Epidemic spreading survey

Zhifeng Sun

February 8, 2009

Lots of real word problems can be motivated and modeled as spreads of epidemics through a network. Prominent examples include the spread of worms and email viruses over the Internet, the spread of disease among the population, and the spread of harmful gossip or panic in a social network. We want to protect computer networks from viruses, prevent disease spreads, and control the leakage of sensitive information and unpleasant gossip. At the same time our resources (anti-virus software, vaccination, influence) are costly and limited, so we are interested in achieving the best possible effect, while allocating the minimum possible resources.

Two epidemic models are commonly used when studying the spreads of epidemics through a network, the susceptible-infected-susceptible (SIS) model and the susceptible-infected-recovered (SIR) model. The main difference between these two models is, a recovered individual can be infected again in the SIS model, but not in the SIR model which assume recovered individuals have life long immunity to the disease. And this difference makes them suitable for different kinds of problems. For instance, when studying childhood diseases which individuals can have long-lasting immunity, either naturally or from vaccination, it is more appropriate to use SIR model; when studying viruses transmitting over the Internet, often times it is more reasonable to use SIS model, since most of the viruses mutate, in which case even if a computer is cured by anti-virus software and therefore not susceptible to the original virus, it is still susceptible to a mutated virus.

Within each model, quite amount of research work has been done, touching different aspects of the problems. This survey is trying to give a comprehensive overview of the research work that has been done in the spreads of epidemics, and discover open problems and further directions. The survey is going to be structured as follows. In section 1, we introduce SIR model formally, browsing different kinds of problems and techniques studied in such a model, and proposing some open problems and further directions in this line of work. In section 2, we use similar approach to introduce SIS model. In section 3, we analyze different kinds of contact graphs, since the contact graph plays an important roll in epidemic problems. Different kinds of graphs have different set of properties, which may help us get different results under SIR/SIS models. More importantly, we need to know for different problems, what kind of contact graph models the real underlying epidemic network well.

1 SIR model

SIR model and its variants are widely used in the analysis of the outbreak and spread of infectious diseases, which is very useful for mass vaccination programme. An infectious disease is said to be endemic when it can be sustained in a population without the need for external inputs. This means that, on average, each infected person is infecting exactly one other person (any more and the number of people infected will grow exponentially and there will be an epidemic, any less and the disease will die out).

The structure of this section is organized as follows. Section 1.1 introduce SIR model definition and its parameters formally. Section 1.2 summaries the analysis results for SIR model. Section 1.3 shows the research work on optimization problems under SIR model. Last, section 1.4 looks at game theory analysis results for epidemic spread problems in SIR model.

1.1 Model setup

SIR model is used to study diseases which individuals can have long-lasting immunity, like most common childhood diseases (measles, mumps, rubella, etc.). So it makes sense to divide the population into those who are susceptible to the disease (S), those who are infected (I) and those who have recovered and therefore are immune (R). These subdivisions of the population are called compartments. The letters (S, I, R) also represent the number of people in each compartment at a particular time. N is the total number of people, i.e. $N = S + I + R$. To indicate that the numbers might vary over time we make the precise numbers a function of time (t): $S(t)$, $I(t)$ and $R(t)$. The transition rate from S to I is β. The transition rate from I to R is υ (recovery rate). Shown in Figure 1.

Figure 1: SIR model.

Originally, the SIR model doesn't have the concept of contact graphs. The assumption there is the population is homogeneous, i.e. individuals make contact at random to the whole population. And people usually study the case where the size of the population is a constant. Under such assumptions, the epidemic can be captured by a set of differential equations.

$$\frac{dS}{dt} = \mu N - \mu S - \beta \frac{I}{N} S \quad (1)$$

$$\frac{dI}{dt} = \beta \frac{I}{N} S - (\mu + \upsilon) I \quad (2)$$

$$\frac{dR}{dt} = \upsilon I - \mu R \quad (3)$$

where μ is the birth rate, which is also death rate, since we assume the population is a fixed constant. By solving these differential equations, we can know at any time t, how many people are susceptible, how many people are infected and how many people are recovered.

