Chapter 1

Introduction

The process of diffusion is the spread of information or some flow in a network through local transmissions. Many real world applications can be modeled as diffusion processes over networks. Some prominent examples include diseases transmitted among humans, viruses transmitted over computer networks, information/ideas spreading over contact networks, and creation of friendships through social networks. Despite the diversity among these applications, there are fundamental similarities in the mathematical models. Understanding the properties of these applications through mathematical models can help us anticipate, exploit, and control the propagation processes.

Based on the information or nature of the commodity that is flowing, we classify diffusion processes into the following two categories, *positive diffusion* and *negative/harmful diffusion*. In positive diffusion, the information or commodities are useful to the nodes, like innovation and ideas, whereas in negative diffusion, the information or commodities are harmful to the nodes, like diseases and viruses. In positive diffusion, we are interested in analyzing the converging time and designing efficient algorithms for fast diffusion. While in negative diffusion, we are interested in analyzing the converging time and the extent of diffusion processes, as well as designing good intervention strategies. Take the spread of a disease or computer virus as an example. Lots of important questions can be asked. Will it become an epidemic? How much time does it take to become an epidemic? Who will get infected? What's the social cost of the epidemic? Once we understand all these, we can design interventions to control the dynamics. For instance, how do we vaccinate or quarantine the population so that the epidemic is controlled? How do we secure computers to enhance the network resilience?

What polices should be applied with budget constraints (limited vaccines or anti-virus software licenses), how should we distribute resources, and how much can we reduce social cost? Often these interventions can be translated into voluntary directives from government, like take vaccines or stay at home. However, people usually don't adhere to such recommendations. Instead, they make decisions based on their specific utilities and objectives. Such decisions happen in a decentralized manner, which makes game theory a natural approach to study these problems. Moreover, people alter their contacts dynamically. For example, a vaccinated person may increase his/her contacts with friends, due to perceived secure feelings. These behavioral changes have a huge impact on the dynamics and the effectiveness of these interventions, so that "good" intervention strategies might be ineffective, depending on the behavioral changes. All these make the analysis of diffusion processes more interesting and challenging.

In the first half of this dissertation, we concentrate on enabling positive diffusion. More interestingly, we focus on the diffusion processes on dynamically changing networks. The networks can be changed by the diffusion process itself or by an adversary. We introduce the problems in these two types of dynamic networks in detail in Section 1.1 and Section 1.2 respectively. In the second half of this dissertation, we switch gear to controlling negative diffusion. We design both centralized and decentralized strategies to control negative diffusion, introduced in Section 1.3. We further consider the effects of individual behavior changes on the design of control strategies, which is introduced in Section 1.4.

1.1 Diffusion under organic dynamics

Many large-scale, real-world networks such as peer-to-peer networks, the Web, and social networks are highly dynamic with continuously changing topology. The evolution of the network as a whole is typically determined by the decentralized behavior of nodes, i.e., the local topological changes made by the individual nodes (e.g., adding edges between neighbors). The dynamics can be captured as diffusion processes in self-altered networks. Understanding the dynamics of such diffusion processes is critical for both analyzing the underlying stochastic phenomena, e.g., in evolution of social networks, the Web and other real-world networks [40, 109, 123], and designing practical algorithms for associated algorithmic problems, e.g., in resource discovery in distributed networks

[75, 92] or in the analysis of algorithms for the Web [45, 51]. In this thesis, we study the dynamics of network evolution that result from *local* gossip-style processes. Gossip-based processes have recently received significant attention because of their simplicity of implementation, scalability to large network size, and robustness to frequent network topology changes; see, e.g., [54, 83, 84, 47, 82, 80, 106, 43] and the references therein. In a local gossip-based algorithm (e.g., [47]), each node exchanges information with a small number of randomly chosen neighbors in each round.[1] The randomness inherent in the gossip-based protocols naturally provides robustness, simplicity, and scalability.

