SI 301: Models of Social Information Processing

Daniel M Romero
School of Information
University of Michigan

Teaching Team

Daniel Romero	drom@umich.edu	Assistant Professor
Danaja Maldeniya	dmal@umich.edu	Ph.D Student
John Posch	posch@umich.edu	Ph.D Student
Muhan Yuan	yuanmh@umich.edu	Master’s Student

Office Hours

Day	Block 1	Block 2
Monday	11am-2pm Maldeniya	2-4pm Yuan
Tuesday	9-10am Yuan	10:30am-12:30pm Posch
Wednesday	9-10am Maldeniya	11am-12pm Romero 3340 North Quad
Thursday	9-10am Yuan	
Friday	9-11am Posch	3-5pm Romero 3340 North Quad

Most office hours will be held in 1277 North Quad.
Additional office hours by appointment
15 hours in total availably Mon-Fri. **Use them!**

Discussion Sections

- Reinforcement of material covered in class and/or readings
- Small group and individual labs
- Opportunity for Q&A and discussion
- Programming exercises and examples
- Attendance is mandatory
- Participation is part of your grade

Assignments

- Most important part of your learning.
- Weekly (with some exceptions).
- Usually due on Tuesday in class.
- Written and programming.
- Solutions will be posted the day the assignment is due.
- No extensions allowed.

Collaborating on Assignments

Allowed:

- Discussing how to solve problems
- Explaining concepts to each other

Not Allowed:

- Copying someone else's solution
- Turning in solutions you do not understand

Collaborating on Assignments

Allowed:

- Discussing how to solve problems
- Explaining concepts to each other

Not Allowed:

- Copying someone else's solution
- Turning in solutions you do not understand

Always write solutions independently

Always turn in original work

Always indicate the names on the students you collaborate with

All cases of plagiarism will be officially reported

Extra Credit on Homework

- Starting with assignment 2, you will have an opportunity to earn extra credit on homework assignments (10 extra credit points out of 100).
- **Team work:**
  - You will be split into two groups: the “early” and “late” team working groups.
  - If you’re in the “**early**” group, you have to work with an assigned partner on the homework assignment in order to get extra credit points on assignments **2-7**.
  - If you’re in the “**late**” group, you have to work with an assigned partner on the homework assignment in order to get extra credit points on assignments **8-13**.
- Pairs will be assigned.
- Each assignment is capped at 100 points.

Life happens!® Insurance Policy

Life happens!®

We are happy to offer *Life happens!®* Insurance Policy:

- Two assignments with the lowest grade will be dropped.
- We allow two missed lectures without penalty.
- We allow two missed discussions without penalty.

The Cost: late assignments will not be accepted and no additional make-up opportunities will be granted. Please do not ask for extension or exceptions.

Exams

1. Midterm: In class on Oct 26th.
2. Cumulative Final: Dec 20th 4-6pm. Location: TBA

Any material covered in:

- Lectures
- Discussions
- Assignments
- Readings

Grades

Grade breakdown:

- Final: 25%
- Weekly Assignments: 30%
- Midterm: 25%
- Discussion Participation: 10%
- Lecture attendance: 5%
- Lecture i-Clicker questions: 5%

A	[90, 100]
B	[80, 90)
C	[70, 80)
D	[60, 70)
E	[0, 60)

See syllabus for +/- grade assignments

i-Clicker

Why i-clicker

- Make class more interactive.
- Allow me to
  - check how well the entire class understands material.
  - track attendance and participation.
- Will be used in lectures not in discussions.
- 5% of grade is based on attendance.
- 5% of grade is based on correctness of i-Clicker answers.
- Don’t forget to bring clickers!

i-Clicker Correctness Points

Get points throughout the semester

- 1 point for a **correct answer** if $\geq$ 50% students get it right.
- 2 points for a **correct answer** if $<$ 50% students get it right.
- 1 point for a **incorrect answer** if $<$ 50% students get it right.

