Chapter 7

Information Diffusion in Social Media

In February 2013, during the third quarter of Super Bowl XLVII, a power outage stopped the game for 34 minutes. Oreo, a sandwich cookie company, tweeted during the outage: "Power out? No Problem, You can still dunk it in the dark." The tweet caught on almost immediately, reaching nearly 15,000 retweets and 20,000 likes on Facebook in less than two days. A simple tweet diffused into a large population of individuals. It helped the company gain fame with minimum cost in an environment where companies spent as much as $4 million to run a 30-second ad. This is an example of *information diffusion*.

Information diffusion is a field encompassing techniques from a plethora of sciences. In this chapter, we discuss methods from fields such as sociology, epidemiology, and ethnography, which can help social media mining. Our focus is on techniques that can model information diffusion.

Societies provide means for individuals to exchange information through various channels. For instance, people share knowledge with their immediate network (friends) or broadcast it via public media (TV, newspapers, etc.) throughout the society. Given this flow of information, different research fields have disparate views of what is an information diffusion process. We define information diffusion as the *process by which a piece of information (knowledge) is spread and reaches individuals* through interactions. The diffusion process involves the following three elements:

1. **Sender(s).** A sender or a small set of senders initiate the information

217

diffusion process.

2. **Receiver(s).** A receiver or a set of receivers receive diffused information. Commonly, the set of receivers is much larger than the set of senders and can overlap with the set of senders.

3. **Medium.** This is the medium through which the diffusion takes place. For example, when a rumor is spreading, the medium can be the personal communication between individuals.

This definition can be generalized to other domains. In a disease-spreading process, the disease is the analog to the information, and infection can be considered a diffusing process. The medium in this case is the air shared by the infecter and the infectee. An information diffusion can be interrupted. We define the process of interfering with information diffusion by expediting, delaying, or even stopping diffusion as *intervention*.

Individuals in online social networks are situated in a network where they interact with others. Although this network is at times unavailable or unobservable, the information diffusion process takes place in it. Individuals facilitate information diffusion by making individual decisions that allow information to flow. For instance, when a rumor is spreading, individuals decide if they are interested in spreading it to their neighbors. They can make this decision either dependently (i.e., depending on the information they receive from others) or independently. When they make dependent decisions, it is important to gauge the level of dependence that individuals have on others. It could be *local dependence*, where an individual's decision is dependent on all of his or her immediate neighbors (friends) or *global dependence*, where all individuals in the network are observed before making decisions.

Local and Global Dependence

In this chapter, we present in detail four general types of information diffusion: *herd behavior*, *information cascades*, *diffusion of innovation*, and *epidemics*.

Herd behavior takes place when individuals observe the actions of *all* others and act in an aligned form with them. An information cascade describes the process of diffusion when individuals merely observe their immediate neighbors. In information cascades and herd behavior, the network of individuals is observable; however, in herding, individuals decide based on global information (global dependence); whereas, in information
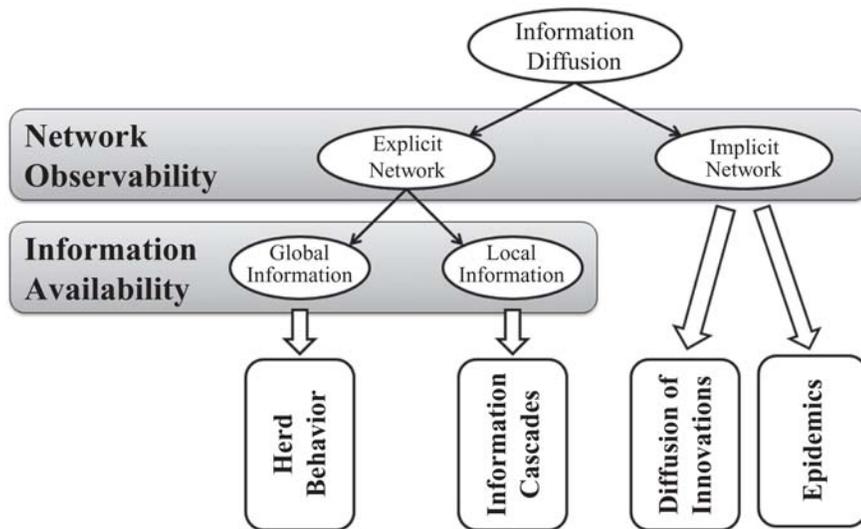
Figure 7.1: Information Diffusion Types.

cascades, decisions are made based on knowledge of immediate neighbors (local dependence).

Diffusion of innovations provides a bird's-eye view of how an innovation (e.g., a product, music video, or fad) spreads through a population. It assumes that interactions among individuals are unobservable and that the sole available information is the rate at which products are being adopted throughout a certain period of time. This information is particularly interesting for companies performing market research, where the sole available information is the rate at which their products are being bought. These companies have no access to interactions among individuals. Epidemic models are similar to diffusion of innovations models, with the difference that the innovation's analog is a pathogen and adoption is replaced by infection. Another difference is that in epidemic models, individuals do not decide whether to become infected or not and infection is considered a random natural process, as long as the individual is exposed to the pathogen. Figure 7.1 summarizes our discussion by providing a decision tree of the information diffusion types.

7.1 Herd Behavior

Consider people participating in an online auction. Individuals are connected via the auction's site where they cannot only observe the bidding behaviors of others but can also often view profiles of others to get a feel for their reputation and expertise. Individuals often participate actively in online auctions, even bidding on items that might otherwise be considered unpopular. This is because they trust others and assume that the high number of bids that the item has received is a strong signal of its value. In this case, *herd behavior* has taken place.

Herd behavior, a term first coined by British surgeon Wilfred [283], describes when a group of individuals performs actions that are aligned without previous planning. It has been observed in flocks, herds of animals, and in humans during sporting events, demonstrations, and religious gatherings, to name a few examples. In general, any herd behavior requires two components:

1. connections between individuals

2. a method to transfer behavior among individuals or to observe their behavior

Individuals can also make decisions that are aligned with others (mindless decisions) when they conform to social or peer pressure. A well-known example is the set of experiments performed by Solomon Asch during the 1950s [17]. In one experiment, he asked groups of students to participate in a vision test where they were shown two cards (Figure 7.2), one with a single line segment and one with three lines, and told to match the line segments with the same length.

Solomon Asch Conformity Experiment

Each participant was put into a group where all the other group members were actually collaborators with Asch, although they were introduced as participants to the subject. Asch found that in control groups with no pressure to conform, in which the collaborators gave the correct answer, only 3% of the subjects provided an incorrect answer. However, when participants were surrounded by individuals providing an incorrect answer, up to 32% of the responses were incorrect.

In contrast to this experiment, we refer to the process in which individuals *consciously* make decisions aligned with others by observing the decisions of other individuals as *herding* or *herd behavior*. In theory, there is

Figure 7.2: Solomon Asch Experiment. Participants were asked to match the line on the left card to the line on the right card that has the exact same length.

no need to have a network of people. In practice, there is a network, and this network is close to a complete graph, where nodes can observe at least most other nodes. Consider this example of herd behavior.

Example 7.1. *Diners Example [23].* *Assume you are visiting a metropolitan area that you are not familiar with. Planning for dinner, you find restaurant **A** with excellent reviews online and decide to go there. When arriving at **A**, you see that **A** is almost empty and that restaurant **B**, which is next door and serves the same cuisine, is almost full. Deciding to go to **B**, based on the belief that other diners have also had the chance of going to **A**, is an example of herd behavior.*

In this example, when **B** is getting more and more crowded, herding is taking place. Herding happens because we consider *crowd intelligence* trustworthy. We assume that there must be private information not known to us, but known to the crowd, that resulted in the crowd preferring restaurant B over A. In other words, we assume that, given this private information, we would have also chosen B over A.

In general, when designing a herding experiment, the following four conditions need to be satisfied:

1. There needs to be a decision made. In this example, the decision involves going to a restaurant.

2. Decisions need to be in sequential order.

3. Decisions are not mindless, and people have private information that helps them decide.

4. No message passing is possible. Individuals do not know the private information of others, but can infer what others know from what they observe from their behavior.

Anderson and Holt [11, 12] designed an experiment satisfying these four conditions, in which students guess whether an urn containing red and blue marbles is majority red or majority blue. Each student had access to the guesses of students beforehand. Anderson and Holt observed a herd behavior where students reached a consensus regarding the majority color over time. It has been shown [78] that Bayesian modeling is an effective technique for demonstrating why this herd behavior occurs. Simply put, computing conditional probabilities and selecting the most probable majority color result in herding over time. We detail this experiment and how conditional probabilities can explain why herding takes place next.