We present two illustrative applications for our study. First, consider a P2P network, where nodes (computers or end-hosts with IDs/IP addresses) can communicate only with nodes whose IP address are known to them. A basic building block of such a dynamic distributed network is to efficiently discover the IP addresses of all nodes that currently exist in the network. This task, called *resource discovery* [75], is a vital mechanism in a dynamic distributed network with many applications [75, 5]: when many nodes in the system want to interact and cooperate they need a mechanism to discover the existence of one another. Resource discovery is typically done using a local mechanism [75]; in each *round* nodes discover other nodes and this changes the resulting network — new edges are added between the nodes that discovered each other. As the process proceeds, the graph becomes denser and denser and will finally result in a complete graph. Such a process was first studied in [75] which showed that a simple randomized process is enough to guarantee almost-optimal time bounds for the time taken for the entire graph to become complete (i.e., for all nodes to discover all other nodes). Their randomized *Name Dropper* algorithm operates as follows: in each round, each node chooses a random neighbor and sends *all* the IP addresses it knows. Note that while this process is also gossip based the information sent by a node to its neighbor can be extremely large (i.e., of size $\Omega(n)$).

Second, in social networks, nodes (people) discover new nodes through exchanging contacts with their neighbors (friends). Discovery of new nodes changes the underlying network — new edges are added to the network — and the process continues in the

[1] Gossip, in some contexts (see e.g., [80, 82]), has been used to denote communication with a random node in the network, as opposed to only a directly connected neighbor. The former model essentially assumes that the underlying graph is complete, whereas the latter (as assumed here) is more general and applies even to arbitrary graphs. The local gossip process is typically more difficult to analyze due to the dependencies that arise as the network evolves.

changed network. For example, consider the *LinkedIn* network[1], a large social network of professionals on the Web. The nodes of the network represent people and edges are added between people who directly know each other — between direct contacts. Edges are generally undirected, but LinkedIn also allows directed edges, where only one node is in the contact list of another node. LinkedIn allows two mechanisms to discover new contacts. The first can be thought of as a *triangulation* process (see Figure 1.1(a)): A person can introduce two of his friends that could benefit from knowing each other — he can mutually introduce them by giving their contacts. The second can be thought of as a *two-hop* process (see Figure 1.1(b)): If *you* want to acquire a new contact then you can use a shared (mutual) neighbor to introduce yourself to this contact; i.e., the new contact has to be a two-hop neighbor of yours. Both the processes can be modeled via gossip in a natural way and the resulting evolution of the network can be studied. This yields insight on the evolution of the social network over time.

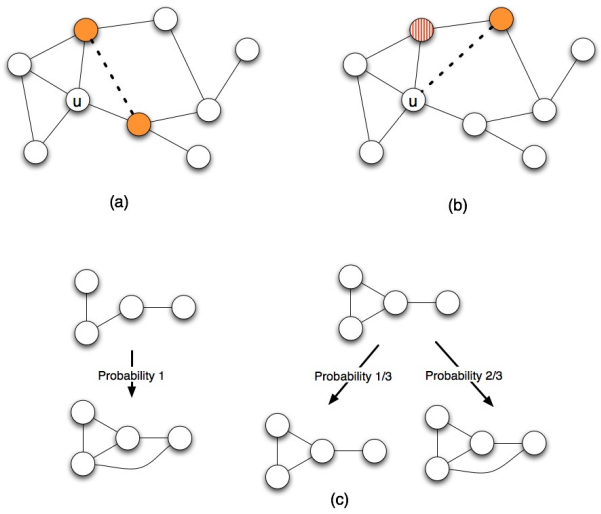

Figure 1.1: (a) Push discovery or triangulation process. (b) Pull discovery or two-hop walk process. (c) Non-monotonicity of the triangulation process – the expected convergence time for the 4-edge graph exceeds that for the 3-edge subgraph.

[1] `http://www.linkedin.com`.

