

University of Massachusetts Medical School

eScholarship@UMMS

---

Population and Quantitative Health Sciences  
Publications

Population and Quantitative Health Sciences

---

2015-02-09

## Cluster-based Epidemic Control Through Smartphone-based Body Area Networks

Zhaoyang Zhang

*University of Massachusetts - Dartmouth*

*Et al.*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/qhs\\_pp](https://escholarship.umassmed.edu/qhs_pp)



Part of the [Digital Communications and Networking Commons](#), [Epidemiology Commons](#), [Health Information Technology Commons](#), [Systems and Communications Commons](#), and the [Theory and Algorithms Commons](#)

---

### Repository Citation

Zhang Z, Wang H, Wang C, Fang H(. (2015). Cluster-based Epidemic Control Through Smartphone-based Body Area Networks. Population and Quantitative Health Sciences Publications. <https://doi.org/10.1109/TPDS.2014.2313331>. Retrieved from [https://escholarship.umassmed.edu/qhs\\_pp/1127](https://escholarship.umassmed.edu/qhs_pp/1127)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Population and Quantitative Health Sciences Publications by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

# Cluster-based Epidemic Control Through Smartphone-based Body Area Networks

Zhaoyang Zhang\*, Honggang Wang\*, Chonggang Wang<sup>‡</sup>, Hua Fang<sup>¶</sup>

\*University of Massachusetts, Dartmouth, MA, USA

<sup>‡</sup> InterDigital Communications, USA

<sup>¶</sup>University of Massachusetts Medical School, Worcester, MA, USA

**Abstract**—Increasing population density, closer social contact and interactions make epidemic control difficult. Traditional offline epidemic control methods (e.g., using medical survey or medical records) or model-based approach are not effective due to its inability to gather health data and social contact information simultaneously or impractical statistical assumption about the dynamics of social contact networks, respectively. In addition, it is challenging to find optimal sets of people to be quarantined to contain the spread of epidemics for large populations due to high computational complexity. Unlike these approaches, in this paper, a novel cluster-based epidemic control scheme is proposed based on Smartphone-based body area networks. The proposed scheme divides the populations into multiple clusters based on their physical location and social contact information. The proposed control schemes are applied within the cluster or between clusters. Further, we develop a computational efficient approach called UGP to enable an effective cluster-based quarantine strategy using graph theory for large scale networks (i.e., populations). The effectiveness of the proposed methods is demonstrated through both simulations and experiments on real social contact networks.

## I. INTRODUCTION

**E**PIDEMIC control becomes more challenging nowadays due to the increasing population density, closer social contact and interactions. The recent epidemic outbreaks have caused a lot of fatalities. For example, about 774 people died due to the outbreak of SARS in 2003, and more than 8870 people died due to the outbreak of H1N1 influenza in 2009, and at least 345 people died due to H5N1 since 2003 [1].

Traditional offline epidemic control strategies typically include both offline control and model-based approaches, which can quarantine or immunize people to contain the epidemic spreading. However, these approaches lack the effectiveness due to either the time-delayed data collection (e.g, using medical survey or medical records) or impractical statistic assumptions about the dynamics of social contact networks. As a type of social networks, social contact network [2] focuses more on physical social contact than online social relationships. To address the effectiveness issue, we proposed a WBAN-smartphone based data collection framework to collect people’s health data and social contact information simultaneously in our previous work [3], [4]. The architecture of this proposed information collection framework is illustrated in Fig. 1. It is integrated by body mounted body sensor networks and smartphones. Note that sensors can also sense surrounding environmental information. In the proposed system, vital signs

of human body such as body temperature, blood pressures, could be collected by the body sensors.

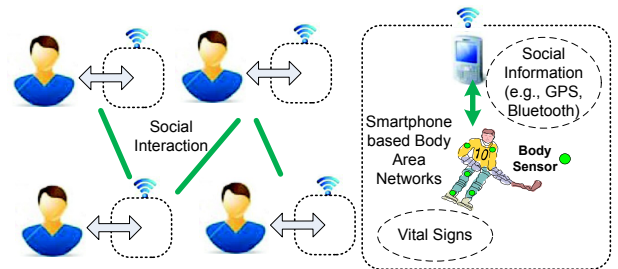


Fig. 1: Vital signs and social interaction information collection

The data collection works as follows: 1) smartphones collect social information. Such information mainly includes physical social contact information (e.g., the social interaction frequency and time interval) and social pattern. 2) Body Sensors collect vital signs such as blood pressure, heart rate and body temperature. 3) smartphones also assist in sensing body behavior or collecting gesture information. 4) smartphones act as gateways for body sensor networks. 5) All sensed information will be delivered to smartphones and then will be forwarded to the remote control center.

Further, as the social information and health information are available, it is challenging to fuse these information together when they are from different information sources. In this paper, we propose to utilize network graphs to represent the data and fuse them together. Network graphs are widely used to represent relations between interacting actors or nodes. They can be used to describe the behavior of epidemics. The edges in the network graph can be used to represent the presence or strength of a relationship between two nodes. Nodes can represent people and the color of nodes can represent health status (i.e., infected or not) of people. Then based on the network graph, we can predict the evolutions of epidemic spread and find optimal sets of people to be removed to contain the spread of epidemics. However, Using the proof similar to [5], finding the optimal sets in a graph has been proved to be a NP-hard problem for large scale networks consisting of smartphones and WBANs. An approximation method with high effectiveness is to divide the populations into multiple clusters, and then realize the control strategy within the clusters in which both inter- and intra- critical node identification (CNI) algorithms can be applied.

In the literature, some approximation algorithms for finding optimal sets such as high degree set targeting and dominating set targeting, have been studied for epidemic control. Unlike these approaches, we focus on applying graph partitioning to find the optimal sets. We also show that the graph partitioning based approach outperforms these existing approaches. In graph theory, there are some existing typical criteria for defining good graph clusters. Spectral algorithms are typically used to discover these clusters where the eigenvector of a matrix related to the adjacency matrix can be used to find a good cut for subgraphs. In this paper, based on the graph theory, a new approach called unequal graph partitioning (UGP) method is introduced to find out the minimal separator of clusters.

The main contributions of this paper can be summarized as follows.

- 1) Develop a new approach to group the populations into clusters for epidemic control;
- 2) Develop effective inter- and intra- cluster based epidemic control strategies based on real social contact networks;
- 3) Provide a graph partitioning based approach for large scale networks (i.e., populations) with reduced computational complexity.

The rest of the paper is organized as follows. Section II presents the literature review; Section III introduces the system design; Section IV shows performance evaluations based on both simulations and real dataset testing; The paper is concluded in Section V.

