# 11. *Broadcast, Diffusion, and Contagion*

*As contagion of sickness makes sickness, contagion of trust can make trust.*

—Marianne Moore

In this chapter, we model the spread of information, technologies, behaviors, beliefs, and diseases throughout a population using models of broadcast, diffusion, and contagion. These models play central roles in communication, marketing, and epidemiology. All three models partition the population into people who know or have some thing and those who do not. Over time, people move between those two groups. Someone moves from being susceptible to a disease to being infected, or from being uninformed about a new product or idea to being informed.

Empirical plots of the number of people who over time catch a disease, buy a product, or know a piece of information (the *adoption curve*) tend to be either concave or S-shaped. How people learn the information or catch the disease—that is, whether it spreads by broadcast or diffusion—determines the shape of that graph. One contribution of this chapter will be to link the micro-level processes of how ideas and diseases spread to the shape of these adoption curves. The chapter begins with an analysis of the broadcast model, which applies when people hear of an idea or catch a disease from a single source. This model produces plots with an r-shape. We then cover the diffusion model, in which spread occurs from contact, as when a disease spreads from person to person. This model produces an S-shaped curve.

Many products, programs, ideas, and pieces of information spread by both broadcast and word of mouth. We can model these environments by allowing for both broadcast and diffusion. The resulting model, known as the Bass model, plays a central role in marketing. Whether it produces more of an r-shape or S-shape depends on the strengths of the two processes. The last model we cover, the SIR model of contagion from epidemiology, includes a rate of recovery. This assumption could capture an immune system fighting off a disease, behaviors or styles dropping out of fashion, or information becoming less worthy of passing on to others.

The SIR model produces a tipping point, where small changes in the attributes of the product or a disease spell the difference between failure and success. A slight reduction in virulence can transform a mass infection into a minor outbreak. A small increase in the probability of spreading word of a hot new band can be the difference between the Beatles and a band that played pubs in Liverpool for a few months in the 1960s.

# The Broadcast Model

All of the models we cover in this chapter assume a *relevant population*, denoted by $N_{\text{POP}}$. This consists of those people who could potentially catch the disease, learn the piece of information, or adopt the product. The relevant population is not the entire population of, say, a city or country. If we are modeling the spread of a continuous aortic suture method, the relevant population is heart surgeons, not everyone in the city of Philadelphia.

At any moment in time, some people have the disease, know the information, or adopt the behavior. We refer to these people as either the *infected* or the *informed* (denoted by $I_t$). The remaining members of the relevant population are *susceptible* (denoted by $S_t$). These people could catch the disease or learn the information or behavior.[1] The relevant population equals the sum of the number of people infected (or informed) plus the number of susceptible people: $N_{\text{POP}} = I_t + S_t$.

# Broadcast Model

$I_{t+1} = I_t + P_{\text{broad}} \cdot S_t$

where $P_{\text{broad}}$ denotes the **broadcast probability**, and $I_t$ and $S_t$ equal the number informed and susceptible at time $t$.

Initially, $I_0 = 0$ and $S_0 = N_{\text{POP}}$.

The *broadcast model* captures the spread of ideas, rumors, information, or technologies through media like television, radio, or the internet. Knowledge of most current events spreads through broadcast. The model captures processes in which a source, which could be the government, a corporation, or a newspaper, spreads information. It could also capture contaminations that spread through a water supply. The model does not apply to diseases or ideas that spread from person to person. As the broadcast model better fits the spread of ideas and information than disease, we refer to the number of people informed, as opposed to the number infected.

The number of informed people in a given time period equals the number informed in the previous period plus the probability that a susceptible person hears of the information multiplied by the number of susceptible people (see box). By convention, the initial population contains only susceptible people. Calculating the number of informed people in all future periods involves plugging the number of informed and susceptible people into the difference equation. The result will be an r-shaped adoption curve.

