Graph Embedding-based Smart Vaccination Using Mobile Data

Saeed Jamshidiha, MohammadMohsen Jadidi, Iman Masroori, Pegah Moslemi, Abbas Mohammadi, Vahid Pourahmadi

Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

ABSTRACT: A novel smart vaccination method is proposed in this paper to distribute a limited number of vaccines among the people of a large community, such as a country, consisting of smaller communities like cities or provinces. The proposed method is comprised of two phases; A vaccine allocation phase and a targeted vaccination phase. In the first phase, the available vaccines are allocated to the communities based on demographics and the effectiveness of each type of vaccine. In the second phase, each community is modelled as a contact graph, and the vaccines available to the community are administered to the individuals whose vaccination has the greatest impact on breaking the chain of transmission. As a result of utilizing the Node2Vec graph embedding algorithm, the complexity of the proposed method increases linearly with the number of people in the community, as opposed to common centrality based methods, the complexities of which increase with the square or cube of the number of individuals. Furthermore, the proposed method can distribute multiple types of vaccines with different probabilities of effectiveness. The performance of the proposed method is comparable to the common centrality based vaccination methods, while its complexity is lower. The results of the simulation show a 20% decrease in the peak number of infected individuals.

1- INTRODUCTION

The COVID-19 pandemic, first diagnosed in Wuhan, China in December 2019, spread rapidly around the world in a short period of time and seriously affected the health of many people in many countries. The World Health Organization (WHO) described the outbreak as a warning to the health of all human beings, and in March 2020 recognized the disease as a global epidemic [1]. The first case of COVID-19 disease in Iran was registered on February 19, 2020 [2].

Due to the high prevalence of COVID-19 and the destructive effects of the disease on communities, various countries and institutions are looking for solutions to reduce the destructive effects. One way to prevent the spread of this disease is vaccination, but since the rollout of the vaccine takes time, and as a result of limited resources, it is not possible to vaccinate everyone immediately. Smart vaccination methods can use limited vaccines most efficiently to disrupt the transmission of the COVID-19 infection.

In order to target individuals for vaccination effectively, it is required to obtain data about the conditions of people, and the contacts among them. The information obtained from this data can be used to determine the extent of involvement of each person in the transmission chain of the disease, and hence their vaccination priority. Wireless communications networks can be leveraged to obtain this information in a convenient

and safe manner, e.g. by using smartphone applications. Other methods for gathering information and contact tracing, have been discussed in [3-11]. This information could also be used to manage the disease, announce warnings, and quarantine people who have been infected, or have been exposed to the infection and are at risk.

The community is divided into small groups in [12], and the individuals with higher vaccination priority are selected by analyzing these groups, without analyzing members themselves. In [13], the communication activities of people in the context of digital networks has been used to find people for targeted vaccinations. Information is gathered from sources such as contact records and Bluetooth measurements of peer-to-peer contacts. Finally, people are selected based on their distances and their contact activities.

The authors of [14, 15] identify priority individuals by gathering contact information in a school and using graph centrality metrics. They install a wireless contact tracing sensor system in a high school and give each student a mobile sensor to record contact information at all times. Finally, they model this information as a graph and use the centrality metrics to target higher priority individuals.

A new algorithm for describing nodes in communities is introduced in [16], which describes nodes in the community based on intra-central and inter-central metric. In [17], in order to minimize contamination of digital networks by
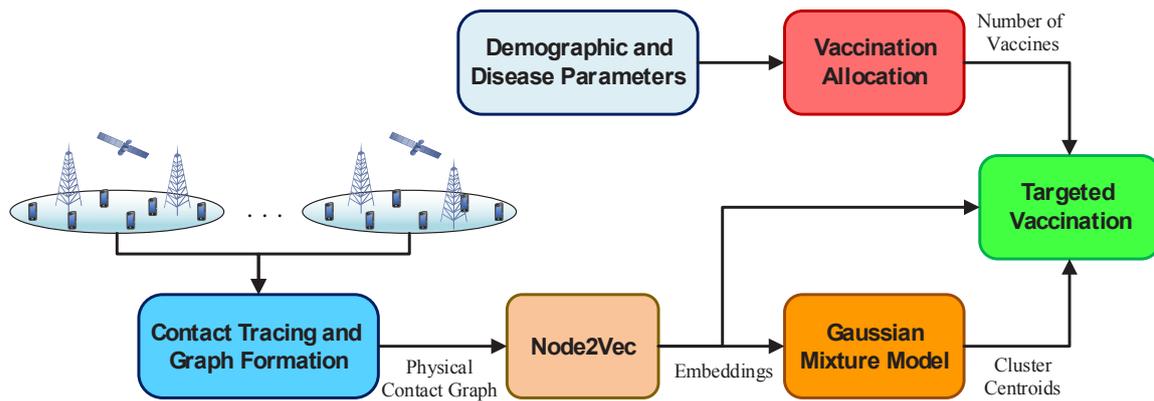
Fig. 1. The proposed system model

malware, several heuristic methods are proposed to distribute resources among network nodes. The authors of [18] hypothesize that the transmission rate of a disease is variable and, considering a simple disease model, investigate the issue of targeted vaccination in scaleless artificial networks.

The equal graph segmentation method is utilized in [19], with resource constraints considerations, to secure the network. A model is suggested in [20] in which a network is initially divided into clusters and security packages are distributed to all or none of the members of a cluster to prevent or counter the effects of computer viruses. A combination of temporary and permanent methods are considered in [21] to prevent the has been considered to prevent the spread of malware on smartphones. The history of infected nodes are communicated with neighbors as a permanent method, and a quarantine is enforced as a temporary method. A vaccination method is proposed in [22] that requires only local information to select a subset of the community for targeted vaccination.

A two step vaccination method is presented in our previous work, [23, 24], comprising of a vaccine allocation phase, and a targeted vaccination phase. In the first step, vaccines are distributed among different communities in a society according to demographics, and in the second step, the contacts among people in each community are modelled as a graph, and the nodes with the highest priorities for vaccination are selected in the graph. A novel metric is introduced to prioritize the nodes for vaccination, taking the centrality of each node, as well as its probability of infection, and probability of infecting others in case of infection into account.

In this paper, we improve upon our previous work by including different types of vaccines with different probabilities of effectiveness in our model, and by improving the computational complexity of the second step of our method (the targeted vaccination phase), by leveraging the Node2Vec algorithm [25] to map the nodes of the graph into a vector space, where the nodes can be clustered and prioritized in a fraction of the time required to calculate the vaccination priorities of the nodes in the graph domain. In this method, the optimal effective vaccination fraction of each community

is calculated in the vaccine allocation step, and the vaccines are allocated to the communities accordingly. In the targeted vaccination phase, the physical contacts among people is first modelled as a graph, which is then mapped into a vector space using the Node2Vec algorithm. The nodes are then clustered in the vector space, using the Gaussian mixture model (GMM), and the nodes closest to the centers of the clusters are selected for vaccination.

The remainder of the paper is organized as follows. The mathematical model for the transmission of the infection, and the system model for the proposed smart vaccination method are discussed in Section 2. The vaccine allocation phase of the proposed method is presented in Section 3, and the targeted vaccination phase in Section 4. Our simulations and results are provided and discussed in Section 5. Section 6 concludes the paper.

