

A Community Based Vaccination Strategy Over Mobile Phone Records

Bowu Zhang¹, Xiuzhen Cheng¹, Rongfang Bie², and Dechang Chen³

¹Computer Science, The George Washington University, DC, USA.

²Information Science and Technology, Beijing Normal University, Beijing, China.

³Division of Epidemiology and Biostatistics, Uniformed Services University of the Health Sciences, MD, USA.

E-mail: bowuzh,cheng@gwu.edu, bierf@163.com, dechang.chen@usuhs.edu

Abstract

Every year, millions of people die from epidemic diseases around the world. The current disease control is mainly conducted by mass vaccination strategies to immunize as many people as possible. However, due to side effects and high costs, it is difficult to be implemented within a large population. In this paper, we address this challenging problem by designing a target vaccination strategy which efficiently prevents diseases from spreading through the population by vaccinating a small fraction of the population. We take advantage of social relationship details extracted from mobile phone records to partition the whole population into communities. A two-level vaccination strategy is then proposed to prevent local and inter-community infections. We evaluate our approach by simulations over real world data and compare its performance with those of other methods. The experimental results show that our approach has the best performance in that it can reduce the number of infections significantly.

1 Introduction

To eliminate or alleviate the threat of epidemic diseases, human beings have taken various measures to protect ourselves. One such important measure is vaccination, where people are given vaccines to obtain immunity against diseases. When one individual gets immunized, he (she) not only effectively protects himself (herself) from being infected, but also indirectly eliminates or lowers the probability of the disease spreading from himself (herself) to others. Current epidemic disease control is mainly conducted by mass vaccination. However, due to side effects and high costs, it is difficult to implement mass vaccination within a large population. Target vaccination [10, 12], on the other hand, could provide a significant improvement in disease control. By im-

munizing a small fraction of the population, target vaccination strategies could not only prevent more infections from happening but also save health costs dramatically. Previous works on target vaccination normally choose individuals at risk of being infected by comparing their contact frequencies, while in fact the probability that an individual becomes infected depends on the nature of the interaction (e.g., time, interaction type) between the individual and the infected person(s). In addition, since target vaccination strategies often require the global network information to select nodes for vaccination, how to collect and analyze such information becomes vital in this approach. (In this paper, a node simply represents an individual.)

Mobile healthcare has attracted great attention from both industrial and research communities with the rise of mobile devices such as cell phones and wearable body sensors in daily lives. With the help of mobile devices, health services can be delivered quickly and conveniently. In this paper, we study how to prevent infections locally and globally by using mobile phone records, which include the communication time. In our approach, we first analyze phone records to capture the interaction pattern among individuals. We then use such an interaction pattern to infer the transmissibility that a disease is transmitted between two individuals. On the basis of the inferred transmissibilities, we partition people into disjoint communities, where people in the same community have a bigger chance to infect each other than people across different communities. A two-level target vaccination strategy is then proposed to provide both local and global control. The main contributions of our work are concluded as follows.

- We determine the transmissibility of disease transmission among individuals by extracting social information from phone records. Coupled with the transmissibilities, a clustering algorithm is employed to partition the population into communities, where people in the same community share a stronger chance to infect each other than people across different communities.
- We introduce the percolation theory to model the diffusion of diseases, from which a percolation threshold is derived to improve the performance of vaccination strategies.
- We propose a two-level target vaccination strategy to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

provide both local and global control over the disease propagation. The communities and the percolation threshold are employed to guide the design of vaccination strategies. In particular, we propose a bridge node identification algorithm to identify key nodes which influence the disease transmission across communities. To the best of our knowledge, this is the first work which is used to prevent inter-community transmissions of diseases.

- Simulations and comparisons are conducted to evaluate the performance of the proposed vaccination strategy. Results show that our vaccination strategy has the best performance in that it can reduce the number of infections significantly.

The rest of the paper is organized as follows. Section 2 presents the most related work. Sections 3 and 4 discuss how to compute the transmissibility of a disease between two nodes and describe how nodes are partitioned into communities. A two-level vaccination strategy is detailed in section 5, and evaluated under various scenarios in section 6. Finally, we conclude our work in section 7.

