Propagation in Networks

- Understanding how information spreads over the network.

- **Epidemic Models:**
- There are many factors that impact whether persons who are exposed to a disease catch it: age, nutritional state, overall health, natural immunity, length of their contact with disease carrier etc.
- But, it is difficult to decide on every factor for each person.
- So, there is a need to simplifying the assumption to model the process.

- **Compartmental models:**
- This categorizes the people according to their state with respect to the disease.

  - S (Susceptible). These are people who have not yet contracted the disease, but who are susceptible to catching it.
  - I (Infected). These people have caught the disease and are actively infected and contagious.
  - R (Recovered). People in this state have recovered from the disease and are no longer contagious or susceptible to reinfection.

- Combining the first letters, there are four models: SI, SIR, SIRS, and SIS
- SI: Person is susceptible, then can become infected, but once infected never recovers.

Table 10.1 A List of Four Disease Models and Example Diseases for Each

SI	SIR	SIS	SIRS
HIV	Chicken Pox	Strep throat	Whooping Cough
Herpes	Mononucleosis	Common Cold	Syphilis
	Mumps		Chlamydia
	Rubella		Salmonella
	Measles		
	Polio		

Threshold models

- In the compartmental models discussed above, a person is either susceptible to a disease or not; there is no consideration given to the level of susceptibility or to what makes a person more or less likely to catch a disease.
- A threshold model considers how many infected individuals a person must be exposed to before becoming infected.
- This is called a k-threshold model, and k represents the number of neighbors who must be infected for a node to catch the disease.
- If someone can become infected from only one neighbor, this is a 1-threshold model. If three neighbors must be sick for the disease to be passed, it is a 3-threshold model.

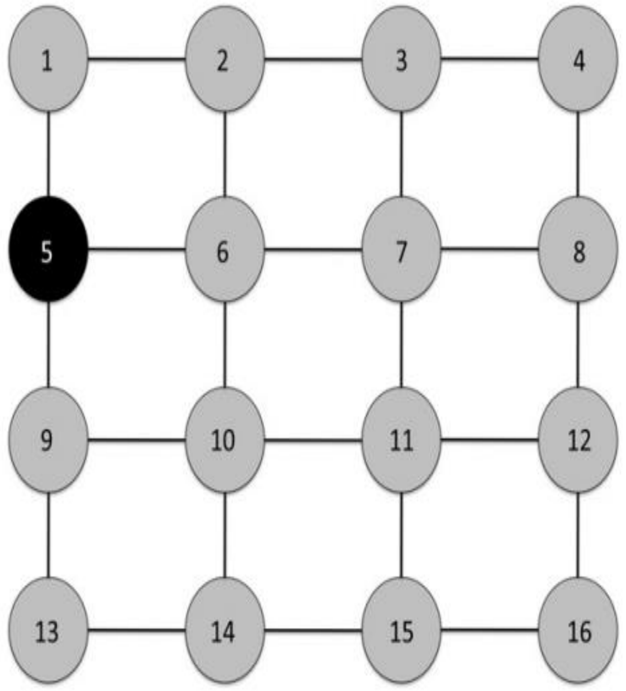


FIGURE 10.1

A simple network with 16 nodes connected in a grid pattern only one of which is infected.

- Node 5, indicated in black, starts off as infected.
- In a 1-threshold model, nodes 1, 6, and 9 could be infected by their sick neighbor, node 5.
- In a 2-threshold model, no nodes can be infected because only node 5 is sick, and thus no node has 2 sick neighbors.
- This is true for any k >1 in this graph since only one node is sick and thus no node will ever have more than one sick neighbor.

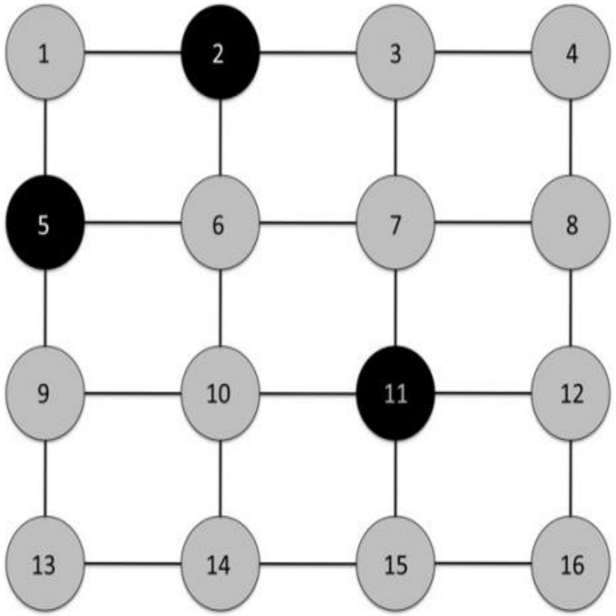


FIGURE 10.2

A grid network with three infected nodes.