Your correctness grade (5%) is:

$$\frac{P}{Q} + 0.10$$

P is the number of points you have at the end of the semester
Q is the total number of i-clicker questions

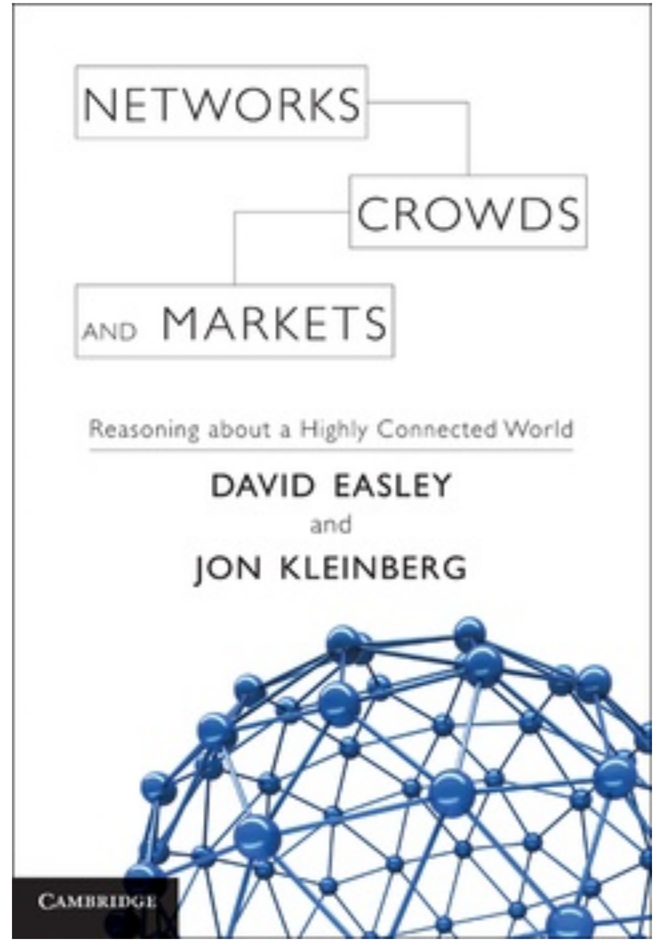
Correctness grade is capped at 100%

Text Book

"Networks, Crowds and Markets: Reasoning about a Highly Connected World" by David Easley and Jon Kleinberg.

Access: You can freely download a pre-publication version of the book at the authors' site:

http://www.cs.cornell.edu/home/kleinber/networks-book/



Python

Review Python during the first few weeks of the semester.

Material covered in SI 106 is enough.

We will mainly use the NetworkX Python library.

Computers in Classroom

Respond using i-Clicker:

A. I am planning on using a laptop to take notes during lecture.

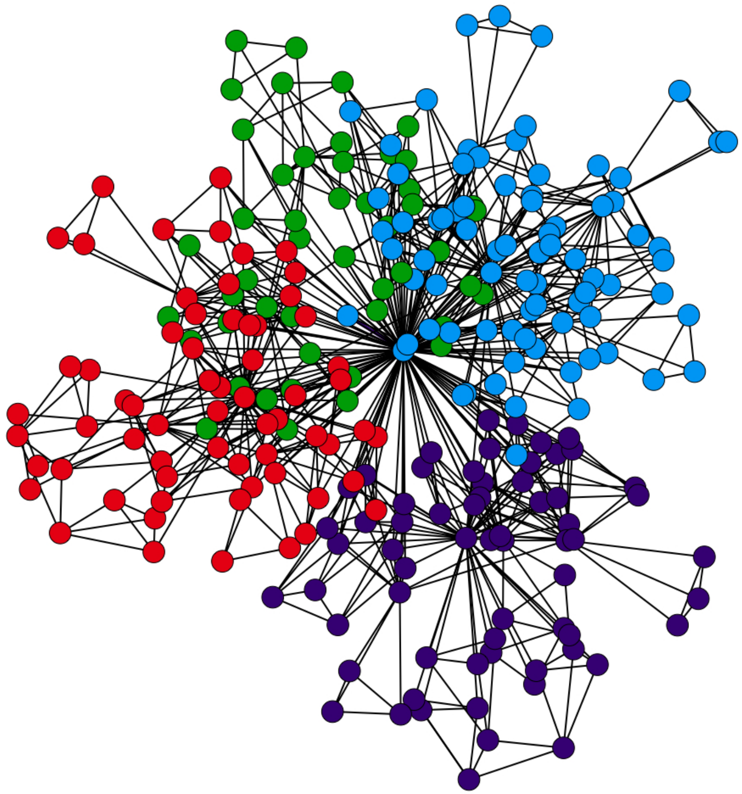
B. I am not planning on using a laptop during lecture but I do not mind if other students do.