However, once we include contact graphs, the population is not homogeneous any more. Different individuals (vertices in the graph) have different neighbors. Those vertices with large degree are more likely to spread the disease than those with small degrees if they get infected. So we cannot use the above differential equations to capture the epidemic process in this case.

1.2 Analysis results

In this section, we present the analysis results of SIR model, which is without the existence of vaccination or anti-dots, how the disease is going to spread, is the epidemic going to last or die out?

Since so far people often study disease spread without considering contact graphs in SIR model and use those results to help mass vaccination programme, the follow analysis results will focus on this restricted model.

In epidemiology, the *basic reproduction number* (sometimes called basic reproductive rate or basic reproductive ratio) of an infection is the mean number of secondary cases a typical single infected case will cause in a population with no immunity to the disease in the absence of interventions to control the infection. It is often denoted as R_0. This metric is useful because it helps determine whether or not an infectious disease will spread through a population. When $R_0 < 1$, the infection will die out in the long run; when $R_0 > 1$, the infection will be able to spread in a population. Large values of R_0 may indicate the possibility of a major epidemic. Generally, the larger the value of R_0, the harder it is to control the epidemic. In particular, the proportion of the population that needs to be vaccinated to provide *herd immunity* and prevent sustained spread of the infection is given by $1 - 1/R_0$.

Based on equations (1), (2) and (3), define *basic reproduction number* to be

$$R_0 = \frac{\beta}{\mu + \upsilon}$$

which has threshold property. In fact, independently from biologically meaningful initial values $(S(0), I(0), R(0))$, one can show that if $R_0 \leq 1$, then

$$\lim_{t \to \infty} (S(t), I(t), R(t)) \to (N, 0, 0)$$

which is called the *Disease Free Equilibrium*; if $R_0 > 1$ and $I(0) > 0$, then

$$\lim_{t\to\infty}(S(t), I(t), R(t)) \to \left(\frac{N}{R_0}, \frac{\mu N}{\beta}(R_0 - 1), \frac{\upsilon N}{\beta}(R_0 - 1)\right)$$

which is called the *Endemic Equilibrium*.

The above result shows if $R_0 > 1$, we need vaccination or anti-dots to eradicate the epidemic. The following is the analysis of Mass Vaccination Program, i.e. how many people we need to vaccinate in order to have herd immunity and prevent sustained spread of the disease. Let us consider a disease for which the newborn are vaccinated (childhood disease) at rate p, and V be the number of vaccinated people. Then we have the following differential equations.

$$\frac{dS}{dt} = \mu N(1-p) - \mu S - \beta\frac{I}{N}S \tag{4}$$

$$\frac{dI}{dt} = \beta\frac{I}{N}S - (\mu + \upsilon)I \tag{5}$$

$$\frac{dV}{dt} = \mu N p - \mu V \tag{6}$$

One can show that if $R_0(1-p) \leq 1$, then

$$\lim_{t\to\infty}(S(t), I(t)) \to (N(1-p), 0)$$

which is disease free equilibrium; if $R_0(1-p) > 1$ and $I(0) > 0$, then

$$\lim_{t\to\infty}(S(t), I(t)) \to \left(\frac{N}{R_0(1-p)}, \frac{\mu N}{\beta}\left(R_0(1-p) - 1\right)\right)$$

In other words, if $p > p^* = 1 - 1/R_0$, then the vaccination program is successful in eradicating the disease, on the contrary it will remain endemic, although at lower levels than the case of absence of vaccinations.

1.3 Optimization problems

As we have seen in section 1.2, when $R_0 > 1$, we need vaccination or anti-dots to eradicate the disease. If we don't consider contact networks, we already know we need to vaccinate $1 - 1/R_0$ percent of the population in order to successfully eradicate the disease. However, if we take contact networks into account (which models the real word problems better), how many people we need to vaccinate to control the epidemic, and how to distribute vaccinations (or anti-dots) to better make sure of these limited resources?