2- INFECTION TRANSMISSION MODEL
2-1- System Model
The block diagram of the proposed system for targeted vaccination can be seen in Fig.1. In this model, vaccines are first distributed among different communities (e.g. cities, states, provinces, etc.) based on demographics and pandemic parameters. Then, in each community, the physical contacts among people are obtained from wireless network data, and modelled as a graph. An unsupervised graph representation method, namely the Node2Vec algorithm [25], is then utilized to map this graph to a vector space. The people are then clustered in this vector space, and those closest to the center of their cluster are selected for vaccination, as they are the people with the largest number of contacts, and the highest probability of transmitting the disease.

2-2- Disease Spread Model
Mathematical modeling can predict the possible outcome of infections effectively. The SIR model is a well-known mathematical model of infection transmission that divides communities into three separate compartments called Susceptible, Infected and Recovered [26-29]. The sum of these compartments is equal to the population of the community at all times. Suppose that P is a set of target communities, then

a)$f_j^1, f_j^2 = 0.15, 0.1, p^1, p^2 = 0.95, 0.9$ b)$f_j^1, f_j^2 = 0.25, 0.15, p^1, p^2 = 0.95, 0.9$ c)$f_j^1, f_j^2 = 0.15, 0.1, p^1, p^2 = 0.80, 0.75$
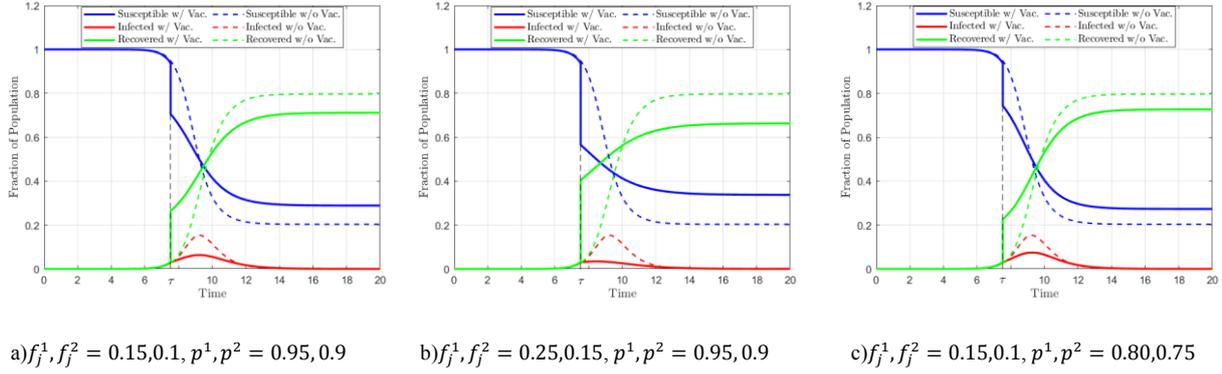
Fig. 2. The value of the fraction of susceptible, infected, and recovered people in community $\mathbf{j}$. Solid lines represented the vaccinated situation, and dashed lines express the unvaccinated situation. It is assumed that $\beta_j = 4, \gamma_j = 2$.

for each community $j \in P$:

$$N_{s_j}(t) + N_{i_j}(t) + N_{r_j}(t) = N_j \qquad (1)$$

Where $N_{s_j}(t), N_{i_j}(t) \, and \, N_{r_j}(t)$ are the number of susceptible, infected and recovered persons at time t in community j, and N_j is number of all person in community j.

In the SIR model, the susceptible people are transferred from the susceptible compartment to the infected compartment with a infection transmission rate β, and the infected people are transferred from the infected compartment to the recovered compartment with a recovery rate γ. In this model, it is assumed that the recovered people, i.e. those who recover from the infection or are vaccinated, are immune to the disease and will no longer be infected. The movements of the people between these compartments are modelled by the following differential equations [26-29]:

$$\frac{ds_j(t)}{dt} = -\beta_j s_j(t) i_j(t) \qquad (2a)$$

$$\frac{di_j(t)}{dt} = \beta_j s_j(t) i_j(t) - \gamma_j i_j(t) \qquad (2b)$$

$$\frac{dr_j(t)}{dt} = \gamma_j i_j(t) \qquad (2c)$$

Where $s_j(t), i_j(t) \, and \, r_j(t)$ are the susceptible, infected and recovered fractions of community j at time t.

3- VACCINE ALLOCATION

In this section, the vaccine allocation step among different communities is discussed. In this step, the goal is to allocate (distribute) the available vaccines among different communities according to the number of people and the disease conditions.

3-1- Vaccination and Mass Immunity

Vaccination is an important measure that can be taken to prevent infectious and contagious pandemics effectively. Vaccination assists in controlling and mitigating the pandemic in two ways. Firstly (which is obvious), people who receive the vaccine become immune to the disease after producing antibodies in their bodies. On the other hand, in the second way, when a significant fraction of the community is vaccinated, the person who has not yet been vaccinated (and is susceptible to the disease) is less likely to get the disease because most of the people they come in contact with have been vaccinated (and do not transmit the disease). In fact, in a community where the vast majority of people have immunity (either due to vaccination or due to recovery from the infection), such people have no share in the spread of the disease, and the chains of infection are disrupted. This indirect effect of vaccination is called mass immunity.

Assume that P is a set of different communities that we want to distribute the available vaccines among them. The number of various type of available vaccines is considered to be K. It is also assumed that vaccines do not surely cause immunity and that a vaccinated person may not be immune to the disease. Now suppose at time τ a fraction of the community $j \in P$ is vaccinated. This fraction is denoted by f_j that $f_j = f_j^1 + f_j^2 + \ldots + f_j^K$. In this relation, f_j^k represents a fraction of the community j that has been vaccinated by the k th type of vaccine that $1 \le k \le K$. Also, p^k represents the probability of the k th vaccine immunity, that is clear $0 < p^k \le 1$.

At time τ, before vaccination, the state of community j is $(s_j(\tau), i_j(\tau), r_j(\tau))$. Immediately after vaccination, the state of the community j alters to $(s_j(\tau) - f_{j,eff}, i_j(\tau), r_j(\tau) + f_{j,eff})$, where $f_{j,eff}$ represents the effective vaccination fraction, defined as $f_{j,eff} = p^1 f_j^1 + p^2 f_j^2 + \ldots + p^K f_j^K$.

