12 – NETWORK ANALYSIS

CS 1656

Introduction to Data Science

Alexandros Labrinidis – http://labrinidis.cs.pitt.edu
University of Pittsburgh

NETWORK CHARACTERISTICS

The importance of networks

- The World Wide Web
- Computer networks
- Communication networks
- Social networks (online and physical)
- Networks of concepts/ideas
- Citation networks
- Organizational networks
- Transportation networks
- Physical infrastructure networks (power, water, gas)
- Environmental interrelationship networks
- Food chain networks

Characteristics of Nodes & Edges

- **Degree Centrality** of node n

  = degree of node n, i.e., number of edges n has

- **Closeness Centrality** of node n

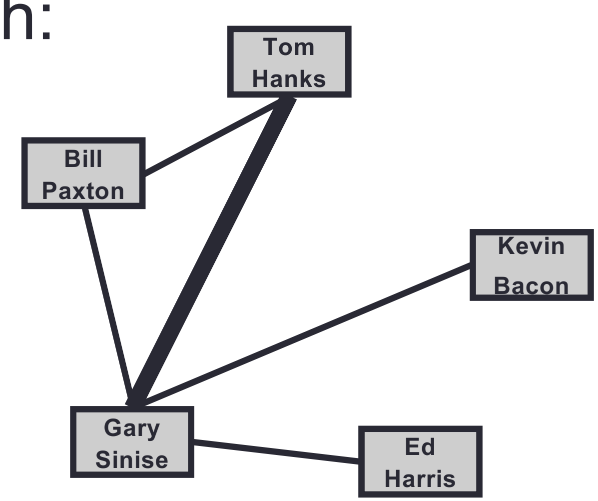
  = average of shortest-path length of n to all other nodes

- **Betweenness Centrality** of node n

  = fraction of shortest-paths among pairs of nodes that include n

Degree Centrality Example

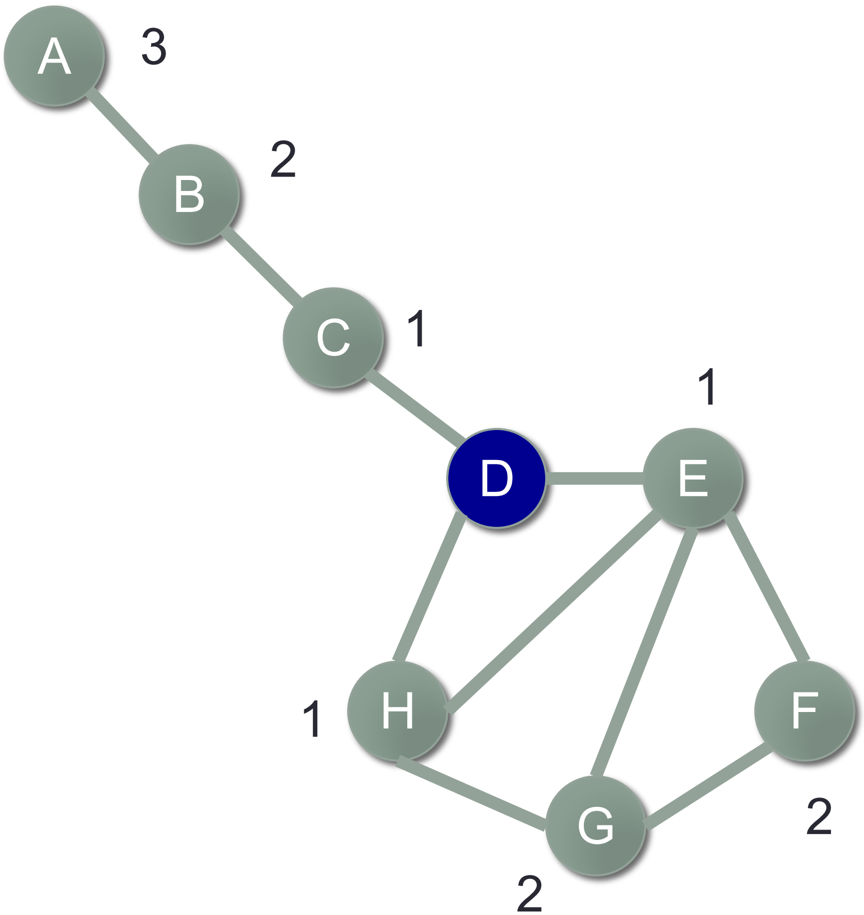
- Assume the following graph:



Actor	Degree
Tom Hanks	2
Bill Paxton	2
Kevin Bacon	1
Gary Sinise	4
Ed Harris	1

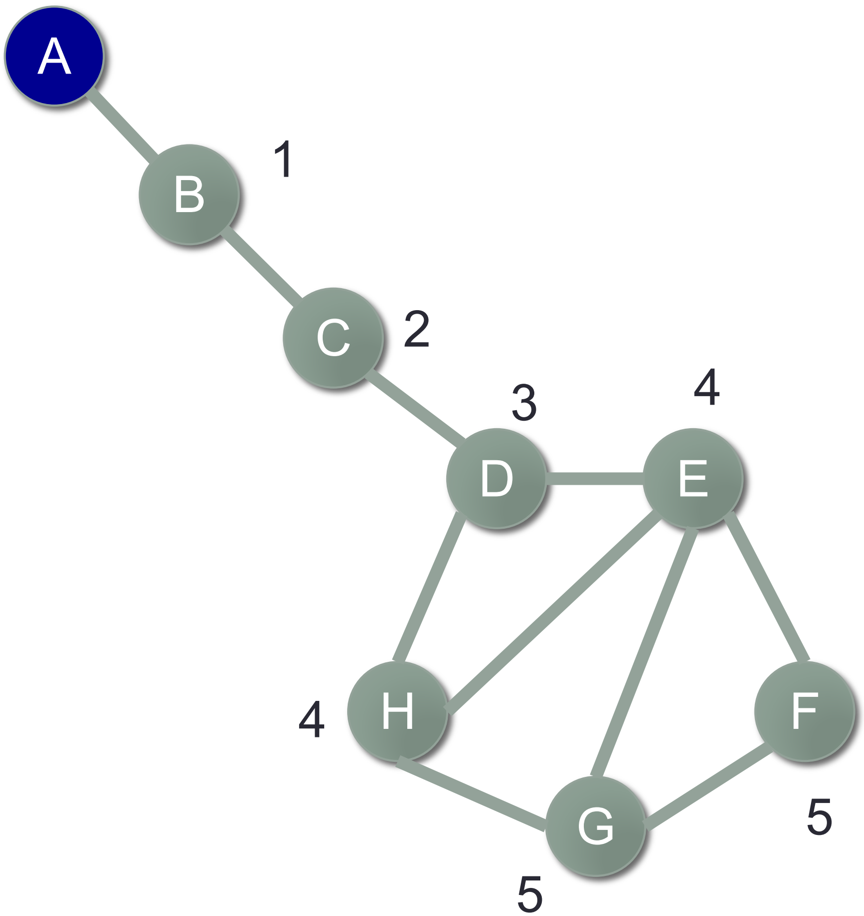
Fun fact: Oracle of Bacon: https://oracleofbacon.org

Closeness Centrality Example



- Shortest paths from D
  - Average shortest path is (3+2x3+1x3)/7 = 12/7 = 1.71

Closeness Centrality Example



- Shortest paths from A
  - Average shortest path is (1+2+3+4x2+5x2)/7 = 24/7 = 3.43