## II. RELATED WORK

Cluster based epidemic control requires effective clustering approach. The clustering methods for network graph have been studied in the literature. The authors in [6] proposed a criterion to combine internal density with external sparsity for social network clustering. In the paper [7], the authors proposed a DP-cluster algorithm based on the distance-preserving subgraphs, which finds the proper distance preserving subgraphs and partitions a graph into an arbitrary number of distance-preserving subgraphs. However, these methods do not consider node's location and their attributes at the same time and thus are not applicable to the epidemic control applications. In the paper [8], [9], the author studied a latent position cluster model (LPCM) for social network clustering method. LPCM clusters the social network based on nodes's locations and the probability of a social contact between two nodes, but it ignores the attributes of nodes such as gender, age, occupation and others that are related with epidemics. In this paper, we study a clustering approach that can consider both social relations and divergence (i.e., infection state) of nodes.

On the other hand, the spread of epidemic disease has been investigated for many years. The modelings of epidemics are used to describe the mechanisms of disease spreading and predict the spread of the epidemic and evaluate the epidemic control strategies. An SIR model is widely used to characterize the epidemic spread for a fixed population. The compartments used in the SIR model are susceptible, infected and recovered. Each individual is at one of the three compartments at a time moment [10]. The author in [11] models the epidemic and

gives the threshold of the epidemics. However, it is designed based on the homogeneous population and does not consider the diversity of individuals.

The spread of epidemic on social networks has been studied in [12]–[14]. The authors in these papers study the spread of epidemic in social networks, but they ignore the fact that the epidemic spreads through physical contact rather than traditional social relationship represented by the social networks such as Facebook and Twitter. Therefore, these model-based approaches have too idea statistical assumption about the dynamics of social contact networks, and thus might not provide effective epidemic control.

Graph partitioning is an NP hard problem for numerous applications. Some algorithms have been proposed to divide the graph into several small clusters. The “Chaco” scheme [15] allows applications to find small edge separators. The authors in [16] proposed a multilevel algorithm to partition graphs which are approximated by a sequence of increasingly smaller clusters. In the graphs, the number of links between clusters is minimized. A general framework of graph clustering is proposed in [17], which balances the overlapped clusters and minimizes the sum of computational and communication time. A K-way partitioning algorithm is presented in [18], which reduces the size of the graph by collapsing vertices and edges. The above algorithms can divide a graph into smaller clusters by minimizing the links between clusters. These algorithms could be potentially used for solving the epidemic control problem. In this paper, the algorithm we proposed is to find the smallest separator node set to divide the graph into small clusters, which minimizes the correlations between clusters and thus is suitable for cluster-based epidemic control. It is widely accepted that the most efficient epidemic control strategies are based on high-degree targeted strategies [19]–[21]. By quarantining or immunizing the high-degree nodes, the network breaks down to smaller clusters, and then the epidemic spreading can be contained. To further improve the performance of epidemic control, a novel unequal graph partitioning algorithm called UGP is proposed in the paper. The details of the UGP methods are introduced in the following section.

## III. SYSTEM DESIGN

### A. Framework

As shown in Fig. 2, the framework of the proposed system includes two steps. At the first step, the critical network (i.e., population) is clustered to many small groups based on their locations and social contacts. At the second step, a cluster based epidemic control will be implemented, which includes critical node identification, critical set identification and an unbalanced graph partitioning method. The social sensing detection in this paper is to find physical social contact pattern. Our study is focused on the collections of physical social interactions (e.g., face-to-face interactions) which may cause epidemics spreading. The social information from traditional social networks could be helpful for detecting social ties among people. However, the information is different from the one obtained from the social sensing in the proposed

scheme. In the proposed system, a social contact network graph called *critical network* [3] in our previous work is employed to collect the vital signs and social interaction information. The research in this paper is based on the data collection framework and the focus is on finding effective epidemic control strategies. The proposed algorithm runs on the critical networks which can quickly capture the critical epidemic information in different environments.

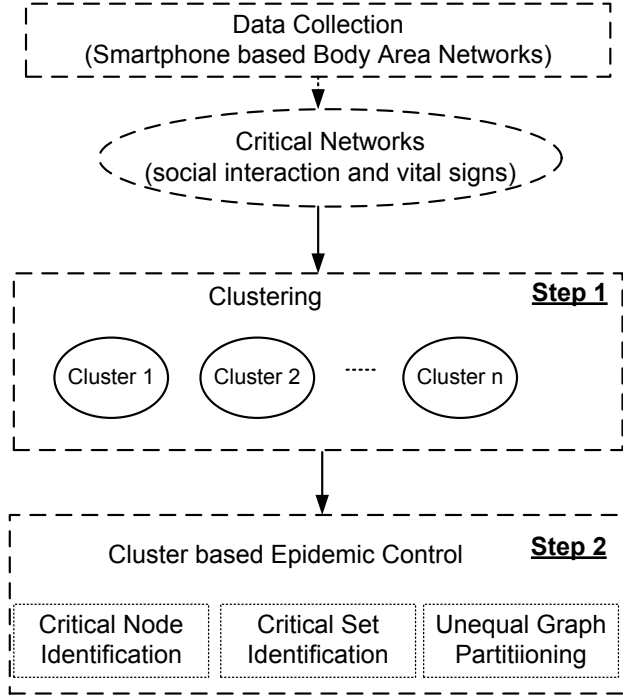


Fig. 2: System framework

The detail of critical network definition can be referred to our previous study in [3]. Basically, the elements of a critical network are nodes, arcs, and a series of planes. A node represents a user of the data collection system. Each node can be white, gray, or black. White nodes denote people who have never been infected and are healthy; Gray nodes denote people who were infected but now have already been recovered; Black nodes denote people who are infected and have not been recovered yet. An arc between two nodes represents the infection relationship; the node pointed by the arc is infected by the one that originates the arc. We also define weights on each arc to indicate how much in terms of possibility the node at the tail of the arc is infected by the one at the head of arc. A plane represents the above information during one time period unit. Planes are also called snapshots of a critical network. The proposed epidemic control strategy includes two major stages as discussed in the following:

1) *The First step: Clustering:* The population represented in critical network is divided into clusters based on their location and attribute similarities. The goal of this division is to partition the graph into  $k$  clusters with short mean distance and homogeneous attribute values. The problem is quite challenging because their location and attribute similarities are two independent objectives we pursued. For the network clustering,

we design a distance function as follows to calculate the distance between two nodes  $v_i$  and  $v_j$

$$d(v_i, v_j) = \phi \cdot d_E(v_i, v_j) + \gamma \cdot d_A(v_i, v_j) \quad (1)$$

in which  $d_E(v_i, v_j)$  is Euclidean distance between vertices  $v_i$  and  $v_j$  and  $d_A(v_i, v_j)$  is attribute similarity (e.g., vital signs and infected status) between them. If the coordinates of  $v_i$  and  $v_j$  are  $(x_1, y_1)$  and  $(x_2, y_2)$ , respectively, the Euclidean distance is  $d_E = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ . Suppose that each vertex has an attribute set  $A = \{A_1, A_2, \dots, A_m\}$  with values  $\{a_1, a_2, \dots, a_m\}$ , e.g., the values of attribute set  $\{A_1, A_2, \dots, A_m\}$  for nodes  $v_i$  are  $a_{1i}, a_{2i}, \dots, a_{mi}$ , then the attribute similarity between vertices  $v_i$  and  $v_j$  is calculated by  $d_A = \sqrt{(a_{1i} - a_{1j})^2 + (a_{2i} - a_{2j})^2 + \dots + (a_{mi} - a_{mj})^2}$ .