Imagine that the mayor of a city with 1 million residents announces a new tax policy. Prior to the announcement, no one could have known about the policy. If we assume the probability that someone hears the news on any given day equals 30% ($P_{broad} = 0.3$), then 300,000 people hear about it the first day. On the second day, 30% of the remaining 700,000 people, or 210,000, hear about it. In each period, the number of informed people increases and does so at a decreasing rate, as shown in figure 11.1.

image

Figure 11.1: The r-Shaped Adoption Curve Produced by the Broadcast Model

In the broadcast model, everyone in the relevant population learns the information or buys the product. Using initial sales data, we can therefore estimate the relevant population size. Suppose that a company introduces a new line of shoes for people who practice tai chi, and in the first week, it receives orders for 20,000 pairs of shoes. If in the second week it receives orders for 16,000 pairs, we can make a crude estimate of eventual total sales, the size of the relevant population, to be 100,000.

# Fitting the Broadcast Model to Data

**Period 1:** $I_1 = 20,000 = P_{\text{broad}} \cdot N_{\text{POP}}$

**Period 2:** $I_2 = 36,000 = 20,000 + P_{\text{broad}} \cdot (N_{\text{POP}} - 20,000)$

**Solution:**[2] $P_{\text{broad}} = 0.2$ and $N_{\text{POP}} = 100,000$

We should not have a great deal of confidence in any estimate based on two data points. The model leaves out any number of real-world features. People might be hearing by word of mouth as well as through media, some people may have bought multiple pairs, or advertising may have targeted likely buyers. Including these features would change the estimates. Caveats aside, the model provides a rough estimate. The firm should not expect to sell exactly 2 million pairs, but it should be confident that they will sell more than 100,000 pairs. As more data arrives, the estimate can be improved. If week three's sales equals 13,000 pairs (the amount the model predicts), then the firm can place more confidence in the initial prediction.

# The Diffusion Model

Most diseases as well as information about many products, ideas, and breakthroughs spread by word of mouth. The *diffusion model* captures such processes. It assumes that when one person adopts a technology or catches a disease, that person has some probability of passing it on to those with whom she comes in contact. In the case of a disease, choice pays no role. The probability a person catches the disease depends on factors such as genetics, the virulence of the disease, and even the temperature. Malaria will spread much faster during a hot, wet season than during a cold, dry one.

The spread of a technology involves a choice on the part of the adopters, so technologies that are more useful will be adopted with a higher probability. We do not explicitly consider choice in the model, however. Therefore, the hipness of the Apple Watch plays the same role as the virulence of a flu strain.

We again emphasize the spread of information, so we will refer to people as informed or uninformed. New people become informed if they meet an informed person and the information spreads between them. These are two distinct events that vary by context. People in cities may have higher contact probabilities than rural people, and information with high salience—say, news that aliens have landed—has a higher sharing probability than news of the reintroduction of pretzel M&M's. Thus, we write the *diffusion probability* as the product of a *contact probability* and a *sharing probability*. We can write the model in terms of the diffusion probability, but when we estimate or apply the model, we must keep track of the two probabilities independently.

The diffusion model assumes *random mixing*, that is, that any two people in the relevant population are equally likely to make contact. This assumption should raise a red flag. It may be an accurate assumption for a model of disease spread in a preschool where children interact with high frequency. It is problematic to apply it to a city-level population. People do not randomly mix. People live and work in neighborhoods; they belong to work teams, families, and social groups. Their interactions are primarily within those groups. Remember, though, an assumption need not be accurate to be part of a useful model. We proceed with the assumption but keep an open mind toward changing it.

# Diffusion Model

image

where $P_{\text{diffuse}} = P_{\text{spread}} \cdot P_{\text{contact}}$.

In this model as well, in the long run everyone in the relevant population learns the information. However, in this model, the adoption curve for the diffusion model has an S-shape. Initially, few people are informed; $I_0$ is small. It follows that the number of susceptible people who meet an informed person must also be small. As the number of informed people grows larger, the number of contacts between informed and uninformed people increases, producing larger increases in the number of informed people. When nearly everyone in the relevant population is informed, the number of newly informed people decreases, forming the top of the S-shape. Technological adoption curves often have this shape. For example, adoption curves for hybrid seed in the last century vary by state (Iowans adopted hybrid seeds faster than Alabamans), but all of the curves have an S-shape.[3]