In Fig. 2, the values of susceptible persons fraction, infected persons fraction, and recovered persons fraction in the with and without vaccination case, assuming the availability of two different types of vaccines, are shown. In this figure, the dashed lines indicate the without vaccination case, and the solid lines indicate the vaccination case. In

a) The mass immunity function

b) contour plot of community j
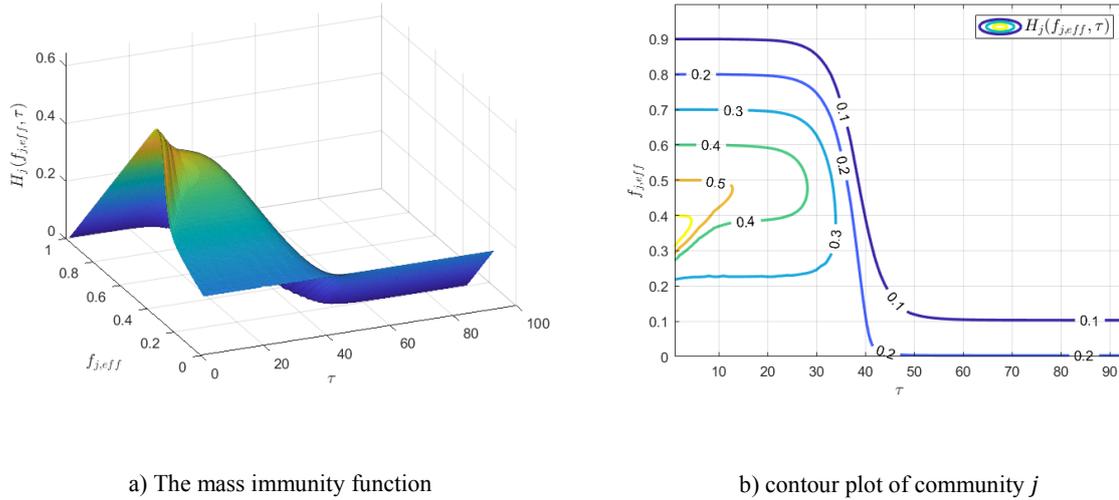
Fig. 3. The mass immunity function of community j

Fig 2a, as can be seen, with vaccination, the summit of the infected persons fraction decreased. On the other hand, the number of persons who did not become infected (the final fraction of susceptible persons) increased. Fig 2b shows that as the vaccination fraction f_j^1 and f_j^2 increase, compared with Fig 2a, the peak of the infected persons fraction further decreases, and the disease mitigates much faster. On the other hand, in Fig 2c, while the probability of vaccine immunity is reduced, as expected with comparing to Fig 2a, the peak of the infected persons fraction is less reduced, and more time is needed for the disease to eradicating.

The final susceptible persons fraction is a beneficial measure of the severity of the pandemic because it represents individuals who have not been infected during the pandemic period. As described, this value actually represents the quantity amount of indirect way of controlling the disease, so the *mass immunity function* $H_j\left(\tau; f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K\right)$ for the community j is defined as follows:

$$H_j(\tau, f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K) = \quad (3)$$
$$\lim_{t \to \infty} s_j(t, \tau; f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K)$$

Where $\tau \geq 0$, $f_j^k, p_j^k \geq 0 \,\forall k = 1, ..., K$ and $f_j^1 + f_j^2 + ... + f_j^K < s_j(\tau)$. Note that the value of the mass immunity function depends on the time of vaccination, vaccination fractions, and their probability of immunity. For convenience, we show the mass immunity function as $H_j\left(f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K\right)$ for fixed vaccination time and $H_j(\tau)$ for fixed vaccination fractions.

Fig. 3 illustrates the mass immunity function with its contour plot for community j regarding vaccination fractions and vaccination time. As can be seen, for a fixed τ, the mass immunity first increasing and then decreasing. Further, for a constant $f_{j,eff}$, the mass immunity always reducing respect to τ.

To facilitate further investigation, the mass immunity

function curves for fixed vaccination fractions and different β_j and γ_j values are presented in Fig. 4a, and the mass immunity function curves for fixed vaccination time and different β_j and γ_j values are presented in Fig 4b. Fig. 4a shows that the mass immunity function is unimodal with respect to $f_{j,eff}$, and is equal to zero at $f_{j,eff} = s_j(\tau)$. Also, for small f, as the disease transmission rate increases more people become infected and mass immunity decreases. On the other hand, while the recovery rate increases mass immunity increases for small $f_{j,eff}$, since more people become immune. The general structure of the mass immunity function is similar for different values of β_j and γ_j. Note that it follows from this figure that injecting more vaccines does not necessarily result in an increase in mass immunity. Fig. 4b shows the mass immunity function is strictly decreasing after τ. It can be concluded that if vaccination is done sooner, mass immunity increases.

3-2- VACCINE ALLOCATION FORMULATION

Assuming the existence of a single community, called community j, the goal of vaccine allocation is to maximize the number individuals who are immunized as a direct or indirect result of vaccination. This goal can be formulated as the following optimization problem:

$$\max_{\mathbf{f}_j} \quad N_j(p^1 f_j^1 + p^2 f_j^2 + ... + p^K f_j^K) + N_j H_j(f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K) \quad (4\,a)$$

$$s.t. \quad N_j f_j^k \leq V^k, \quad k = 1, 2, ..., K, \quad (4b)$$

$$f_j^k \geq 0, \quad k = 1, 2, ..., K \quad (4c)$$

$$f_j^1 + f_j^2 + ... + f_j^K \leq s_j(\tau) \quad (4d)$$

Where V^k is the available number vaccines of type k available, and $\mathbf{f_j} = [f_j^1, f_j^2, ..., f_j^K]$ is the vector of vaccination fractions for community j. The first term in the objective function is equal to the number of people in community j who are immunized directly by vaccination, and the second term is equal to the number of people in community j who
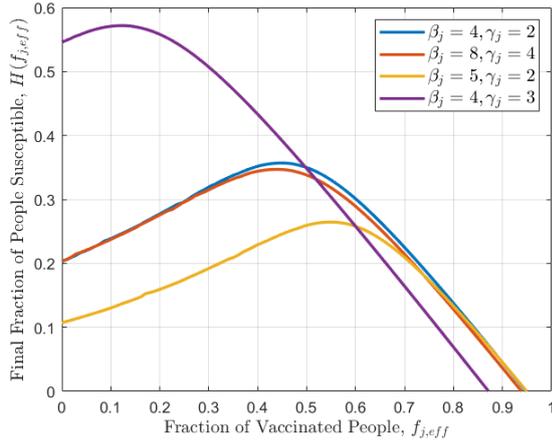
a) constant τ, $i_j(\tau) = 0.025$
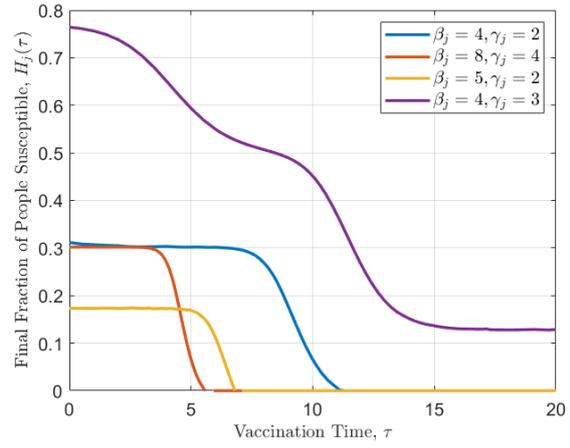
b) constant $f_{j,eff}$ = 0.25

Fig. 4. The mass immunity function

are immunized indirectly by being surrounded by immune people (mass immunity). Inequality (4b) indicates a limit on the number of available vaccines. Inequality (4c) states that the vaccination fractions must be non-negative. Since vaccines are only administered to the susceptible individuals, the sum of vaccination fractions should be less than the susceptible fraction at the time of vaccination (Inequality (4d)).

