Prediction of asymptomatic COVID-19 infections based on complex network

Yili Chen¹ | Haoming He²,³ | Dakang Liu⁴ | Xie Zhang⁵ | Jingpei Wang⁶ | Yixiao Yang⁷

¹School of Automation and Key Laboratory of Intelligent Information Processing and System Integration of IoT (GDUT), Ministry of Education, Guangdong University of Technology, Guangzhou, China
²111 Center for Intelligent Batch Manufacturing Based on IoT Technology (GDUT), Guangdong University of Technology, Guangzhou, China
³Guangdong Key Laboratory of IoT Information Technology (GDUT), Guangdong University of Technology, Guangzhou, China
⁴Guangdong-Hong Kong-Macao Joint Laboratory for Smart Discrete Manufacturing (GDUT), Guangdong University of Technology, Guangzhou, China
⁵School of Electric Power, South China University of Technology, Guangzhou, China
⁶College of Control Science and Engineering, Zhejiang University, Hangzhou, China
⁷School of Software, Tsinghua University, Beijing, China

Correspondence
Xie Zhang, School of Electric Power, South China University of Technology, Guangzhou, China.
Email: zhangxie@scut.edu.cn

Funding information
National Key R&D Program of China, Grant/Award Number: 2020AAA0108304; National Natural Science Foundation of China, Grant/Award Numbers: 62073088, 62003275, U1911401, U1701261

Abstract
Novel coronavirus pneumonia (COVID-19) epidemic outbreak at the end of 2019 and threaten global public health, social stability, and economic development, which is characterized by highly contagious and asymptomatic infections. At present, governments around the world are taking decisive action to limit the human and economic impact of COVID-19, but very few interventions have been made to target the transmission of asymptomatic infected individuals. Thus, it is a quite crucial and complex problem to make accurate forecasts of epidemic trends, which many types of research dedicated to deal with it. In this article, we set up a novel COVID-19 transmission model by introducing traditional SEIR (susceptible-exposed-infected-removed) disease transmission models into complex network and propose an effective prediction algorithm based on the traditional machine learning algorithm TrustRank, which can predict asymptomatic infected individuals in a population contact network. Our simulation results show that our method largely outperforms the graph neural network algorithm for new coronary pneumonia prediction and our method is also robust and gives good results even if the network information is incomplete.

KEYWORDS
complex network, COVID-19, machine learning, Trustrank algorithm
1 | INTRODUCTION

Since December 2019, there has been an outbreak of novel coronavirus pneumonia (COVID-19), which is highly infectious and rapidly spreading. The non-effectively suppressed in the early stages of transmission is due to the extreme irregularity of COVID-19. In the process of epidemic control, epidemic prevention, and control are severely tested due to the large randomness of human population movement and the diversity of population contact methods. For scientific simulation and prediction, researchers have proposed various mathematical models to assist in decision making and planning. Current research on the epidemic spread and development trends lies mainly in the work of predicting infectious nodes, and the more common models focus on dynamical models as well as statistical models. In the field of biomathematics, for single population infectious diseases, there are traditional dynamical models such as SIR, SEIR, and models derived or improved on this basis. However, these traditional dynamical models are not able to capture and address the problem of asymptomatic infected individuals. At present, the reason why COVID-19 is so difficult to control is that there are a large number of asymptomatic infections. The main approach taken by most government agencies today is to strictly control the social distance of the population to minimize exposure.

Current research on epidemic transmission and development trends lies mainly in the work on predicting infectious nodes, and the more common models focus on kinetic models as well as statistical models. In the field of biomathematics, there are several basic kinetic models such as SIR and SEIR. In addition, there are also single population infectious disease models derived from these models. For example, Zareie effectively predicted the spread of the COVID-19 epidemic in Iran by constructing the SIR model for the Iranian COVID-19 epidemic; a time-dependent susceptibility-infection-recovery (SIR) model was proposed to track the transmission and recovery rates at specific times; Wang et al. added different types of time-varying quarantine strategies to the traditional SIR model; Anca and Kieran targeted specific dynamic intervals and epidemiological parameters of COVID-19 based on traditional SEIR models; Johns Hopkins University combined SEIR and regression models for predicting changes in the spread of COVID-19.

The statistical model was based on a machine learning model to extract statistical data and then used to predict the spread of the epidemic. For example, Ahmed et al. used a logistic model to predict the scale of the epidemic in Turkey and Iraq, and the prediction results verified the validity of the model; Li et al. used Gaussian theory to study the spread of COVID-19 based on official epidemic data; De Moraes et al. used support vector machine (SVM) and data from emergency care admission exams to detect COVID-19 cases.

In addition, multilayer perceptron models, artificial intelligence models, and social phenomenology models have also been applied to the analysis of foreign epidemic spread prediction. Although effective epidemic spread prediction results can be obtained by the aforementioned prediction models, both kinetic and statistical models only predict epidemics by setting prediction parameters. It will lead to the problem of difficult and complicated solutions when there are too many parameters. Besides, the influence of the suspected population in the epidemic spread is not considered. While relevant artificial intelligence models (e.g., multilayer perceptron) tend to fall into the problem of local minima, and there is barely any research on domestic intelligent prediction models related to the spread of epidemics.