In the broadcast model, estimating the relevant population size from data is straightforward. The initial numbers of adopters correlates strongly with the relevant population. In contrast, estimating the size of the relevant population using data from a diffusion model can be difficult. The same increases in product sales could result from a large diffusion probability among a small relevant population or a small diffusion probability among a large population. Figure 11.2 shows data for two hypothetical smartphone applications. On the first day, one hundred people bought each application. For each of the next five days, the first application realizes both higher total sales and larger increases in sales. Absent a model, we would probably predict the first application to have the larger market. Fitting the model to the two data streams shows the opposite to be true.

image

Figure 11.2: Two Adoption Curves for Sales of an Application

The first application fits a diffusion probability of 40% and a relevant population of 1,000 people, while the second application corresponds to a diffusion probability of 30% and a relevant population of 1 million people.[4] Within a few days, we would come to realize that there is a larger relevant population for the second application. Nevertheless, absent the model, we would make the incorrect inference about total sales if we based it on just five days of data.

When using the diffusion model to guide action, we unpack the probability of diffusion as the product of the probability of sharing and the probability of making contact. To increase the speed of an application's sales, its developer could increase the rate at which people meet or increase the probability that they share information about the application. Changing the first probability would be difficult. To increase the second probability, the developer could provide incentives for signing up new buyers, which many developers do. A game developer might give points within the game to users who sign up new buyers. Though this would increase the speed of diffusion, it would not affect total sales, at least not according to the model. As mentioned, total sales equals the relevant population size, regardless of the probability of sharing. Increasing the rate of sales produces no long-term effect.

Most consumer goods and information spread through both diffusion and broadcast. Our next model, the *Bass model,* combines the two processes in a single model.[5] The difference equation in the Bass model equals the sum of the difference equations from the broadcast model and the diffusion model. The adoption curve of the Bass model will be more S-shaped the larger the probability of diffusion. The adoption curves for televisions, radios, cars, computers, telephones, and cell phones all are combinations of r-shapes and S-shapes.

# Bass Model

image

where $P_{\text{broad}}$ = probability of broadcast and $P_{\text{diffuse}}$ = probability of diffusion.

# The SIR Model

In the models that we have covered so far, a person who adopts a technology never abandons it. That makes sense for the adoption of technologies like electricity, the dishwasher, and television; we never reverse our adoptions. That assumption does not hold for all things that spread by diffusion. After we catch a disease, we recover. When we adopt a fashion or fad, such as a particular style of dress or a dance step, we can abandon it. Following convention, we refer to people who drop an adoption as *recovered*. The resulting model, the *SIR model* (susceptible, infected, recovered), occupies a central position in epidemiology.

Given the model's origins and that recovery occurs more naturally in diseases, we describe the model using the spread of a disease, such as COVID. To avoid overcomplicating the mathematics, we assume that people who recover reenter the susceptible pool, that being cured of the disease does not create future immunity.

# SIR Model

image

where $P_{\text{spread}}$, $P_{\text{contact}}$, and $P_{\text{recover}}$ equal the probability of spreading the disease, the probability of contact, and the probability of recovery.

Epidemiologists keep separate track of the probability of contact and the probability of spreading, so we will as well. Contact rates depend on how the disease passes from one person to another. HIV spreads through sexual contact. Diphtheria spreads through saliva. Flu viruses spread through the air. Thus flu has a higher contact probability than diphtheria, which has a higher contact probability than HIV. Once contact occurs, the probability of spread also varies. Pertussis (whooping cough) transfers to another person more readily than COVID.

The SIR model produces a *tipping point* at what is known as the *basic reproduction number ($R_0$),* the ratio of the probability of contact times the probability of spread to the probability of recovery. A disease with an $R_0$ greater than 1 can spread through the population. Diseases with $R_0$'s less than 1 dissipate. In this model, the information, or in this case the disease, need not spread to the entire relevant population. Whether or not it does depends on the value of $R_0$. Hence, government agencies like the Centers for Disease Control rely on estimates of $R_0$ to guide policy.[6]

## $R_0$: The Basic Reproduction Number

image

As shown in the table below, measles, which can spread through the air, has a higher reproduction number than HIV, which spreads through sexual contact and needle sharing.