2 Related Work

Prior work on detecting communities mainly focuses on comparing the intensity of human connections. In [13], the authors form communities based on the number of nodes' connections. They design a security strategy where nodes with a larger number of connections in each community get security patches earlier than other nodes, so that the worm won't infect the entire network. Similarly, the work in [9] detects overlapping communities by comparing the number of connections which have two ending points in two distinct communities with a pre-defined threshold. Few studies on community detection consider the variations in closeness of connections. In this paper, we partition people into disjoint communities by using the strength of the relationship among individuals.

Motivated by recent advances of mobile phones, mobile health care has been actively explored on various applications of mobile devices such as remote health monitoring [4, 5, 11], diagnostic imaging and telemedicine [1], and text messaging [7]. However, few attention has been given to investigating the use of mobile phones in vaccination strategies. So far the only related work is given by [10]. The authors there first extract "kernels" and "communities" by checking the contact frequency on mobile phone bluetooth records. Then they propose a vaccination strategy, where nodes in the kernel sharing more contacts with the infected nodes have a higher priority to be vaccinated, and nodes in the community sharing less contacts with the infected nodes would receive vaccines only after every node in the kernel has been vaccinated. This work is similar to ours in that we both consider extracting social communities from phone records. However, [10] ignores inter-community connections that could lead to a huge number of infections across communities. In addition, [10] does not specify the variations in closeness of relationships, which play a key role in modeling disease spreading. Finally, our vaccination strategy presented in this paper is completely different from that in [10].

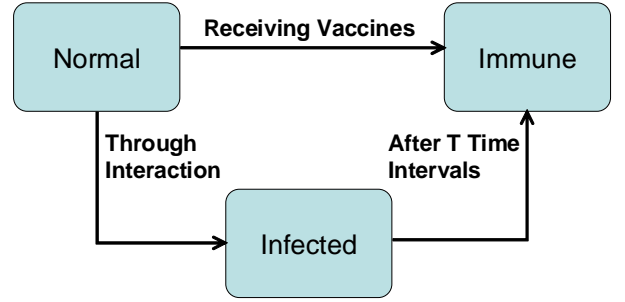


Figure 1. The SIR epidemic model.

Our strategy consists of two phases with one focusing on local prevention and the other on global prevention, while the strategy in [10] relies on the kernels.

3 Percolation Theory

Consider an undirected weighted graph $G(V, E, \tau)$, where V denotes the node set, E denotes the edge set and τ represents the time interval during which the communication data are recorded. A node $v \in V$ represents an individual, while an edge e_{ij} between nodes v_i and v_j indicates that there exist interactions between the two nodes of the pair. Note that an interaction can be a phone call, a bluetooth connection, or a message, depending on the nature of the mobile device communications. Attached to the edge e_{ij} is the weight w_{ij} , which is used to quantify the interaction between v_i and v_j . If the interaction is made through phone calls, then w_{ij} can be the total time used in the phone conversations. If the interaction is done through messages, then w_{ij} can include information such as the number of messages and sizes of messages.

3.1 The SIR Model

In this paper, we employ a SIR epidemic model, in which individuals may be either susceptible to the disease, or infected with the disease, or recovered and immune to the disease. The model has 3 health states: *Normal*, *Infected* and *Immune*. Assume an infected person would stay *Infected* for T consecutive time intervals, where all intervals are of the same length. After these T intervals, the infected node becomes *Immune*, which means the node can not further get infected or infect others. Ideally, people who get vaccinated become *Immune* to the disease automatically. During each time interval, we assume the strength of the relationship between v_i and v_j can be represented by a number $p_{ij} \in [0, 1]$. If one node is infected and the other is normal, p_{ij} could be thought as the "probability" that the normal node gets infected due to the influence of the infected node. The definition of p_{ij} is discussed below. State transitions of SIR are described in Fig.1.

3.2 The Definition Of p_{ij}

The quantity p_{ij} will be based on distances, numbers of friends, and interactions. Specifically, we define p_{ij} ($i \neq j$) in the following way:

$$p_{ij} = f(d_{ij}, N_i, N_j, |\tau|, w_{ij}, \sum_{k \in V, k \neq i} w_{ik}, \sum_{k \in V, k \neq j} w_{jk}) \quad (1)$$

where f is a multivariate function ranging from 0 to 1, d_{ij} is the average physical distance between v_i and v_j , N_i repre-

sents the set of nodes which communicate with v_i directly, and $|\tau|$ denotes the size of the time interval. Below are some notes regarding the above definition of p_{ij} :

- In practice it is easier to compute f as a product of f_1, f_2 and f_3 , that is,

$$\begin{aligned} & f(d_{ij}, N_i, N_j, |\tau|, w_{ij}, \sum_{k \in V, k \neq i} w_{ik}, \sum_{k \in V, k \neq j} w_{jk}) \\ = & f_1(d_{ij}) \times f_2(N_i, N_j) \times \\ & \times f_3(|\tau|, w_{ij}, \sum_{k \in V, k \neq i} w_{ik}, \sum_{k \in V, k \neq j} w_{jk}), \end{aligned}$$

where f_1, f_2 and f_3 are functions ranging from 0 to 1. Clearly, this decomposition allows an easy integration of any additional factor into the process in computing p_{ij} .

