Containment Policies for Transmissible Diseases

Shirish Tatikonda, Sameep Mehta, and Srinivasan Parthasarathy*

Department of Computer Science and Engineering,
The Ohio State University, Columbus, OH 43210, USA

1. ABSTRACT

In this paper, we propose several strategies to curb the transmissible diseases from spreading. Our policies identifies individuals and locations which play a vital role in spreading the disease. Our analysis is based on simulation data generated by Episims system. We model this data as People-People Contact Network and People-Locations Activity Graph. We also evaluate our proposed strategies under different practical resource constraints.

2. INTRODUCTION

Outbreak of highly pathogenic $H5N1$ avian influenza or bird flu was first reported in 2003 in South-East Asia. This infectious disease of birds caused by the type A strains of influenza virus is highly lethal and has now spread to 13 different countries in Asia and Europe. Till date, a total of 170 human cases are reported causing 92 deaths[1]. In the year 2006, bird flu is recored in Turkey and Iraq for the first time. Though the cause of all reported human cases is from direct contact with diseased poultry, a fully transmissible pandemic might arise via process of re-assortment events and adaptive mutations in influenza virus. Furthermore, with the increase in global transport, urbanization and overcrowded conditions, it is likely that the virus spreads around the world quickly.

Given such lethal transmissible influenza pandemic threat, it is essential to be prepared for the potential outbreak of pandemic influenza. However, lack of enough knowledge about the dynamics of the virus strain makes it difficult to design effective controlling strategies. Nascent state of $H5N1$ vaccines and anti-viral drugs makes the design of containment policies even more challenging. Large scale vaccination policies can significantly reduce the threat of pandemic but the high cost make these policies impractical.

It is thus important to implement adequate measures before the occurrence of pandemic. Analysis of interactions among people and their movement in co-located geographic regions is the key for assessing the transmissibility of the virus. Eubank et. al [3], [4] have developed disease outbreak models for generating large-scale synthetic data. They also proposed fast algorithms for computing basic structural properties such as clustering coefficients and shortest paths distribution. Ferguson et. al [5] have proposed strategies to contain the emerging influenza pandemic.

In this paper, we primarily focus on developing algorithms for finding individuals who are highly infectious. We also find locations at which people are more susceptible to the infection. This paper is organized as follows: In Section 3, we briefly discuss about the data format and models used to represent the data. Section 4 presents our various containment policies. Finally, we discuss the results in Section 5 followed by conclusions.

3. DATA AND MODELS

We use the simulation data generated from the model described in [6]. The data describes the interactions between individuals in Portland, USA. The data includes the geographical locations of places in Portland, demographic information of individuals, their daily activities at different locations and their contact network based on daily activities. Furthermore, it also provides the details of disease outbreak and spread that is simulated by EpiSims system. The data consists of 1.6 million people spread over $246,000$ locations. At the start of the simulation, 100 individuals are selected and are marked as diseased. The system is then run to simulate 100 days. When a person gets infected, the time and place at which the infection occurred is recorded by the system. With no containment policies, $565,000$ people ($\sim 35\%$ of whole population) are infected with the disease, at the end of 100^{th} day.

The given simulation data can be modeled in many different ways. Such models should capture the interactions among people and interplays between people and locations. Our models are generic in nature and hence many graph theoretic algorithms can be easily applied to them. We use two representations, *People - Locations Activities Graph* (PLA) and *People - People Contacts Network* (PPC) (Figure 1). Though these two models seem to capture different interactions, one model can be derived from the other. In our discussion, we chose to differentiate these two models for the ease of exposition. The PLA graph, (V_P, V_L, E_P), is a bi-partite graph and models the relationship between people and locations in the city. V_P and V_L represents the set of all people and locations, respectively. Here, an edge $(P_i, L_j) \in E_P$ implies that the person P_i is *connected* to the location L_j to perform some activity like going to school. Each edge is weighted with the type, start time, and duration of the corresponding activity. The PPC network, (V_C, E_C) captures the social interactions between people in the form of contacts. Edges in PPC network are weighted with type, start time and duration of the contact. Contact type is the purpose for which the contact is made. It is one of *Home (id: 0), Work (1), Shop (2), Visit (3), Social/Recreation*