image

Estimates of $R_0$ do not assume that people change their behavior in response to a disease. People responded to the COVID pandemic by wearing masks, keeping distances from others, and avoiding large crowds. If these actions lower the probabilities of contact and spread sufficiently, they create an *effective reproduction number* which will be less than the basic reproduction number. If the effective reproduction number, denoted by $R_t$, falls below one, the virus will cease to spread. Quarantines offer a sure fire way reduce $R_t$ below one, but they are costly.[7]

Vaccines also stop the spread of disease. a vaccine exists, then vaccination can prevent disease spread. Disease spread can be prevented even without vaccinating everyone. The proportion of people who must be vaccinated, the *vaccination threshold,* is given by the formula image which we can derive from the model.[8]

The vaccination threshold increases with $R_0$. To prevent the spread of polio, which has an $R_0$ of 6, the vaccine must cover image of the population. To stop the spread of measles, which has an $R_0$ of 15, the vaccine must cover image of the population. The mathematical derivation of the vaccination threshold provides guidance to policymakers. If too few people are vaccinated, then the disease will spread, so governments vaccinate more than the threshold amount estimated by the model. For diseases with high basic reproduction numbers, such as measles and polio, governments try to vaccinate everyone.

Some people worry about side effects of vaccines and choose not to participate in vaccination programs. If these people constitute a small percentage of the population, the vaccination of others prevents them from catching the disease. Epidemiologists call this phenomenon *herd immunity*. The people who choose not to get the vaccine *free ride* off the vaccinations of others. We study free riding in greater detail later in the book.[9]

## $R_0$, Superspreaders, and Degree Squaring

The derivation $R_0$, the basic reproduction number, assumes random mixing: in each time step, individuals in the population randomly meet one another. As noted above, the random mixing assumption may approximate airborne diseases or diseases spread by touch, but it makes less sense for diseases that spread through sexual contact.

If we embed the SIR model on a network, we see the importance of the degree distribution to disease spread. Here, we compare a *rectangular grid network* (a checkerboard)—where each node is connected to the nodes to the north, south, east, and west—to a *hub-and-spoke network* where a hub node connects to all other nodes.

Assume that a disease randomly occurs at a node. We set $P_{contact} = 1$ within the network so that each person makes contact with everyone to whom he is connected. In the next period, the disease potentially spreads to each neighbor independently with a given probability corresponding to the virulence of the disease.

First consider the rectangular grid network. In each period, the disease can spread to any of the four nodes to the north, south, east, and west. If the probability of the disease spreading exceeds image, we would expect the disease to spread. If we look ahead one period, we see that if one new node caught the disease, then that node has three possible neighbors who could catch the disease. If two neighbors, those to the north and east of the original node, caught the disease, then there exist five new nodes to which the disease could spread. This network, in this case, does not have much of an effect on the likelihood of disease spread.

image

Next, consider the hub-and-spoke network. The first node to get the disease could be the hub or a spoke. If the hub catches the disease, then it could spread the disease to any one of the spokes. We would expect the disease to spread, even for a low probability of spreading. If a spoke caught the disease, then the only possible node that could catch the disease is the hub. And as we just learned, if the hub catches the disease, the disease will spread even for low probabilities of spreading.

For the hub-and-spoke network, $R_0$ is less informative because if the hub catches the disease, the disease will spread. Epidemiologists refer to high-degree hub people as *superspreaders*. Superspreaders contributed to the early spread of both HIV and SARS.[10] A superspreader need not be extremely social or well connected. A superspreader may have an occupation—tollbooth operator, bank teller, dental hygienist—that puts him in contact with people who belong to distinct social networks. Mary Mallon (Typhoid Mary) worked as a cook in New York at the turn of the twentieth century. She moved from family to family infecting each with typhoid fever. Once discovered as the source, Mary was quarantined against her will.