7.1.1   Bayesian Modeling of Herd Behavior

In this section, we show how Bayesian modeling can be used to explain herd behavior by describing in detail the urn experiment devised by Anderson and Holt [11, 12]. In front of a large class of students, there is an urn that has three marbles in it. These marbles are either blue (B) or red (R), and we are guaranteed to have at least one of each color. So, the urn is either majority blue (B,B,R) or majority red (R,R,B). We assume the probability of being either majority blue or majority red is 50%. During the experiment, each student comes to the urn, picks one marble, and checks its color in private. The student predicts majority blue or red, writes the prediction on the blackboard (which was blank initially), and puts the marble back in the urn. Other students cannot see the color of the marble taken out, but can see the predictions made by the students regarding the majority color and written on the blackboard. Let the BOARD variable denote the sequence of predictions written on the blackboard. So, before the first student, it is

BOARD: {}

We start with the first student. If the marble selected is red, the prediction will be majority red; if blue, it will be majority blue. Assuming it was blue, on the board we have

BOARD: {B}

The second student can pick a blue or a red marble. If blue, he also predicts majority blue because he knows that the previous student must have picked blue. If red, he knows that because he has picked red and the first student has picked blue, he can randomly assume majority red or blue. So, after the second student we either have

BOARD: {B,B} or BOARD: {B,R}

Assume we end up with BOARD: {B, B}. In this case, if the third student takes out a red ball, the conditional probability is higher for majority blue, although she observed a red marble. Hence, a herd behavior takes place, and on the board, we will have BOARD: {B,B,B}. From this student and onward, independent of what is being observed, everyone will predict majority blue. Let us demonstrate why this happens based on conditional probabilities and our problem setting. In our problem, we know that the first student predicts majority blue if $P(\text{majority blue}|student's\ obervation) > 1/2$ and majority red otherwise. We also know from the experiments setup that

$$P(\text{majority blue}) \ = \ P(\text{majority red}) = 1/2, \tag{7.1}$$
$$P(\text{blue}|\text{majority blue}) \ = \ P(\text{red}|\text{majority red}) = 2/3. \tag{7.2}$$

Let us assume that the first student observes blue; then,

$$P(\text{majority blue}|\text{blue}) \ = \ \frac{P(\text{blue}|\text{majority blue})P(\text{majority blue})}{P(\text{blue})} \tag{7.3}$$

$$P(\text{blue}) \ = \ P(\text{blue}|\text{majority blue})P(\text{majority blue})$$
$$+ P(\text{blue}|\text{majority red})P(\text{majority red}) \tag{7.4}$$
$$= \ 2/3 \times 1/2 + 1/3 \times 1/2 = 1/2. \tag{7.5}$$

Therefore, $P(\text{majority blue}|\text{blue}) = \frac{2/3 \times 1/2}{1/2} = 2/3$. So, if the first student picks blue, she will predict majority blue, and if she picks red, she will predict majority red. Assuming the first student picks blue, the same argument holds for the second student; if blue is picked, he will also predict majority blue. Now, in the case of the third student, assuming she has picked red, and having BOARD: {B,B} on the blackboard, then,

$$
\begin{aligned}
P(\text{majority blue}|\text{blue, blue, red}) &= \frac{P(\text{blue,blue,red}|\text{majority blue})}{P(\text{blue,blue,red})} \\
&\quad \times P(\text{majority blue}) \quad\quad (7.6) \\
P(\text{blue, blue, red}|\text{majority blue}) &= 2/3 \times 2/3 \times 1/3 = 4/27 \quad\quad (7.7) \\
P(\text{blue, blue, red}) &= P(\text{blue, blue, red}|\text{majority blue}) \\
&\quad \times P(\text{majority blue}) \\
&\quad + P(\text{blue, blue, red}|\text{majority red}) \\
&\quad \times P(\text{majority red}) \quad\quad (7.8) \\
&= (2/3 \times 2/3 \times 1/3) \times 1/2 \\
&\quad + (1/3 \times 1/3 \times 2/3) \times 1/2 = 1/9.
\end{aligned}
$$

Therefore, $P(\text{majority blue}|\textit{blue,blue,red}) = \frac{4/27 \times 1/2}{1/9} = 2/3$. So, the third student predicts majority blue even though she picks red. Any student after the third student also predicts majority blue regardless of what is being picked because the conditional remains above 1/2. Note that the urn can in fact be majority red. For instance, when $\textit{blue, blue, red}$ is picked, there is a $1 - ^2/_3 = ^1/_3$ chance that it is majority red; however, due to herding, the prediction could become incorrect. Figure 7.3 depicts the herding process. In the figure, rectangles represent the board status, and edge values represent the observations. Dashed arrows depict transitions between states that contain the same statistical information that is available to the students.

7.1.2 Intervention

As herding converges to a consensus over time, it is interesting how one can intervene with this process. In general, intervention is possible by providing private information to individuals that was not previously available. Consider an urn experiment where individuals decide on majority red over time. Either (1) a private message to individuals informing them that the
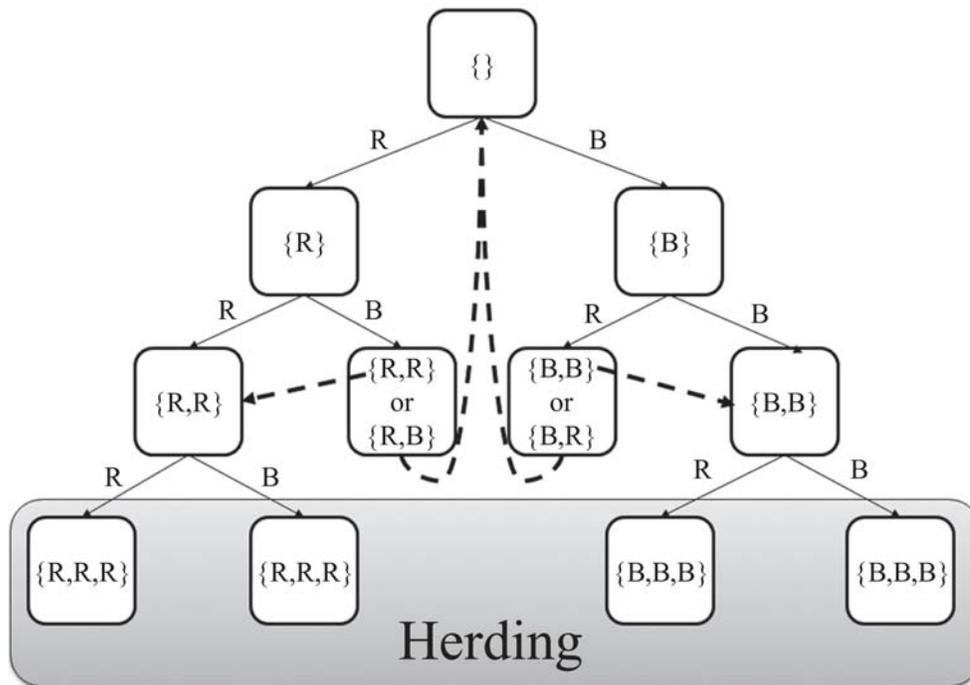
Figure 7.3: Urn Experiment. Rectangles represent student predictions written on the blackboard, and edge values represent what the students observe. Rectangles are filled with the most likely majority, computed from conditional probabilities.

urn is majority blue or (2) writing the observations next to predictions on the board stops the herding and changes decisions.

7.2 Information Cascades

In social media, individuals commonly repost content posted by others in the network. This content is often received via immediate neighbors (friends). An *information cascade* occurs as information propagates through friends.

Formally, an information cascade is defined as a piece of information or decision being cascaded among a set of individuals, where (1) individuals are connected by a network and (2) individuals are only observing

decisions of their immediate neighbors (friends). Therefore, cascade users have less information available to them compared to herding users, where almost all information about decisions are available.

There are many approaches to modeling information cascades. Next, we introduce a basic model that can help explain information cascades.

7.2.1 Independent Cascade Model (ICM)

In this section, we discuss the independent cascade model (ICM) [146] that can be utilized to model information cascades. Variants of this model have been discussed in the literature. Here, we discuss the one detailed by Kempe et al. [146]. Interested readers can refer to the bibliographic notes for further references. Underlying assumptions for this model include the following:

- The network is represented using a directed graph. Nodes are actors and edges depict the communication channels between them. A node can only influence nodes that it is connected to.

- Decisions are binary – nodes can be either *active* or *inactive*. An active nodes means that the node decided to adopt the behavior, innovation, or decision.

- A node, once activated, can activate its neighboring nodes.

- Activation is a *progressive* process, where nodes change from inactive to active, but not vice versa.[1]

Sender-Centric Model

Considering nodes that are active as senders and nodes that are being activated as receivers, in the independent cascade model (ICM) senders activate receivers. Therefore, ICM is denoted as a *sender-centric* model. In this model, the node that becomes active at time t has, in the next time step $t+1$, one chance of activating each of its neighbors. Let v be an active node at time t. Then, for any neighbor w, there is a probability $p_{v,w}$ that node w gets activated at $t+1$. A node v that has been activated at time t has a single chance of activating its neighbor w and that activation can only happen at $t+1$. We start with a set of active nodes and we continue until no further activation is possible. Algorithm 7.1 details the process of the ICM model.

---

[1]This assumption can be lifted [146].