*Contact email: srini@cse.ohio-state.edu

[1] http://www.who.int/csr/disease/avian_influenza/en/

(4), Other (5), Pick up (6), School (7), College (8). Contact hours $(0-23)$ is the duration for which the contact is made. An edge $(P_i, P_j) \in E_C$ weighted with c, s, h implies that the person P_i is in contact with P_j for the purpose of c and for the duration of h hours, starting at s.
We observe that PLA graph and PPC networks are scale-free networks. Their degree distributions follow the power law i.e., number of nodes with degree k falls as $k^{-\alpha}$ for some constant α. Scale-free networks and power laws are well addressed in [1] and [7]. For example, in PLA graph, we have observed the value of α to be between $1.7-2.1$ for locations and around 1.8 for people.

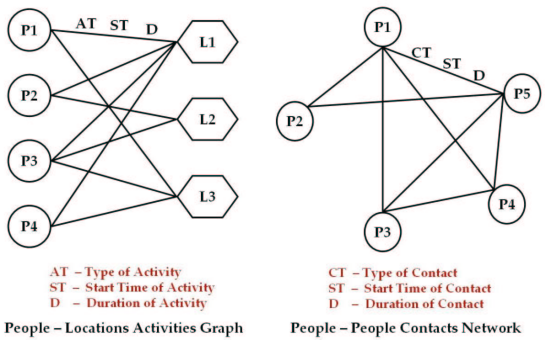


Figure 1: Data Models (a) People - Locations Activities graph and (b) People - People Contacts network

We now present the effect of demographic attributes like age, household size etc. on disease spread. Such an analysis can help in developing probability functions for gaining better insights into the data. We concentrate on four attributes viz.,*household size, contact type, age,* and *hours of contact*. Figure 2 illustrates the effect of these attributes on disease outbreak. X-axis in each graph shows the type of attribute and Y-axis gives the number of people who got infected. Not surprisingly, an infected person in a larger household is more infectious than an infected person in a smaller household, because the person in larger household can spread the infection to more number of people. For example, an infected person in an household of size 14 can infect 80% of the other household members. Figure 2 (b) illustrates how the contact type effects the probability of a person to get infected. All types of contacts are not equally infectious. Contacts at schools and work places are far more infectious than other contacts. Similarly, a person's age can effect the susceptibility of a person to the disease. We can infer that the people who are between age of 5 and 60 are more susceptible to the infection than the people from other age groups. Along with the type of activity, the duration of contact also effects the person's susceptibility to the disease. Chances of getting the infection is directly proportional to the time spent with the infected person.
Based on this analysis, we developed probability functions as follows. Let us consider the example of contact types. From figure 2 (b), we can infer that 24% of the people going to the work and 21% of the people attending the school are infected. By normalizing these susceptibility percentages over all activities, we can model the effect of contact type as a discrete probability mass function. In similar spirit, we can construct probability density functions for other attributes also. Based on a person's attributes (age, household size etc.) and the above developed attribute specific functions, we can derive the probability of that person to get infected.

4. STRATEGIES

A naive strategy to contain the disease is by vaccinating the whole population. However, difficulty in implementing and high cost renders this scheme practically infeasible. Another seemingly correct approach will be to vaccinate all the household members of the 100 initially infected people. This strategy is easy to implement and cheap but not effective. It could only prevent $66,419$ cases out of a total of $565,685$ cases. Therefore there is a need for more sophisticated methods which take into account interaction among people, effect of locations, activity type and number of contact hours. In this section, we propose five such policies. Each of our five strategies chooses a set of people to vaccinate from the whole population. Intuitively, we would like to choose and vaccinate the people who are more probable to get the disease in future.

4.1 Random Vaccination

This strategy randomly chooses the set of people to vaccinate. A simple random sampling without replacement is employed for this purpose. Random numbers are generated using a Uniform distribution.