To derive the effect of high-degree nodes, we note first that a high-degree node is both better able to spread the disease and more likely to catch it. A person with three times as many friends as another will be approximately three times as likely to catch the disease and able to spread it three times as widely. His total contribution to the spread of the disease will therefore be nine times that of the other. Thus, a node's contribution to the spread of a disease (or an idea) correlates with the square of the node's degree. If node $A$ has a degree $K$ times larger than node $B$, then node $A$ will be $K$ times as likely to spread the disease and spread it to $K$ times as many others as $B$. Its total effect will be $K^2$ times larger than $B$'s, a phenomenon known as *degree squaring*.

# One-to-Many

Though the SIR model was designed to examine the spread of diseases, we can apply it to social phenomena that spread by diffusion and then fade: books, songs, dance steps, phrases, websites, diets, and exercise regimens. We can estimate probabilities of contact, spread, and recovery and basic reproduction numbers in these contexts as well. The model implies that small changes in these probabilities could spell the difference between success and failure by moving the basic reproduction number above zero. Success can hinge on what John Updike, in describing Ted Williams's last at-bat, called the "tissue-thin difference between a thing done well and a thing done ill."[11] Suppose that you think up a new joke. Making the joke a little bit funnier might push the basic reproduction number above 1 and cause the joke to spread. The same logic applies to the stickiness of ideas. If an idea sticks in people's minds a little longer, the recovery rate will be lower, increasing the basic reproduction number.

Not all cases lie on the threshold. The Beatles had enormous talent. Their reproduction number surely exceeded 1 by a large amount. That is of course conjecture. For current pop stars, we can use internet downloads to estimate basic reproduction numbers. Pop star Justin Bieber had an estimated $R_0$ of 24, making him more virulent than the measles.[12]

In the SIR model, we derived two critical thresholds, $R_0$ and the vaccination threshold. These thresholds are *contextual tipping points*, at which small changes in the environment (the context) have large effects on the outcomes. These differ from *direct tipping points*, where small actions at a particular moment in time forever alter the path of a system. Direct tips occur at unstable points, such as when a ball is perched atop a hill. A small push in either direction sends the ball down one side of the hill or the other. That small push is a direct tip.[13]

At a contextual tipping point, a change in a parameter changes how the system behaves. At a direct tipping point, the trajectory of future outcomes takes a sharp turn. A kink, such as the first bend in the S-shaped adoption curve produced by the diffusion model, satisfies neither definition of tipping point. The kink in the adoption curve corresponds to the point where the slope has maximal increase. At that point, the diffusion is well under way. No tip occurs.

Figure 11.3 shows the number of users of Google+ in its first two weeks.[14] A kink in the graph occurs six days after the release. By that time, the process of diffusion was well under way. It is not the case that Google+ struggled early and that a direct tip occurred on day six, with the result that within two weeks Google+ had over 16 million users. This conflation of tips with sharp upturns leads to an overuse of the term *tipping point*. Moments identified in the news media and on the internet as tipping points rarely satisfy the formal definition.

image

Figure 11.3: A Kink (Not a Tip) in the Number of Google+ Users

We can also think of obesity as an epidemic. Though people cannot catch obesity the way they might catch a cold, they can be socially influenced to adopt behaviors that contribute to obesity.[15] To reverse the obesity epidemic we must lower its basic reproduction number, which can be accomplished by decreasing the probabilities of contact or sharing or increasing the probability of recovery. The SIR model applied to obesity, school dropout rates, or crime is not better than economic or sociological models. It is a different model, so it produces different explanations and predictions. It also possibly points to different actions or policies. It contributes to an ensemble of models that help us make sense of the world. It is not a golden arrow that will solve the problem.

In applying models of broadcast, diffusion, and contagion to social phenomena, we may find that some assumptions hold and others do not. In the spread of a disease, each contact has an independent probability of spreading the disease. In social domains, contagion may become more likely with more exposure because adoption is a choice. We do not choose the flu. We catch it. We do choose to buy tight-fitting jeans. As more people wear tight jeans, we may become more likely to as well. Similar logic applies to the choice to become involved in a social movement, to adopt a new technology, or to get a tattoo. In these cases, as well as in the contagion of beliefs or of trusting behavior, we may have to emend the model to allow for the possibility that the probability of adoption per exposure increases with more exposures.[16] Such modifications are often necessary when broadening the set of applications of a model.