Gossip-based discovery. Motivated by the above applications, we analyze two natural gossip-based discovery processes (also diffusion processes). We assume that we start with an arbitrary undirected connected graph and the process proceeds in synchronous rounds. Communication among nodes occurs only through edges in the network. We further assume that the size of each message sent by a node in a round is at most $O(\log n)$ bits, i.e., the size of an ID.

1. Push discovery (triangulation): In each round, each node chooses two random neighbors and connects them by "pushing" their mutual information to each other. In other words, each node adds an undirected edge between two of its random neighbors; if the two neighbors are already connected, then this does not create any new edge. Note that this process, which is illustrated in Figure 1.1(a), is completely local. To execute the process, a node only needs to know its neighbors; in particular, no two-hop information is needed.

2. Pull discovery (two-hop walk): In each round, each node connects itself to a random neighbor of one of its randomly chosen neighbors, by "pulling" a random neighboring ID from a random neighbor. Alternatively, one can think of each node doing a two-hop random walk and connecting to its destination. This process, illustrated in Figure 1.1(b), can also be executed locally: a node simply asks one of its neighbors v for an ID of one of v's neighbors and then adds an undirected edge to the received contact.

Both the above processes are local in the sense that each node only communicates with its neighbors in any round, and lightweight in the sense that the amortized work done per node is only a constant per round. Both processes are also easy to implement and generally oblivious to the current topology structure, changes or failures. It is interesting also to consider variants of the above processes in directed graphs. In particular, we study the two-hop walk process which naturally generalizes in directed graphs: each node does a two-hop directed random walk and adds a *directed* edge to its destination. We are mainly interested in the time taken by the process to converge to the transitive closure of the initial graph, i.e., till no more new edges can be added.

Our results. We present almost-tight bounds on the number of rounds it takes for the push and pull discovery processes to converge.

- **Undirected graphs:** In Sections 2.2 and 2.3, we show that for *any* undirected n-node graph, both the push and the pull discovery processes converge in $O(n \log^2 n)$ rounds with high probability. We also show that $\Omega(n \log n)$ is a lower bound on the number of rounds needed for almost any n-node graph. Hence our analysis is tight to within a logarithmic factor.

- **Directed graphs:** In Section 2.4, we show that the pull process takes $O(n^2 \log n)$ time for any n-node directed graph, with high probability. We show a matching lower bound for weakly connected graphs, and an $\Omega(n^2)$ lower bound for strongly connected directed graphs. Our analysis indicates that the directionality of edges can greatly impede the resource discovery process.

Applications. The gossip-based discovery processes we study are directly motivated by the two scenarios outlined above, namely algorithms for resource discovery in distributed networks and analyzing how discovery process affects the evolution of social networks. Since our processes are simple, lightweight, and easy to implement, they can be used for resource discovery in distributed networks. The original resource discovery algorithm of [75] was helpful in developing systems like Akamai. Unlike prior algorithms for the discovery problem [75, 92, 91, 5], the amortized work done per node in our processes is only constant per round and hence this can be efficiently implemented in bandwidth and resource-constrained networks (e.g., peer-to-peer or sensor networks). In contrast, the *Name Dropper* algorithm of [75], can transfer up to $\Theta(n)$ information per edge per round and hence may not be scalable for large-scale networks. We note that, however, because there is essentially no restriction on the bandwidth, the number of rounds taken by the *Name Dropper* algorithm is $O(\log^2 n)$. (We note that in our model, $\Omega(n)$ is a trivial lower bound). Our analyses can also give insight into the growth of real-social networks such as LinkedIn, Twitter, or Facebook, that grow in a decentralized way by the local actions of the individual nodes. For example, it can help in predicting the sizes of the immediate neighbors as well as the sizes of the second and third-degree neighbors (e.g., these are listed for every node in LinkedIn).

An estimate of these can help in designing efficient algorithms and data structures to search and navigate the social network.

