a). The system consists of Data pre-processing, model settings, simulation, and optimal model fitting.

The target area is divided into hextiles using grid geographical data with a user-defined hextile size. In the pre-processing step, the user can apply additional information, including road networks, airline networks, and weather conditions. After applying a geographical setting, the epidemic model and spreading model are applied. When the user clicks a certain hextile, the selected hextile is infected, and the disease spreading starts. The protection plan provides isolation, 60% vaccination, and 30% vaccination. The optimal model fitting is the way to analyze disease from the spreading data. Since the combination of epidemic models and network structures is too complex and heuristic to set optimal epidemic parameters, our optimal model fitting enables the users to discover the optimal parameters in the epidemic and spreading models.

Table 1 presents the functional comparison of existing epidemic analytics systems. It is challenging to analyze the disease spreading patterns using the epidemic model only due to the complexity of epidemic patterns. Some research applies an additional network (GLEAMviz and STEM) and a protection plan (EpiSIM and STEM) to complement this limitation. To analyze the disease spreading pattern in a visual analytics system, the system should allow the users to apply complex networks, utilize simple visualization and transmit information efficiently for the analysis, and enable various parameters with feedback interactions. Most epidemic analysis systems partially contain these capabilities. However, it is difficult to analyze the spreading patterns without full functionalities. Our system provides all features needed for the comprehensive disease spreading analysis.



**Figure 3. Geometric settings. (a) presents grids of grid area data. (b) is the hextile mapping after grid mapping. The population map is applied in (c). Then, the user can apply additional network as shown in (d).**

## 5. Models for Disease Analysis

The real epidemic is affected by various conditions, including scientifically unknown factors. Epidemiologists investigate disease outbreak, disease spread, monitoring, and treatments. However, most research is focused only on discovering disease characteristics, disease models, or treatment methods separately. Even the research on disease spreading does not provide enough insight due to the ambiguous simulation systems. In this section, we propose a disease spreading model combining the spreading model and a pseudo-agent-based model as an extension to the epidemic model.

### 5.1. Geometric Settings

Many studies of geometric analysis use tessellations with small polygons, such as triangles, rectangles, and circles. Despite many types of polygons, we use the hexagonal polygons in our system. The advantages of the hextile mapping are as follows: 1)it is possible to cover the entire map without any empty space, whereas, other geometry mappings, such as circle and octagon, generate uncovered empty spaces; 2)there are many adjacent neighbors, and it is possible to transmit the disease toward various directions compared to the triangle and rectangle.

We apply the geographical information with the grid population and the grid area data produced by the SEDEC(Socioeconomic data and application center) data of NASA. The grid population and area data are formatted in the regular grid with fixed-size cells, and each cell has latitude and longitude information at the upper-left point, as shown in Figure 3 (a). When the user

selects an area, our system calculates the latitudes and longitudes that cover the selected area. Then, the system generates hextiles with the user-defined hextile size, as presented in Figure 3 (b). The population information for each hextile is computed utilizing the grid population and area data from SEDEC in Figure 3 (c). After applying geographical information, we connect the adjacent network in the data structure. Once the hextile map is built for a region, we apply the additional network conditions, as shown in Figure 3 (d).