In this article, we propose a method for effective prediction using a non-pharmacological intervention approach in a detection tool for asymptomatic infected individuals in a contact network. By customizing and modifying the search strategy, we obtain a model suitable for detecting the asymptomatic infected person, and present a series of improvements to the baseline model.

First, based on empirical observations of the spread of new coronavirus, we combine the disease transmission process with the topology of complex networks and propose a set of simulated transmission frameworks in which we can observe the state of each node in the network at each time slice. The framework can serve the purpose of capturing and tracking the probability of transmission between individuals with different infection states in the network, and further using and transforming it to extend it into an efficient method that can effectively identify asymptomatic individuals. As stated earlier, asymptomatic infected individuals play a critical role in the spread of new coronavirus. Our transmission model is based on an improvement of the traditional SEIR transmission model. Thus we assume that a person can have four different states in the transmission of COVID-19: susceptible (S), asymptomatic infection (A), symptomatic infection (I), and recovery (R). More clearly, in the simulation framework, each node could be in one of the four states mentioned above.
Then, after simulating the spread of the disease, we propose a prediction method for asymptomatic infected nodes in the network based on the traditional machine learning TrustRank. Throughout this work, we extracted information about the state of each node in the network at a certain moment in time. Then after learning by an improved algorithm, we use the information obtained from machine learning as a key metric to rank the nodes in the network to determine how much propagation a node has. In our targeted simulation results, we found that the method used outperformed most graph neural network algorithms in predicting asymptomatic infected individuals.

Finally, after obtaining the prediction results, we also propose isolation measures in combination with our approach based on the results and also demonstrate that our isolation measures are more effective than most of the current measures. Non-pharmacological interventions are extremely important in the control of new coronavirus outbreaks, and our work provides a more credible solution to assist in the tracking of infected individuals. In particular, with the increased awareness of the population and the technological advances of managers, the improved access to information on the internet can now better allow algorithms to help.

2 RELATED WORK

2.1 SEIR epidemic model

The free spread of an epidemic can be described by the SEIR model of “susceptible state—disease free state—infected state—recovered state”, where $S$ represents the susceptible population, $E$ represents the population that does not show symptoms after being infected, $I$ represents the population that shows symptoms of infection, and $R$ represents the population that has been cured and immunized or effectively quarantined or died of the disease, and so forth, and no longer has an impact on the epidemic transmission dynamics. Assuming that an infected individual (state $I$) comes into contact with a susceptible individual (state $S$), the probability that the susceptible individual is infected but does not show symptoms (transmission rate) is $\beta$; an asymptomatic individual in state $E$ will change to state $I$ with probability $\gamma_1$ per unit time; an individual in state $I$ will change to state $R$ with probability $\gamma_2$ per unit time (unit time in days). Clearly, the SEIR propagation process can be described by the following differential equations:

\[
\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N},
\]

\[
\frac{dE(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \gamma_1 E(t),
\]

\[
\frac{dI(t)}{dt} = \gamma_1 E(t) - \gamma_2 I(t),
\]

\[
\frac{dR(t)}{dt} = \gamma_2 I(t),
\]

where $S(t)$, $E(t)$, $I(t)$, and $R(t)$ denote the number of individuals in the susceptible, asymptomatic, infected, and recovered states, respectively. And $N$ denotes the total number of individuals in the system and $N = S(t) + E(t) + I(t) + R(t)$.

2.2 Detection method based on machine learning

2.2.1 MLP

Multilayer perceptrons, also known as multilayer feedforward neural networks, have excellent nonlinear matching and generalization capabilities. MLPs are trained using a back propagation algorithm that reduces the global error between the output data and the actual desired data. MLPs have very good nonlinear mapping capabilities, high parallelism, and the ability for global optimization.

Although the MLP architecture has many advantages, its efficiency in high-dimensional space is relatively low, and it can even lead to overfitting of the model during the training process. The presence of hidden layers increases the number of hyperparameters, which makes MLPs need to handle a high amount of computation during the training process. The
data input that can be received by a single neuron in the traditional MLP real-valued model is a single real number, which usually does not achieve satisfactory results when it is subjected to the multidimensional signal input.

2.2.2 | GCN

Existing graph convolutional neural networks\(^2\textsuperscript{7-29}\) are divided into two categories: spectral methods, which define the graph convolution from the spectral domain using the convolution on graph theorem, and spatial methods, which aggregate each central node and its neighboring nodes by defining an aggregation function from the node domain. The application scenarios targeted by graph data modeling are very broad, which makes the tasks handled by graph data modeling more diverse. In modeling graph convolutional neural networks, researchers focus on how to build convolution operators on graphs.