The probability that a social contact (edge) exists between nodes  $v_i$  and  $v_j$  is modeled by

$$P = \text{logit}(p(e_{i,j} = 1 | d_{i,j})) \quad (2)$$

where the *logit* of a probability  $p$  is  $\text{logit}(p) = \log(\frac{p}{1-p})$ . The model accounts both observed attributes and Euclidean distance between vertices  $v_i$  and  $v_j$ . To identify the scale of the positions  $z_i$  and weight  $\phi$  and  $\theta$ , the positions are restricted to have unit root mean square

$$\sqrt{\left(\frac{1}{n} \sum_i^N |z_i|^2\right)} = 1 \quad (3)$$

where  $N$  is the number of nodes in the critical networks (i.e., population). We assume that the population could be divided into  $M$  clusters (groups), and each cluster satisfies a multivariate normal distribution with a mean vector and spherical covariance matrix which differ between clusters. Then

$$z_i \sim \sum_{l=1}^M \lambda_l \mathcal{N}_d(\mu_l, \Sigma_l) \quad (4)$$

where  $\lambda_m$  is the probability that a vertex  $v_i$  belongs to the  $l$ th cluster,  $\lambda_m \geq 0$  and  $\sum_{l=1}^M \lambda_l = 1$ .  $\mathcal{N}_d$  is multivariate normal distribution of a  $d$ -dimensional random vector.  $\mu_l$  and  $\Sigma = \sigma_l^2 Q_d$  are the mean vector and covariance matrix for  $l$ th cluster, respectively.  $Q_d$  is the  $d \times d$  identity matrix. Based on the distance function and probability function, we can divide the nodes of social contact networks (i.e., populations) into multiple clusters.

2) *Epidemic Spreading:* Let  $S(t)$ ,  $I(t)$  and  $R(t)$  respectively denote the number of susceptible nodes, infective nodes and recovered nodes at the time  $t$ , and suppose that at time  $t = 0$  these numbers are given by  $S(0) = n - m$ ,  $I(0) = m$  and  $R(0) = 0$ . We use the a typical  $q$ -influence model to describe the epidemic spreading. It has been often used to simulate the spread of computer virus in computer network [22] and spread of epidemic diseases in social network [4]. In a  $q$ -influence model, a susceptible node could be infected by one of its neighbors with a fixed probability of  $q$ . In the model, the influences from its neighbors are independent. Suppose that a susceptible node  $v_i$  has  $n$  neighbors and  $m$  of them are

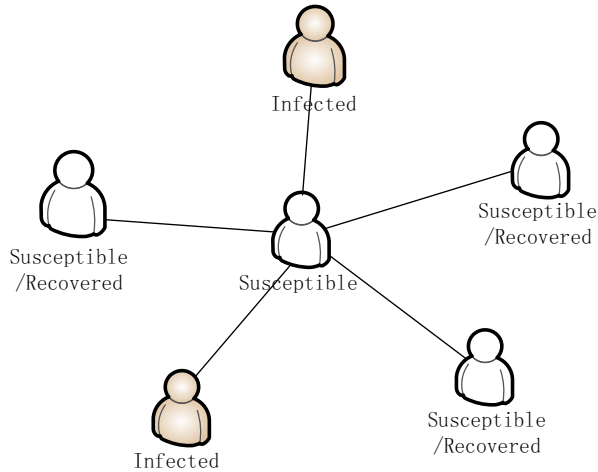


Fig. 3: An case for epidemic spreading

infective nodes, the probability that the node will be infected by its infective neighbors is:

$$p = 1 - (1 - q)^m \quad (5)$$

As shown in 3, a susceptible individual has five neighbors whom the individual has contact with, and two of them are infective nodes. If the influence parameter  $q = 0.1$ , the probability that the node will be infected by its infective neighbors is  $p = 1 - (1 - 0.1)^2 = 0.19$ . The dynamics of the model are defined as follows. An infectious individual infects a susceptible person through social contacts. People who were infected start spreading the disease according to the same rule. Infected individual remains infectious for a fixed time period and then they stop being infectious, recover and become immune to the diseases. The epidemic starts at time  $t = 0$ . As the epidemic evolves, new individuals get infected, then recover, and there will be no infective nodes in the community after time  $T$ . When no further individuals can get infected, the epidemic spread stops. The final state of the epidemic is described by the ultimate infected number of nodes  $R(T)$ . It is worthy to notice that there are different infection models such as  $m$ -threshold influence model and majority rule influence model given in [23]. We chose  $q$ -influence because of its simplicity and being widely adopted. In addition, the epidemic spreading could highly depend on the initial location and number of epidemic sources.

#### IV. THE SECOND STEP: CLUSTER-BASED EPIDEMIC CONTROL

The proposed cluster-based epidemic control includes two tier control: intra-cluster and inter-cluster control.

##### A. Intra-cluster epidemic control

The goal of intra-cluster control is to remove minimal number of nodes in a cluster so that the epidemic spread within the cluster can be effectively contained. Intra-cluster critical nodes identification algorithm is designed based on the  $q$ -influence model. It will be easy to extend our model

to other typical epidemic models. The focus of the algorithm is to find a set of critical nodes to be removed to minimize the overall number of infected nodes within a specified time period  $T$ .

Logically, if the infection possibility for each individual is available, it is highly possible to deduce whom will be more (or less) likely to be infected in some future time. First, the number of infected nodes at the next time moment can be predicted by using the epidemic spread model. As long as the social contact information and vital signs (i.e., infected status) are available, the potential infected nodes at the next time moment are predictable.