- Infections spread in steps; there is an initially infected node, then its neighbors are infected, then those nodes' neighbors, and so on.
- More formally, these stages can be treated as time steps. The network begins at time 0, t = 0.
- Then, at the next step (t =1), the first set of neighbors are infected.
- At time t = 2, the next set are infected, and so on.
- Nodes 2, 5, and 11 are infected at time t =0.
- In a 1-threshold model, all the neighbors of these nodes will be infected at step t = 1 - nodes 1, 3, 6, 7, 9, 10, 12, and 15.
- At time t =2, all the remaining nodes will be infected because their neighbors will be sick.

In a 2-threshold model:

Nodes with two sick neighbors will be infected.

At first, that will only be nodes 1 and 6 (infected by 2 and 5).

Then, at time t = 1, nodes 7 and 10 will be infected by the combination of 6 and 11. and so on.

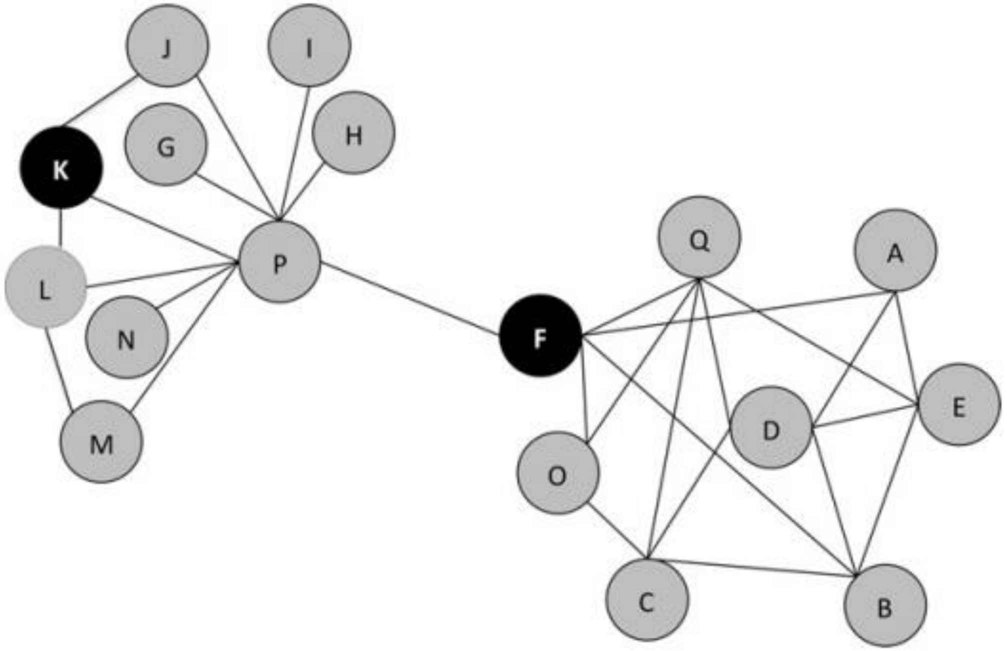


FIGURE 10.3

A more complex network with two infected nodes: F and K.

The firefighter problem

- Depending on the value of k, not all of node P’s neighbors will necessarily become infected, but the possibility of infection is higher when node P is sick. If we wanted to stop the infection from spreading, eliminating the possibility of node P becoming sick would have a big impact.

- Can certain nodes be “vaccinated” or removed from the network so that they are no longer susceptible to infection and thus will never be able to pass it on?
- Conversely, if we want to spread something, like a viral video, are there nodes whose network position allows them to reach a wider audience than other nodes.

