

Epidemic Vulnerability Index for Effective Vaccine Distribution Against Pandemic

Hunmin Lee ¹, Mingon Kang ², Donghyun Kim ³, Daehee Seo ⁴, and Yingshu Li ⁵

Abstract—COVID-19 vaccine distribution route directly impacts the community's mortality and infection rate. Therefore, optimal vaccination dissemination would appreciably lower the death and infection rates. This paper proposes the Epidemic Vulnerability Index (EVI) that quantitatively evaluates the subject's potential risk. Our primary aim for the suggested index is to diminish both infection rate and death rate efficiently. EVI was accordingly designed with clinical factors determining the mortality and social factors incorporating the infection rate. Through statistical COVID-19 patient dataset analysis and social network analysis with an agent-based model that is analogous to a real-world system, we define and experimentally validate the capability of EVI. Our experiments consist of nine vaccination distribution scenarios, including existing indexes which estimate the risk and stochastically proliferate the contagion and vaccine in a 300,000 agent-based graph network. We compared the outcome and variation of the three metrics in the experiments: infection case, death case, and death rate. Through this assessment, vaccination by the descending order of EVI has shown to have a significant outcome with an average of 5.0% lower infection cases, 9.4% lower death cases, and 3.5% lower death rate than other vaccine distribution routes.

Index Terms—COVID-19, epidemic vulnerability index, agent-based model, clinical and social data analysis, simulation-based inference, statistical proliferation estimation, vaccine distribution strategy

1 INTRODUCTION

CORONAVIRUS disease (COVID-19) is a novel pandemic infected through the SARS-CoV-2 virus starting in late 2019 [1], [2]. Multiple types of variants have been propagated by mutation inside the host, rapidly escalating the infection and death cases in many countries. Vaccines have been developed and inoculated throughout the world, but several nations suffer from the increasing demand and the limited supply. The distribution strategy of the vaccination is critical, for its effective propagation route will significantly decrease the casualties, including death and infection cases. On the contrary, an improper vaccination dispatch strategy would extend the virus dissemination, increasing the damage.

In this paper, we suggest a novel index that estimates the potential risk of the subject based on the clinical and social factors, which shows to have significant interconnection with their mortality and infection rate, respectively. Clinical factors are the interior features of the subject that have been observed to determine the intensity of disease exacerbation,

such as types of underlying disease, age, physical vulnerability, etc. In contrast, the social factor is an external property that affects the infection rate. By assessing the importance of the subject among its affiliated community, it offers the information to quantify the potential risk of being infected. We compute this through Social Network Analysis (SNA) and centrality among the real-world based social network structure, which explains the influence of the subject among the associated neighborhood when it is infected. We collect and analyze the statistical dataset involving past COVID-19 patients and social network datasets to analyze external and internal factors.

This study proposes the novel Epidemic Vulnerability Index (EVI), representing the overall potential risk assessment. We utilize the EVI to discover the effective vaccination route among the social network. The possibility of applying EVI as a criterion to denote the necessity of vaccination is reasonable because it is specifically designed to estimate the internal and external risk for the subject, which incorporates its property and heterogeneous circumstances. We construct the Agent-Based Model (ABM) consisting of 300,000 agents with a directed graph structure for validation. Each node contains certain features such as age, comorbidities, gender, locational factors, centrality, etc. The node represents an individual subject, and the edges illustrate the physical contacts within the associated network. The internal properties of the node and the edges are allocated via real-world statistics such as physical contact per age, comorbidity ratio, age group, and gender population. Implementing the established ABM, we conduct nine vaccine distribution simulations based on existing criteria or index that measures the subject's risk and compare the numeric results of the three metrics: infection cases, death

- Hunmin Lee is with the Department of Computer Science and Engineering, University of Minnesota, Twin Cities, MN 55455 USA. E-mail: lee03915@umn.edu.
- Mingon Kang is with the Department of Computer Science, University of Nevada, Las Vegas, NV 89154 USA. E-mail: Mingon.Kang@unlv.edu.
- Donghyun Kim and Yingshu Li are with the Department of Computer Science, Georgia State University, Atlanta, GA 30302, Georgia. E-mail: donghyun.david.kim@gmail.com, yili@gsu.edu.
- Daehee Seo is with the Department of Computer Engineering, Sangmyung University, Seoul 31006, Republic of Korea. E-mail: daehee@smu.ac.kr.

Manuscript received 19 February 2022; revised 3 July 2022; accepted 7 August 2022. Date of publication 15 August 2022; date of current version 26 December 2023.

(Corresponding author: Hunmin Lee.)

Digital Object Identifier no. 10.1109/TCBB.2022.3198365

cases, and death rates after a virus and vaccines were proliferated within a certain period of time.

The major contributions of this paper are as follows:

1. We propose the EVI that quantitatively evaluates the risk of mortality and infection rate based on the subject's clinical and social factors after multi-perspective analyses. EVI could be optimized by tuning the parameters in suitable scenarios and objectives accordingly.
2. We design the probabilistic simulation algorithm operated in the ABM that was built analogous to real-world conditions. Virus and vaccine proliferation in the ABM in diverse settings, such as initial spreaders, regional-based propagation, stochastic infection rate, etc., offers inference and explainability from statistical estimation to predictive analysis.
3. We simulate the nine vaccine distributions among the ABM and compare the corresponding dissemination effect through statistics of infection cases, death cases, and death rate, validating the efficiency of EVI.

Estimating the propagation of infectious disease and vaccination is a complex problem dynamically being affected by miscellaneous variables such as economic issues, vaccination effect, regionality, and political agenda. Not only is the impact randomized dispersion in terms of time, but it also varies in spatial aspects because locational environments that influence the infection have their unique property. Thus, selecting a successful distribution route in given circumstances is an elusive task. However, diverse feasible simulation models on real-world based population sets provide us with predictive cases of currently viable response options. Therefore, intensive analysis and various empirical studies offer references and insights to devise optimal solutions. This research aims to derive the new aspect concerning distributions of COVID-19 vaccination, examining the influential factors to successfully minimize the infection and death rates.

The contents covered in this paper are organized as follows. Section 2 illustrates the existing works regarding risk evaluation on COVID-19 patients and our discriminative aspect compared to previously suggested studies. Section 3 measures the individual potential mortality based on the statistical, clinical dataset involving internal properties. Section 4 focuses on the infection rate, constructing the ABM analogous to the real-world system and analyzing the effect of the centrality types of individual subjects. Adopting the ABM, we select the best propagation criteria and validate the selection through experiments. Experiments are conducted in Section 5, where we suggest the vaccine dispersion simulation algorithm and compare the number of casualties and death rate with nine distribution scenarios. Finally, Section 6 concludes our study, explaining the availability and limitations.

2 RELATED WORKS

After global-scale pandemics such as severe acute respiratory syndrome (SARS), Ebola, and COVID-19, studies that estimate the potential risk were proposed along with auxiliary criteria [3], [4], [5], [6], [7], [8], [9], [10], [11] that accommodate the response strategy [12], [13], [14], [15], [16]. In a

similar context, corresponding medical supply (*e.g.*, vaccine, mask, self-test kits) distribution schemes and challenges [17], [18], [19], [20] were studied. The existing auxiliary indexes mainly estimate the risk primarily focused on the intrinsic biometric properties of the subject, accumulated from past patients who suffered from the pandemic or other diseases. This section explains past studies with respect to the vaccine distribution strategies and suggests indexes that quantify the patient's risk.