Technical contributions. Our main technical contribution is a probabilistic analysis of localized gossip-based discovery in arbitrary networks. While our processes can be viewed as graph-based coupon collection processes, one significant distinction with past work in this area [6, 11, 56] is that the graphs in our processes are constantly changing. The dynamics and locality inherent in our process introduces nontrivial dependencies, which makes it difficult to characterize the network as it evolves. A further challenge is posed by the fact that the expected convergence time for the two processes is *not monotonic*; that is, the processes may *take longer* to converge starting from a graph G than starting from a subgraph H of G. Figure 1.1(c) presents a small example illustrating this phenomenon. This seemingly counter-intuitive phenomenon is, however, not surprising considering the fact that the cover time of random walks also share a similar property. One consequence of these hurdles is that analyzing the convergence time for even highly specialized or regular graphs is challenging since the probability distributions of the intermediate graphs are hard to specify. Our lower bound analysis for a specific strongly connected directed graph in Theorem 15 illustrates some of the challenges. In our main upper bound results, we overcome these technical difficulties by presenting a uniform analysis for all graphs, in which we study different local neighborhood structures and show how each lead to rapid growth in the minimum degree of the graph.

1.2 Diffusion under adversarial dynamics

In an adversarial dynamic network, nodes and communication links can appear and disappear at will over time. Emerging networking technologies such as ad hoc, wireless, sensor, mobile, and peer-to-peer networks are inherently dynamic, resource-constrained, and unreliable. This necessitates the development of a solid foundation to design efficient, robust, and scalable algorithms for diffusion processes in adversarial networks, and to understand the power and limitation of distributed computing on such networks. Such a foundation is critical to realize the full potential of these large-scale dynamic communication networks.

As a step towards understanding the fundamental computation power of such dynamic networks, we investigate dynamic networks in which the network topology changes arbitrarily from round to round. We first consider a worst-case model that was introduced by Kuhn, Lynch, and Oshman [89] in which the communication links for each round are chosen by an online adversary, and nodes do not know who their neighbors for the current round are before they broadcast their messages. (Note that in this model, only edges change and nodes are assumed to be fixed.) The only constraint on the adversary is that the networks should be connected at each round. Unlike prior models on dynamic networks, the model of [89] does not assume that the network eventually stops changing and requires that the algorithms work correctly and terminate even in networks that change continually over time.

We study a fundamental diffusion process, information spreading (also known as gossip) in such dynamic network. In gossip, or more generally, k-gossip, there are k pieces of information (or tokens) that are initially present in some nodes and the problem is to disseminate the k tokens to all nodes. By just gossip, we mean n-gossip, where n is the network size. Information spreading is a fundamental primitive in networks which can be used to solve other problems such as leader election.

The focus of this thesis is on the power of *token-forwarding* algorithms, which do not manipulate tokens in any way other than storing and forwarding them. Token-forwarding algorithms are simple, often easy to implement, and typically incur low overhead. In a key result, [89] showed that under their adversarial model, k-gossip can be solved by token-forwarding in $O(nk)$ rounds, but that any deterministic online token-forwarding algorithm needs $\Omega(n \log k)$ rounds. They also proved an $\Omega(nk)$ lower bound for a special class of token-forwarding algorithms, called knowledge-based algorithms. Our main result is a new lower bound on *any* deterministic online token-forwarding algorithm for k-gossip.

- We show that every deterministic online token-forwarding algorithm for the k-gossip problem takes $\Omega(nk/\log n)$ rounds. Our result applies even to centralized (deterministic) token-forwarding algorithms that have a global knowledge of the token distribution.

This result resolves an open problem raised in [89], significantly improving their lower bound, and matching their upper bound to within a logarithmic factor. Our lower

bound also enables a better comparison of token-forwarding with an alternative approach based on network coding due to [72, 73], which achieves a $O(nk/\log n)$ rounds using $O(\log n)$-bit messages (which is not significantly better than the $O(nk)$ bound using token-forwarding), and $O(n+k)$ rounds with large message sizes (e.g., $\Theta(n \log n)$ bits). It thus follows that for large token and message sizes there is a factor $\Omega(\min\{n,k\}/\log n)$ gap between token-forwarding and network coding. We note that in our model we allow only one token per edge per round and thus our bounds hold regardless of the token size.