C. I am not planning on using a laptop during lecture and I am distracted when other students do.

Networks

Networks: A set of things (nodes) with interconnections (edges).

Why study networks?

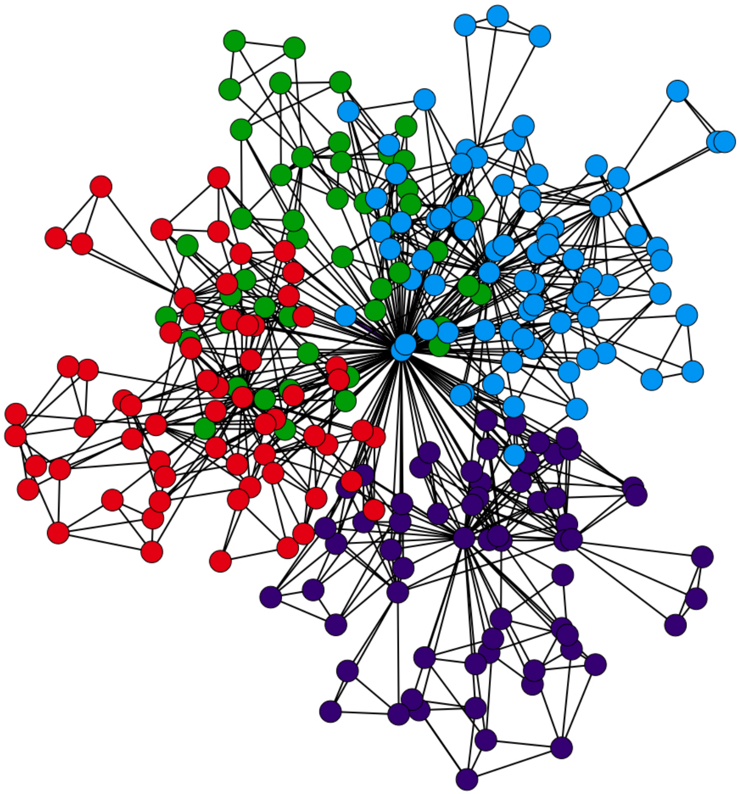


Networks

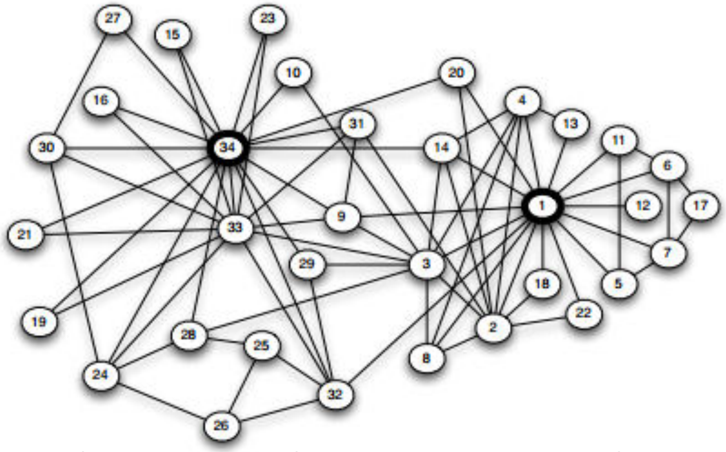
Networks: A set of things (nodes) with interconnections (edges).

Why study networks?