The vaccine is one of the encouraging options that we can effectively respond to contain the infectious disease among the penetrable host. However, the vaccine supply [16] is limited compared to the substantial demand, especially during the initial phase after the vaccine is invented. Therefore, locating an effective and efficient distribution trajectory is crucial after a multi-perspective analysis. Bubar et al. [12] suggested five vaccination distribution schemes analyzed based on the standard of age stratification and showed the coinciding result efficacy after vaccination simulations with SEIR [21] (Susceptible, Exposed, Infectious, Recovered) model. Our research is similar to what this manuscript presents, concerning the derivatives throughout the time of major factors such as infection, death, and recovered cases. Our simulation involves other significant properties of the subjects, especially social factors based on the thorough analysis, and more experimental cases were reflected compared to [21] to organize a realistic simulation framework. Matrajt et al. [22] analyzed the vaccination result on selected time periods based on distribution by age. It shows evidence-based guidance for vaccination from the perspective of the age group. For the optimal vaccination route, various empirical studies on the population that simulates the variation were illustrated. One frequently utilized scheme is the Agent-Based Model [23], [24], which is constructed of multiple individual agents that indicate the entity and their connection to other agents with the stochastic interaction. Kerr et al. [15] suggest COVID-19 agent-based simulator, covering several locational properties to approximate its propounded system into the real-world scenario. Silva et al. [25] implemented the seven settings in social activities within the regional community and observed the number of deaths along with the economic impact using their formulated ABM. Several studies targeted the specific region reflecting their certain characteristics, such as Gomez et al. [26] designed an ABM for the city of Bogotá, Columbia, and Hoertel et al. [27] for New York City, USA. On account of the intrinsic regional distinctive properties, the ABM simulated the transmission of COVID-19, offering future variations and predictability. Hinch et al. [28] suggested the ABM model that traces the age-stratified subject's contact based on U.K. statistics. Our study also involves a similar context to this work, for physical contact in age groups is one of the decisive statistics that has been validated [29]. The limitation of recent vaccination distribution manuscripts is that the analysis only focuses on the patient's age group. Most open-sourced COVID-19 datasets are aligned with the age and comparatively easy to preprocess. Although age does imply a significant correlation concerning mortality, more investigation could be found. This study mainly considers age-stratified, 22 types of comorbidities, gender, centrality, and locational factors in order to score

the danger level, then reflect the results for devising an optimal vaccine distribution route.

There have been instructive attempts to quantitatively estimate the entity's potential risk in given conditions from the perspective of individual conditions and geographical properties. These studies evaluated the current risk situation and offered a prognosis of future statistics to take successful countermeasures to stagnate the disease proliferation. The Centers for Disease Control (CDC) devised a Social Vulnerability Index (SVI) [5] for assessing the future impact of the disaster, as the socially vulnerable are highly prone to be adversely affected by natural disasters. SVI was validated through the statistical dataset of Hurricane Katrina-related casualties. Moreover, Frailty Index (FI) [9] suggests the vulnerability of a given subject, and it is especially established as a measure for elderly citizens to determine their physical and mental health. Studies are emerging regarding the relationships between the COVID-19 casualties and the FI [10], [11], and several recent studies have illustrated that FI is not sufficiently suitable for COVID-19 analysis. Decaprio et al. [4] proposed the C-19 index, which quantifies the vulnerability of individuals based on past respiratory disease patients, utilizing the XGBoost machine learning algorithm. Although the relationship between COVID-19 and existing respiratory diseases is shown to be positively proportional, the disuse of the COVID-19 patient database regards this as a proxy measure. Amram et al. [3] suggested the COVID-19 diagnosis rate in the counties of Washington State in the US. Their study visualized the county-level risk with geographical maps that offer intuitive risk scoring to compare. Similarly, Marvel et al. [6] presented a Pandemic Vulnerability Index (PVI) with geographical visualizations with a dashboard on county-level risk scores among given states of the US. PVI's risk scoring algorithm uses a Bayesian machine learning model with a county-level statistical COVID-19 patient dataset. Most propounded indexes are county-unit risks that evaluate the adverse effect among affiliated geographical units [30]. Apart from those risk criteria, other studies have been presented to estimate the economic and financial impact [8] of the pandemic, such as the Global Fear Index [7]. The existing studies that designed an index mostly estimated the internal risk of mortality based on the clinical dataset only and aggregated the statistics to express the general risk by geographical location boundaries. Our study considers the social factors and estimates the risk, including the clinical properties that the subject possesses. Based on the quantified risk, we validate our estimation in a real-world based ABM and compare the vaccination performance to existing indexes that present the credibility of EVI. This approach is to satisfy the ultimate objective that both infection rate and death rate must decline simultaneously.

3 MORTALITY AND CLINICAL FACTORS

The subject's mortality is directly affected by its internal conditions. This section mainly focuses on three influential factors to consider the mortality: age-stratified groups, underlying diseases, and gender. We define the subject's susceptibility based on the past COVID-19 patient

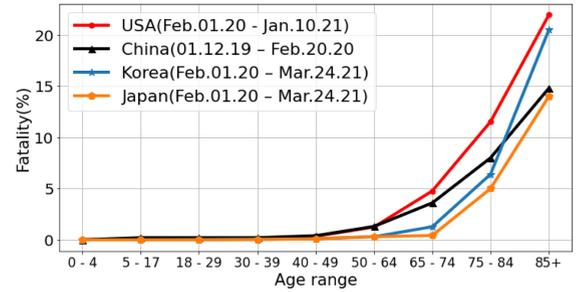


Fig. 1. COVID-19 Mortality by given age-stratified groups in corresponding countries by given period.

benchmark dataset. Let s_i be an individual subject where population S is expressed as $S = 0x22C3_{\forall i} s_i$, and we set the mortality of s_i as $M(\cdot)$, such that $M(s_{\exists i})0x225Dr/S$, with r indicating number of deaths and i indicates the index of the individual subject.

3.1 Age-stratified Groups and Gender in Mortality

Authoritative medical institutions such as CDC and WHO have declared that age and gender are crucial properties in assessing the risk [31], [32]. Statistically, subjects with higher age groups have significantly higher mortality, as shown in Fig. 1. Let age group $A = \{\alpha_k | 1 \leq k \leq 9, k \in \mathbb{N}\}$ where k refers to an index of age group, and $\alpha_{1 \leq k \leq 9}$ denotes age range between 0-4, 5-17, 18-29, 30-39, 40-49, 50-64, 65-74, 75-84, 85+ sequentially. $\alpha_{\exists k} = 0x22C3_{\forall i'} s_{i'}$ where $s_{i'} \in S$ is a subject that has the age range of corresponding α_k and i' denotes the arbitrary index among $0x22C3_{\forall i'} i$ where $1 \leq i' \leq \sum_{\forall i} 1$. $n(S) = n(A) = \sum_{\forall k} n(\alpha_k)$ and $M(s_{(k,i)}) = r_k/\alpha_k$ where $s_{(k,i)}$ indicates the subject of age range within α_k .

The mortality dataset that our team acquired is the COVID-19 open dataset that is publicly accessible, provided by the CDC [33] and NYTimes [34]. The dataset illustrates the statistical numbers of infection cases, death cases, demographic trends, age groups of the patient, each US state and county, comorbidities of the passed away patient, gender, and racial type, in a daily and weekly period. In order to discover practical information within the collected COVID-19 patient dataset, we process the dataset to generalize the parameters with the following approach. The accumulated number of COVID-19 confirmed cases in males and females as of March 31st, 2021 are 6,277,679 and 5,750,585, where we parametrize these values as I_F, I_M respectively. Statistics show that $I_F = I_M \cdot 1.09$, and $I_F : I_M = 52.2 : 47.8$, with 4.4% higher for I_F as the studies have specified in [31]. Based on conditional probability $P(r_G | (I_F 0x222AI_M))0x225DM(s_G)$ where G is a binary variable that indicates gender, where $G \in \{F, M\}$, s_G refers to the subject with corresponding sex and r_G denotes the number of deaths with given sex, the ratio of $M(s_F) : M(s_M) = 0.559 : 0.441$. The proportional difference is 0.5 ± 0.059 , and we denote the term $\pm c := P_G$ in (1). The mortality of the subject with the age index k and sex G is $M(s_{(k,G)}) = M(s_G) \cdot (P_G + 1)$, and we normalize the component with min-max normalization, between the range of [0,1]. In spite that the min-max normalization scheme is prone to be ineffective when outliers exist, it is suitable for our case since G is deterministic. Therefore, the statistical risk of $M(s_{(k,G)})$ is shown in (2).