- In the **Firefighter Problem**, think of nodes as trees in a forest and as the spreading disease being a fire.
- Trees can catch on fire if neighboring trees are on fire, unless a firefighter is there to prevent it.
- Using a k-threshold model, a tree will catch on fire if k of its neighbors are on fire, unless there is a firefighter.
- At time t = 0, an initial set of fires is present in the network.
- Then, at time t = 1, we can place some number (call it n) of firefighters onto trees in the network.
- At time t = 2, the fire will spread to susceptible trees that are not protected by firefighters.
- At time t = 3, we can place n more firefighters, and at t = 4, the fire spreads again. The problem progresses in alternating turns until the fire is stopped or all nodes are on fire (i.e., infected).

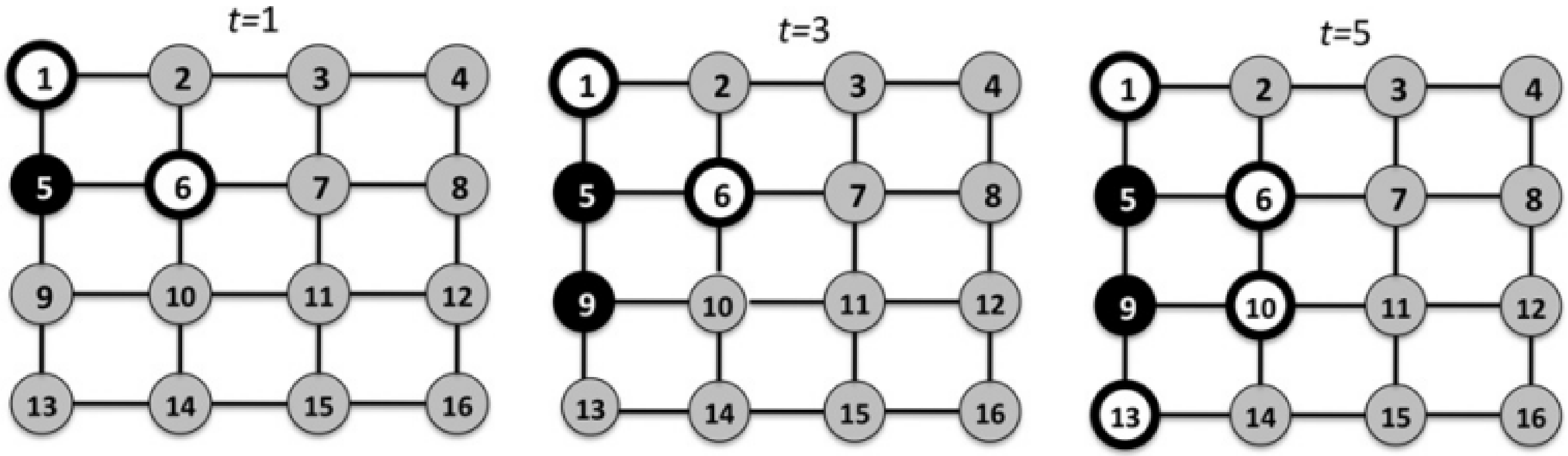


FIGURE 10.4

The placement of firefighters (white circles with black outlines) and progression of the infection (black nodes) at time steps 1 3. After time $t = 3$, there are no susceptible nodes adjacent to the infected nodes, so the disease stops spreading.