Because they are everywhere

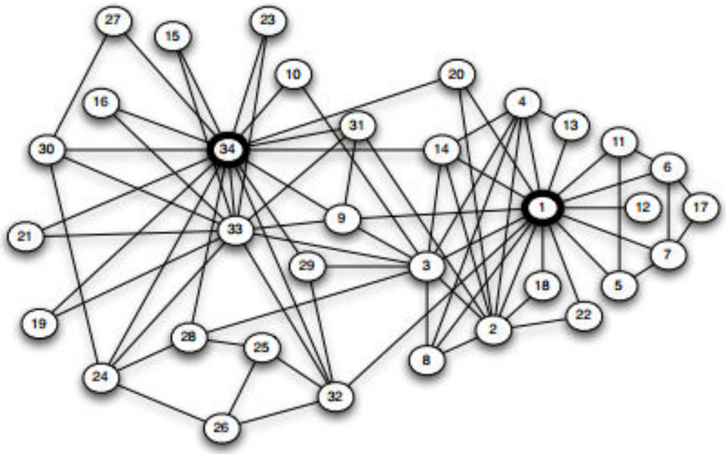


Social Networks

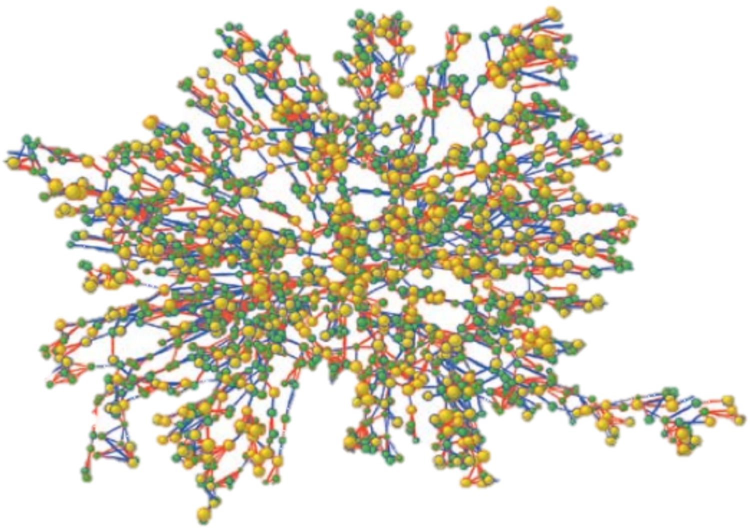


Friendship network in a 34-person karate club

Social Networks

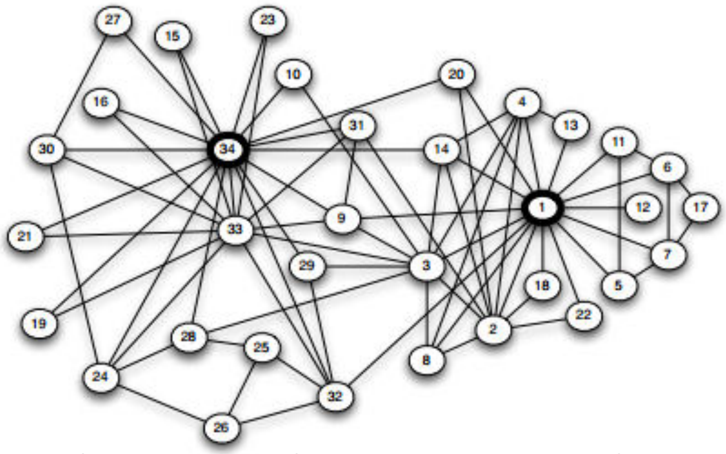


Friendship network in a 34-person karate club

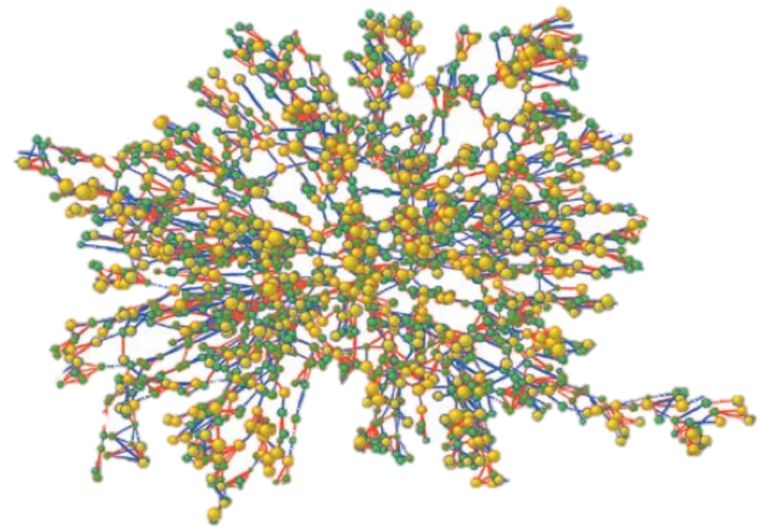