## 5.2. Epidemic Model

Among the studies on the disease spread analysis, Bernoulli proposed a mathematical model for inoculating against smallpox [29] in 1760, which preceded modern epidemiology. The most famous epidemic model is SIR proposed by Kermack and McKendrick [4] in 1927. People in the area are classified as susceptible, infected, and recovered. **SIR model** is described as a differential equation in the following.  $\frac{dS}{dt} = -\beta \cdot S \cdot I$ ,  $\frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I$ ,  $\frac{dR}{dt} = \gamma \cdot I$ , where  $\beta$  is the parameter controlling how often a susceptible-infected contact results in a new exposure and  $\gamma$  is the rate of moving into the recovered phase from the infected status. **SEIR model** has an additional parameter, exposed or latent period of the disease ( $E(t)$ ), on top of the SIR model. The SEIR model is presented as follows:  $\frac{dS}{dt} = B - \beta \cdot S \cdot I - \mu \cdot S$ ,  $\frac{dE}{dt} = \beta \cdot S \cdot I - \epsilon \cdot E - \mu \cdot E$ ,  $\frac{dI}{dt} = \epsilon \cdot E - \gamma \cdot I - \mu \cdot I$ ,  $\frac{dR}{dt} = \gamma \cdot I - \mu \cdot R$ , where  $\epsilon$  is the rate at which an exposed person becomes infected and  $\mu$  is the natural mortality rate unrelated to the disease. **MSEIR model** includes a natural-born passive immunity as an additional parameter,  $M(t)$ , on top of SEIR model. MSEIR model is formulated as follows:  $\frac{dM}{dt} = B - \delta \cdot M - \mu \cdot M$ ,  $\frac{dS}{dt} = \delta \cdot M - \beta \cdot S \cdot I - \mu \cdot S$ , where  $\delta$  indicates the average temporary immunity period.  $\frac{dE}{dt}$ ,  $\frac{dI}{dt}$ , and  $\frac{dR}{dt}$  are same as ones in SEIR model. Our system utilizes these three epidemic models for simulation and modeling of disease diffusion. The epidemic models are computed for each hextile cell containing the population information and the geographical information.

## 5.3. Spreading Model

The epidemic model presented in the previous section is a mathematical model under certain ideal conditions, whereas many other factors control the real epidemic. When an outbreak of a new unknown pandemic occurs, researchers do not have

**Table 2. Controllable parameters in our models**

	Parameters	Description
Geometry	Hextile size	Geographical aggregation level
Epidemic model	Types	Selectable epidemic model using SIR, SEIR, MSEIR
	$\beta$	Contact rate
	$1/\gamma$	Average infectious period
	$1/\epsilon$	Average latent period
	$\mu$	Average death rate
Spreading weight	$\delta$	Average immunity period
	Weights	weight parameters for environments
Network	Road Airline	The additional information for infectious path
Protection plan	Isolation/ vaccination	The government actions to prevent the spread

any information on scientific attributes. Although parameters of a new model are complicated, there is a reasonably simple regulation for the disease propagation, which is affected by the adjacent areas containing infected people. In this work, we propose a spreading model to determine the diffusion rules and probabilities.

We denote the current area as  $i$  and adjacent area as  $j$ . Given time  $t$ , the infection degree in the current area  $d_{(i,t)}$  is determined as follows:

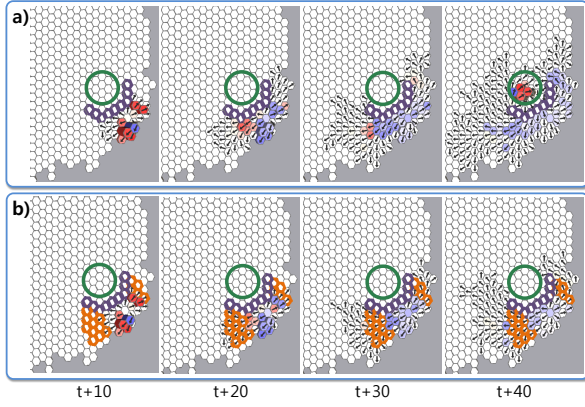
$$d_{(i,t)} = \sum_j F_{EPI} (c_{(i,t-1)} \cdot (d_{(i,t-1)} + s_{(i,t-1)}))$$

$F_{EPI}$  is a certain epidemic model, such as SIR, SEIR, and MSEIR as presented in Section 5.2. The infection in the current area is related to the disease spreading in the current area at the previous time,  $d_{(i,t-1)}$ , and the disease spreading from an adjacent area at the previous time,  $s_{(i,t-1)}$ . Also the inherited disease characteristics,  $c_{(i,t)}$ , of a disease affect the infection degree. The characteristic of the inherited disease is defined as follows:  $c_{(i,t)} = r_{cond} \cdot w_{cond(i,t)}$ .

where  $r_{cond}$  is a user-defined local rate in a specific area, and  $w_{cond}$  is a weight for the spreading condition. Generally, a disease is diffused via a specific medium, such as mite, mosquito, and virus. When these kinds of medium exist densely, the weight,  $w_{cond}$ , becomes large. As another condition, if climate conditions, such as temperature and humidity, are suitable, a disease is diffused faster.  $w_{cond}$  is related to the disease condition associated with environmental factors for the spread of disease.