4.2 Contacts Driven Vaccination

Another intuitive vaccinating policy is to select people who are directly in contact with an infected person. One can continue to higher levels by choosing the people who are in direct contact with the contacts of infected person, and so on. In graph theoretic terms, this algorithm is called as Breadth-First Search (BFS). We start the BFS from each of the 100 people who are initially infected (*sources* of BFS). In level i, we select nodes which are connected to the *source* by $i-1$ nodes or i edges.

4.3 Sociability Driven Vaccination

As mentioned in section 3, PPC network is a scale-free network. The distinguishing feature of a scale free networks is the presence of centrally located and highly connected nodes known as *hubs*. In the PPC network, vertex degrees (i.e., number of contacts) varies from 0 to 277 with the $\mu = 39$ and the $\sigma^2 = 33.28$. There are 956 people (0.06% of population) with number of contacts greater than or equal to 200 and hence they act as hubs. These hubs corresponds to people who are more societal compared to others. A straight forward strategy will be to vaccinate these hubs to control the outbreak. But, vaccinating these 956 people could only prevent 4655 cases, which is less than 1% of total number of infected people. Vaccinating people with number of contacts ≥ 150 could prevent only 12% of the cases. Therefore, vaccinating just the hubs will not help in containing the disease from spreading. Instead of concentrating on just the hubs, we propose to vaccinate all the nodes with degree greater than a fixed *cutoff degree*. If the cutoff degree is small, then the set of nodes chosen for vaccination would be large. Hence, it can lead to better prevention rate. We discuss the detailed results in section 5.

4.4 Profile Based Vaccination

In this strategy, we use random walks on the PPC network to select the set of people to vaccinate. Random walk is a random process consisting of sequence of discrete steps of

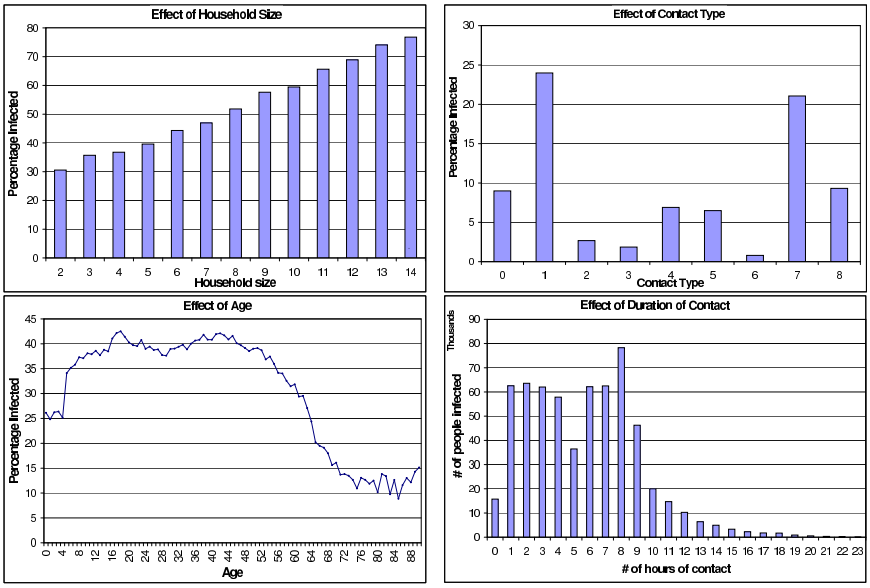


Figure 2: Effect of attributes on disease outbreak (a) Household size (b) Contact Type (c) Age (d) Contact Hours

fixed length. In our context, each discrete step corresponds to traversing an edge of the graph. Random walk starts at a node (i.e., person) referred to as *source* of the walk. At each step, next node in the walk is chosen randomly from all the nodes, which are adjacent to the current node. Multiple random walks from a given node can be performed using restarts i.e., at each step during the walk we can restart from the source with certain probability known as *restarting probability*. We say that two people have similar profiles if their contact networks are similar. Sun et. al [8] have demonstrated, in context of bi-partite graphs, that random walk with restarts can be used to determine nodes which are most relavent (i.e., more similar) to the source node. If the source is an infected person, random walk visits a set of people who have similar profile as of the source. Therefore, people visited during such random walks are more susceptible to the infection. We treat each of the 100 initially infected people as 100 sources of a random walk.

