Neural Computing and Applications (2022) 34:14859–14879
https://doi.org/10.1007/s00521-022-07210-8

ORIGINAL ARTICLE

Who are the 'silent spreaders'?: contact tracing in spatio-temporal memory models

Yue Hu[1] · Budhitama Subagdja[2] · Ah-Hwee Tan[2] · Chai Quek[3] · Quanjun Yin[1]

Received: 26 October 2021 / Accepted: 29 March 2022 / Published online: 14 May 2022

Abstract

The COVID-19 epidemic has swept the world for over two years. However, a large number of infectious asymptomatic COVID-19 cases (*ACCs*) are still making the breaking up of the transmission chains very difficult. Efforts by epidemiological researchers in many countries have thrown light on the clinical features of ACCs, but there is still a lack of practical approaches to detect ACCs so as to help contain the pandemic. To address the issue of ACCs, this paper presents a neural network model called Spatio-Temporal Episodic Memory for COVID-19 (*STEM-COVID*) to identify ACCs from contact tracing data. Based on the fusion Adaptive Resonance Theory (*ART*), the model encodes a collective spatio-temporal episodic memory of individuals and incorporates an effective mechanism of parallel searches for ACCs. Specifically, the episodic traces of the identified positive cases are used to map out the episodic traces of suspected ACCs using a weighted evidence pooling method. To evaluate the efficacy of STEM-COVID, a realistic agent-based simulation model for COVID-19 spreading is implemented based on the recent epidemiological findings on ACCs. The experiments based on rigorous simulation scenarios, manifesting the current situation of COVID-19 spread, show that the STEM-COVID model with weighted evidence pooling has a higher level of accuracy and efficiency for identifying ACCs when compared with several baselines. Moreover, the model displays strong robustness against noisy data and different ACC proportions, which partially reflects the effect of breakthrough infections after vaccination on the virus transmission.

Keywords Asymptomatic coronavirus carriers · ART-based spatio-temporal episodic memory · Weighted evidence pooling · COVID-19 simulation · Realistic scenarios

1 Introduction

In stark contrast to SARS and MERS, the two highly infectious coronaviruses caused respiratory diseases, a much larger proportion of COVID-19 patients never develop symptoms like fever and cough but are still contagious while shedding the severe acute respiratory syndrome coronavirus 2 (*SARS-CoV-2*) [1–3]. The presence of symptoms facilitates case detection, however, the asymptomatic COVID-19 cases (or *ACCs*) are usually unaware of their infectiousness. Therefore, many researchers termed them as 'silent spreaders' of SARS-CoV-2 [4]. Currently many countries have limited measures to contain the large number of transmissions brought about by ACCs, making it hard to control the pandemic [2, 5].

Recently many researchers have started to focus on issues about ACCs such as the proportion of ACCs over all infected cases, the duration of virus shedding and the

✉ Yue Hu
huyue11@nudt.edu.cn

Budhitama Subagdja
budhitamas@smu.edu.sg

Ah-Hwee Tan
ahtan@smu.edu.sg

Chai Quek
ashcquek@ntu.edu.sg

Quanjun Yin
yin_quanjun@163.com

[1] College of Systems Engineering, National University of Defense Technology, Changsha, Hunan 410073, China

[2] School of Computing and Information Systems, Singapore Management University, 178902 Singapore, Singapore

[3] School of Computer Science and Engineering, Nanyang Technological University, 639798 Singapore, Singapore

temporal dynamics of infectiousness [4, 6–11]. Nevertheless, given that quick population screening for all positive cases is technically very difficult and financially crippling for many countries to implement, alternative proactive and cost-effective approaches are urgently needed to identify those silent spreaders and isolate them from the rest of the population.

In reality, many countries and regions like Singapore, Israel, South Korea, Australia and Taiwan are implementing digital contact tracing [12–18]. With suitable privacy protection, a large database of contact tracing data can potentially be used to identify ACCs. It is important to note that picking ACCs from a population is a new problem different from other contact tracing applications. Most of the existing systems using contact tracing serve to provide warnings to individuals who shared the same time and space with positively identified COVID-19 cases so they can be wary of the possibility of being infected [12, 19–21]. In contrast, the task studied in this paper is to identify invisible asymptomatic cases who may have infected those positively identified cases in the first place.

To address this important issue, we propose an episodic memory-based computational model named STEM-COVID to identify ACCs by learning and reasoning over the spatiotemporal trajectories of a population. The basic consideration is that people sharing the same space at the same time have a greater chance of infecting each other. Therefore, if one has appeared at the same places and at the same periods as many others who were later tested positive with the coronavirus, he/she could be considered as a plausible source.

Episodic memory, as a form of long-term memory, is a record of sequential events associated with contextual information, e.g., observations, activities, emotions, times and places. Fusion Adaptive Resonance Theory (ART) networks (see Sect. 5.1), as a class of biologically-inspired self-organized neural networks, are an effective computational tool to encode and recall episodic memory [22–24]. For efficient memory encoding and ACCs search, we have extended one such model called Spatio-Temporal Episodic Memory (STEM) [24] to encode the collective spatiotemporal traces and the COVID-19 positivity of tested individuals [25]. The model manifests potential efficacy for identifying ACCs out of the entire population.

However, this previous model suffers from some limitations in ACC search. In particular, it merges the episodic traces of all tested-symptomatic COVID-19 cases (*t-SCCs*) into a Boolean evidence vector. It can only indicate whether any *t*-SCC ever appeared within a spatiotemporal context, but not the quantum of *t*-SCCs present. To overcome this shortcoming, this paper proposes a weighted evidence pooling mechanism to extract more detailed information from the contact history between individuals. Specifically, the weights on different events in the evidence vector are proportionate to the number of subsequently diagnosed positive cases under the respective spatiotemporal contexts. Then the pooled memory trace of *t*-SCCs is used to search among the untested cases for individuals with similar episodic traces. Those with higher similarities are deemed to have greater likelihoods to be asymptomatic spreaders of the virus.

Additionally, given the difficult access to real-world contact tracing data due to privacy issue, this paper designs an agent-based modeling (*ABM*) of COVID-19 spreading to simulate the transmission of the coronavirus among people. The model is much more sophisticated and realistic than the one proposed in the previous work [25]; in terms of the infectiousness profile of ACCs and SCCs, the incubation period of SCCs, the duration of viral shedding of ACCs and the proportion of ACCs over all cases. For performance evaluation, experiments are conducted by running the simulation model under multiple scenarios in adherence to different specifications of infection and population sizes. We also conduct more rigorous benchmark experiments where the contact tracing data provided to the STEM-COVID model can be noisy or incomplete, mirroring that people in the real world often miss the registrations of some activities. The comparison with several other typical algorithms strongly shows that STEM-COVID can identify ACCs with a fairly high level of accuracy, robustness and efficiency.

In summary, the main novelties and technical contributions of this paper include: (a) an efficient search process for ACCs with a weighted evidence pooling method to unify the episodic traces of positive COVID-19 cases and searching over the untested cases in parallel; (b) a realistic spatiotemporal data simulation model of COVID-19 spreading designed based on recent epidemiological findings; (c) a rigorous benchmark setup using incomplete input data.

The rest of the paper is organized as follows. Section 2 summarizes recent studies on ACCs and epidemiological models and Sect. 3 formulates our study objective and describe the problem. Section 4 describes the simulated model and associated settings. The proposed model and detection algorithm are presented in Sect. 5. Experimental configurations, results and discussion are reported in Sect. 6. Section 7 concludes the paper with a discussion of contributions and future work.

2 Related epidemiological studies

2.1 Discussions on asymptomatic coronavirus cases

Since early January in 2020 when an asymptomatic infection of a 10-year-old child occurred in a family cluster who have traveled to Wuhan [26], the following months in 2020, 2021 and 2022 have witnessed occurrences of ACCs around the world including China, South Korea, Japan, Vietnam, Singapore, America, Italy and Iceland [9–11, 27–30].

As suggested by Mizumoto et al. [10], the proportion of asymptomatic cases and their infectiousness highlighted the level of difficulty to contain the disease. The estimations of the proportion of ACCs over all infected people varies between 5% and 80% of people diagnosed positive with the SARS-CoV-2 [29]. A report by the World Health Organization (WHO), published in March, suggested that about 80% of all infections would be asymptomatic or only mild [31]. In early April, China's National Health Commission has identified 130 asymptomatic cases out of 166 new infections [32]. An isolated village in northern Italy also reported that 50%-75% of infected people were asymptomatic [28]. However, there are also many observations that have presented much lower numbers. An observational cohort study on 36 children in Zhejiang, China found 28% of cases were asymptomatic [9]. Another family clustering-based study carried out in Zhejiang reported 54 (14%) cases with no symptoms out of 391 in total [33]. The infections on the Diamond Princess cruise ship included 17.9% of asymptomatic cases [10]. Tian *et al.* analyzed the data of 262 patients among whom 13 (5%) cases were asymptomatic [11]. This percentage was reported to be 23% in Singapore by the National Centre for Infectious Diseases (NCID) based on tracing of about 2500 household close contacts [27].

Despite the differences in the ratios of ACCs/SCCs, the epidemiological community has agreed that asymptomatic carriers can also be a significant source of transmission [1, 7, 8, 34]. Specifically, researchers observed transmissions among family members caused by asymptomatic cases, with some of whom developed severe pneumonia [7, 8]. Further in-depth study is needed to gain a greater insight on the dynamics of such transmissions as well as the clinical characteristics of ACCs. Studies conducted both in South Korea and northern Italy suggested that the viral load in asymptomatic patients is similar to that in symptomatic cases, implying similar potential degree of infectiousness [6, 34]. At the same time, evidence shows that the viral load in ACCs reduces at a much slower pace [6]. Similar observations can be found in [4] where the authors inferred a significantly longer duration of viral shedding in the asymptomatic carriers than the symptomatic group. They also observed a weaker immune response to the coronavirus infection in ACCs with lower levels of immunoglobulin IgG and neutralizing antibody.

The asymptomatic transmissions are challenging traditional symptom-based public health strategies [35, 36] and are regarded as the Achilles' heel in the COVID-19 control effort [37]. As advised by Oran and Eric [30], more innovative and cost-effective approaches are urgently required to detect the asymptomatic cases and disrupt the transmission chains.

2.2 Popular epidemiological models

To facilitate the prediction of how an epidemic would evolve under certain setting, experts in theoretical epidemiology have developed mathematical models to describe the transmission dynamics [38]. Among them, the popularly used compartmental models assign people in a population with different labels and represent the progress of the people in different compartments. Originated from an early work of Kermack and McKendrick in 1927 [39], the SIR (Susceptible-Infectious-Recovered) model, the basic form of this family [40], consists of three coupled ordinary differential equations to delineate the change of numbers of susceptible, infectious and recovered individuals. The dynamics of a considered disease demonstrated by the model are heavily dependent on some important parameters, which reflect not only the characteristics of the disease itself, but also the effectiveness of public intervention measures. Due to the hardship to obtain analytic solutions, approximation methods are usually used to compute the evolution of the numbers over time. By relaxing the assumptions in the original SIR model, researchers extend it into more sophisticated variants such as SEIR, SEIS and SLIS models and have applied them in describing the outbreak of SARS [41] and SARS-CoV-2 [42]. However, this kind of equation-based models can only represent an epidemic on a macro level due to their aggregated nature, but not at a human-to-human contact history in a population.