Network of friendship, marital tie, and family tie among 2200 people

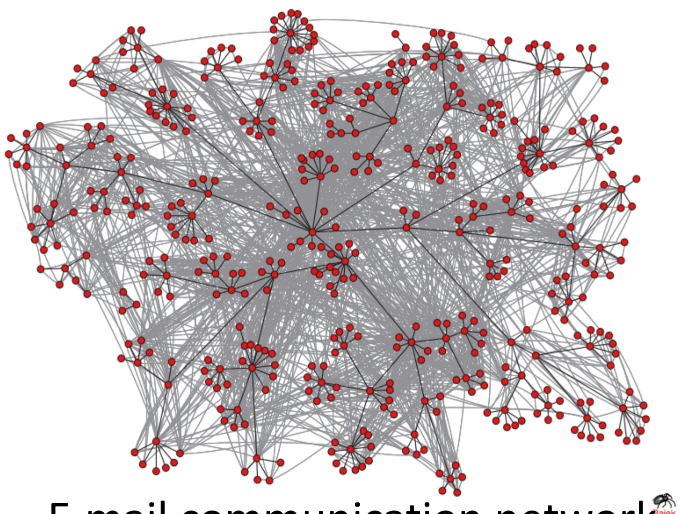
Social Networks



Friendship network in a 34-person karate club

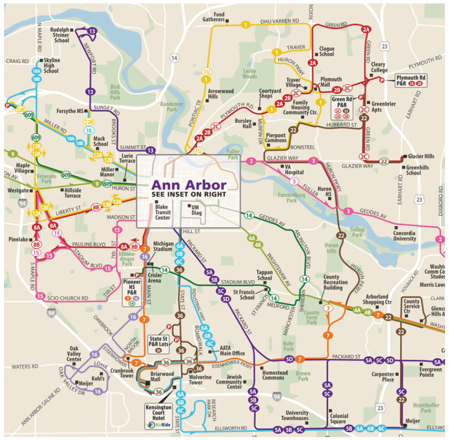


Network of friendship, marital tie, and family tie among 2200 people



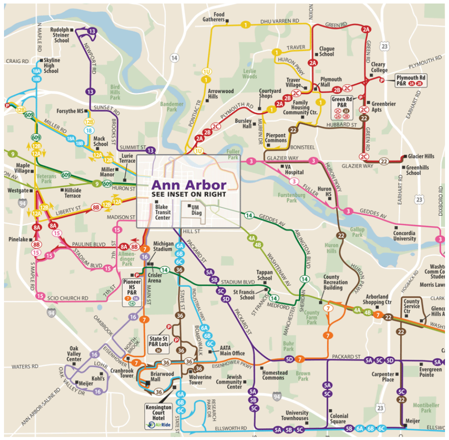
E-mail communication network among 436 HP employees