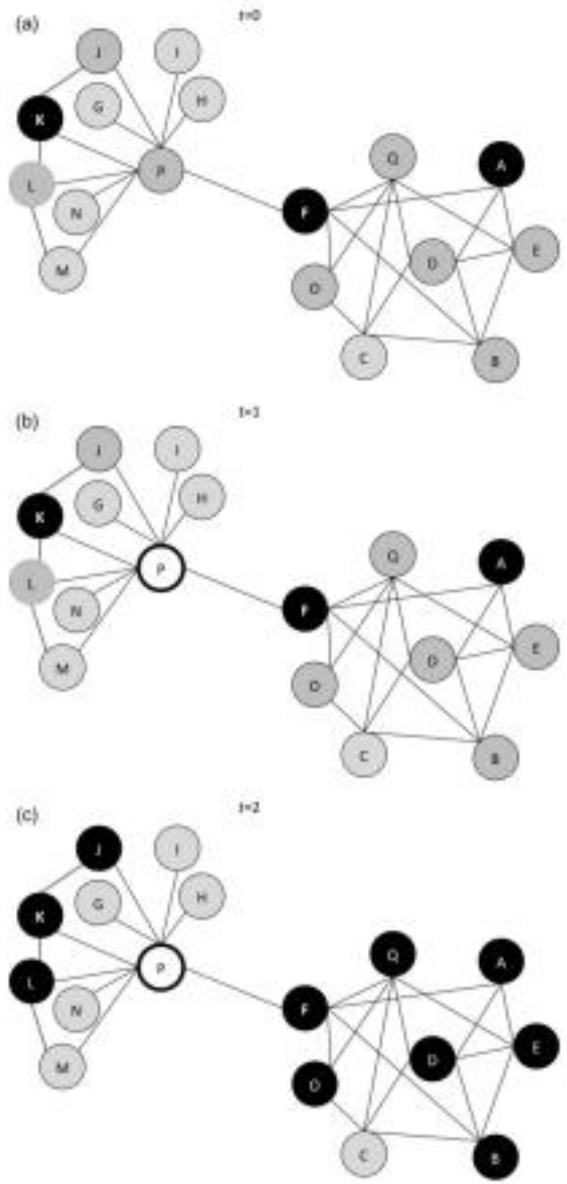


FIGURE 10.5

The spread of infection and placement of firefighters in a more complex network.

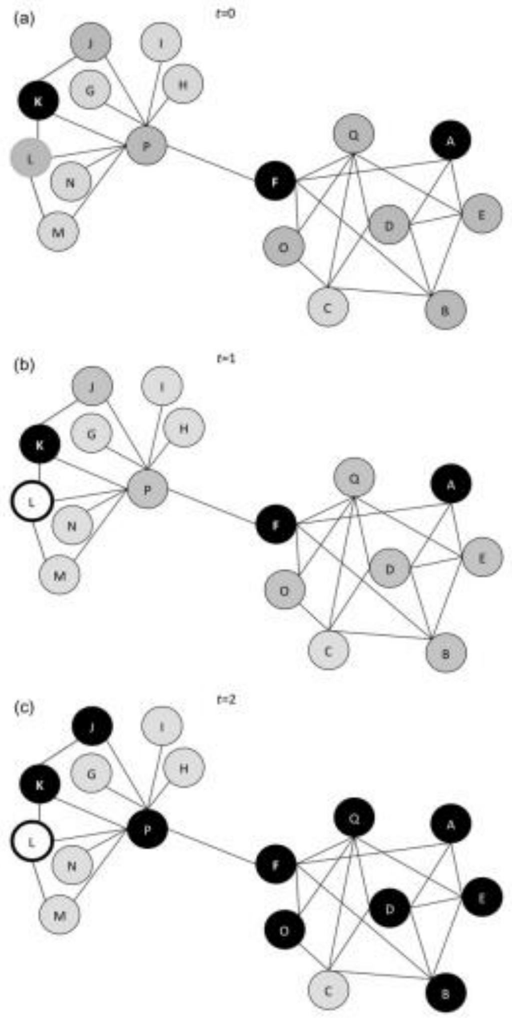


FIGURE 10.6

The spread of infection in the network when different nodes are protected.

Stochastic models

- The compartmental models and threshold models discussed above are deterministic models; they assume that any susceptible individual who is exposed to a infected person (or to k infected people) will become infected.
- Stochastic models introduce probabilities into the models.
- Let **p** be the probability that a disease is transmitted from an infected person to a susceptible person at a given time step.
- The value for p will range from 0 to 1, where 0 means there is a 0% chance of transmission, and 1 means a 100% chance of transmission. If p = 0.6, the chance of transmission is 60%.

$$\begin{aligned} PI(B) * p &= PI(C) \\ 0.8 * 0.8 &= 0.64 \end{aligned}$$

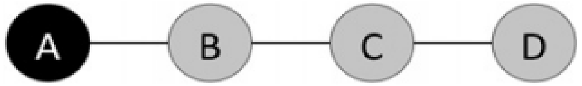


FIGURE 10.7

A small network, where node A begins as infected and nodes B, C, and D are susceptible.