In traditional random walk, probability of an edge to be taken is same for all the edges incident on a given node. From Figure 2 (b), we know that the contact type and hours of contact effects a person's susceptibility to infection. For example, probability of getting infected at a work place is 0.29 whereas at a shopping place it is 0.033. We construct these probability distributions for attributes, contact type and the duration of contact (as described in section 3). Each edge in the contact network is weighted based on the distributions constructed above. Assume that v is an intermediate node in the random walk and $E_v = \{e_1, e_2, ..., e_n\}$ is the set of edges incident on v. Let edge e_i corresponds to one of the v's contact with type X and duration H. Assume that p_x and p_h are probabilities of getting infected for contact type X and duration of contact H, respectively. Each edge e_i is then weighted with the product $p_x \times p_h$. We normalize the weights of all the edges incident on v in order to make the sum to 1.

4.5 Location Based Vaccination

Till now, we have presented four vaccination policies which are based on the contact network of people. In this section, we propose a policy based on locations at which people perform their activities. In order to study the people-locations relationship, we make use of PLA graph. We first identify the critical locations based on the number of cases reported at each location. Assume that a location, L is connected to P_L number of people (i.e., the degree of L in PLA graph) out of which I_L are infected. We calculate the measure *Infection Ratio* at location L as on day D, $IR_L^D = \frac{I_L}{P_L}$ to analyze how infectious a location is. A location is declared as *critical* if IR_L^D exceeds a certain threshold value, $Threshold_{IR}$. D is referred to as *Policy Effective Date* (PED). Even though IR_L^D can be calculated every time I_L changes, we calculate IR_L^D at the end of each day for computational efficiency. It is important to note that the cases reported till policy effective day *can not* be prevented. Once the critical locations are identified, people can be vaccinated in various ways. One possible way is to vaccinate all the people who are connected to critical locations through some activity.

This method can easily be extended to make more sophisticated policies. For example, the concept of IR_L^D can be extended to take the type, start time and duration of the activity into consideration. And, instead of calculating IR_L^D for each location, we can group locations that are collocated into regions to identify *critical regions* over space. Such grouping of locations can either be *static* or *dynamic*. Static grouping can be based on geographic boundaries or any other given constraints. In dynamic grouping, set of locations constituting a group can change over time.

5. RESULTS

In this section, we first present the effectiveness of proposed policies and compare them. We then evaluate our strategies under different constraints like limited number of anti-viral drugs and delay in response time. We compare the effectiveness of different methods using the measure, *Percentage Prevention*. It is the percentage of people who are prevented from the disease. To calculate this measure, we make use of disease evolution data from EpiSims simulation system

[6]. This data provides insight into when, where and from whom a person got infected. Simulation is started at $t = 0$ with 100 initially infected people. When any person is infected during the simulation, system records the time and location at which the person got infected along with the list of *already* infected people who are currently in the same location. Assume that a person, P is infected at time T_P during the simulation. With no containment policies, P can infect other people in the contact network. Assume that P infects a person Q at T_Q ($> T_P$). Now, assume that using one of the disease containment strategies, we vaccinate P at time T_V ($T_V < T_P$). Since P is vaccinated, P is prevented from the disease *directly*. But vaccinating P in turn also prevents Q from getting infected because P is no longer infectious. We say that Q is prevented from the disease *indirectly*. Therefore, vaccinating a person not only prevents that person but can also prevent other people indirectly. *Percentage Prevention* includes total number of people who are prevented from infection both directly and indirectly. Another closely related measure is the *cost incurred* by the containment strategy. It is inversely proportional to number of cases prevented (both directly and indirectly) per vaccine.

