Dynamic Immunization Node Model for Complex Networks Based on Community Structure and Threshold

Ronghua Shang, *Member, IEEE*, Weitong Zhang, *Member, IEEE*, Licheng Jiao, *Fellow, IEEE*, Xiangrong Zhang, *Senior Member, IEEE*, and Rustam Stolkin, *Member, IEEE*

***Abstract*—In the information age of big data, and increasingly large and complex networks, there is a growing challenge of understanding how best to restrain the spread of harmful information, for example, a computer virus. Establishing models of propagation and node immunity are important parts of this problem. In this article, a dynamic node immune model, based on the community structure and threshold (NICT), is proposed. First, a network model is established, which regards nodes carrying harmful information as new nodes in the network. The method of establishing the edge between the new node and the original node can be changed according to the needs of different networks. The propagation probability between nodes is determined by using community structure information and a similarity function between nodes. Second, an improved immune gain, based on the propagation probability of the community structure and node similarity, is proposed. The improved immune gain value is calculated for neighbors of the infected node at each time step, and the node is immunized according to the hand-coded parameter: immune threshold. This can effectively prevent invalid or insufficient immunization at each time step. Finally, an evaluation index, considering both the number of immune nodes and the number of infected nodes at each time step, is proposed. The immune effect of nodes can be evaluated more effectively. The results of network immunization experiments, on eight real networks, suggest that the proposed method can deliver better network immunization than several other well-known methods from the literature.**

***Index Terms*—Dynamic propagation model, immune threshold, node immunization, propagation probability.**

Manuscript received December 21, 2018; revised April 17, 2019; accepted April 15, 2020. Date of publication May 21, 2020; date of current version March 14, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61773304, Grant 61671350, Grant 61371201, Grant 1772399, and Grant 61876141, and in part by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT1170. This article was recommended by Associate Editor P. P. Angelov. *(Corresponding author: Weitong Zhang.)*

Ronghua Shang, Weitong Zhang, Licheng Jiao, and Xiangrong Zhang are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: rhshang@mail.xidian.edu.cn; zwt@stu.xidian.edu.cn; lchjiao@mail.xidian.edu.cn; xrzhang@mail.xidian.edu.cn).

Rustam Stolkin is with the Extreme Robotics Lab, University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: r.stolkin@cs.bham.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.2989427.

Digital Object Identifier 10.1109/TCYB.2020.2989427

I. Introduction

AN EARLY use of the term "network" appears in electrical systems: a circuit or part of it consisting of several elements is called a network. The network enables electrical signals to be transmitted according to a certain requirement. In computing, network sometimes refers to a set of connected computers, or more recently to the system composed of several individuals connected via the World Wide Web, such as a social network. Other forms of the network include transportation networks, political networks, the contagion of populations by biological pathogens, and many other examples. By abstracting these network systems into complex network structures composed of nodes and edges, it is helpful to study the characteristics, functions, and security of the network systems [1]. Information propagation networks [2] and computer networks are a societally important infrastructure. Therefore, it is particularly important to control harmful information [3] in information propagation networks, such as inappropriate social media content (preventing nasty images/videos from spreading via Facebook), or virus propagation in computer networks. In order to ensure the continuous spread of harmless information and the normal operation of network systems in information propagation networks and computer networks, while simultaneously preventing the spread of harmful information, there is increasing interest within the research community in developing efficient and accurate methods of network immunization or node immunization [4]. The purpose of node immunization is to achieve the optimal effect of harmful information or virus control under the condition of minimal node immunization.

In order to study the characteristics of information or virus propagation and network immune methods, a variety of propagation network models has emerged. The threshold model [5], [6], proposed to describe collective behavior, has been widely used in solving the propagation threshold problem in complex systems [7]. More recently, threshold models have been extended to research multilayer networks [8] and time networks [9]. The independent cascade (IC) model was originally proposed in research on marketing models [10]. By introducing a time-delay parameter into the IC model, a new continuous-time independent cascade (CTIC) model [11] was proposed. Adding the dynamic time variation to the CTIC model, the time-based asynchronous IC model

(T-BASIC) [12] was proposed. The most widely used epidemic model [13], [14] was first proposed by Kermack and McKendrick. There is also a susceptible-infectious model (SIm), [15] suitable for modeling the initial outbreak of the virus. The susceptible-infectious-recovered (SIR) model [16] was proposed by adding a removed state to the SIm model. Considering that each infected individual may be transformed back into a susceptible individual after recovery, with a fixed probability, the susceptible-infectious-susceptible (SIS) model [17] was proposed. A stable Cox–Ingersoll–Ross (SCIR) model [18] was proposed by introducing a new nodal contacted state. Other network models include the adaptive networks model [19], [20] and the activity-driven networks model [21]. These are two popular models that are used to describe the network structure and the propagation dynamics process. Because these models predominantly ignore the original connection between nodes in the real network, they typically cannot explain the essential law of information or virus propagation in the real network. To overcome this problem, this article proposes a new network information and virus propagation model, which describes the essential relationship between the network topology information and the individuals in the real network, based on the network community structure information and the similarity between nodes. Our proposed model can be used to study the nodal immunization problem which inhibits the propagation of harmful information.

In order to better inhibit the spread of harmful information, many network or node immunization methods have been proposed. The purpose of node immunization is to immune as few nodes as possible to achieve the optimal immune effect. Classic immunization strategies include random immunization [22], acquaintance immunization [23], and target immunization [24].

Random immunization is also known as uniform immunity. The immunized nodes in the network are selected randomly without any information of the network or nodes. This method is the simplest, but requires a priori knowledge of the minimum number of nodes requiring immunization to ensure that harmful information does not continue to spread. This method does not exploit models or rules of information or virus propagation in the network.

Acquaintance immunization belongs to the class of methods known as decentralized immunization. Some nodes in the network are randomly selected, and then some neighbor nodes of these nodes are randomly selected for immunization. This method requires much less network information. Nodes with a large node degree are selected preferentially over nodes with a small node degree. Due to incorporation of the additional knowledge about node degree, the acquaintance immunization method is typically more effective and efficient than the random immunization.

Target immunization methods improve efficiency by exploiting the knowledge of the attributes of the nodes, node degrees, betweenness, and other information, e.g., selecting the core nodes in the network, which have a higher node degree, and have a higher influence on other nodes. This is more effective in suppressing the spread of harmful information. This method can reduce the number of nodes that need to be immunized, by better targeting the most critical nodes.

Most of the existing nodal immunization methods are based on the advance selection of k nodes for immunization. There are three main approaches given as follows.

1) Static immunization (SI) [25], [26] selects k nodes for immunization according to certain rules at the initial moment of propagation, namely, before the spread of harmful information.
2) Uniform immunization (UI) [27] attempts to immunize during propagation. According to certain rules, the number of k/T nodes in each time step is selected for immunization, where k is the total number of immune nodes, and T is the estimated time of the propagation process.
3) Exponent immunization (EI) [28] also performs immunization during the propagation process. According to certain rules, $2^{-t} * k$ nodes are selected for immunization at each time step, where t is the number of time steps. Determining an appropriate number of immune nodes is another problem to be solved in suppressing the spread of harmful information.

Because information propagation is often dynamic, dynamic immune models are now attracting increasing attention from the research community. This article proposes a dynamic immune model, incorporating an immune method based on the immune gain threshold. The main contributions of this article are as follows.

1) Considering the situation of the real network, the node carrying harmful information is abstracted as a new node in the network. In the real network, harmful information is more likely to spread within closely connected small groups. Therefore, in this article, based on the community structure information in the network and the similarity between nodes, the propagation probability of the harmful information between nodes is calculated.
2) In order to judge more efficiently and accurately whether any particular node should be immunized, an improved immune gain based on the propagation probability of the community structure and node similarity is proposed. The improved immune gain value is calculated for the neighbors of the infected node, at each time step, and the node is immunized according to the given threshold. This can effectively prevent invalid or insufficient immunization at a certain time step.
3) In order to better evaluate the results of immunization algorithms, a novel evaluation index is proposed in this article. By considering both the number of immunized nodes, and the nodes finally infected, the immune effect of nodes and the efficiency of the immunization approach, can be evaluated more accurately.

II. Information or Virus Propagation

A. Propagation Model

The information or virus propagation model was originally abstracted from the problem of maximizing the influence of

information propagation on social networks [29]. The connection between nodes is abstracted by establishing the network structure. The propagation probability between nodes is used as the weight of the edge to describe the process or characteristics of the propagation of information or virus. A set of initially infected nodes is generated and harmful information or viruses are disseminated at each step. The nodes that successfully transmit harmful information or viruses are called activated nodes or infected nodes [30]. Contrary to the problem of maximization of influence, this article studies the methods to effectively suppress the spread of harmful information or viruses. The propagation model in the maximization of influence can also be used to suppress nodal immunity in the propagation of harmful information or virus. There are three main types of common information or virus propagation models: 1) linear threshold (LT) model [31]; 2) IC model [32], [33]; and 3) SIR model [34].