- An estimate of d_{ij} could be computed by using built-in GPS or cell phone tower. f_1 should be defined as a non-increasing function of the distance.
- One choice for f_2 is $\frac{|N_i \cap N_j| + 1}{|N_i \cup N_j| + 1}$. The reasoning is based on the observation that if two individuals share a big percentage of friends (compared to the total number of friends of these two individuals), then the relationship between these two individuals is strong.
- An example of constructing f_1, f_2 and f_3 is seen in section 6.
- Each node can be treated in the way that it is most likely to influence itself. Therefore, we define $p_{ii} = 1$.

3.3 Epidemic Threshold

Disease propagation can be viewed as a growing bond percolation process on G . In this process, each edge of G is labeled either infectious or not-infectious. For a pair of nodes v_i and v_j , if v_i is infected, the link e_{ij} would become infectious with a transmissibility P_{ij} . In other words, v_j would be infected by v_i with the transmissibility P_{ij} . The number of people infected is growing when the above process repeats.

Consider a pair of two nodes v_i and v_j who are connected in G . Assume v_i is infected. Also assume that the infection for v_j at one time interval is independent of that at another time interval. Then we see that the transmissibility P_{ij} that v_j would be infected in the next following T time intervals is

$$P_{ij} = 1 - (1 - p_{ij})^T, \quad (2)$$

where p_{ij} is given in (1). We can easily see the average value of all P_{ij} is

$$\bar{P} = 1 - \frac{\sum_i \sum_j (1 - p_{ij})^T}{N^2}, \quad (3)$$

where N is the number of nodes in V . According to the percolation theory [8], when the average reaches a critical value P_c , called the percolation threshold, the number of nodes infected would be infinite. Following [8], we can obtain the following percolation threshold

$$P_c = \frac{\sum_k n_k z}{\sum_k n_k k(k-1)}, \quad (4)$$

where n_k is the number of nodes each having k edges, $z = \frac{\sum_k k n_k}{\sum_k n_k}$ is the average number of edges of a node. When \bar{P} is

above P_c , even a single introduced infection would result in an epidemic. When \bar{P} is below P_c , the disease would die out.

4 Social Community

Social network studies [9, 13] indicate that human beings tend to form communities. People who are in the same community are often intensively connected, and people who are in different communities are less connected. Since most diseases are transmitted through human contacts, once an individual becomes infected, those who are in the same community with the infected individual are more likely to be infected. Therefore, communities could provide a rough estimation of the number of people at risk and enables quicker reactions than considering all the people in the given entire network. Prior approaches [6, 9] form communities based on the graph connectivity, without taking into account the indirect contacts between nodes or specifying the variations in closeness of relationships. In this paper, we partition the given network into disjoint communities by using p_{ij} and an exemplar-based clustering algorithm Affinity Propagation (AP) [3].

We apply AP to the complete graph $G'(V, E', \tau)$, where $E \subseteq E'$. For each edge e'_{ij} in G' , we assign p_{ij} to it as the weight. Note that e'_{ij} may only exist in G' . In other words, there might be no edge connecting v_i and v_j in G . This reflects the scenario in our real world where two people, who have no direct contacts, may infect each other if they work in the same building, or share the same type of transportation, as diseases could be transmitted both by direct and indirect contacts. One well known example is the biggest SARS outbreak in Hong Kong in 2003, for which the infection originated from the plumbing in an apartment where a man sick with the disease lived. Thus we believe that even v_j is not in v_i 's direct contact list, v_j still has a probability to get infected by v_i if p_{ij} is sufficiently large.