The proposed intra-cluster critical nodes identification algorithm works as follows. (i) The inputs of the algorithm are current time  $t$ , network graph  $G_t$ , susceptible nodes  $S(t)$ , infective nodes  $I(t)$ , recovered nodes  $R(t)$ , and the isolation (i.e., immunization or treatment) capability  $\beta$ .  $n$  is the number of individuals in the network and  $k$  is the number of infected individuals. Then, we initialize the variables to 0 ( $i, j = 0$ ). (ii) Remove one infected node from the network and predict the epidemic spread at the next time  $I_0(t+1)$ , and then remove the infected nodes one by one to get a set of predicted infective nodes at the next time  $\{I_0(t+1), I_1(t+1), \dots, I_k(t+1)\}$ . In the process, a critical node  $s$  is the one that leads to minimized number of infective nodes after it is removed. (iii) Put the critical nodes into the critical nodes set  $\Gamma$  and increase the count of critical nodes by 1  $j = j + 1$ . (iv) Check whether the algorithm has found all critical nodes. If  $j < \beta$ , set  $i = 0$  and the algorithm goes to (ii). Otherwise, the algorithm outputs the critical nodes set  $\Gamma$  and exits.

1) *Inter-Cluster Epidemic Control*: The major idea for inter-cluster control is to identify the critical clusters and to remove the links among clusters. Under the limited medical resources, we want to identify the most significant clusters to be isolated so that the spread of epidemics can be effectively contained. The inter-cluster method is similar to the intra-cluster epidemic control method. In this method, each cluster is viewed as a node. Comparing to the critical nodes identification algorithm above, the critical clusters identification algorithm contains a set of big “nodes”.  $Z_m, (m = 1, 2, \dots, M)$ .

$$Z_m = \frac{1}{N_m} \sum_{i=1}^{N_m} z_i \quad (6)$$

where  $N_m$  denotes the number of nodes in  $m$ th cluster. If node  $v_i$  is in  $m$ th cluster and node  $v_j$  is in  $n$ th cluster, the probability that a social contact exist between vertices  $v_i$  and  $v_j$  is  $P(e_{i,j} = 1 | d_{i,j})$ . Suppose that the  $m$ th cluster  $C_m$  and the  $n$ th cluster consist of  $N_m$  and  $N_n$  nodes, respectively, then the probability that a social contact exists between clusters  $C_m$  and  $C_n$  is defined to be the mean value of all edge probabilities between the two clusters. The inter-cluster critical node identification algorithm is described in Fig. 4.

$$P(E_{m,n} = 1 | D_{m,n}) = \frac{1}{N_m N_n} \sum_i^{N_m} \sum_j^{N_n} P(e_{i,j} = 1 | d_{i,j}) \quad (7)$$

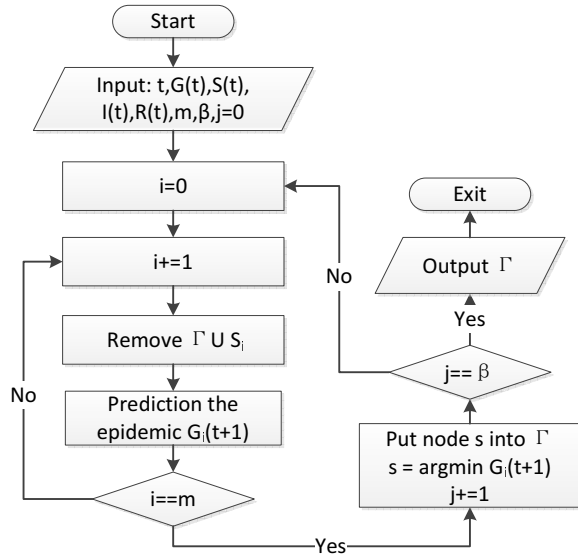


Fig. 4: Critical clusters identification algorithm

where  $E_{m,n}$  and  $D_{m,n}$  are the edge and distance between the  $m$  and the  $n$ th clusters, respectively.

### B. Privacy Issue and Missing Information

The privacy issue is critical for data collection. For example, some people carrying body sensor networks may not want to share their vital sign information with others due to their privacy concerns. Their privacy concerns may restrict them from taking advantage of the full benefits from the proposed system. There are many arguments for and against various forms using privacy control of BANs and smartphones. The security and privacy protection for saved data or data transmission is an unresolved concern for BANs and/or smartphones [24]. There have been ongoing research works such as [24], [25] that attempt to address the concern using access control, authentication, cryptography and etc.. The challenges for solving this issue not only come from the development of more innovative techniques, but also from the enforcement of public regulations. We argued that effective privacy control for BANs highly depends on the specific medical applications. For emergent epidemic outbreaks, privacy would become less concerned, as being alive and not infected are more important.

In addition, the information of some nodes may not available due to several reasons, such as privacy concerns, failure of the device, and even people who are infected turn off the BANs to escape from quarantining. Thus, it is important to accurately estimate the status of the unknown nodes. We proposed to use the social contact information of a node's neighbor to estimate its health status. As show in the Fig. 5, the information of two nodes, A and B, are lost. We need to estimate the status of the two nodes to fully understand the spread of endemic disease. It is easy to use an N-threshold voting method to classify the center node. By the N-threshold, a node is infected when last least N of its neighbors are infected. Suppose we set the threshold  $N = 3$ , then node A is estimated not infected and B is infected, because A only has two infected neighbors but B has four infected neighbors. Although the N-threshold voting

method works well, it is sensitive to the number of neighbors. In the proposed social contact network, we use a local majority estimation method, by which an unknown node is infected if over  $\eta$  fraction of its neighbors are infected. Compared with the above N-threshold voting method, the threshold number of nodes in majority varies automatically according to the number of neighbors the unknown nodes have. If we set majority fraction  $\eta = 0.5$ , both the node A and B in Fig. 5 are estimated to be infected.

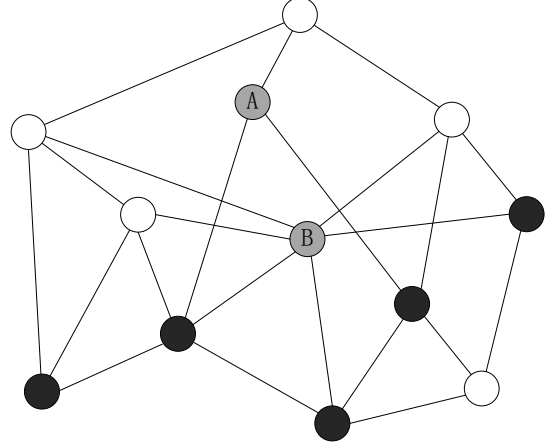


Fig. 5: Local majority estimation for missing information

### C. Improved Control via Unequal Graph Partitioning

Running critical node/cluster identification algorithm for large scale networks (i.e., population) is not realistic because it is a NP-hard problem to find the best strategy to quarantine or immunize (receiving vaccine) a group with a minimal number of people. The proposed cluster based control algorithm has high computation complexity issues for large scale networks. We further proposed to use graph theory to address the issue. In order to find a optimal immunization set, an unequal graph partitioning method is proposed in this work.