Algorithm 7.1 Independent Cascade Model (ICM)

Require: Diffusion graph $G(V, E)$, set of initial activated nodes A_0, activation probabilities $p_{v,w}$

1: **return** Final set of activated nodes A_∞
2: $i = 0$;
3: **while** $A_i \neq \{\}$ **do**
4:
5:　　$i = i + 1$;
6:　　$A_i = \{\}$;
7:　　**for all** $v \in A_{i-1}$ **do**
8:　　　　**for all** w neighbor of $v, w \notin \cup_{j=0}^{i} A_j$ **do**
9:　　　　　　rand = generate a random number in [0,1];
10:　　　　　　**if** rand $< p_{v,w}$ **then**
11:　　　　　　　　activate w;
12:　　　　　　　　$A_i = A_i \cup \{w\}$;
13:　　　　　　**end if**
14:　　　　**end for**
15:　　**end for**
16: **end while**
17: $A_\infty = \cup_{j=0}^{i} A_j$;
18: Return A_∞;

Example 7.2. *Consider the network in Figure 7.4 as an example. The network is undirected; therefore, we assume $p_{v,w} = p_{w,v}$. Since it is undirected, for any two vertices connected via an edge, there is an equal chance of one activating the other. Consider the network in step 1. The values on the edges denote $p_{v,w}$'s. The ICM procedure starts with a set of nodes activated. In our case, it is node v_1. Each activated node gets one chance of activating its neighbors. The activated node generates a random number for each neighbor. If the random number is less than the respective $p_{v,w}$ of the neighbor (see Algorithm 7.1, lines 9–11), the neighbor gets activated. The random numbers generated are shown in Figure 7.4 in the form of inequalities, where the left-hand side is the random number generated and the right-hand side is the $p_{v,w}$. As depicted, by following the procedure after five steps, five nodes get activated and the ICM procedure converges.*
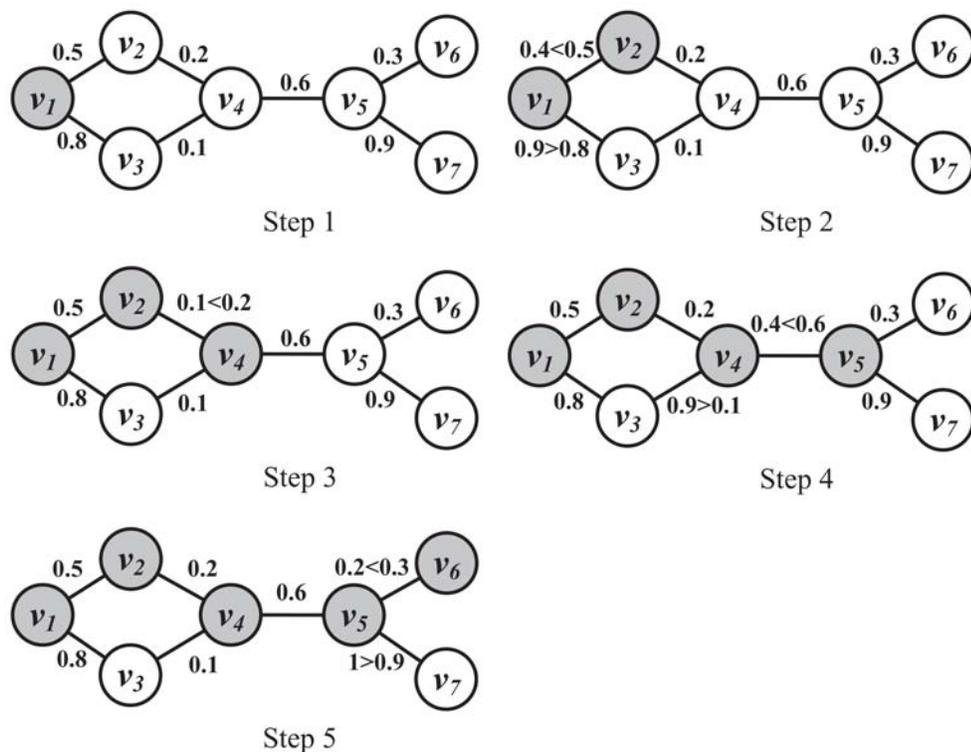
Figure 7.4: Independent Cascade Model (ICM) Simulation. The numbers on the edges represent the weights $p_{v,w}$. When there is an inequality, the activation condition is checked. The left number denotes the random number generated, and the right number denotes weight $p_{v,w}$.

Clearly, the ICM characterizes an information diffusion process.[2] It is sender-centered, and once a node is activated, it aims to activate all its neighboring nodes. Node activation in ICM is a probabilistic process. Thus, we might get different results for different runs.

One interesting question when dealing with the ICM model is that given a network, how to activate a small set of nodes initially such that the final number of activated nodes in the network is maximized. We discuss this next.

---

[2]See [112] for an application in the blogosphere.

7.2.2 Maximizing the Spread of Cascades

Consider a network of users and a company that is marketing a product. The company is trying to advertise its product in the network. The company has a limited budget; therefore, not all users can be targeted. However, when users find the product interesting, they can talk with their friends (immediate neighbors) and market the product. Their neighbors, in turn, will talk about it with their neighbors, and as this process progresses, the news about the product is spread to a population of nodes in the network. The company plans on selecting a set of initial users such that the size of the final population talking about the product is *maximized*.

Formally, let S denote a set of initially activated nodes (seed set) in ICM. Let $f(S)$ denote the number of nodes that get ultimately activated in the network if nodes in S are initially activated. For our ICM example depicted in Figure 7.4, $|S| = 1$ and $f(S) = 5$. Given a budget k, our goal is to find a set S such that its size is equal to our budget $|S| = k$ and $f(S)$ is maximized.

Since the activations in ICM depend on the random number generated for each node (see line 9, Algorithm 7.1), it is challenging to determine the number of nodes that ultimately get activated $f(S)$ for a given set S. In other words, the number of ultimately activated individuals can be different depending on the random numbers generated. ICM can be made deterministic (nonrandom) by generating these random numbers in the beginning of the ICM process for the whole network. In other words, we can generate a random number $r_{u,w}$ for any connected pair of nodes. Then, whenever node v has a chance of activating u, instead of generating the random number, it can compare $r_{u,w}$ with $p_{v,w}$. Following this approach, ICM becomes deterministic, and given any set of initially activated nodes S, we can compute the number of ultimately activated nodes $f(S)$.

Before finding S, we detail properties of $f(S)$. The function $f(S)$ is nonnegative because for any set of nodes S, in the worst case, no node gets activated. It is also monotone:

$$f(S \cup \{v\}) \geq f(S). \tag{7.9}$$

This is because when a node is added to the set of initially activated nodes, it either increases the number of ultimately activated nodes or keeps them the same. Finally, $f(S)$ is submodular. A set function f is submodular if for any finite set N, **Submodular function**

$$\forall S \subset T \subset N, \forall v \in N \setminus T, f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T). \tag{7.10}$$

The proof that function f is submodular is beyond the scope of this book, but interested readers are referred to [146] for the proof. So, f is non-negative, monotone, and submodular. Unfortunately, for a submodular non-negative monotone function f, finding a k element set S such that $f(S)$ is maximized is an NP-hard problem [146]. In other words, we know no efficient algorithm for finding this set.[3] Often, when a computationally challenging problem is at hand, approximation algorithms come in handy. In particular, the following theorem helps us approximate S.

Theorem 7.1 (Kempe et al. [146]). *Let f be a (1) non-negative, (2) monotone, and (3) submodular set function. Construct k-element set S, each time by adding node v, such that $f(S \cup \{v\})$ (or equivalently, $f(S \cup \{v\}) - f(s)$) is maximized. Let S^{Optimal} be the k-element set such that f is maximized. Then $f(S) \geq (1 - \frac{1}{e})f(S^{\text{Optimal}})$.*

This theorem states that by constructing the set S greedily one can get at least a $(1 - 1/e) \approx 63\%$ approximation of the optimal value. Algorithm 7.2 details this greedy approach. The algorithm starts with an empty set S and adds node v_1, which ultimately activates most other nodes if activated. Formally, v_1 is selected such that $f(\{v_1\})$ is the maximum. The algorithm then selects the second node v_2 such that $f(\{v_1, v_2\})$ is maximized. The process is continued until the kth node v_k is selected. Following this algorithm, we find an approximately reasonable solution for the problem of cascade maximization.

Example 7.3. *For the following graph, assume that node i activates node j when $|i - j| \equiv 2 \ (mod\ 3)$. Solve cascade maximization for $k = 2$.*

To find the first node v, we compute $f(\{v\})$ for all v. We start with node 1. At time 0, node 1 can only activate node 6, because

$$|1 - 6| \equiv 2 \ (mod\ 3), \tag{7.11}$$

---

[3]Formally, assuming $P \neq NP$, there is no polynomial time algorithm for this problem.