The weight of the the spreading is defined as a multiplication of a probability and spreading weights from the adjacent areas as presented in the following:

$$s_{(i,t)} = \sum_{j=1}^6 P_{(i,j)} \cdot d_{(j,t)} \cdot (r_{net(i)} \cdot w_{net(j,t)} + r_{vec(i)} \cdot w_{vec(j,t)})$$



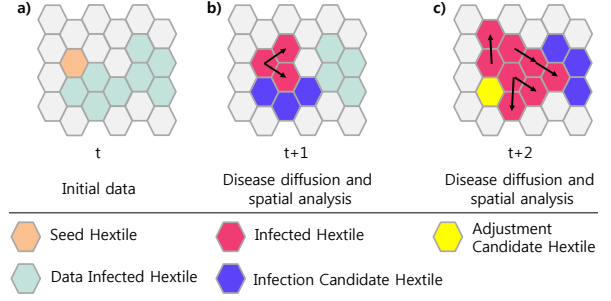
**Figure 4. Comparison of the disease diffusion patterns for the isolation plan only with the additional vaccination plan.**

where  $j$  is an adjacent area associated with spatial adjacency and discursive network connections, such as road, subway, and airline network.  $P_{(i,j)}$  is a probability of the user-defined parameters.  $r_{net}$  and  $r_{vec}$  are the user-defined local rates for the network weight,  $w_{net}$ , and the vector field weight,  $w_{vec}$ , respectively.  $w_{net}$  is set for each network since each network characteristic is different. For example, a road network has a speed limit, a subway network has hourly usages, and an airline network has departure and destination information.  $w_{vec}$  refers to the vector field information, such as the wind information and the infected carrier information.

The system performs a simulation to predict the disease spreading once the model is determined. The dispersion direction is defined by combining the parameters in the adjacent cells and additional network settings. Assuming that the local cell and adjacent cell are indexed as  $i$  and  $j$ , respectively, the diffusing direction is obtained by the probability of transmission,  $P_{(i,j)}$ , as presented in the following:

$$P_{(i,j)} = \frac{w_1}{6} \cdot \sum_{s=1}^6 \frac{\alpha_s \cdot e^{-n_{(i,j)}}}{n_{(i,j)}} + w_2 \cdot |\vec{W}_{i \rightarrow j}| \cdot |\vec{L}_{i \rightarrow j}| + w_3 \cdot \frac{e^{-(T_i - R_T)^2}}{\sqrt{2\pi}} + w_4 \cdot \frac{e^{-(H_i - R_H)^2}}{\sqrt{2\pi}}$$

$w$  is the weight for each transmission parameter,  $n_{(i,j)}$  is the sum of population in  $(i, j)$ ,  $\alpha_s$  is the population level,  $W_{i \rightarrow j}$  and  $L_{i \rightarrow j}$  represent the wind vector and the location vector, respectively.  $(T_i - R_T)$  and  $(H_i - R_H)$  are the results by subtracting the reference values from the measured values. In other words, these represent appropriate degrees of temperature and humidity in  $i$  area. The direction in a hextile is computed as the average diffusion directions for the neighbors transmitting to the hextile. The diffusion flow fields vary along the simulation time, and users can analyze how the disease spreads in the space.



**Figure 5. The optimal network model.**

## 5.4. Pseudo-agent-based Model

The epidemic model has an assumption that there is no disease spread and population movement out of the analysis area. In order to develop more flexible simulations, our system combines the spreading model with an agent-based model. The original agent-based model consists of dynamically interacting rule-based agents, and each agent indicates each person in the society commonly. However, it is difficult to run the original agent-based model due to the prodigious computational power requirement. To overcome this limitation, our system utilizes a pseudo-agent-based model that can be applied to the hextile-level agent instead of the people-level agent. The population in each hextile is computed by the population exchange between neighboring hextiles in every system clock. This interchange exerts stochastic effects on the spreading factor, such as population, wind, temperature, and humidity. Due to the population exchange, the pseudo-agent-based model can determine the spreading patterns more appropriately by considering the population movements that the traditional spreading model cannot accommodate.