Now assume the existence of multiple communities. In this scenario, the aim is to maximize the sum total of the people immunized as a direct or indirect result of vaccination in all communities. By defining the vaccination fraction matrix as $\mathbf{F} = [\mathbf{f_1}, \mathbf{f_2}, ..., \mathbf{f_{|P|}}]^\mathbf{T}$ and the vaccine immunity probability vector as $\mathbf{p_j} = [p^1, p^2, ..., p^K]$ for all $j \in \{1, 2..., |P|\}$, the vaccine allocation optimization problem is formulated as follows:

$$\max_{\mathbf{F}} \sum_{j \in P} N_j (p_j^1 f_j^1 + p_j^1 f_j^2 + ... + p_j^K f_j^K) + \sum_{j \in P} N_j H_j(f_j^1, f_j^2, ..., f_j^K, p_j^1, p_j^2, ..., p_j^K) \quad (5a)$$

$$s.t. \sum_{j \in P} N_j f_j^k \leq V^k, \quad k = 1, 2, ..., K \quad (5b)$$

$$\qquad (5c)$$

$$f_j^k \geq 0, \quad \forall j \in P, k = 1, 2, ..., K \quad (5d)$$

$$f_j^1 + f_j^2 + ... + f_j^K \leq s_j(\tau), \quad \forall j \in P.$$

Which can be written in the vector form:

$$\max_{\mathbf{F}} \sum_{j \in P} N_j (\mathbf{p} \mathbf{f}_j^T) + \sum_{j \in P} N_j H_j(\mathbf{f}_j, \mathbf{p}) \quad (6a)$$

$$s.t. \sum_{j \in P} N_j f_j^k \leq V^k, \quad k = 1, 2, ..., K \quad (6b)$$

$$\mathbf{f}_j \succeq 0, \quad \forall j \in P \quad (6c)$$

$$\mathbf{1}^T \mathbf{f}_j \leq s_j(\tau), \quad \forall j \in P. \quad (6d)$$

Similar to (4), the first term in the objective function is equal to the sum of people from all communities who

have been vaccinated and immunized against the disease, and the second term is equal to the sum of people from all communities who are not vaccinated but are immune as a result of being surrounded by immune individuals. Inequality (5b) states that the number of distributed vaccines of each type cannot be greater than the number of available vaccines of that type. The other two constraints are similar to (4c) and (4d).

To solve this optimization problem, the objective function is examined first. As mentioned earlier, vaccination increases immunity to the disease in two ways (direct and indirect). Function F, which indicates the total effect of vaccination on community j, $f_j^1 + f_j^2 + ... + f_j^K < s_j(\tau)$, is defined as follows:

$$F_j(f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K) =$$
$$H_j(f_j^1, f_j^2, ..., f_j^K, p^1, p^2, ..., p^K) + \quad (7)$$
$$(p^1 f_j^1 + p^2 f_j^2 + ... + p^K f_j^K)$$

This function depends on the vaccination fractions in community j, and the effectiveness of the available vaccines (their immunity probabilities). The first and the second terms on the right of equality indicate the indirect and the direct effects of vaccination, respectively. Note that the objective function of the optimization problem (6) differs from the total effect of vaccination by a constant coefficient (number of persons in community, N_j) only. There is no closed form expression for the mass immunity function, and it must be evaluated numerically, as a result, there is no closed form expression for the total effect of vaccination either. As proven in [23, 30], the function representing the total effect of vaccination is concave and monotonically increasing. An example plot of this function (related to the $H(f_{j,eff})$ of Fig. 4b for $\beta_j = 4$ and $\gamma_j = 2$) is given in Fig. 5. In this figure, $f_{j,max,eff}$ is the maximum effective vaccination fraction and is equal to the susceptible fraction of community j at the
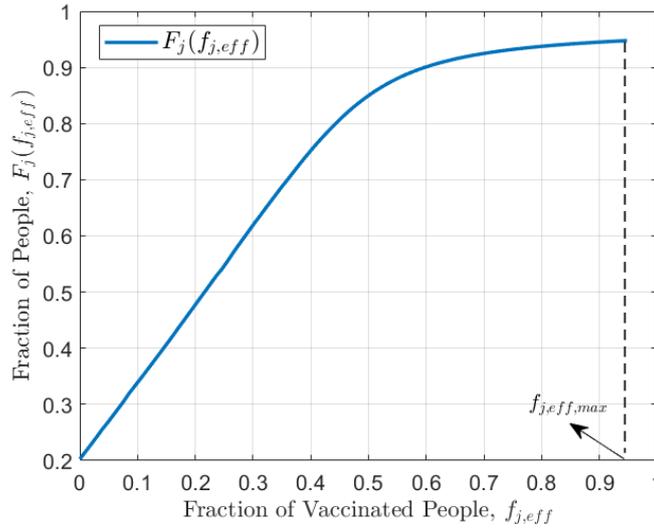
Fig. 5. The total effect of vaccination in community $\mathbf{j}$ for $f_j^1 + f_j^2 + \cdots + f_j^K < s_j(\tau)$ and $f_{j,eff}$

vaccination time. Obviously, by increasing vaccination fraction, immunity always increases, and as a result, it is best to set the vaccination fraction equal to f_j, which, of course, is not possible in practice due to the limited number of vaccines (constraint (5b)).

As mentioned, the objective function of the optimization problem (5) can be written in terms of the weighted sum of the total effect of the vaccination functions. Each community has an increasing and concave total effect of the vaccination function, so the objective function will also be increasing and concave. The constraints are also linear with respect to vaccination fractions, so the optimization problem is convex. To solve this optimization for two community, inspired by the gradient descent method and the algorithm presented in [23], Algorithm (1) is proposed. In this algorithm, first the fractions $f_{j,eff}$ are obtained for each community and then the $\mathbf{f}_j$ vectors are determined from them. To solve a problem for more number of community, ratio of effective vaccination fraction can found for any two communities applying Algorithm 1 and eventually distribute the vaccines based on these ratios.

4- TARGETED VACCINATION

In this section, the physical contact graph is introduced, and a targeted vaccination algorithm is presented that maps the graph into a vector space, and selects target individuals for vaccination by clustering them in that vector space.

4-1- Physical Contact Graph

In order to model the interactions among the people in each community effectively, a graphical model $G = (V, E)$ is utilized, where the set of vertices $V = \{v_1, v_2, ..., v_{N_V}\}$ is comprised of the people in the community, and each edge e_i in the set of edges $E = \{e_1, e_2, ..., e_{N_E}\}$ signifies contact among the two people represented by the node that are linked by the edge e_i, and N_V and N_E represent the number of people in the community

and the number of contacts between them, respectively. Since (physical) contact among people can transmit the disease in both directions, the graph is undirected.

4-2- Graph Embedding

The goal of graph embedding is to find the following mapping in such a way that similarity in the graph domain results in similarity in the vector space domain:

$$f : v_i \in G \rightarrow z_i \in \mathbb{R}^d \tag{8}$$

Where z_i is a vector representation assigned to the node v_i. In order to find such a mapping, similarity measures must be defined for both the graph domain, and the vector space.

A common definition of similarity in the graph domain is having identical neighbors. Measures of similarity in the vector space domain include Euclidean distance and dot product.

Random walk based methods of graph embedding are one of the most widely used families of methods in this fields. In these methods, a label is assigned to each node, and then strings of nodes are produced in the following manner. A node is first selected. Then, one of its neighbors is selected at random, followed by one of the neighbors of the second node, and so on. The selection of the next node in each step could follow a uniform or non-uniform distribution, depending on the specific algorithm.