TABLE 1
Comorbidity Types and Corresponding Number of Casualties in Each Age-stratified Groups

Comorbidity	Index	0-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
Influenza and pneumonia	C_1	175	853	2,174	6,220	15,295	25,749	30,258	30,239
Chronic lower respiratory diseases	C_2	36	88	203	594	2,334	5,577	7,551	6,749
Adult respiratory distress syndrome	C_3	86	328	895	2,527	5,582	8,367	7,667	5,659
Respiratory failure	C_4	153	613	1,582	4,726	11,840	21,324	25,811	24,738
Respiratory arrest	C_5	10	38	88	234	580	1,062	1,427	1,837
Other respiratory diseases	C_6	25	89	211	530	1,273	2,267	2,717	2,787
Hypertensive disease	C_7	25	163	667	2,261	6,375	11,807	14,966	17,828
Ischemic heart disease	C_8	5	38	154	692	2,640	6,085	9,088	10,804
Cardiac arrest	C_9	76	278	678	1,930	4,344	6,869	7,442	7,944
Cardiac arrhythmia	C_{10}	14	36	107	375	1,238	3,107	5,240	7,144
Heart failure	C_{11}	9	54	126	444	1,442	3,113	5,130	7,899
Cerebrovascular diseases	C_{12}	9	41	137	466	1,452	2,885	3,737	4,209
Other circulatory diseases	C_{13}	65	160	361	845	1,945	3,380	3,970	4,688
Sepsis	C_{14}	56	223	560	1,697	4,030	6,652	6,177	4,435
Malignant neoplasms	C_{15}	39	54	150	438	1,650	3,242	3,711	3,88
Diabetes	C_{16}	64	257	899	2,723	6,714	11,198	11,391	8,427
Obesity	C_{17}	123	430	960	1,670	2,418	2,446	1,321	358
Alzheimer disease	C_{18}	0	0	0	6	77	670	2,848	6,308
Vascular and unspecified dementia	C_{19}	0	1	2	31	413	2,605	8,548	17,068
Renal failure	C_{20}	29	180	463	1,388	3,426	5,891	6,249	5,574
Injury poisoning other events	C_{21}	90	262	337	550	1,103	1,759	2,225	3,017
Other conditions and causes	C_{22}	435	1,175	2,631	6,997	17,581	31,390	36,470	38,705

$$P_G := \frac{M(s_G)}{\sum_{\forall G} s_G} - 0.5 \tag{1}$$

$$M(s_{\exists k, G}) := \frac{M(s_{\exists k}) \cdot (1 + P_G) - \min(0x22C3_{\forall k} M(s_{\forall k}) \cdot (1 + P_G))}{\max(0x22C3_{\forall k} M(s_{\forall k}) \cdot (1 + P_G)) - \min(0x22C3_{\forall k} M(s_{\forall k}) \cdot (1 + P_G))} \tag{2}$$

3.2 Comorbidity in Mortality

Comorbidity is a critical factor that may intensify the patient’s condition and escalate mortality. Various past research [35], [36], [37], [38], [39] have asserted a strong correlation between mortality and the underlying disease, where statistics validate this claim. It shows that almost up to 90% of hospitalized COVID-19 patients suffered from the underlying illness [35] before the COVID-19 infection. The statistics indicate that the patients possess higher mortality if contracted diseases, especially respiratory, cardiac, and cardiovascular-related illnesses. Therefore, examining the patient’s comorbidity is imperative to make the prognosis. The statistical dataset [33] used in our study categorizes the disease types into 22 large-scale categories displayed in Table 1. The initial classified types could be partitioned further with the specific disease based on the WHO’s International Statistical Classification of Disease and Related Health Problems (ICD) [40].

Let $D = 0x22C3_{j=1}^{22} C_j$ (Table 1), and $M(s_{(D,G,k)})$ be the mortality of the subject with three properties of D, G, k . In $\exists C_j$, we measure the correlation with the Pearson correlation coefficient among each age group, with $\rho_{(D,k)} := Cov(M(s_D), M(s_k)) / (\sigma(M(s_D)) \cdot \sigma(M(s_k)))$ and $\rho_{(D',k)} \geq 0.9$ where $\sigma(\cdot)$ expresses the standard deviation and $D' = D - \{C_{17}, C_{18}, C_{19}\}$ as shown in Fig. 2. C_{17} indicates the obesity where age group between 55~64 had the most significant number of casualty and C_{18}, C_{19} (i.e., Alzheimer and Dementia) have exponentially increased with a linear

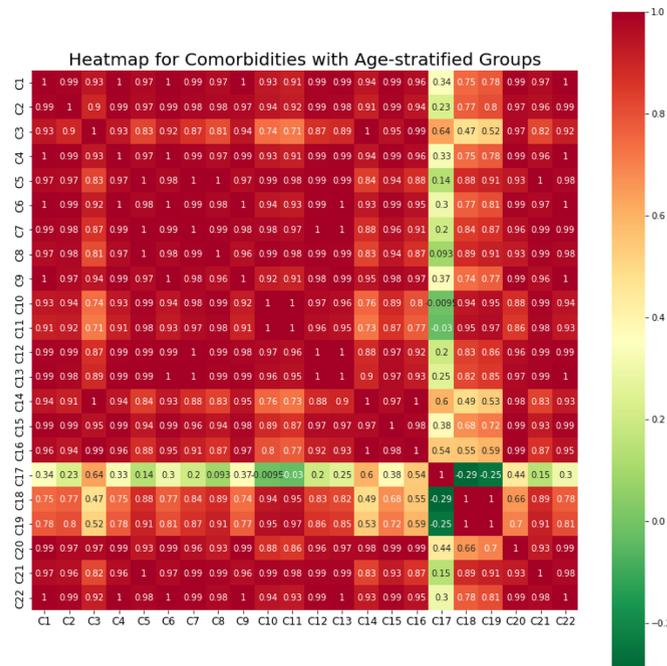


Fig. 2. Heatmap indicating the correlations between each age group based on $M(s_k)$.

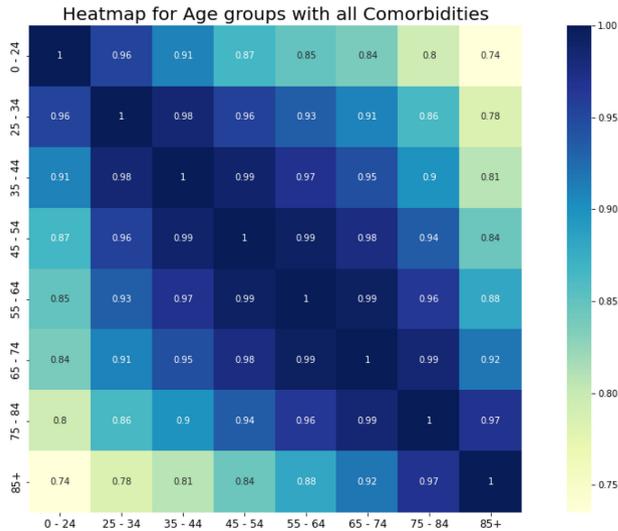


Fig. 3. Heatmap indicating the correlations between each age group based on $M(s_D)$.

increment of age groups. This asserts that most $M(s_D)$ have a linear correlation with $M(s_k)$, and until 55~64 the $\Delta s(D',k,k+1)$ tend to increase where $\Delta s(D',k,k+1) = |M(s(D',k)) - M(s(D',k+1))|$. After that point, the gradient starts to decrease, having $\Delta s(D',\bar{k},\bar{k}+1) > \Delta s(D',\bar{k},\bar{k}+1)$ where \bar{k} indicates the index of age group over 65, and k refers to the index of age group under 65. Furthermore, Fig. 3 shows that the correlation between given $M(s_D)$ sorted by order of k , which implicates that $\mu(M(s_{(D,\exists k)}))$ and $\mu(M(s_{(D,\exists k+L)}))$ where $\mu(\cdot)$ denotes the average set of inputs and arbitrary $L \neq 0 \in (\mathbb{R}^+ 0x222AR^-)$, $1 \leq (k+L) \neq k \leq 9$ becomes distinct as $|L|$ gets larger. Table 1 shows the number of COVID-19 deaths with the corresponding comorbidity sorted with age groups, and we estimate the fatality based on statistics. The fatality $\delta(s_{(D,k)})$ is expressed in (3).

$$\delta(s_{(D,k)}) := \frac{M(s_{(D,k)})}{M(s_D) \cdot M(s_k)} \quad (3)$$

Since the exact number of diseases that each patient infected was not publicly opened, we randomly allocated the comorbidity number $q_i \in [0, 3]$ and $q_i \in \mathbb{N}$, and this setting will be explained in detail in Section 4.3. Equation (4) specifies the linear aggregation of mortality in given features, for each comorbidity is independent and thus linearly escalates the risk. We scale the parameters through min-max normalization and compute the final mortality $M(s_{(D,G,k)})$.

$$M(s_{(D,G,k)}) := \frac{M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)}) - \min(0x22C3_{\forall k} M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)}))}{\max(0x22C3_{\forall k} M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)})) - \min(0x22C3_{\forall k} M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)}))} \quad (4)$$

4 INFECTION RATE AND NETWORK CENTRALITY

As infection rate is another critical factor when examining the pandemic's analytical properties, various past studies were conducted with practical perspectives to model and

defined the infection. Fundamentally, the current real-world system consists of numerous entities that dynamically commit transition with diverse dependent and independent variables, which cannot be intuitively defined nor easily predicted. Therefore, our study implements the simulation-based approach to estimate the propagation in a scenario that is specifically built to express the infection route via social interaction statistics, which is closely akin to a real-world topology.