Algorithm 7.2 Maximizing the spread of cascades – Greedy algorithm

---

Require: Diffusion graph $G(V, E)$, budget k

1: **return** Seed set S (set of initially activated nodes)
2: $i = 0$;
3: $S = \{\}$;
4: **while** $i \neq k$ **do**
5:     $v = \arg\max_{v \in V \setminus S} f(S \cup \{v\})$;
      or equivalently $\arg\max_{v \in V \setminus S} f(S \cup \{v\}) - f(s)$
6:     $S = S \cup \{v\}$;
7:     $i = i + 1$;
8: **end while**
9: Return S;

$$|1 - 5| \not\equiv 2 \ (mod \ 3). \tag{7.12}$$

At time 1, node 1 can no longer activate others, but node 6 is active and can activate others. Node 6 has outgoing edges to nodes 4 and 5. From 4 and 5, node 6 can only activate 4:

$$|6 - 4| \equiv 2 \ (mod \ 3) \tag{7.13}$$

$$|6 - 5| \not\equiv 2 \ (mod \ 3). \tag{7.14}$$

At time 2, node 4 is activated. It has a single out-link to node 2 and since $|4 - 2| \equiv 2 \ (mod \ 3)$, 2 is activated. Node 2 cannot activate other nodes; therefore, $f(\{1\}) = 4$. Similarly, we find that $f(\{2\}) = 1$, $f(\{3\}) = 1$, $f(\{4\}) = 2$, $f(\{5\}) = 1$, and $f(\{6\}) = 4$. So, 1 or 6 can be chosen for our first node. Let us choose 6. If 6 is initially activated, nodes 1, 2, 4, and 6 will become activated at the end. Now, from the set $\{1, 2, 3, 4, 5, 6\} \setminus \{1, 2, 4, 6\} = \{3, 5\}$, we need to select one more node. This is because in the setting for this example, $f(\{6, 1\}) = f(\{6, 2\}) = f(\{6, 4\}) = f(\{6\}) = 4$. In general, one needs to compute $f(S \cup \{v\})$ for all $v \in V \setminus S$ (see Algorithm 7.2, line 5). We have $f(\{6, 3\}) = f(\{6, 5\}) = 5$, so we can select one node randomly. We choose 3. So, $S = \{6, 3\}$ and $f(S) = 5$.

7.2.3 Intervention

Consider a false rumor spreading in social media. This is an example where we are interested in stopping an information cascade in social media.

Intervention in the independent cascade model can be achieved using three methods:

1. By limiting the number of out-links of the sender node and potentially reducing the chance of activating others. Note that when the sender node is not connected to others via directed edges, no one will get activated by the sender.

2. By limiting the number of in-links of receiver nodes and therefore reducing their chance of getting activated by others.

3. By decreasing the activation probability of a node ($p_{v,w}$) and therefore reducing the chance of activating others.

7.3 Diffusion of Innovations

Diffusion of innovations is a phenomenon observed regularly in social media. A music video going viral or a piece of news being retweeted many times are examples of innovations diffusing across social networks. As defined by Rogers [239], an innovation is "an idea, practice, or object that is perceived as new by an individual or other unit of adoption." Innovations are created regularly; however, not all innovations spread through populations. The theory of diffusion of innovations aims to answer why and how these innovations spread. It also describes the reasons behind the diffusion process, the individuals involved, and the rate at which ideas spread. In this section, we review characteristics of innovations that are likely to be diffused through populations and detail well-known models in the diffusion of innovations. Finally, we provide mathematical models that can model the process of diffusion of innovations and describe how we can intervene with these models.

7.3.1 Innovation Characteristics

For an innovation to be adopted, the individual adopting it (adopter) and the innovation must have certain qualities.

Innovations must be highly *observable*, should have a *relative advantage* over current practices, should be *compatible* with the sociocultural

paradigm to which it is being presented, should be observable under various trials (*trialability*), and should not be highly *complex*.

In terms of individual characteristics, many researchers [239, 127] claim that the adopter should adopt the innovation earlier than other members of his or her social circle (*innovativeness*).

7.3.2 Diffusion of Innovations Models

Some of the earliest models for diffusion of innovations were provided by Gabriel Tarde in the early 20th century [281]. In this section, we review basic diffusion of innovations models. Interested readers may refer to the bibliographical notes for further study.

Ryan and Gross: Adopter Categories

Ryan and Gross [242] studied the adoption of hybrid seed corn by farmers in Iowa [266]. The hybrid seed corn was highly resistant to diseases and other catastrophes such as droughts. However, farmers did not adopt it because of its high price and the seed's inability to reproduce. Their study showed that farmers received information through two main channels: mass communications from companies selling the seeds and interpersonal communications with other farmers. They found that although farmers received information from the mass channel, the influence on their behavior was coming from the interpersonal channel. They argued that adoption depended on a combination of information from both channels. They also observed that the adoption rate follows an *S*-shaped curve and that there are five different types of adopters based on the order in which they adopt the innovations: (1) *Innovators* (top 2.5%), (2) *Early Adopters* (13.5%), (3) *Early Majority* (34%), (4) *Late Majority* (34%), and (5) *Laggards* (16%). Figure 7.5 depicts the distribution of these adopters as well as the cumulative adoption S-shaped curve. As shown in the figure, the adoption rate is slow when innovators or early adopters adopt the product. Once early majority individuals start adopting, the adoption curve becomes linear, and the rate is constant until all late majority members adopt the product. After the late majority adopts the product, the adoption rate becomes slow once again as laggards start adopting, and the curve slowly approaches 100%.
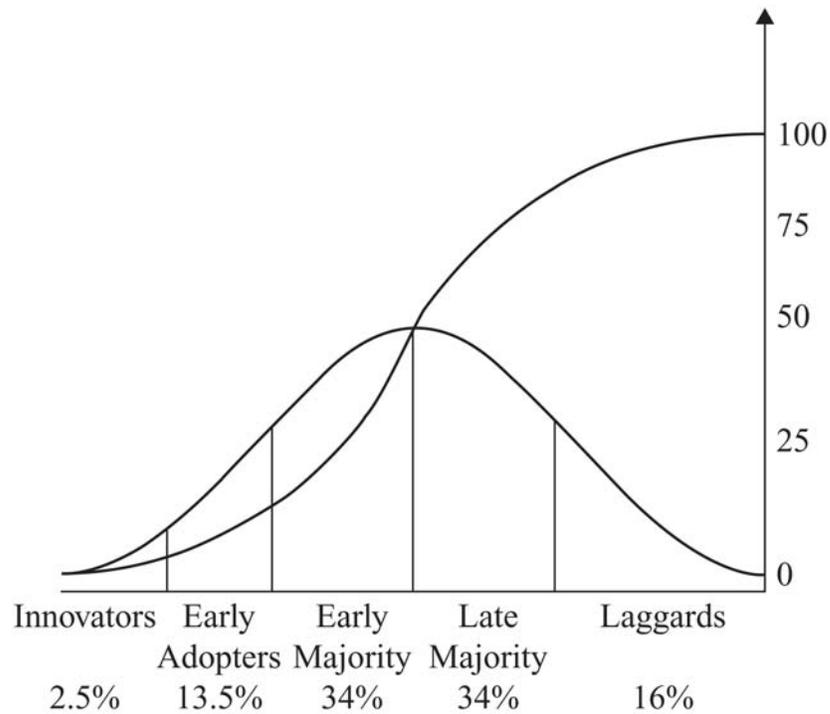
Figure 7.5: Types of Adopters and S-Shaped Cumulative Adoption Curve.

Katz: Two-Step Flow Model

Elihu Katz, a professor of communication at the University of Pennsylvania, is a well-known figure in the study of the flow of information. In addition to a study similar to the adoption of hybrid corn seed on how physicians adopted the new tetracycline drug [59], Katz also developed a *two-step flow* model (also known as the *multistep flow model*) [143] that describes how information is delivered through mass communication. The basic idea is depicted in Figure 7.6. Most information comes from mass media and is then directed toward influential figures called *opinion leaders*. These leaders then convey the information (or form opinions) and act as hubs for other members of the society.

234
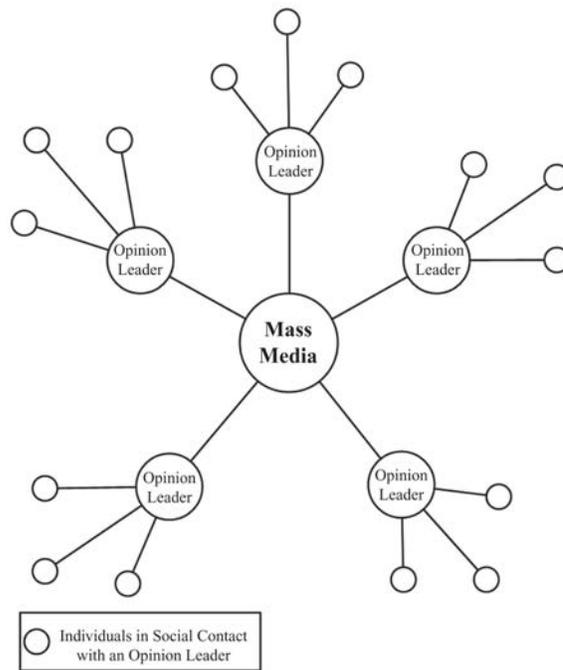
Figure 7.6: Katz Two-Step Flow Model.

Rogers: Diffusion of Innovations Process

Rogers in his well-known book, *Diffusion of Innovations* [239], discusses various theories regarding the diffusion of innovations process. In particular, he describes a five stage process of adoption:

1. **Awareness**: In this stage, the individual becomes aware of the innovation, but her information about the product is limited.

2. **Interest**: The individual shows interest in the product and seeks more information.

3. **Evaluation**: The individual imagines using the product and decides whether or not to adopt it.

4. **Trial**: The individual performs a trial use of the product.

5. **Adoption**: The individual decides to continue the trial and adopts the product for full use.