Fun fact: Oracle of Bacon: https://oracleofbacon.org

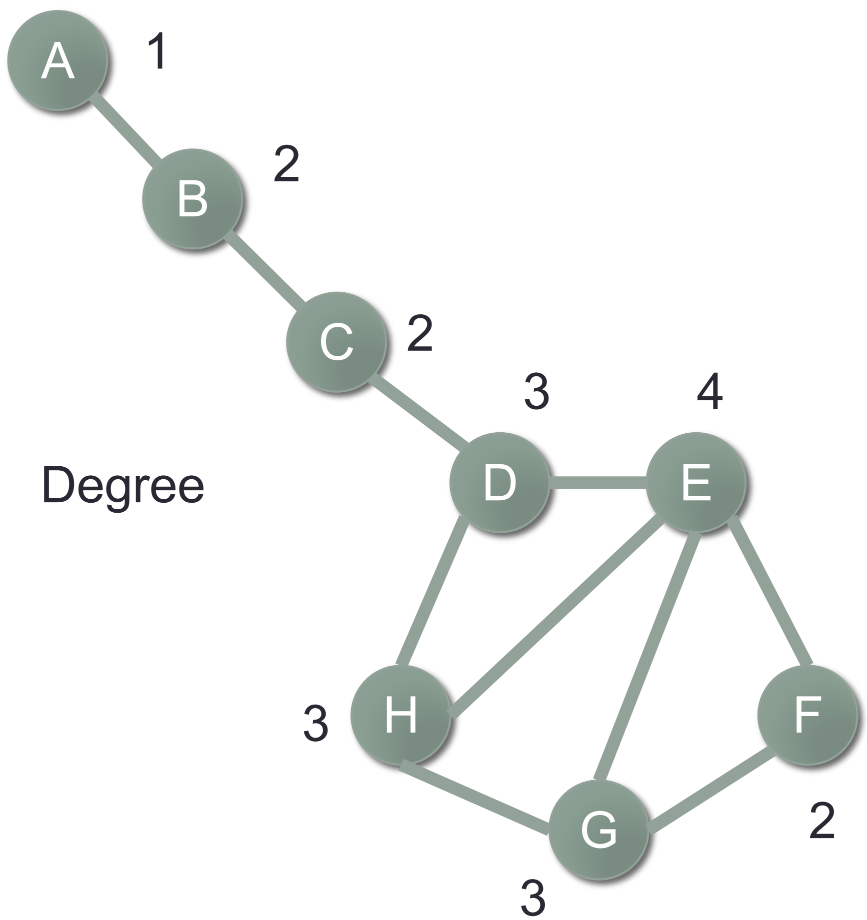
Characteristics of Networks

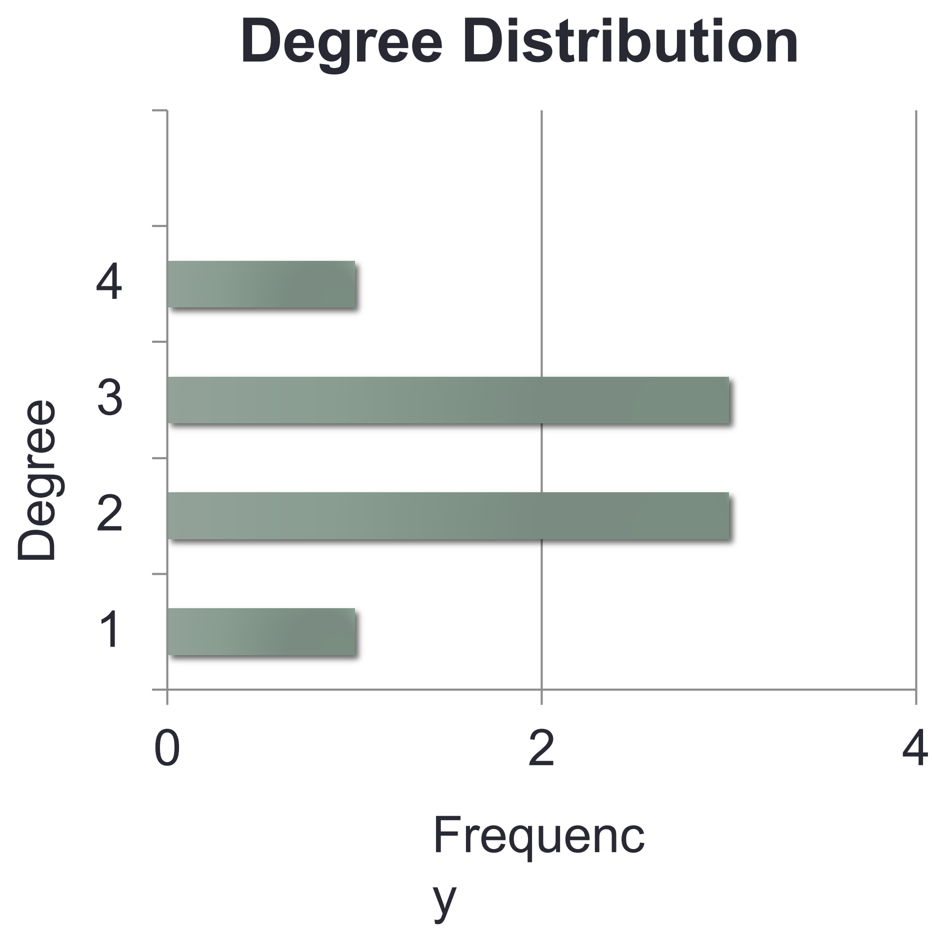
- **Degree Distribution** for a network
  = how many nodes have each possible degree

- **Density** of a network
  = how connected the network is,
  i.e., number of edges / number of possible edges

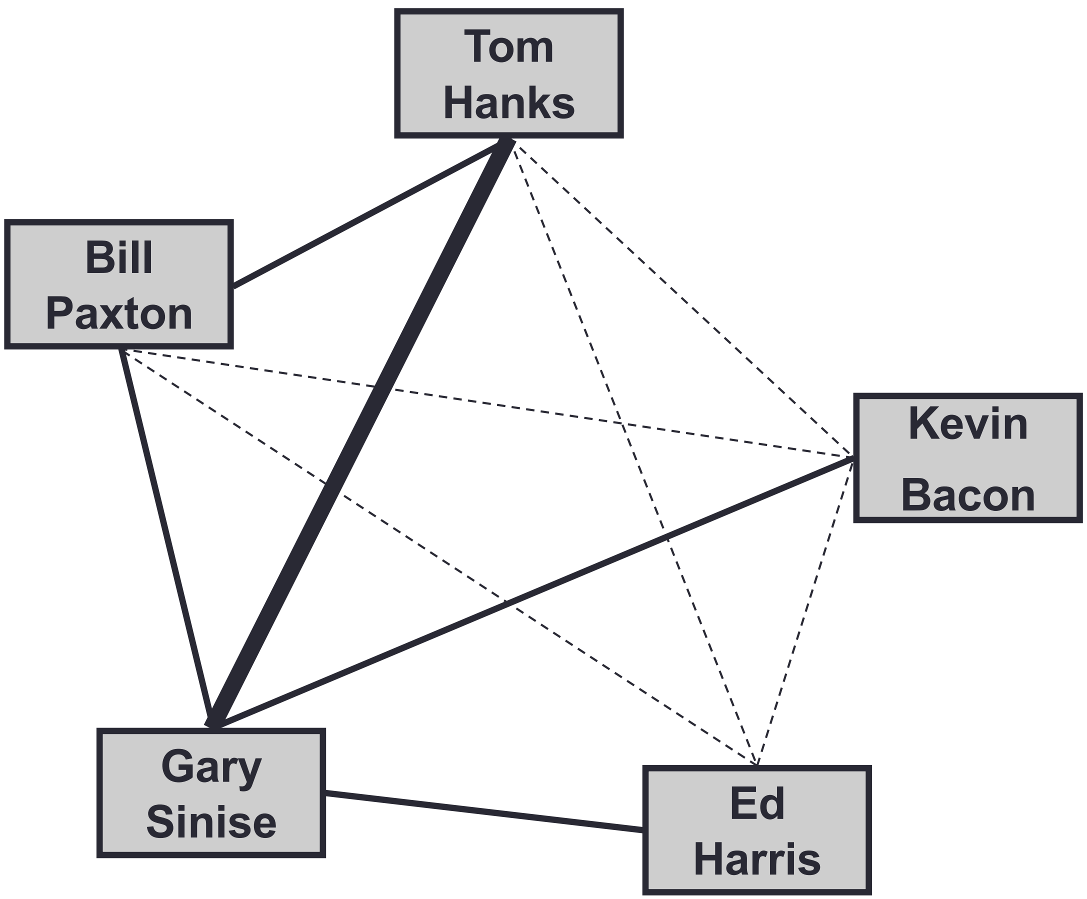
- **Connectivity** of a network
  = minimum number of nodes that would have to be removed
  before the graph becomes disconnected
  (i.e., no longer a path from each node to every other node)