Besides compartmental models, agent-based modeling (ABM) is a currently popular approach for epidemiologists to study the effect of social behaviors on the dynamics of infectious diseases like Ebola [43]. Researchers have also developed an agent based micro-simulation of the COVID-19 epidemic in France and concluded that only consistently performing measures of lockdown, physical distancing and mask wearing can effectively protect the society during the pandemic [44]. The ABM models presented in [19, 45–47] were used to test the effect of "forward" or/and "backward" contact tracing on the epidemic control. However,

these models provide no means for identifying the invisible ACCs from the population based on the generated contact tracing data.

2.3 Motivation

As discussed above, a large number of ACCs pose extra difficulty to the containment of the COVID-19 epidemic. The traditional symptom-based public health strategies require expensive medical testing, and massive population screening for all positive cases cannot be very responsive to the fast evolving epidemic. Moreover, theoretic epidemiological models can only provide insights into the development of the disease, but not address the problem of tracking asymptomatic cases. In comparison, it would be a more cost-effective way to exploit the digital contact tracing data to swiftly identify potential transmission sources. It will undoubtedly help contain the transmission of the coronavirus in a continual manner as the contact tracing data is being continuously fed into the system. Moreover, to prove the efficacy of the proposed model, a realistic agent-based simulation is needed and it is expected to coincide with current epidemiological evidence on the epidemic.

3 Problem formulation and overall procedure

This paper aims to provide a computational solution to the problem of ACC identification from the perspective of cognitive computation. The approach taken is to model the spatiotemporal trajectories of different individuals in a collective manner and identify ACCs through comparing the collective episodic traces of *t*-SCCs and the untested individuals.

Consider a population of size N, each spatiotemporal data point of an individual can be represented by an *event* in the form of $\varepsilon = (t, p)$, where t is a time stamp in the unit of hours and p is the identifier of a place. We consider a quantized spatial and temporal scale, i.e., $0 \leq t < T$ and $0 \leq p < P$, where T is the duration during which the data points are collected and P is the number of all possible places in the environment. An *episodic trace* of an individual is then specified by a sequence of events described as $e = <\varepsilon_0, \varepsilon_1, ..., \varepsilon_{T'}>$.

All the individuals in the simulation can be classified into four categories according to their respective COVID-19 positivity as follows.

(1) The first group consists of symptomatic COVID-19 patients who have been diagnosed positive with the virus and are isolated after the onset of symptoms. They are thus referred to as tested-symptomatic COVID-19 cases (*t*-SCCs for short), or known cases. The lengths of episodic traces of *t*-SCCs could be $T' \leq T$ due to potential halfway isolation.

(2) The other three groups are untested cases having no sign of symptoms. Among them,

   (a) some are asymptomatic COVID-19 cases (ACCs),
   (b) some will be SCCs but are currently in their pre-symptomatic stage [48],
   (c) and the rest are uninfected people.

Therefore, the COVID-19 positivity of an individual with index $0 \leq i < N$ can be $CP_i \in \{0, 1\}$ where 1 indicates a positive and symptomatic patient while 0 is for an untested individual.

Given the input data as the episodic trace e_i of each individual i with the CP_i label, the target outputs of the proposed model are the likelihood of each untested individual i to be an ACC, defined as $L_i = Prob(ACC|CP_i = 0)$.

Based on the formulation, Fig. 1 illustrates the idea of COVID-19 data simulation and the process of building the collective episodic memory in STEM-COVID and identifying ACCs from the memory model. For generating input data for the proposed model by simulation, Sect. 4 presents the spatial representation, life cycle of agents and dynamics

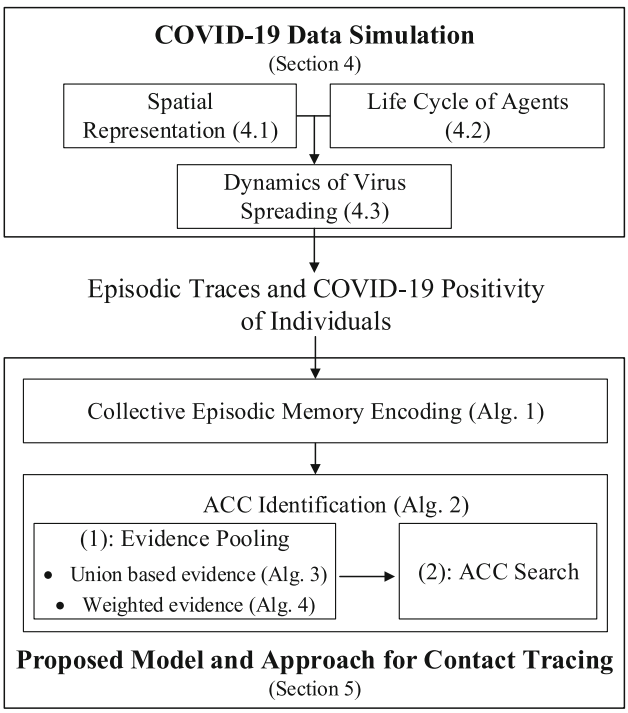


Fig. 1 A flowchart for the whole procedure of simulation and proposed model

of virus spreading in the simulation model. Based on the produced episodic traces and COVID-19 positivity of N individuals from simulation, STEM-COVID supports a two-step process of ACC identification as show in Algorithm 2. For the first step, this paper presents two different methods as shown in Algorithm 3 and 4, respectively, namely union-based evidence pooling and weighted evidence pooling. Subsequently, ACCs are screened through a comparison of similarities between the unified evidence vector and the episodic traces of untested individuals.

4 COVID-19 data simulation

For empirical performance evaluation, a simulation model of the real world, where no public health interventions are exerted, is designed to emulate the spreading of coronavirus in communities. This section will briefly introduce the spatial representation, the life cycle of individual agents, the temporal dynamics of virus transmissibility and the key parameter settings. Compared to the previous work [25], we extend the simulation model with more realistic distributions for incubation periods of SCCs and duration of viral shedding of ACCs, and more delicate models of the infectiousness profile of the positive cases.

4.1 Spatial representation

We discretize the simulated environment into a network of places that are reachable from each other. As shown in Fig. 2, the neighboring places can be connected by walkways or roads, and the whole space can be seen as a graph.

People undergo different levels of risks to be infected by a coronavirus carrier at different types of places, typically influenced by the crowd density, the degree of ventilation and many other factors. This is the reason why most of the transmission chains were found in family clusters where members infected outside then transmitted the virus to each others [8, 49]. In view of such findings, we categorize the places into different groups. Places with very high infection risk are mostly homes, nursing rooms, etc. Places such as cafes, restaurants and shopping malls are highly risky for consumers to be infected. Workplaces like offices and factories can be considered with medium risk and open spaces like playgrounds and parks carry low risk to people.

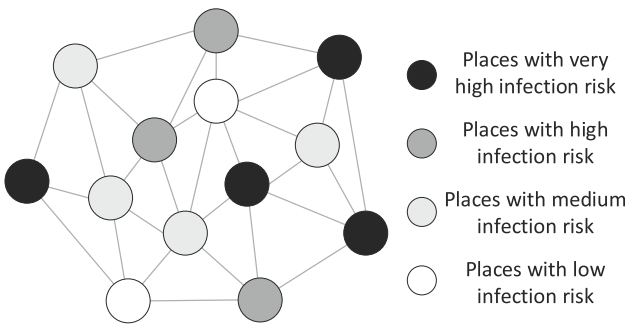


Fig. 2 A graphic representation of the spatial model

4.2 Life cycle of agents

To simulate the dynamic activities of individuals between different places, we model each individual as an agent who is assigned a daily life cycle, wherein they stay at various places for different periods of time. For example, the agents stay at home (very high-risk places) for about 10 hours, work at offices (middle-risk) for 8 hours, go shopping or socialize (high-risk) for 3 hours and do outdoor activities (low-risk) for 3 hours. For each day in the simulation, we sample the duration (in hours) for each agent to stay at different places using discrete uniform distributions as given in Eq. (1):

$$\begin{cases} T(vh) \sim U[9, 11] \\ T(m) \sim U[7, 9] \\ T(h) \sim U[1, 24 - T(vh) - T(m) - 1] \\ T(l) = 24 - T(vh) - T(m) - T(h), \end{cases} \tag{1}$$

where vh, h, m, l are abbreviations for the four types of places, $T(\cdot)$ represents the duration of staying at certain type of places for an agent in one day, and $U[a, b]$ is a sampling from a discrete uniform distribution based on the set $\{a, a+1, ..., b\}(a, b \in \mathbb{N}, a \leq b)$. The durations vary from day to day and from person to person.

Besides the duration of staying at different types of places, the simulation model additionally needs to address the places that the agents can visit and how those agents transit between places. For this, we divide all the P places into different numbers (i.e., P_{vh}, P_h, P_m and P_l) of various types of places through setting the ratio among them. For example, a typical life style would resolve one very high-risk place (home), four high-risk places (two for going out with family members while the other two for socializing with colleagues or friends), one middle-risk place (workplace) and one low-risk place (park). We design various simulation scenarios with different settings of the mentioned parameters, which are described in Sect. 6. Combined with the settings of $T(.)$ in one day and place assignment, we use a simulation paradigm of uniform time-step advancement (an hour as a unit) to drive the agents to transit between the assigned places in the order of very high, middle, high and low-risk. For example, after a duration of $T(vh)$, the agent will be moved to its assigned middle-risk place to stay for $T(m)$. Though more possibilities of life styles exist in reality, such as more types of duration and mixed visit order between places, our assumption is sensible because it represents a typical daily

life cycle for many people. And it is adequate to the efficacy of our approach.

4.3 Dynamics of virus spreading

Along with the transitions between locations, a simulation of virus spreading requires some additional consideration for the temporal dynamics in clinical features of the infected people. As many studies reported [2, 4, 8], SCCs are infectious not only after symptom onset, but also in their pre-symptomatic stage. Figure 3 illustrates the dynamics based on the epidemiological findings in [2]. As shown in the top plot in Fig. 3, the authors inferred that very few transmissions would occur before five days prior to the symptom onset, and the infectiousness monotonically increased to a peak around the onset of symptoms and declined quickly within several days. Although these works mainly focused on symptomatic cases, many other researches suggested similar viral loads and infectiousness development in ACCs [6, 34]. Therefore, this paper simulates the infectiousness profiles of both ACCs and SCCs based on the temporal dynamics mentioned above.

Since we assume that SCCs will be identified and isolated once they show any symptom, their contagiousness after the symptom onset will not be considered in the simulation. For simplification, this paper uses piecewise and linear functions to represent the evolving infection rates of SCCs and ACCs during their incubation periods (time from infection to symptom onset) and periods of viral shedding, respectively, as shown in Fig. 3. A simpler alternative in the previous work is just step functions to describe the infectiousness profiles [25]. In the beginning, i.e., 10%, of respective periods, both types of coronavirus cases are not infectious at all in the simulation because of low levels of viral loads. Each data point in the lines indicates the probability of infecting others at places with different levels of infection risks. The next subsection will elaborate on the settings of incubation periods of SCCs, duration of viral shedding of ACCs and the infection rates.