Such problem can be formulated as follows. Given a graph $G = (V, E)$ which represents the contact network, and a number k which is the number of vaccinations available, how to distribute these vaccinations such that the epidemic size is minimized. If a node is vaccinated and the vaccination is successful, then this node is immune to the disease and removed from graph G.

In [2], Aspnes et al studied one special case of this problem. The model they considered is for highly infectious computer virus (or human diseases), i.e. if one node in G is infected, all the other nodes in the same connected component will get infected too. The computer virus (or disease) starts from a single node, randomly chosen from graph G. So they are trying to minimize the expected size of connected component. And they showed that even in such restricted model, such problem is NP-hard. And they give a $O(\log^{1.5} n)$ approximation algorithm, where n is the number of vertices in graph G.

[11] studied network immunization against virus spread with limited immunization budget using simulation approach. They considered such problem in a discrete-time special case of SIR model (called *independent-cascade model* proposed by Kempe et al [14]), i.e. at time $t = 0$ the adversary plants r viruses to some vertices of the graph. Then if a vertex i becomes infected for the first time at time t, it is given a single chance to infect each of its neighbors j that is currently uninfected. The probability that vertex i succeeds in infecting vertex j is p_{ij}. They propose an algorithm with good simulation performance, but no theoretical proof.

1.4 Game theory problems

The optimization problem explained in section 1.3 only models the situation where there is a centralized authority, like US government, who has some limited resources (certain number of vaccinations, anti-dots, anti-virus software, etc.). While in real word, it's often not the case. For example, in volunteer vaccination program, as more individuals become vaccinated, the remaining unvaccinated individuals are increasingly unlikely to become infected, because of herd immunity. For a population with sufficiently high vaccine coverage, a disease can be eradicated without vaccinating everyone. Therefore, as coverage increases, there is a greater individual incentive not to vaccinate, since non-vaccinators can gain the benefits of herd immunity without the risk of vaccine complications. In such settings, we need *game theory* to capture such individual behaviors, which formalizes strategic interactions in a group where individuals attempt to maximize their payoffs.

In this section, we are going to introduce the game theory study in epidemic spreading problems. In section 1.4.1, we look at the one shot game in SIR model without contact graph. In section 1.4.2 we present research work on one shot game in SIR model with contact graph. Lastly, in section 1.4.3, we explore studies on multi-round game in SIR model without contact graph.

1.4.1 One round game without contact graph

In this section, we look at the most simply game theory setup. It is a single round game, all players make choices simutaneous based on their strategies, and they recieve certain payoff at the end. Each person knows all the global information, like vaccination uptake level, probability of getting infected, etc. Also there is no contact graph, i.e. the population is homogenous.

[4] studies the problem of voluntary vaccination policies for childhood diseases. The kind of diseases they considered is once you get recovered or vaccinated, you will be immune during life time. Since if a sufficient proportion of the population is already immune, either by vaccination or naturally, then risk associated with vaccination will outweight the risk from infection, which means parents will tend to not vaccinate their children. In their model, each person takes vaccination with probability P, and its expected payoff is

$$E(P, p) = P(-r_v) + (1 - p)(-r_i)\pi_p$$

where p denotes the vaccination uptake level in the population, r_v denotes the morbidity risk of taking vaccine, r_i denotes the morbidity risk of getting infected, and π_p denotes the probability that an unvaccinated individual will eventually be infected if the vaccine coverage level in the population is p. In such model, they showed Nash equilibrium always exists. And the calculation shows it is impossible to eradicate a disease through voluntary vaccination when individuals act according to their own interests.