In these methods, the frequency of the co-occurrence of nodes in different random walks on the graph is defined as a measure of similarity of the nodes in the graph domain, and the dot product of the vectors assigned to them is defined as the measure of similarity in the vector space.

It has been observed [31] that the probability distribution of the occurrence of the nodes in the strings obtained by performing random walks on graphs is very similar to the distribution of the occurrence of words in text. Motivated

Algorithm 1: Optimization algorithm

Input: $N_1, N_2, V^1, ..., V^K, p^1, ..., p^K, \Delta > 0, \varepsilon > 0$

Output: $\mathbf{F}^* = [\mathbf{f}_1^*, \mathbf{f}_2^*]^T$

1 **Initialization**: $f_1^1, ..., f_1^K, f_2^1, ..., f_2^K$ such that $N_1 f_1^k + N_2 f_2^k = V^k, \forall k$

2 $\quad\quad\quad\quad i = 0;$

3 $f_{1,eff} = p^1 f_1^1 + ... + p^K f_1^K;$

4 $f_{2,eff} = p^1 f_2^1 + ... + p^K f_2^K;$

5 $F(\mathbf{f}_{eff}) = F(f_{1,eff}, f_{2,eff}) = N_1 F_1(f_{1,eff}) + N_2 F_2(f_{2,eff});$

6 $\mathbf{d} = (1, \dfrac{-N_1}{N_2});$

7 **while** $g > \varepsilon$ **do**

8 $\quad g_1 = F(\mathbf{f}^i_{eff} + \Delta.\mathbf{d}) - F(\mathbf{f}^i_{eff});$

9 $\quad g_2 = F(\mathbf{f}^k - \Delta.\mathbf{d}) - F(\mathbf{f}^i_{eff});$

10 $\quad$ **if** $g_1 > g_2$ and $g_1 > 0$ **then**

11 $\quad\quad \mathbf{f}^{i+1}_{eff} = \mathbf{f}^i_{eff} + \Delta.\mathbf{d}$

12 $\quad$ **else if** $g_2 > g_1$ and $g_2 > 0$ **then**

13 $\quad\quad \mathbf{f}^{i+1}_{eff} = \mathbf{f}^i_{eff} - \Delta.\mathbf{d}$

14 $\quad$ **else**

15 $\quad\quad$ **break**;

16 $\quad$ **end**

17 $\quad g = \max(g_1, g_2);$

18 $\quad i = i + 1;$

19 **end**

20 $\mathbf{f}^*_{eff} = \mathbf{f}^{i+1}_{eff}$

21 **find** $\mathbf{F}^*$ *s.t.* $(\mathbf{1}_{2\times1}\mathbf{p})\mathbf{F}^* = \mathbf{f}^{*T}_{eff}$

---

by this observation, random walk based methods utilize the Word2Vec transform [32] which is widely used in the natural language processing literature, to map the nodes to a vector space. This mapping is obtained by leveraging a shallow neural network.

In this paper, we have utilized the Node2Vec algorithm [25] to map the nodes to a vector space. This algorithm performs a biased random walk on the graph to be able to model both local and global structures of the graph effectively [25].

In order to visualize the graph embedding process, a randomly generated graph is depicted in Fig. 6. This graph is mapped to a two dimensional embedding space using the Node2Vec algorithm, resulting in the vector space presented in Fig. 7.

Since the similarity measure in the vector space is defined as the dot product of the vectors assigned to the nodes, there are only two completely distinctive clusters of nodes possible in a two dimensional embedding space, and as it can be seen in Fig. 7, the nodes are mapped to an arc, and the users are differentiated by the angle between their vector representations.

Since the users might be grouped into more than two clusters in practice, the embedding space utilized must have higher dimensionality. Furthermore, in order to change the similarity measure from dot product to Euclidean distance, we utilize kernel principle component analysis (PCA), with cosine kernel, and keep the output dimensionality same as the input dimensionality.

Fig 8. presents a two dimensional embedding space obtained by applying Kernel PCA to a ten dimensional Node2Vec representation and keeping the first two basis vectors. As mentioned earlier, all basis vectors are kept in practice, and this two dimensional representation is only intended for visualization.

4-3- Clustering

In order to target the individuals for vaccination effectively, we cluster the people in the representation domain to obtain smaller communities of tightly-knit groups of people. Then, in these small communities, we vaccinate the people who are closest to the center of their cluster. This way, we can ensure
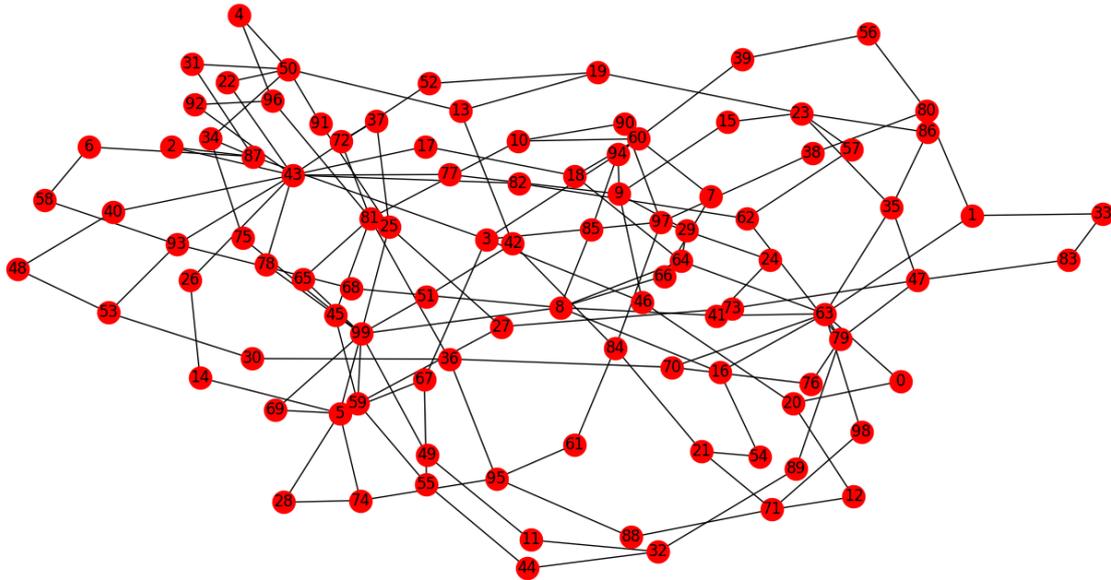
Fig. 6. A randomly generated graph
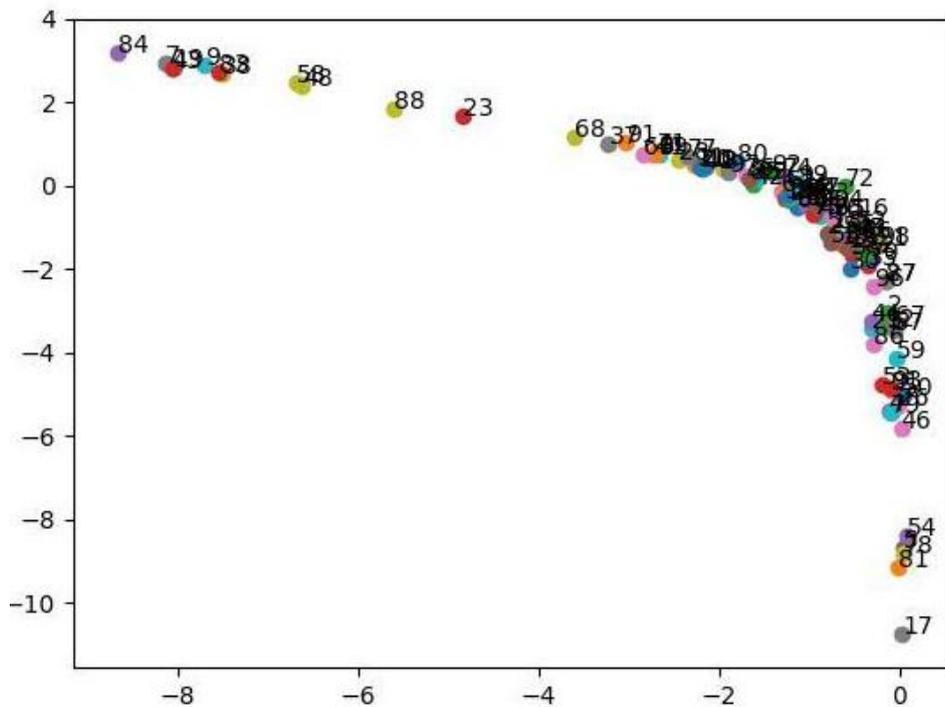
Fig. 7. 2D representation of the graph in Fig. 6 obtained using 2D Node2Vec