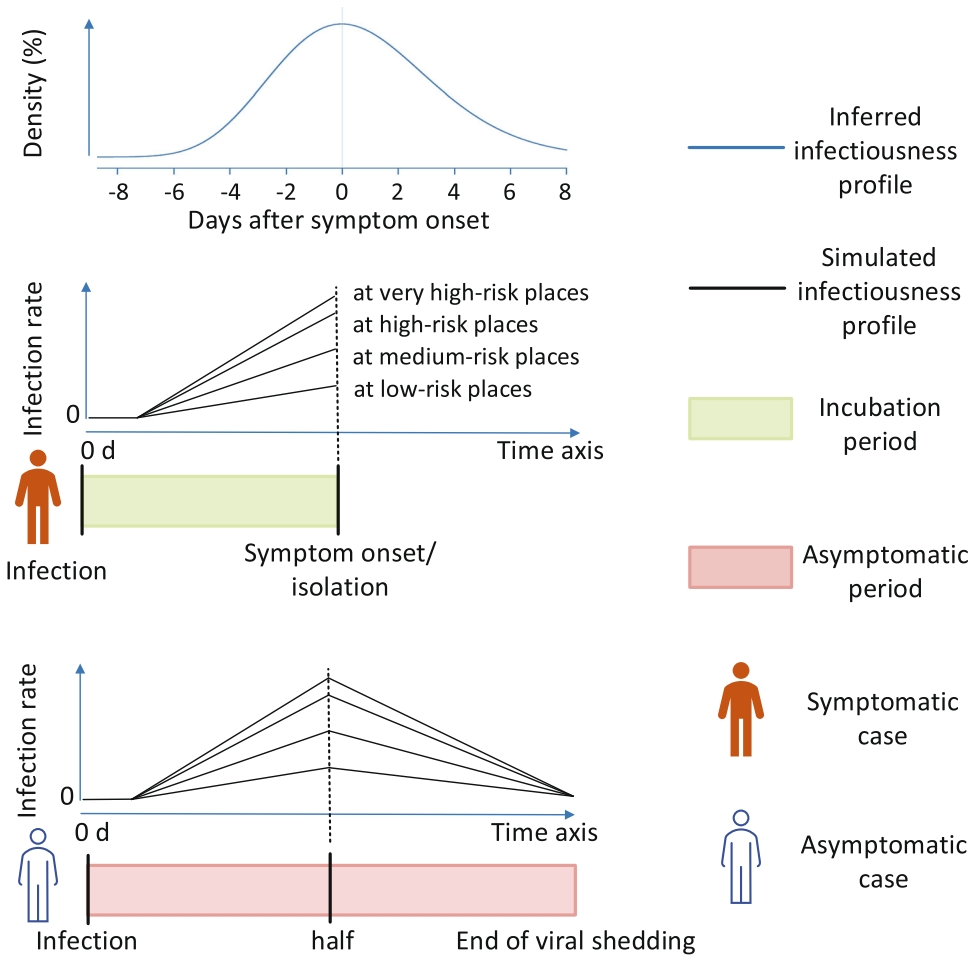


Fig. 3 The temporal dynamics of SCCs and ACCs, partly adapted from [2]

4.4 Parameter settings

Table 1 lists the key parameters used in the simulation model. The detailed settings for some of them are given as follows.

4.4.1 Incubation period

A number of studies have reported different estimations of the duration of incubation periods of SCCs [2, 50–53]. Among them, a highly cited work reported the mean incubation period to be 5.2 days, with the 95th percentile of the distribution at 12.5 days. In view of the estimations, we use a long-tailed Gamma distribution $T_{\text{in}} \sim Gamma(4.0, 1.3)$ to model an incubation period for each SCC, as shown in Fig. 4a. The mean value and the 0.95 quartile of 100,000 samples from this distribution are roughly 5.2 and 10.1 days, respectively, which are very close to the investigation in the real world.

4.4.2 Duration of viral shedding

According to the statistical results from [4], the group of asymptomatic cases shed the SARS-CoV-2 virus for a significantly longer duration than symptomatic patients. Specifically, the median duration of ACCs could be 19 days with the interquartile range (*IQR*) being from 15 to 26 days. As illustrated in Fig. 4b, another Gamma distribution is applied to sample a duration of viral shedding for each ACC, as $T_{\text{vs}} \sim Gamma(10.0, 1.9)$ whose median, Q1 and Q3 duration are roughly 18.4, 14.7 and 22.6 days, respectively.

4.4.3 Proportion of ACCs

As mentioned in Sect. 2, the existing estimations of the proportion of asymptomatic cases over all infections can be varied from 5 to 80% [29, 31]. But a large number of them fall into the range from 10 to 30%, such as the studies on cases in Zhejiang, China, the Diamond Princess cruise ship and Singapore [9, 10, 27]. In view of these epidemiological findings, this paper sets a percentage of 20% of all infections as ACCs.

4.4.4 Infection rates

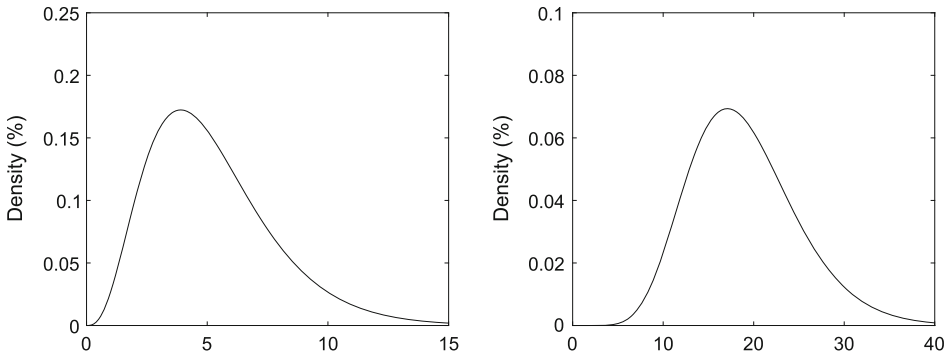
We set the infection rates in a unit of hours, at the peak of the infectiousness of a positive case, for the mentioned four types of places as $r_{vh} = 0.003, r_h = 0.0026, r_m = 0.002$ and $r_l = 0.0002$, respectively. Thus, the infection rate of a coronavirus carrier at any given time and place can be determined by these values and the piecewise linear function illustrated in Fig. 3. Note that these settings remain the same no matter whether the source of transmissions is symptomatic or asymptomatic, in view of the similar viral loads found in the different types of cases [6]. These numbers indicate the probabilities that a coronavirus carrier infects any other healthy individual within a time step in the simulation. With $k > 1$ coronavirus carriers at the same space and the same time period, we simply add their infection probabilities as long as the sum is not over 1. We will verify the settings through the comparison between the results of simulation and recent epidemiological findings about the epidemic.

5 Proposed approach: spatio-temporal data modeling and reasoning

This section reviews the basic of fusion ART and the STEM model. Subsequently, the architecture of the STEM-COVID model is presented, followed by the algorithms for collective episodic memory encoding and ACC search. Finally, an analysis of the computation complexity is provided.

Table 1 Summary of key parameters in the simulation model

Description of parameters	Values
[0.8pt] simulation duration T (unit: days)	20
number of individuals in a family N_f	4
incubation period T_{in} for each SCC (unit: days)	$Gamma(4.0, 1.3)$
duration of viral shedding T_{VS} of each ACC (unit: days)	$Gamma(10.0, 1.9)$
proportion of ACCs over all infected cases	20%
highest hourly infection rate in very high-risk places	0.3%
highest hourly infection rate in high-risk places	0.26%
highest hourly infection rate in medium-risk places	0.2%
highest hourly infection rate in low-risk places	0.02%



(a) Days from infection to symptom onset for SCCs **(b)** Duration when shedding the virus for ACCs

Fig. 4 Gamma distributions for incubation periods of SCCs and duration of viral shedding of ACCs

5.1 Fusion ART networks

The Adaptive Resonance Theory (ART) is a neural network architecture used to emulate how brains process information for pattern recognition and prediction, given bottom-up sensory information and top-down expectation [54, 55]. As shown in Fig. 5, fusion ART is a form of multi-channel ART which encodes multimodal input patterns as cognitive nodes (marked as ellipses) in the top F_2 layer and enables recognition and recall of stored patterns through the connections (marked as black semicircles) between the F_1 and F_2 layers [56–59].

Input vectors: Let $\mathbf{I}^k = (I_1^k, I_2^k, ..., I_l^k)$ be an input vector to the channel k, where $I_i^k \in [0,1]$ for $i = 1, 2, ..., l$ and $k = 1, 2, ..., n$.

Input fields: Let F_1^k be the kth input field in the bottom F_1 layer and $\mathbf{x}^k = (x_1^k, x_2^k, ..., x_l^k)$ be the activity vector of this field after receiving $\mathbf{I}^k$. We have $x_i^k = I_i^k$ in common cases. If complement coding is applied to prevent the "code proliferation" problems [60], the activity vector is further augmented with a complemented vector $\overline{\mathbf{x}}^k$, where $\overline{x}_i^k = 1 - x_i^k$ for $i = 1, 2, ..., l$. For more comprehensive description about complement coding, please refer to [56].

Category fields: Let $F_i (i > 1)$ be one category field, where the learned cognitive nodes are stored. The standard multichannel ART has only one category field F_2. Let $\mathbf{y} = (y_1, y_2, ..., y_m)$ be the activity vector of F_2, formed from the activation values of the cognitive nodes, where m is the number of currently learned categories.

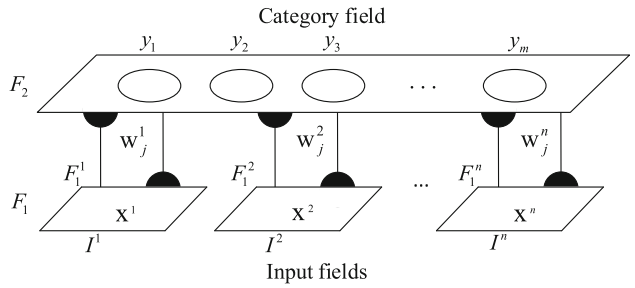


Fig. 5 The architecture of fusion ART

Weight vectors: Let $\mathbf{w}_j^k$ be the weight vector associated with the jth node in F_2 for learning the input patterns of F_1^k.

Parameters: The dynamics of each input field k is determined by its choice parameter $\alpha^k \geq 0$, learning rate β^k, contribution parameter γ^k and vigilance ρ^k, where $\beta^k, \gamma^k, \rho^k \in [0,1]$.

The dynamics of fusion ART using fuzzy ART operations are briefly described as follows. A more detailed description can be found in [58].

The **Code Activation** process on the jth node in F_2, when all input fields are activated with $\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^n)$, is controlled by the choice function given in Eq. (2):

$$T_j = \sum_{k=1}^{n} \gamma^k \frac{|\mathbf{x}^k \wedge \mathbf{w}_j^k|}{\alpha^k + |\mathbf{w}_j^k|}, \tag{2}$$

where the fuzzy AND operation $\wedge$ is defined by $(\mathbf{p} \wedge \mathbf{q})_i \equiv \min(\mathbf{p}_i, \mathbf{q}_i)$ and the norm $|\cdot|$ is defined by $|\mathbf{p}| \equiv \sum_i \mathbf{p}_i$.

A **Code Competition** process then selects an F_2 node J with the highest T_j followed by a **Template Matching** to check if resonance occurs between the current input pattern $\mathbf{x}$ and the selected node, following the rule given in Eq. (3):

$$m_J^k = \frac{|\mathbf{x}^k \wedge \mathbf{w}_J^k|}{|\mathbf{x}^k|} \geq \rho^k, 1 \leq k \leq n. \tag{3}$$

If no selected node in F_2 meets the vigilance in exhaustive iterations of code competition and template matching, an uncommitted node will be recruited in F_2 as a new category node so that the network automatically grows.