7.3.3 Modeling Diffusion of Innovations

To effectively make use of the theories regarding the diffusion of innovations, we demonstrate a mathematical model for it in this section. The model incorporates basic elements discussed so far and can be used to effectively model a diffusion of innovations process. It can be concretely described as

$$\frac{dA(t)}{dt} = i(t)[P - A(t)]. \tag{7.15}$$

Here, $A(t)$ denotes the total population that adopted the innovation until time t. $i(t)$ denotes the coefficient of diffusion, which describes the innovativeness of the product being adopted, and P denotes the total number of potential adopters (until time t). This equation shows that the rate at which the number of adopters changes throughout time depends on how innovative is the product being adopted. The adoption rate only affects the potential adopters who have not yet adopted the product. Since $A(t)$ is the total population of adopters until time t, it is a cumulative sum and can be computed as follows:

$$A(t) = \int_{t_0}^{t} a(t)dt, \tag{7.16}$$

where $a(t)$ defines the adopters at time t. Let A_0 denote the number of adopters at time t_0. There are various methods of defining the diffusion coefficient [185]. One way is to define $i(t)$ as a linear combination of the cumulative number of adopters at different times $A(t)$,

$$i(t) = \alpha + \alpha_0 A_0 + \cdots + \alpha_t A(t) = \alpha + \sum_{i=t_0}^{t} \alpha_i A(i), \tag{7.17}$$

where α_i's are the weights for each time step. Often a simplified version of this linear combination is used. In particular, the following three models for computing $i(t)$ are considered in the literature:

$$
\begin{aligned}
i(t) &= \alpha, && \text{External-Influence Model} && (7.18)\\
i(t) &= \beta A(t), && \text{Internal-Influence Model} && (7.19)\\
i(t) &= \alpha + \beta A(t), && \text{Mixed-Influence Model} && (7.20)
\end{aligned}
$$

External Influence Factor

where α is the *external-influence factor* and β is the *imitation factor*. Equation

236

7.18 describes $i(t)$ in terms of α only and is independent of the current number of adopters $A(t)$; therefore, in this model, the adoption only depends on the external influence. In the second model, $i(t)$ depends on the number of adopters at any time and is therefore dependent on the internal factors of the diffusion process. β defines how much the current adopter population is going to affect the adoption and is therefore denoted as the *imitation* factor. The mixed-influence model is a model between the two that uses a linear combination of both previous models.

Imitation Factor

External-Influence Model

In the external-influence model, the adoption coefficient only depends on an external factor. One such example of external influence in social media is when important news goes viral. Often, people who post or read the news do not know each other; therefore, the importance of the news determines whether it goes viral. The external-influence model can be formulated as

$$\frac{dA(t)}{dt} = \alpha[P - A(t)]. \tag{7.21}$$

By solving Equation 7.21,

$$A(t) = P(1 - e^{-\alpha t}), \tag{7.22}$$

when $A(t = t_0 = 0) = 0$. The $A(t)$ function is shown in Figure 7.7. The number of adopters increases exponentially and then saturates near P.

Internal-Influence Model

In the internal-influence model, adoption depends on how many have adopted the innovation in the current time step.[4] In social media there is internal influence when a group of friends join a site due to peer pressure. Think of a group of individuals where the likelihood of joining a social networking site increases as more group members join the site. The internal influence model can be described as follows:

$$\frac{dA(t)}{dt} = \beta A(t)[P - A(t)]. \tag{7.23}$$
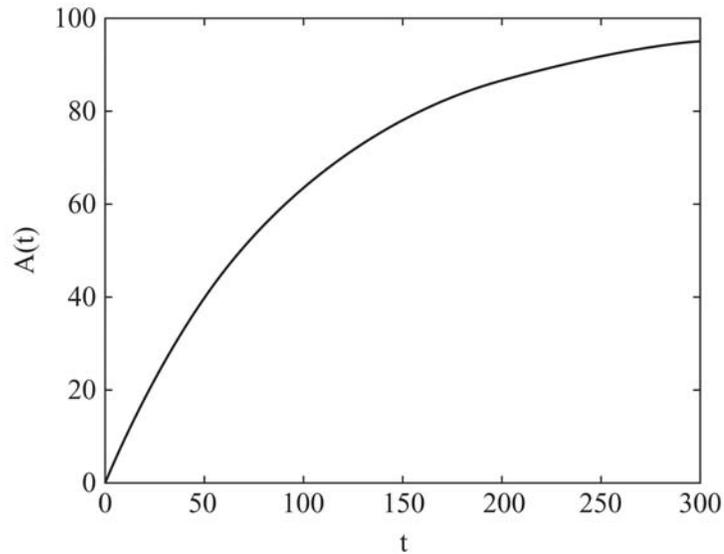
Figure 7.7: External-Influence Model for $P = 100$ and $\alpha = 0.01$.

Since the diffusion rate in this model depends on $\beta A(t)$, it is called the *pure imitation* model. The solution to this model is defined as

$$A(t) = \frac{P}{1 + \frac{P-A_0}{A_0}e^{-\beta P(t-t_0)}}, \tag{7.24}$$

where $A(t = t_0) = A_0$. The $A(t)$ function is shown in Figure 7.8.

Mixed-Influence Model

As discussed, the mixed influence model is situated in between the internal- and external-influence models. The mixed-influence model is defined as

$$\frac{dA(t)}{dt} = (\alpha + \beta A(t))[P - A(t)]. \tag{7.25}$$

By solving the differential equation, we arrive at

$$A(t) = \frac{P - \frac{\alpha(P-A_0)}{\alpha+\beta A_0}e^{-(\alpha+\beta P)(t-t_0)}}{1 + \frac{\beta(P-A_0)}{\alpha+\beta A_0}e^{-(\alpha+\beta P)(t-t_0)}}, \tag{7.26}$$

---

[4]The internal-influence model is similar to the SI model discussed later in the section on epidemics. For the sake of completeness, we provide solutions to both. Readers are encouraged to refer to that model in Section 7.4 for further insight.
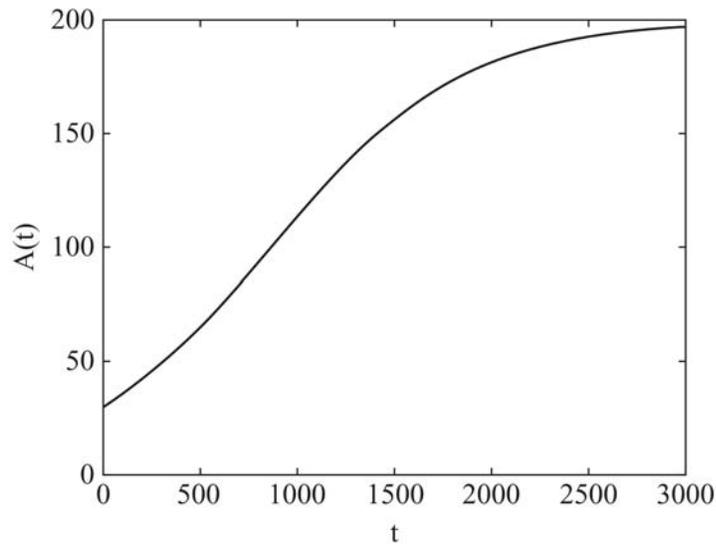
Figure 7.8: Internal-Influence Model for $A_0 = 30$, $\beta = 10^{-5}$, and $P = 200$.

where $A(t = t_0) = A_0$. The $A(t)$ function for the mixed-influence model is depicted in Figure 7.9.

　　We discussed three models in this section: internal, external, and mixed influence. Depending on the model used to describe the diffusion of innovations process, the respective equation for $A(t)$ (Equations 7.22, 7.24, or 7.26) should be employed to model the system.

7.3.4　Intervention

Consider a faulty product being adopted. The product company is planning to stop or delay adoptions until the product is fixed and re-released. This intervention can be performed by doing the following:

- *Limiting the distribution of the product or the audience that can adopt the product.* In our mathematical model, this is equivalent to reducing the population P that can potentially adopt the product.

- *Reducing interest in the product being sold.* For instance, the company can inform adopters of the faulty status of the product. In our models, this can be achieved by tampering α: setting α to a very small value in Equation 7.22 results in a slow adoption rate.

239
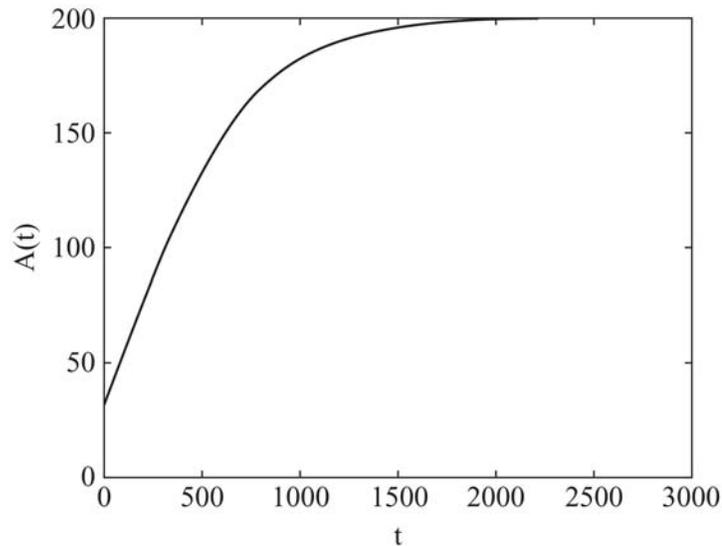
Figure 7.9: Mixed-Influence Model for $P = 200$, $\beta = 10^{-5}$, $A_0 = 30$, and $\alpha = 10^{-3}$.