4.1 Graph Network With Agent-Based Model

In the infectology system, physical connectivity matters when the host is infected. Infections are perpetrated when a virus has transferred from the infected to the non-infected, and we define this relationship with a graph network structure.

Let $G = (V, E)$ where V is a set of vertices, and E is a set that is consisted of edges where $V = \{v_i | 1 \leq i \leq N, i \in \mathbb{N}\}$, N indicates the total number of clients (nodes; agents) in G and $E = \{\langle v_a, v_b \rangle | (v_a, v_b) \in V \times V, 1 \leq (a, b) \in \mathbb{N} \times \mathbb{N} \leq \sum_{\forall i} 1\}$. The elements of G are finite, and $\langle v_a, v_b \rangle$ indicates the directed edge with initial v_a and destination v_b . The infection process has a particular direction, where infected agents transmit the virus to the connected non-infected node, which verifies a source and destination. Thus, utilizing a directed graph is reasonable. The visualization of our graph structure sample is shown in Fig. 4.

Let $E_a = 0x22C3_{\forall b} \langle v_a, v_b \rangle$, and $n(E_a)$ indicates the number of edges that begins with v_a and is connected towards $\exists v_b$ where b specifies the neighbor $v_{\exists i}$ of the v_a . When v_a with $\max n(E_a)$ is infected such that $v_a = True$, the virus's diffusion pace is likely to increase with a high probability since its importance among the affiliated community is significant. To measure the importance of individual agents among the associated group, we implement the network centrality [41], [42] metrics to determine such potential influence. We applied five following well-known centrality metrics: Degree Centrality [43], Closeness Centrality [42], Betweenness Centrality [44], Eigenvector Centrality [45], and PageRank [46]. Among those centralities, we examine the worst-case scenario that proliferates the virus within the shortest period in an equivalent population setting.

4.2 Selecting Optimal Centrality

To select the worst-case centrality, we compare the virus propagation period it takes to infect the virus to all the nodes inside the ABM, initiating the proliferation from the number of initial spreaders. At the outset, we construct the three ABM types with $n(V) = 1,000, 5,000, \text{ and } 10,000$. For

$n(E) = \sum_{\forall a} n(E_a)$, we implemented the statistics of physical contact per day by age groups suggested by Del Valle et al. [26], considering that the infection spreads through close contact exposure. Therefore, each agent is allocated

TABLE 2
Population Ratio and Corresponding Contact Frequency

Age	Population ratio	Contact Frequency	RCF
0-4	0.068	10.216	10
5-9	0.061	14.812	15
10-14	0.063	18.224	18
15-19	0.064	17.582	18
20-29	0.137	13.573	14
30-39	0.135	14.142	14
40-49	0.123	13.830	14
50-59	0.129	12.308	12
60-69	0.116	9.216	9
70+	0.112	6.898	7

with a certain age, based on the ratio of U.S. population demographics by age in 2020 [47], and v_b are randomly selected for v_a as much as the corresponding number of contacts per day. The contact frequency for each age group is listed in Table 2, and the final number of edges for each vertex is randomly designated between the range of $1 \leq n(E_a) \leq 2 \cdot RCF$ where RCF is an abbreviation of Round-up Contact Frequency in the coinciding age group. The infection rate was set to 100% since our aim is to discern the propagation speed. In each time step t , any v_i that are connected to the infected agent is designated as a newly infected agent, where $\langle v_a, v_b \rangle$ refers to the virus transmission from infected v_a to v_b . Visualization for intuitive interpretation is shown in Fig. 4.

Each transmission through a single edge equals a one-time step, and the random walk propagation algorithm was designed based on Recursion and Depth-First-Search, which we explain in detail through pseudocode in Section 5. Assuming that the initial spreaders with higher centrality would trigger faster dissemination (*i.e.*, empirical proofs are shown in Section 4.3.), we select the ten initial spreaders $\tilde{p} = 0x22C3_{10}\tilde{v}_i$ where $\tilde{v}_i = \max(0x22C3_{\psi_i}\varphi(v_i))$, and $\varphi(v_i)$ denotes the centrality of the input v_i and \tilde{v}_i is sampled without replacement. We experimentally validate the selection of $\max(0x22C3_{\psi_i}\varphi(v_i))$ in the next section. The virus propagation result is shown in Table 3 after 50 trials on each centrality on three types of total nodes. Each value is the average value of 50 trials. Note that every attempt has

TABLE 3
Average Period Duration Among Given Centrality

Centrality	Number of Nodes			Average
	1,000	5,000	10,000	
Degree	271.8 (±4.2)	331.4 (±3.3)	332.3 (±2.9)	311.8 (±3.5)
Closeness	129.5 (±20.7)	145.5 (±22.1)	143.2 (±12.0)	139.4 (±18.3)
Betweenness	129.5 (±20.7)	145.5 (±22.1)	143.2 (±12.0)	139.4 (±18.3)
Eigenvector	130.4 (±19.9)	144.4 (±31.1)	156.4 (±11.2)	143.8 (±20.7)
PageRank	116.9 (±19.4)	139.4 (±14.3)	103.7 (±10.7)	123.1 (±14.8)

Each value is the average value of 50 trials (new graph structure in each trial), and the number inside the parenthesis illustrates the standard deviation.

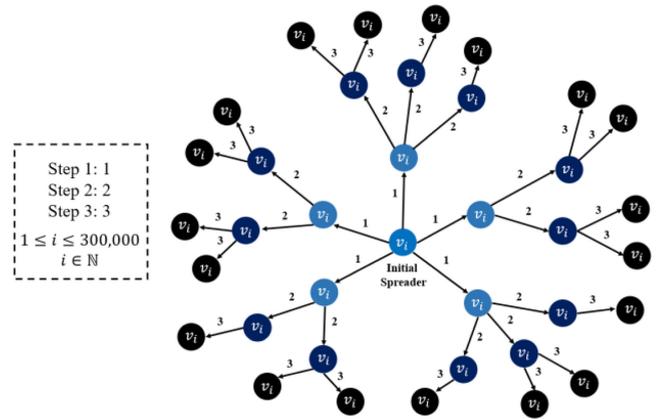


Fig. 4. Time steps $t (=1,2,3)$ of spreading disease in given graph network.

experimented with a new randomly generated graph structure with an identical number of the vertex. The outcome shows that PageRank consumes the shortest period when spreading the disease among all existing nodes. Therefore, we take PageRank centrality into account to measure the infection rate. PageRank is an algorithm that ranks the entity in graph structure widely implemented in the Google web search engine [46]. It judges the node's importance by aggregating the quality and quantity of other connected nodes. The original equation is shown in (5), where $PR(\cdot)$ is a PageRank value of the designated node $v_i, i \neq i'$, and we set the damping factor $q = 0.85$.

$$PR(v_i) = \frac{1 - q}{n(0x22C3_{\psi_i}v_i)} + q \cdot \sum_{v_{i'} \in InLink(v_i)} \frac{PR(v_{i'})}{NumOutLinks(v_{i'})} \tag{5}$$

4.3 ABM Construction by PageRank and EVI

Fig. 4. Time steps $t (= 1,2,3)$ of spreading disease in given Graph Network.