1.4.2 One round game with contact graph

In this section, the game theory model considers the underlying contact graph. So the disease will transmit through this contact network. Different factors, and utility models can be take into consideration. Like, is the vaccination reliable or not? What is the disease transmission probability? Does each person know the global information or only the situation of his neighbors in the contact graph? The following references study this kind of game theory problems.

[2] studies the spread of viruses in general undirected graph. The viruses are highly infectious, i.e. if one vertex gets infected, all the other vertices that are in the same connected component will eventually get infected. If a vertex installs anti-virus software, which has cost C, it will be immune from viruses. If a vertex does not install anti-virus software and gets infected, it will experience a loss of L. Each vertex has all the global information and attempts to calculate the best strategy for itself. They showed pure Nash equilibrium always exists in such setting, and can be found in polynomial time. However, both computing the pure Nash equilibrium with lowest social cost and computing the pure Nash equilibrium with highest social cost are NP-hard. They also consider the centralized version of this problem, and obtain a polynomial time $O\left(\log^{1.5} n\right)$-approximation algorithm for computing social cost, where n is the total number of vertices in the graph.

[3] studies the game between a virus and an alert over a network. Initially, randomly select a small set of detector nodes, and a single node to start infection. In every round thereafter, each infected node sends out a constant number (β) of worms to other nodes, and each alerted node sends out a constant number (α) of alerts. Infected nodes can send out worms to any node in the graph, however alerted nodes can only send out alerts through a previously determined, constant degree overlay network. If a worm is received by a node that is not a detector and is not alerted, that node becomes infected; If a worm is received by a node

that is a detector, that node becomes alerted; If an alert is received by a node that is not infected, that node becomes alerted. The question is, assuming that infected nodes are omniscient, is there a strategy for the alerted nodes that ensures only a vanishingly small fraction of nodes become infected? In such settings, they show, by using a simple alert strategy (i.e. each alerted node send out alerts to α nodes selected uniformly at random without replacement from its neighbors in the overlay graph), if $d \geq \alpha$ and $\frac{\alpha}{\beta(1-\gamma)} > \frac{2d}{c}$, then with high probability only $o(n)$ nodes get infected, where d is the degree of the overlay graph, c is the node expansion of the overlay graph, γ is the probability that each node is a detector.

1.4.3 Multi-round game without contact graph

The game theoretical analyses introduced in previous sections rely upon classical game theory, assuming a static game where individuals know all the global information. In reality, people change their strategies over time, and individuals don't have global information. Moreover, humans adopt new strategies through learning, by imitating others who appear to have adopted more successful strategies. This means we should develop dynamics in game theory model.

[5] set up a dynamic game theory model for childhood disease vaccination program. The payoff for vaccinated people is $f_v = -r_v$, where r_v is the perceived probability of significant morbidity from the vaccine (perfect vaccine efficacy is assumed). The payoff for non-vaccinated people depends on the perceived probability r_i of suffering significant morbidity upon infection, and the perceived probability of eventually becoming infected, which is assumed to increase linearly with the current disease prevalence $I(t)$, which models the fact that each individual doesn't know what is the exact probability he will be infected. So payoff for non-vaccinated people is $f_n(I) = -r_i m I$ where parameter m quantifies the sensitivity of vaccinating behavior to changes in prevalence. For the dynamic to imitate other people's strategies, they assumed that individuals randomly sample other members of the population at some constant rate. If the strategy of the sampled member provides a higher payoff, then his strategy is adopted with a probability proportional to the expected gain in payoff.

2 SIS model

SIS model is originally used to study disease outbreak and spread, where individuals don't have long-lasting immunity, like common cold. This means an individual who is recovered from infection can be infected again. Later on, this model is also used for modeling virus spread over the Internet or email networks, since most of the viruses mutate. So even if a computer is cured by anti-software and therefore not susceptible to the original virus, it is still susceptible to the mutated viruses.