- *Reducing interactions within the population.* Reduced interactions result in less imitation of product adoptions and a general decrease in the trend of adoptions. In our models, this can be achieved by setting β to a small value.

7.4 Epidemics

In an epidemic, a disease spreads widely within a population. This process consists of a *pathogen* (the disease being spread), a population of *hosts* (humans, animals, and plants, among others), and a spreading mechanism (breathing, drinking, sexual activity, etc.). Unlike information cascades and herding, but similar to diffusion of innovations models, epidemic models assume an implicit network and unknown connections among individuals. This makes epidemic models more suitable when we are interested in global patterns, such as trends and ratios of people getting infected, and not in who infects whom.

In general, a complete understanding of the epidemic process requires substantial knowledge of the biological process within *each* host and the im-

mune system process, as well as a comprehensive analysis of interactions among individuals. Other factors such as social and cultural attributes also play a role in how, when, and where epidemics happen. Large epidemics, also known as *pandemics*, have spread through human populations and include the Black Death in the 13th century (killing more than 50% of Europe's population), the Great Plague of London (100,000 deaths), the smallpox epidemic, in the 17th century (killing more than 90% of Massachusetts Bay Native Americans) and recent pandemics such as HIV/AIDS, SARS, H5N1 (Avian flu), and influenza. These pandemics motivated the introduction of epidemic models in the early 20th century and the establishment of the epidemiology field.

There are various ways of modeling epidemics. For instance, one can look at how hosts contact each other and devise methods that describe how epidemics happen in networks. These networks are called *contact networks*. A contact network is a graph where nodes represent the hosts and edges represent the interactions between these hosts. For instance, in the case of the HIV/AIDS epidemic, edges represent sexual interactions, and in the case of influenza, nodes that are connected represent hosts that breathe the same air. Nodes that are close in a contact network are not necessarily close in terms of real-world proximity. Real-world proximity might be true for plants or animals, but diseases such as SARS or avian flu travel between continents because of the traveling patterns of hosts. This spreading pattern becomes clearer when the science of epidemics is employed to understand the propagation of computer viruses in cell phone networks or across the internet [229, 214]. Contact Networks

Another way of looking at epidemic models is to avoid considering network information and to analyze only the rates at which hosts get infected, recover, and the like. This analysis is known as the *fully mixed* technique, assuming that each host has an equal chance of meeting other hosts. Through these interactions, hosts have random probabilities of getting infected. Though simplistic, the technique reveals several useful methods of modeling epidemics that are often capable of describing various real-world outbreaks. In this section, we concentrate on the fully mixed models that avoid the use of contact networks.[5] Fully Mixed Technique

Note that the models of information diffusion that we have already discussed, such as the models in diffusion of innovations or information

---

[5]A generalization of these techniques over networks can be found in [126, 125, 212].

cascades, are more or less related to epidemic models. However, what makes epidemic models different is that, in the other models of information diffusion, actors decide whether to adopt the innovation or take the decision and the system is usually fully observable. In epidemics, however, the system has a high level of uncertainty, and individuals usually do not decide whether to get infected or not. The models discussed in this section assume that (1) no contact network information is available and (2) the process by which hosts get infected is unknown. These models can be applied to situations in social media where the decision process has a certain uncertainty to it or is ambiguous to the analyst.

7.4.1 Definitions

Since there is no network, we assume that we have a population where the disease is being spread. Let N define the size of this crowd. Any member of the crowd can be in either one of three states:

1. **Susceptible**: When an individual is in the susceptible state, he or she can potentially get infected by the disease. In reality, infections can come from outside the population where the disease is being spread (e.g., by genetic mutation, contact with an animal, etc.); however, for simplicity, we make a *closed-world assumption*, where susceptible individuals can only get infected by infected people in the population. We denote the number of susceptibles at time t as $S(t)$ and the fraction of the population that is susceptible as $s(t) = S(t)/N$.

   Closed-world Assumption

2. **Infected**: An infected individual has the chance of infecting susceptible parties. Let $I(t)$ denote the number of infected individuals at time t, and let $i(t)$ denote the fraction of individuals who are infected, $i(t) = I(t)/N$.

3. **Recovered (or Removed)**: These are individuals who have either recovered from the disease and hence have complete or partial immunity against the infection or were killed by the infection. Let $R(t)$ denote the size of this set at time t and $r(t)$ the fraction recovered, $r(t) = R(t)/N$.

Clearly, $N = S(t) + I(t) + R(t)$ for all t. Since we are assuming that there is some level of randomness associated with the values of $S(t)$, $I(t)$, and $R(t)$,
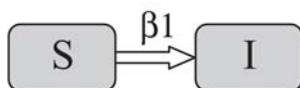
Figure 7.10: SI Model.

we try to deal with expected values and assume S, I, and R represent these at time t.

7.4.2   SI Model

We start with the most basic model. In this model, the susceptible individuals get infected, and once infected, they will never get cured. Denote β as the contact probability. In other words, the probability of a pair of people meeting in any time step is β. So, if $\beta = 1$, everyone comes into contact with everyone else, and if $\beta = 0$, no one meets another individual. Assume that when an infected individual meets a susceptible individual the disease is being spread with probability 1 (this can be generalized to other values). Figure 7.10 demonstrates the SI model and the transition between states that happens in this model for individuals. The value over the arrow shows that each susceptible individual meets at least βI infected individuals during the next time step.

Given this situation, infected individuals will meet βN people on average. We know from this set that only the fraction S/N will be susceptible and that the rest are infected already. So, each infected individual will infect $\beta NS/N = \beta S$ others. Since I individuals are infected, βIS will be infected in the next time step. This means that the number of susceptible individuals will be reduced by this factor as well. So, to get different values of S and I at different times, we can solve the following differential equations:

$$\frac{dS}{dt} = -\beta IS, \tag{7.27}$$

$$\frac{dI}{dt} = \beta IS. \tag{7.28}$$

Since $S + I = N$ at all times, we can eliminate one equation by replacing S with $N - I$:

$$\frac{dI}{dt} = \beta I(N - I). \tag{7.29}$$

243

The solution to this differential equation is called the *logistic growth function*,

$$I(t) = \frac{NI_0 e^{\beta t}}{N + I_0(e^{\beta t} - 1)}, \tag{7.30}$$

where I_0 is the number of individuals infected at time 0. In general, analyzing epidemics in terms of the number of infected individuals has nominal generalization power. To address this limitation, we can consider infected fractions. We therefore substitute $i_0 = \frac{I_0}{N}$ in the previous equation,

$$i(t) = \frac{i_0 e^{\beta t}}{1 + i_0(e^{\beta t} - 1)}. \tag{7.31}$$

Note that in the limit, the SI model infects all the susceptible population because there is no recovery in the model. Figure 7.11(a) depicts the logistic growth function (infected individuals) and susceptible individuals for $N = 100$, $I_0 = 1$, and $\beta = 0.003$. Figure 7.11(b) depicts the infected population for HIV/AIDS for the past 20 years. As observed, the infected population can be approximated well with the logistic growth function and follows the SI model. Note that in the HIV/AIDS graph, not everyone is getting infected. This is because not everyone in the United States is in the susceptible population, so not everyone will get infected in the end. Moreover, there are other factors that are far more complex than the details of the SI model that determine how people get infected with HIV/AIDS.

(a) SI Model Simulation   (b) HIV/AIDS Infected Population Growth
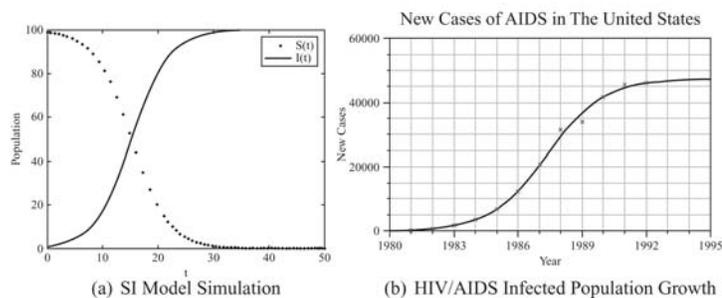
Figure 7.11: SI model simulation compared to the HIV/AIDS growth in the United States.

Figure 7.12: SIR Model.

7.4.3 SIR Model

The SIR model, first introduced by Kermack, and McKendrick [148], adds more detail to the standard SI model. In the SIR model, in addition to the *I* and *S* states, a recovery state *R* is present. Figure 7.12 depicts the model. In the SIR model, hosts get infected, remain infected for a while, and then recover. Once hosts recover (or are removed), they can no longer get infected and are no longer susceptible. The process by which susceptible individuals get infected is similar to the SI model, where a parameter β defines the probability of contacting others. Similarly, a parameter γ in the SIR model defines how infected people recover, or the recovering probability of an infected individual in a time period Δt.