Transportation and Mobility Networks

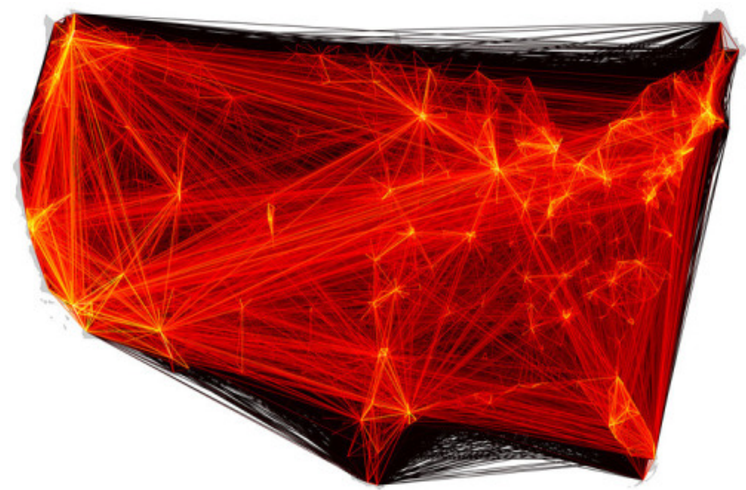


Ann Arbor bus transportation network

Transportation and Mobility Networks

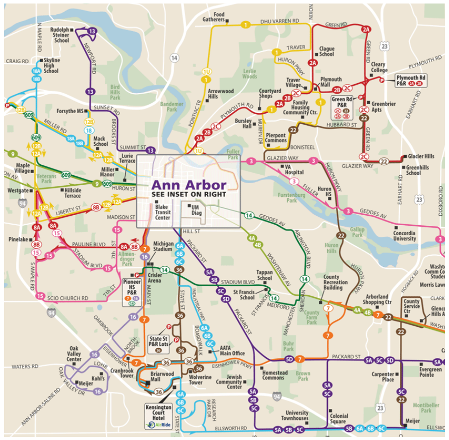


Ann Arbor bus transportation network

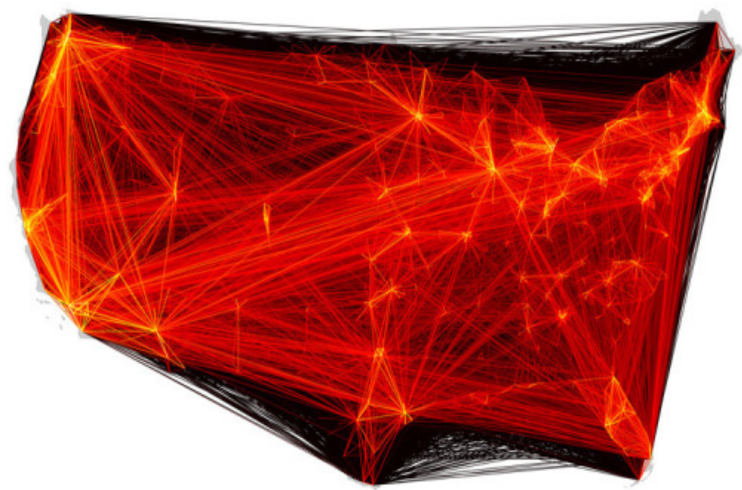


Human mobility network based
on location of dollar bills (Where's George)

Transportation and Mobility Networks



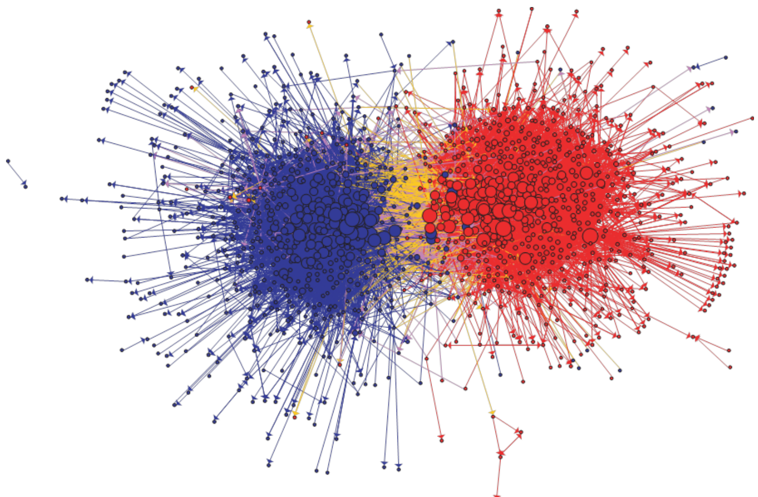
Ann Arbor bus transportation network



Human mobility network based on location of dollar bills (Where's George)