Assume in a weighted graph  $G$  (i.e., a critical network), the nodes in  $G$  are a set of nodes to be grouped into  $k$  clusters. The most direct way to construct a partition is to solve the min-cut problem [26]. However, mincut always causes imbalance graph partitioning (e.g. one partition only contains one vertex), and we need to guarantee that sets  $A_1, \dots, A_k$  are "reasonably large". The common objective function which can implement this requirement is the normalized cut:  $\sum_{i=1}^k \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}$ . while the size is measured by the weights of its edge  $vol(A_i)$ . A well-known solution to such a problem is obtained by computing the top  $k$  eigenvectors of the Laplacian matrix [26]. However, the partitioning method itself cannot solve the optimal set problem for epidemic control. Therefore, we proposed an algorithm that combines the graph partitioning and optimal searching algorithm of removal nodes. The algorithm is described in Algorithm 2 (including Algorithm 1), which can identify minimized set of nodes to be removed so that the spread of epidemics can be effectively contained.

The above method uses both social contact information and vital signs to optimize the epidemic control. However, in some

---

**Algorithm 1** 2-way Partitioning Algorithm

---

**Require:** Graph  $G(V, E)$ , Target fraction  $\eta \geq 0.5$ **Ensure:** Subgraph  $G_1, G_2$ , Separator  $S$ 

- 1: Generate the connected component  $G'$  of graph  $G$
  - 2: Randomly assign nodes in  $G'$  into two clusters  $G_1$  and  $G_2$ , the size ratio of  $G_1$  and  $G_2$  is approximately equal to  $\eta$
  - 3: Move all nodes in  $G_1$  ( $G_2$ ) that are connected to  $G_2$  ( $G_1$ ) to a third cluster  $S$
  - 4: **repeat**
  - 5:   Swap a node in  $S$  with that in  $G_1$  or  $G_2$
  - 6:   **if** the size ratio of  $G_1$  and  $G_2$  is approximately equal to  $\eta$  **and** The size of  $S$  reduces **and** there are no connection between  $G_1$  and  $G_2$  **then**
  - 7:     The swap is accepted
  - 8:   **end if**
  - 9: **until** No further swap can reduce the size of  $S$
- 

---

**Algorithm 2** Unequal Graph Partitioning Algorithm

---

**Require:** Graph  $G(V, E)$ , Target fraction  $\eta$ **Ensure:** Separator  $S$ 

- 1: **if**  $\eta \geq 0.5$  **then**
  - 2:   **call** Algorithm 1 with  $G(V, E), \eta$
  - 3: **else**
  - 4:   **call** Algorithm 1 with  $G(V, E), \eta = 0.5$
  - 5:   **repeat**
  - 6:     **if** Size ratio of  $G_i, i = 1, 2$  is bigger than  $\eta$  **then**
  - 7:       **call** Algorithm 1 with  $G_i, \eta = 0.5$
  - 8:     **end if**
  - 9:   **until** the size ratios of all subgraphs are smaller than  $\eta$
  - 10:   the separator  $S$  is the union of all the generated separators
  - 11: **end if**
  - 12: **return** Separator  $S$
- 

scenarios, either the social contact information or vital signs may not be available due to some reasons such as privacy concern. As a result, the incomplete health and social data resulting from the potential unavailability of WBANs and mobile phones (e.g., due to their limited deployment) makes the prediction more challenging. To address the issue, we could consider some estimation algorithms such as majority voting algorithm, which enable the estimation of the missing health information (i.e., infectious state). It could be also possible for us to predict the epidemic spreading through a certain portion of data collected from a relatively small group of monitored people. However, the accuracy of the prediction algorithms will significantly be impacted by the level of missing data. In addition, to deal with the privacy issue during the data collections, the identity of each individual people could be mapped to a processed ID through one-way encryption algorithm and all the identity information will be deleted after the mapping process. There are no ways to covert the processed ID back to each individual's identity. In the proposed scheme, the identity information is not needed for

the epidemic predictions.

## V. PERFORMANCE EVALUATION

## A. Epidemic spreading on a scale-free network

A scale-free graph is a graph whose node degree follows a power law distribution. In a scale-free graph, the fraction of the nodes that have  $k$  neighbors, denoted by  $P(k)$ , is proportional to  $k^{-\alpha}$  for large values of  $k$ , or

$$P(k) \propto k^{-\alpha} \quad (8)$$

where  $\alpha$  is a positive constant with typical value in range of  $2 < \alpha < 3$ . Each node in  $G(V, E)$  presents a person and a node can influence the other nodes if there are connections, denoted by edges  $e$ , between them. For example, if node  $i$  is infected, it may infect its neighbors. Obviously, one can have different infection models such as q-influence model, m-threshold influence model and majority rule influence model given in [27]. Our study is focused on the statistical properties of the scale-free random social networks. First, we construct a local mean field (LMF) of an arbitrary node in  $G$ . LMF is a transformation of  $G$  and we can use it to model the correlation structure on local neighborhoods. Also, LMF provides an asymptotic behavior when the number of nodes of a sparse random graph goes to infinity with a given asymptotic degree distribution  $P_0(k)$  [28].

To construct LMF of  $G$ , we randomly choose a node,  $r \in V$ , as the root node of the local mean field. Because  $r$  is randomly chosen, it should have  $deg(r)$  neighbors,  $[v_1, v_2, \dots, v_{deg(r)}]$ , where  $deg(r)$  follows the power law distribution,

$$\text{Prob}(deg(r) = k) = Ck^{-\alpha} \quad (9)$$

where  $C$  is a constant and  $C \sum_{k=1}^{\infty} k^{-\alpha} = 1$ . Then, for an infinite random power law graph, the probability that node has degree  $k$  is

$$P_1(k) = \frac{k^{1-\alpha}}{\zeta(\alpha - 1)} \quad (10)$$

where  $\zeta(x) = 1/C = \sum_{k=1}^{\infty} k^{-\alpha}$  is the Riemann zeta function. Thus, the degree distribution of the descendants of node  $r$  follows a shifted power law distribution  $P_1(k)$ . Then the local mean field of  $G$  is completely determined. It is recursive and free of loops. In the following, we will use the model to study the epidemic spread.

We first build a social contact network graph  $G$  with  $N$  nodes and  $E$  edges using the above social network modeling. Then we assume that  $N_t$  nodes are infected at the beginning time  $t = 0$ . If we assume that the nodes have no self-recovery capability (the recovery period time  $T_{rec}$  is bigger than the sample duration), and the system has limited immunization capability. The parameters used in the simulation are shown in Table I.