The structure of this section is organized as follows. Section 2.1 introduce the definition and parameters in SIS model. Section 2.2 summaries the analysis

results in SIS model, like disease die out time and epidemic size. Section 2.3 presents the research work on the optimization problems under SIS model. For example, with certain amount of anti-dotes, how to distribute them in order to reduce the disease die out time most effectively? Last, section 2.4 shows those results which study epidemic spreads problems from game theory point of view.

2.1 Model setup

Similar to SIR model, SIS model divides the population into two groups, those who are susceptible to the disease (S) and those who are infected (I). These subdivisions of the population are called compartments. The letters (S, I) also represent the number of people in each compartment at a particular time. N is the total number of people, i.e. $N = S + I$. To indicate that the numbers might vary over time we make the precise numbers a function of time (t): $S(t)$ and $I(t)$. The infection rate (from state S to state I) is β, and recover rate (from state I to state S) is υ. Shown in figure 2.

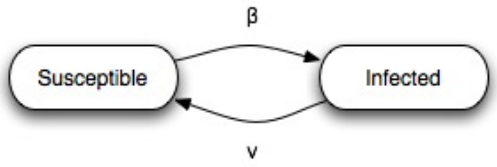

Figure 2: SIS model.

If we don't consider contact graphs and assume the population is homogeneous, then SIS model can be captured by the following equations.

$$\frac{dS}{dt} = -\beta S\frac{I}{N} + \upsilon I \tag{7}$$

$$\frac{dI}{dt} = \beta S\frac{I}{N} - \upsilon I \tag{8}$$

The same as SIR model, once we introduce contact graph, we loss the homogeneous property, hence we cannot use those differential equations to model such process any more. With contact graph, we can describe SIS model as follows. Given graph (contact network) $G = (V, E)$, V represents the set of people, and E represents their contacts. At time t, the state can be represented by a vector $X(t)$ such that if person i is infected at time t if and only if $X_i(t) = 1$, and if person i is healthy at time t if and only if $X_i(t) = 0$. Infected nodes contaminate their neighbors at rate β, and recover at rate υ. Then

$$\begin{aligned} X_i : 0 \to 1 \quad & \text{at rate } \beta \textstyle\sum_{(j,i)\in E} X_j \\ X_i : 1 \to 0 \quad & \text{at rate } \upsilon \end{aligned}$$

We want to see under such random process, what is epidemic going to be like.

2.2 Analysis results

In this section, we present the analysis results of SIS model, which is without the existence of vaccination or anti-dots, how the disease is going to spread, is the epidemic going to last or die out?

If we don't consider contact graphs, we can use differential equation approach by solving equation (7) and (8). Because we also have $S + I = N$, substituting $I = N - S$ in equation (8). We have

$$\frac{dI}{dt} = \beta(N - I)\frac{I}{N} - \upsilon I = (\beta - \upsilon)I - \frac{\beta}{N}I^2$$

Solving $dI/dt = 0$, we see that there are two possible equilibria for this SIS model, one with $I = 0$ and the other with $I = \frac{\beta-\upsilon}{\beta}N$. Define *basic reproductive number* as $R_0 = \beta/\upsilon$. If $R_0 \leq 1$, then $\lim_{t\to\infty} I(t) \to 0$, which is disease free equilibrium; if $R_0 > 1$, then $\lim_{t\to\infty} I(t) \to N(\beta - \upsilon)/\beta$, which is endemic equilibrium.

In the following section, we present the analysis results on SIS model with contact graphs. There are two natural measures, one is time (how long the epidemic is going to last) and the other is size (how many people are infected).