5.1 Effectiveness of Policies

In every strategy, percentage prevention increases as more people are vaccinated. Random vaccination policy gives the theoretical upper bound on number of vaccines needed to achieve the given prevention rate (Figure 4 (a)). In practice, this strategy will not be effective as it can not take any extra knowledge about the data or the disease into account.

Figure 4 (b) shows how the number of vaccines and percentage prevention changes as we change the number of levels of BFS. Exponential increase in number of nodes visited (i.e., number of vaccines) by BFS illustrates shorter path lengths of small-world PPC network. Although we prevent 98% of cases at level 3, the cost incurred is very high because we vaccinate 71% of population. From level 2 to level 3, number of cases prevented per vaccine dropped from 162 to 48. This method is very costly and therefore may not be very practical. We later show that even with a fixed number of vaccines, it performs poorly when compared to the profile based and sociability driven strategies. Note that, BFS with 1 level will vaccinate the direct contacts of an infected person. Since a person can pass on the infection only to direct contacts, BFS with 1 level should achieve 100% prevention rate. But, this is not true for the given data. This is due to the presence of people who are neither infected from others nor in the set of initially infected people. We attribute them as people who got infected naturally but not from contacts made with an infected person.

Figure 4 (c) illustrates the variation in percentage prevention and number of vaccines given as we vary the cutoff degree. Trend representing the number of vaccines expounds the power-law degree distribution in PPC network. Presence of highly connected nodes (*hubs*) can be seen from the slow increase in the number of vaccines given, initially. A quick increase after the cutoff degree of 100 is due to large number of nodes with smaller degrees of contact. Hence, a smaller cutoff degree leads to higher percentage prevention. Sociability driven prevention strategy offers the lower bound on number of vaccines to be spent for achieving a given percentage of prevention. Prevention of 99.21% is achieved at cutoff degree of 30. This can be achieved by vaccinating *at least* 55% of population.

Effectiveness of the profile based vaccinating policy with varying number of steps is shown in figure 4 (d). We have set the restarting probability to be 0.15 for our experiments. As we increase the number of steps taken during the walk, number of nodes visited and hence the prevention rate goes up. As more number of nodes are visited, more number of vaccines are used. Therefore, there is a trade-off between the cost incurred and the percentage prevention achieved. Thus, constraints such as availability of vaccines and other resources should be taken into account when determining the number of steps. It is important to note that the increase in percentage prevention is slow when compared to increase in number of vaccinations. In other words, number of cases prevented per vaccine reduces as the number of steps increases. In practice, profile based strategy might work better than any sociability driven strategy. Because, sociability driven methods require the exact knowledge of contacts of a person and they assume the contact network to be static. In real life, it is very difficult to keep track of exact information of contacts as the contact network changes over time. Hence, in such cases, we can expect the profile based policies to be more effective than others.

5.2 Location Based Vaccination

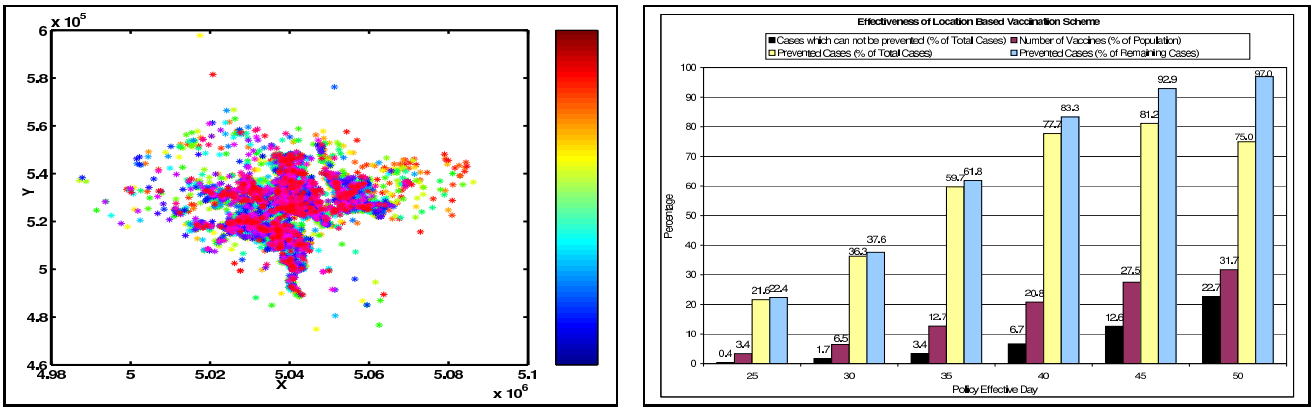
Figure 3 (a) shows the effectiveness of location based vaccination scheme for various values of policy effective day. We have set $Threshold_{IR}$ to 0.01 i.e., if 1% of people connected a location are infected then that location is declared as critical. Let T be the set of people who got infected during the 100 day simulation. And, let S is the set of people who got infected before PED. Hence, people present in S can not be prevented from disease. We define a set R as $T - S$ that represents the set of people who are yet to be prevented. Percentage prevention can be calculated based on both T and R. We refer to prevention rate as a percent of T and R as PP_T and PP_R, respectively. As mentioned earlier, we vaccinate all the people connected to critical locations.