that the people who are in contact with most of the other people of their group will be vaccinated.

In this paper, the Gaussian mixture model (GMM) has been used to cluster the people in the representation domain. . The problem of clustering using GMM can be formulated as follows:

$$\max_{k} \quad \pi_k . \Pr\{\mathbf{z_i} \mid c_i = k\}$$

$$\text{s.t.} \quad \Pr\{\mathbf{z_i} \mid c_i = k\} \sim N(\boldsymbol{\mu_k}, \boldsymbol{\sigma_k}^2)$$

$$0 \le \pi_k \le 1 \tag{9}$$

$$\sum_{k=1}^{N_c} \pi_k = 1$$
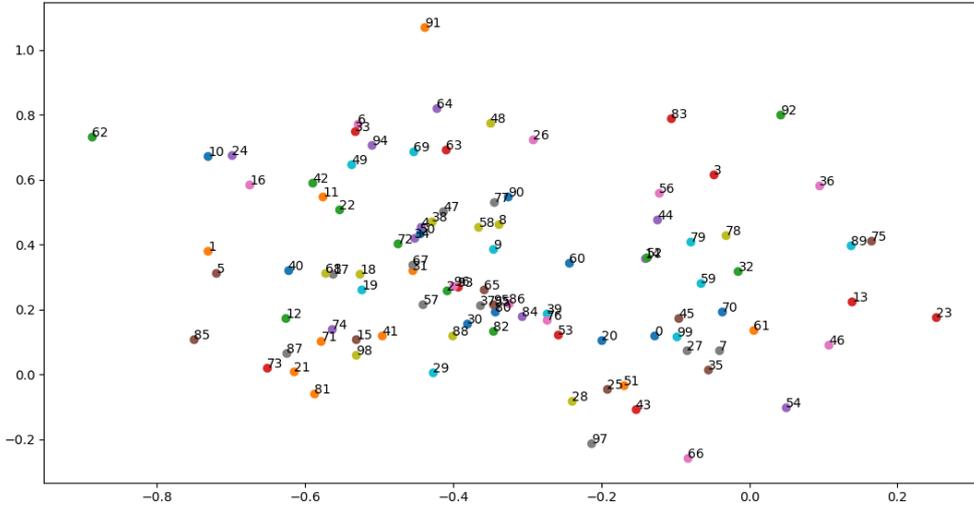
Fig. 8. 2D representation of the graph in Fig. 6 obtained using 10D Node2Vec and kernel PCA.

a) No Vaccination   b) Random Vaccination   c) Targeted Vaccination
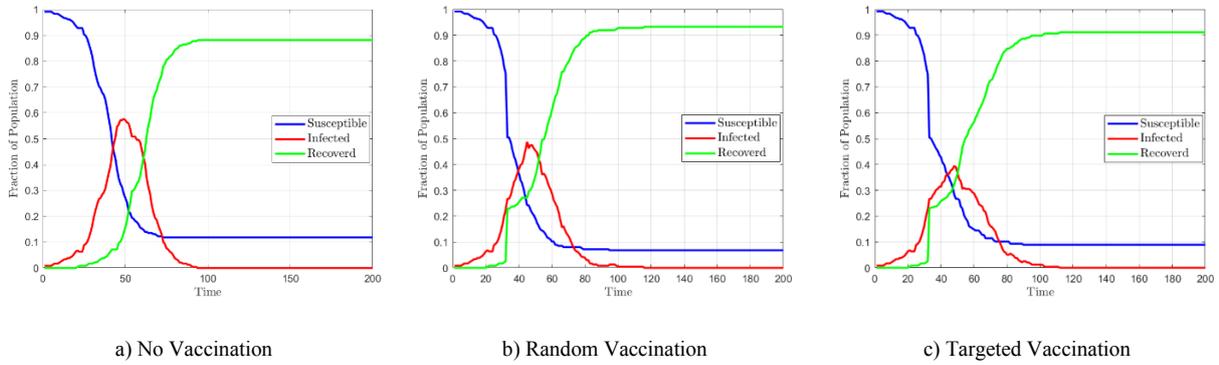
Fig. 9. Susceptible, Infected and Recovered fractions, with $f = \alpha = 0.2$

where z_i is the vector representation of node i, obtained by applying kernel PCA to the Node2Vec embedding space, c_i is the cluster to which node i belongs, $N(\mu_k, \sigma_k)$ is a multi-dimensional normal (Gaussian) distribution with mean μ_k and standard deviation σ_k for cluster k, π_k is the weight associated to cluster k in the GMM, and N_c is the number of clusters in the GMM.

It is also worth mentioning that since the complexity of the Node2Vec algorithm is linear in the number of nodes [25], the complexity of the GMM is linear in the number of datapoints, and the proposed targeted vaccination requires $N_v * N_c$ comparisons, the overal complexity of the proposed targeted vaccination method is linear in the number of nodes in the graph. In comparison, the complexities of the degree-based methods employed in the literature [14, 15, 23] are usually $O(N^2)$ or higher.

5- SIMULATION AND RESULTS

The pandemic has been simulated using the "Primary school - cumulated networks" dataset [8], which is a database of daily contacts among 242 individuals in a primary school. It is assumed that 1% of the nodes of the graph are infected on the first day of the simulation, and all other nodes are susceptible to the infection. Then, the disease starts to spread in the graph, infected more people as time goes on. When the number of infected people reaches the threshold αN_v, where α is a tuning parameter and N_v is the number of nodes in the graph, fN_v vaccines are distributed in the graph, where f is the vaccination fraction. The simulation continues until the eradication of the disease.