2.2.1 Epidemic time

One nature measure for epidemic is how much time it takes for a disease to die out. This problem has been studied in the probability community [18], but where it is usually studied on bounded-degree graphs. The most important general result in the context is the existence of epidemic thresholds. For infinite graphs it has shown that there exist two epidemic thresholds $\lambda_1 \leq \lambda_2$. If the infection ratio $\lambda = \beta/\upsilon$ is larger than λ_2, then with positive probability the epidemic can spread and survive at any point of the graph. If $\lambda_1 < \lambda < \lambda_2$, the epidemic survives with positive probability, but every vertex almost surely eventually heals without being reinfected. If $\lambda < \lambda_1$, the epidemic dies out almost surely. (see [18] and [23, 22])

However, for finite graphs, it is easy to see that the infection will eventually die out with probability 1. In this case people say that the infection becomes an epidemic if the time that it takes to die out is super-polynomial in the number of vertices of the the graph.

[12] studies the relationship between topological properties of the graph and epidemic time. Their results are, in finite graphs, (1) a sufficient condition for a quick die out (i.e. the die out time is logarithmic of the size of the graph) is that the ratio of infection rate to cure rate not exceed the spectral radius of the adjacency matrix of the underlying topology graph; (2) a sufficient condition for slow die out (i.e. the die out time is exponential of the size of the graph) is that the ratio of infection rate to cure rate be larger than the isoperimetric constant

associated with the graph. [24] also studies the same problem, the relationship between threshold and graph properties. And they have similar results.

The other effort studying the epidemic time is trying to reduce the epidemic die out time by making recover rate non-uniform among vertices [8]. The motivation for this is we can distribute antidotes unevenly to control epidemics. The more antidotes one gets, the sooner it recovers from infection. With limited amount of antidotes, what is a good way to distribute them in order to make the epidemic die out quickly? Their results show that if we distribute antidotes proportional to vertex's degree, the epidemic will die out quickly.

2.2.2 Epidemic size

Although a lot of work has been done on epidemic time under SIS model, few results are known on epidemic size. [15] studies such problem, the mean number of infections. But they didn't involve the concept of contact graphs. Instead they use differential equation approach, which I think is the same as the contact graph is a complete graph.

2.3 Optimization problems

Since there are two natural measure of epidemic (time and size) under SIS model, it's optimization problem can be characterized as follows. Given a graph $G = (V, E)$, and a number k which represents the amount of resources available (vaccination, anti-dots, anti-virus software, etc.), remove k vertices from graph G to decrease epidemic time (or size) as much as possible.

[8] studied such optimization problem with anti-dots as cure. They showed that by distributing anti-dots to vertices proportional to their degrees can reduce the epidemic time to $O(\log n)$ where n is the total number of population. [9] looked at the same problem using simulation approach. And they came up with similar conclusion.

[11] studied network immunization against virus spread with limited immunization budget using simulation approach. They considered such problem under SIS model, and gave a simple heuristic which out perform the greedy algorithm, where deleting the vertex with maximum degree at each step, in simulation.

2.4 Game theory problems

3 Contact graphs

The underlying contact graph plays an important roll in the spreads of epidemics. Different kinds of graphs can provide different set of properties, which are crucial in the analysis of epidemic problems. More important questions would be what kinds of graphs should we use to model the real word problems? And what kinds of results can we derive from that? A lot of work has been devoted in this area.

3.1 Power-law graphs

Power-law graph is one where the number of nodes with degree k is proportional to $k^{-\gamma}$ for some $\gamma > 1$. Ever since [10] discovers that the Internet AS-level graph exhibits a power law degree distribution, lots of other large scale social networks are discovered to have power-law distribution, even human sexual contact networks [19]. So it is nature to study epidemic problems under power-law graphs, as well as the interesting properties of such graphs.

3.1.1 Robustness and vulnerability

Robustness and vulnerability are important connectivity properties of power-law graph, which basically means power-law graphs are robust under random attack (i.e. randomly delete vertices from the graph) but vulnerable under targeting attack (i.e. choose vertices to delete).

[1] studies the robustness of power-law graph under random errors and targeting attacks. Its simulation results show that power-law graphs is very robust under random errors (i.e. by randomly deleting nodes the graph remains high connectivity) but vulnerable under targeting attacks (i.e. by removing a small fraction of high degree nodes we can break the graph into small pieces).