Degree Distribution Example



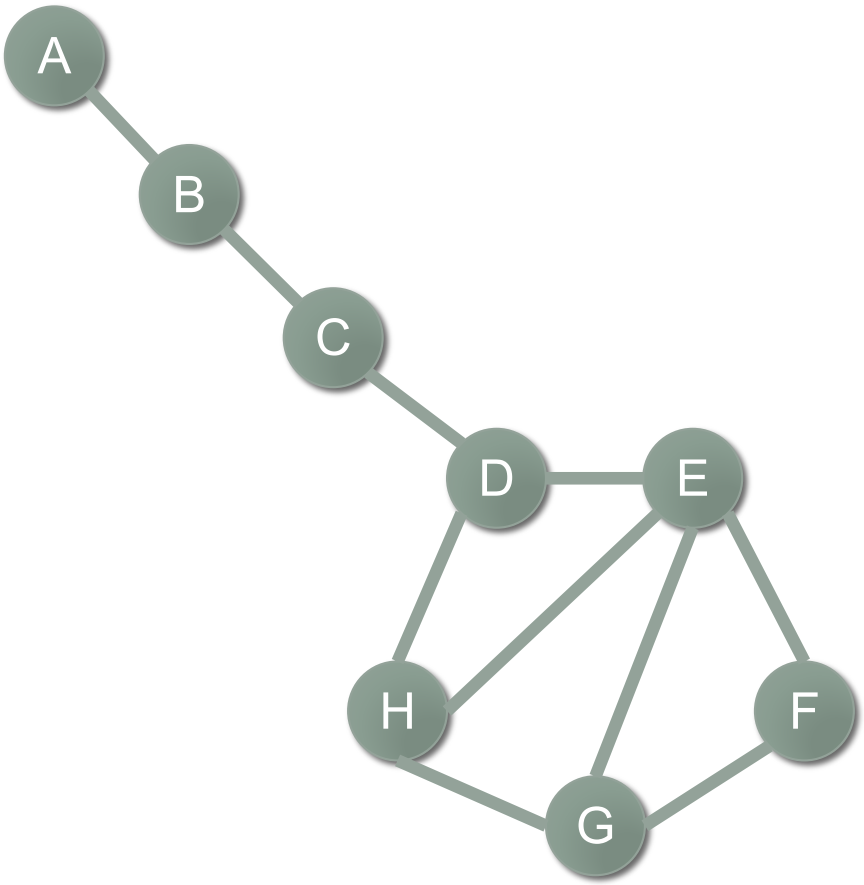


Density Example

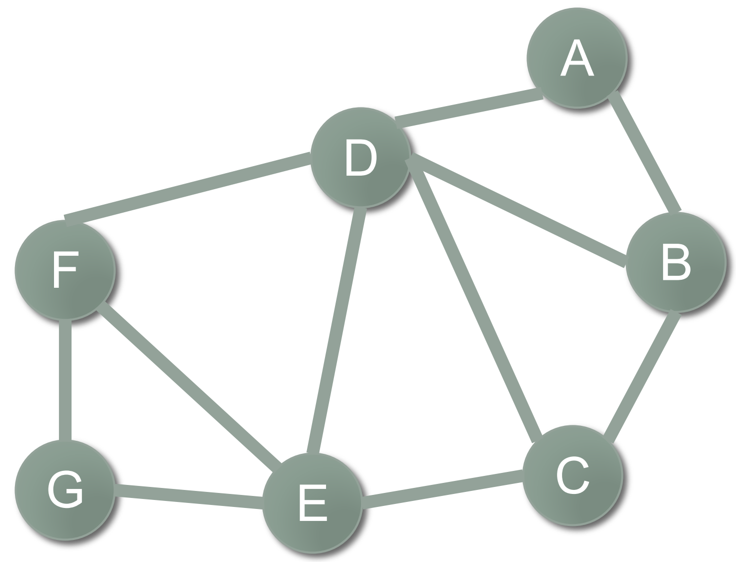


- Edges = 5
- Total possible edges = 10
- General formula:
  - Undirected graph:
    TPE = n * (n - 1) / 2
  - Directed graph:
    TPE = n * (n - 1)
- Density = 5 / 10 = 0.5

Connectivity Examples



- Connectivity = 1 (B, C, D)



- Connectivity = 2

PROPAGATION IN NETWORKS

Epidemic Models

One way to model disease spread is with compartmental models

Categorize people according to their state with respect to the disease:

- **S (Susceptible)** – have not yet contracted disease, but susceptible to catching it
- **I (Infected)** – have caught disease, actively infected and contagious
- **R (Recovered)** – have recovered from disease, no longer contagious, and not susceptible to reinfection

Epidemic models (cont)

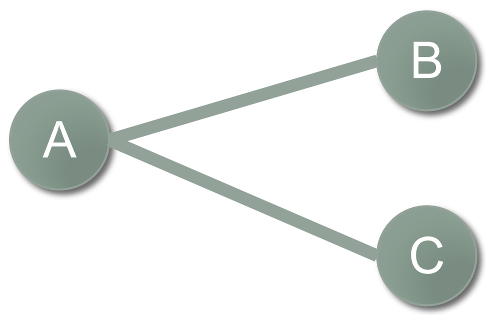
- **S**usceptible – **I**nfected – **R**ecovered

- **SI**: susceptible, infected, never recovers (e.g., HIV)
- **SIR**: susceptible, infected, recovers, then immune (e.g., Chicken Pox)
- **SIRS**: susceptible, infected, recovers (for a period of time), then susceptible again (e.g., Whooping Cough)
- **SIS**: susceptible, infected, then susceptible again (e.g., Common Cold)

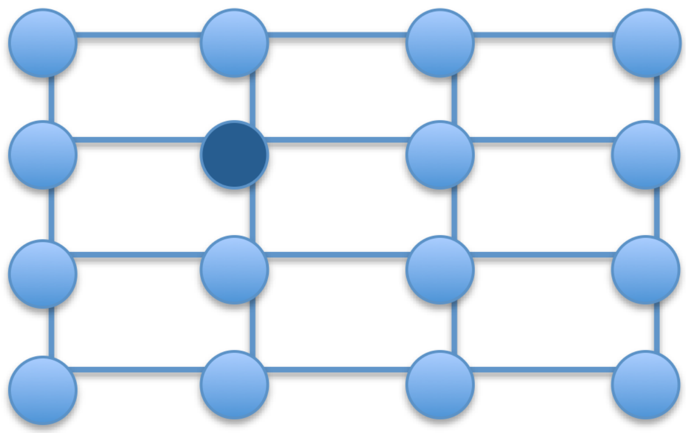
Threshold Models

- Consider how many infected individuals a person must be exposed to before becoming infected

- **k-threshold model**:
  k is the number of neighbors that must be infected for a node to become infected
  - 1-threshold: a node can become infected from only one neighbor
  - Example: node A can become inflected if B OR C are infected



Understanding Question / Q1

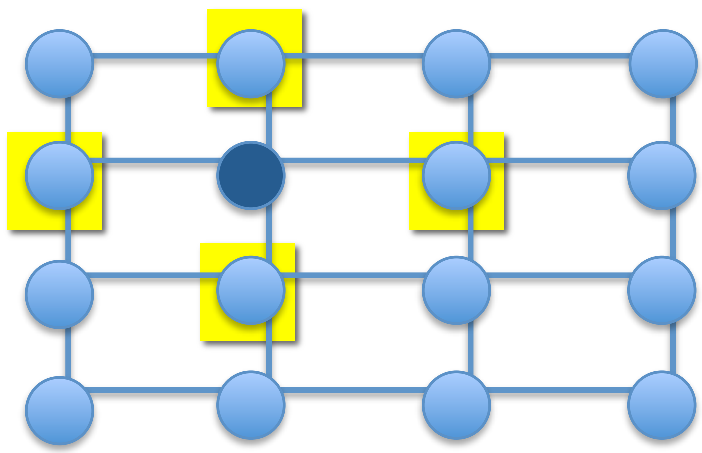


- **Question**:
- Assume dark node is infected. How many time steps does it take for the entire network to be infected?