Fig. 9 presents the course of the pandemic, following three scenarios: no vaccination (Fig. 9a), vaccinating fN_v individuals randomly (Fig. 9b), and targeted vaccination using the proposed method (Fig. 9c). The blue, red, and green curves represent the susceptible, infected, and recovered fractions, respectively. As expected, the peak number of infected people is reduced from around 0.6 in the no vaccination scenario to around 0.5 in the random vaccination scenario, and to around 0.4 in the targeted vaccination scenario (the proposed method).

75

a) $f = \alpha = 0.2$ b) $f = 0.4, \alpha = 0.2$
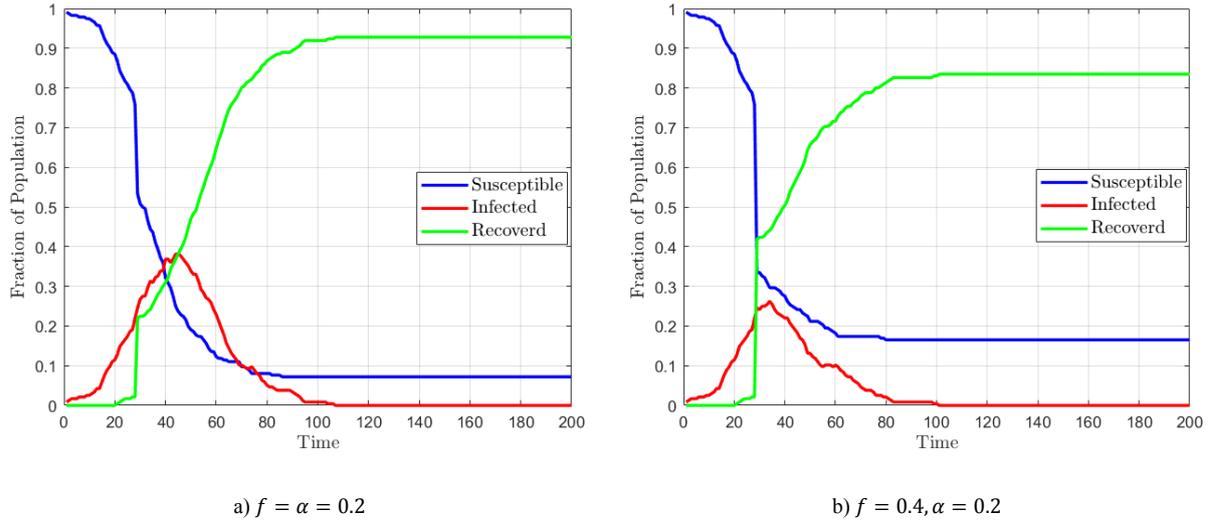
Fig. 10. Susceptible, Infected and Recovered fractions with $\mathbf{f} = 0.4$ and $\mathbf{f} = 0.2$
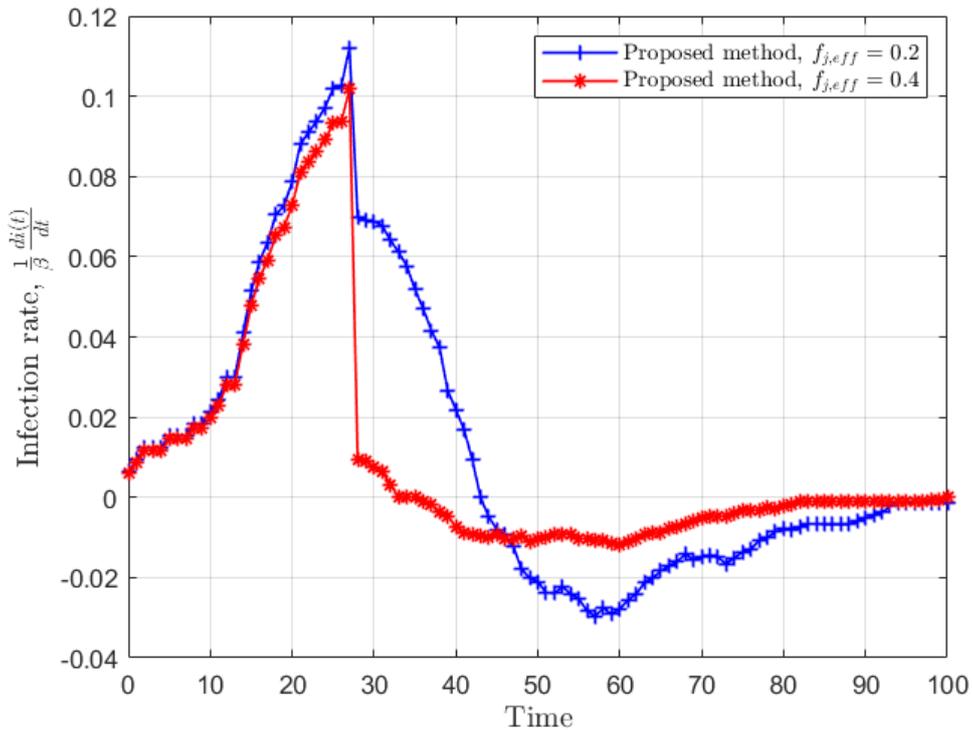
Fig. 11. The Rates of the Spread of the Infection

Fig 10. presents the simulated course of the pandemic with targeted vaccination using the proposed method, following two scenarios where $f = 0.2$ (Fig. 10a), and $f = 0.2$ (Fig. 10b). As expected, more vaccines result in a lower infection peak, and faster eradication of the infection.

The rates of the spread of the infection in the scenarios presented in Fig. 10, can be seen in Fig. 11. These rates are defined as $\frac{d_i(t)}{dt}$, normalized over β. The higher these rates are,

the faster the infection spreads in the community. Negative values indicate that the rate of recovery exceeds the rate of infection. As evident in the figure, larger values of f result in a sharper decline of the rate of the spread of the infection immediately after vaccination.

The proposed method has been compared to various degree based methods for targeted vaccination in Fig. 12. As it can be seen, the proposed method results in the sharpest
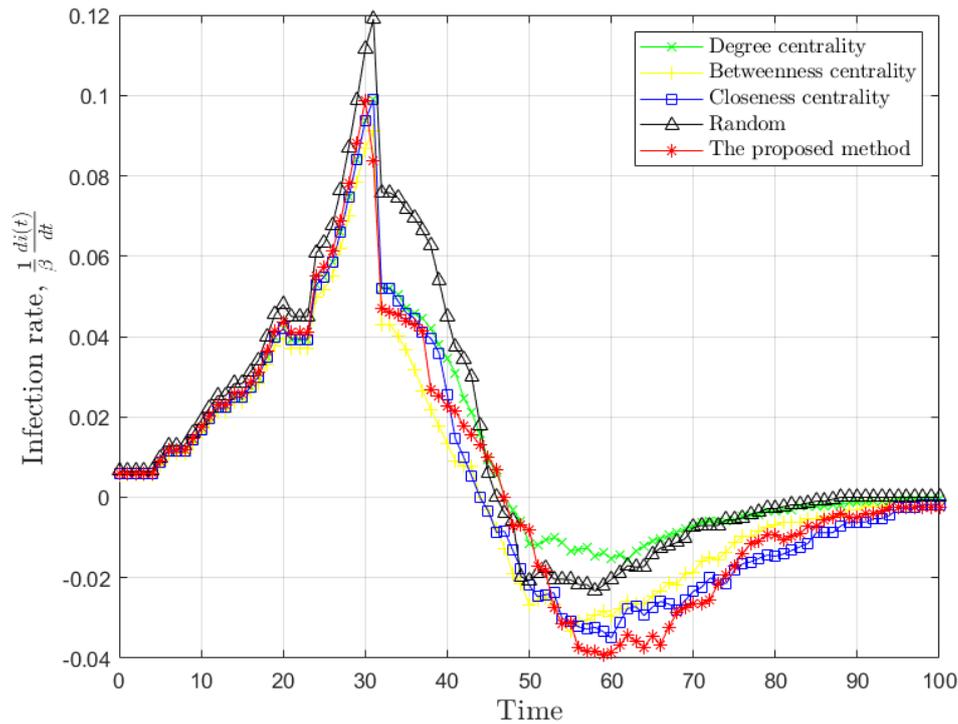
Fig. 12. Comparison of the Proposed Method with Degree-based Methods