When applying AP, our objective is to find a partition such that the sum of p_{ij} between a node and the center of the cluster the node belongs to is maximized. Specifically, let x_i be the center of the cluster to which node v_i belongs (Each center is also a node). The AP algorithm seeks a solution to the following optimization problem,

$$\max \left\{ \sum_{i=1}^N p_{ix_i} \right\}.$$

Details about the AP algorithm can be found in [3]. The algorithm divides V into k disjoint communities, where k can be automatically determined by AP or previously specified. For large scale networks, the map-reduce programming model [2] could be employed in a distributed fashion in order to effectively apply AP. The whole network would be partitioned into subsets, each of which would be processed with AP independently. The results from all subsets would then be merged.

5 Vaccination Strategy

Once a node becomes infected, it would report to the sever. Then the vaccination strategy would be triggered. In this section, we propose a target vaccination strategy which

chooses a small group of individuals for vaccination in order to prevent a disease outbreak from developing into epidemic. The strategy consists of two phases: 1) emergency health care, aiming at preventing local infections without requiring the global knowledge of the network, 2) community health care, aiming at stopping disease spreading across communities. Here communities refer to those identified in section 4 or the existing ones. For convenience, we consider a sequence of consecutive time intervals each having a fixed length $|\tau|$. At the beginning of each time interval vaccination strategies are applied if possible. And at the end of each time interval, cell phone communication records for the interval are assumed to be available.

5.1 Emergency Health Care Scheme

Algorithm 1 Emergency Alerting(v_i, c_1, P_c)

Input:

- The infected node v_i ;
- The percolation threshold P_c ;
- The remedy factor c_1 ;

```

1: function EMERGENCY ALERTING( $v_i, c_1, P_c$ )
2:   Create a queue  $Q$ , push  $v_i$  into  $Q$ , mark  $v_i$  as visited, and set  $P_i = 1$ .
3:   while  $Q$  is not empty do
4:     Pop  $t$  from  $Q$ .
5:     for all  $v_j$  in  $t$ 's contact list do
6:        $P_j = P_t \times P_{ij}$ .
7:       if  $v_j$  is not visited, not infected and  $P_j > c_1 P_c$  then
8:         Send an alert to  $v_j$ .
9:         Mark  $v_j$  as visited.
10:        Enter  $v_j$  into  $Q$ .
11:      end if
12:    end for
13:  end while
14: end function

```

In order to provide efficient health services, the emergency health care scheme targets a group of “local” nodes which are close to the infected nodes, without requiring the global health information. At the beginning of each execution of the proposed scheme, we assume each individual v_i keeps a recent record of the quantity w_{ij} regarding interactions with other people. Then we can compute p_{ij} , P_{ij} and P_c according to equations Eq.(1), Eq.(2), and Eq.(4), respectively. These numbers are saved in every mobile device. Our proposed scheme then makes its recommendation of vaccination in the following way. Once a node v_i gets infected, his (her) mobile device would send an alert to all of his (her) friends v_j with whom $P_{ij} \geq c_1 P_c$, where $c_1 \leq 1$ is a predefined positive constant. A node v_j receiving an alert from v_i would also send an alert to v_j 's friends v_k with whom $P_{ij} \times P_{jk} \geq c_1 P_c$. This procedure repeats until none of the potential product of transmissibilities is at least $c_1 P_c$. All nodes who receive alerts are recommended for vaccination. At the end of each time interval, p_{ij} will be updated in light of available mobile phone records. The proposed emergency health care is described in Algorithm 1. This Breadth First Search based algorithm would quickly spread the news to nodes who are at risk, without waiting for the information of the entire network.

Algorithm 2 Community Alerting Scheme(P, v_i, c_2, P_c)

Input:

- The partition P ;
- The infected node v_i ;
- The percolation threshold P_c ;
- The remedy factor c_2 ;

```

1: function COMMUNITY ALERTING SCHEME( $P, v_i, c_2, P_c$ )
2:   Find the community  $V$  in which  $v_i$  is.
3:   for every node  $v_j$  in  $V$  do
4:      $P_j = 0$ .
5:     for every other community  $V_s$  in  $P$ ,  $V_s \neq V$  do
6:        $P_j^+ = \sum_{m \in V_s} P_{jm}$ .
7:     end for
8:      $P_j = P_{ij} \times P_j$ 
9:   end for
10:  Send alerts to  $v_j$  in  $V$  if  $v_j$  is not infected and  $P_j \geq c_2 P_c$ .
11: end function