A **Template Learning** process is applied to the connection weights once resonance occurs. For each channel k, the weight vector $\mathbf{w}_J^k$ is modified by the learning rule given in Eq. (4):

$$\mathbf{w}_J^{k(\text{new})} = (1 - \beta^k)\mathbf{w}_J^{k(\text{old})} + \beta^k(\mathbf{x}^k \wedge \mathbf{w}_J^{k(\text{old})}). \tag{4}$$

The weights of a newly committed node J are initialized by an overwriting learning as $\mathbf{w}_J^k = \mathbf{x}^k, 1 \le k \le n$. In turn, the learned pattern in a selected node J in F_2 may perform an **Activity Readout** to all corresponding F_1^k fields by $\mathbf{x}^{k(\text{new})} = \mathbf{w}_J^k$ as the output.

5.2 The STEM architecture

Previous models of ART-based episodic memory include Episodic Memory-ART (EM-ART) [22, 23] and STEM [24]. While EM-ART aims to learn the temporal ordering of events within an episode by a gradient encoding scheme, STEM encodes and recalls events by explicitly representing the multimodal activity patterns of events without an episode layer.

As shown in Fig. 6a, the original STEM architecture includes four input fields to represent the context of an event. The *object* field represents the presence of people and physical objects appeared in the event, the *activity* field encodes the occurred activity and the *time* field stores the time of occurrence. In particular, the *place* property of an event in STEM is characterized by the real-world coordinates and landmark specification such as "reception" and "lift". Through the generalization by fusion ART, each node learned in the place field represents a group of visited locations. On top of the four fields, each category node in F_3 represents an event in response to the multimodal information presented in F_2.

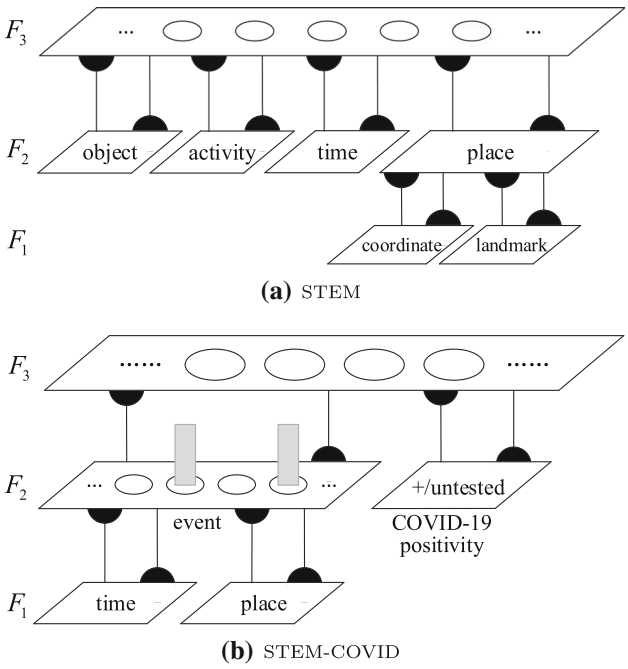


Fig. 6 The architectures of STEM and STEM-COVID

Besides efficient encoding of events, experiments on a public data set also demonstrated the capability of STEM for robust retrieval of events under partial or noisy cues.

5.3 The STEM-COVID architecture

Learning episodic memory to identify ACCs, as the task in this paper, requires explicitly encoding the spatiotemporal contextual information of events as in STEM. However, the new model differs much from STEM in several respects. First, the real-world coordinates of places are of less significance than the transitions of agents between places in the task. Second, besides generalizing the input patterns of single events as performed in STEM, the new memory model in this paper needs to learn the episodic trace of each individual, which consists of patterns of multiple different events. Third, on the same level of episodic traces, different individuals need to be further categorized by the COVID-19 positivity conditions, as formulated in Sect. 3. Fourth, this paper is concerned about the episodic memory of a population instead of a single person and thus a higher-level representation is in demand to learn the collective memory.

In view of these considerations, this paper designs the STEM-COVID model as shown in Fig. 6b, which can be viewed as 2 two-channel fusion ART networks stacked in a hierarchical manner. The bottom network can be seen as a simplified edition of STEM, which encodes *event nodes* in the *event* field in the F_2 layer based on the time and place attributes. In this middle layer, the model also incorporates another specialized field to indicate the COVID-19 positivity of an individual. In turn, the top network aggregates the episodic trace of an individual with his/her COVID-19 positivity into a single node in the F_3 layer. This learning scheme bears some similarity with that in EM-ART that represents an episode as a cognitive node in its F_3 layer. However, besides the absence of the positivity field, EM-ART differs also by its gradient encoding method for sequential memory encoded in the activity vector in F_2.

As an intermediate memory buffer, the *event* field in STEM-COVID serves as both a category field of the bottom network and an input field of the top one. After the events on the episodic trace of an individual are presented and learned by the bottom network, the episodic trace is represented by the activity vector $\mathbf{y}$ of the event field, denoted as $\mathbf{e}$ here for clarity. Each position in this vector indicates a single event experienced by any individual whose spatiotemporal trajectory has been encoded. Specifically, values of 1s in $\mathbf{e}$ correspond to events experienced by the current individual, marked by the gray bars in Fig. 6b, while the rest are zeros.

Besides the event field, the other field in F_2 indicates the COVID-19 positivity of an individual i by a complement

coded positivity vector $\mathbf{c} = (CP_i, \overline{CP_i})$, where $\overline{CP_i} = 1 - CP_i$. In the next step, the vectors $\mathbf{e}$ and $\mathbf{c}$ are used as the activity vectors of the two input fields for the top network, which learns a unique code for each individual in the F_3 layer. In general, the event field and F_3 layer (i.e., the two category fields of the bottom and top network) continuously grow in response to novel patterns of events and individuals.

5.4 Collective episodic memory encoding

As mentioned, episodic memory considers an individual episode as a sequence of events as $e = <\varepsilon_0, \varepsilon_1, ..., \varepsilon_{T'} >$, where an event is formalized as $\varepsilon_l = (t_l, p_l)$. We present a hierarchical scheme for constructing collective episodic memory based on the STEM-COVID architecture, as shown in Algorithm 1.

Algorithm 1 Incremental construction of collective spatio-temporal episodic memory.

Require: episodes of all N individuals $\{e_i | 0 \leq i < N\}$
Require: positivity of all individuals: $\{CP_i\}$
Ensure: learned episodic memory in STEM-COVID

1: **for** $e_i =< \varepsilon_0, \varepsilon_1, ..., \varepsilon_{T'} >$ **do**
2: initialize a set of activated codes $J\text{s} \leftarrow \varnothing$;
3: ▷ learning in bottom network
4: **for** $\varepsilon_l = (t_l, p_l)$ **do**
5: initialize a code flag: $J \leftarrow -1$;
6: normalize: $t' \leftarrow t_l/T, p' \leftarrow p_l/P$;
7: present $\mathbf{I} = (\mathbf{I}^t, \mathbf{I}^p) = ((t'), (p'))$ to F_1;
8: **for** code j in episode field in F_2 **do**
9: **if** $|(\mathbf{w}_j^t, \mathbf{w}_j^p) - \mathbf{I}| < \delta$ **then**
10: $J \leftarrow j$;
11: break;
12: **end if**
13: **end for**
14: **if** $J == -1$ **then**
15: create a new code J with $\mathbf{w}_J^t = (t')$ and $\mathbf{w}_J^p = (p')$;
16: expand the field as an input to F_3;
17: **end if**
18: $J\text{s} \leftarrow J\text{s} \cup \{J\}$;
19: **end for**
20: ▷ learning in top network
21: form a binary activity vector $\mathbf{e}$: $\mathbf{e}_j = 1$, if $j \in J\text{s}$;
22: form an activity vector of positivity: $\mathbf{c} = (CP_i, 1 - CP_i)$;
23: present the activity vectors $(\mathbf{e}, \mathbf{c})$ to F_2;
24: recruit an uncomitted code J in F_3;
25: do an overwrite learning: $\mathbf{w}_J^e = \mathbf{e}$ and $\mathbf{w}_J^c = \mathbf{c}$;
26: **end for**

The algorithm consists of a loop through the episodic traces of all individuals in the considered population (Line 1-25). To encode an event, the algorithm first normalizes the time and place attributes into real values t' and $p' \in [0, 1)$, respectively (Line 6), presented to the input fields in the F_1 layer (Line 7). Note that the input vectors $\mathbf{I}^1$ and $\mathbf{I}^2$ in fusion ART are here denoted as $\mathbf{I}^t$ and $\mathbf{I}^p$ for clarity. A standard learning process in fusion ART networks will then involve code activation and resonance search whereby nodes in the F_2 layer are visited through some iterations of code competition and template matching. However, since an event can be considered unique with a distinct combination of time stamp and place ID, the search for resonance in STEM-COVID can be implemented by a simplified code activation and competition without template matching as given in Eq. 3 (Line 8-12). During a loop through activating each existing event node in the F_2 layer, once a match between an event node and the current input pattern is perfectly reached (Line 9), it is considered that the node has won the code competition without exhausting all remaining nodes (Line 10-11). This is because no any other can match the input pattern any better. Then the input can be categorized into the found

event node j represented by the weights $\mathbf{w}_j^t$ and $\mathbf{w}_j^p$. Furthermore, the matching is implemented by a linear function instead of the standard choice function (Line 9).

On the other hand, if the loop through all existing event nodes comes to no match, a new uncommitted event node will be recruited to encode the novel input (Line 15). In such a case, the event field will automatically grow, followed by the expansion of the associated weights $\mathbf{w}_j^e$ of all F_3 nodes j (Line 16).

As the entire trajectory and positivity status of an individual are formed as the activity vectors $\mathbf{e}$ and $\mathbf{c}$ of the two corresponding fields in F_2, the information can be directly encoded and stored as a unique node with weights $\mathbf{w}_J^e$ and $\mathbf{w}_J^c$ in the F_3 layer (Line 21-25). Therefore, the learning in the top network also avoids invoking iterations of bottom-up activation and top-down matching to seek resonance.

5.5 ACC identification based on STEM-COVID

Based on the collective episodic memory learned in the STEM-COVID model, a two-step process is proposed to identify unknown ACCs based on the spatiotemporal trajectories of the known SCCs, as shown in Algorithm 2. The first step aims to readout and combine the episodic traces of all t-SCCs into a specialized episode vector called evidence vector $\mathbf{e}^{ev}$ (Line 1-6), which is used as a template episodic trace in the second step to compute the similarity with the trace of each untested individual (Line 7-13). To this end, the positivity field in F_2 is the only contributive field in the first step for searching for all t-SCCs, whereas the second step needs both episode and positivity fields while the pooled trace $\mathbf{e}^{ev}$ and $\mathbf{c} = (0, 1)$ are being presented.

5.5.1 Evidence pooling

For pooling the episodic traces of known COVID-19 positive cases, the contribution parameters of the two fields in F_2 are set to $\gamma^e = 0$ and $\gamma^c = 1$ (Line 1). The specific episode vector of each individual, learned in the connection weights $\mathbf{w}_j^e$, is thus ignored when the corresponding node is activated. Consequently, all nodes in F_3, encoding COVID-19 positive cases, are then activated by the choice function wherein the positivity vector $\mathbf{c} = (1, 0)$ is presented (Line 2-4). Based on the activated F_3 nodes, we present two methods to pool their encoded episodic traces into a single template vector (Line 6) described as follows.

Algorithm 2 A two-step framework for identifying ACCs.