The initial spectral methods had the drawback of high spatio-temporal complexity, and ChebNet and GCN parametrized the convolution kernel in the spectral methods to greatly reduce the spatio-temporal complexity. These two methods, although classified as spectral methods, have started to define the weight matrix of the nodes from a spatial perspective. Inspired by these two methods, spatial methods were born and people started to consider modeling the weights between nodes in the node domain with attention mechanisms, serialization models, and so forth.

The computational equation between the GCN layers is:

\[
H^{(l+1)} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)
\]

\[
\tilde{A} = A + I
\]

where \(I\) is the unit matrix, \(D\) is the degree matrix of \(\tilde{A}\), \(H\) is the feature of each layer, and \(\sigma\) is the nonlinear activation function.

2.2.3 | Node2vec

The Node2vec\(^3\textsuperscript{0}\) is a network representation method based on the idea of word2vec to solve the complex problems caused by the high-dimensional and sparse network structure, and its core is to introduce potential variables to model the global relational structure pattern and transform the network information into low-dimensional and dense real vectors.

Unlike the uniform random wandering of DeepWalk,\(^3\textsuperscript{1}\) Node2vec introduces the jump probability parameter \(p\) and parameter \(q\) to control the breadth-first sampling and depth-first sampling strategies in the random sequence generation process. Depth-first or breadth-first alone are only the extremes in the network, Node2vec proposes to learn and obtain the optimal parameter \(p\) and \(q\) values by semi-supervised network form, assign weights to the edges in the graph, and then guide the wandering process by the magnitude of the weights to achieve an optimal balance between breadth-first and depth-first, balancing the local and global information of the network.

2.3 | PageRank

The PageRank\(^3\textsuperscript{2,33}\) algorithm was first used by Google to rank the importance of web pages in its search engine by borrowing the idea of traditional citation analysis, which is widely used for ranking in the field of international graphical intelligence. The PageRank algorithm assumes that the importance and quality of a web page can be measured by the number of hypertext links to it from other web pages.

At present, many important link analysis algorithms are derived from the PageRank algorithm. PageRank is a method used by Google to identify the rank/importance of a webpage, and it is the only standard used by Google to measure the quality of a website. Google uses PageRank to adjust the results so that those web pages with more “rank/importance” will be ranked in the search results and the ranking of the site will be improved, thereby improving
the search results. The higher the PR value, the more popular (more important) the web page is. Its calculation formula is shown in the following equation.

\[
PR(p) = \frac{(1 - d)}{N} + d \sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)}
\]

where \(PR(p)\) denotes the page level of page \(p\); \(T_i\) denotes other pages pointing to page \(p\); \(C(T_i)\) denotes the number of links pointed outward by page \(T_i\); \(\frac{PR(T_i)}{C(T_i)}\) is the PR value given to page \(p\) by page \(T_i\); \(d\) is the probability of randomizing to a page, between 0 and 1, usually set to 0.85.

### 2.4 TrustRank

The TrustRank algorithm is a semi-automatic classification method for describing the trust level of web pages and thus ranking them. TrustRank assumes that good sites rarely link to bad sites, while anti-TrustRank assumes that pages that link to spam pages are likely to be spam pages as well. Since there are links from good pages to spam pages on the web, that is, good-to-bad links, using either TrustRank or anti-TrustRank alone will result in some pages getting unreasonable rank values, so propagating a combination of both is an effective method.

The TrustRank algorithm organizes the web pages in the network into a Web graph \(G_w = < P, R >\) based on interlinking relationships, where \(P\) is the set of web pages in the network, which is the set of vertices in the Web graph, and \(R\) is the set of links between web pages and web pages. In the Web graph \(G_w\), to represent the link relationship between web pages, define \(R_{pq} \in R\) to denote the link from web page \(p\) to web page \(q\). Denote by \(N_{Op}\) the number of links pointed out by web page \(p\), that is, the out-degree of web page \(p\) in the Web graph, and denote by \(O_{pi}\) the links pointed out by web page \(p\), where \(i \in \left(0, N_{Op}\right)\). Denote by \(N_{inp}\) the number of links pointing to web page \(p\), that is, the in-degree of web page \(p\) in the Web graph, and denote by \(IN_{pi}\) the links pointing to web page \(p\), where \(i \in \left(0, N_{inp}\right)\).

\[
TR(p) = (1 - d) + d \sum_{i=1}^{N_{inp}} \frac{TR(q_i)}{N_{O_{qi}}}
\]

where \(TR(p)\) is the TrustRank value of the page, \(d\) is the damping factor and \(q_i\) is the \(i\)th page containing a link to page \(p\).

### 3 METHOD

#### 3.1 Data

##### 3.1.1 Infectious stay away network (ISA)

A web-based dataset was used for the simulation, which was downloaded from http://konect.cc/networks/sociopatterns-infectious/. The dataset is a contact network exhibited at the eInfectious Stay Awaye exhibition in Dublin on July 15, 2009. The network consists of 410 nodes and 2847 links with an average degree of 13. Each link represents a real-life face-to-face contact between two people, and the links can be analogized for the spread of the virus in COVID-19. In our approach, we assume that when an infected node is connected to a susceptible node, the susceptible node will have a certain probability to turn into an infected node.