In the LT model, the probability of node j being activated is calculated by the state of its neighbor nodes and the propagation probability between them. Given a fixed or randomly generated threshold [35] between $(0, 1)$, determine whether node j is activated or not. The condition that node j is activated is determined by the following formula:

$$\sum_{i \in N_{in}(j)} p(i,j) \geq \theta \tag{1}$$

where j is the node to be activated, node i is the neighbor node of node j, $N_{in}(j)$ is the neighbor node of node j in the set of infected nodes, and $p(i,j)$ is the probability of disseminating information from node i to node j. θ is the threshold to determine whether the node is activated. It is important to note that

$$\sum_{i \in N(j)} p(i,j) \leq 1 \tag{2}$$

where $N(j)$ is the set of neighbor nodes of node j.

In the IC model, when node i is activated, it obtains an opportunity to activate its neighbor node. When there is more than one infected node in the set of neighbor nodes of node j, the node j will be activated in the random order and cannot be repeated. The weighted independent cascade (WIC) model [36] is an extension of the IC model. In the WIC model, the weight of the edge between nodes represents the propagation probability between nodes, which is non-negative and independent of the network structure. Compared with the SIm model [37], the SIR model adds a recovery state which can reflect some special viruses in the real network. For example, if a man has smallpox and is cured, he will not be infected again, nor will he pass it on to others. In the SIR model, nodes are classified into three states: 1) susceptible (S); 2) infected (I); and 3) recovered (R). S state indicates that the nodes are not affected by the virus. I state indicates that the nodes are affected by the virus and it can transmit the virus to other nodes. R state means that the nodes are neither infected nor infect other nodes.

In the traditional LT model, IC model, SIR model, and its derivative model, the propagation probability between nodes are usually only related to the states of nodes and their neighbors. In real-world networks, the probability of spreading information or virus between nodes is closely related to the relationship and degree between nodes. Presently, there is no suitable propagation model that starts from the connection between nodes in the network. So they cannot explain the essential rule of information or virus transmission in the real network. In this article, a new network information and virus propagation model is proposed based on the network topology information and the essential relationship between individuals in the real network. It can be used to study the problems that how to suppress the spread of harmful information or viruses. In the part of the experiment, the model in this article will be analyzed and verified in detail.

B. Concepts in Complex Networks

In this section, the network community structure and the similarity between network nodes mentioned in this algorithm are briefly introduced.

1) Community Structure in Complex Networks: In order to study the principles and functions of complex systems better, complex systems are often abstracted into complex networks. The complex network is composed of nodes abstracted by individuals and edges abstracted by links between individuals. Complex networks have the properties of scale-free [38], small world, aggregation, and power-law distribution of degree. There are differences in the degree of closeness between nodes in complex networks. The edges in the network have diversity and heterogeneity. It may be unidirectional or bidirectional. The weights of the edges in the network may be different. Studying these characteristics of complex networks is helpful to understand and analyze the structural characteristics of various real network systems [39]. There are many substructures in the real network structure, among which the community structure is widely studied and analyzed because of its contribution to the study of network functions and behavior patterns [40], [41]. Community structure is a kind of subnetwork structure with tight internal connection and sparse external connection [42]. It can be seen from this intuitive definition that the community structure in the network is usually a set of nodes with some common characteristics or similar function to the entire system. In the scientific research cooperative network [43], the research group that the researchers work together can be divided according to the unit where the researcher belongs or according to the region. Community structure in the network is formed by the partitioning results. In the jazz musicians' partnership model [44], the network can be divided into black musicians' group and white music group by race. In Belgian mobile communication networks [45], users speak the same language form the community structure in the network. In the protein network [46], the community structures are the sets of proteins with consistent functions.

2) Node Similarity: The node similarity in the network can represent the influence ability between the two nodes. It can be calculated according to the information of connected nodes in the network. There are many kinds of functions for node similarity measurement. Several common similarity functions

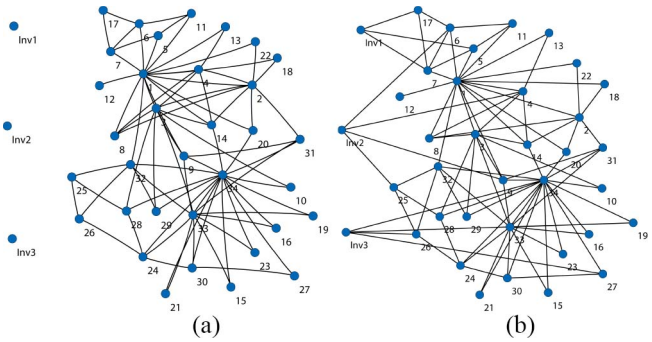


Fig. 1. Example of a simple node intrusion model. (a) Before invasion. (b) After invasion.

include RA index [47], Jaccard index [48], hub promoted index [49], cosine similarity function [50], and so on.

III. Dynamic Node Immunization Under the New Dynamic Propagation Model

A. Intrusion Node With Harmful Information and Virus

It is relatively simple to select nodes randomly in the network structure as infected nodes carrying harmful information or viruses. But in real social networks or virus propagation networks, foreign persons or sources that carry harmful information or viruses usually invade the network. There may be one or more new nodes. The new node may be connected to one or more nodes in the network. The spread of the harmful information or virus carried by the new node may end at the first moment, or it may still appear in the following moment. Harmful information or virus disseminated between nodes may be unidirectional or bidirectional. These depend on the actual applications. In this article, a dynamic network model is established, in which nodes carrying harmful information or viruses are abstracted into new nodes in the network. The number of new nodes and the way of building the edges between the new nodes and the original nodes in the network can be changed according to the needs of different networks. Fig. 1 shows an example of a simple node intrusion model. The right part of Fig. 1(a) is an existing network structure. The left three nodes Inv1, Inv2, and Inv3 of Fig. 1(a) are intrusion nodes with harmful information or viruses that suddenly appear at the initial moment. Suppose that at the second moment, all three intrusion nodes complete the propagation of harmful information or virus as shown in Fig. 1(b). The intrusion node Inv1 transmits the harmful information or virus to the nodes 5, 7, and 17, the intrusion node Inv2 transmits the harmful information or virus to the nodes 4, 6, 9, and 25, and the intrusion node Inv3 transmits the harmful information or virus to the nodes 19, 27, and 34 in the original network structure. Then at the next moments, the harmful information or virus that has already intruded in the network will continue to spread between nodes. The object node that the intrusion node spreads the virus may be randomly selected, also may have some kinds of rules. For example, in the social network, the harmful information is easier to disseminate through the known person.

In the network propagation model shown in Fig. 1, the new node terminates the infection at the end of the initial moment. According to different reality, the number of nodes and the propagation mode of the network propagation model can be changed. For example, at the following moments, the edges between the foreign nodes and the original nodes in the network are still generated according to the propagation condition of the foreign nodes. This article assumes that foreign nodes only disseminate harmful information or virus at the initial moment, and the harmful information or virus propagation between nodes is bidirectional.

B. New Propagation Model Based on Propagation Probability According to Community Structure and Node Similarity

In this article, a new network propagation model is established, in which the foreign or the infected source carrying harmful information or virus is abstracted as the new node in the network structure. The connection between the new nodes and the original network structure is established by the edge between the new nodes and the original nodes. When there are intrusion nodes in the network, namely, harmful information or virus, the initial affected or infected nodes will continue to spread harmful information or virus at the next moment. It can be seen from Section II-A that the commonly used information propagation network models often assign fixed propagation probability to nodes in the network or simply assign the propagation probability according to the number of node neighbors. This is divorced from the relationship between individuals and network topology information in real networks. For example, there are 50 students in a class. Suppose there are several small groups of well-connected students. Then these small groups can be seen as societies in the network structure of this class. When a student learns about some news or finds something, he is most likely to tell his classmates in his group of friends, at least tell them first. It can be seen that because of the closer relationship between the members of the community structure, when intrusion nodes carry harmful information or virus in these network structures, the infection and propagation within the communities will be more frequent and more serious.