In this section, we experimentally validate that vaccination through descending order of PageRank suppresses the proliferation of virus in the ABM, and we assemble the existing factors to compute the final EVI. In order to corroborate the descending order of the proliferation effect, we designed the ABM with 300,000 nodes via real-world statistics. Each agent consisting of ABM has individual unique clinical properties, allocated based on the statistical proportion of features [33], [34]. For example, when the population of the male age group of 35~44 takes $x\%$, we allocate $n(V) \cdot x \cdot 0.01$ nodes with a male and random age between 35~44 with uniform distribution and Gaussian noise. The underlying disease dataset applied in Section 3.2 indicates the collective groups from the death patient data, and we compute the percentage of deaths among the infection cases and adopt it to V . The original mortality is 1.45%, but it was intentionally increased by five times (7.25%) since distinct computation for the value was too sparse to effectively observe the variation of death cases. Comorbidities were allocated to the agents with their corresponding statistical ratio. For the number of comorbidities per agent, considering that there may be the possibility of patients who possess multiple illnesses, recall that the maximum number of the

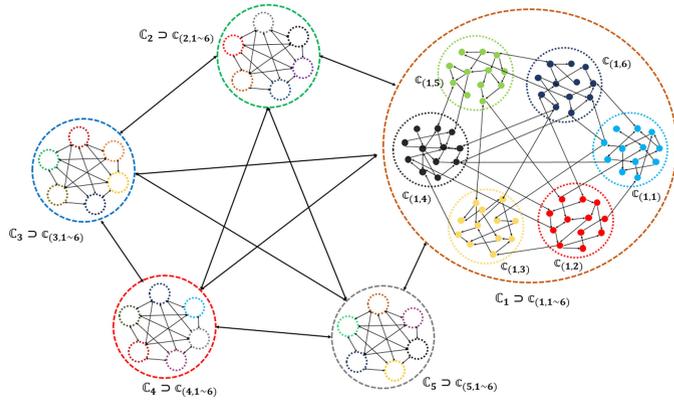


Fig. 5. The constructed graph network structure for simulations.

underlying disease was restricted to three and randomly selected between $0 \sim 3$ for the amount of ratio that includes the comorbidity. Edges were connected identical fashion with the built graph network in Section 4.2. Moreover, V was divided into five clusters: $C_1 \sim C_5$, and each cluster is subdivided into six smaller regions: $c_1 \sim c_6$ as shown in Fig. 5. This is primarily because the virus proliferation is initially bounded by the geographical regions, and most of the physical contact occurs inside the affiliated community where interaction between the group to group is not as frequent as the internal contact frequency. To accommodate the diversity of regional properties, each region consists of a statistical ratio of age groups, and areas are connected intermittently by 20% of the original edges.

We compared the two following cases: the first case selected the $\bar{\rho} = 0x22C3_{i=0}^{19}\bar{L}$ and the second case with $\bar{\rho} = 0x22C3_{i=0}^{19}\bar{L}$ as initially infected agents where $\bar{L} = \text{sort}(0x22C3_{\varphi_i}\varphi(v_i))$ based on the order of $\text{argmin}(0x22C3_{\varphi_i}\varphi(v_i))$, and $\bar{L} = \text{sort}(0x22C3_{\varphi_i}\varphi(v_i))$ based on the order of $\text{argmax}(0x22C3_{\varphi_i}\varphi(v_i))$, and $\varphi(v_i)$ specify a PageRank centrality computation. The experiment measured the cumulative period until every node was infected where initial agents are designated without replacement. The result shows that when $\bar{\rho}$, the average period was 456.607-time steps (Blue graph in Fig. 6.) and 485.642-time steps (Red graph in Fig. 6.) for $\bar{\rho}$ after 70 trials, respectively. Each period density distribution is shown in Fig. 6. This entails that the highest to lowest order distribution (*i.e.*, $\sum_{\forall t} \bar{\rho} \ll \sum_{\forall t} \bar{\rho}$) brings faster proliferation speed; therefore, vaccination based on \bar{v}_i implicates the effectiveness. We compute the individual's EVI after aggregating the mortality $M(s_{(D,G,k)})$ and the PageRank $\varphi(v_i)$ in (6), normalizing each factor using the standard normalization, and then scaling the accumulation again with min-max normalization.

As (6) shows, the EVI is to both minimize the death rate and infection rate, involving the portions of $W_1 \cdot \text{Infection rate} + W_2 \cdot \text{Death rate}$. We can calibrate the weights (W_1, W_2) that determine the level of consideration of each term by adjusting values where $W_1 + W_2 = 1$ and $(W_1, W_2) \in \mathbb{R}^+$. Similar to Fig. 7, which implies the linear correlation coefficient between the infection rate and death rate is -0.3, and this infers the trade-off between the infection rate and the mortality among the identical population. The labels of F1~F11 are as follows: {population, population

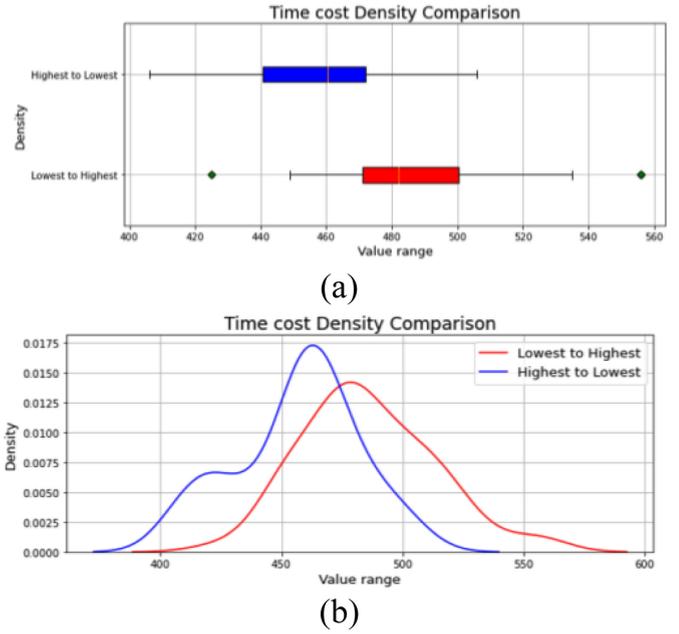


Fig. 6. Distribution results of period visualization using box plot (a) and density plot (b).

density, land area, infection cases, death cases, infection rate, death rate, GDP per capita, Median age, race variance, age variance} in respect of 50 state-level datasets of United States. From the age perspective, the younger generations are mostly correlated with spreading the virus, and the older generation occupies high morality. This is expected to trigger the trade-off when selecting a vaccination candidate, where vaccination towards the young-aged group may decrease the infection rate but escalate the overall death rate, and vaccination focusing on the older generation would lead to an opposite result. Therefore, propagating vaccines by examining those two factors in the optimal ratio will balance the casualties in the long run.

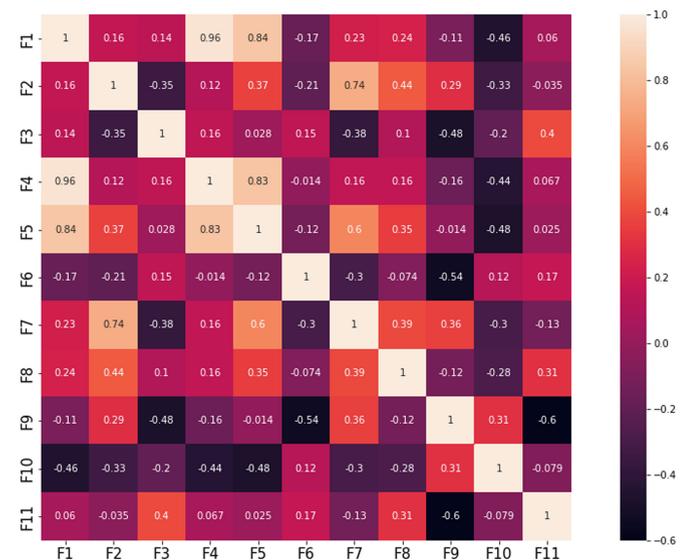


Fig. 7. Heatmap that visualizes the Pearson correlation coefficient of the combination of factors.

$$\begin{aligned}
EVI &\stackrel{def}{=} \frac{W_1A + W_2B - \min(W_1A + W_2B)}{\max(W_1A + W_2B) - \min(W_1A + W_2B)} \\
s.t. \quad A &= \frac{M(s_{(D,G,k)}) - \mu \left(M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)}) \right)}{\sigma \left(M(s_{(G,k)}) + \sum_{\forall q_i} \delta(s_{(D,k)}) \right)} \\
and \quad B &= \frac{\phi(v_i) - \mu(\phi(V))}{\sigma(\phi(V))} \quad (6)
\end{aligned}$$

5 EXPERIMENTS

This section presents the experiment and its consequential results to validate the vaccine distribution effect through propagating the vaccine with the highest to the lowest order of EVI while the virus is being disseminated in the ABM. The evaluation criteria are the number of infection cases, death cases, and the mortality rate in the ABM consisting of 300,000 agents, which incorporates their unique features such as age, comorbidity types, gender, PageRank value, region, mortality rate, EVI, etc. We compare nine vaccination routes and observe the variation of three metrics.