[7] studies the same problem as [1], but more rigorously. It shows mathematically that (1) if vertices are deleted at random, then as long as any positive proportion remains, the graph induced on the remaining vertices has a component of order of n (which is the total number of vertices) (2) if the deleted vertices are chosen maliciously, a constant fraction less than 1 can be deleted to destroy all large components, and the the vertices they targeted are those added in the early stage of preferential attachment process.

3.1.2 Virus transmission

Since the Internet, WWW, email networks, and human sexual networks are well modeled as power-law graph, it's nature to study virus transmission in such graph. And the standard model in viral infections is SIS model. Let β be virus transmission probability, υ be recovery probability of an infected node. And define an effective spreading rate to be $\lambda = \beta/\upsilon$.

In some graphs, like regular lattice, there is threshold λ_c such that if $\lambda > \lambda_c$ the virus will persist, while if $\lambda < \lambda_c$ the virus will die out quickly. However, [21] shows that for power-law graphs the threshold vanishes, i.e. $\lambda_c = 0$. This implies that on such networks even weakly infectious viruses can spread and prevail.

So [9] studies virus spreading with cures. This means we can eradicate virus from the node to which the cure is applied, but the cure doesn't offer a permanent protection. Its result shows that biased strategies, which cures the hubs with higher probability than the less connected nodes, can restore the threshold, i.e. $\lambda_c > 0$.

Borgs et al studied this problem more rigorously. In [6], they showed virus with a positive rate of spread from a node to its neighbors has a non-vanishing

chance of becoming epidemic. Quantitatively, for a virus with effective spread rate λ, if the infection starts at a typical vertex, then it develops into an epidemic with probability $\lambda^{\Theta\left(\frac{\log(1/\lambda)}{\log\log(1/\lambda)}\right)}$, but on average the epidemic probability is $\lambda^{\Theta(1)}$, which is consistent with the simulation results in [21]. Here becoming an epidemic means the time that it takes for infection to die out is super-polynomial in the number of vertices of the graph, and typical vertex means those with degree less than λ^{-2}.

Furthermore, in [8], Borgs et al rigorously analyzed that by distributing anti-dots proportional to vertex's degree can largely decrease disease die out time.

3.1.3 Other properties

[20] shows preferential connectivity model have constant conductance, which means it is a constant expander.

3.2 Small world graphs

Small world graph proposed by Kleinberg [17, 16] is considered to be a good model for social networks. It models the *small-world phenomenon*. A social network exhibits the small-world phenomenon if, roughly speaking, any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances. Milgrams basic small-world experiment remains one of the most compelling ways to think about the problem. The goal of the experiment was to nd short chains of acquaintances linking pairs of people in the United States who did not know one another. In a typical instance of the experiment, a source person in Nebraska would be given a letter to deliver to a target person in Massachusetts. The source would initially be told basic information about the target, including his address and occupation; the source would then be instructed to send the letter to someone she knew on a rst-name basis in an effort to transmit the letter to the target as efficaciously as possible. Anyone subsequently receiving the letter would be given the same instructions, and the chain of communication would continue until the target was reached. Over many trials, the average number of intermediate steps in a successful chain was found to lie between ve and six, a quantity that has since entered popular culture as the "six degrees of separation principle" [13].

The following is the small world graph model. Start with a set of nodes (representing individuals in the social network) that are identied with the set of lattice points in an $n \times n$ square, $\{(i,j) : i \in \{1,2,\ldots,n\}, j \in \{1,2,\ldots,n\}\}$. Dene the lattice distance between two nodes (i,j) and (k,l) to be the number of lattice steps separating them, $d\left((i,j),(k,l)\right) = |k-i|+|l-j|$. For a universal constant $p \geq 1$, the node u has a directed edge to every other node within lattice distance p (these are its local contacts). For universal constants $q \geq 0$ and $r \geq 0$, we also construct directed edges from u to q other nodes (the long-range contacts) using independent random trials: the ith directed edge from u has endpoint v with probability proportional to $[d(u,v)]^{-r}$.