- Node A starts out as infected, and nodes B, C, and D are susceptible.
- Assume a 1- threshold model, and let p =0.8 (an 80% transmission rate).
- At time t = 1, node B can be infected by node A.
- Since p = 0.8, there is an 80% chance that node B gets infected.
- At time t = 2, node C can become infected if node B was infected at time t =1. There are two possibilities.
  - If node B was infected, there is an 80% chance it will pass the disease on to node C.
  - If node B was not infected (which happens 20% of the time), there is no chance node C can catch it, since node B does not have the disease.
- Thus, there are three possible things that can happen:
  - Node B is infected and passes the disease on to C.
  - Node B is infected and does not pass the disease on to C.
  - Node B is not infected and thus cannot pass the disease on to C

To find out how likely it is that node C becomes infected, the probability of each option has to be considered.

- The other two possible outcomes also have a probability of happening. The probability that B is sick but does not pass on the disease is

$$PI(B) * (1 - p)$$
$$0.8 * (1 - 0.8) = 0.8 * 0.2 = 0.16$$

- The probability of node D becoming infected follows clearly. Node C is the only node that can infect D. The probability that C is infected is 0.64. The probability of transmission is 0.8. Thus, the chance node D is infected is:

$$\begin{aligned} PI(C) * p &= PI(D) \\ 0.64 * 0.8 &= 0.512 \end{aligned}$$

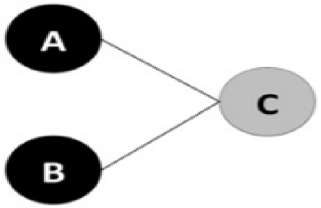


FIGURE 10.8

A network where nodes A and B are infected and node C is susceptible.

Threshold model and the probability that nodes A and B are infected is 1 (a 100% chance of infection). Let p = 0.6. What is the chance that C becomes infected?

There are three scenarios where C can be infected in a 1-threshold model:

- Node A passes the disease, Node B passes the disease.
- Node A passes the disease, Node B does not pass it.
- Node A does not pass the disease, Node B passes it.

Note that the other possibility is that neither node passes the disease, but in that case, node C does not get infected.

- Node A passes the disease, Node B passes the disease $0.6 * 0.6 = 0.36$
- Node A passes the disease, Node B does not pass it $0.6 * 0.4 = 0.24$
- Node A does not pass the disease, Node B passes it $0.4 * 0.6 = 0.24$

