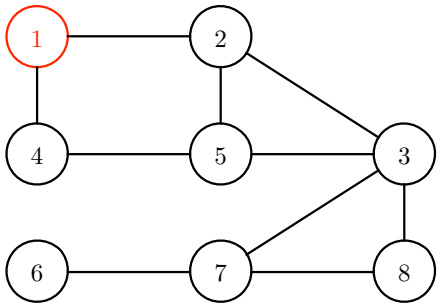
11.6 PROJECTS

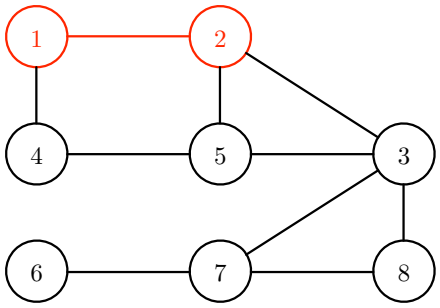
Project 11.1 Diffusion of ideas and influence

In this project, you will investigate how memes, ideas, beliefs, and information propagate through social networks, and who in a network has the greatest influence. We will simulate diffusion through a network with the ***independent cascade model***. In this simplified model, an idea originates at one node (called the seed) and each of the neighbors of this node adopts it with some fixed probability (called the *propagation probability*). This probability measures how influential each node is. In reality, of course, people have different degrees of influence, but we will assume here that everyone is equally influential. Once the idea has propagated to one or more new nodes, each of these nodes gets an opportunity to propagate it to each of their neighbors in the next round. This process continues through successive rounds until all of the nodes that have adopted the idea have had a chance to propagate it to their neighbors.

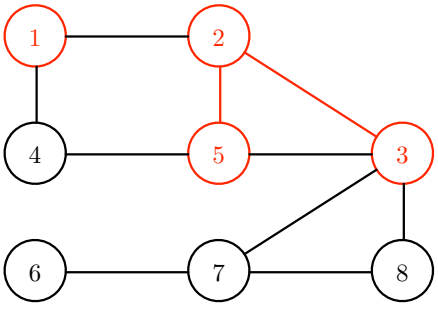
For example, suppose node 1 in the network below is the seed of an idea.



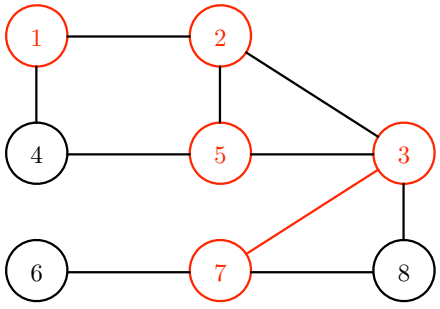
Initial state



Round 1



Round 2



Round 3

In round 1, node 1 is successful in spreading the idea to node 2, but not to node 4. In round 2, the idea spreads from node 2 to both node 3 and node 5. In round 3, the idea spreads from node 3 to node 7, but node 3 does not successfully influence node 8. Node 5 also attempts to influence node 4, but is unsuccessful. In round 4 (not shown), node 7 attempts to spread the idea to nodes 6 and 8, but is unsuccessful, and the process completes.

Copyright Taylor and Francis, 2021

Part 1: Create the network

For this project, you will use a network that is an anonymized version of a very small section of the Facebook network with 333 nodes and 2,519 links. The file containing this network is available on the book website. The format of the file is the same as the format described in Exercise 11.1.7: each node is represented by a number, and each line represents one link. If you have not already done so in Exercise 11.1.7, write a function

```
readNetwork(fileName)
```

that reads in a network file with this format, and returns an adjacency list representation of the network.

Part 2: Simulate diffusion with a single seed

Simulating the diffusion of an idea through the network, as described above, is very similar to a breadth-first search, except that nodes are visited probabilistically. Just as visited nodes are not revisited in a BFS, nodes that have already adopted the idea are not re-influenced. Using the `bfs` function from Section 11.2 as a starting point, write a function

```
diffusion(network, seed, p)
```

that simulates the independent cascade model starting from the given seed, using propagation probability `p`. The function should return the total number of nodes in the network who adopted the idea (i.e., were successfully influenced).

Part 3: Rate each node's influence

Because this is a random process, the `diffusion` function will return a different number every time it is run. To get a good estimate of how influential a node is, you will have to run the function many times and average the results. Therefore, write a function

```
rateSeed(network, seed, p, trials)
```

that calls `diffusion` `trials` times with the given values of `network`, `seed` and `p`, and returns the average number of nodes influenced over that many trials.

Part 4: Find the most influential node(s)

Finally, write a function

```
maxInfluence(network, p, trials)
```

that calls the `rateSeed` function for every node in the network, and returns the most influential node.

Combine these functions into a complete program that finds the most influential node(s) in the small Facebook network, using a propagation probability of 0.05. Then answer the following questions. You may write additional functions if you wish.

Question 11.1.1 *Which node(s) turned out to be the most influential in your experiments?*

Question 11.1.2 *Were the ten most influential nodes the same as the ten nodes with the most friends?*

Question 11.1.3 *Can you find any relationship between a person's number of friends and how influential the person is?*