```

5.2 Community Health Care Scheme

Despite relatively low interactions among communities, the number of people potentially being influenced by inter-community communication could be very large. A disease can be transmitted to another community via just a single or a few interactions. Once a disease is introduced to other communities, it could dramatically spread out. Take the H1N1 flu in 2009 as an example. The flu first began in Mexico and had been slowly spreading out for months before it became globally epidemic. After the first patient was diagnosed in California, many cases were detected in the US in a very short period of time. Therefore we need to prevent diseases spreading across communities. We define an infected community as a community which has at least one infected member. We also define a bridge node as a node that belongs to an infected community and has close connections with nodes in other communities. Bridge nodes are viewed as messengers transmitting diseases among communities. Assume there are k communities: V_1, \dots, V_k . If v_i in V_r gets infected, we define v_j in V_r to be a bridge node if $P_{ij} \times \sum_{v_m \notin V_r} P_{jm} \geq c_2 P_c$, where $c_2 \leq 1$ is a predefined positive constant. Once the bridge nodes are detected, the server would send alerts to them regarding vaccination recommendations. Our community health care scheme focuses on detecting all the bridge nodes in the population. The scheme is described in Algorithm 2.

6 Simulation

6.1 Simulation Set-up

We validate the proposed vaccination strategy by using real phone records from MIT reality (<http://crawdad.cs.dartmouth.edu/meta.php?name=mit/reality>), where the cell phone bluetooth connections of 103 people are recorded. We use the data of the first 15 days (from Oct. 1, 2004 to Oct. 15, 2004) for constructing communities, and computing the initial p_{ij} and P_{ij} . The data of the next 10 days are used to evaluate the strategy that consists of Algorithms 1 and 2. For convenience, we treat every day as a time interval. In this paper, we use the bluetooth dataset to capture the details of social relationship. Note that collecting

information regarding the social relationship can also be done by using other communication records, i.e, phone calls and twitter re-tweets.

Since the available dataset does not provide any information regarding locations of individuals, we initially compute p_{ij} by using

$$p_{ij} = \frac{|N_i \cap N_j| + 1}{|N_i \cup N_j| + 1} \times \frac{\max\{w_{ij}, 1\}}{\min\{|\tau|, \sum_{k \in V, k \neq i} w_{ik}, \sum_{k \in V, k \neq j} w_{jk}\}}.$$

Here, w_{ij} is equal to the average daily number of minutes used for communications between node v_i and node v_j during the first 15 days (from Oct. 1, 2004 to Oct. 15, 2004). $|\tau| = 24 \times 60 = 1440$ minutes. Both $|N_i \cap N_j|$ and $|N_i \cup N_j|$ represent the average number from the first 15 days. The number 1 appearing on each of the two numerators is used to prevent p_{ij} from becoming 0. This action is based on the thought that a node always has a potential influence on any other node. When updating p_{ij} , we use the information from the current day and previous 14 days and repeat the above approach.

To verify the strength of the proposed strategy, we implement two other strategies for performance comparisons. One strategy employs a random vaccine distribution in the sense that a number of individuals are randomly chosen to be vaccinated according to the number of available vaccines. This strategy has been widely used in the literature and will be denoted as RD in this paper. Another implemented method is the “kernel-community” strategy [10], which has been reported very recently. Here we implement its centralized version, which has proven to have a better performance than its distributed version. Following the setting in [10], two parameters w_1 and w_2 of “kernel-community” scheme are set at 1 and 3, to determine members in kernels and communities. We will use KC to denote the “kernel-community” scheme in our simulations. We execute our proposed strategy for 50 runs. The average number of vaccines used by our strategy will serve as one input of the other two strategies (i.e., RD and KC).

We define final infection ratio as ratio of the total number of infected people during the time of evaluation (from Oct. 16, 2004 to Oct. 26, 2004 in this study) to the size of the entire population (103 in this study). The final infection ratio will be used as the primary performance metric for the evaluation of vaccination strategies in our simulations. We will examine these strategies by varying different parameters such as the initial infection ratio, the recovery cycle T , as well as c_1 and c_2 . The initial infection ratio is defined as the total number of infected persons on the first day (i.e., Oct. 16, 2004) divided by the size of the population. The initial infected persons are chosen randomly in our simulations. We report our experimental results by an average of 50 runs.