- **Possible Answers**:
  - 2, 3, 4, 5, 6?

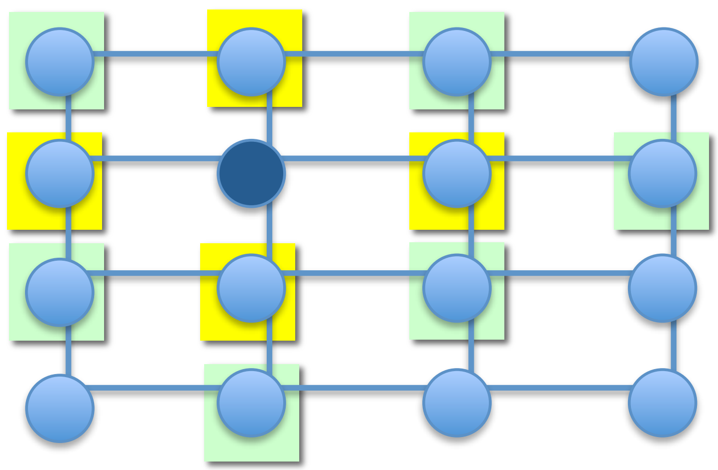
1-threshold example – Step 1

Assume dark node is infected. How many time steps does it take for the entire network to be infected?



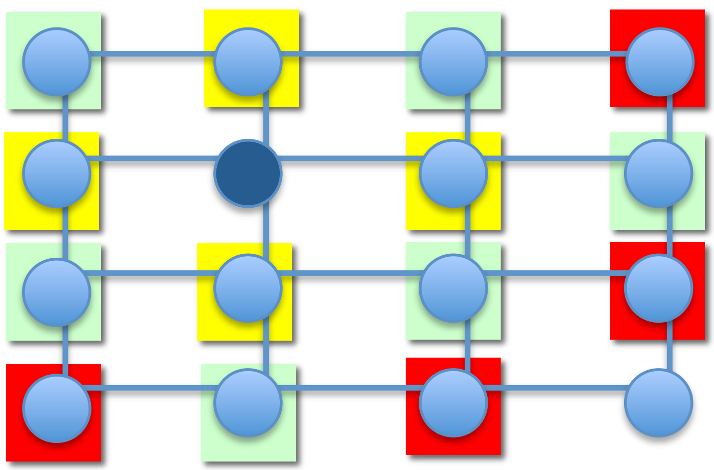
1-threshold example – Step 2

Assume dark node is infected. How many time steps does it take for the entire network to be infected?



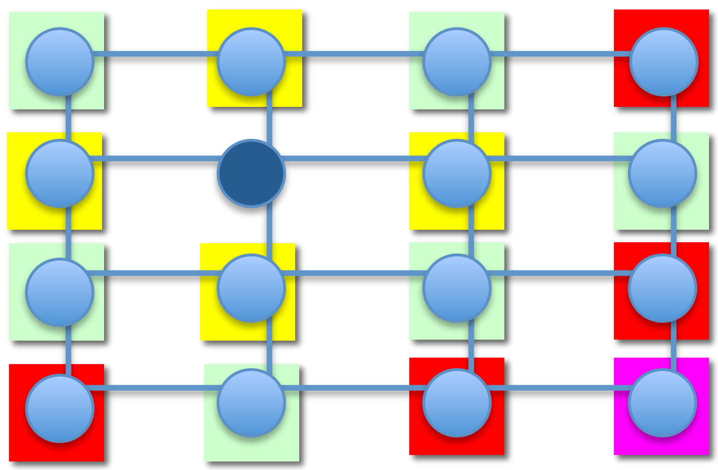
1-threshold example – Step 3

Assume dark node is infected. How many time steps does it take for the entire network to be infected?



1-threshold example – Step 4

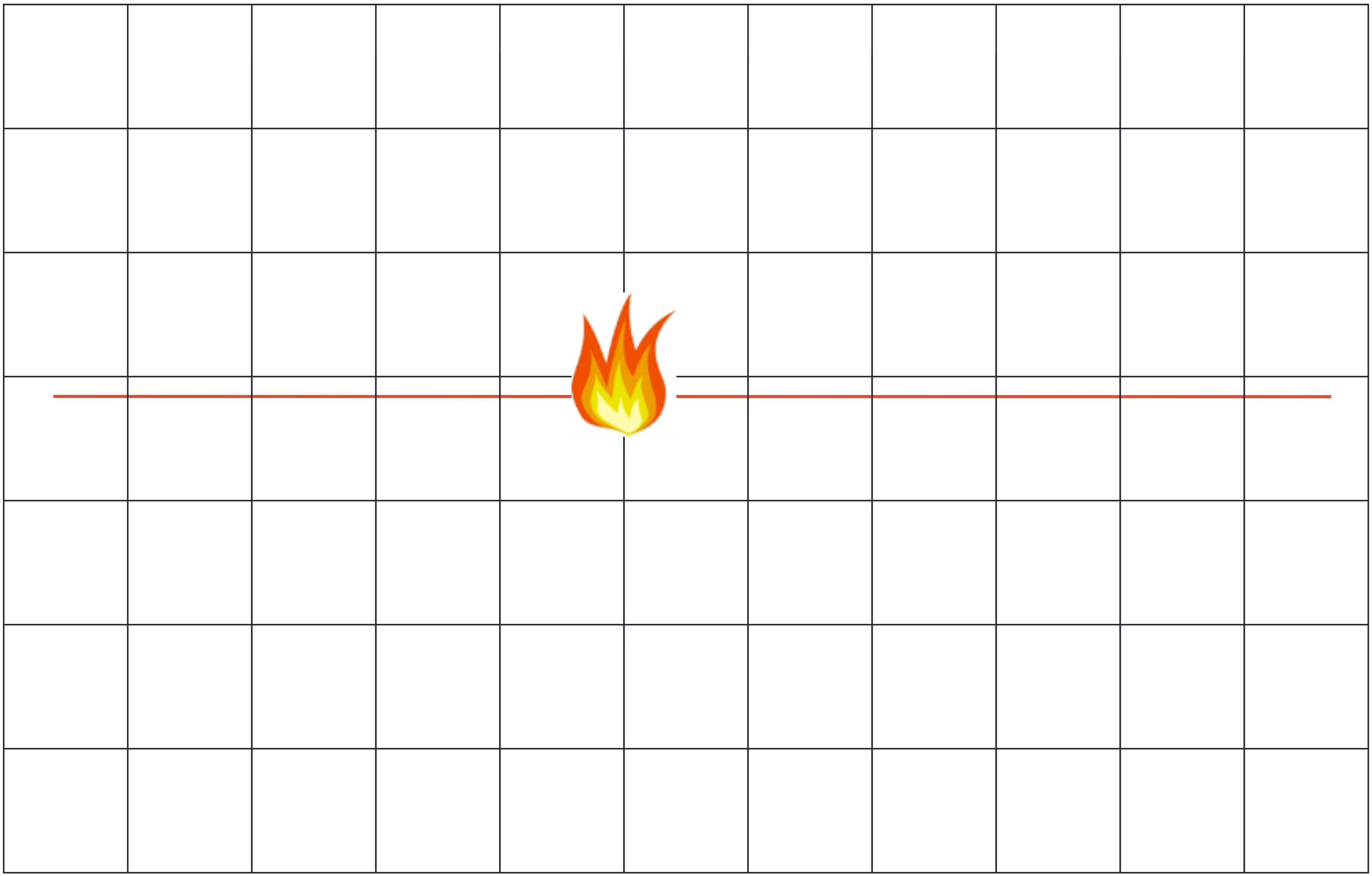
Assume dark node is infected. How many time steps does it take for the entire network to be infected?