In terms of differential equations, the SIR model is

$$\frac{dS}{dt} = -\beta IS, \tag{7.32}$$

$$\frac{dI}{dt} = \beta IS - \gamma I, \tag{7.33}$$

$$\frac{dR}{dt} = \gamma I. \tag{7.34}$$

Equation 7.32 is identical to that of the SI model (Equation 7.27). Equation 7.33 is different from Equation 7.28 of the SI model by the addition of the term γI, which defines the number of infected individuals who recovered. These are removed from the infected set and are added to the recovered ones in Equation 7.34. Dividing Equation 7.32 by Equation 7.34, we get

$$\frac{dS}{dR} = -\frac{\beta}{\gamma} S, \tag{7.35}$$

and by assuming the number of recovered at time 0 is zero ($R_0 = 0$),

$$\log \frac{S_0}{S} = \frac{\beta}{\gamma} R. \tag{7.36}$$

245

$$S_0 = Se^{\frac{\beta}{\gamma}R} \tag{7.37}$$

$$S = S_0 e^{-\frac{\beta}{\gamma}R} \tag{7.38}$$

Since $I + S + R = N$, we replace I in Equation 7.34,

$$\frac{dR}{dt} = \gamma(N - S - R). \tag{7.39}$$

Now combining Equations 7.38 and 7.39,

$$\frac{dR}{dt} = \gamma(N - S_0 e^{-\frac{\beta}{\gamma}R} - R). \tag{7.40}$$

If we solve this equation for R, then we can determine S from 7.38 and I from $I = N - R - S$. The solution for R can be computed by solving the following integration:

$$t = \frac{1}{\gamma} \int_0^R \frac{dx}{N - S_0 e^{-\frac{\beta}{\gamma}x} - x}. \tag{7.41}$$

However, there is no closed-form solution to this integration, and only numerical approximation is possible. Figure 7.13 depicts the behavior of the SIR model for a set of initial parameters.

The two models in the next two subsections are generalized versions of the two models discussed thus far: SI and SIR. These models allow individuals to have temporary immunity and to get reinfected.

7.4.4 SIS Model

The SIS model is the same as the SI model, with the addition of infected nodes recovering and becoming susceptible again (see Figure 7.14). The differential equations describing the model are

$$\frac{dS}{dt} = \gamma I - \beta IS, \tag{7.42}$$

$$\frac{dI}{dt} = \beta IS - \gamma I. \tag{7.43}$$

By replacing S with $N - I$ in Equation 7.43, we arrive at

$$\frac{dI}{dt} = \beta I(N - I) - \gamma I = I(\beta N - \gamma) - \beta I^2. \tag{7.44}$$

246
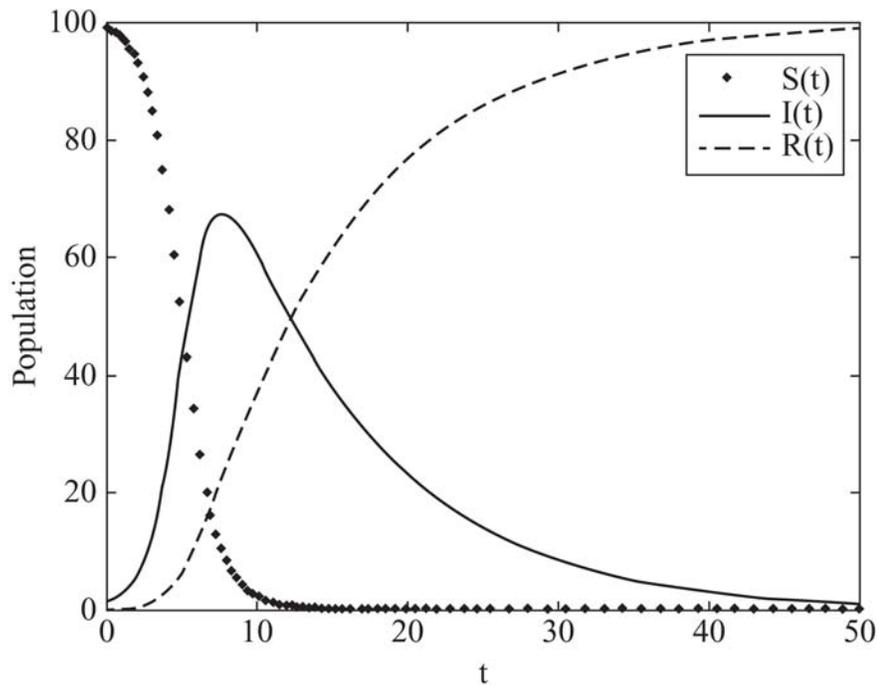
Figure 7.13: SIR Model Simulated with $S_0 = 99$, $I_0 = 1$, $R_0 = 0$, $\beta = 0.01$, and $\gamma = 0.1$.

When $\beta N \leq \gamma$, the first term will be negative or zero at most; hence, the whole term becomes negative. Therefore, in the limit, the value $I(t)$ will decrease exponentially to zero. However, when $\beta N > \gamma$, we will have a logistic growth function as in the SI model. Having said this, as the simulation of the SIS model shows in Figure 7.15, the model will never infect everyone. It will reach a steady state, where both susceptibles and infecteds reach an equilibrium (see the epidemics exercises).
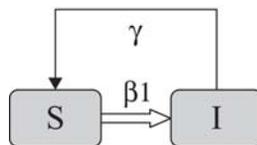
Figure 7.14: SIS Model.
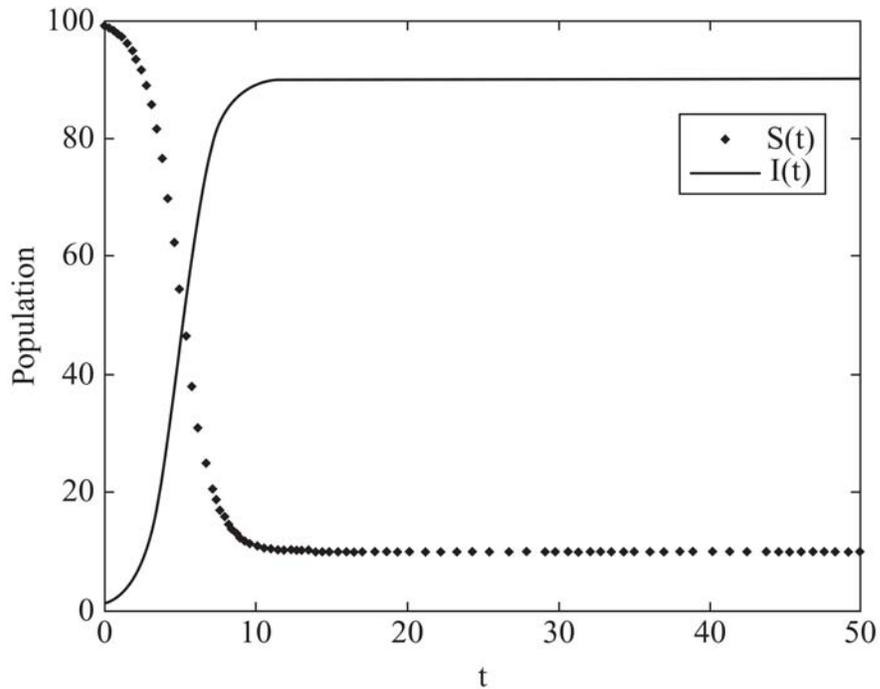
Figure 7.15: SIS Model Simulated with $S_0 = 99$, $I_0 = 1$, $\beta = 0.01$, and $\gamma = 0.1$.

7.4.5 SIRS Model

The final model analyzed in this section is the SIRS model. Just as the SIS model extends the SI, the SIRS model extends the SIR, as shown in Figure 7.16. In this model, the assumption is that individuals who have recovered will lose immunity after a certain period of time and will become susceptible again. A new parameter has been added to the model λ that defines the probability of losing immunity for a recovered individual. The
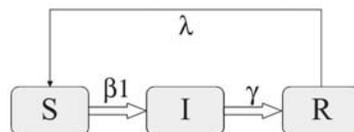
Figure 7.16: SIRS Model.

248

set of differential equations that describe this model is

$$\frac{dS}{dt} = \lambda R - \beta IS, \tag{7.45}$$

$$\frac{dI}{dt} = \beta IS - \gamma I, \tag{7.46}$$

$$\frac{dR}{dt} = \gamma I - \lambda R. \tag{7.47}$$

Like the SIR model, this model has no closed-form solution, so numerical integration can be used. Figure 7.17 demonstrates a simulation of the SIRS model with given parameters of choice. As observed, the simulation outcome is similar to the SIR model simulation (see Figure 7.13). The major difference is that in the SIRS, the number of susceptible and recovered individuals changes non-monotonically over time. For example, in SIRS, the number of susceptible individuals decreases over time, but after reaching the minimum count, starts increasing again. On the contrary, in the SIR, both susceptible individuals and recovered individuals change monotonically, with the number of susceptible individuals decreasing over time and that of recovered individuals increasing over time. In both SIR and SIRS, the infected population changes non-monotonically.