In addition to the connections among the members of the communities, each pair of connected nodes in the whole network has different degrees of connection. When a node i in the network carries harmful information or virus, it will disseminate the harmful information or virus to its neighbor nodes by probability. It is obvious that node i is more likely to disseminate harmful information or virus to its neighbor nodes in the same community than the neighbor nodes out of the same community. From the community point of view, assuming that node i is equally likely to disseminate harmful information or virus to its neighbor nodes in the same community, the probability that node i disseminates harmful information or virus to node j in the same community is the reciprocal of the number of neighbor nodes of node i. From the point of view of the similarity between nodes, the probability of disseminating harmful information or virus between node i and node j is related to their common neighbors. In conclusion, when

Algorithm 1 Procedure of the Community Integration Strategy

Input: n: Total number of network nodes, Node connection information;
Output: Network partition result f;
1: Update the core node set ← calculate node degree;
2: Pre-partitioning result ← calculate node similarity between each core node and its neighbors; $\Delta D = 0$;
3: **while** $\Delta D \geq 0$ **do**
4: Arrange the communities in a descending order according to their external connection number;
5: **for** $i = 1$: Number of current communities **do**
6: Calculate ΔD_u ← for the neighbor community u of the community i;
7: Calculate ΔD_v ← find the community v corresponding to the $\max(\Delta D_u)$;
8: **if** $\max(\Delta D_u) \geq \max(\Delta D_v)$ **then**
9: label(v) =label(u) ← merge u and v;
10: **end if**
11: **end for**
12: Calculate ΔD.
13: **end while**

node j and node i belong to the same community, the probability of node i disseminating harmful information or virus to node j is related to the similarity between the two nodes. It is also related to the number of neighbors in the community which node i belongs to. When node j and node i are not in the same community, the probability of node i transmitting harmful information or virus to node j is only related to the similarity between node i and node j.

In order to deal with the problem of probability assignment of the large-scale propagation network model, the community integration strategy based on an improved modularity density increment for large-scale networks is used in this article to detect the network structure [51]. An improved modularity density increment was proposed as the objective function for community integration in this method. The global judgment is added to the local integration process, which can improve the resolution of the modularity density function effectively and reduce the probability of error integration. The experimental results showed that the method can obtain more detailed and accurate community partition results on large-scale networks. The modularity density increment function ΔD is represented as follows:

$$\Delta D = \left[\frac{l(u) - l_o(u) - l_o(v) + 3l_{uv}}{d_s(u) + d_s(v)}\right] - \left[\frac{l(u) - l_o(u)}{d_s(u)} + \frac{l(v) - l_o(v)}{d_s(v)}\right] \tag{3}$$

where u indicates any community in the network, v indicates the neighbor community of community u, $l(u)$ indicates the number of connections within the community u, $l_o(u)$ indicates the number of connections between the community u and the outside, l_{uv} represents the number of connections between u and v, $d_s(u)$ represents the sum of node degree of nodes in community u, and $d_s(v)$ represents the sum of node degree of nodes in community v.

The flow of the specific community detection method is shown in Algorithm 1.

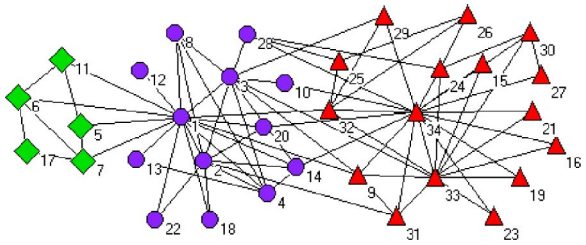


Fig. 2. Community detection results on the karate network.

After getting the results of community detection, based on the dynamic propagation model, the propagation probability of harmful information or virus between each pair of connected nodes is calculated according to the community structure and the similarity between nodes. In order to avoid the low influence of community attributes on propagation probability, the similarity function value in propagation probability is divided by 2. The probability of node i transmitting harmful information or virus to its neighbor node j is calculated as follows:

$$p(i,j) = \begin{cases} \frac{\mathrm{Sim}(i,j)}{2} + \frac{1}{|N_i|}, & \text{if } j \in N_i \\ \frac{\mathrm{Sim}(i,j)}{2}, & \text{otherswise} \end{cases} \tag{4}$$

where Sim denotes the degree of similarity between nodes i and j, i and j denote any two nodes in the network, and N_i represents a set of neighbor nodes within the same community as node i. As described in Section II-C, there are many methods to calculate the similarity between nodes. Cosine similarity is used as follows:

$$\mathrm{Sim}(i,j) = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)||N(j)|}} \tag{5}$$

where n denotes the total number of nodes in the network, $N(i)$ denotes the set of all neighbor nodes of node i, and $N(j)$ represents the set of all neighbor nodes of node j.

It can be seen that when the connected node i and node j belong to the same community, the probabilities of transmitting harmful information and virus of them are equal. Taking the karate network [52] as an example, the division of community and the calculation of propagation probability are given. Shang *et al.* [52] observed and studied a karate club from 1970 to 1972 and established the karate network model. During the observation period, club activities include regular karate classes for members in club and social affairs (party, dance, etc). The karate network consists of 34 nodes and 78 edges.

The result of community detection on the karate network is shown in Fig. 2. It can be seen from Fig. 2 that the community integration strategy based on the improved modularity density increment divides the karate network into three groups. The square, circle, and triangle represent three community structures, respectively. As can be seen, the association structure within the members of the relationship is relatively close.

Based on the above results, the propagation probability of harmful information or virus is calculated for each pair of connected nodes in the karate network. Table I shows the probability of node 1 transmitting harmful information or virus to its neighbor nodes in the karate networks.

TABLE I
PROBABILITY OF NODE 1 PROPAGATING TO ITS NEIGHBOR NODES IN KARATE NETWORKS

Nodes	2	3	4	5	6	7	8	9
P	0.4333	0.3372	0.4062	0.2165	0.1875	0.1875	0.35	0.1118
Nodes	11	12	13	14	18	20	22	32
P	0.2165	0.225	0.2768	0.3236	0.2768	0.2443	0.2768	0.051

TABLE II
PROPAGATING PROBABILITY OF NODE 2 AND NODE 3 TO THEIR NEIGHBOR NODES IN KARATE NETWORKS

Nodes	1	3	4	8	14
P	0.4583	0.3885	0.4652	0.4583	0.4231
Nodes	18	20	22	31	/
P	0.3607	0.3175	0.3607	0.0833	/
Nodes	1	2	4	8	9
P	0.3800	0.4064	0.4656	0.4591	0.2121
Nodes	10	14	28	29	33
P	0.2547	0.4257	0.2219	0.0913	0.0913

From Table I, we can see that nodes 6 and 7 have the same node degree, that is, equal neighbor nodes. Although they are not in the same community of node 1, they have an equal number of common neighbors with node 1. Therefore, it is not difficult to understand that the propagation probability of node 1 to nodes 6 and 7 is equal. Similarly, the propagation probabilities of node 1 to nodes 13, 18, and 22 are equal.

Table II shows the probability of nodes 2 and 3 spreading harmful information or virus to their neighbors.

It can be seen from Table II that the spread of information or virus is directed. That is, in this propagation model, the propagation probability of node 1 to node 2 is not equal to that of node 2 to node 1.

In summary, the propagation probability is calculated for each pair of connected nodes in the network, and the new network propagation model is established. In the process of node immunization, it is ensured that the immune nodes will not be infected, and the propagation probability between other nodes will not be affected, so it is not necessary to calculate the propagation probability between nodes repeatedly in the process of dynamic propagation, and the algorithm complexity is greatly reduced. In this article, a propagation threshold pt is given. The probability of the neighbor node of the infected node being activated is calculated according to the propagation probability. When the activation probability is greater than the propagation threshold, the neighbor node is transmitted harmful information or virus. The activation probability is calculated as follows:

$$AP(i) = 1 - \prod_{w \in N_{in}(i)} (1 - p(w, i)) \tag{6}$$

where i and w are any two nodes in the network, $AP(i)$ denotes the activation probability of node i, $N_{in}(i)$ is the neighbor node set of node i in the infected node set, and $p(w, i)$ represents the probability of propagating harmful information or virus from node w to node i.

In this article, it is assumed that the new nodes only propagate harmful information or virus at the initial time, and randomly select the nodes in the network to propagate. Therefore, the community structure in the network is first detected and the propagation probability is allocated. The procedure of generating the propagation network model is shown in Algorithm 2.

Algorithm 2 Procedure of Generating the Propagation Network Model

Input: f: Network community partition results, n: total number of network nodes, pt: propagation threshold, a, b: the number of new nodes and edges, T: maximum time step, Node connection information;
Output: Infected nodes set I;

```
for each pair of connected nodes do
    Calculate Sim ← the similarity of connected nodes;
end for
Initial infected nodes ← Randomly select a * b nodes;
for t = 1:T do
    for i = 1:n do
        Calculate the activation probability AP ← for each neighbor node in the initial infected node set;
        if AP(i) > pt then
            Add node i to the set of infected nodes set I;
        end if
        i = i + 1;
    end for
    t = t + 1.
end for
```

C. Modified Immunization Gain

This section will introduce the modified immune gain (MIG) based on the proposed network propagation framework. The immune gain represents the changes in the probability of neighbors being infected after the node is immunized.