## 5.5. Diffusion Model and Additional Network

Diffusion model parameters and additional network parameters can be manipulated on the fly. Table 2 shows a summary of the adjustable parameters. The diffusion model parameters consist of epidemic model parameters and spreading model parameters. In the case of the epidemic model, the epidemic model parameters can be adjusted by pressing the corresponding epidemic tab. Spreading parameters affect disease diffusion probability and velocity. Every weight of spreading parameters are considered as the normalized rate that has a range (0,1). The additional network consists of the flight network and the road network. In the case of the road network, the real road information is obtained from the web-based map API, such as google maps API and mapbox API.

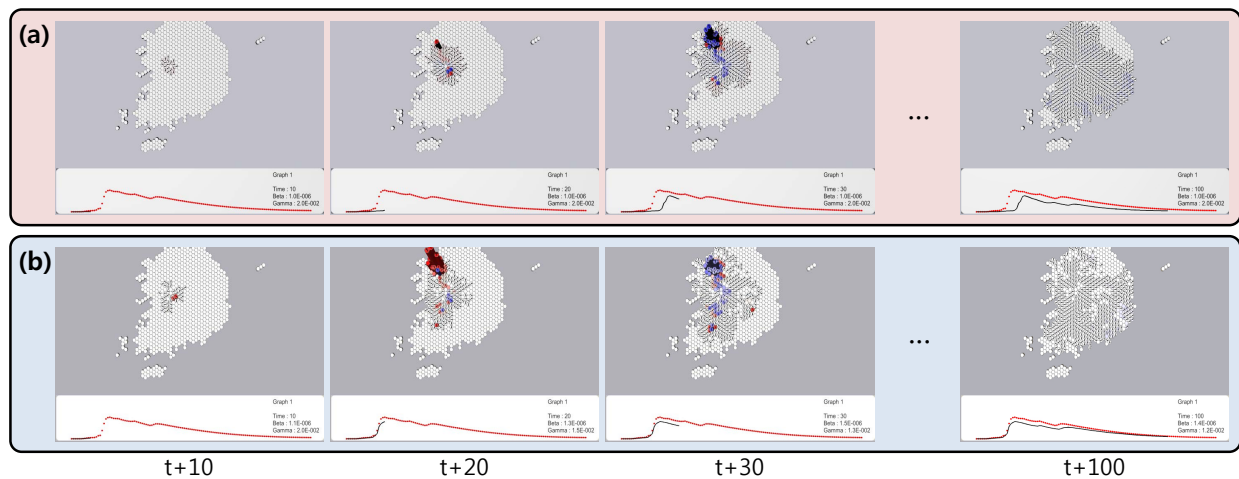


Figure 6. The result of the optimal modeling using a synthetic data.

## 5.6. Protection Plan

Epidemic model and spreading model can be more realistic to represent the disease diffusion patterns by removing the disease spread risk, such as quarantine, vaccination, and isolated area. In order to reflect these disease risk management, we add a protection model to the simulation. Our system provides two different protection plans, isolation, and vaccination. To add the protection plans, users select a protection type then drag the mouse on the map to select the hextiles for the protection. An example of our protection plan simulation is shown in Figure 4. The isolation is marked as purple hexagons, and any disease can not be transmitted through the isolated hexagons, as seen in (a). The isolation is to prevent the disease from spreading by regional blockades. The blockade regions have no network to infiltrate the disease to their neighbors, which indicates that  $w_n$  is set to zero. The other protection plan is to vaccinate regions, which increases the disease resistance and recovery rate. The vaccination marked as orange hexagons in Figure 4 (b) can also be applied by mouse dragging on the map, and the ratio of vaccinated people can be adjusted. We use 60% and 30% of vaccination rates. The effect of the vaccination is to mitigate disease infection and transmission. Since we apply the incubation period and recovery ratio, if the higher vaccination ratio is provided, the entire people in the vaccinated areas could be recovered entirely quickly before they transmit the disease to the neighbors, as seen in Figure 4 (b). When the vaccination plan is activated in a region,  $\beta$  decreases, and  $\gamma$  increases in the epidemic models introduced in Section 5.2. As shown in Figure 4, the diffusion through the vaccinated areas becomes slow or null in the end; therefore, our protection plans are adequate to slow down the disease diffusion.