Based on the above model, Kleinberg answered the following two questions: (1) Why should there exist short chains of acquaintances linking together arbitrary pairs of strangers? (2) Why should arbitrary pairs of strangers be able to nd short chains of acquaintances that link them together? Basically he study "decentralized" algorithms by which individuals, knowing only the locations of their direct acquaintances, attempt to transmit a message from a source to a target along a short path.

And his results to those questions are: (1) There is a constant α_0, depending on p and q but independent of n, so that when $r = 0$, the expected delivery time of any decentralized algorithm is at least $\alpha_0 n^{2/3}$; (2) There is a decentralized algorithm A and a constant α_2, independent of n, so that when $r = 2$ and $p = q = 1$, the expected delivery time of A is at most $\alpha_2 (\log n)^2$; (3) Let $0 \leq r < 2$. There is a constant α_r, depending on p, q, r, but independent of n, so that the expected delivery time of any decentralized algorithm is at least $\alpha_r n^{(2-r)/3}$. Let $r > 2$. There is a constant α_r, depending on p, q, r, but independent of n, so that the expected delivery time of any decentralized algorithm is at least $\alpha_r n^{(r-2)/(r-1)}$. (see [17])

References

[1] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, July 2000.

[2] James Aspnes, Kevin L. Chang, and Aleksandr Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*, 72(6):1077–1093, 2006.

[3] James Aspnes, Navin Rustagi, and Jared Saia. Worm versus alert: Who wins in a battle for control of a large-scale network? In *Principles of Distributed Systems: 11th International Conference, OPODIS 2007*, volume 4878 of *Lecture Notes in Computer Science*, pages 443–456. Springer, Dec 2007.

[4] C. T. Bauch and D. J. D. Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Science*, 101:13391–13394, September 2004.

[5] Chris T. Bauch. Imitation dynamics predict vaccinating behaviour. In *Proceeding of The Royal Society*, 2005.

[6] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. In *Proceedings of SODA 2005*, 2005.

[7] Bela Bollobas and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, May 2004.

[8] Christian Borgs, Jennifer Chayes, Ayalvadi Ganesh, and Amin Saberi. How to distribute antidote to control epidemics. 2008.

[9] Zoltan Dezso and Albert-Laszlo Barabasi. Halting viruses in scale-free networks. *Physical Review E*, 65:055103, 2002.

[10] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Proceeding of SIGCOMM 1999*, 1999.

[11] E. Terzi P. Tsaparas G. Giakkoupis, A. Gionis. Models and algorithms for network immunization. *Technical Report C-2005-75, Department of Computer Science, University of Helsinki*, 2005.

[12] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceeding of INFOCOM 2005*, 2005.

[13] J. Guare. Six degrees of separation: A play. *Vintage Books*, 1990.

[14] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM.

[15] David A. Kessler. Epidemic size in the sis model of endemic infections. *J. Appl. Probab.*, 43:757–778, 2008.

[16] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798), August 2000.

[17] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *in Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 163–170, 2000.

[18] T. M. Liggett. Stochastic interacting systems: contact, voter and exclusion processes. *Springer-Verlag*, 1999.

[19] F. Liljeros, C. Edling, L. Amaral, H. Stanley, and Y. Aberg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[20] Milena Mihail, Christos Papadimitriou, and Amin Saberi. On certain connectivity properties of the internet topology. In *Proceedings of FOCS 2003*, 2003.

[21] Romualdo Pastor-satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks, 2000.

[22] A. Stacey. The existence of an intermediate phase for the contact process on trees. *Ann. Probab.*, 1996.

[23] A. Stacey. The contact process on finite homogeneous trees. *Probab. Theory Related Fields*, 2001.

[24] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *In SRDS*, pages 25–34, 2003.