7.4.6 Intervention

A pressing question in any pandemic or epidemic outbreak is how to stop the process. In this section, we discuss epidemic intervention based on a recent discovery [55]. In any epidemic outbreak, infected individuals infect susceptible individuals. Although in this chapter we discussed random infection in the real world, what actually takes place is quite different. Infected individuals have a limited number of contacts and can only infect them if said contacts are susceptible. A well-connected infected individual is more dangerous to the epidemic outbreak than someone who has no contacts. In other words, the epidemic takes place in a network. Unfortunately, it is often difficult to trace these contacts and outline the contact network. If this was possible, the best way to intervene with the epidemic outbreak would be to vaccinate the highly connected nodes and stop the epidemic. This would result in what is known as *herd immunity* and would stop the epidemic outbreak. Herd immunity entails vaccinating a population inside a herd such that the pathogen cannot initiate an outbreak
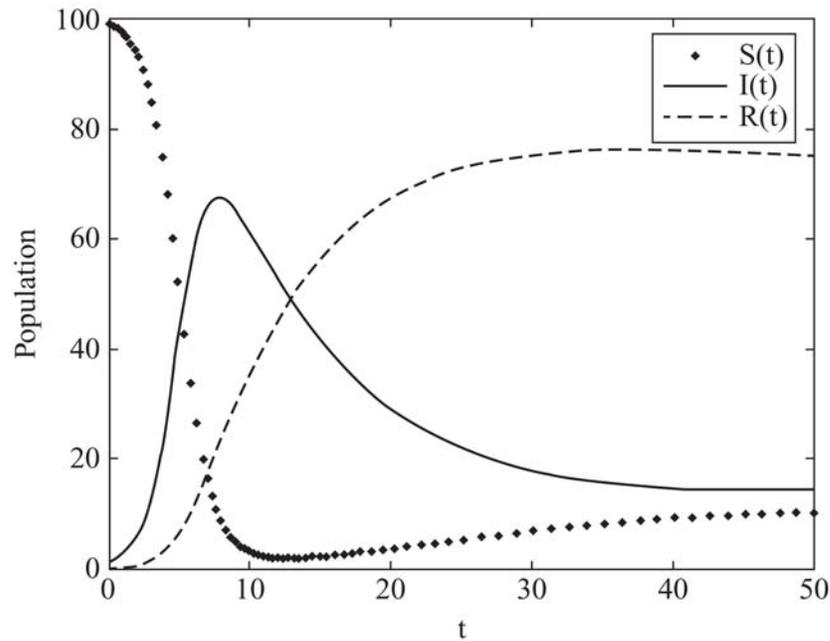
Figure 7.17: SIRS Model Simulated with $S_0 = 99$, $I_0 = 1$, $R_0 = 0$, $\gamma = 0.1$, $\beta = 0.01$, and $\lambda = 0.02$.

inside the herd. In general, creating herd immunity requires at least a **random** sample of 96% of the population to be vaccinated. Interestingly, we can achieve the same herd immunity by making use of friends in a network. In general, people know which of their friends have more friends. So, they know or have access to these higher-degree and more-connected nodes. Researchers found that if a random population of 30% of the herd is selected and then these 30% are asked for their highest degree friends, one can achieve herd immunity by vaccinating these friends. Of course, older intervention techniques such as separating those infected from those susceptible (quarantining them) or removing those infected (killing cows with mad cow disease) still work.

7.5 Summary

In this chapter, we discussed the concept of information diffusion in social networks. In the herd behavior, individuals observe the behaviors of others and act similarly to them based on their own benefit. We reviewed the well-known diners example and urn experiment and demonstrated how conditional probabilities can be used to determine why herding takes place. We discussed how herding experiments should be designed and ways to intervene with it.

Next, we discussed the information cascade problem with the constraint of sequential decision making. The independent cascade model (ICM) is a sender-centric model and has a level of stochasticity associated with it. The spread of cascades can be maximized in a network given a budget on how many initial nodes can be activated. Unfortunately, the problem is NP-hard; therefore, we introduced a greedy approximation algorithm that has guaranteed performance due to the submodularity of ICM's activation function. Finally, we discussed how to intervene with information cascades.

Our next topic was the diffusion of innovations. We discussed the characteristics of adoption both from the individual and innovation point of view. We reviewed well-known theories such as the models introduced by Ryan and Gross, Katz, and Rogers, in addition to experiments in the field, and different types of adopters. We also detailed mathematical models that account for internal, external, and mixed influences and their intervention procedures.

Finally, we moved on to epidemics, an area where decision making is usually performed unconsciously. We discussed four epidemic models: SI, SIR, SIS, and SIRS; the two last models allow for reinfected individuals. For each model we provided differential equations, numerical solutions, and closed-form solutions, when available. We concluded the chapter with intervention approaches to epidemic outbreaks and a review of herd immunity in epidemics. Although a 96% random vaccination is required for achieving herd immunity, it is also possible to achieve it by selecting a random population of 30% and then vaccinating their highest degree friends.

7.6 Bibliographic Notes

The concept of the herd has been well studied in psychology by Freud (crowd psychology), Carl Gustav Jung (the collective unconscious), and Gustave Le Bon (the popular mind). It has also been observed in economics by Veblen [288] and in studies related to the *bandwagon effect* [240, 259, 165]. The behavior is also discussed in terms of *sociability* [258] in sociology.

Herding, first coined by Banerjee [23], at times refers to a slightly different concept. In herd behaviour discussed in this chapter, the crowd does not necessarily start with the same decision, but will eventually reach one, whereas in herding the same behavior is usually observed. Moreover, in herd behavior, individuals decide whether the action they are taking has some benefits to themselves or is rational, and based on that, they will align with the population. In herding, some level of uncertainty is associated with the decision, and the individual does not know why he or she is following the crowd.

Another confusion is that the terms "herd behavior/herding" is often used interchangebly with "information cascades" [37, 299]. To avoid this problem, we clearly define both in the chapter and assume that in herd behavior, decisions are taken based on global information, whereas in information cascades, local information is utilized.

Herd behavior has been studied in the context of financial markets [60, 74, 38, 69] and investment [250]. Gale analyzes the robustness of different herd models in terms of different constraints and externalities [93], and Shiller discusses the relation between information, conversation, and herd behavior [256]. Another well-known social conformity experiment was conducted in Manhattan by Milgram et al. [195].

Other recent applications of threshold models can be found in [307, 295, 296, 285, 286, 252, 232, 202, 184, 183, 108, 34]. Bikhchandani et al. [1998] review conformity, fads, and information cascades and describe how observing past human decisions can help explain human behavior. Hirshleifer [128] provides information cascade examples in many fields, including zoology and finance.

In terms of diffusion models, Robertson [238] describes the process and Hagerstrand et al. [118] introduce a model based on the spatial stages of the diffusion of innovations and Monte Carlo simulation models for diffusion of innovations. Bass [30] discusses a model based on differential equations. Mahajan and Peterson [187] extend the Bass model.

Instances of external-influence models can be found in [119, 59] and internal-influence models are applied in [188, 111, 110]. The Gompertz function [189], widely used in forecasting, has a direct relationship with the internal-influence diffusion curve. Mixed-influence model examples include the work of Mahajan and Muller [186] and Bass model [30].

Midgley and Dowling [193] introduce the *contingency model*. Abrahamson and Rosenkopf [3] mathematically analyze the bandwagon effect and diffusion of innovations. Their model predicts whether the bandwagon effect will occur and how many organizations will adopt the innovation. Network models of diffusion and thresholds for diffusion of innovations models are discussed by Valente [286, 287]. Diffusion through blogspace and in general, social networks, has been analyzed by [112, 169, 306, 310].

For information on different pandemics, refer to [220, 31, 230, 68, 77, 53, 113, 206]. To review some early and in-depth analysis of epidemic models, refer to [21, 13]. Surveys of epidemics can be found in [124, 125, 126, 72]. Epidemics in networks have been discussed [212, 201, 144] extensively. Other general sources include [171, 78, 212]; [28]. A generalized model for contagion is provided by Dodds and Watts [73] and, in the case of best response dynamics, in [202].

Other topics related to this chapter include wisdom of crowd models [104] and swarm intelligence [79, 82, 42, 147]. One can also analyze *information provenance*, which aims to identify the sources from which information has diffused. Barbier at al. [25] provide an overview of information provenance in social media in their book.

7.7 Exercises

1. Discuss how different information diffusion modeling techniques differ. Name applications on social media that can make use of methods in each area.

Herd Effect

2. What are the minimum requirements for a herd behavior experiment? Design an experiment of your own.

Diffusion of Innovation

3. Simulate internal-, external-, and mixed-influence models in a program. How are the saturation levels different for each model?

4. Provide a simple example of diffusion of innovations and suggest a specific way of intervention to expedite the diffusion.

Information Cascades

5. Briefly describe the independent cascade model (ICM).

6. What is the objective of cascade maximization? What are the usual constraints?

7. Follow the ICM procedure until it converges for the following graph. Assume that node i activates node j when $i - j \equiv 1 \ (mod\ 3)$ and node 5 is activated at time 0.
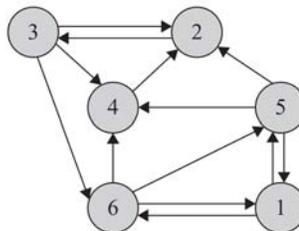
Epidemics

8. Discuss the mathematical relationship between the SIR and the SIS models.

9. Based on our assumptions in the SIR model, the probability that an individual remains infected follows a standard exponential distribution. Describe why this happens.

10. In the SIR model, what is the most likely time to recover based on the value of γ?

11. In the SIRS model, compute the length of time that an infected individual is likely to remain infected before he or she recovers.

12. After the model saturates, how many are infected in the SIS model?