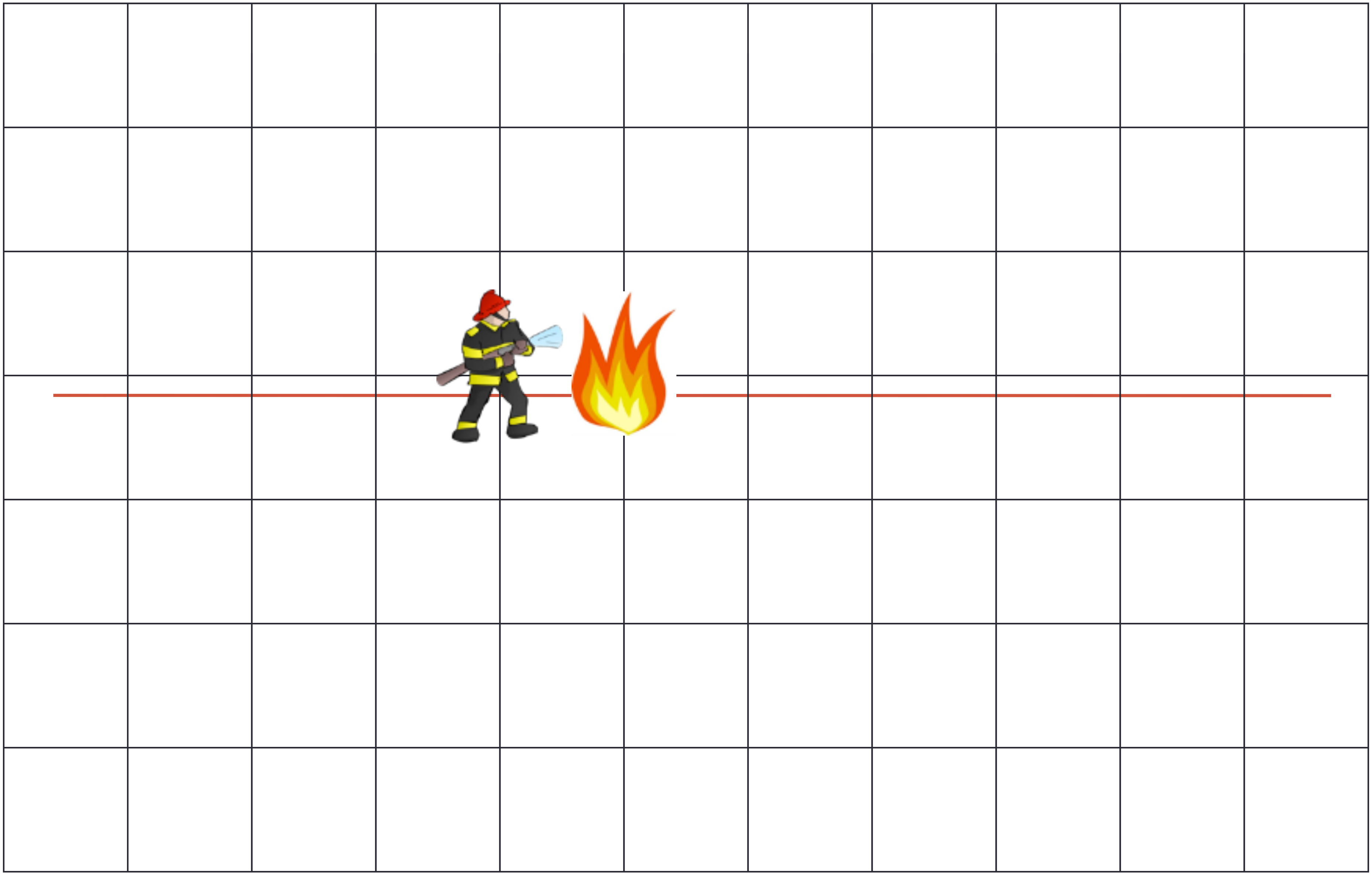


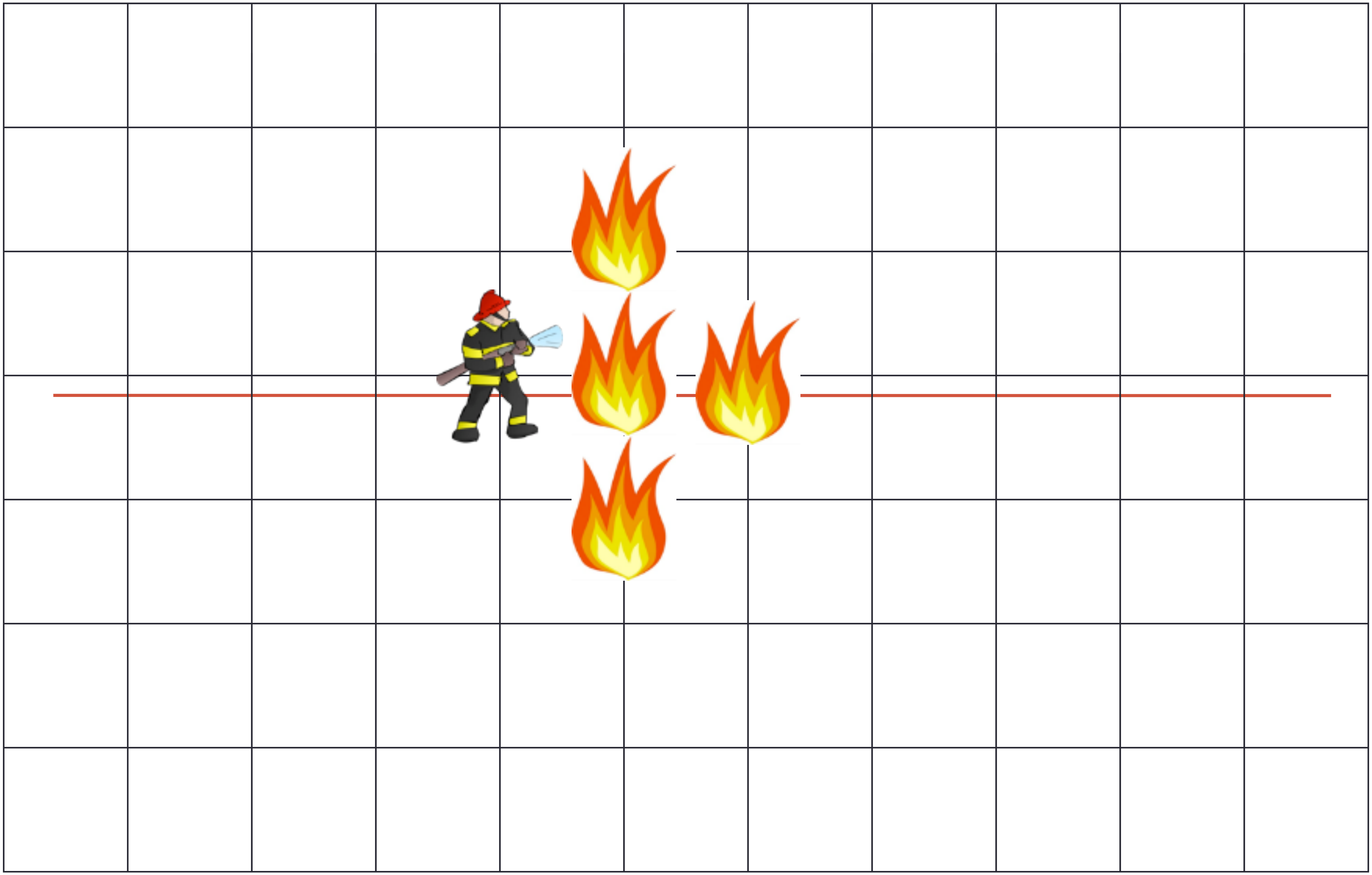
Firefighter Problem

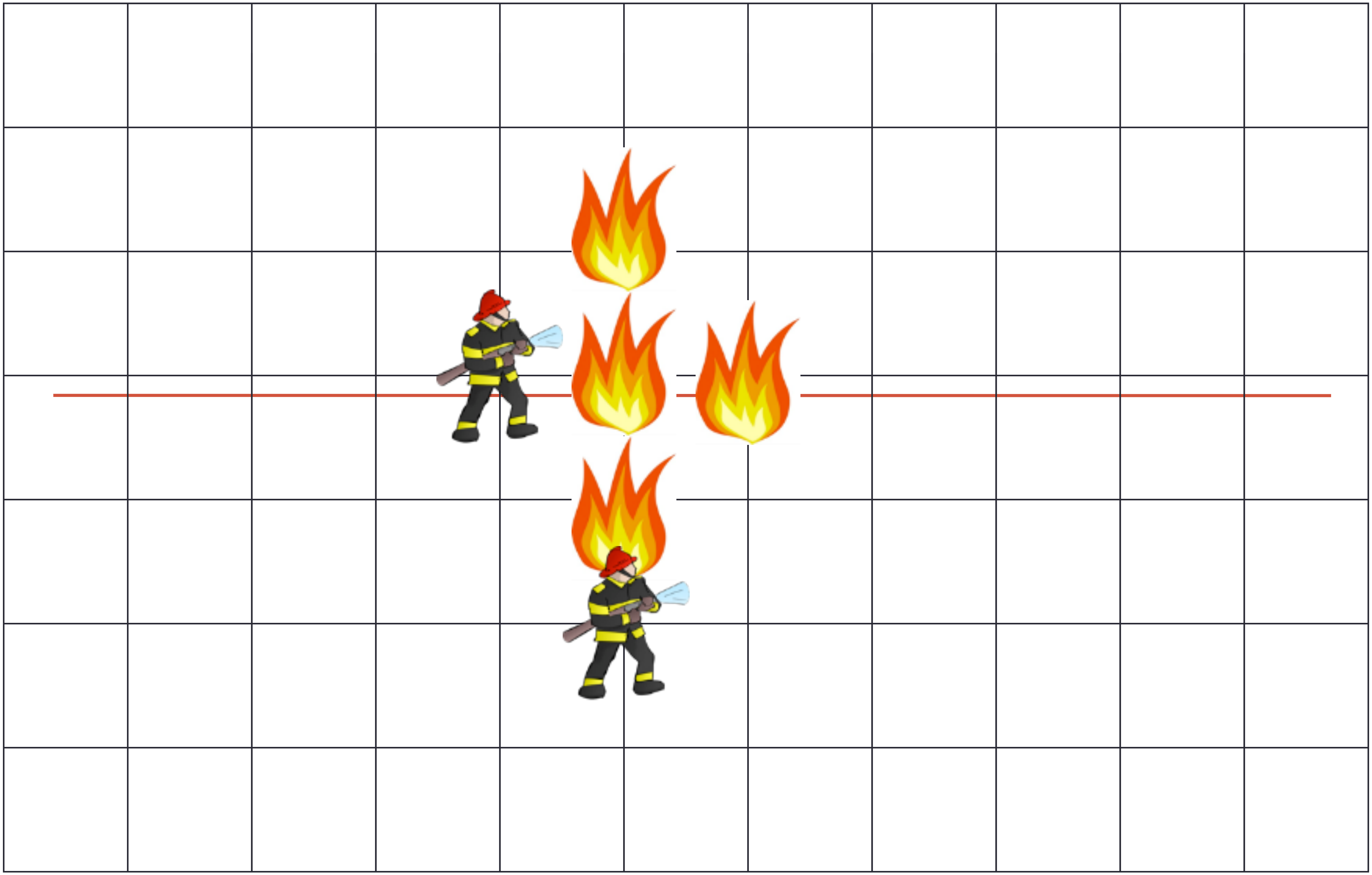
A simple network - a grid where each intersection point is a node.