Fig. 6 and Fig. 7 show the epidemic control performance when the self-recovery (i.e., immunized) is considered or not. In Fig. 6, we use symbol "[C1, C2, C3, C4]" to indicate the number of nodes to be removed in cluster C1, C2, C3, C4. Both figures show that removing the nodes from different sets could significantly impact the performance of epidemic control. In Fig. 6, removing nodes in the setting of "[2,2,0,0]"

TABLE I: Notations and Parameters

Symbol	Notations	value
$G$	Social contact network graph	-
$N$	Total number of participant nodes in $G$	484
$E$	Number of edges between nodes in $G$	1440
$N_0$	Number of initial infected nodes	1
$N_t$	Number of infected nodes at time $t$	-
$T$	Simulated epidemic spreading time	35
$S, I, R$	Susceptible, Infected and Recover states	-
$T_{rec}$	Recovery period	x
$q$	Influence probability	0.1
$\theta$	Infection threshold	0.5
$\beta$	Number of removed nodes at each turn	1
$\eta$	Majority fraction of a node	0.5

can achieve the best epidemic control performance. It indicates the nodes in C1 and C2 clusters have more significant impact on the epidemic spreading. Therefore, it is critical to determine which group of nodes should be removed if the number of removal nodes is constrained. Even with the consideration of epidemic recovery capability, the setting of [2,2,0,0] still achieve better epidemic control performance.

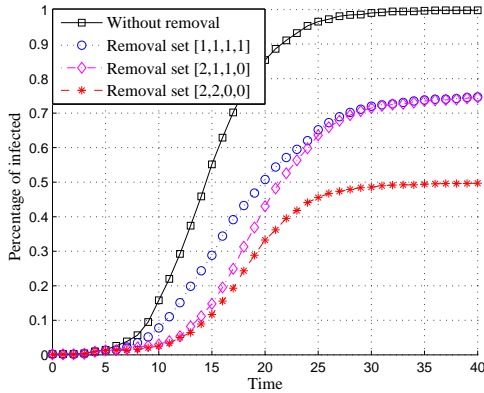


Fig. 6: Epidemic control when self-recovery is not considered

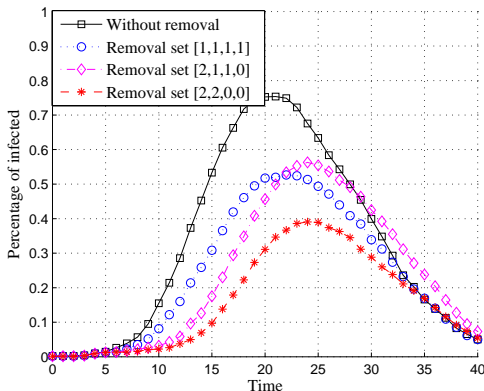


Fig. 7: Epidemic control when self-recovery is considered

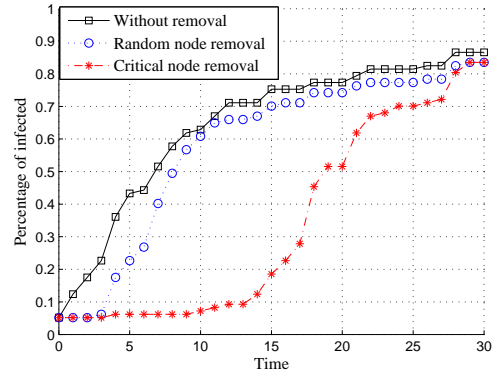


Fig. 9: Epidemic control using MIT Reality dataset

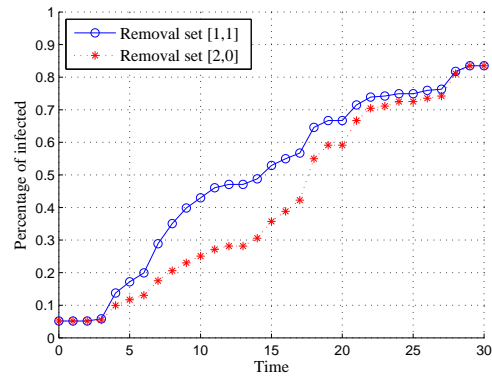


Fig. 10: Cluster-based epidemic control using MIT Reality dataset

### B. A Case Study based on MIT Reality Mining Dataset

We further evaluate the performance of the proposed algorithms based on the real contact network dataset. The Reality Mining project was conducted from 2004-2005 at the MIT Media Laboratory. The Reality Mining study following ninety-four subjects using mobile phones pre-installed with several pieces of software that recorded and sent the researcher data about call logs, Bluetooth devices in the proximity of approximately five meters and other context information. In the Reality Mining project, when a Bluetooth device conducts a discovery scan, other Bluetooth devices within a range of 5-10 meters respond with their unique infirmations, such as user defined name, the device type, and MAC address. When a subject's MAC address is discovered by a periodic Bluetooth scan performed by another subject, it indicates that the two devices are 5-10 meters close to each other. Based on the gathered information about the nearby devices, we could build a dynamic contact network. As shown in Fig. 9, at each time unit (i.e., day), there is a contact graph which may be significant different from others. Then we assume that an epidemic disease outbreaks at the beginning and 5 nodes are initiated to be infected. Fig 9 and Fig 9 shows the spread of the epidemic on the MIT reality's dynamic contact network, which demonstrated the effectiveness of our approaches.