First, influence ability Ia of node i on the neighbor node j outside the infected node set is calculated. The influence ability represents the difference in the probability of neighbor j being infected before and after node i is immunized. In this article, the probability of node j being infected is not only calculated according to its neighbor nodes in the infected node set but also to all its neighbor nodes. The influence of node i on node j is calculated according to the activation probability of node i and the probability that all neighbor nodes of node j propagate to it. The formula is as follows:

$$Ia(i, j) = \left(\prod_{v \in N(j), v \neq i} (1 - AP(v) * p(v, j)) \right) - \left(\prod_{v \in N(j)} (1 - AP(v) * p(v, j)) \right) \tag{7}$$

where $Ia(i, j)$ is the influence ability of node i on node j, $AP(v)$ is the activation probability of node v, $p(v, j)$ is the probability of node v spreading harmful information or virus to node j, and $N(j)$ is the neighbor node set of node j. The immune gain of node i is calculated as follows:

$$\mathrm{MIG}(i) = \sum_{j \in N_{\mathrm{out}}(i)} Ia(i, j) \tag{8}$$

where $\mathrm{MIG}(i)$ is the immune gain obtained after immunizing the node i, $Ia(i, j)$ is the influence ability of node i on node j,

and $N_{\text{out}}(i)$ is all neighbor nodes of node i outside the set of infected nodes.

D. Node Immunization

Node immunization in the network propagation model is to reduce the spread of harmful information or virus. In order to reduce the number of infected nodes in the whole network at the end of the propagation, the nodes that are most likely to be infected or play an important role in the propagation process should be quarantined in advance. According to the introduction in Section II-B, the immune nodes will no longer participate in the propagation of harmful information or virus in the proposed propagation model. That is, the immune nodes will no longer be infected or spread harmful information or virus to other nodes. The nodes in the infected node set will propagate harmful information or virus to the nonimmune nodes in the network at each time according to the propagation probability. Based on the proposed network propagation model and the improved node immune gain, an effective node immune method is proposed, which does not require the number of immune nodes in advance. The efficiency and accuracy of the method are improved by immunization against the neighbors of infected nodes.

The node immunization methods should not only restrain the propagation of harmful information or virus but also needs to maintain the original function and information exchange of the network. Previous algorithms always need the number of nodes to be immunized in advance. It is one sided to achieve control of immune nodes in each time step only by a fixed number. For example, the number of influential nodes that needed to be immunized in a certain time step may be more or less than the number of nodes that will be immunized in this time step. This will reduce the accuracy and the final effect of the entire immunization process. The number of necessary immune nodes in each time step should be judged more flexibly because of the randomness of the initial infection set during each propagation of harmful information or virus. The immune method proposed in this article determines the number of immunized nodes in each time step by a given threshold value. It will be controlled from the point of view of the attribute of the node itself, so as to improve the immune accuracy.

The initial infection node set is determined according to the network propagation model introduced in Sections III-A and III-B. In this article, it is assumed that foreign nodes only spread harmful information or virus at the initial moment. After the initial infected node set is determined, the improved immune gain value is calculated for the neighbor nodes of nodes in the infected node set at each time. Then, we need to determine whether or not to immunize the node against a given threshold, and ensure that immune nodes no longer participate in the propagation of harmful information or virus. The procedure of the node immune algorithm is shown in Algorithm 3.

In summary, the overall flowchart of the proposed algorithm is shown in Fig. 3.

Algorithm 3 Procedure of the Node Immune Algorithm

Input: it: Immune threshold, I_0: Initial infected nodes set, T: maximum time step, Propagation model, Node connection information;
Output: Immune nodes set IM, Infected nodes set I;

```
for t = 0:T do
    for j = 1 : length(I(t)) do
        Calculate MIG(j) ← each neighbor node j of the nodes i;
        if MIG(j) > it then
            Add the node j to IM;
        end if
        i = i + 1;
    end for
    I(t) ← Run propagation model;
    t = t + 1.
end for
```

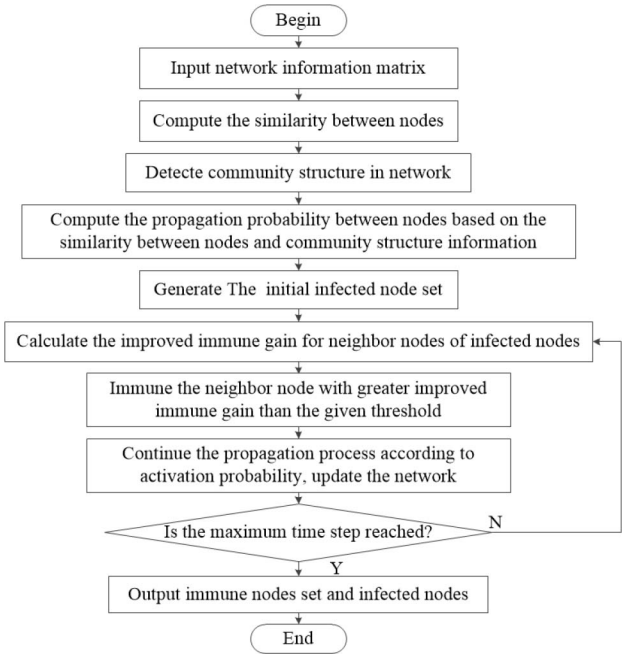


Fig. 3. Overall flow chart of the proposed algorithm.

IV. Experimental Results and Analysis

A. Comparison Algorithms and Datasets

In order to validate the effectiveness of this article, the NICT is compared with the following algorithms. The network propagation framework of the following algorithms is based on the framework given in this article. All the immune gain involved is based on the improved immune gain given in this article.

1) *Formal Position-Based Acquaintance-Degree Algorithm (FPAD):* In [53], an algorithm combining acquaintance algorithm with the formal position was proposed. Since the proposed network propagation model is also based on the similarity between nodes and community structure information, this method is applied to node immunization as one of the comparative algorithms. In the problem of node immunization, formal posts are regarded as the core nodes in the network [54]. In the acquaintance algorithm, nodes with a higher degree in the core node neighborhood will be immunized before

TABLE III
REAL NETWORK INFORMATION

Networks	Node	Edge
Karate	34	78
Dolphin	62	159
Football	115	613
SFI	118	200
Netscience	1589	2742
Power	4941	6594
PGP	10680	24340
Internet	22963	48436

propagation begins. The total number of immune nodes is k.

2) *Dynamic Immunization (DI) Algorithm:* The method of dynamic node immunization given in [55] is adopted. The immune gain between the current time step and the next time step is compared before the node is immunized.
3) *Static Immunization (SI) Algorithm:* The most classical static immunization algorithm is used to immunize k nodes in the network before propagation begins, where k is the number of nodes immunized in advance.
4) *UI Algorithm:* The UI method is a kind of the dynamic method. There are k/T nodes immunized in each time step during propagation, where T is the total number of propagation time steps.
5) *EI Algorithm:* The EI method is also a kind of dynamic method. $2^{-t} * k$ nodes are immunized in each time step during propagation, where t is the number of the current propagation time step.

The methods are tested on eight real-world networks ranging from small scale to large scale. The real network information is shown in Table III.

B. Evaluation Index

The common evaluation method of the node immune effect is to compare the number of infected nodes in the network at the end of the propagation. When the number of immune nodes is the same, the smaller the number of infected nodes is, the better the immune effect is. The algorithm in this article does not need to give the number of immune nodes in advance, so in order to better compare the algorithm with the contrast algorithm, a specific calculation method is proposed. The evaluation index node immune effect (NIE) is presented as follows:

$$\mathrm{NIE} = \sqrt{\frac{n}{I} + \frac{n}{IM}} \tag{9}$$

where n is the total number of nodes in the network, IM is the number of immune nodes at the last moment, and I is the number of infected nodes at the last moment. It can be seen that for the same network, when the number of immune nodes is the same, the larger the number of infected nodes at the final moment is, the smaller the value of NIE is, and the worse the immune effect is. When the number of immune nodes is the same, the smaller the number of infected nodes is, the greater the value of NIE is, and the better the immune effect is.

It is need to note that in theory I and IM will not be zero. When I or IM is zero, the value of NIE will be positive infinity. Because in the propagation framework, the number of propagation time steps has been given in advance, it has no significant effect on the propagation when the immune nodes are too small. The infected node may not increase due to the limitation of the number of steps, thus the NIE value will be larger. Therefore, the NIE function is not suitable for the results with too few immune nodes.

C. Parameter Analysis

This section will analyze the parameters in this algorithm. The main parameters in the propagation model are propagation threshold pt and immune threshold it.