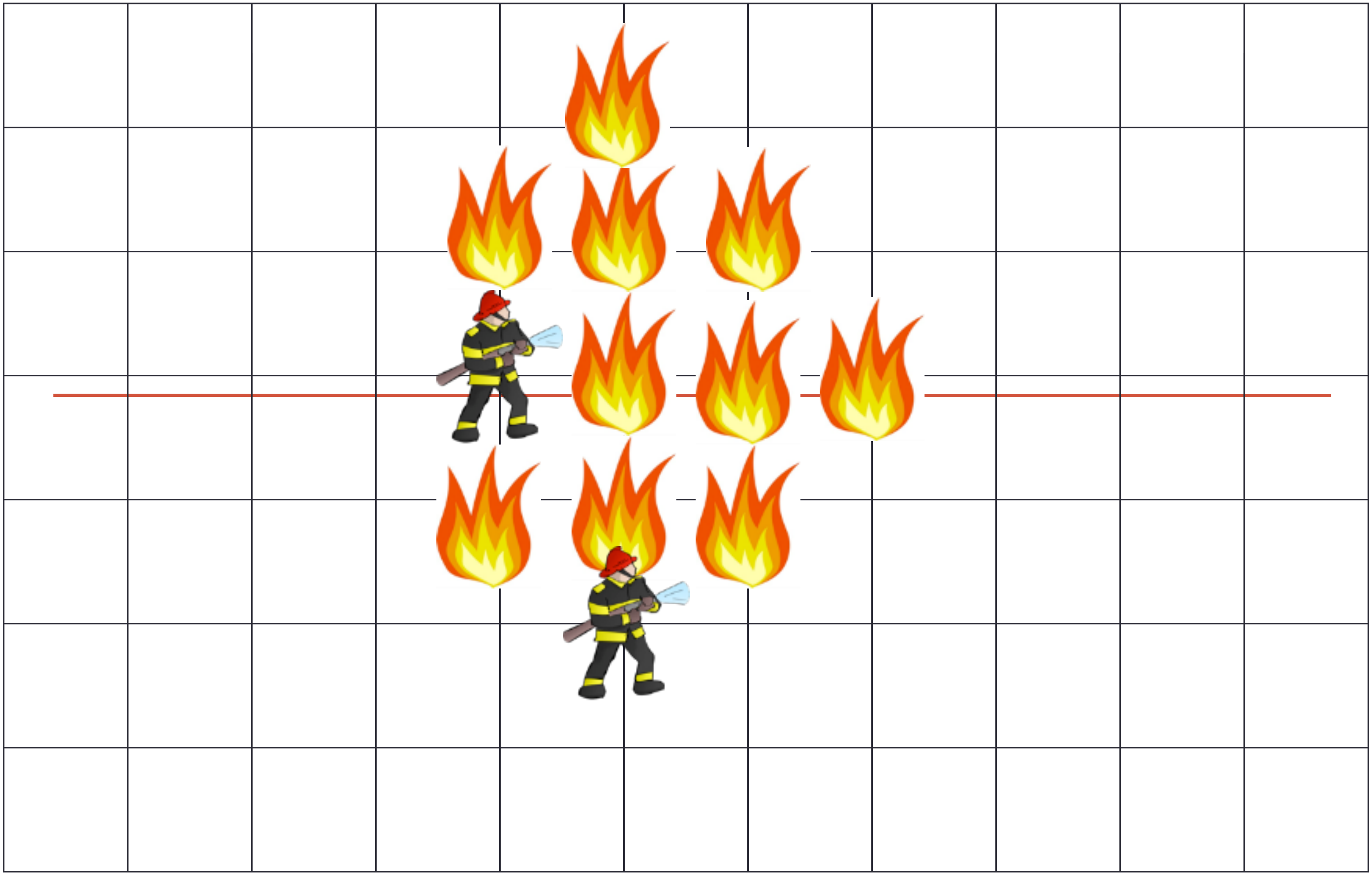
1. Fire starts at one (or more) point(s)
2. 1 Firefighter can be deployed to protect a point at each time step
3. Fire spreads to all unprotected adjacent vertices in the next time step
4. Repeat