Require: episodic memory learned in STEM-COVID
Ensure: a set of identified individuals

▷ Step 1: Evidence activity pooling

1: set contribution parameters: $\gamma^e \leftarrow 0, \gamma^c \leftarrow 1$;
2: present a positivity vector to F_2: $\mathbf{c} \leftarrow (1, 0)$;
3: **for** code j in F_3 **do**
4: $\quad T_j = \gamma^c \frac{|\mathbf{c} \wedge \mathbf{w}_j^c|}{\alpha^c + |\mathbf{w}_j^c|}$;
5: **end for**
6: $\mathbf{e}^{ev} \leftarrow$ `evidence_pooling`();

▷ Step 2: ACC Search

7: set contribution parameters: $\gamma^e \leftarrow 1, \gamma^c \leftarrow 1$;
8: set a positivity vector: $\mathbf{c} \leftarrow (0, 1)$;
9: present $\mathbf{e}^{ev}$ and $\mathbf{c}$ to F_2;
10: **for** code j in F_3 **do**
11: $\quad T_j = \gamma^e \frac{|\mathbf{e}^{ev} \wedge \mathbf{w}_j^e|}{\alpha^e + |\mathbf{w}_j^e|} + \gamma^c \frac{|\mathbf{c} \wedge \mathbf{w}_j^c|}{\alpha^c + |\mathbf{w}_j^c|}$;
12: **end for**
13: return $\{L_j = T_j\}$;

Algorithm 3 Union based evidence pooling.

Require: episodic memory learned in STEM-COVID
Require: the activation values T_j of F_3 nodes
Ensure: an evidence vector
1: initialize a binary evidence vector $\mathbf{e}^{ev} \leftarrow \mathbf{0}$;
2: **for** code j in F_3 with $0 < T_j \leq 1$ **do**
3: readout node j to episode field: $\mathbf{x}^e \leftarrow \mathbf{w}_j^e$;
4: update the evidence vector: $\mathbf{e}^{ev} \leftarrow \mathbf{x}^e \vee \mathbf{e}^{ev}$;
5: **end for**
6: return $\mathbf{e}^{ev}$;

Union-based evidence pooling Our previous work developed a union-based scheme of evidence pooling to unify the episodic traces of all *t*-SCCs [25]. As shown in Algorithm 3, it merges the binary episode vectors $\mathbf{e}$ of F_3 nodes representing *t*-SCCs into a binary evidence vector $\mathbf{e}^{ev}$ (Line 1-4). Specifically, the episode vector of each *t*-SCC is readout from the weights to the activity vector $\mathbf{x}^e$ in the event field (Line 3). All such vectors are incrementally integrated into the initially empty vector $\mathbf{e}^{ev}$ based on a simple fuzzy OR operation wherein $(\mathbf{x}^e \vee \mathbf{e}^{ev})_i = \max(\mathbf{x}_i^e, \mathbf{e}_i^{ev})$ for $i = 1, 2, ..., n_\varepsilon$, where n_ε is the number of existing event nodes in the event field and the length of $\mathbf{e}^{ev}$ as well (Line 4). In such a way, the final $\mathbf{e}^{ev}$ encodes which events all the *t*-SCCs have experienced in a collective manner.

Weighted evidence pooling

Although the method described above is simple and straightforward, each binary element $\mathbf{e}_i^{ev}$ of the resultant evidence vector is only capable of implying whether any *t*-SCC ever experienced the corresponding *i*th event, but offers no hint about the number of *t*-SCCs appearing in the same context. This piece of information, however, is important for identifying ACCs in the sense that if more COVID-19 patients have shared the same space at the same time with one individual, it is highly plausible to consider the individual as a more likely source of infection.

In view of this consideration, this work proposes a weighted evidence pooling scheme in Algorithm 4. By looping through all F_3 nodes with positive activation values, this method combines all the encoded binary episode vectors with an element-wise accumulation so that $\mathbf{e}_i^{ev}$ equals to the number of *t*-SCCs who experienced the corresponding spatiotemporal context (Line 2-4). In the end, the evidence vector is normalized by its maximum element so that $\mathbf{e}_i^{ev} \in [0, 1]$ is a real number instead of a Boolean flag (Line 6-7). In this way, the weights on different spatiotemporal contexts in the pooled evidence vector are proportionate to the numbers of known coronavirus carriers who have experienced the respective events. When searching for ACCs, high similarities to the evidence vector should be assigned to the episodic traces of individuals who (a) have experienced many of the same events as *t*-SCCs and (b) have ever appeared in some spatiotemporal context where more than one otherwise healthy person got infected.

Algorithm 4 Weighted evidence pooling.

Require: episodic memory learned in STEM-COVID
Require: the activation values T_j of F_3 nodes
Ensure: an evidence vector
1: initialize evidence vector $\mathbf{e}^{ev} \leftarrow \mathbf{0}$;
2: **for** code j in F_3 with $0 < T_j \leq 1$ **do**
3: readout node j to episode field: $\mathbf{x}^e \leftarrow \mathbf{w}_j^e$;
4: update the evidence vector: $\mathbf{e}^{ev} \leftarrow \mathbf{x}^e + \mathbf{e}^{ev}$;
5: **end for**
6: $e^{max} \leftarrow \max_{\forall i=1,2,..,n_\varepsilon} \mathbf{e}_i^{ev}$;
7: normalize the evidence vector: $\mathbf{e}^{ev} \leftarrow \mathbf{e}^{ev}/e^{max}$;
8: return $\mathbf{e}^{ev}$;

5.5.2 ACC search

Following evidence pooling, searching for ACCs can be done by computing the similarities between the unified trace and the traces of all untested individuals. Accordingly, the contribution parameters in the F_2 layer are set as $\gamma^e = \gamma^c = 1$ and the positivity vector $\mathbf{c}$ is set as (0, 1) ignoring all known positive cases (Line 7-8 in Algorithm 2). With the activity vectors $\mathbf{e}^{ev}$ and $\mathbf{c}$, every single node in the F_3 layer will be activated by the choice function (Line 9-11).

The choice function is essentially an evaluation of the similarity between the current input pattern and a cognitive code with respect to the already learned pattern. The activation value for an untested individual indicates the likelihood of this individual to be an ACC. Such a high value can only be obtained by a larger value of $|\mathbf{e}^{ev} \wedge \mathbf{w}_j^e|$, indicating that:

(1) the number of events shared by the corresponding untested individual and any SCC is large, or
(2) in some spatiotemporal contexts experienced by the individual, many otherwise healthy people were infected with the coronavirus.

The potential ACCs can be identified by selecting the top k individual nodes based on the activation values (i.e., the respective likelihoods to be an ACC), where k is a user-defined parameter. When a community or a city only has limited medical resources to test for potentially positive cases, the officials in the health section may determine a relatively small value of k. Otherwise, choosing a larger value of k can help screen more hidden spreaders of the virus.

5.6 Complexity analysis

This subsection analyzes the space and time complexities of the episodic memory construction and the ACC search algorithms presented above. The involved parameters include the population size N, the time duration of data collection T, the number of different places in the simulation P, the number of untested individuals N_u, the number of t-SCCs $N_t = N - N_u$ and the number of current event nodes in STEM-COVID n_ε. For space complexity, we are mainly concerned about the space requirement for storing all the event nodes and individuals nodes. Then we consider the time complexity of learning the memory model by Algorithm 1 and searching for ACCs by Algorithm 2. Finally for comparison, we present a baseline of ACC identification and give its time complexity.

5.6.1 Space complexity of STEM-COVID

The worst-case space complexity occurs when all individuals share no common events so that the number of event nodes can be $T \cdot N$. However, there can only be $T \cdot P$ combinations of time and place indicators. Therefore, the worst-case space complexity for the event field is $O(T \cdot \min(N, P))$. For encoding a population of size N, the total space complexity in F_3 is given by $O(NT\min(N, P))$, considering that the connection weights between each individual node in F_3 and the event field in F_2 involve all event nodes.

5.6.2 Time complexity of STEM-COVID

Constructing the memory model We first analyze the time complexity of learning a single event or individual node when the number of existing event nodes is n_ε. In the standard fusion ART model, the worst-case time complexity for learning a category node is $O(mn^2)$ if no matched node is found, where m is the number of attributes in the input fields and n is the number of existing category nodes [24]. The quadratic component is incurred by the repeated code competition and template matching for resonance search. However, we substitutes a loop of top-down matching through all existing event nodes for such iterations (Line 8-12 in Algorithm 1). By this simplification, the worst-case time complexity is reduced to $O(mn_\varepsilon)$, where $m = 2$ given the two single-attribute fields in F_1. Due to no matching required, the time complexity of learning a single individual node in F_3 (Line 21-25 in Algorithm 1) is just $O(n_\varepsilon)$.

Next we look into the growing complexity of learning event and individual nodes because of the dynamic increase of n_ε as the ongoing learning. In the worst case, all individuals share no common events and the lengths of all episodic traces are T. When the episodic trace of the ith individual is presented, there would have been iT event nodes in F_2. As such, the time complexity of learning all the T events of this individual is calculated as $2\{iT + (iT + 1) + ... + [(i + 1)T - 1]\}$ according to the analysis in the last paragraph. After the entire episodic trace is learned, encoding the ith individual node in F_3 incurs the time complexity of $O((i + 1)T)$ given that $n_\varepsilon = (i + 1)T$ now.

As a result, the overall time complexity of learning the event and individual nodes for all the individuals $(i = 0, 1, ..., N - 1)$ is then $O(N^2T^2)$. Given that the final total of event nodes is limited to $O(T\min(N, P))$ as analyzed above, the time complexity of constructing the memory model by Algorithm 1 should be $O(NT^2\min(N, P))$.

Algorithm 5 A baseline of computing trace similarities.

Require: merged events $\{\varepsilon_t\}$ of N_t t-SCCs
Require: episodes of N_u untested individuals
Ensure: similarity of each untested individual
1: **for** untested individual $a_{i=0,1,...,N_u-1}$ **do**
2: intialize a counter $c \leftarrow 0$;
3: retrieve the episode $e_i =< \varepsilon_0, \varepsilon_1, ..., \varepsilon_{T-1} >$;
4: **for** ε_j in e_i **do**
5: **for** $\varepsilon_l \in \{\varepsilon_t\}$ **do**
6: **if** $\varepsilon_j == \varepsilon_l$ **then**
7: $c \leftarrow c + 1$;
8: break;
9: **end if**
10: **end for**
11: **end for**
12: calculate the trace similarity $s_i \leftarrow c/T$;
13: **end for**

Completing the ACC detection Algorithm 2 includes time complexity of two steps. For evidence pooling, N individual nodes are activated (Line 4 in Algorithm 2) and activity readout and fuzzy OR operations or element-wise additions are performed for N_u times (Line 2-4 in Algorithm 3 or Line 2-4 in Algorithm 4). Since activating or reading-out each code involves operations on the weights for n_ε event nodes, the time complexity in the first step can be written as $O((N_u + N)n_\varepsilon)$. Similarly, the second step, when N nodes in F_3 are activated, would have a time complexity of $O(Nn_\varepsilon)$ (Line 11 in Algorithm 2). Overall, the time complexity of STEM-COVID for ACC identification is $O(NT\min(N, P))$.