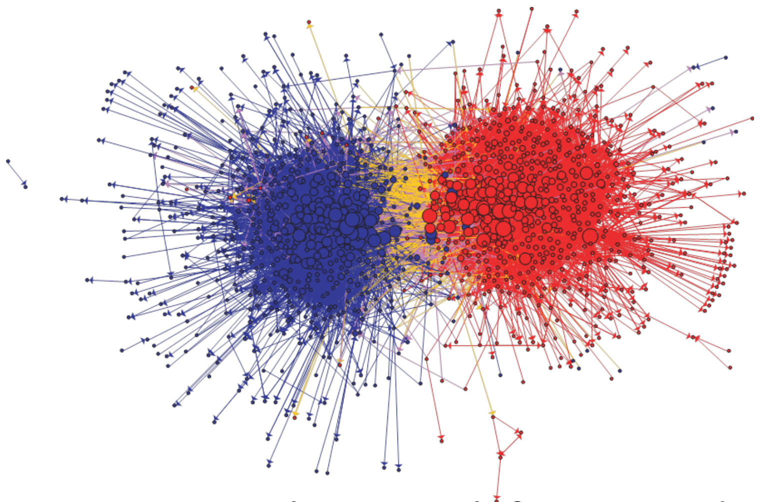
Network of direct flights around the world

Information Networks

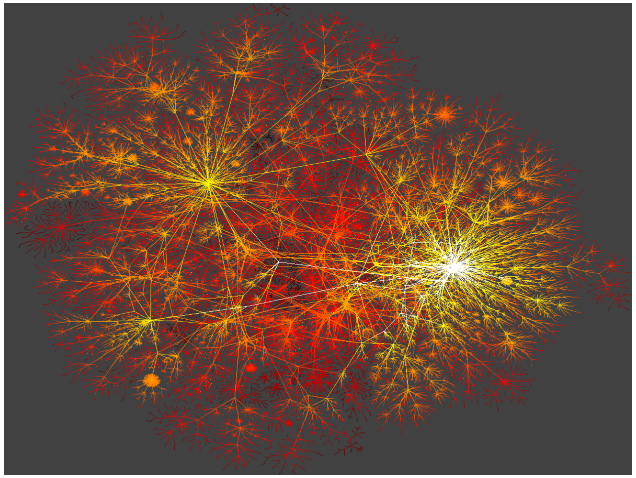


Communication between left-wing and right-wing political blogs (Adamic and Glance)

Information Networks

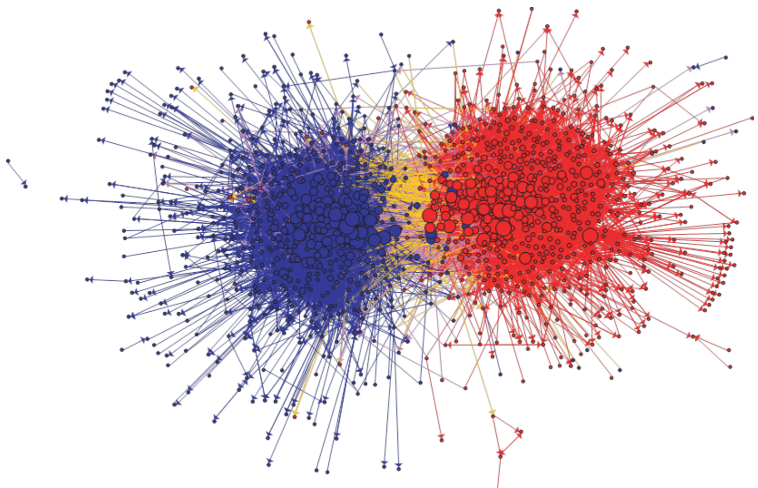


Communication between left-wing and right-wing political blogs (Adamic and Glance)