Our lower bound indicates that one cannot obtain efficient (i.e., subquadratic) token-forwarding algorithms for gossip in the adversarial model of [89]. Furthermore, for arbitrary token sizes, we do not know of any algorithm that is significantly faster than quadratic time. This motivates considering other weaker (and perhaps, more realistic) models of dynamic networks. In fact, it is not clear whether one can solve the problem significantly faster even in an offline setting, in which the network can change arbitrarily each round, but the entire evolution is known to the algorithm in advance. Our next contribution takes a step in resolving this basic question for token-forwarding algorithms.

- We present a polynomial-time offline token-forwarding algorithm that solves the k-gossip problem on an n-node dynamic network in $O(\min\{nk, n^{1.5}\sqrt{\log n}\})$ rounds.
- We also present a polynomial-time offline token-forwarding algorithm that solves the k-gossip problem in a number of rounds within an $O(n^\epsilon)$ factor of the optimal, for any $\epsilon > 0$, assuming the algorithm is allowed to transmit $O(\log n)$ tokens per round.

The above upper bounds show that in the offline setting, token-forwarding algorithms can achieve a time bound that is within $O(\sqrt{n \log n})$ of the information-theoretic lower bound of $\Omega(n+k)$, and that we can approximate the best token-forwarding algorithm to within a $O(n^\epsilon)$ factor, given logarithmic extra bandwidth per edge.

1.3 Controlling negative diffusion

In this section, we motivate our problems using computer virus as an example. However the study of intervention strategies can be easily applied to other fields like epidemiol-

ogy.

Over the recent decades, there has been an explosive growth in the use of personal digital devices of various kinds, which are connected to the Internet through new technologies, such as Bluetooth and Wi-Fi to allow ubiquitous access. This has, unfortunately, been accompanied by significant increase in worm attacks that exploit bugs in these new technologies, and which have new and growing "medium" to spread on - recent attacks, e.g., Cabir and CommWorm, that span multiple networks are expected to become increasingly prevalent in future. While, effective anti-virus software and patches are readily available, the average user is very independent and does not often care to be proactive about installing the most effective anti-virus software, and downloading the latest patches, partially because of the cost of the software and the effort involved, which we refer to as the *security cost*. Indeed, a large fraction of devices are estimated to be without adequate anti-virus protection. If a user does not install protective software, they would incur a cost if his device gets attacked, due to downtime, loss of revenue, and cost of re-installing systems; we refer to these as the *infection cost*. If enough other nodes in the network are secured, the likelihood of a specific device getting infected would go down (as a result of the "herd immunity"), leading to a natural game theoretic scenario. A number of different non-cooperative game formulations have been developed to study this basic problem, e.g., [15, 16, 35, 67, 71, 94, 126]; one issue with many of these formulations is that they involve utility functions that require quite a lot of non-local information to compute, and it is not clear how implementable such games might be.

In this thesis, we present a generalized network security game model GNS(d), which incorporates arbitrary contact networks through an undirected graph G and heterogeneous nodes with individual security and infection costs. Our model is parametrized by network locality parameter d, which represents the distance within the network that a given infection can spread. Equivalently, the parameter d in the game GNS(d) could represent the extent of neighborhood information that is available to a node when making strategic security decisions, which is a departure from earlier models which require global information for making decisions. Qualitatively, we consider three important cases with respect to d. The case $d = 1$, which we refer to as the *local infection model*, is most well-suited for ad hoc wireless networks and social networks, when certain actions initiated by an insecure node could adversely affect immediate neighbors, friends,

or email contacts. For this case, our model can be viewed as a variant of the IDS model of [81]. The case $d = \infty$, which we refer to as the *global infection model*, is most well-suited for the highly infectious worms and viruses in the Internet that can be transmitted in an hop-unlimited manner through unsuspecting insecure nodes, under the assumption that individual nodes have complete information. Our GNS(∞) model is a generalization of the elegant model of [15]. The intermediate case $1 < d < \infty$ applies to the majority of network security hazards where the transmission may be hop-limited and nodes may only have limited local information about the topology and security decisions taken by others. Our main results are the following.