## 5.7. Optimal Modeling

Researchers publish the epidemic parameters for the known disease after studying the diffusion patterns. The parameters can be easily applied to the simulation for the future diffusion patterns. However, when a new disease outbreaks, it is not easy to discover the epidemic parameters since there are so many complex factors. To aid researchers, we develop an algorithm for optimal modeling based on the data collected from patients and hospitals. Our optimal modeling algorithms adjust to the epidemic model and the network model. The optimal epidemic model allows us to obtain epidemic parameters, including  $I$ ,  $R$ ,  $\beta$ , and  $\gamma$ . The optimal network model is designed to discover additional infection carriers, such as wind, car, and airplane. In the optimal epidemic model, our system, first, reads the data and computes references of infection ratio and recover ratio as  $Inf_{ref}$  and  $Recv_{ref}$  from  $I_{data}/Population$  and  $R_{data}/Population$ , respectively. Similarly, infection and recover ratio of current hextile are calculated as  $Inf_{cur}$  and  $Recv_{cur}$ . Then, our system compared between reference and current. Our system increases and decreases of  $I$  and  $R$  based on the difference between the reference and current parameters. Optimal training is to adjust the parameters in the epidemic model by comparing it with the infection time, degree, and area in the given data. For example, the  $\beta$  and  $\gamma$  in SIR model are adjusted for the appropriate values according to the data fitting. If the overestimate or underestimate of infection area occurs, the system adjusts the network parameter. For example, when the infection rate does not vary during the several time steps in the expected infection region, the system proceeds to set the infection in the region during the simulation.

Figure 5 illustrates our optimal network model from time  $t$  to  $t + 2$ . The orange hextile in Figure 5 (a) is the first seed area where the disease diffusion starts, and the mint color hextile indicates the infected area marked in the input data. At the time step  $t + 1$ , the disease spreads to the adjacent areas from the seed area by following the diffusing algorithm introduced in Section 5.3. The red color hextile represents the infected area, whereas, the purple color hextile is the infection candidate area. The infection candidate areas indicate the tiles infected at the time step  $t + 1$  in the data. If the infection candidate tiles are not infected during several time steps, these tiles are classified as adjustment candidate hextiles. The adjustment candidate tiles are marked as the yellow color tile in Figure 5 (c). The infection candidate hextiles at  $t + 2$  are marked as the same way as mentioned before. In this way, the simulation is corrected and adjusted with the optimal network model by comparing the actual data and simulation results. We apply the optimal modeling to the synthetic data in order to validate our technique. The synthetic data has complicated characteristics since the data is obtained with SEIR model with various additional network parameters and a few protection plans. We have tested the optimal model with SIR model. Figure 6 presents the result from the optimal modeling. The red graph is the reference infection ratio in the data, and the black graph is the result of our optimal modeling. A set of Figure 6 (a) present the simulation results without the optimal modeling. Figure 6 (b) presents the simulation results with the optimal modeling. The different dominant result is found in the infection time. Disease dispersion time in (b) is faster than in (a). Another difference is seen in the infection rate over time. Our optimal modeling provides more accurate simulation results compared to the original data.

## 6. Evaluation and Discussion

We have tested the system on Intel Core i7-4790 with 16GB RAM and NVidia GeForce GTX 960. Our epidemic simulation and modeling are computed and visualized in real-time, and user interactions can be applied on the fly during the simulation. We utilize the map information for the road network using Mapbox API. For the hextile interaction, we create the box collider and apply the raycasting algorithm. When the mouse interaction occurs, the system checks the box collider using the raycasting. Our system uses a 2D list type data structure containing location, population, epidemic model parameters, adjacent cell list, infection state, dispersion direction, ratio, and weight. In the following sections, we present two different simulations

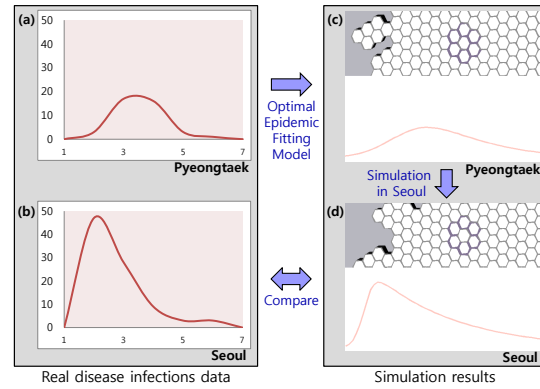


Figure 7. The result of MERS disease analysis.

to evaluate our system and algorithm. We apply our technique to the real statistical data in MERS and FMD.