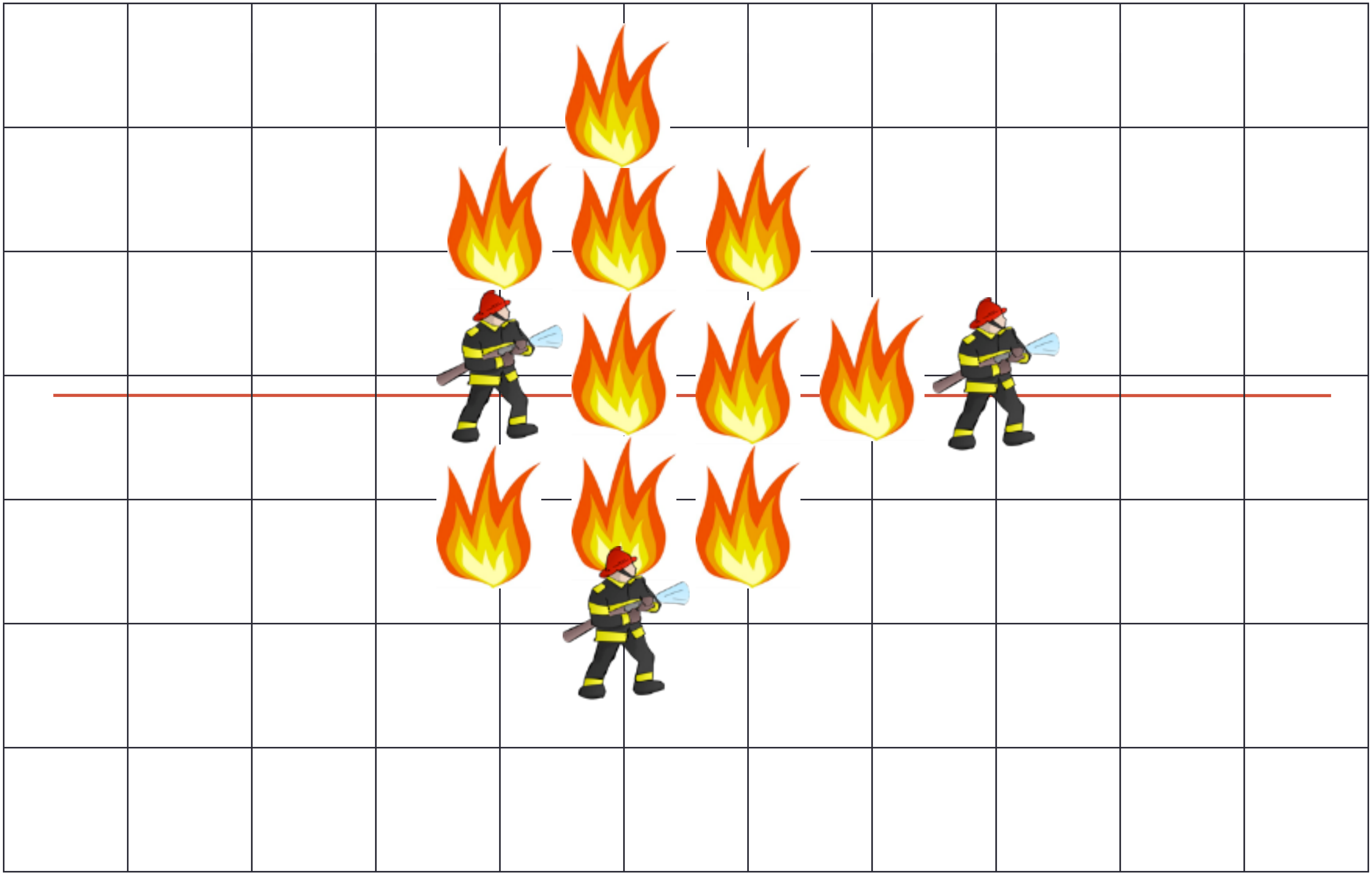




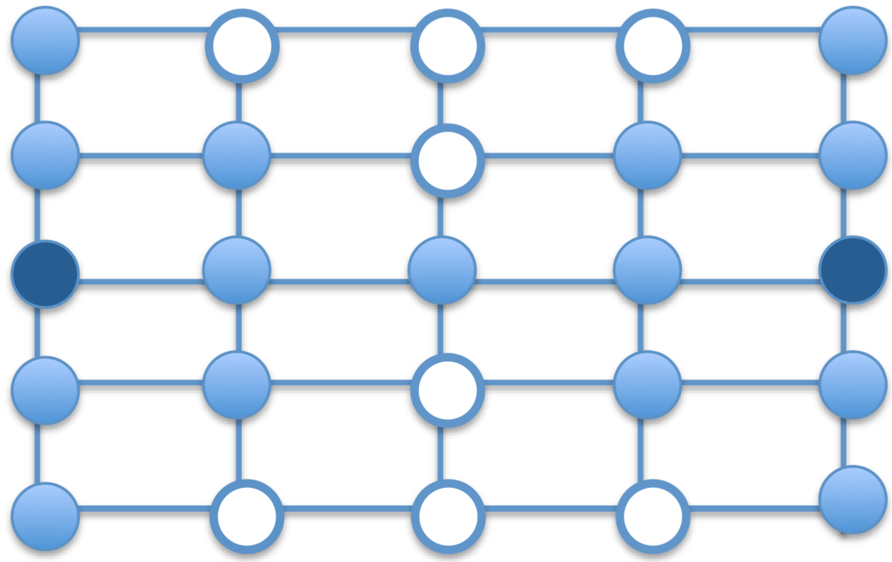






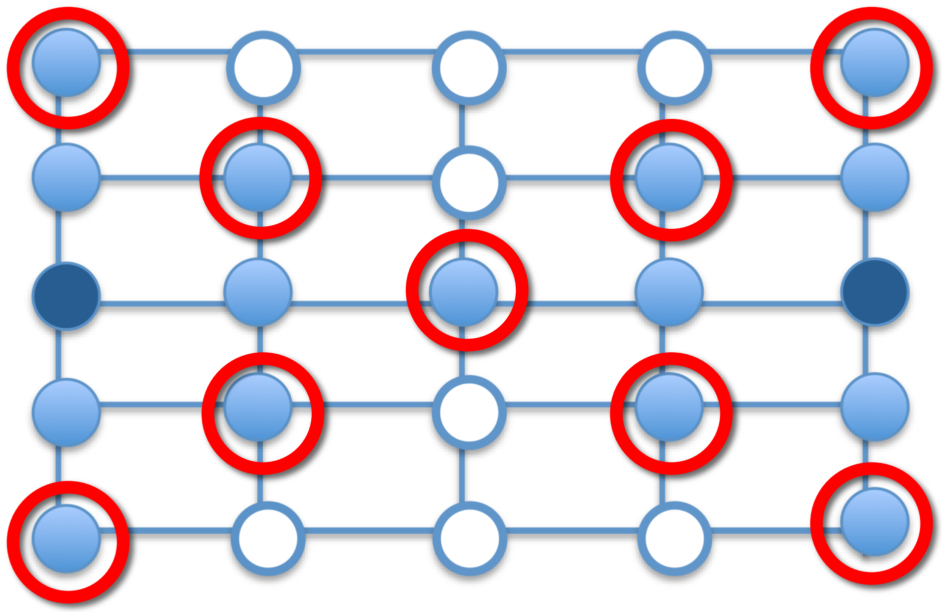


Understanding Question / Q2



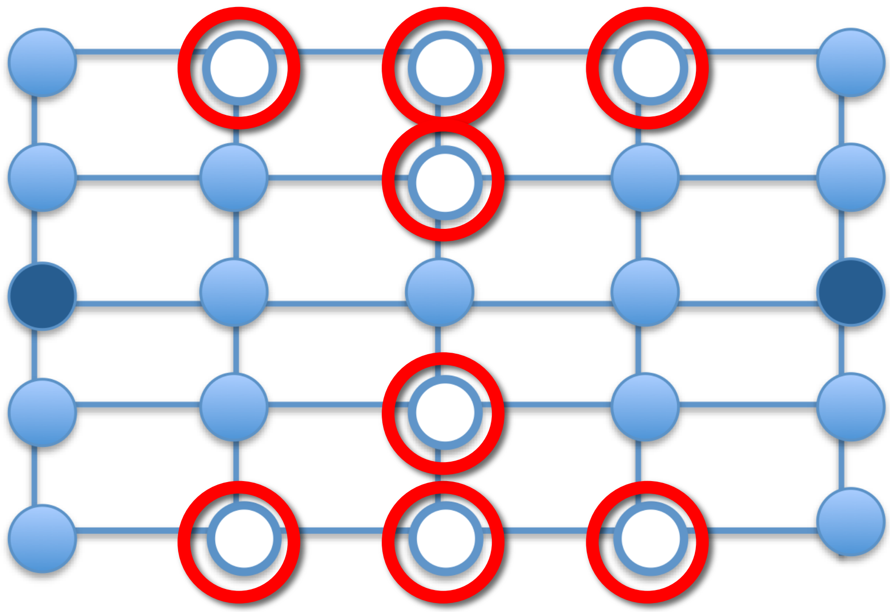
- **Question**:
- Assume you are playing a variant of the firefighter game. Dark nodes are fires. What is the least number of firefighters that you would need to deploy at once in order for the fire to not reach the white nodes?
- **Possible Answers**:
  - 3, 6, 8, 9, or 10?

Understanding Question



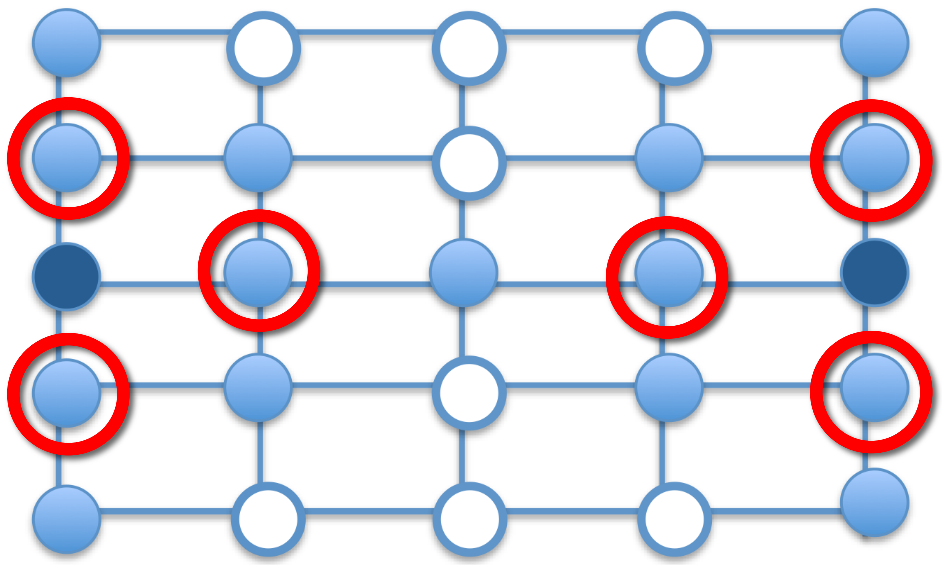
- **One Answer**: 9

Understanding Question



- **Another Answer**: 8

Understanding Question



- **Correct Answer**: 6

STOCHASTIC MODELS

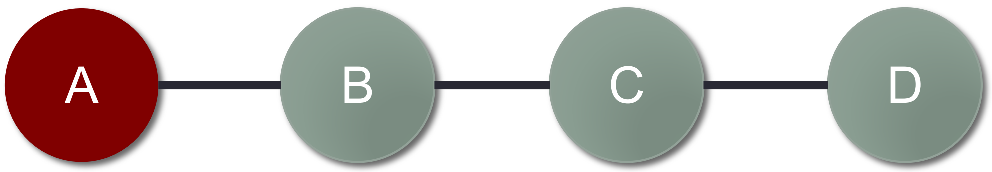
Stochastic Models

- Stochastic models introduce probabilities into the models i.e., once in contact with an infected person you will not always become infected

- Assume p is probability the disease spreads from person to person
  - p=0 → 0% chance of transmission
  - p=1.0 → 100% chance of transmission

Stochastic Model – Example

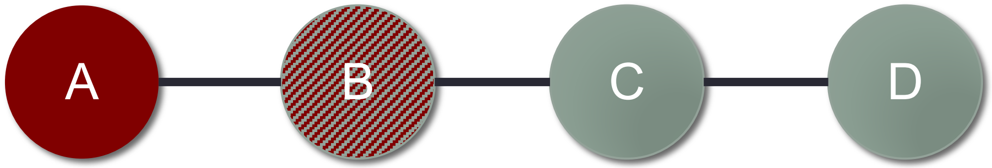
- Assume simple network:



- Assume node A is infected, B/C/D are susceptible, 1-threshold model, probability of infection p is **80%**

- Time t=1, node B infected with probability 80%
- Time t=2, node C may become infected. What is the chance?

Stochastic Model – Example (cont)



- Case 1: Node B is infected and passes disease to C