The propagation threshold pt is used to determine whether the node can be activated during the propagation process. That is, the activation probability of the node is compared with the propagation threshold. The propagation model in this article is tested on several real networks. There are 34 nodes in the karate network and 115 nodes in the football network. The initial number of infected nodes $a * b$ is set to 6 and the propagation time step T is set to 10. There are 1589 nodes in the netscience network. The initial number of infected nodes $a * b$ is set to 20, and the propagation time step T is set to 50. There are 10 680 nodes in the PGP network. The initial number of infected nodes $a * b$ is set to 50, and the propagation time step T is set to 100. After 20 runs, the average numbers of final infected nodes in several real networks under different propagation thresholds pt are shown in Fig. 4.

Fig. 4 shows that when the propagation threshold pt is set to 0.1 or 0.2, all nodes in the karate network are infected at the end of the propagation. When the propagation threshold pt is set to 1, no node is infected during propagation. In the football network, when the propagation threshold pt is set to 0.1 to 0.3, all nodes in the network will be infected at the end. When the propagation threshold pt is set to 0.9 or 1, few nodes are infected during propagation. In the netscience network, when the propagation time step T is set to 50 and the propagation threshold pt is set to 0.1 or 0.2, nearly one-third of the nodes in the network are infected at the end. When the propagation threshold pt is set to 1, only about 0.3% of the nodes are infected during propagation. In the PGP network, when the propagation threshold pt is set to 1, only about 0.2% of the nodes are infected during the propagation process. That is, nearly no harmful information is propagated. In conclusion, when the propagation threshold pt is too high or too low, the propagation model cannot reflect the harmful information or virus propagation process well. Considering the performance on both small-scale and large-scale networks, the propagation threshold pt of the propagation model is set to 0.5.

The immune threshold it directly controls the number of immune nodes in each time step in the immune algorithm. The immune threshold should be selected based on different network structures and sizes. Since the propagation probability in this article is determined by the similarity between nodes and the community structure information, and the similarity between nodes is a cosine similarity function determined by

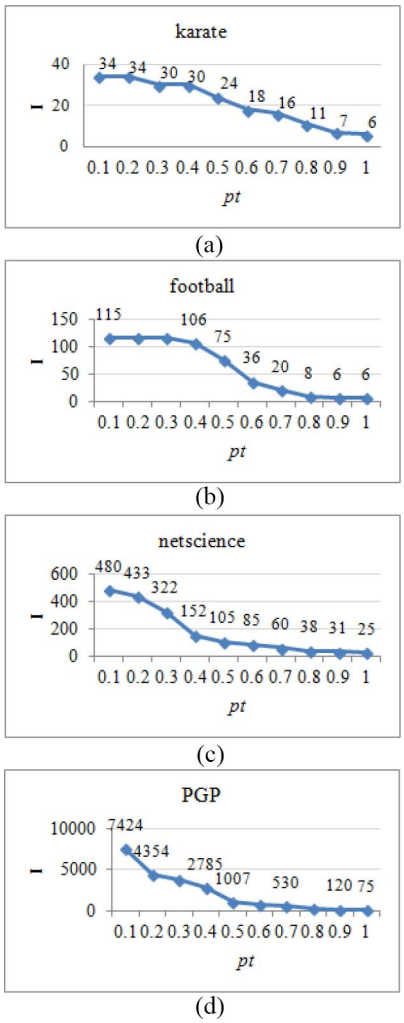


Fig. 4. Average number of infected nodes in four real networks under different propagation thresholds *pt*. (a) Karate network. (b) Football network. (c) Netscience network. (d) PGP network.

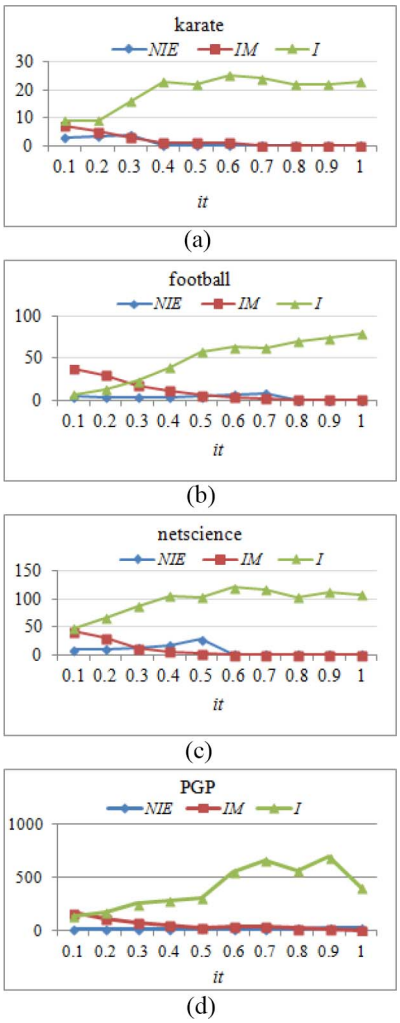


Fig. 5. Final immune situations of four real networks under different immune thresholds. (a) Karate network. (b) Football network. (c) Netscience network. (d) PGP network.

the node degree, the determination of the immune threshold can be preliminarily determined according to the average node degree of the network. The method is tested on several real networks. There are 34 nodes in the karate network, and the average node degree is 4.59. There are 115 nodes in the football network, and the average node degree is 10.66. There are 1589 nodes in the netscience network, and the average node degree is 3.45. There are 10 680 nodes in the PGP network. The average node degree is 4.55. The initial number of infected nodes $a*b$ is set to 50, and the propagation time step T is set to 100. After 20 runs, the final immune situations of four real networks under different immune thresholds are compared as shown in Fig. 5.

Fig. 5 shows that when the immune threshold is set to 0.2, a relatively high NIE value can be obtained when the number of immune nodes is small. This indicates that the immune nodes play an important role in the transmission of harmful information. When they were immunized, the number of infected nodes decreased significantly at subsequent times. When the immune threshold is above 0.4, the number of immune nodes is 0. In the football network, when the immune threshold is set to 0.4, a relatively high NIE value can be obtained when the number of immune nodes is small. When the immune threshold is above 0.5, it is easy to generate too few immune nodes. In the netscience network, when the immune threshold is set to 0.3, a relatively high NIE value can be obtained when the number of immune nodes is small. In the PGP network, when the immune threshold is set to 0.3, a relatively high NIE value can be obtained when the number of immune nodes is small.

In summary, the immune threshold setting is related to the average node degree of the network. Through the analysis and preliminary test of the average node degree of the network, the specific value of the immune threshold in this experiment is shown in Table IV.

D. Model Analysis

In order to verify the validity of the propagation model proposed in this article, the comparison with the WIC model in the karate network is analyzed. There are 34 nodes in the karate network. As can be seen from Fig. 2, the karate network

TABLE IV
AVERAGE NODE DEGREE AND IMMUNE THRESHOLD SETTING OF EIGHT REAL NETWORKS

Networks	Average degree	it
Karate	4.59	0.2
Dolphin	5.13	0.3
Football	10.66	0.4
SFI	3.38	0.2
Netscience	3.45	0.3
Power	2.67	0.1
PGP	4.55	0.3
Internet	4.22	0.3

is divided into three community structures by community integration strategy adopted in this article. The initial infected nodes are all set to 6, and the propagation time step is set to 10. In the WIC model, the edge weight between nodes i and j is $w(i,j) = 1/|N_{in}(j)|$, where $N_{in}(j)$ is the neighbor node set of node j in the infected node set. Fig. 6 shows the information dissemination of the WIC model in the karate network. Fig. 7 shows the information dissemination of the model proposed in this article. pt is set to 0.5. The nodes whose color changes in each time step are the activated nodes at the end of the time step. The red nodes when $t = 0$ are the nodes randomly selected as the initial active nodes. As can be seen from Fig. 6, the number of activated nodes in the WIC model reaches its maximum at $t = 3$. In the end of the time step, only two nodes in the network were not activated. The number of nodes activated at $t = 1$ was the largest and significantly more than other times. The propagation probability only depends on the state of the neighbor nodes. This model does not take into account the actual laws of the network structure and information propagation. As can be seen from Fig. 7, the propagation model presented in this article no longer produces active nodes after $t = 4$. In the end, 71% of the total number of nodes in the network are activated by setting the parameter propagation threshold $pt = 0.5$. At each time step, the number of activated nodes is more uniform. The propagation depends on the degree of closeness between nodes. In the propagation process, the activated node still has the opportunity to disseminate information to any neighbor node, which is more in line with the actual situation of the real network. Therefore, the model proposed in this article based on network community structure information and node similarity can better explain the essence of information or virus dissemination in the real network.