- **Existence of pure Nash equilibria (NE):** We show that the locality parameter d plays a significant role in the structure of the resulting games. Both the extremes of GNS(1) and GNS(∞) turn out to be ordinal potential games, and a pure NE can be computed by best response dynamics – that is, every sequence of best response steps by the individual players converges to a pure NE. However, for every d in the range $(1, \infty)$, there exists an instance of GNS(d) that does not have a pure equilibrium. The price of anarchy for a GNS(1) game is at most the maximum degree of the contact graph, while that for GNS(∞) is inversely proportional to the vertex expansion of the contact graph.

- **Complexity of computing pure NE:** While there is a simple combinatorial characterization for the existence of pure NE in GNS(d) for all d, we show that for $1 < d < \infty$, deciding if an arbitrary instance of GNS(d) has a pure NE is NP-complete. For GNS(1), we show that finding a pure NE of least cost is NP-complete; a corresponding result for GNS(∞) is in [15].

- **Approximating the social optimum:** We show that computing the social optimum is NP-complete for a GNS(d) game, for any d; the case of $d = \infty$ was shown by [15]. We design a general framework for finding a strategy vector for the players in polynomial time, whose cost is at most $2d$ times that of the optimal, for any fixed d. In particular, this implies that for $d = 1$, we obtain a 2-approximation. For $d = \infty$, we provide a different algorithm within the framework that yields an $O(\log n)$-approximation, where n is the number of nodes in the network; this improves on the approximation bound of $O(\log^{1.5} n)$ of [15] achieved for a special case of the GNS(∞).

- **Empirical results:** We study the characteristics of NE empirically in two distinct classes of graphs: random geometric graphs and power law graphs. For $d = 1$, we find that the convergence time for best response is sub-linear in the number of nodes in both the classes of graphs, while it is linear for $d = \infty$. Also, for $d = 1$, we find that the cost of the pure NE obtained is very close to that of the social optimum, indicating that the pure NE obtained in real-world networks approximate social optimum very well. For $d = \infty$, we observe that there may be a significant gap between the cost of the pure NE and that of the social optimum, even for small networks. Finally, we study the performance of our approximation algorithms for the social optimum, and find that the approximation guarantees in practice are much smaller than our theoretical bounds.

Pure NE represent stable operating points for a system with selfish users. Therefore, for a network planner, understanding and controlling the quality of equilibria reached is an important issue. Our results suggest locality characteristics of the network or the amount of information available to the strategic network players have a significant impact on the existence of equilibria. The non-monotonicity in the existence of NE, with respect to d, is somewhat surprising and suggests a closer examination of the impact of information on pure NE in such games. While our theoretical analysis indicates that pure NE may be significantly inferior to the optimum in terms of social optimum in the worst-case, our experiments suggest that for real-world network models pure NE obtained by uncoordinated best response dynamics have low cost relative to the social optimum, especially in the case of $d = 1$. Additionally, our results on the price of anarchy suggest natural heuristics to aid a network planner in enforcing efficient equilibria. Finally, the approximations achieved by our approximation algorithms, both in theory and experiments, indicate that our proposed algorithms are viable candidates wherever centralized decisions can be made on network protection mechanisms.

1.4 Controlling negative diffusion in the presence risk behavior changes

The study in Section 1.3 assumes that the behavior of each individual remains the same before and after taking interventions. However, this is not an accurate assumption in some real world scenarios. For instance, people expose themselves more to the public

after taking vaccinations. This behavior change is often referred as risk behavior change. And it is very common, specially in epidemiology. Since vaccination is not 100% reliable, this kind of behavior change has the potential to increase the likelihood of disease transmission. In our study, it is important to consider the impact of risk behavior to good intervention strategies. In our discussion below, we use disease transmission as example, but risk behavior is not limited to epidemiology.