### 6.1. MERS Disease Analysis

The respiratory disease, MERS, first confirmed on May 20, 2015, in South Korea, was brought into the country by a person who visited the Middle East, and MERS resulted in 186 confirmed cases and 38 deaths until January 26, 2016. More than 70 cases have been traced back to the hospital. It is believed that the patients picked up the virus from an infected person who waited for days in various parts of the emergency ward, potentially exposing the virus to an estimated 900 staff, patients, and visitors. Korea immediately isolated the confirmed patient and associated persons to prevent the spread of infection [30]. We analyze the disease diffusion with the MERS data using our optimal model to approximate epidemic parameters. Then, we apply the estimated parameters to another region in order to predict disease diffusion. When a certain unknown disease occurs, this analysis helps the analysts predict disease dispersion quickly and determine proper actions to prevent the disease spread. Figure 7 (a) shows the number of infected patients in a region of Pyeongtaek between May 20 and 26, 2015. Figure 7 (b) presents the number of patients infected in a region of Seoul between June 4 and 10, 2015. First, we cut off the road network using the isolation protection plan in order to make isolate condition. We then let our system learn the epidemic parameters using our optimal epidemic model with the data in the region of Pyeongtaek. We simulated our model by setting the infected seed in the region of Seoul. Figure 7 (c) presents the infection pattern using our optimal epidemic model. Figure 7 (d) presents the infection patterns in the region of Seoul using the parameters obtained from the cases in the region of Pyeongtaek. Although the infection rates seem slightly different between the real data and simulation results



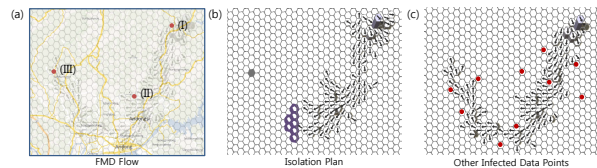
due to the various other factors that we do not consider in our model, the simulation results are very similar to the original data. This indicates that our model can capture the disease properties as the parameters and the simulations are correctly computed to represent the real data.

## 6.2. Foot and Mouth Disease Analysis

Foot-and-mouth disease or hoof-and-mouth disease (Aphthae epizootic) is an infectious and sometimes fatal viral disease that affects cloven-hoofed animals. Foot-and-mouth disease (FMD) has severe implications for animal farming. It is highly infectious and can be spread by infected animals through aerosols, through contact with contaminated farming equipment, vehicles, clothing, or feed by domestic and wild predators [31]. Since the spreading occurs dominantly along with the vehicle and human movements, we apply the real road network to disease-spreading simulations. We predict the disease dispersion when there are infection carriers on the road and, then, we simulate the suitable protection plans to prevent the disease from spreading. To evaluate the road network model and the protection plans, we used the FMD data in Andong, South Korea, between November 28 and December 4, 2010. The data consists of infected time, location information, livestock information, and farm size. The total number of foot-and-mouth disease cases is 38. First, we apply the filter for cattle farms with more than 4000 in order to sample events. Figure 8 (a) shows the map with three large cattle farms. In order to analyze the FMD data, we apply road networks (yellow lines), as shown in Figure 8 (a). Then, we set the disease seed in the farm (I) and simulate the disease spreading model. Figure 8 (b) is the simulation result with the protection plan to prevent further diffusion. Figure 8 (c) is the simulation result indicating that FMD mostly spreads along with the road networks. The number of FMD cases occurring during the simulation is 27, which is 71.05% out of total cases along the path. FMD outbreaks are found in the farm (I) on November 28, in the farm (II) on December 1, and in the farm (III) on December 7. In our simulation, FMD is observed in the farm (II) at the time step 10, and FMD reached the farm (III) at the time step 24. The temporal and spatial spreading ratios are similar between the real data and the simulation results. This indicates that our model can well predict the disease spreading spatially and temporally.