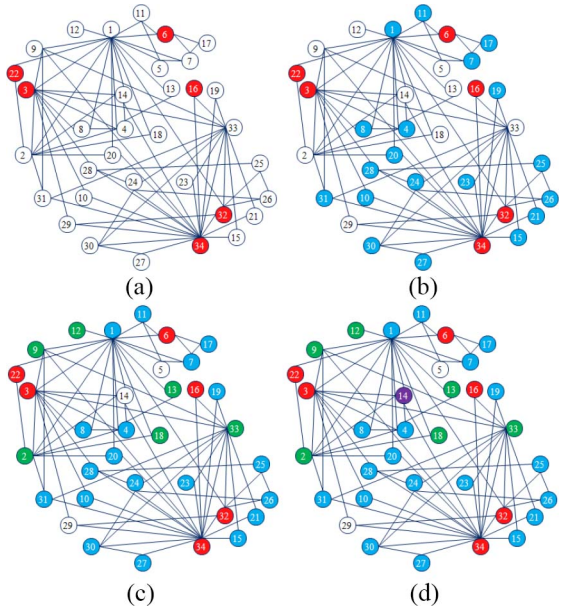


Fig. 6. Information dissemination of the WIC model in the karate network. (a) $t = 0$. (b) $t = 1$. (c) $t = 2$. (d) $t = 3$.

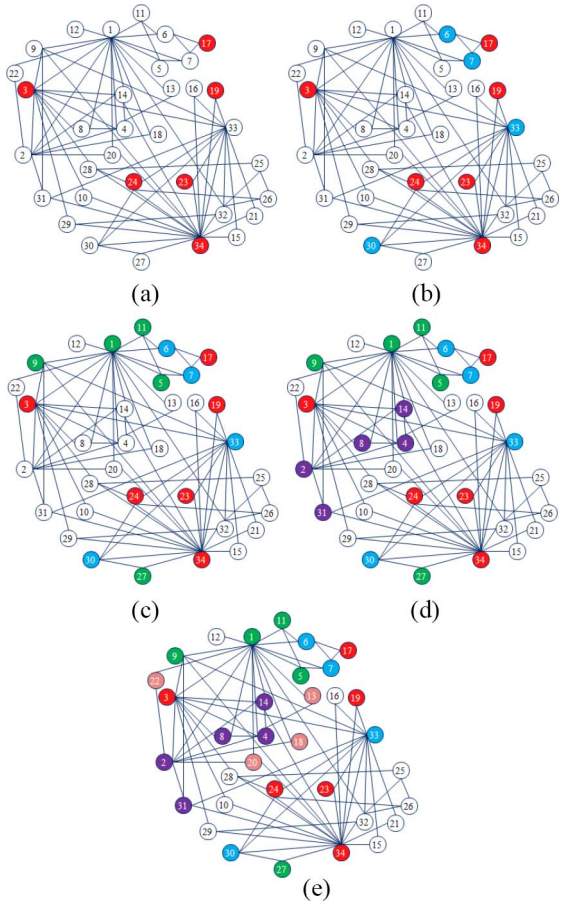


Fig. 7. Information dissemination of the model proposed in this article in the karate network. (a) $t = 0$. (b) $t = 1$. (c) $t = 2$. (d) $t = 3$. (e) $t = 4$.

E. Performance Comparison

In this section, the node immune effect of this algorithm and comparison algorithms are analyzed in detail. In order to compare several algorithms fairly, all the algorithms are tested on the propagation model proposed in this article. Assume that the new nodes only spread harmful information at the initial time. For small-scale networks (karate, dolphin, football, and SFI), the number of initial infected nodes is set to 6, and the maximum value of propagation time step is set to 10. For medium-scale networks (netscience and power), the initial number of infected nodes is set to 20, and the maximum propagation time step is set to 50. For large-scale networks (PGP and Internet), the initial number of infected nodes is set to 50, and the maximum propagation time step is set to 100.

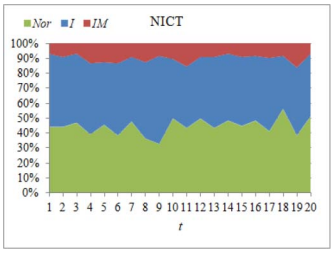


Fig. 8. Node immunization result of the NICT running 20 times on the power network.

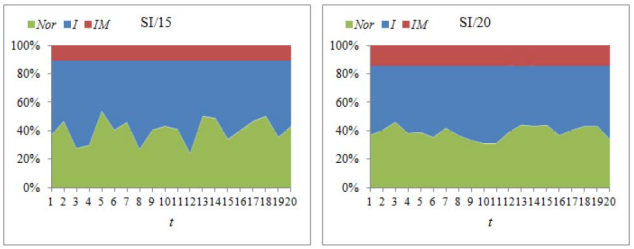


Fig. 9. Immunization result of the SI algorithm running 20 times on the power network. (a) $k = 15$. (b) $k = 20$.

In order to prove the effectiveness of the NICT, the node immunization results of the NICT and the comparison algorithms in the power network are analyzed. The three parts of the graph from top to bottom represent the number of final immune nodes, the number of infected nodes, and the number of normal nodes. In order to compare the algorithms intuitively, the normal nodes in all result graphs are subtracted from the same number according to the concrete situation of the power network. Fig. 8 shows the node immunization result of the NICT after 20 runs on the power network. Fig. 8 shows that the NICT controls the number of the final infected nodes between 50 and 83 in the power network. The number of immune nodes is between 10 and 23. Although the number of immune nodes in 20 runs of the NICT is different, the number of final infected nodes in the network is relatively stable.

We set the number of immune nodes on the power network to 15 and 20, respectively, for each of the five comparison algorithms. Fig. 9 shows the immunization result of the SI algorithm running 20 times on the power network.

As can be seen in Fig. 9, when the number of immune nodes is set to 15, the best result of the SI algorithm in the power network controls the number of infected nodes to 50, and the worst result is to control the number of infected nodes to 92. The immune effect is very unstable. When the number of immune nodes is set to 20, the best result of the SI algorithm in the power network is to control the number of infected nodes to 56, and the worst result is to control the number of infected nodes to 77. At the cost of increasing the number of immune nodes, the worst case of node immunity is improved in the SI algorithm, but the overall effectiveness is not obvious.

Fig. 10 shows the immunization result of the UI algorithm running 20 times on the power network.

As shown in Fig. 10, when the number of immune nodes is set to 15, the result of the UI algorithm in the power network

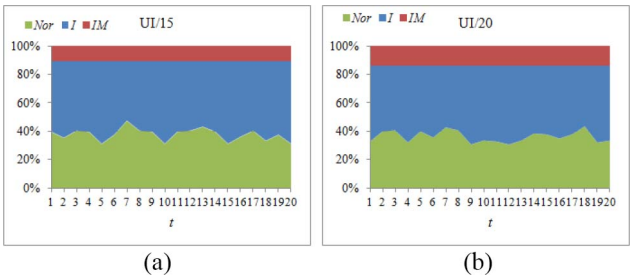


Fig. 10. Immunization result of the UI algorithm running 20 times on the power network. (a) $k = 15$. (b) $k = 20$.

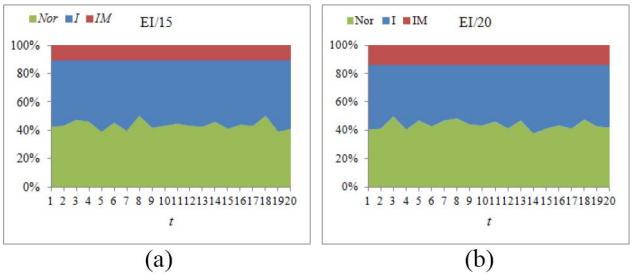


Fig. 11. Immunization result of the EI algorithm running 20 times on the power network. (a) $k = 15$. (b) $k = 20$.

Fig. 12. Immunization result of the DI algorithm running 20 times on the power network. (a) $k = 15$. (b) $k = 20$.

controls the number of infected nodes to 59–82. When the number of immune nodes is set to 20, the result of the UI algorithm in the power network is to control the number of infected nodes to 60–78. Compared with the SI algorithm, the stability of the UI algorithm is better, but the overall effect is not high.

Fig. 11 shows the immunization result of the EI algorithm running 20 times on the power network. Fig. 11 shows that when the number of immune nodes is set to 15, the result of the EI algorithm in the power network controls the number of infected nodes to 55–71. When the number of immune nodes is set to 20, the result of the EI algorithm in the power network controls the number of infected nodes to 51–68. It can be seen that the immune effect of the EI algorithm is relatively stable. But at the cost of increasing the number of immune nodes, the immune effect of the EI algorithm in the power network is not obviously improved.

Fig. 12 shows the immunization result of the DI algorithm running 20 times on the power network. Fig. 12 shows that when the number of immune nodes is set to 15, the result of the DI algorithm in the power network controls the number of

(a) (b)

Fig. 13. Immunization result of the FPAD algorithm running 20 times on the power network. (a) $k = 15$. (b) $k = 20$.