##### 3.1.2 The parameters of COVID-19

- **Basic reproduction number** \(R_0\): The basic number of infections, defined as the average number of individuals an infected person would infect if no external forces intervened and everyone was not immune. The \(R_0\) we use is 3.5.
• **The coefficient of proportion of asymptomatic infections $p$:** We set $p = 0.2$ in the main text, a parameter obtained from another empirical study. We also set up controlled experiments for different proportions of $p$.

• **Distribution of infection duration $f_I(d)$:** For symptomatic infected patients, we fitted $f_I(d)$ using a normal distribution with $\mu_I = 8.8$ and $\sigma_I = 3.88$.

• **Distribution of asymptomatic infection $f_A(d)$:** In general, the exact time of tract infection and recovery is not known for asymptomatic infections. In combination with the actual COVID-19 situation, we used a normal distribution with $\mu_A = 20$ and $\sigma_A = 5$ to fit $f_A(d)$. We also set up controlled experiments to discuss other values of $\mu_A$ and $\sigma_A$.

### 3.2 Simulation of COVID-19 spreading

In this section, we will describe in detail how to simulate the transmission process of new coronavirus. To validate the framework, we simulate a realistic transmission process of COVID-19 in a population contact network-ISA network, where we obtain details of each infected node in the network, such as when it became infected and when it recovered, within a finite number of days. It can derive the exact infection history of nodes without infection symptoms from the detailed infection information of infected nodes.

The new coronavirus transmission model in our study is very similar to the traditional susceptible-exposure-infection-recovery model (also known as the SEIR model), but with several major differences. First, the asymptomatic potential state of the real situation is introduced. That is, when a susceptible node faces the situation of being infected, it will become asymptomatic with a probability of $p$, or become infected with a probability of $1 - p$. Second, unlike the differential equations of transmission mechanisms in traditional disease transmission models, the complex network-based transmission is more realistic; the infectious disease model as a research tool we use is to develop solutions to related problems, which is of more important practical value.

In this model, the implementation of the dissemination process on the ISA network that lasted several days is presented. When $T = 0$, any number of nodes in the network are selected; then as time $T$ increases, according to the realistic parameters of COVID-19, there are some laws here that can be used to simulate the propagation, such as a mechanism to determine the state of each node in the network with probability $p$ presenting no symptoms and probability $1 - p$ presenting an infected state; meanwhile, a node that has been infected node will last for several days, and this number of days will be determined by a specific probability distribution. The simulation of each day corresponds to a dynamical propagation implementation based on the network structure, and we can obtain the infection status of individual nodes in each snapshot.

In our simulation framework, we introduce the following parameters:

- $s_n(t)$ represents the state of the $n$th node in the network located on day $t$. In the ISA network, the range of values of $n$ is $(1, 410)$. In our simulation scheme, the range of values of $t$ is $(1, 60)$. That is, we simulated the state change of 410 nodes in the network for 60 days from the discovery of the first infected person to the next.

- $r_{n_{state}}$ represents the time that the $n$th node in the network stays in the current state. For example, when $state = S$, $r_{n_S}$ represents the number of days that the $n$th node stays in state $S$; the same is true when $state$ takes $A$, $I$, and $R$. This parameter helps us to determine whether the node should be transformed to the next state in the current time slice.

- $\alpha_n$ represents the set of neighboring nodes of the $n$th node in the network. In the update rule presented next, we will traverse all neighboring nodes of the current node, which means that the current infected node will have a certain probability of spreading the new coronavirus.

- $f_A(t)$ represents the time distribution of asymptomatic infected individuals in the infection state in the real situation, and we take $\mu_A = 20$ and $\sigma_A = 5$. In the update rule presented next, we will discretize the continuous time distribution $f_A(t)$ so as to identify a number of days of continuous stay for the node; specifically, we can sample uniformly over the interval $[1, 60]$ to obtain:

$$\{X_1, \ldots, X_T\}$$

Obtain the set of discrete values based on the sample:

$$\{f(X_1), \ldots, f(X_T)\}$$
It is sufficient to take one from the set of discrete values as the duration of the current asymptomatic infected node by a certain strategy.

- $f_I(t)$ represents the distribution of time in the infected state of the dominant infected person in a representative realistic situation, which we fit using a normal distribution with $\mu_I = 8.8$ and $\sigma_I = 3.88$. Similarly, we take the sampling method mentioned above to obtain the duration of the dominantly infected node’s continuous stay.

- Our model involves an infection rate $\beta$, which can be derived from the basic regeneration number $R_0$ as follows:

$$\beta = \frac{R_0}{\langle k \rangle \cdot \lambda}$$

In this equation, $\langle k \rangle$ is the average number of exposures per person in the network and is defined as the average time a susceptible person carries the virus.