5.1 Experiment Settings

In this vaccine propagation simulation, we compare the variation of three metrics (infection cases, death cases, mortality rate) when vaccination through descending order of EVI and eight other criteria (SVI [5], PVI [6], CVI [4], Age, Comorbidity risk, Age + Comorbidity risk, Random distribution, no vaccination) where every agent contains its unique corresponding index. The reason for indicating only three metrics is that other indexes can be inferred from these three. (e.g., death cases + cured cases = infection cases). The propagation through EVI is shown in Algorithm 1, which accords with the stochastic random walk process. In Algorithm 1, $R([a, b], w = c)$ denotes the function that outputs the stochastically random integer value between $[a, b]$, with the corresponding possibility (or weight; $w \in [0, 1]$, $w \in \mathbb{R}$) of selection on each candidate. \ominus and \oplus symbol indicates the remove and appends function respectively, for example, $D \ominus d$ specifies the element $d \in D$ is being removed from the list D , and $D \oplus d$ refers to append element d to list D . In each time step, the algorithm iteratively summons the recursive function and stores/updates the individual agent's status to the auxiliary buffer (i.e., partially visualized in Fig. 4). Each vaccination simulation type propagated the virus through an average of 50% infection rate mixed with Gaussian noise ($\mu = 0.5$, $\sigma = 0.1$; i.e., no particular dataset has been released concerning contracting the disease when the host encountered the virus). It was tested for 70 trials with 20 initial patients and 500 vaccines per time unit. At each trial, properties of the individual node were altered randomly, creating a new allocation of edges.

5.2 Experiment Result

Fig. 8a displays the visualization of cumulative numerical values of each metric (death, cured, no virus, vaccinated) in time-series format, and (b) shows the enlargement view of the sequential time-step variation values where $|x_t - x_{(t+1)}|$, having x as the numerical value of designated metrics.

During the initial steps, the virus propagation tends to increase sharply until a certain point, as the variation graph (Fig. 8b) involves a long-tail distribution format, extending the tail to the right side (as the graph suggests an enlarged view). During those initial periods, it asserts that the augmented contagious level reflects the graph structure, and the reality in most of the top 30 GDP countries' pandemic dynamics follows the suggested trend with a slight time difference (i.e., 1~3 months) on a regional basis, for each variant (e.g., Delta+(B.1.617.2.1, 2021. June), Omicron(B.1.1.529, 2021. November)) has rampantly struck the national-level communities with at least one explosive transmission. Figures (a, b) are triggered mainly by the static graph structure. The reason that figures (a, b) entail the low residual noises or random patterns is that we used the static ABM topology, whereas the real-world is a structure that dynamically varies with diverse variables, which is impalpable to define. Moreover, no viable benchmark dataset exists that represents the dynamic interaction. Instead, we incorporate the diversity with 70 trials on each random graph structure, where the outcomes are shown in Fig. 8(c ~ h).

Algorithm 1: Vaccination_EVI

Input: Graph g , list of integers $patients$, list of strings $stat$, list of list $Dataset$, integer vpt (vaccine per time)

Output: list of integers $time_unit$, list of strings $stat$

```

1:  $(D, C, V) \leftarrow 0, (node\_lst, D\_lst, C\_lst, V\_lst) \leftarrow$  empty list
2: While  $V \neq vpt$  do
3:   if  $EVI\_sort = \emptyset$  then
4:     Break
5:   else
6:     if  $stat[EVI\_sort[cnt]] = 'No\ Virus'$  then
7:        $stat[EVI\_sort[cnt]] \leftarrow 'V'$  // Vaccinated
8:        $(V++)$  and  $(EVI\_sort \ominus EVI\_sort[cnt])$ 
9:   for  $node \in \{1, 2, \dots, n(patients)\}$  do
10:    if  $stat[patients[node]] = 'No\ Virus'$  then
11:      if  $R([0, 1], w = [0.5, 0.5]) = 0$  then // infection rate 50%
12:         $tmp \leftarrow R([0, 1], w = \text{death rate in } Dataset)$ 
13:        if  $tmp = 0$  then
14:           $stat[patients[node]] \leftarrow 'D'$  // Dead
15:           $(D++)$  and  $(EVI\_sort \ominus EVI\_sort[cnt])$ 
16:        Else
17:           $stat[patients[node]] \leftarrow 'C'$  // Cured
18:           $(C++)$  and  $(EVI\_sort \ominus EVI\_sort[cnt])$ 
19:        for  $n \in \{1, 2, \dots, \# \text{ of neighbors of } patients[node]\}$  do
20:           $node\_lst \oplus \text{list of neighbors of } patients[node][n]$ 
21:    $(D\_lst \oplus D)$  and  $(C\_lst \oplus C)$  and  $(V\_lst \oplus V)$ 
22:   for  $node \in \{1, 2, \dots, n(patients)\}$  do
23:     for  $neighbor \in \{1, 2, \dots, n(node\_lst)\}$ 
24:       if  $stat[node\_lst[neighbor]] = 'D'$  or  $'C'$  or  $'V'$  then
25:         pass
26:       else
27:          $time\_unit \oplus 'O'$  //  $n(time\_unit) = \text{time passed}$ 
28:          $Vaccination\_EVI(g, patients, stat, Dataset, vpt)$ 

```

Bar graphs in (c ~ h) show the vaccination result for each criterion after the proliferation ceases. In the figures (c ~ k), the non-vaccination outcome was omitted for almost 100% of was infected. A y-axis is a relative unit, where 1.0 is the value of a non-vaccination scenario. For example, no