TABLE V
NUMBER OF IMMUNE NODES OF THE COMPARISON ALGORITHM ON EIGHT REAL NETWORKS

Networks	Range of k values in NICT	k of comparison algorithms
Karate	$3 \sim 10$	7
Dolphin	$3 \sim 17$	10
Football	$2 \sim 18$	10
SFI	$5 \sim 23$	10
Netscience	$11 \sim 59$	30
Power	$10 \sim 23$	15
PGP	$48 \sim 193$	100
Internet	$36 \sim 59$	50

infected nodes at 66–96. When the number of immune nodes is set to 20, the result of the DI algorithm in the power network controls the number of infected nodes to 64–93.

Fig. 13 shows the immunization result of the FPAD algorithm running 20 times on the power network. As can be seen in Fig. 13, when the number of immune nodes is set to 15, the result of the FPAD algorithm in the power network controls the number of infected nodes to 57–124. When the number of immune nodes is set to 20, the result of the FPAD algorithm in the power network controls the number of infected nodes at 64–97. That is, at the cost of increasing the number of immune nodes, the worst case of node immunity in the power network is improved by the FPAD algorithm, but the overall immune effect of nodes is not significantly improved. And the immune effect is very unstable.

Compared with Fig. 8, it can be seen that the NICT cannot only determine the number of immune nodes more flexibly but also obtain better immune effect than five algorithms when the number of immune nodes is less.

Obviously, the more the immune nodes is, the less the number of infected nodes is. But this is contrary to the original intention of nodal immunization. When the number of immune nodes increases to a certain number, the immune effect of some immune nodes may be poor. This will lead to the decrease of the impact on the number of infected nodes. Therefore, according to the approximate number of immune nodes by the NICT in each network, the number of immune nodes in the comparison algorithm is set as shown in Table V. The best NIE values of the NICT compared with those of the five comparison algorithms running 20 times in eight real networks are shown in Table VI. The average NIE values of the NICT compared with those of the five comparison algorithms running 20 times in eight real networks are shown in Table VII. Bold numbers denote the optimal values.

TABLE VI
BEST NIE VALUES OF SIX ALGORITHMS RUNNING 20 TIMES IN EIGHT REAL NETWORKS

Networks	SI	UI	EI	DI	FPAD	NICT
karate	3.65	2.87	3.24	2.94	3.41	**3.95**
dolphin	4.06	3.17	3.73	3.73	3.88	**5.13**
football	4.51	4.12	4.75	4.04	3.97	**8.75**
SFI	5.15	4.99	5.43	4.57	5.15	**6.19**
netscience	9.02	8.64	9.67	8.94	8.50	**12.82**
power	20.69	20.32	20.47	20.10	20.39	**24.04**
PGP	14.44	11.41	13.89	12.98	11.18	**17.03**
Internet	24.29	22.79	24.17	22.87	26.67	**29.66**

TABLE VII
AVERAGE NIE VALUES OF SIX ALGORITHMS RUNNING 20 TIMES IN EIGHT REAL NETWORKS

Networks	SI	UI	EI	DI	FPAD	NICT
karate	**3.05**	2.62	2.81	2.63	2.97	3.00
dolphin	3.34	2.86	3.32	2.98	3.29	**3.46**
football	3.78	3.67	4.09	3.75	3.70	**4.41**
SFI	4.47	4.18	4.73	4.23	**4.79**	4.24
netscience	8.64	8.24	**9.21**	8.39	8.28	9.09
power	20.07	19.94	20.16	19.45	19.86	**20.98**
PGP	12.90	10.91	12.83	11.10	10.84	**12.92**
Internet	23.91	22.72	23.81	22.77	26.60	**26.98**

Compare the results of each algorithm in Figs. 8–13 on the power network. It can be found that the NICT has the best control of the number of infected nodes in the network without giving the number of immune nodes in advance. The node immune effect of the EI algorithm is optimal and very stable. It can be found from Tables VI and VII that the most of maximum and average NIE values obtained by the NICT are optimal. Although the maximum NIE value of the EI algorithm is not optimal compared with that of the NICT and SI, the average NIE value of obtained by EI is just lower than that obtained by the NICT. So the effect of the algorithm is stable. The validity of the NIE evaluation index proposed in this article is further verified. By calculating the maximum value of NIE, we can compare the optimal situation of the number of infected nodes that the algorithm can achieve. By calculating the average value of NIE, the stability of the algorithm can be better reflected. Thus, the node immune effect of the algorithm can be evaluated and compared more effectively. As shown in Tables VI and VII, the NICT can get the highest maximum NIE values in eight real networks compared with the five comparison algorithms. The NICT can get five best results of the average NIE values in the eight real-world networks. Therefore, the NICT has a significant improvement compared with the comparison algorithms on the optimal results of the node immunization without giving the number of the immune nodes in advance.Compared with the comparison algorithms, the stability of the NICT is also dominant on the whole. Moreover, in large-scale networks, the advantages of the NICT will be more obvious because the number of immune nodes is more difficult to determine.

V. CONCLUSION

In this article, a dynamic node immune model based on the community structure and threshold (NICT) has been proposed. In real networks, harmful information is more likely to spread

within closely connected small groups. Therefore, we propose a method for calculating the probability of propagation of the information or virus between nodes, based on the community structure information in the network and the similarity between nodes. In addition, in order to be more flexible to immune network nodes, we have proposed an improved immune gain based on the propagation probability of the community structure and node similarity. The improved immune gain value is calculated for the neighbors of infected nodes at each time step, and the immunization of the nodes is determined according to the given threshold. This can effectively prevent the situation of ineffective or too little immunization at certain time steps. Finally, an evaluation index of the node immune effect has been proposed. This evaluation index takes into consideration the numbers of immune nodes and infected nodes at the same time, facilitating the evaluation of the immune effect and the efficiency of the node immunization process. The experimental results show that our proposed NICT method yields a better node immunization effect, and verifies the effectiveness of the NIE index.

However, we note that the NIE function is not suitable for situations in which the number of immune nodes is too small. In future work, this problem will be studied further.

ACKNOWLEDGMENT

The authors would like to express our sincere appreciation to the editors and the anonymous reviewers for their insightful comments, which have greatly helped us improve the quality of this article.

REFERENCES

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, nos. 4–5, pp. 175–308, Feb. 2006.

[2] E. Bakshy, I. Rosenn, C. Marlow, and L. A. Adamic, "The role of social networks in information diffusion," in *Proc. 21st Int. Conf. World Wide Web*, Apr. 2012, pp. 519–528.

[3] J. Wang, Y. Chen, Y. Tang, and Q. Li, "The effect of rumor clarification on Chinese stock markets," in *Proc. PACIS*, Jun. 2016, p. 298.

[4] G. Giakkoupis, A. Gionis, E. Terzi, and P. Tsaparas, "Models and algorithms for network immunization," Dept. Comput. Sci., Univ. Helsinki, Helsinki, Finland, Rep. C-2005-75, 2005.

[5] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Soc.*, vol. 83, no. 6, pp. 1420–1443, May 1978.

[6] A. Nematzadeh *et al.*, "Optimal network modularity for information diffusion," *Phys. Rev. Lett.*, vol. 113, no. 8, Aug. 2014, Art. no. 088701.

[7] D. Centola, V. M. Eguĺluz, and M. W. Macy, "Cascade dynamics of complex propagation," *Physica A Stat. Mech. Appl.*, vol. 374, no. 1, pp. 449–456, Jan. 2007.

[8] C. D. Brummitt, K. M. Lee, and K. I. Goh, "Multiplexity-facilitated cascades in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 85, no. 4, Apr. 2012, Art. no. 045102.

[9] O. Yağan and V. Gligor, "Analysis of complex contagions in random multiplex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 86, no. 3, Sep. 2012, Art. no. 036103.

[10] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Market. Lett.*, vol. 12, no. 3, pp. 211–223, Aug. 2001.

[11] Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: A martingale approach," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, Jun. 2015, pp. 1539–1554.

[12] L. Liu, B. Chen, B. Qu, L. He, and X. Qiu, "Data driven modeling of continuous time information diffusion in social networks," in *Proc. IEEE 2nd Int. Conf. Data Sci. Cybersp. (DSC)*, Jun.–Aug. 2017, pp. 655–660.

[13] W. O. Kermack and A. G. Mckendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc.*, vol. 115, no. 772, pp. 700–721, Aug. 2003.

[14] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics. II. The problem of endemicity," *Proc. Roy. Soc. London A*, vol. 138, no. 834, pp. 55–83, Oct. 1932.

[15] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ, USA: Princeton Univ. Press, 2011.

[16] P. Y. Chen, S. M. Cheng, and K. C. Chen, "Optimal control of epidemic information propagation over networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2316–2328, Dec. 2014.