## 7. Conclusion

In this paper, we have proposed an interactive multimodal epidemic visual analytics and modeling



**Figure 8. The result of Foot and Mouth Disease (FMD) analysis.**

system. To improve the conventional epidemic model, we have defined a spreading model and a protection model, and these models were incorporated with the epidemic model. The spreading model was developed with environmental conditions, such as population, transportation, weather condition, and transmission rate. The protection model was used to isolate or vaccinate in certain areas. Our system has been designed as an interactive tool for the simulation and modeling the disease diffusion based on data. To evaluate our development, we apply real statistical data, MERS and FMD, in the system. However, our system has some limitations. In order to apply the optimal model, a pre-processing, such as data conversion, is required since most of the data are geo-located as districts instead of exact latitude and longitude. This limitation hinders the ease of system usage. We plan to implement an automatic data load module converting districts to geographic coordinates. Since we use random numbers in the disease spreading algorithm, the slightly different results are obtained sometimes. We are currently developing a web-based system to share our work. We plan to add the COVID-19 dataset to our web system. As future work, we will study the uncertainty visualization for the simulation results or apply Monte Carlo simulation for the spreading. Currently, the optimal parameter setting focuses only on the disease spreading speed; therefore, we plan to develop an algorithm for the optimal parameters to extract the network or vector field without additional inputs.

## Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2019-0-00242, Development of a Big Data Augmented Analysis Profiling Platform for Maximizing Reliability and Utilization of Big Data) and (2019-0-00136, Development of AI-Convergence Technologies for Smart City Industry Productivity Innovation), and (2019-0-00795, Development of integrated cross-model data processing platform supporting a unified analysis of various big data models). Yun Jang is the corresponding author.