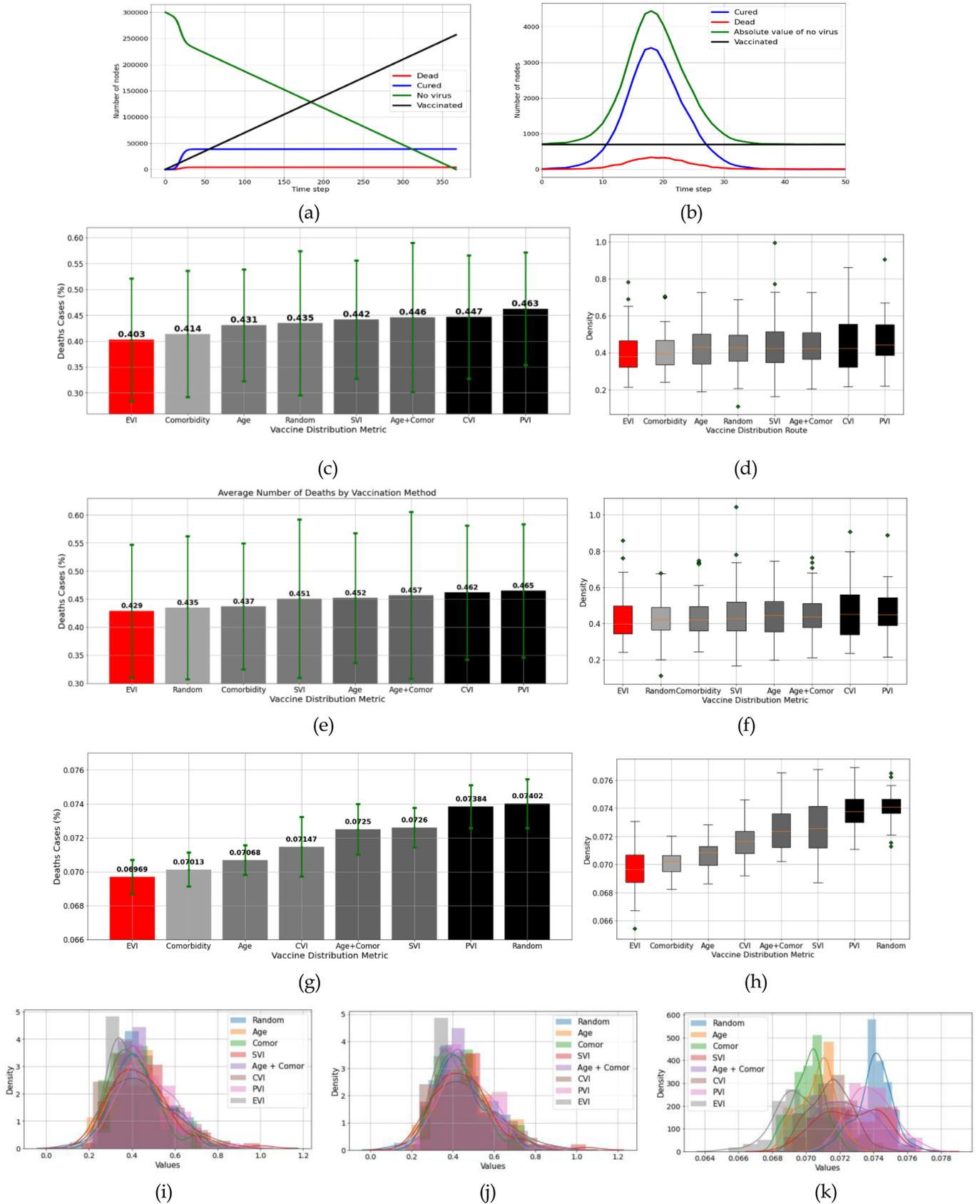


Fig. 8. (a) displays the cumulative result for each metric throughout the vaccination in EVI. (b) shows the variation of sequential metric values. (c), (e), and (g) are the average and standard deviation of death cases, infection cases, and death rate, respectively. (d), (f), and (h) indicate the distribution range for 70 trials for each vaccination route. Each value in (c) ~ (h) denotes a relative value where 1.0 is an outcome without the vaccine distribution. (i) ~ (k) indicate the distribution estimations of 70 trials in each metric (*i.e.*, (i) death cases, (j) infection cases, (k) death rate).

vaccine scenario is 1.0; 100% and EVI 0.403; 40.3% when measuring death cases. Fig. 8(c, e, g) respectively compares the death case, infection case, and mortality result of initiated vaccine distribution simulations, and their standard deviation is shown. Fig. 8(d, f, h) respectively displays the density of death cases, infection cases, and mortality results after 70 trials in the box plot. To verify the statistical significance, we implemented the ANOVA test by aggregating the 70 simulation outcomes for each vaccine distribution case. The p-values indicated that it converges to 0, such that the death case specifies $2.415e-156$, infection case with $8.907e-147$, and death rate shows to have $8.663e-127$, where the distribution plot is shown in Fig. 8(i ~ k) for each metrics; death case, infection case, and death rate respectively. Vaccine distribution through EVI had shown an average of 9.4% lower in death cases, 5.0% lower in infection cases, and 3.5% lower in death rates than other distribution criteria.

6 CONCLUSION

This paper presents the Epidemic Vulnerability Index that quantitatively assesses the potential risk of the individual subject based on their clinical conditions, which estimate the mortality and social centrality that determine their infection rate. Our experiments entail that the vaccination simulation through descending order of EVI had the best performance of containing the casualties among the ABM formulated based on the real-world statistical dataset with 300,000 nodes. Compared to the eight criteria that determine the risk, EVI vaccination represents 5.0% lower infection cases, 9.6% lower death cases, and 3.5% lower mortality.

EVI suggests efficiency through empirical vaccination simulations to evaluate the vaccination distribution routes. To accommodate an accurate assessment, modeling the propagation approximating the real-world scenario is critical. However, virus proliferation among convoluted collected groups is a dynamic system, for numerous and diverse factors are intertwined, including social atmosphere, economic, political aspects, vaccine performance, randomness, etc. Therefore, estimating future conditions and patterns with accurate predictions is an arduous objective. The limitation of our work lies in this domain, for our simulation was primarily conducted on the diverse static graph structures, where the real-world system evolves dynamically throughout the time-domain.

As we have shown that infection and death rate hold a negative correlation, simultaneously seizing both factors is a complex task. However, fine-tuning each weight by deliberating the current and future trends would yield the optimal response. Various new perspectives and approaches are required to construct a productive vaccine distribution strategy. In reality, the decision-makers select feasible candidates in the solution process according to multi-perspective analysis and establish a combined policy. This study could be applied to the response plan for other possible pandemics, such as regional vaccine distribution scenarios, predicting the medical supply among the community, adopting the new factors, and organizing the optimal strategy at the right time and place.

REFERENCES

- [1] World Health Organization, 2020. Accessed: Mar. 15, 2020. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [2] World Health Organization, "WHO director-general's opening remarks at the media briefing on COVID-19 - 11 March 2020," 2020. Accessed: Mar. 15, 2020. [Online]. Available: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>
- [3] O. Amram, S. Amiri, R. B. Lutz, B. Rajan, and P. Monsivais, "Development of a vulnerability index for diagnosis with the novel coronavirus, COVID-19, in Washington state, USA," *Health Place*, vol. 64, no. 4, 2020, Art. no. 102377.
- [4] D. DeCaprio et al., "Building a COVID-19 vulnerability index," *J. Med. Artif. Intell.*, vol. 3, 2020.
- [5] B. E. Flanagan, E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, and B. Lewis, "A social vulnerability index for disaster management," *J. Homeland Secur. Emerg. Manage.*, vol. 8, no. 1, 2011.
- [6] S. W. Marvel et al., "The COVID-19 pandemic vulnerability index (PVI) dashboard: Monitoring county-level vulnerability using visualization, statistical modeling, and machine learning," *Environ. Health Perspectives*, vol. 129, no. 1, p. 017701, 2021.
- [7] A. A. Salisu and O. A. Lateef, "Constructing a global fear index for the COVID-19 pandemic," *Emerg. Markets Finance Trade*, vol. 56, no. 10, pp. 2310–2331, 2020.
- [8] H. Pinglin et al., "Accounting index of COVID-19 impact on Chinese industries: A case study using big data portrait analysis," *Emerg. Markets Finance Trade*, vol. 56, no. 10, pp. 2332–2349, 2020.
- [9] S. D. Searle, A. Mitnitski, E. A. Gahbauer, T. M. Gill, and K. Rockwood, "A standard procedure for creating a frailty index," *BMC Geriatrics*, vol. 8, no. 1, pp. 1–10, 2008.
- [10] G. Bellelli, R. Paola, and G. Citerio, "The role of frailty in COVID-19 patients," *Intensive Care Med.*, vol. 46, no. 10, pp. 1958–1959, 2020.
- [11] G. Bellelli et al., "Frailty index predicts poor outcome in COVID-19 patients," *Intensive Care Med.*, vol. 46, pp. 1634–1636, 2020.
- [12] K. M. Bubar et al., "Model-informed COVID-19 vaccine prioritization strategies by age and serostatus," *Science*, vol. 371, no. 6532, pp. 916–921, 2021.
- [13] L. Corey, J. R. Mascola, A. S. Fauci, and F. S. Collins, "A strategic approach to COVID-19 vaccine R&D," *Science*, vol. 368, no. 6494, pp. 948–950, 2020.
- [14] A. R. Tuite, L. Zhu, D. N. Fisman, and J. A. Salomon, "Alternative dose allocation strategies to increase benefits from constrained COVID-19 vaccine supply," *Ann. Intern. Med.*, vol. 174, no. 4, pp. 570–572, 2021.
- [15] C. C. Kerr et al., "Covasim: An agent-based model of COVID-19 dynamics and interventions," *PLoS Comput. Biol.*, vol. 17, no. 7, p. e1009149, 2021.
- [16] Centers for Disease Control and Prevention, 2009. Accessed: Nov. 11, 2021. [Online]. Available: www.cdc.gov/h1n1flu/vaccination/vaccinesupply.html
- [17] A. Coustasse, C. Kimble, and K. Maxik, "COVID-19 and vaccine hesitancy: A challenge the United States must overcome," *J. Ambulatory Care Manage.*, vol. 44, no. 1, pp. 71–75, 2021.
- [18] J. Wang, Y. Peng, H. Xu, Z. Cui, and R. O. Williams, "The COVID-19 vaccine race: Challenges and opportunities in vaccine formulation," *AAPS PharmSciTech*, vol. 21, no. 6, pp. 1–12, 2020.
- [19] Y. A. Adebisi, G. I. Oke, P. S. Ademola, I. G. Chinemelum, I. O. Ogunkola, and D. E. Lucero-Prisno, "SARS-CoV-2 diagnostic testing in Africa: Needs and challenges," *Pan Afr. Med. J.*, vol. 35, no. 4, 2020.
- [20] M. Rastegar, M. Tavana, A. Meraj, and H. Mina, "An inventory-location optimization model for equitable influenza vaccine distribution in developing countries during the COVID-19 pandemic," *Vaccine*, vol. 39, no. 3, pp. 495–504, 2020.
- [21] S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, and R. Mboogo, "SEIR model for COVID-19 dynamics incorporating the environment and social distancing," *BMC Res. Notes*, vol. 13, no. 1, pp. 1–5, 2020.
- [22] L. Matrajt, J. Eaton, T. Leung, and E. R. Brown, "Vaccine optimization for COVID-19: Who to vaccinate first?," *Sci. Adv.*, vol. 7, no. 6, p. eabf1374, 2021.
- [23] C. Wolfram, "An agent-based model of COVID-19," *Complex Syst.*, vol. 29, no. 1, pp. 87–105, 2020.