[17] A. J. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan, "A stochastic differential equation SIS epidemic model," *SIAM J. Appl. Math.*, vol. 71, no. 3, pp. 876–902, Jun. 2011.

[18] F. Xiong, Y. Liu, Z.-J. Zhang, J. Zhu, and Y. Zhang, "An information diffusion model based on retweeting mechanism for online social media," *Phys. Lett. A*, vol. 376, nos. 30–31, pp. 2103–2108, Jun. 2012.

[19] D. Guo, S. Trajanovski, R. van de Bovenkamp, H. Wang, and P. Van Mieghem, "Epidemic threshold and topological structure of susceptible–infectious–susceptible epidemics in adaptive networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 88, no. 4, Oct. 2013, Art. no. 042802.

[20] T. Gross, C. J. D. D'Lima, B. Blasius, "Epidemic dynamics on an adaptive network," *Phys. Rev. Lett.*, vol. 96, no. 20, May 2006, Art. no. 208701.

[21] N. Perra, A. Baronchelli, D. Mocanu, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, "Random walks and search in time-varying networks," *Phys. Rev. Lett.*, vol. 109, no. 23, Dec. 2012, Art. no. 238701.

[22] R. Pastor-Satorras and A. Vespignani, "Epidemics and immunization in scale-free networks," 2002. [Online]. Available: arxiv.cond-mat.0205260.

[23] R. Cohen, S. Havlin, and D. Ben-Avraham, "Efficient immunization strategies for computer networks and populations," *Phys. Rev. Lett.*, vol. 91, no. 24, Dec. 2003, Art. no. 247901.

[24] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 11, pp. 3747–3752, Mar. 2004.

[25] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, and A. Vàzquez, "Distance-d covering problems in scale-free networks with degree correlations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 3, Mar. 2005, Art. no. 035102.

[26] L. K. Gallos, F. Liljeros, P. Argyrakis, and A. Vàzquez, "Improving immunization strategies," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 75, no. 4, Apr. 2007, Art. no. 045104.

[27] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 65, no. 3, Feb. 2002, Art. no. 036104.

[28] B. Wang, G. Chen, L. Fu, and X. Wang, "DRIMUX: Dynamic rumor influence minimization with user experience in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2168–2181, Oct. 2017.

[29] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Proc. Int. Workshop Web Internet Econ.*, 2007, pp. 306–311.

[30] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Aug. 2003, pp. 137–146.

[31] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. IEEE 10th Int. Conf. Data Min. (ICDM)*, Dec. 2010, pp. 88–97.

[32] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Min. Knowl. Disc.*, vol. 25, no. 3, pp. 545–576, Nov. 2012.

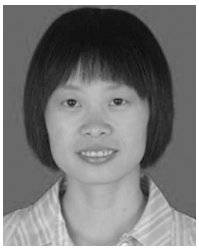
[33] J. Goldenberg, B. Libai, and E. Muller, "Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata," *Acad. Market. Sci. Rev.*, vol. 9, no. 9, pp. 1–18, 2001.

[34] D. Gruhl, R. V. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proc. ACM 13th Int. Conf. World Wide Web*, May 2004, pp. 491–501.

[35] E. Berger, "Dynamic monopolies of constant size," *J. Comb. Theory B*, vol. 83, no. 2, pp. 191–200, Nov. 2001.

[36] Y. Wang, H. Wang, J. Li, and H. Gao, "Efficient influence maximization in weighted independent cascade model," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, Mar. 2016, pp. 49–64.

[37] T. Zhou, J. G. Liu, W. J. Bai, G. Chen, and B. H. Wang, "Behaviors of susceptible–infected epidemics on scale-free networks with identical infectivity," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 5, Nov. 2006, Art. no. 056109.
[38] M. Feng, H. Qu, Z. Yi, X. Xie, and J. Kurths, "Evolving scale-free networks by Poisson process: Modeling and degree distribution," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1144–1155, May 2016.
[39] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, Feb. 2010.
[40] C. Liu, J. Liu, and Z. Jiang, "A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2274–2287, Dec. 2014.
[41] W. Wang and Y. Jiang, "Community-aware task allocation for social networked multiagent systems," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1529–1543, Sep. 2014.
[42] R. Shang, H. Liu, L. Jiao, and A. M. G. Esfahani, "Community mining using three closely joint techniques based on community mutual membership and refinement strategy," *Appl. Soft Comput.*, vol. 61, pp. 1060–1073, Dec. 2017.
[43] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, Feb. 2004, Art. no. 026113.
[44] P. M. Gleiser and L. Danon, "Community structure in jazz," *Adv. Complex Syst.*, vol. 6, no. 4, pp. 565–573, 2003.
[45] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 10, no. 10, Oct. 2008, Art. no. P10008.
[46] A. C. F. Lewis *et al.*, "The function of communities in protein interaction networks at multiple scales," 2009. [Online]. Available: arXiv:0904.0989.
[47] T. Zhou, L. Lü, and Y. C. Zhang, "Predicting missing links via local information," *Eur. Phys. J. B Condensed Matter Complex Syst.*, vol. 71, no. 4, pp. 623–630, Oct. 2009.
[48] L. Leydesdorff, "On the normalization and visualization of author cocitation data: Salton's Cosine versus the Jaccard index," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 1, pp. 77–85, Oct. 2008.
[49] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.
[50] L. Donetti and M. A. Munoz, "Detecting network communities: A new systematic and efficient algorithm," *J. Stat. Mech. Theory Exp.*, vol. 10, no. 10, Oct. 2004, Art. no. P10012.
[51] R. Shang, W. Zhang, L. Jiao, R. Stolkin, and Y. Xue, "A community integration strategy based on an improved modularity density increment for large-scale networks," *Physica A Stat. Mech. Appl.*, vol. 469, pp. 471–485, Mar. 2017.
[52] R. Shang, J. Bai, L. Jiao, and C. Jin, "Community detection based on modularity and an improved genetic algorithm," *Physica A Stat. Mech. Appl.*, vol. 392, no. 5, pp. 1215–1231, Mar. 2013.
[53] G. F. Chami, S. E. Ahnert, N. B. Kabatereine, and E. M. Tukahebwa, "Social network fragmentation and community health," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 36, Jul. 2017, Art. no. E7431.
[54] Z. Lin, X. Zheng, N. Xin, and D. Chen, "CK-LPA: Efficient community detection algorithm based on label propagation with community kernel," *Physica A Stat. Mech. Appl.*, vol. 416, pp. 386–399, Dec. 2014.
[55] D. Yang, X. Liao, H. Shen, X. Cheng, and G. Chen, "Dynamic node immunization for restraint of harmful information diffusion in social networks," *Physica A Stat. Mech. Appl.*, vol. 503, pp. 640–649, Aug. 2018.



Ronghua Shang (Member, IEEE) received the B.S. degree in information and computation science and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

She is currently a Professor with Xidian University. Her current research interests include optimization problems, evolutionary computation, image processing, and data mining.

Weitong Zhang (Member, IEEE) received the B.E. degree from the School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun, China, in 2013, the M.S. degree from the School of Electronics and Communication Engineering, Xidian University, Xi'an, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Circuits and Systems.

Her current research interests include complex networks, intelligent optimization, and deep learning.

Licheng Jiao (Fellow, IEEE) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

From 1990 to 1991, he was a Postdoctoral Fellow with the National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, where he has been a Professor with the School of Electronic Engineering since 1992. He is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. He has led 40 major scientific research projects, and published more than 20 monographs and 100 papers in international journals and conferences. His research interests include image processing, natural computation, machine learning, and intelligent information processing.

He is a member of the IEEE Xi'an Section Executive Committee and the Chairman of Awards and Recognition Committee, a Vice Board Chairperson of the Chinese Association of Artificial Intelligence, the Councilor of the Chinese Institute of Electronics, the Committee Member of the Chinese Committee of Neural Networks, and an expert of the Academic Degrees Committee of the State Council.

Xiangrong Zhang (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science and technology and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 1999, 2003, and 2006, respectively.

She is currently a Professor with the School of Artificial Intelligence, Xidian University. Her research interests include visual information analysis and understanding, pattern recognition, and machine learning.

Rustam Stolkin (Member, IEEE) received the M.Eng. degree in engineering science from the University of Oxford, Oxford, U.K., in 1998, and the Ph.D. degree in computer vision from University College London, London, U.K., in 2004.

He is currently the Director and the Royal Society Industry Fellow of the National Centre for Nuclear Robotics, Birmingham, U.K., and a Professor of robotics with the University of Birmingham, Birmingham, where he is the Founder and the Director of the Extreme Robotics Lab. He is also the Director of Spinout Company A.R.M Robotics Ltd., Barakaldo, Spain. He is highly interdisciplinary, with research interests spanning computer vision and image processing, machine learning and AI, robotic grasping and manipulation, and human–robot interaction.