As we increase the PED, number of cases which can not be prevented increases quickly. For example, if we delay the policy for 50 days then almost 23% of the cases can not be prevented. As the time progresses, disease spreads among the people and so the number of locations with infected people increases. Since we vaccinate all the people connected to infected locations, number of vaccines given and, hence, the PP_R increases. Note that the amount of increase in PP_R reduces as the PED increases. But the PP_R does not give the overall effectiveness of the vaccinating policy. To analyze the exact behavior or to compare against other policies, one should use PP_T. As we change PED from 45 to 50, difference between PP_T and PP_R will become evident. Though the PP_R increases from 92.9% to 96%, PP_T actually decreases from 81.2% to 75%. This is due to the quick increase in number of cases which can not be prevented, from 12.6% to 22.7%.

Figure 3 (b) shows the distribution of infectiousness across various locations. For example, Locations in red color have more infected people. Spatial clustering algorithms like DBSCAN [2] can be applied on this graph to group critical locations into critical regions.

5.3 Effect of Resource Constraints

Anti-viral drugs are often limited in number because of vari-



(a) (b)

Figure 3: Location Based Vaccination. (a) Distribution of infectious locations at PED = 35 (b) Effectiveness

ous reasons like mass production cost, inventory cost etc. In this section, we fix the number of vaccines available and evaluate how the percentage prevention changes across different strategies. Such a comparison will enable us in analyzing the effectiveness and feasibility of various strategies given resource constraints. Figure 5 shows the differences in effectiveness if we fix the number of anti-viral drugs to be used. For all the strategies, total number of prevented cases increases and number of cases prevented per vaccine decreases as the number of vaccines used increases. As we vaccinate larger section of population, difference between their effectiveness decreases. *For any given number of vaccines, sociability driven vaccination policy is the clear winner retaining high levels of effectiveness over all other methods of vaccinating.* As shown in section 4.2, the reach of BFS is very high but for a fixed number of anti-viral drugs, its effectiveness falls behind the sociability driven and profile based vaccination policies.

5.4 Effect of Delay in Response

It is not practical to assume that the disease containment policies can be implemented as soon as the first case of the infection is reported. The delay can be due to various constraints like distance between anti-viral drug inventory and the location at which the infected person resides or might be due to late diagnosis. Therefore, it is important to evaluate the tolerance levels of our strategies to the delay in response after the first case has been reported. Figure 5 shows the effectiveness of different methods as a function of reaction time (in terms of days). Number of anti-viral drugs has been fixed at $350,000$ courses for this experiment. Clearly, number of cases prevented goes down as we delay the implementation of containment policies. From the graph, it can be inferred that, in general, *delay up to 35 to 40 days is acceptable for sociability driven and profile based vaccination policies without foregoing significant prevention percentage.* For other two policies, number of people prevented from infection continuously decreases as the response time increases. After 40 days, prevention rate decreases at a faster rate especially in sociability driven scheme. Difference in effectiveness between degree and profile based policies decreases as the delay increases.