## References

- [1] “WHO pandemic phase descriptions and main actions by phase.” [https://www.who.int/influenza/resources/documents/pandemic\\_phase\\_descriptions\\_and\\_actions.pdf](https://www.who.int/influenza/resources/documents/pandemic_phase_descriptions_and_actions.pdf), accessed on November 15, 2020.
- [2] R. M. Anderson, R. M. May, and B. Anderson, *Infectious diseases of humans: dynamics and control*, vol. 28. 1992.
- [3] J. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, vol. 5. John Wiley & Sons, 2000.
- [4] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 115, no. 772, pp. 700–721, 1927.
- [5] G. Webb, “A reaction-diffusion model for a deterministic diffusive epidemic,” *Journal of Mathematical Analysis and Applications*, vol. 84, no. 1, pp. 150–161, 1981.
- [6] R. M. May and R. M. Anderson, “Population biology of infectious diseases: Part ii,” *Nature*, vol. 280, no. 5722, pp. 455–461, 1979.
- [7] L. J. Allen, “Some discrete-time SI, SIR, and SIS epidemic models,” *Math. Biosciences*, vol. 124, no. 1, pp. 83–105, 1994.
- [8] M. Y. Li and J. S. Muldowney, “Global stability for the seir model in epidemiology,” *Mathematical biosciences*, vol. 125, no. 2, pp. 155–164, 1995.
- [9] F. A. Milner and R. Zhao, “SIR model with directed spatial diffusion,” *Math Popul Stud*, vol. 15, no. 3, pp. 160–181, 2008.
- [10] F. Zhang, Z.-z. Li, and F. Zhang, “Global stability of an SIR epidemic model with constant infectious period,” *Applied Mathematics and Computation*, vol. 199, no. 1, pp. 285–291, 2008.
- [11] L. Scrucca, “Model-based SIR for dimension reduction,” *Comput. Stats. & Data Analysis*, vol. 55, no. 11, pp. 3010–3026, 2011.
- [12] H. Inaba, “Age-structured homogeneous epidemic systems with application to the mseir epidemic model,” *Journal of mathematical biology*, vol. 54, no. 1, pp. 101–146, 2007.
- [13] CDC, “Availability of influenza pandemic preparedness planning fluaid, 2.0,” *J. of the American Medical Assoc.*, vol. 284, no. 14, 2000.
- [14] K. Carley, D. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave, “Biowar: scalable agent-based model of bioattacks,” *IEEE Trans. on Sys., Man and Cyb.*, vol. 36, no. 2, pp. 252–265, 2006.
- [15] D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. Longini, Jr, “Flute, a publicly available stochastic influenza epidemic simulation model,” *PLoS Comput Biol*, vol. 6, 01 2010.
- [16] S. M. Mniszewski, S. Y. Del Valle, P. D. Stroud, J. M. Riese, and S. J. Sydoriak, “Episims simulation of a multi-component strategy for pandemic influenza,” in *Spring Simulation Multi-conference*, pp. 556–563, 2008.
- [17] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, “Multiscale mobility networks and the spatial spreading of infectious diseases,” *National Academy of Sciences*, vol. 106, pp. 21484–21489, Dec. 2009.
- [18] P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Gonçalves, J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. V. den Broeck, V. Colizza, and A. Vespignani, “Modeling vaccination campaigns and the fall/winter 2009 activity of the new a(H1N1) influenza in the northern hemisphere,” *Emerging Health Threats Journal*, vol. 2, 2011.
- [19] B. Lee, S. Brown, P. Cooley, M. Potter, W. Wheaton, R. Voorhees, S. Stebbins, J. Grefenstette, S. Zimmer, R. Zimmerman, T. Assi, R. Bailey, D. Wagener, and D. Burke, “Simulating school closure strategies to mitigate an influenza epidemic,” *J. of public health man. and prac.*, vol. 16, no. 3, pp. 252–261, 2010.
- [20] N. Andrienko and G. Andrienko, “Visual analytics of movement: An overview of methods, tools and procedures,” *Inf. Vis.*, vol. 12, no. 1, pp. 3–24, 2013.
- [21] D. Guo, “Visual analytics of spatial interaction patterns for pandemic decision support,” *Int. J. Geogr. Inf. Sci.*, vol. 21, no. 8, p. 859, 2007.
- [22] A. Ramanathan, C. Steed, and L. Pullum, “Verification of compartmental epidemiological models using metamorphic testing, model checking and visual analytics,” in *BioMedical Computing*, pp. 68–73, 2012.
- [23] S. Afzal, R. Maciejewski, and D. Ebert, “Visual analytics decision support environment for epidemic modeling and response evaluation,” in *2011 IEEE Conf. on Visual Analytics Science and Technology*, pp. 191–200, 2011.
- [24] “Anylogic.” <http://www.anylogic.com>, accessed on November 15, 2020.
- [25] S. B. Edlund, M. A. Davis, and J. H. Kaufman, “The spatiotemporal epidemiological modeler,” in *Proceedings of the ACM Int. Health Inf. Sym.*, pp. 817–820, 2010.
- [26] J. J. Grefenstette, S. T. Brown, R. Rosenfeld, J. DePasse, N. T. Stone, P. C. Cooley, W. D. Wheaton, A. Fyshe, D. D. Galloway, A. Sriram, *et al.*, “Fred (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations,” *BMC public health*, vol. 13, no. 1, p. 940, 2013.
- [27] W. Van den Broeck, C. Gioannini, B. Gonçalves, M. Quagiotto, V. Colizza, and A. Vespignani, “The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale,” *BMC Infectious Diseases*, vol. 11, no. 37, 2011.
- [28] C. Bryan, X. Wu, S. Mniszewski, and K.-L. Ma, “Integrating predictive analytics into a spatiotemporal epidemic simulation,” in *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, pp. 17–24, 2015.
- [29] D. Bernoulli, “Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir,” *Mem. Math. Phys. Acad. R. Sci.*, pp. 1–45, 1766.
- [30] “MERS.” [https://en.wikipedia.org/wiki/Middle\\_East\\_respiratory\\_syndrome\\_coronavirus](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome_coronavirus), accessed on November 15, 2020.
- [31] “Foot-and-mouth disease.” [https://en.wikipedia.org/wiki/Foot-and-mouth\\_disease](https://en.wikipedia.org/wiki/Foot-and-mouth_disease), accessed on November 15, 2020.