- [24] M. S. Shamil, F. Farheen, N. Ibtehaz, I. M. Khan, and M. S. Rahman, "An agent-based modeling of COVID-19: Validation, analysis, and recommendations," *Cogn. Computation*, pp. 1–12, 2021.
- [25] P. CL. Silva et al., "COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions," *Chaos Solitons Fractals*, vol. 139, p. 110088, 2020.
- [26] J. Gomez, J. Prieto, E. Leon, and A. Rodriguez, "INFEKTA: A general agent-based model for transmission of infectious diseases: Studying the COVID-19 propagation in bogotá-Colombia," *PLoS One*, vol. 16, no. 2, p. e0245787, 2021.
- [27] N. Hoertel et al., "Facing the COVID-19 epidemic in NYC: A stochastic agent-based model of various intervention strategies," *MedRxiv*, 2020.
- [28] R. Hinch et al., "OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing," *PLoS Comput. Biol.*, vol. 17, no. 7, p. e1009146, 2021.
- [29] S. Y. Del Valle, J. M. Hyman, H. W. Hethcote, and S. G. Eubank, "Mixing patterns between age groups in social networks," *Social Netw.*, vol. 29, no. 4, pp. 539–554, 2007.
- [30] M. M. Hughes, "County-level COVID-19 vaccination coverage and social vulnerability," *Morbidity Mortality Weekly Rep.*, vol. 70, no. 12, pp. 431–436, 2021.
- [31] N. D. Yanez, N. S. Weiss, J. A. Romand, and M. M. Treggiari, "COVID-19 mortality risk for older men and women," *BMC Public Health*, vol. 20, no. 1, pp. 1–7, 2020.
- [32] S. Mukherjee and K. Pahan, "Is COVID-19 Gender-sensitive?," *J. Neuroimmune Pharmacol.*, vol. 16, no. 4, pp. 38–47, 2021.
- [33] Centers for Disease Control and Prevention, 2021. Accessed: Apr. 12, 2021. [Online]. Available: <https://data.cdc.gov>
- [34] NYTimes, 2021. Accessed: Apr. 06, 2021. [Online]. Available: <https://github.com/nytimes/covid-19-data>
- [35] Medscape, "Almost 90% of COVID-19 admissions involve comorbidities," 2020. Accessed: Apr. 18, 2021. [Online]. Available: <https://www.medscape.com/viewarticle/928531>
- [36] A. Sanyaolu et al., "Comorbidity and its impact on patients with COVID-19," *SN Comprehensive Clin. Med.*, vol. 8, pp. 1–8, 2020.
- [37] B. Wang, R. Li, Z. Lu, and Y. Huang, "Does comorbidity increase the risk of patients with COVID-19: Evidence from meta-analysis," *Aging*, vol. 12, no. 7, pp. 6049–6057, 2020.
- [38] S. Jakhmola et al., "Comorbidity assessment is essential during COVID-19 treatment," *Front. Physiol.*, vol. 11, p. 984, 2020.
- [39] F. Xiaoyu et al., "Epidemiological, comorbidity factors with severity and prognosis of COVID-19: A systematic review and meta-analysis," *Aging*, vol. 12, no. 13, pp. 12 493–12 503, 2020.
- [40] World Health Organization, 2021. Accessed: Apr. 08, 2021. [Online]. Available: <https://www.who.int/standards/classifications/classification-of-diseases>
- [41] A. H. Dekker, "Network centrality and super-spreaders in infectious disease epidemiology," in *Proc. 20th Int. Congr. Modelling Simul.*, 2013.
- [42] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.
- [43] D. Sharma, A. Surolia, W. Dubitzky, O. Wolkenhauer, K. H. Cho, and H. Yokota, "Degree centrality," in *Encyclopedia of Systems Biology*. New York, NY, USA: Springer, 2013.
- [44] L. Leydesdorff, "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 9, pp. 1303–1319, 2007.
- [45] P. D. Straffin, "Linear algebra in geography: Eigenvectors of networks," *Math. Mag.*, vol. 53, no. 5, pp. 269–276, 1980.
- [46] D. Sullivan, "What is google pagerank? A guide for searchers & webmasters," *Search Engine Land*, 2007.
- [47] Statista, 2020. Accessed: Aug. 15, 2021. [Online]. Available: <https://www.statista.com/statistics/241488/population-of-the-us-by-sex-and-age/>



Hunmin Lee received the BE degree in computer engineering from Chonnam National University, Gwangju, South Korea, in 2020, and the MS degree from the Department of Computer Science, Georgia State University, in 2022. He is currently working toward the PhD degree with the University of Minnesota, Twin Cities, MN. His current research interests include span optimizations in a distributed environment (federated learning), data science, and system design based on AI.



Mingon Kang received the BE degree in computer and engineering from Hanyang University, South Korea, in 2006, and the MS and PhD degrees from the Department of Computer Science and Engineering, University of Texas at Arlington, in 2010 and 2015, respectively. He is currently an assistant professor with the Department of Computer Science, University of Nevada, Las Vegas. His research interests include bioinformatics, healthcare informatics, machine learning, data mining, and Big Data analytics.



Donghyun Kim received the BS degree in electronic and computer engineering, and the MS degree in computer science and engineering from Hanyang University, Ansan, South Korea, in Feb. 2003 and Feb. 2005, respectively, and the PhD degree in computer science from the University of Texas at Dallas, Richardson, TX, in May 2010. He is currently an assistant professor with the Department of Computer Science, Georgia State University (GSU), Atlanta, GA. He is a senior member of ACM. He has served as a TPC

co-chair for several international conferences, most recently IPCCC 2020 and COCOON 2020.



Daehee Seo received the BS degree in electronic and electrical engineering from Dongshin University, Naju, South Korea, in Feb. 2001, the MS degree in computer science and engineering and PhD degree in computer science from Soonchunhyang University, South Korea, in Feb. 2003 and Feb. 2006, respectively. He is currently an assistant professor with the Faculty of Artificial Intelligence and Data Engineering, Sangmyung University (SMU), Seoul, South Korea.



Yingshu Li received the PhD and MS degrees from the Department of Computer Science and Engineering, University of Minnesota-Twin Cities. He is currently a professor with the Department of Computer Science and an affiliated faculty member with the INSPIRE Center, Georgia State University. Her research interests include privacy-aware computing, management of big sensory data, Internet of Things, social networks, and wireless networking. He is the recipient of the NSF CAREER Award.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.