  - Probability = P(B infected) * p = 0.8 * 0.8 = **0.64 = P(C infected)**
- Case 2: Node B is infected and does not pass disease to C
  - Probability = P(B infected) * (1-p) = 0.8 * 0.2 = 0.16
- Case 3: Node B is not infected, thus cannot pass to C
  - Probability = P(B not infected) = 1 – 0.8 = 0.2
- Extending idea for D:
  - P(D infected) = P(C infected) * p = 0.64 * 0.8 = 0.512

INFLUENCE MAXIMIZATION PROBLEM

Influence Maximization Problem

Given:

- a social network G
- a propagation model M
  - Independent cascade model
  - Linear threshold model
- find k nodes in G that maximize influence over entire network

- For a given propagation model M and a node set S, we can define the expected size of propagation

[Source: http://edbt-school-2013.imag.fr/wp-content/uploads/2013/09/EDBTsummerschool.pdf]

Independent Cascade Model

- Every arc (u, v) has associated probability **p(u, v)** of node **u** influencing node **v**
- Time proceeds in discrete steps
- At time t, nodes that were active at time t-1 activate their inactive neighbors, with probability p(u, v)

Linear Threshold Model

- Every arc (u,v) has an associated **weight b(u,v)**, such that the sum of all incoming weights in each node is <=1
- Time proceeds in discreet steps
- Each node v picks a random threshold **0<= theta(v) <= 1**
- Node v becomes active when the sum of incoming weights from active neighbors reaches **theta(v)**