5.6.3 Time complexity of baseline algorithm

To show the efficiency of our algorithm for ACC detection, we present a naive algorithm as a baseline for computing the similarity between the unified events of t-SCCs and the trace experienced by each untested individual. It can be regarded as a brute force version of STEM-COVID with the union based evidence pooling, so they share similar accuracy of recognizing ACCs but with different time complexity. As shown in Algorithm 5, the baseline simply counts the number of events commonly experienced by an untested individual and any t-SCC and considers the ratio of this count to the number of all events experienced by the untested individual as a metric of similarity (Line 4-12). The whole process applies three nested loops, thus involving $N_u \cdot T \cdot n_t$ times of event matching, where $n_t = |\{\varepsilon_t\}|$, at most equaling to $N_t \cdot T$. Hence the worst-case time complexity of this baseline is $O(N_u N_t T^2)$, which is much higher in most cases than the time complexity of STEM-COVID for ACC identification.

6 Experiments

This section will report on the performance of STEM-COVID in terms of its effectiveness, robustness and efficiency for identifying ACCs. The settings of eleven different scenarios and performance measures will be first described, before the quantitative results and discussion.

6.1 Simulation scenarios

All simulations run for twenty days, i.e., $T = 480$ h. For simplicity, every four agents are clustered in one household and extension to more complex family structures can easily be applied in the model. The simulation model is run under eleven different scenarios and the key settings of four scenarios are listed in Table 2. Besides the population size N and the numbers of different types of places (i.e., P_{vh}, P_h, P_m and P_l), N_{u_0} agents are randomly selected as the index cases (also referred to as 'patient zero') who are asymptomatic and infectious from the very beginning of the simulation.

The infectiousness profiles of ACCs and SCCs in scenario S200N, a simplified edition of S200P, are represented by a step function, where the infection rates switch from 0 directly to half of the peaking infection rates, instead of continuously increasing from 0 to the peak. This follows the setting in the previous work [25]. The scenario S200SE differs from S200P, in which a superspreader event is

Table 2 Summary of settings in the four scenarios

Metrics	**S200N**	**S200P**	**S200SE**	**S1000P**
N	200	200	200	1000
N_{u_0}	1	1	1	5
P_{vh}	50	50	50	250
P_h	2	2	2	10
P_m	10	10	10	50
P_l	2	2	2	10
infectiousness profile	step function	piecewise,linear	piecewise, linear	piecewise, linear
superspreader event	×	×	√	×

arranged on the 6th day in the simulation, when 50 individuals are involved in a crowded indoor gathering and 5 out of them are then coronavirus carriers [61]. The reason why we have different settings of infection rates in S200N, S200P and S200SE should be attributed to the current shift of dominant SARS-CoV-2 variants (e.g., Alpha found in UK and Delta first identified in India) [1] [2], which have increasing transmissibility. Especially, the superspreader event can be manifested as the change of dominant variants from Vanilla to Alpha to Delta to Omicron now. We set these scenarios trying to mimic the realistic change and investigate the effectiveness of our model under different situations. The last scenario S1000P listed in Table 2, with a population size and spatial scale fivefold as those in S200P, is used to investigate the effect of problem size on performance.

Besides the four scenarios described above, four other sets of experiments, named S200P-20, S200P-40, S200P-60 and S200P-80, are set to assess the robustness of our model in a realistic world where the contact tracing data are incomplete. They employ the same simulation data as S200P, but with a varying percentage (20%, 40%, 60% and 80%) of data points on the spatiotemporal trajectories of individuals randomly discarded. This represents different degrees of noisy and incomplete data.

In addition, although we have set the proportion of ACCs among all positive cases to be 20% in the main experiments based on the literature, this paper also tests the effects of proportions of ACCs among all positive cases on the identification accuracy of the proposed model. The rationale behind this is that according to many recent reports from agencies like Reuters[3] and CNBC[4], there are many breakthrough infections after being partially or fully vaccinated are asymptomatic in countries such as Indonesia and the US. To this end, the experiments include another three scenarios, named S200P-ACC10, S200P-ACC30 and S200P-ACC40, which change the proportion of ACCs in S200P(-ACC20) to 10%, 30% and 40%.

The simulation model runs for a total of 15 times under each scenario described above with different random seeds to generate multiple simulation instances for statistical results.

6.2 Performance measures

Based on the collective memory, STEM-COVID would select top-k untested individuals with the highest similarities of the episodic traces to the pooled evidence vector. A run of ACC detection based on the simulation and STEM-COVID is considered successful if one of the top-k individuals is a true ACC. Since each simulation scenario will be run 15 times, the *top-k success rate* for ACC detection, i.e., the ratio of the number of successful runs to 15, is then used to demonstrate the detection accuracy of the model under that scenario. Beyond normal ACCs, it will be also helpful if the index cases can be recognized, since they could be superspreaders who can potentially infect many other. Therefore, we also have a similar *top-k success rate* for detecting index cases.

Recall that we have theoretically compared the time complexity of STEM-COVID and the baseline for ACC search. In practice, we will present the time costs in seconds required by them to complete the computation of trace similarities in the simulations. All the experiments are run on a laptop with Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz.

6.3 Results and discussion

Before looking into the performance on ACC detection, we first summarize the generated data of the simulations.

[1] https://www.bbc.com/news/health-57431420.

[2] https://www.bbc.com/news/health-55659820.

[3] https://www.reuters.com/world/asia-pacific/hundreds-indonesian-doctors-contract-covid-19-despite-vaccination-dozens-2021-06-17/.

[4] https://www.cnbc.com/2021/06/25/covid-breakthrough-cases-cdc-says-more-than-4100-people-have-been-hospitalized-or-died-after-vaccination.html.

Table 3 Summary of simulation data in the four scenarios

scenarios	# ACC	# *t*-SCC	# SCC	# Untested cases
S200N	17.5	32.3	65.9	116.6
S200P	14.8	26.7	56.0	129.2
S200SE	24.8	61.2	100.2	75.0
S1000P	80.7	151.3	305.5	613.8

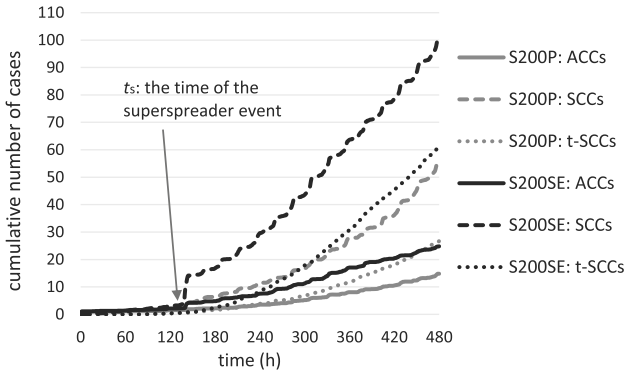


Fig. 7 Growth of COVID-19 cases in S200P and S200SE scenarios

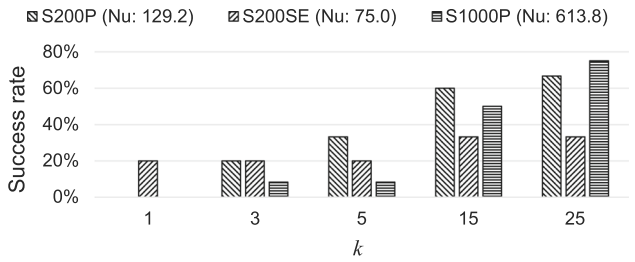


Fig. 9 Top-*k* success rates for identifying index cases. As a reference, there are 1, 1 and 5 such cases in S200P, S200SE and S1000P, respectively

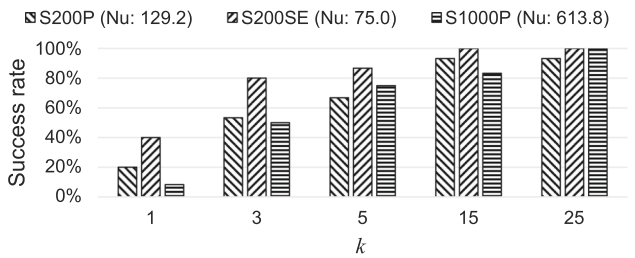


Fig. 8 Top-*k* success rates for identifying ACCs. As a reference, there are 14.8, 24.8 and 80.7 such cases in S200P, S200SE and S1000P, respectively

Finally, we will compare the time overheads required by STEM-COVID and the baseline in identifying ACCs.

6.3.1 Summary of simulations

Table 3 presents the mean numbers of different types of cases in the four scenarios listed in Table 2, averaged over the 15 simulations. With more realistic piecewise functions to simulate the development of infectiousness, there are slightly larger numbers of infected individuals in S200P than in S200N. Based on S200P, S200SE sees a substantial increase of numbers of both ACCs and SCCs due to an indoor gathering. As expected, the results of all metrics in S1000P are roughly five times of those in S200P.

More specifically, Fig. 7 shows the growth of ACCs, *t*-SCCs and all SCCs during the simulation under scenarios S200P and S200SE. The exponential curves imply a rapid increase of cases. Notice that it takes about 170 hours in S200SE to increase the number of SCCs from 50 to 100. The result tallies very well with many epidemiological investigations on early transmissions of COVID-19 that its basic reproductive number (R_0) is over 2 and the doubling time of the cases is around 7 days [50, 62, 63].

Moreover, the superspreader event in S200SE, occurred at t_s marked in Fig. 7, produces a surge in the number of coronavirus carriers and leaves the eventual numbers of ACCs and SCCs roughly double of those in S200P. The striking contrast between the two scenarios manifests that indoor gatherings of large numbers of people can easily drive the epidemic out of control.

6.3.2 Effectiveness for ACC identification

Recall that we presented a weighted evidence pooling approach in this paper, as an enhanced version of the union based scheme [25]. Table 4 reports the top-*k* success rates

Table 4 Comparison of effectiveness of the two evidence pooling methods, where 'wt' indicates the weighted scheme

success rates (%)	**S200N**		**S200P**		**S200SE**		**S1000P**	
	union	wt	union	wt	union	wt	union	wt
top-1	0.0	**13.3**	6.7	**20.0**	**40.0**	**40.0**	6.7	**8.3**
top-3	**33.3**	**33.3**	13.3	**53.3**	73.3	**80.0**	26.7	**50.0**
top-5	46.7	**60.0**	33.3	**66.7**	80.0	**86.7**	46.7	**75.0**
top-15	73.3	**93.3**	73.3	**93.3**	**100.0**	**100.0**	**93.3**	83.3
top-25	**100.0**	93.3	**100.0**	93.3	**100.0**	**100.0**	93.3	**100.0**

Bold values indicate the better results obtained by the two schemes in each scenario

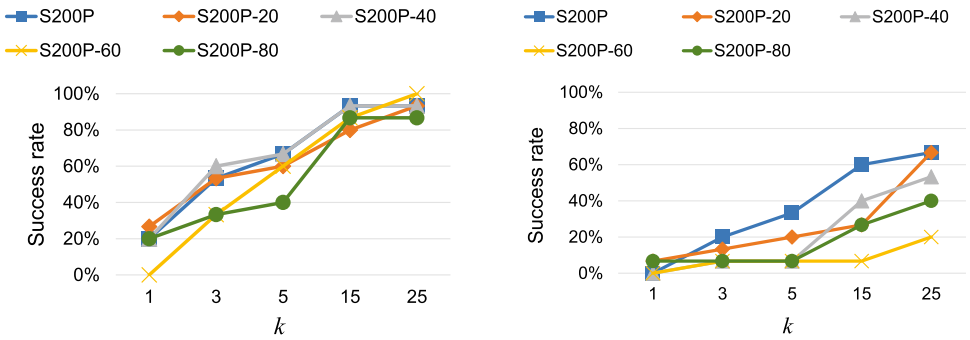


Fig. 10 Performance of STEM-COVID with incomplete input data

(a) Results of identifying ACCs **(b)** Results of identifying index cases

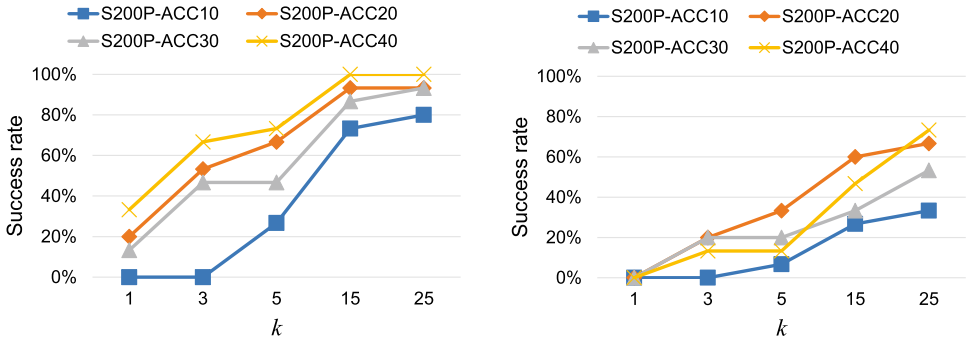


Fig. 11 Performance of STEM-COVID in scenarios with different ACC proportions

(a) Results of identifying ACCs **(b)** Results of identifying index cases

of searching ACCs under the two schemes. The results show remarkable advantages of the enhanced version in effectiveness of pooling the spatiotemporal traces of the known SCCs. It is thus able to obtain a much higher level of accuracy, especially when $k < 15$. A noticeable result is that the top-5 success rates of STEM-COVID with weighted evidence pooling in all scenarios are never below 60%, given such a small k value.