6.2 Simulation Result

Fig. 2 reports the final infection ratio vs. the initial infection ratio, when the recovery cycle $T = 6$, $c_1 = 1$, and $c_2 = 1$. For each initial infection ratio, the corresponding number of vaccines required by our vaccination strategy is given in Fig. 3. As discussed in the above subsection, the number of vaccines required by our method is taken as the

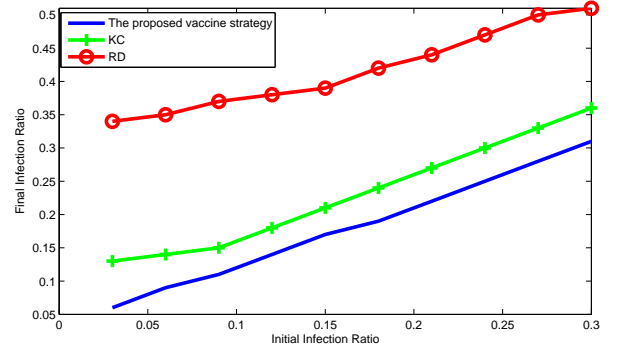


Figure 2. Final Infection Ratio vs. Initial Infection Ratio ($T = 6, c_1 = 1, c_2 = 1$).

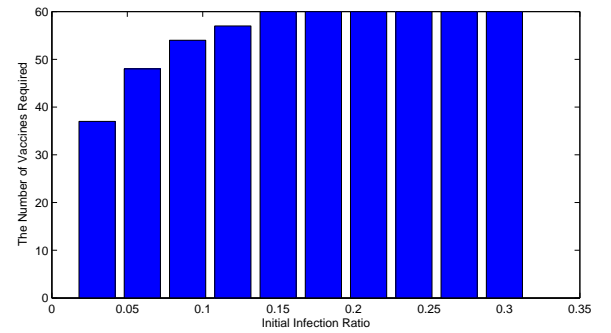


Figure 3. The Number of Vaccines Used vs. Initial Infection Ratio ($T = 6, c_1 = 1, c_2 = 1$).

number of available vaccines for KC and RD. Several observations can be made from Fig. 2. First, we see that the final infection ratio increases as the initial infection ratio increases for all strategies. Second, our proposed method almost prevents all further infections except the initial ones, because the final infection ratio is slightly higher than the initial infection ratio. And third, our proposed strategy and KC achieve a better performance than RD, which is consistent with the reported work in [10, 12]. In Fig. 3, the number of vaccines required first increases as the initial infection ratio increases, and then remains at 60. This fact suggests that a fixed number of available vaccines could suffice to prevent the disease spreading, no matter how big the initial infection ratio is.

Fig. 4 reports the final infection ratio vs. recovery cycle (T) when the initial infection ratio is set at 0.2, $c_1 = 1$ and $c_2 = 1$. The corresponding number of vaccines used by our proposed vaccination strategy is shown in Fig. 5. Several observations drawn from both figures are listed below. First, both the final infection ratio and the number of vaccines required increase as the recovery cycle T increases. This is simply because a person infected with a long recovery cycle tends to infect more people. Second, we observe that compared to other strategies, our strategy achieves a lower final infection ratio under all T s. And third, the performance

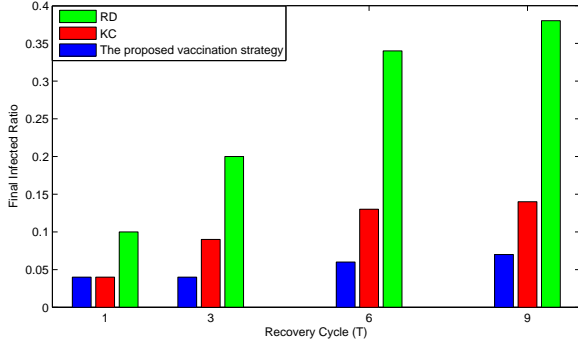


Figure 4. Final Infection Ratio vs. Recover Cycle (The initial infection ratio = 0.03, $c_1 = 1, c_2 = 1$).

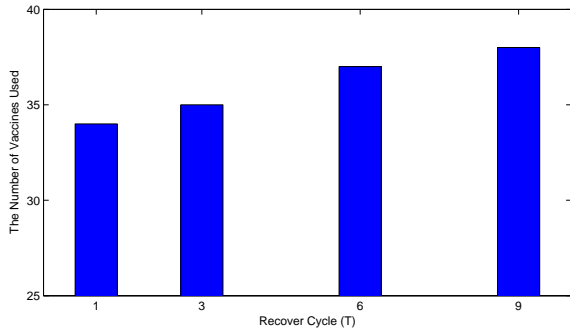


Figure 5. The Number of Vaccines Used vs. Recover Cycle (The initial infection ratio = 0.03, $c_1 = 1, c_2 = 1$).