For many diseases, such as influenza and HIV, prophylactic interventions using anti-virals and vaccinations are commonly used to control the spread of the diseases, and are usually universally recommended, barring individual constraints. Recent studies have shown significant benefits of anti-retrovirals for reducing the spread of HIV [61]. Such treatments have varying levels of efficacy (25-75% in the case of HIV [61, 70] and between 10-80% in the case of influenza [1], depending on the demographics and the specifics of the flu strain). However, people are not very well aware of this limitation, and studies often over-estimate the efficacy of vaccines [114]. Indeed, the perceived protection from infection might cause behavioral changes, leading to an increase in contact by a treated individual; such a behavioral change following vaccination could also be a natural evolutionary response [86, 105], and has also been documented recently in the context of flu vaccines [115]. Regardless of the underlying reasons, failure of prophylactic interventions in conjunction with increased social behavior can have significant unexpected effects on the disease dynamics. In a series of important papers [37, 98] Blower and her collaborators demonstrated that risk behavior change, in the context of HIV vaccinations, could lead to perverse outcomes. Subsequently, several independent studies have confirmed this phenomenon of perversity in the use of HIV vaccines and anti-virals, and vaccines for the human papillomavirus (HPV) [119, 36, 128, 124, 70, 121, 13, 52, 74, 100, 44, 64].[1]

A fundamental question in mathematical epidemiology is to determine the fraction of the population that needs to be vaccinated or treated with anti-virals in order to minimize the impact of the disease, especially when the supply is limited. Modern epidemiological analysis is largely based on an elegant class of models, called SIR

[1] The phenomenon of an increase in risky behavior following protection is also referred to as "moral hazard" and has been studied extensively in a number of areas, such as insurance (e.g., [103]); in the epidemiology literature, this is referred to more commonly as "risk behavior" (e.g., [37]), and we will fix on this terminology for most of the thesis.

(susceptible-infected-recovered), which was first formulated by Reed and Frost in the 1920s, and developed over the subsequent decades. The SIR model and its variants have been highly influential in the study of epidemics [129, 97, 99, 79, 69, 98]. These models, however, do not attempt to capture the rich structure of the contact network over which interactions occur. Network structure has a direct effect on both the spread of diseases as well as the nature of interactions, which has been observed by a number of researchers, e.g. [108, 76]. In the emerging area of contact network epidemiology, an underlying contact graph captures the patterns of interactions which lead to the transmission of a disease [113, 55, 95, 101, 102, 110, 127, 67]. Many studies have predicted the spread of diseases through networks using mathematical analysis or simulations. As we have argued above, moral-hazarding/risky behavior clearly plays an important role in the effectiveness of such interventions. While the impact of risky behavior on prophylactic treatments has been studied in previous work, the extent of the perversity and its dependence on network structure as well as the precise nature of the behavior change has remained largely unknown. [1]

In this thesis, we study the impact of risk behavior change on the spread of diseases in networks and observe a rich and complex structure dependent both on the underlying network characteristics as well as the nature of the change in behavior. We use a discrete-time SIR model of disease transmission on a contact network. The contact network is an undirected graph with each edge having a certain probability of disease transmission. An infected node is assumed to recover in one time step. We consider both uniform random vaccination (where each node is vaccinated independently with the same probability) as well as targeted vaccinations (where nodes are vaccinated based on their degree of connectedness). Vaccines are assumed to fail uniformly and randomly.[2] We model risk behavior change by an increase in the disease transmission probability. A significant aspect of our work is the consideration of the "sidedness" of risk behavior change. We classify risk behavior as one-sided or two-sided based on whether the increase in disease transmission probability requires an increase in

[1] Similar issues arise in the context of the spread of malware through infected computers. Several studies, e.g., [2], have found that computer and smart-phone users do not relate bot infections to risky behavior, such as downloading spam mails, though a large fraction of users have updated anti-virus software. It is plausible that such phenomena can also be associated with risky behavior in many cases.