Besides, Fig. 8 shows the top-k identification success rates of the STEM-COVID model for identifying ACCs in scenarios S200P, S200SE and S1000P. The mean numbers of untested cases N_u averaged over all simulation instances are reported in the corresponding brackets. Overall, STEM-COVID achieves a fairly high level of success rates in all scenarios. For instance, a success rate of over 50% is obtained in S200P with $k = 3$ when there are only 14.8 ACCs out of all 129.2 untested cases on the average. The best results of identifying ACCs are seen in S200SE with the largest proportion of ACCs over all untested cases (24.8/75.0). Furthermore, the performance of STEM-COVID in S1000P is very close to that in S200P, indicating that our approach can scale up well to larger scenarios without significant loss in effectiveness.

Among the ACCs, index cases are extremely difficult to be screened because of their small proportion over the whole population. Nevertheless, as shown in Fig. 9, the success rates for identifying the sole index case in S200P increase rapidly with the larger k values used, where the top-3 and top-5 hit rates are no less than 20%, and even higher with $k = 15$ and 25. Comparatively, the top-k success rates in S200SE do not go beyond 40% in recognizing the index cases. A plausible explanation may be that fewer times of contacts are required in S200SE to cause the same number of infections, which weakens the effectiveness of our contact tracing-based approach. This hypothesis can be verified by the fact that the index cases in S200SE infected 11.9 persons on average, in contrast to only 7.0 in S200P within the same duration of simulation. In addition, the model performs well in S1000P considering that the top-15 and top-25 success rates to detect the index cases in S1000P are much higher than the top-3 and top-5 success rates in S200P.

6.3.3 Robustness of STEM-COVID

The top-k success rates for identifying ACCs and index cases in S200P-20, S200P-40, S200P-60 and S200P-80 are shown in Fig. 10. Compared against the performance in S200P, the effectiveness of STEM-COVID to identify normal ACCs is marginally affected by the incomplete data, especially that the success rates in S200P-20 and S200P-40 are very close to those in S200P. Comparatively, the success rates of detecting index cases are generally higher when less data is lost. To be specific, the top-5

Table 5 Comparison of prediction accuracy and time costs among STEM-COVID (S-C for short), the brute-force baseline (B-F for short) presented in Sect. 5.6, kNN and SOM

		S200N				**S200P**			
		S-C	B-F	kNN	SOM	S-C	B-F	kNN	SOM
success rates of identifying ACCs (%)	top-1	**13.3**	0.0	0.0	0.0	20.0	6.7	13.3	**26.7**
	top-3	**33.3**	**33.3**	13.3	26.7	**53.3**	13.3	33.3	26.7
	top-5	**60.0**	46.7	20.0	33.3	**66.7**	33.3	53.3	40.0
	top-15	**93.3**	73.3	66.7	73.3	**93.3**	73.3	66.7	80.0
	top-25	93.3	**100.0**	73.3	73.3	93.3	**100.0**	73.3	86.7
success rates of identifying index cases (%)	top-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	top-3	**6.7**	0.0	0.0	0.0	**20.0**	0.0	6.7	0.0
	top-5	**13.3**	0.0	0.0	0.0	**33.3**	0.0	6.7	0.0
	top-15	**20.0**	0.0	13.3	13.3	**60.0**	13.3	13.3	6.7
	top-25	**26.7**	13.3	13.3	**26.7**	**66.7**	20.0	20.0	6.7
time (s)	mean	2.5	13.7	5.5	**0.1**	2.5	13.4	5.6	**0.1**
	std	**0.0**	2.8	0.8	**0.0**	**0.0**	3.5	1.0	**0.0**

		S200SE				**S1000P**			
		S-C	B-F	kNN	SOM	S-C	B-F	kNN	SOM
success rates of identifying ACCs (%)	top-1	**40.0**	**40.0**	20.0	20.0	6.7	6.7	**20.0**	0.0
	top-3	**80.0**	73.3	53.3	33.3	**53.3**	26.7	20.0	20.0
	top-5	**86.7**	80.0	80.0	53.3	**73.3**	53.3	46.7	46.7
	top-15	**100.0**	**100.0**	93.3	93.3	86.7	**93.3**	53.3	80.0
	top-25	**100.0**	**100.0**	93.3	**100.0**	**100.0**	93.3	80.0	93.3
success rates of identifying index cases (%)	top-1	**20.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	top-3	**20.0**	0.0	0.0	0.0	**6.7**	0.0	0.0	0.0
	top-5	**20.0**	0.0	0.0	0.0	**6.7**	0.0	0.0	0.0
	top-15	**33.3**	0.0	13.3	6.7	**46.7**	6.7	6.7	13.3
	top-25	**33.3**	0.0	20.0	20.0	**66.7**	13.3	6.7	20.0
time (s)	mean	2.5	15.5	4.4	**0.1**	64.2	409.6	137.4	**2.7**
	std	**0.0**	0.7	0.9	**0.0**	1.4	38.1	6.9	**0.4**

Bold values indicate the better results obtained by the two schemes in each scenario

success rates for identifying index cases in S200P and S200P-20 are 33.3% and 20.0%, which are much higher than those in S200P-40, S200P-60 and S200P-80 (i.e., 6.7% in all the three scenarios). In a nutshell, despite incomplete contact tracing data, the STEM-COVID model preserves a reasonably high level of efficacy, indicating its strong robustness. This desirable property also provides the possibility of randomly discarding part of the contract tracing data to retain the computational efficiency of building the model and searching for ACCs in very large scenarios.

Apart from the robustness against incomplete input data, this paper also demonstrates the impact of the proportions of ACCs on the model performance in Fig. 11. Generally, it would be difficult to identify an ACC from the population when the proportion of ACCs among all infected cases is very small. This is the case in Fig. 11 where under S200P-ACC10, no ACC can be identified when $k = 1$ or 3. However, the performance remains relatively steady with the proportions from 20 to 40%.

6.3.4 Comparison with baselines

To demonstrate the strength of STEM-COVID in identifying hidden ACCs, this paper compares its performance with that of three competitors, listed as follows.

- The baseline presented in Algorithm 5 in Sect. 5.6: It can be considered as a brute-force approach to realizing the ACC identification algorithm with union-based evidence pooling.

- An adapted version of k Nearest Neighbors (*kNN*): Each individual is represented as a vector consisting of the sequence of location specifiers. The model is used to sort and identify an untested case among others based on the number of SCCs among its k nearest neighbors. For two cases with the same number, the one with closer distances to the tested positive cases will have a higher order. Based on prior experiments, euclidean distance is used and k is set to be 5 when $N = 200$ while $k = 25$ when $N = 1000$.
- Self-Organizing Map (*SOM*) [64]: Constructed using unsupervised learning, all input vectors (i.e., traces as used in kNN) are clustered into a small number of output nodes. The model is used to sort and identify an untested case among others based on the number of SCCs, which share the same best matching unit (*BMU*) with it. For two cases with the same number, the one closer to the SCCs is preferred. We adopt a one-dimensional map with 10 nodes and 5000 training steps when $N = 200$, or 50 nodes and 25000 steps when $N = 1000$. To measure the distance between two vectors, SOM also applies the euclidean distance.

As formulated in Sect. 3, the application in this paper is a weakly labeled problem from a perspective of classification. Under each scenario, there is no data with labels for offline training but only individual traces to be leveraged or identified online, so methods that require training before being used are not applicable. Moreover, the traces of different individuals are not independent due to their social contact, whereas conventional classification methods assume the data to be independently distributed. Besides, there is no explicit features in each trace to be leveraged for classification. These points are the reasons why this paper uses the algorithms mentioned above, with reasonable strategies to adapt them in the problem, instead of traditional classification methods, for performance comparison.

Table 5 reports the comparison among STEM-COVID, the brute-force baseline (B-F for short), kNN and SOM in terms of their identification accuracy and costs of CPU time (in seconds). The statistical results under each scenario include the top-k accuracy of identifying ACCs and index cases, the mean value and standard variation of the running times, which are obtained by executing the algorithms once on each simulation instance under the corresponding scenario. It is clear that no matter which scenario the algorithms are applied to, the proposed approach not only is more accurate than B-F, but also performs much better than kNN and SOM. The performance gaps between STEM-COVID and the peer algorithms are even larger for recognizing index cases. Moreover, the proposed method requires considerably less time to complete the computation than B-F and kNN, even though it contains the steps of both evidence pooling and ACC search. While under each scenario, the time costs of the two competitors vary much under different simulation instances, the superiority of our approach remain relatively consistent. However, since SOM builds a small number of output nodes out of the individual traces, it saves much time to sort the cases.

7 Conclusion and discussion

This paper has presented an episodic memory-based neural model called STEM-COVID, which is able to encode the collective spatiotemporal trajectories of a population and support an efficient search algorithm to identify asymptomatic COVID-19 cases. Compared against the previous model [25], the model with a weighted evidence pooling method is able to better capture the contact relationships between individuals in the evidence vector. The experimental results based on realistic simulation scenarios have demonstrated the effectiveness and efficiency of this episodic memory-based computational model when compared with some classical models. Moreover, we also demonstrated its strong robustness against incomplete input data. Notably, this paper tries to mimic the current situation of COVID-19 spread, such as the shift of dominant variants and the changes of ACC proportions as the bulk of the vaccinated breakthrough infections by aggressive variants are asymptomatic. These are all taken into account in our scenario design. In practical situations where mass screening is not possible, the STEM-COVID model can thus help to greatly improve the effectiveness of selective COVID-19 testing.