decline in the rate of the spread of the disease upon the administration of the vaccines.

6- CONCLUSION

The proposed smart vaccination method comprised of a vaccine allocation phase, and a targeted vaccination phase. In the vaccine allocation phase, multiple vaccine types with multiple effectiveness probabilities are distributed among multiple communities, and in the targeted vaccination phase, the contacts among people are modeled as a graph, which is mapped to a vector space, where the individuals are clustered and the closest people to the cluster centers (in the vector space domain) are vaccinated. The proposed method lowers the peak number of infected people by 20% compared to random vaccination. The complexity of the proposed method is linear in the number of people in the community, which is an improvement upon the commonly used centrality based methods.

ACKNOWLEDGMENT

REFERENCES

[1] World health organization website. (2020, October 5, 2020). Available: https://www.who.int

[2] Iranian Ministry of Health and Medical Education Website. (2020). Available: https://behdasht.gov.ir

[3] N. Saeed, A. Bader, T. Y. Al-Naffouri, and M.-S. Alouini, "When wireless communication faces COVID-19: Combating the pandemic and saving the economy," *arXiv preprint arXiv:2005.06637,* 2020.

[4] S. Sharma, G. Singh, R. Sharma, P. Jones, S. Kraus, and Y. K. Dwivedi, "Digital health innovation: exploring adoption of COVID-19 digital contact tracing apps," *IEEE Transactions on Engineering Management,* 2020.

[5] S. M. Kissler, P. Klepac, M. Tang, A. J. Conlan, and J. R. Gog, "Sparking" The BBC Four Pandemic": Leveraging citizen science and mobile phones to model the spread of disease," *bioRxiv,* p. 479154, 2020.

[6] P. Klepac, S. Kissler, and J. Gog, "Contagion! the bbc four pandemic–the model behind the documentary," *Epidemics,* vol. 24, pp. 49-59, 2018.

[7] S. afroj Moon and C. Scoglio, "Contact Tracing Evaluation for COVID-19 Transmission during the Reopening Phase in a Rural College Town," *medRxiv,* 2020.

[8] J. Stehlé *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PloS one,* vol. 6, no. 8, p. e23176, 2011.

[9] P. C. Ng, P. Spachos, and K. Plataniotis, "COVID-19 and Your Smartphone: BLE-based Smart Contact Tracing," *arXiv preprint arXiv:2005.13754,* 2020.

[10] D. J. Leith and S. Farrell, "Coronavirus contact tracing: Evaluating the potential of using bluetooth received signal strength for proximity detection," *ACM SIGCOMM Computer Communication Review,* vol. 50, no. 4, pp. 66-74, 2020.

[11] G. Grekousis and Y. Liu, "Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review," *Sustainable*

cities and society, vol. 71, p. 102995, 2021.

[12] Z. Lu, Y. Wen, and G. Cao, "Community detection in weighted networks: Algorithms and applications," in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2013, pp. 179-184: IEEE.

[13] E. Mones, A. Stopczynski, A. S. Pentland, N. Hupert, and S. Lehmann, "Optimizing targeted vaccination across cyber–physical networks: an empirically based mathematical simulation study," *Journal of The Royal Society Interface,* vol. 15, no. 138, p. 20170783, 2018.

[14] X. Sun, Z. Lu, X. Zhang, M. Salathé, and G. Cao, "Targeted vaccination based on a wireless sensor system," in *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2015, pp. 215-220: IEEE.

[15] X. Sun, Z. Lu, X. Zhang, M. Salathé, and G. Cao, "Infectious disease containment based on a wireless sensor system," *Ieee Access,* vol. 4, pp. 1558-1569, 2016.

[16] Z. Lu, X. Sun, Y. Wen, G. Cao, and T. La Porta, "Algorithms and applications for community detection in weighted networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 26, no. 11, pp. 2916-2926, 2014.

[17] B. A. Prakash, L. Adamic, T. Iwashyna, H. Tong, and C. Faloutsos, "Fractional immunization in networks," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, 2013, pp. 659-667: SIAM.

[18] X. Fu, M. Small, D. M. Walker, and H. Zhang, "Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization," *Physical Review E,* vol. 77, no. 3, p. 036113, 2008.

[19] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, "Finding a better immunization strategy," *Physical review letters,* vol. 101, no. 5, p. 058701, 2008.

[20] Z. Zhu, G. Cao, S. Zhu, S. Ranjan, and A. Nucci, "A social network based patching scheme for worm containment in cellular networks," in *Handbook of optimization in complex networks*: Springer, 2012, pp. 505-533.

[21] F. Li, Y. Yang, and J. Wu, "CPMC: An efficient proximity malware coping scheme in smartphone-based mobile networks," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1-9: IEEE.

[22] P. Holme, "Efficient local strategies for vaccination and network attack," *EPL (Europhysics Letters),* vol. 68, no. 6, p. 908, 2004.

[23] M. Jadidi, S. Jamshidiha, I. Masroori, P. Moslemi, A. Mohammadi, and V. Pourahmadi, "A two-step vaccination technique to limit COVID-19 spread using mobile data," *Sustainable Cities and Society,* vol. 70, p. 102886, 2021.

[24] M. M. Jadidi, P. Moslemi, S. Jamshidiha, I. Masroori, A. Mohammadi, and V. Pourahmadi, "Targeted Vaccination for COVID-19 Using Mobile Communication Networks," in *2020 11th International Conference on Information and Knowledge Technology (IKT)*, 2020, pp. 93-97: IEEE.

[25] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855-864.

[26] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.

[27] I. Z. Kiss, J. C. Miller, and P. L. Simon, "Mathematics of epidemics on networks," *Cham: Springer,* vol. 598, 2017.

[28] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A time-dependent SIR model for COVID-19 with undetectable infected persons," *IEEE Transactions on Network Science and Engineering,* vol. 7, no. 4, pp. 3279-3294, 2020.

[29] I. Cooper, A. Mondal, and C. G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," *Chaos, Solitons & Fractals,* vol. 139, p. 110057, 2020.

[30] E. Westerink-Duijzer, *Mathematical Optimization in Vaccine Allocation*. 2017.

[31] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701-710.

[32] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *arXiv preprint arXiv:1310.4546,* 2013.