6. CONCLUSIONS

In this paper, we proposed and evaluated five different prophylaxis policies to contain the disease from spreading. Four policies are based on PPC network and one is based on PLA graph. Over all, degree based policy gave the best performance in terms of Percentage Prevention. But in practice, profile based strategy might perform better than any sociability strategy as the contact network would be changing. Moreover, It is extremely difficult to obtain the exact contact network of a person. We have shown that implementation of these policies can be delayed up to 35 days without effecting the percentage prevention. In future, We would like to use density-based spatial clustering algorithms like DBSCAN to group collocated locations with similar infection ratios to yield critical regions.

7. ACKNOWLEDGMENTS

This work is supported by NSF grants CAREER-IIS-0347662, RI-CNS-0403342, and NGS-CNS-0406386. The authors would like to thank Sitaram Asur, Duygu Ucar and Greg Buehrer for their useful comments and suggestions.

8. REFERENCES

[1] R. Albert and A. Barabasi. Statistical mechanics of complex networks. In *Review Modern Physics*, 2002.

[2] Martin Ester, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, 1996.

[3] S. Eubank, H. Guclu, V. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modeling disease outbreaks in realistic urban social networks. In *Nature*, volume 429, pages 180–184, 2004.

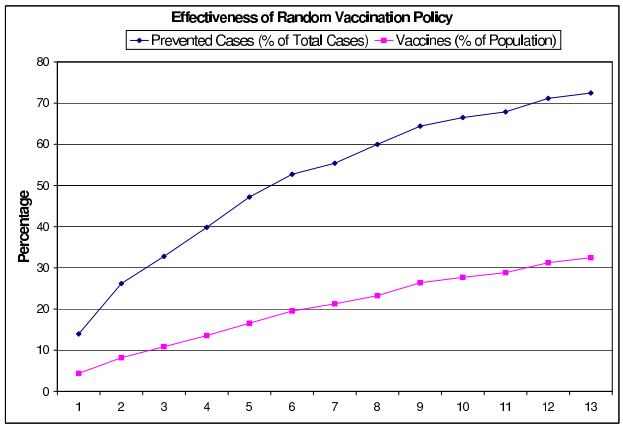
[4] S. Eubank, V. Anil Kumar, M. Marathe, A. Srinivasan, and N. Wang. Structural and algorithmic aspects of massive social networks. In *Symposium on Discrete Algorithms*, 2004.

[5] N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. Strategies for containing an emerging influenza pandemic in southeast asia. In *Nature*, volume 437, pages 209–214, 2005.

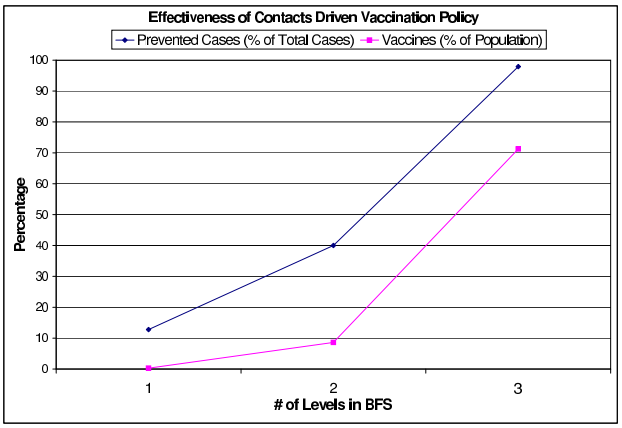
[6] Synthetic Data Products for Societal Infrastructures and NDSSL-TR-06-006 Proto-Populations: Data Set 1.0. Network dynamics and simulation science laboratory, virginia polytechnic institute and state university.

[7] M. Newman. The structure and function of complex networks. In *SIAM Review*, 2003.