As shown in Algorithm 1, we show the steps to simulate the COVID-19. Similar to the traditional SEIR model, in the population contact network, we introduce states for each node, taking one of the values $S$ (susceptible state), $A$ (asymptomatic state), $I$ (infected state), and $R$ (cured state). In each time step, it need to update the status of each node dynamically. The rules for updating are as follows:

1. If the current node $i$ is susceptible to infection, $s_i(t) = S$, we determine whether the node should be transformed to state $A$. We assume that it will be infected by the neighboring infected persons around it, the infection in the crowd contact network is passive, in the sense that we only consider the infection of the node when we encounter a susceptible node. So we traverse the set $a_i$ of all neighboring nodes of the current node and count the number of infected nodes in $a_i$ (including both infected nodes and asymptomatic infected nodes), which is denoted as $k$. The probability that the current node stays in state $S$ is $(1 - \beta)^k$. Then the probability that the current node will be transformed to symptom-free state $A$ next is:

$$p_{StoA} = 1 - (1 - \beta)^k \quad (1)$$

In the concrete implementation, we generate a random probability $p_{\text{Random}}$ to determine whether or not to infect. We can consider the transfer of states as a change of state in a finite state machine:

$$s_i(t + 1) = \begin{cases} 
A & p_{\text{Random}} > p_{StoA} \\
S & \text{other}
\end{cases} \quad (2)$$

Meanwhile, when the node needs to be transformed into state $A$, we randomly sample the number of days $t_A$ that the node stays in state $A$ from the asymptomatic infection distribution $f_A(t)$ and update $r_A^i$:

$$r_A^i = t_A \quad (3)$$

And while the node remains in state $S$, we only update $r_S^i$:

$$r_S^i = r_S^i + 1 \quad (4)$$

2. If the current node $i$ is asymptomatic infected, $s_i(t) = A$, we determine whether the node should change to state $I$. Since the number of days the node stays in state $A$ has already been determined by the time distribution earlier, we only need to determine whether the current number of days $t$ exceeds $r_A^i + r_A^i$.

$$s_i(t + 1) = \begin{cases} 
I & t \geq r_A^i + r_A^i \\
A & \text{other}
\end{cases} \quad (5)$$

Also, when the node needs to be transformed to state $I$, we randomly sample the number of days $t_I$ that the node stays in state $I$ from the asymptomatic infection distribution $f_I(t)$ and update $r_I^i$:

$$r_I^i = t_I \quad (6)$$
Algorithm 1. Propagation simulation method of COVID-19

Require: $G(V, E)$: The whole network with node set $V$ and edge set $E$
$\zeta$: The initial infected set of network
$s_n(t)$: The state of the $n$th node in the network at day $t$
$r_n^{state}$: The dwell time of the $n$th node in the network in the current state state
$a_n$: The set of neighbor nodes of the $n$th node in the network
$f_A(t)$ (or $f_I(t)$): The time distribution of state $A$ (or $I$)
$\beta$: The infection rate

Ensure: $s_i(T)$: The state of any node $i \in V$ at the moment $T$

1: initial $s_i(0)$ to state $S$ for all $i$

2: for $j \in \zeta$ do
3:    let $p_{random}$ be a random number from 0 to 1
4:    if $p_{random} < 0.2$ then
5:        $s_j(0)$ is Infected
6:        let $r_I^j$ be a random number from $f_I$ as the Infected duration
7:    end if
8: end for

9: for $t \in [1, T]$ do
10:    for each $i \in V$ do
11:        if $s_i(t - 1)$ is Susceptive then
12:            for each $n \in a_i$ do
13:                let $k$ be the number of infected nodes, $k = 0$
14:                if $s_n(t - 1)$ is Infected then
15:                    $k = k + 1$
16:            end if
17:        end for
18:        let $p_{random}$ be a random number from 0 to 1
19:        if $p_{random} > 1 - (1 - \beta)^k$ then
20:            let $s_i(t)$ be Exposed, let $r_A^i$ be a random number from $f_A$ as the asymptomatic duration and let $r_A^i = t_A$
21:        end if
22:        if $p_{random} \leq 1 - (1 - \beta)^k$ then
23:            let $r_S^i = r_S^i + 1$
24:        end if
25:    end if
26:    if $s_i(t - 1)$ is Asymptomatic then
27:        if $t > r_S^i + r_A^i$ then
28:            let $s_i(t)$ be Infected, let $r_I^i$ be a random number from $f_I$ as the infected duration and let $r_I^i = t_I$
29:        end if
30:        if $t \leq r_S^i + r_A^i$ then
31:            let $r_A^i = r_A^i + 1$
32:        end if
33:    end if
34:    if $s_i(t)$ is Infected then
35:        if $t > r_S^i + r_A^i + r_I^i$ then
36:            let $s_i(t)$ is Recovery
37:        end if
38:        if $t \leq r_S^i + r_A^i + r_I^i$ then
39:            let $r_I^i = r_I^i + 1$
40:        end if
41:    end if
42: end for
43: end for
3. If the current node \(i\) is infected, \(s_i(t) = I\), we determine whether the node should be transformed to state \(R\). We only need to determine whether the current number of days \(t\) exceeds \(r_i^I + r_i^A + r_i^S\):

\[
s_{i}(t + 1) = \begin{cases} 
    R & \text{if } t \geq r_i^I + r_i^A + r_i^S \\
    I & \text{otherwise}
\end{cases}
\]  