[2] Though we focus on vaccinations and disease transmission, the basic results apply to other prophylactic treatments such as anti-virals, and other phenomena such as malware spread.

risk behavior of both the infector and the infectee or just the infector. As examples: influenza (H1N1) may be modeled as a one-sided disease since a vaccinated individual may be motivated to behave more riskily (going to crowded places, traveling on planes etc.,) thus increasing the chance of infecting all the individual comes in contact with; whereas AIDS (HIV) may be modeled as a two-sided disease since the increase in disease transmission requires both the individuals participating in the interaction to engage in risky behavior. Of course, these examples are simplistic and most diseases have elements of both one-sided as well as two-sided risk behavior.

Our main findings are threefold.

- First, we find that the *severity of the epidemic varies non-monotonically as a function of the vaccinated fraction*. The specific dynamics depend on the nature of risky behavior, as well as the efficacy of the vaccine (the less reliable the vaccine, the greater the non-monotonicity) and the contagiousness of the disease, but in general, we observe that increased vaccination does not immediately imply less severity; in some cases, the severity could increase by as much as a factor of two.

- Second, we find that *one-sided risk behavior change leads to perverse outcomes at low levels of vaccination, while two-sided risk behavior change leads to perverse outcomes at high levels of vaccination*. Our analysis indicates that effective prophylactic interventions against diseases with one-sided risk behavior change need to have sufficiently high coverage; on the other hand, for diseases with two-sided risk behavior change, it is essential to combine prophylactic treatments with education programs aimed at reducing risky behavior.

- Our third and, perhaps, most surprising finding is that *interventions that target highly connected individuals can be strictly worse than random interventions* for the same level of coverage and that this phenomenon occurs both for one-sided as well as two-sided risk behavior change. Given prior work on targeting vaccine distributions, this finding flies in the face of intuition that expects that targeted vaccination would confer greater benefits.

Our results have direct implications for public policy on containing epidemic spread through prohylactic interventions. Implications of risk behavior in public health have been examined earlier, e.g. [37, 98]. These prior studies are based on differential

equation models, which divide the population into a fixed set of groups and model the interaction between different groups in a uniform way. The epidemic spread is then characterized by the "reproductive number," denoted by R_0, with the expected epidemic size exhibiting a threshold behavior in terms of R_0. In contrast, we use a network model that captures the fine structure of interactions between (an arbitrary number of) individuals rather than (a fixed set of) groups, and find that the network structure has a significant impact on the resulting dynamics. The heterogenous network model extends to a larger range of real-life situations but the increased fidelity comes at a price. The outcomes are more complicated and varied and the general approach of lowering R_0 does not appear to be directly applicable. Another new contribution of our study is the focus on the sidedness inherent in risk behavior change, which has not been considered before. Prior research has implicitly assumed one-sided risk behavior change where vaccinated individuals engage in risky behavior increasing the chances of infection of those they come in contact with. This work explicitly treats both one-sided and two-sided risk behavior changes and shows that their differing impact needs to be considered in public intervention policies.

1.5 Overview

In this thesis, we design efficient algorithms to enable positive diffusion and good intervention strategies to control negative diffusion. In Chapter 2, we study two nature diffusion processes under organic dynamics, and show an almost tight upper bound for both of these processes. In Chapter 3, we study similar problems as in Chapter 2, but under adversarial dynamics. We show an lower bound for any token-forwarding algorithms under online adversarial model, and design two efficient algorithms under offline adversarial model. In Chapter 4, we study both centralized and decentralized intervention strategies. We give an $O(\log n)$ approximation algorithm for optimal centralized intervention strategy. Then we show the existence of intervention strategies in decentralized settings and compare their costs with the optimal centralized strategy. In Chapter 5, we extend the study in Chapter 4 to the presence of risk behavior changes, and observe two interesting phenomena: 1) less interventions can be more effective, and 2) targeted intervention strategy can be worse than random intervention strategy. We conclude in Chapter 6.