- Thus, the chance that C is infected is $0.36 + 0.24 + 0.24 = 0.84$

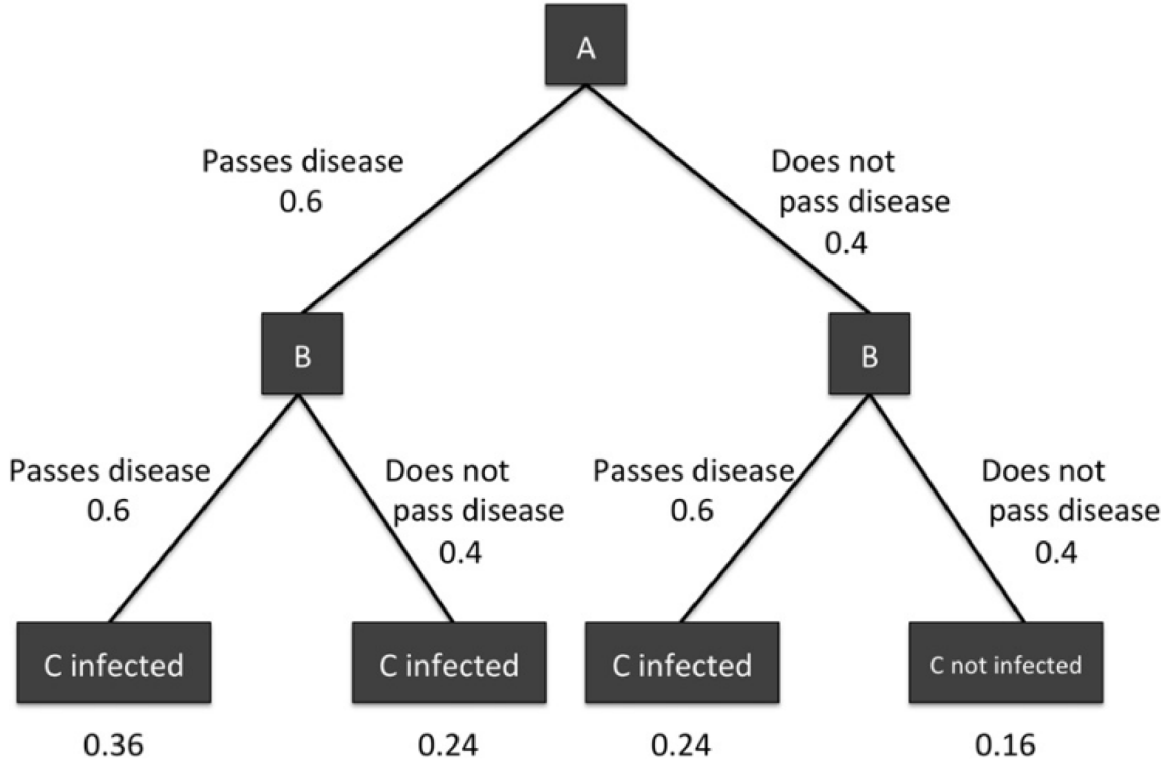


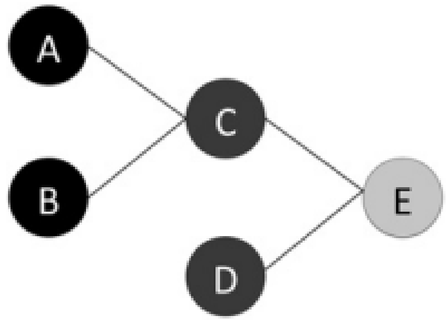
FIGURE 10.9

A tree showing the possibilities of infection from node A, and then from node B. The bottom row shows all four possibilities and the probability of each happening. Note that the probabilities on the bottom are the product of the values on the edges leading to that option.

- If this same network is considered with a 2-threshold model, then both nodes connected to C must be infected and pass on the disease in order for C to catch the disease.

- None of the other scenarios will lead to infection in a 2-threshold model, so the chance that C becomes infected is simply 0.36.

An extension of the network



- It includes node C, which has a 0.84 chance of being infected, and two additional nodes: D and E.
- Node D also has a probability of being infected. Let PI(D) = 0.7.
- *Knowing this, what is the probability that E becomes infected?*

- In a 1-threshold model, there are several more scenarios that will lead to infection

- In a 1-threshold model, there are several more scenarios that will lead to infection. First, we compute the probability for each of those scenarios. That is the probability of the scenario for C (infected or not) times the probability of the scenario for D (infected or not):

  - Nodes C and D are both infected: 0.84 * 0.7 = 0.588
  - Node C is infected but node D is not: 0.84 * 0.3 = 0.252
  - Node C is not infected but node D is: 0.16 * 0.7 = 0.112
  - Neither node is infected: 0.16 * 0.3 = 0.048

If C and D are indeed infected, the same three scenarios presented in the scenario with A and B infecting C apply:

- Node C passes the disease, Node D passes the disease $0.6 * 0.6 = 0.36$
- Node C passes the disease, Node D does not pass it $0.6 * 0.4 = 0.24$
- Node C does not pass the disease, Node D passes it $0.4 * 0.6 = 0.24$

So, if C and D are infected, the chance that the disease is passed to node E is 0.84. However, nodes C and D are not necessarily infected. Thus, we have to multiply 0.84 by the probability that both nodes are infected.

$$0.588 * 0.84 = 0.494$$

If node C is infected but node D is not, the chance that C passes the disease is simply p (0.6). So the total chance of the disease passing in this case is the chance of the scenario happening times the chance of infection:

$$0.252 * 0.6 = 0.151$$

The same applies to the scenario where node D is infected, but node C does not:

$$0.112 * 0.6 = 0.067$$

To get the final probability of node E becoming infected, we add the probability of it being infected when both nodes are sick, plus the probability of being infected if only C is sick, plus the probability of being infected if only D is sick:

$$0.494 + 0.151 + 0.067 = 0.712$$

In a 2-threshold model, the case is again simpler. The only way to be infected is if both C and D are infected and if they both pass it. The chance that both nodes are infected, as computed above, is 0.588. The chance that both nodes pass the disease in this case is 0.36 (again, calculated above). Thus, the chance that node E is infected is:

$$0.588 * 0.36 = 0.212$$

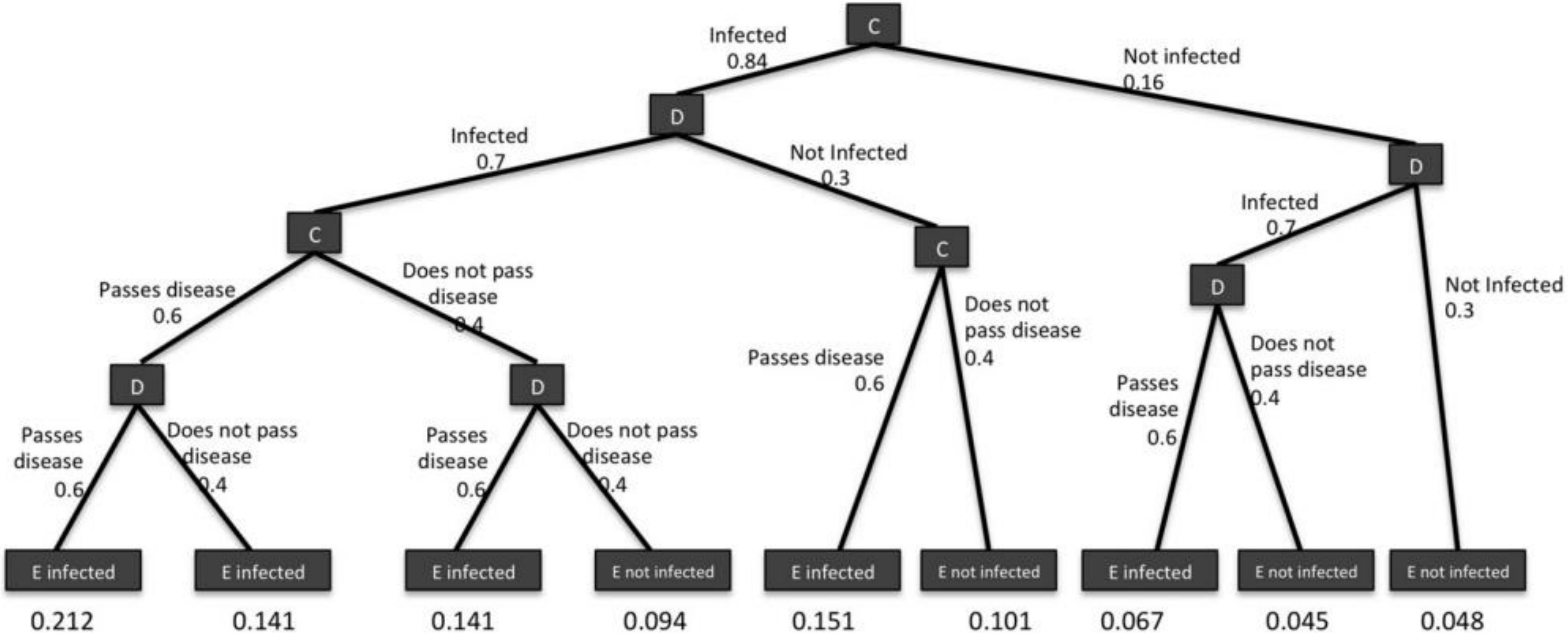


FIGURE 10.11

The full set of scenarios for nodes C and D being infected and passing the disease to node E.

Applications of epidemic models to social media

- Viral marketing is a topic of great interest to companies, especially online retailers, because it spreads information about products at no cost to the seller. This is good if the product reviews and feedback are positive, but it can backfire if negative information spreads virally.
- A method of marketing whereby consumers are encouraged to share information about a company's goods or services via the internet.
- Researchers (Leskovec et al., 2007) investigated “viral” recommendations that users send to one another and their impact on purchasing behavior.
- Nodes in the network were users of an e-commerce system, and there was an edge between users if one recommended an item to the other.
- The researchers used an SIR model, where susceptible individuals have not purchased an item, an “infected” person buys an item based on a recommendation, and after purchase, a person is “recovered” and not susceptible to buying the item again.
- Using real data from the system, the researchers looked at a threshold model to understand how recommendations spread in a network.
- They found that, overall, recommendations did not spread much in a viral way, but for some products the recommendations did have more of a reach into the network.