(7)

3.3 Containment strategy based on TrustRank

After simulating the spread of the disease, we will investigate how our approach can be used to more effectively control COVID-19. In the current challenging environment, many countries and regions have adopted approaches such as isolating sick people and tracking contacts in an attempt to gain control of the COVID-19 outbreak. In several experiments it can be found that the strategy of tracking contacts reduces the size of simulated outbreaks more than the strategy of tracking only contacts, but this strategy also has a disadvantage: it leads to the isolation of most of the population at the same point in time. The detection and isolation of released non-infectious individuals led to an increase in the size of the outbreak, suggesting that contact tracing and isolation may be most effective as a “local lockdown” strategy when exposure rates are high.

To study this effect, a simulated transmission experiment of COVID-19 can be performed again on the ISA network, starting from day \(T = 8\), where asymptomatic infected individuals will be tracked using our method; as time \(T\) increases, we tracked 4% of the number of network nodes each day, and infected individuals (including asymptomatic infected individuals) were immediately quarantined and the node switched accordingly to the status \(R\) (considered as a cured state or quarantine state).

Here we take three different measures to discuss how to contain the spread of COVID-19:

1. **Static control:** this is the simplest and most intuitive isolation control measure. At the time \(T = t\), we traverse the network to find nodes that have shown symptoms of COVID-19 (i.e., infected nodes) and place them in quarantine at the subsequent time \(T = t + 1\). This step is repeated every day, and nodes in other states remain in their original state to participate in the subsequent propagation of COVID-19.

2. **Local control:** On the basis of static control, we can selectively quarantine some nodes. One of the strategies is to randomly select the nearest neighbors of the current infected person for detection. Of course, it is impractical to detect neighbors for all infected nodes, we can randomly select 2% (or 4%) of the number of nodes in the whole network for local control. The number of nodes in the network is denoted as \(N\).

3. **Dynamic control (our method):** this is also the measure we have chosen in combination with our algorithm. Based on the static control, we use our method to screen potentially dangerous nodes (i.e., asymptomatic infected persons in a potential state). Specifically, on each day of the update traversal, we can select the top 2% (or 4%) number of nodes with the highest probability of infection for control.

At the moment \(T = t\), we take the nodes with the highest screening risk according to the TrustRank algorithm. The matrix iteration form of the TrustRank algorithm is as follows:

\[
r = c \cdot U \cdot r + (1 - c) \cdot d
\]

(8)

where \(c\) stands for decay factor, \(U\) stands for inverse transition matrix, \(d\) stands for seed vector, and \(r\) stands for the final derived score vector. Here, seed means a node that deserves to be noticed, which can be interpreted as a training set.

In our approach, all nodes that have state \(I\) at the current moment \(t\) can be considered as seed nodes. Before iterative computation, the number of neighboring nodes of the seed node is first calculated, specifically, if node \(i\) is a node with state \(I\), we traverse the set of neighbors \(\alpha_i\) and count the total number of nodes in \(\alpha_i\) (denoted as \(\text{neighbor}(i)\)), followed by the sum of nodes with state \(S\) and state \(A\) in \(\alpha_i\) (denoted as \(\text{outdegree}(i)\)). \(d\) is given by the following equation:

\[
d(i) = \begin{cases} 
    \frac{1}{\text{outdegree}(i)}, & \text{if node } i \text{ is seed} \\
    0, & \text{other}
\end{cases}
\]

(9)
And the vector \( r \) has the same initial value as \( d \), but it can be seen from Equation (8) that \( r \) is updated with iteration, while \( d \) is fixed. The formula for \( U \) is as follows:

\[
U(p, q) = \begin{cases} 
\frac{1}{\text{neighbor}(q)}, & \text{if } pq \text{ is connected} \\
0, & \text{other}
\end{cases}
\]

(10)

After iterative computation of the TrustRank algorithm, the \( M \) nodes with the highest risk of being filtered out can be obtained. If there are infected but asymptomatic nodes among these filtered nodes (i.e., with status \( A \)), the infected nodes can be added to the training set (i.e., they are considered as neutron nodes in the next day’s update) and their status is set to \( R \) (i.e., they are considered as isolated by our isolation measure).

### 3.4 Robustness test

As mentioned in the background elaboration above, the collection of contact network information cannot be complete because this requires more transparent cooperation from the population in addition to the degree of technical perfection, so in the actual epidemic statistics, most of the contact network information collected by the survey is incomplete. However, although the network is incomplete, it still has valuable information value, so in order to fully use the exposure network information and maximize the screening of individuals with the highest probability of infection from the network, we conducted a robustness test on the designed algorithm. In this test, we imposed some measures to perform information incompleteness simulation to test the effectiveness of the algorithm. We randomly removed the edges with the highest 80% in the ISA network and tested the accuracy in the remaining network respectively. As shown in the figure, our algorithm is able to reliably predict and filter infected nodes in various difficult scenarios where we artificially remove contact edges.