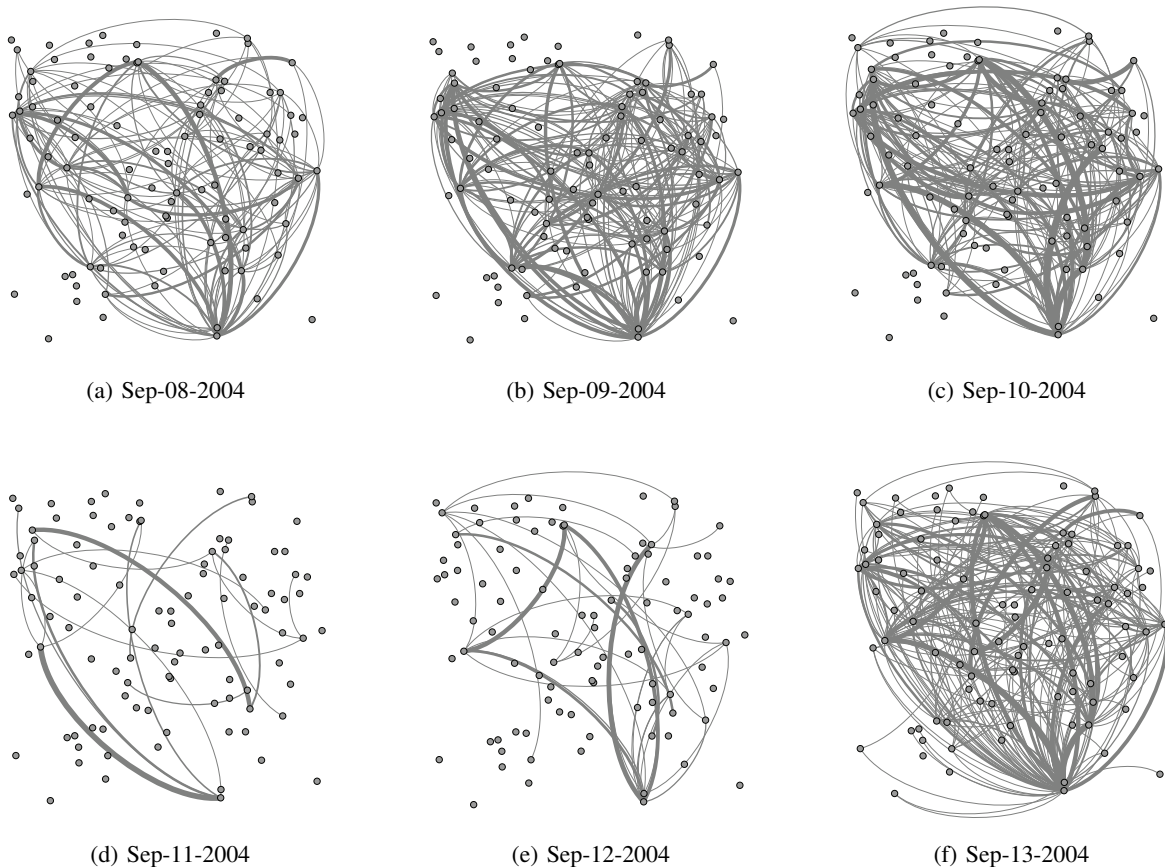


Fig. 8: Contact networks for six consecutive days.

### C. Performance evaluation of UGP

We test the effectiveness of the UGP strategy on scale free network models. Based on a social network model and  $q$ -influence contagious model, we investigated four algorithms: (a) random selection, i.e., randomly remove (i.e., quarantine or immunize)  $M$  nodes; (b) high degree selection (“High degree” in Fig. 11 and Fig. 12), i.e., remove top- $M$  nodes with higher degrees; (c) dominating set selection (“ $k$ -DS” in Fig. 11 and Fig. 12), i.e., remove  $M$  nodes from  $k$ -dominating set and (d) a heuristic UGP graph partitioning method, i.e., separate a network into two unequal-size clusters with a minimal number of nodes to be removed. All four algorithms separate the whole graph to a certain number of clusters after the selected nodes are removed. As shown in Fig. 11 and Fig. 12, in the worst case, UGP outperforms other three algorithms in terms of maximal infected people by removing the same number of nodes. Fig. 11 and Fig. 12 show the significant benefits of employing graph partition methods for the epidemic control problem. However, the UGP may not be the best algorithm to solve the graph partitioning problem, and the performance of cluster-based epidemic control also depends on optimality of the algorithm of removing edges among clusters.

## VI. CONCLUSION

In this paper, we proposed cluster based epidemic control based on social contact networks (i.e., critical networks) which can collect both vital signs and social interaction information.

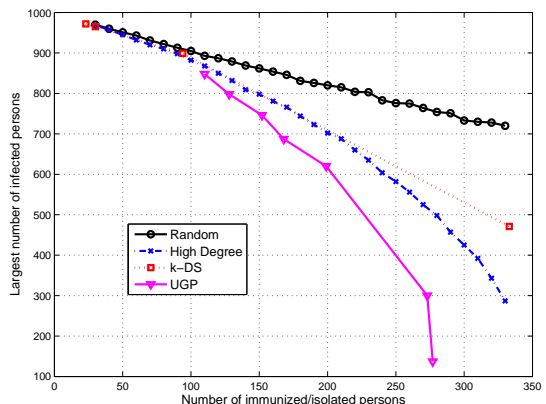


Fig. 11: Performance of UGP on fixed network topology

Unlike either common offline control or traditional model based approaches, our approach is designed based on the real time social contact and health information and can determine optimal number/set of nodes to be removed so that the epidemic spreading can effectively contained. Our approach can work for large scale networks where the optimization problem is an NP-hard problem. The proposed cluster based control approach includes two steps. In the first step, we group the population into multiple clusters. In the second step, critical node/set identification algorithms are applied within or among

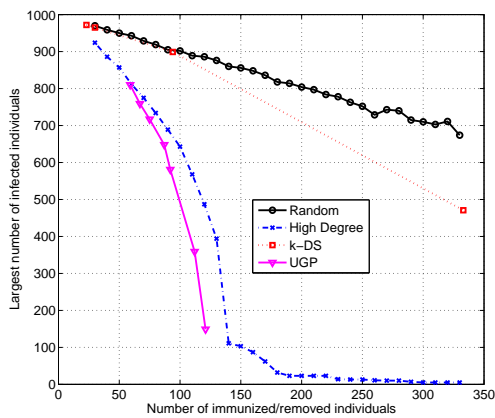


Fig. 12: Performance of UGP on scale free network

clusters. However, the proposed cluster based epidemic control still may have limitations in some application scenarios. As we discussed above, some information of either social contact or vital signs may not be collectable. For example, some people are not willing to carry body sensors or smartphones. We understand that the limitations could exist in the real applications. Some potential statistical estimation algorithms such as missing data estimation algorithms could be applied to mitigate the impact of the unavailability of the information, which would be the part of our future studies. Our major contributions are the innovations of inter- and intra cluster based epidemic control methods and exploring graph theory for epidemic control in large scale networks (i.e., populations). Our research work opens a new vista of epidemic control using smartphone based WBANs.

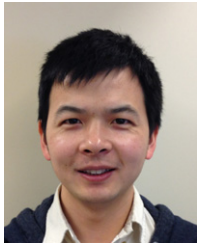
## VII. ACKNOWLEDGEMENTS

The project described was supported by the pilot project program awarded to Dr. Hua Fang from the National Center for Research Resources UL1RR031982 and NIH/ NIDA award 1R01DA033323-01.

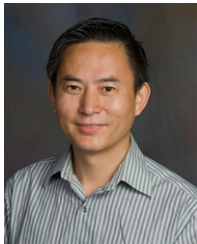
## REFERENCES

- [1] C. Fraser, C. A. Donnelly, S. Cauchemez, and et al., "Pandemic potential of a strain of influenza a (h1n1)," *Science*, vol. 324, no. 5934, pp. 1557–1561, 2009.
- [2] Y. Wang, D. Zeng, Z. Cao, Y. Wang, H. Song, and X. Zheng, "The impact of community structure of social contact network on epidemic outbreak and effectiveness of non-pharmaceutical interventions," *Lecture Notes in Computer Science*, pp. 108–120, 2011.
- [3] Z. Zhang, H. Wang, X. Lin, H. Fang, and D. Xuan, "Effective epidemic control and source tracing through mobile social sensing over wbans," in *Proc. the 32nd IEEE International Conference on Computer Communications (IEEE INFOCOM 2013), mini-conference*, 2013.
- [4] Z. Zhang, K. Lee, H. Wang, D. Xuan, and H. Fang, "Epidemic control based on fused body sensed and social network information," in *32nd International Conference on Distributed Computing Systems Workshops (2012 ICDCSW)*, june 2012, pp. 285–290.
- [5] C. Kuhlman, V. Anil Kumar, M. Marathe, S. Ravi, and D. Rosenkrantz, "Finding critical nodes for inhibiting diffusion of complex contagions in social networks," in *Machine Learning and Knowledge Discovery in Databases*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, vol. 6322, pp. 111–127.
- [6] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan, "Clustering social networks," in *Algorithms and Models for the Web-Graph, 5th International Workshop, WAW 2007, 2007*, pp. 56–67.