Serving as a starting point, this work can be further extended and applied to very large-scale real-world contact tracing data as a cost-effective way to contain the epidemic, in addition to traditional symptom-based strategies. This would require national agencies and resources to exploit this approach in addressing such asymptomatic cases as it involves privacy issues. Under acceptable protocols of privacy protection, the health officials can use the GPS, Bluetooth or check-in registration data from users' mobile phone to gather or infer their travel routes and then identify potential transmission sources of the coronavirus [12–17]. A good example is the SafeEntry and TraceTogether systems widely used in Singapore [65]. For real-world applications, nevertheless, there remain some limitations in our work that require further research. For example, the demographic and geographic models will need to be much expanded to reflect the real world. We also would like to incorporate more practical models of temporal dynamics of COVID-19 transmissions. Finally, to deal with the modeling of super large-scale scenarios, like epidemic prevention in a metropolis of millions of people,

multi-resolution and hierarchical modeling of spatiotemporal data could become necessary.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 61273300, 62102432, 62103420, 62103428 and 62103425, the Natural Science Fund of Hunan Province under Grant 2021JJ40697 and 2021JJ40702, the China Scholarship Council, the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), and the Singapore Ministry of Education Academic Research Fund (AcRF) Tier-1 under Grant 19-C220-SMU-023.

Declarations

Conflict of interest All authors declare that no conflict of interest exists.

References

1. Wang Y, Tong J, Qin Y, Xie T, Li J, Li J, Xiang J, Cui Y, Higgs ES, Xiang J et al (2020) Characterization of an asymptomatic cohort of SARS-CoV-2 infected individuals outside of Wuhan, China. Clinical Infectious Diseases, 1–7
2. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X et al (2020) Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Med 26(5):672–675
3. Furukawa NW, Brooks JT, Sobel J (2020) Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. Emerg Infect Dis 26(7):1–6
4. Long Q-X, Tang X-J, Shi Q-L, Li Q, Deng H-J, Yuan J, Hu J-L, Xu W, Zhang Y, Lv F-J et al (2020) Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nature Med 26(8):1200–1204
5. Zhan C, Chen J, Zhang H (2021) An investigation of testing capacity for evaluating and modeling the spread of coronavirus disease. Inf Sci 561:211–229
6. Lee S, Kim T, Lee E, Lee C, Kim H, Rhee H, Park SY, Son H-J, Yu S, Park JW et al (2020) Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with SARS-CoV-2 infection in a community treatment center in the Republic of Korea. JAMA Internal Medicine, 1–6
7. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, Ma H, Chen W, Lin Y, Zheng Y et al (2020) Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci China Life Sci 63(5):706–711
8. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of COVID-19. JAMA 323(14):1406–1407
9. Qiu H, Wu J, Hong L, Luo Y, Song Q, Chen D (2020) Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. Lancet Infect Dis 20(6):689–696
10. Mizumoto K, Kagaya K, Zarebski A, Chowell G (2020) Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Eurosurveillance 25(10):2000180
11. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, Chen H, Wang D, Liu N, Liu D et al (2020) Characteristics of COVID-19 infection in Beijing. J Infect 80(4):401–406
12. Cho H, Ippolito D, Yu, YW (2020) Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv:2003.11511
13. Kim MJ, Denyer S (2020) A travel log of the times in South Korea: Mapping the movements of coronavirus carriers. https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d_story.html
14. Koh D (2020) Australia's COVIDSafe contact tracing app sees 2 million downloads within 1 day of launch. https://www.mobihealthnews.com/news/asia-pacific/australia-s-covidsafe-contact-tracing-app-sees-2-million-downloads-within-1-day
15. MacKay J (2020) Tracking your location and targeted texts: how sharing your data could help in New Zealand's level 4 lockdown. https://theconversation.com/tracking-your-location-and-targeted-texts-how-sharing-your-data-could-help-in-new-zealands-level-4-lockdown-134894
16. Wang CJ, Ng CY, Brook RH (2020) Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. JAMA 323(14):1341–1342
17. Abbas R, Michael K (2020) The coronavirus contact tracing app won't log your location, but it will reveal who you hang out with. https://theconversation.com/the-coronavirus-contact-tracing-app-wont-log-your-location-but-it-will-reveal-who-you-hang-out-with-136387
18. Tan L, Yu K, Bashir AK, Cheng X, Ming F, Zhao L, Zhou X (2021) Toward real-time and efficient cardiovascular monitoring for COVID-19 patients by 5G-enabled wearable medical devices: a deep learning approach. Neural Comput Appl, 1–14
19. Almagor J, Picascia S (2020) Exploring the effectiveness of a COVID-19 contact tracing app using an agent-based model. Sci Rep 10(1):1–11
20. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, Lu FS, Huybers P, Resch B, Havas C et al (2021) An early warning approach to monitor Covid-19 activity with multiple digital traces in near real time. Sci Adv 7(10):6989
21. Ouyang L, Yuan Y, Cao Y, Wang F-Y (2021) A novel framework of collaborative early warning for Covid-19 based on blockchain and smart contracts. Inf Sci 570:124–143
22. Wang W, Subagdja B, Tan A-H, Starzyk JA (2012) Neural modeling of episodic memory: encoding, retrieval, and forgetting. IEEE Trans Neural Netw Learn Syst 23(10):1574–1586
23. Subagdja B, Tan A-H (2015) Neural modeling of sequential inferences and learning over episodic memory. Neurocomputing 161:229–242
24. Chang P-H, Tan A-H (2017) Encoding and recall of spatio-temporal episodic memory in real time. In: International Joint Conference on Artificial Intelligence, pp. 1490–1496
25. Hu Y, Subagdja, B, Tan, A-H, Quek C, Yin Q (2020) Who are the 'silent spreaders'?: Contact tracing in spatio-temporal memory models. arXiv:2010.00187
26. Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, Xing F, Liu J, Yip CC-Y, Poon RW-S et al (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 395(10223):514–523
27. Min CH (2020) Almost a quarter of infected household members of COVID-19 patients were asymptomatic, Singapore study finds. https://www.channelnewsasia.com/news/singapore/infected-household-members-covid-19-asymptomatic-12803192. 04 June
28. Day M (2020) Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ: British Medical Journal (Online) **368**
29. Carl Heneghan, Jon Brassey TJ (2020) COVID-19: What proportion are asymptomatic? https://www.cebm.net/covid-19/covid-19-what-proportion-are-asymptomatic/. 6 April

30. Oran DP, Topol EJ (2020) Prevalence of asymptomatic SARS-CoV-2 infection: A narrative review. Annals of Internal Medicine, 1–7
31. World Health Organization (2020) Q&A: Influenza and COVID-19 - similarities and differences. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza. 17 March
32. Day M (2020) Covid-19: four fifths of cases are asymptomatic, China figures indicate. British Medical Journal Publishing Group
33. Sun W, Ling F, Pan J, Cai J, Miao Z, Liu S, Cheng W, Chen E (2020) Epidemiological characteristics of 2019 novel coronavirus family clustering in Zhejiang Province. Chin J Prevent Med 54:027–027
34. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, Rossi L, Manganelli R, Loregian A, Navarin N et al (2020) Suppression of COVID-19 outbreak in the municipality of Vo, Italy. medRxiv
35. Loey M, Manogaran G, Khalifa NEM (2020) A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. Neural Comput Appl 1–13
36. Madhavan MV, Khamparia A, Gupta D, Pande S, Tiwari P, Hossain MS (2021) Res-covnet: an internet of medical health things driven Covid-19 framework using transfer learning. Neural Comput Appl 1–14
37. Gandhi M, Yokoe DS, Havlir DV (2020) Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. New Engl J Med 382:2158–2160
38. Rahimi I, Chen F, Gandomi AH (2021) A review on Covid-19 forecasting models. Neural Comput Appl 1–11
39. Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc lond Ser A Contain Papers Math Phys Character 115(772):700–721
40. Lymperopoulos IN (2021) # stayhome to contain covid-19: neuro-sir-neurodynamical epidemic modeling of infection patterns in social networks. Expert Syst Appl 165:113970
41. Zhang Z (2007) The outbreak pattern of SARS cases in China as revealed by a mathematical model. Ecol modell 204(3–4):420–426
42. D'Arienzo M, Coniglio A (2020) Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. Biosaf Health 2(2):57–59
43. Venkatramanan S, Lewis B, Chen J, Higdon D, Vullikanti A, Marathe M (2018) Using data-driven agent-based models for forecasting emerging infectious diseases. Epidemics 22:43–49
44. Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, Limosin F, Leleu H (2020) A stochastic agent-based model of the SARS-CoV-2 epidemic in France. Nature Med 26(9):1417–1421
45. Aleta A, Martin-Corral D, y Piontti AP, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini IM Jr, Merler S et al (2020) Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Hum Behaviour 4(9):964–971
46. Bradshaw WJ, Alley EC, Huggins JH, Lloyd AL, Esvelt KM (2021) Bidirectional contact tracing could dramatically improve COVID-19 control. Nature Commun 12(1):1–9
47. Kojaku S, Hébert-Dufresne, L, Mones E, Lehmann S, Ahn Y-Y (2021) The effectiveness of backward contact tracing in networks. Nature Physics, 1–7
48. Ahamad MM, Aktar S, Rashed-Al-Mahfuz M, Uddin S, Liò P, Xu H, Summers MA, Quinn JM, Moni MA (2020) A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. Expert Syst Appl 160:113661
49. Tong Z-D, Tang A, Li K-F, Li P, Wang H-L, Yi J-P, Zhang Y-L, Yan J-B (2020) Potential presymptomatic transmission of SARS-CoV-2, Zhejiang province, China, 2020. Emerg Infect Dis 26(5):1052
50. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New Engl J Med 382(13):1199–1207
51. Backer JA, Klinkenberg D, Wallinga J (2020) Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. Eurosurveillance 25(5):2000062
52. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Internal Med 172(9):577–582
53. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S-M, Yuan B, Kinoshita R, Nishiura H (2020) Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. J Clin Med 9(2):538
54. Carpenter GA, Grossberg S (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vis Graph Image Process 37(1):54–115
55. Meng L, Tan A-H, Miao C (2019) Salience-aware adaptive resonance theory for large-scale sparse data clustering. Neural Netw 120:143–157
56. Tan A-H, Carpenter GA, Grossberg S (2007) Intelligence through interaction: Towards a unified theory for learning. In: International Symposium on Neural Networks, pp. 1094–1103. Springer
57. Teng T-H, Tan A-H, Zurada JM (2014) Self-organizing neural networks integrating domain knowledge and reinforcement learning. IEEE Trans Neural Netw Learn Syst 26(5):889–902
58. Tan A-H, Subagdja B, Wang D, Meng L (2019) Self-organizing neural networks for universal learning and multimodal memory encoding. Neural Netw 120:58–73
59. Gao S, Tan A-H, Setchi R (2019) Learning adl daily routines with spatiotemporal neural networks. IEEE Trans Knowl Data Eng 33(1):143–153
60. Carpenter GA, Grossberg S, Rosen DB (1991) Fuzzy art: fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Netw 4(6):759–771
61. Woodward A (2020) Coronavirus super-spreader events all have notable similarities – and they reveal the types of gatherings we should avoid for years. https://www.businessinsider.com/coronavirus-super-spreader-events-reveal-gatherings-to-avoid-2020-5
62. Wu JT, Leung K, Leung GM (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet 395(10225):689–697
63. Volz E, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, Cuomo-Dannenburg G, Donnelly CA Dorigatti I, FitzJohn R, et al (2020) Report 5: phylogenetic analysis of SARS-CoV-2. Imperial College London COVID-19 Response Team
64. Kohonen T (2013) Essentials of the self-organizing map. Neural Netw 37:52–65
65. Woo J (2020) Policy capacity and Singapore's response to the COVID-19 pandemic. Policy Soc 39(3):345–362