In this section, we specifically focus on the virus propagation parameters that tend to change during actual propagation. We tested the sensitivity of these parameters in the ISA network used:

1. the integrity of network structure.
2. the basic reproduction number \( R_0 \).
3. the distribution of asymptomatic infection \( f_A(d) \).

### 4 RESULT

#### 4.1 Result of simulation validation

##### 4.1.1 Screening performance assessment

This part of the experiment compares our method with several graph neural network algorithms. Many people use graph neural network algorithm to do node prediction of complex network because it can well retain the structure of graph network and save the connection information between points in the network, but its flexibility is poor, and it is difficult to be different. It is difficult to reflect the change process of the dynamic network in the graph structure. In contrast, our method can make better use of the connection information of nodes in the network. The experimental results are shown in Figure 1.

Specifically, we first defined five symptomatically infected nodes on the original network, and then we started from the symptomatically infected nodes of the current network and performed 1000 times on our network propagation framework (60 per simulation days) simulation experiments, and then save the sorting results of the remaining sets of symptomatic infections removed in each simulation experiment. For the node ranked first, we predict that it has the greatest probability of being an asymptomatic infection. Therefore, we count the number of times that the first node in the 1000 experiment results is an infected person (that is, the predicted success), which is recorded as \( n_E \), and the number of times that is not an infected person is recorded as \( n_{notE} \). The accuracy is recorded as \( \text{acc} \), and the recall rate is the normalized result of the accuracy \( \text{acc} \), which is recorded as \( \text{recall} \):
1612  

CHEN et al.

FIGURE 1  Method performance evaluation. (A) The accuracy of predicting the level of infection (proportion of non-susceptible individuals in the ranking list). (B) The accuracy of our proposed algorithm is compared with the accuracy of other algorithms under the relative accuracy of the machine learning algorithm. (C) The ratio of successfully identified non-susceptible individuals to all network individuals. (D) The relative recall rate of the machine learning algorithm. Compared with the dynamic algorithm, the recall rate of the static algorithm on the entire network.

\[
\text{acc} = \frac{n_E}{n_E + n_{\text{not}E}}
\]

\[
\text{recall}_i = \frac{\text{acc}_i - \text{acc}_{\text{min}}}{\text{acc}_{\text{max}} - \text{acc}_{\text{min}}}
\]

4.1.2  Effectiveness of control measure

According to the $T = 8$ of the new crown pneumonia transmission simulation on the ISA network, we have adopted different methods to contain the pandemic, including the recommended method (i.e., dynamic containment), infection containment and neighbor containment. For neighbor containment and the proposed method, we select a total of $2\%N$ individuals for screening and then isolate those individuals who test positive at each time step.

The experimental results, which are shown in Figure 2, show that our method can effectively find the infected nodes, which can effectively inhibit the infection.
4.2 | Result of robustness test

4.2.1 | Network structure

In this part, we randomly remove a small part of the links in the ISA network and then use the targeted screening scheme on the remaining networks.

As shown in Figure 3, we randomly retain the edges of different proportions (20%, 40%, 60%, 80%, 100%) in the graph structure to test the performance of our method on the missing edge network. From the graph, we can see that even if there are only 20% edges left in the graph structure, our method can still maintain good performance and maintain a relatively high accuracy rate.
4.2.2 Basic reproduction number

In this part, we test different basic reproduction number $R_0$ on our method. As shown in Figure 4, we can still maintain good performance for different basic regeneration numbers $R_0$. $R_0$ is a key indicator of epidemic infectious diseases. The robustness of our method to $R_0$ proves that our method can be effectively transplanted to different epidemic infectious diseases.

4.2.3 Distribution of asymptomatic infection

For infected persons in the asymptomatic state, we test the performance of our static screening method for different asymptomatic infection time distributions $f_A(d)$. So we change the average time $\mu_A$ and the variance $\sigma_A$. Figure 5 and Figure 6 show us that our method is not sensitive to the value of $\mu_A$ (or $\sigma_A$), and can still maintain good performance for different $\mu_A$ (or $\sigma_A$).
FIGURE 6 Performance of our method for different $\sigma_A$. (A) The relationship between the accuracy rate and different basic reproduction number $\sigma_A$. (B) The relationship between the recall rate and different basic reproduction number $\sigma_A$.

5 | CONCLUSION

In this article, based on the spreading law of COVID-19 and the spreading mechanism of the TrustRank algorithm, we studied the tracking behavior of asymptomatic infections in the population contact network. Our method can infer the potential dangers in the crowd contact network. Compared with isolating direct contacts, our isolation method is more timely and effective. This research is not only important for controlling the epidemic and solving the current spread of the virus, but also for relevant personnel to provide better and faster auxiliary decision-making methods for prevention and control plans when major health incidents may occur in the future. Besides, we tested the robustness of the simulation on the network, and the experiment proved that even if the network structure is incomplete or the information collected is wrong, the method is still robust. We believe that this theoretical framework has important practical significance for finding potential infections in COVID-19.