Table 1. The initial infection ratio and the number of required vaccines under different c_1 and c_2 (The initial infection ratio = 0.03 and $T = 3$).

c_1	c_2	Final Infection Ratio	# Required Vaccines
0.7	0.7	0.07	47
0.9	0.9	0.07	44
1	1	0.07	42
0.9	0.7	0.07	45
0.7	0.9	0.07	47
0.7	1	0.07	48
1	0.7	0.07	43
0.9	1	0.07	42
1	0.9	0.07	44

difference between the proposed method and KC or RD is growing larger with an increasing recovery cycle. This fact shows the final infection ratio based on our proposed strategy is less sensitive to the length of recovery cycle than the other two strategies.

Table 1 reports the impact of c_1, c_2 on the performance of the proposed vaccination strategy. We choose c_1, c_2 from $\{0.7, 0.9, 1\}$ independently. The recovery cycle is fixed at 3, the initial infection ratio is set at 0.03. We observe that the number of vaccines used by the proposed strategy changes

along with c_1, c_2 with a maximum number of vaccines 48 and a minimum number of vaccines 42. These changes are due to the fact that both c_1 and c_2 are used to find the candidates for vaccination. This suggests the final infection ratio could be minimized by using suitable values of c_1 and c_2 , therefore a better performance be obtained.

7 Conclusion

In this paper, we propose a two-level target vaccination strategy based on mobile phone records. It is used to prevent diseases from spreading through the entire population by vaccinating a small group of individuals. Simulation results on real world data show that our proposed strategy is superior to other methods used for comparisons in several ways. It reduces the number of infections dramatically. The number of final infections after execution of our strategy is less sensitive to the recovery cycle. And with our strategy, the same number of vaccines can be used effectively for a range of large values of the initial infection ratio. Our future work includes detailed studies on computing the influences among individuals and deriving variations of the proposed strategy.

8 References

- [1] D. N. Breslauer, R. N. Maamari, N. A. Switz, W. A. Lam, and D. A. Fletcher. Mobile phone based clinical microscopy for global health applications. *PLoS ONE*, 4(7):e6320, 2009.
- [2] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, 2008.
- [3] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315:972–976, 2007.
- [4] R. Istepanian, E. Jovanov, and Y. Zhang. Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity. *Information Technology in Biomedicine, IEEE Transactions on*, 8(4):405–414, 2004.
- [5] R. S. H. Istepanian, S. Laxminarayan, and C. S. Pattichis. *M-Health: emerging mobile health systems*. Topics in Biomedical Engineering International Book Series. Springer, Dordrecht, 2006.
- [6] C. Lee, F. Reid, A. McDaid, and N. Hurley. Detecting highly overlapping community structure by greedy clique expansion. *KDD SNA 2010*, 10:10, 2010.
- [7] M. S. C. Lim, J. S. Hocking, M. E. Hellard, and C. K. Aitken. Sms sti: a review of the uses of mobile phone text messaging in sexual health. *International Journal of STD AIDS*, 19(5):287–90, 2008.
- [8] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.
- [9] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai. Overlapping communities in dynamic networks: their detection and mobile applications. In *Proceedings of the 17th annual international conference on Mobile computing and networking, MobiCom '11*, pages 85–96, 2011.
- [10] Y. Ren, J. Yang, M. C. Chuah, and Y. Chen. Mobile phone enabled social community extraction for controlling of disease propagation in healthcare. In *IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS 2011)*, 2011.
- [11] A. T. van Halteren, R. G. A. Bults, K. E. Wac, D. Konstantas, I. A. Widya, N. T. Dokovski, G. T. Koprnikov, V. M. Jones, and R. Herzog. *The Journal on Information Technology in Healthcare*, 2(5):365–373, 2004.
- [12] P. Wang, M. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding the spreading patterns of mobile phone viruses. *Science*, 324:1071–1076, 2009.
- [13] Z. Zhu, G. Cao, S. Zhu, S. Ranjan, and A. Nucci. A social network based patching scheme for worm containment in cellular networks. In *INFOCOM 2009, IEEE*, pages 1476–1484, 2009.