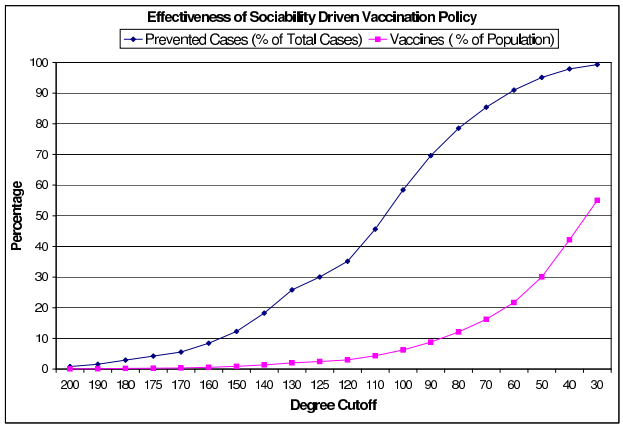
[8] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In *Fifth IEEE International Conference on Data Mining*, 2005.



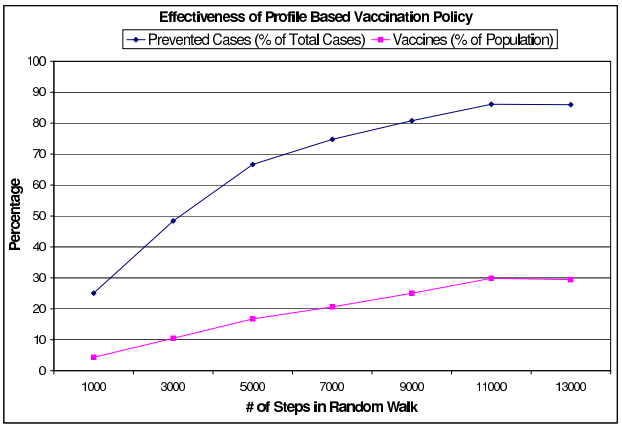
(a) Random vaccination



(b) Contacts driven vaccination

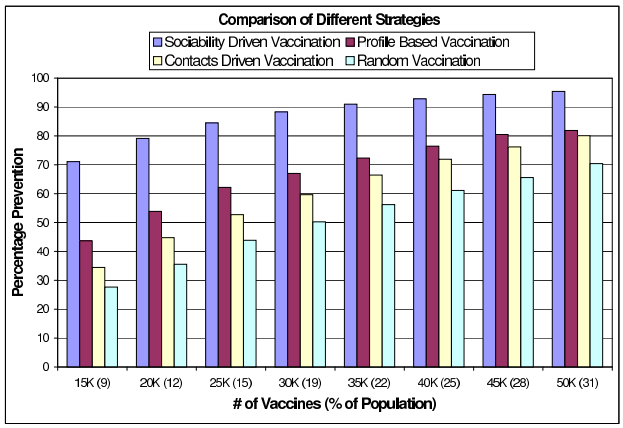


(c) Sociability driven vaccination

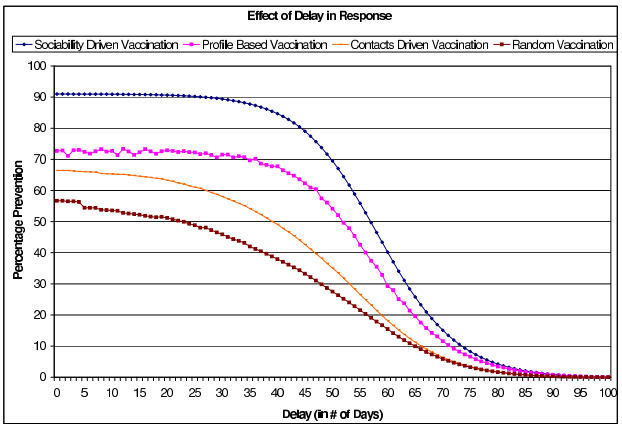


(d) Profile based vaccination

Figure 4: Effectiveness of various vaccination policies



(a) Limited anti-viral drugs



(b) Delay in response time

Figure 5: Effectiveness under constraints