Further works still need to be done in order to analyze the stability of the model, chaos or delay effect, and the consistency of this model to the real data.

ACKNOWLEDGMENT

This work was supported in part by the Grants of National Key R&D Program of China (2020AAA0108304), and National Natural Science Foundation of China (62073088, 62003275, U1911401, U1701261).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Xie Zhang https://orcid.org/0000-0003-1398-5075

REFERENCES

1. Sohrabi C, Alsafi Z, O’Neill N, et al. World health organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg. 2020;76:71-76.
2. Anirudh A. Mathematical modeling and the transmission dynamics in predicting the Covid-19 - what next in combating the pandemic. Infect Dis Model. 2020;5:366-374. doi:https://doi.org/10.1016/j.idm.2020.06.002
3. Nokes D, Anderson R. The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. Epidemiol Infect. 1988;101(1):1-20.
4. Calafiore GC, Novara C, Possieri C. A modified SIR model for the COVID-19 contagion in Italy. Proceedings of the 2020 59th IEEE Conference on Decision and Control (CDC); 2020:3889-3894.
5. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett. 2001;86(14):3200.
6. Boguñá M, Pastor-Satorras R, Vespignani A. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys Rev Lett*. 2003;90(2):028701.
7. Small M, Tse CK. Small world and scale free model of transmission of SARS. *Int J Bifurcat Chaos*. 2005;15(05):1745-1755.
8. Zhan C, Tse CK, Fu Y, Lai Z, Zhang H. Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data. *PLoS One*. 2020;15(10):e0241171.
9. Huang S, Sun J, Feng L, Xie J, Wang D, Hu Y. Identify hidden spreaders of pandemic over contact tracing networks; 2021. arXiv preprint arXiv:2103.09390.
10. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Soliton Fract*. 2020;139:110057.
11. Kruijshaar ME, Barendregt JJ, Hoeymans N. The use of models in the estimation of disease epidemiology. *Bull World Health Organ*. 2002;80:622-628.
12. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369.
13. Prem K, Liu Y, Russell TW, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*. 2020;5(5):e261-e270.
14. Zareie B, Roshani A, Mansournia MA, Rasouli MA, Moradi G. A model for COVID-19 prediction in Iran based on China parameters; 2020. MedRxiv.
15. Liu TH. A time-dependent SIR model for COVID-19 with undetectable infected persons; 2020. arXiv q-bio.
16. Wang L, Zhou Y, He J, et al. An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China. *J Data Sci*. 2020;18(3):409-432.
17. Radulescu A, Williams C, Cavanagh K. Management strategies in a SEIR model of COVID 19 community spread; 2020. arXiv preprint arXiv:2003.11150.
18. Gupta R, Pandey G, Chaudhary P, Palc S. SEIR and regression model based COVID-19 outbreak predictions in India; 2020.
19. Ahmed A, Salam B, Mohammad M, Akgul A, Khoshnaw S. Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model. *Aims Bioeng*. 2020;7(3):130-146.
20. Li L, Yang Z, Dang Z, et al. Propagation analysis and prediction of the COVID-19. *Infect Dis Model*. 2020;5:282-292.
21. de Moraes Batista AF, Miraglia JL, Donato TH, Chiavegatto Filho AD COVID-19 diagnosis prediction in emergency care patients: a machine learning approach; 2020. medRxiv.
22. Wang L, Lin ZQ, Wong A. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci Rep*. 2020;10(1):1-12.
23. Li L, Qin L, Xu Z, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*. 2020.
24. Sethy PK, Behera SK, Rath PK, Biswas P. Detection of coronavirus disease (COVID-19) based on deep features and support vector machine; 2020.
25. Zheng N, Du S, Wang J, et al. Predicting COVID-19 in China using hybrid AI model. *IEEE Trans Cybern*. 2020;50(7):2891-2904.
26. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE Rev Biomed Eng*. 2020.
27. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks; 2016. arXiv preprint arXiv:1609.02907.
28. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs; 2017. arXiv preprint arXiv:1706.02216.
29. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks; 2017. arXiv preprint arXiv:1710.10903.
30. Grover A, Leskovec J. node2vec: scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016:855-864.
31. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: online learning of social representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2014:701-710.
32. Klicpera J, Bojchevski A, Günneßmann S. Predict then propagate: graph neural networks meet personalized pagerank; 2018. arXiv preprint arXiv:1810.05997.
33. Gleich DF. PageRank beyond the web. *SIAM Rev*. 2015;57(3):321-363.
34. Gyongyi Z, Garcia-Molina H, Pedersen J. Combating web spam with trustrank. Proceedings of the 30th International Conference on Very Large Data Bases; 2004.

**How to cite this article:** Chen Y, He H, Liu D, Zhang X, Wang J, Yang Y. Prediction of asymptomatic COVID-19 infections based on complex network. *Optim Control Appl Meth*. 2023;44(3):1602-1616. doi: 10.1002/oca.2806