- [7] R. Nussbaum, A.-H. Esfahanian, and P.-N. Tan, "Clustering Social Networks Using Distance-Preserving Subgraphs," in *International Conference on Advances in Social Networks Analysis and Mining, 2010*, aug, pp. 380–385.
- [8] M. S. Handcock, A. E. Raftery, and J. M. Tantrum, "Model-based clustering for social networks," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 170, no. 2, pp. 301–354, 2007.
- [9] P. N. Krivitsky and M. S. Handcock, "Fitting latent cluster models for networks with latentnet," *Journal of Statistical Software*, vol. 24, no. 5, pp. 1–23, 5 2008.
- [10] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London Series A*, vol. 115, no. 772, pp. 700–721, 1927.
- [11] T. Britton, "Stochastic epidemic models: A survey," *Mathematical Biosciences*, vol. 225, no. 1, pp. 24–35, 2010.
- [12] M. E. J. Newman, "Spread of epidemic disease on networks," *Physical Review E*, vol. 66, no. 1, p. 16128, 2002.
- [13] J. C. Miller, "Percolation and epidemics in random clustered networks," *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, vol. 80, no. 2 Pt 1, p. 020901, 2009.
- [14] T. Zhou, Z.-Q. Fu, and B.-H. Wang, "Epidemic dynamics on complex networks," *Progress in Natural Science*, vol. 16, no. 5, p. 12, 2005.
- [15] B. Hendrickson and R. Leland, "The chaco users guide version 2.0," Technical Report SAND95-2344, Sandia National Laboratories, Tech. Rep., 1995.
- [16] B. Hendrickson and R. W. Leland, "A multi-level algorithm for partitioning graphs," *SC*, vol. 95, p. 28, 1995.
- [17] A. Pinar, "Partitioning for complex objectives," in *In Proceedings of International Parallel and Distributed Processing Symposium*, 2001, pp. 121–123.
- [18] G. Karypis and V. Kumar, "Multilevel k-way partitioning scheme for irregular graphs," *Journal of Parallel and Distributed computing*, vol. 48, no. 1, pp. 96–129, 1998.
- [19] R. Cohen, S. Havlin, and D. ben Avraham, "Efficient immunization strategies for computer networks and populations," *Physical Review Letters*, vol. 91, no. 24, p. 247901, 2002.
- [20] F. Nian and X. Wang, "Efficient immunization strategies on complex networks," *Journal of Theoretical Biology*, vol. 264, no. 1, pp. 77–83, 2010.
- [21] M. Salath and J. H. Jones, "Dynamics and control of diseases in networks with community structure," *PLoS Computational Biology*, vol. 6, no. 4, p. 11, 2010.
- [22] C. C. Zou, D. Towsley, and W. Gong, "Email worm modeling and defense," in *Proceedings. 13th International Conference on Computer Communications and Networks, 2004. ICCCN 2004.* IEEE, Oct. 2004, pp. 409–414.
- [23] B. Zhao, Y. K. Li, J. C. S. Lui, and D. ming Chiu, "Mathematical modeling of advertisement and influence spread in social networks."
- [24] M. Li, W. Lou, and K. Ren, "Data security and privacy in wireless body area networks," *Wireless Communications, IEEE*, vol. 17, no. 1, pp. 51–58, 2010.
- [25] D. Halperin, T. Kohno, T. Heydt-Benjamin, K. Fu, and W. Maisel, "Security and privacy for implantable medical devices," *Pervasive Computing, IEEE*, vol. 7, no. 1, pp. 30–39, 2008.
- [26] J. Hao and J. Orlin, "A faster algorithm for finding the minimum cut in a directed graph," *Journal of Algorithms*, vol. 17, no. 3, pp. 424 – 446, 1994. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0196677484710431>
- [27] B. Zhao, Y. Li, and J. C. Lui, "Mathematical modeling of advertisement and influence spread in social networks," in *ACM Workshop on the Economics of Networked Systems (NetEcon'09)*, 2009, pp. 90–98.
- [28] M. Lelarge and J. Bolot, "Network externalities and the deployment of security features and protocols in the internet," in *Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, ser. SIGMETRICS '08, 2008, pp. 37–48.

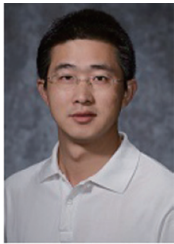


**Zhaoyang Zhang** received the B.S. degree in science and the M. S. degree in electrical engineering from Xidian University, Xian, China, in 2007 and 2010, respectively. He is currently pursuing his Ph.D. degree at the Department of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, MA, USA. His current research interests include wireless healthcare, wireless body area networks, and cyber-physical systems.



**Honggang Wang** received the Ph.D. degree in Computer Engineering at the University of Nebraska-Lincoln in 2009. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Massachusetts Dartmouth, USA.

His research interests include wireless communication and networking, sensor networks, multimedia communications, social networks and wireless healthcare. He has published more than 90 papers in his research areas. He serves as a chair/co-chair for several international conferences and on the editorial board for several journals. He is a Lead Guest Editor of IEEE Journal of Biomedical and Health Informatics (J-BHI) (previous IEEE Transactions on Information Technology in Biomedicine (TITB)) special issue on "Emerging Wireless Body Area Networks (WBANs) for Ubiquitous Healthcare" in 2013.



**Chonggang Wang** received the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), China in 2002. He is currently a Senior Staff Engineer with InterDigital Communications with focuses on Internet of Things (IoT).

He has (co-) authored more than 100 journal/conference articles. He is a senior member of IEEE and on the editorial board for several journals. He is the editor-in-chief of IEEE IoT Journal. He also serves as the vice-chair of IEEE ComSoc Multimedia Technical Committee (MMTC).



**Hua Fang** is an Assistant Professor in the Department of Quantitative Health Sciences' Division of Biostatistics and Health Services Research at the University of Massachusetts Medical School. Dr. Fangs research interests include computational statistics, research design, statistical modeling and analyses in clinical and translational research. She is interested in developing novel methods and applying emerging robust techniques to enable and improve the studies that can have broad impact on the treatment or prevention of human disease.