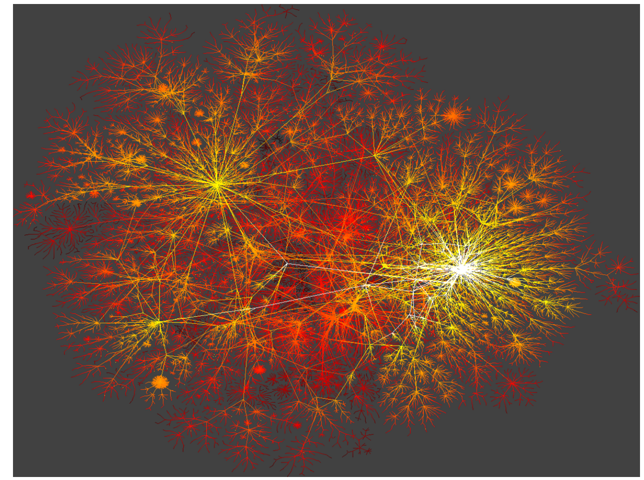


Internet Connectivity

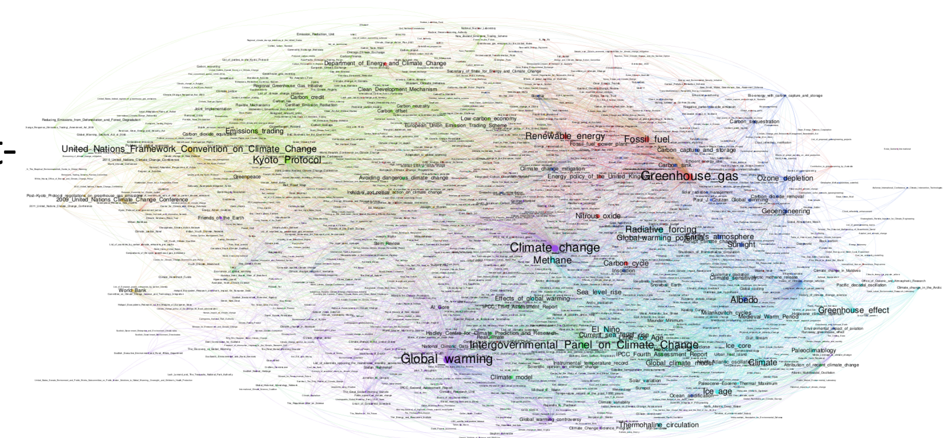
Information Networks



Communication between left-wing and right-wing political blogs (Adamic and Glance)

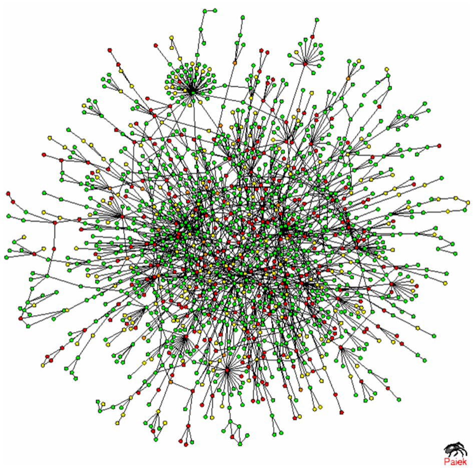


Internet Connectivity



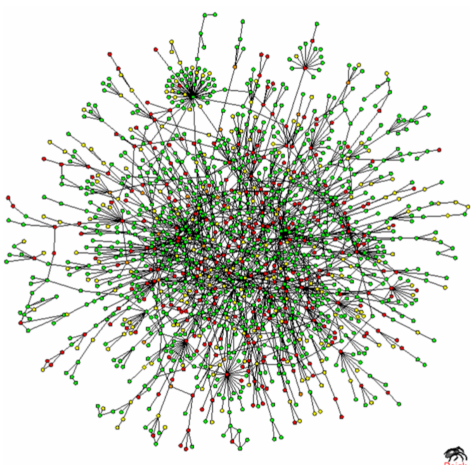
Network of Wikipedia articles about climate change (EMAPS)

Biological Networks

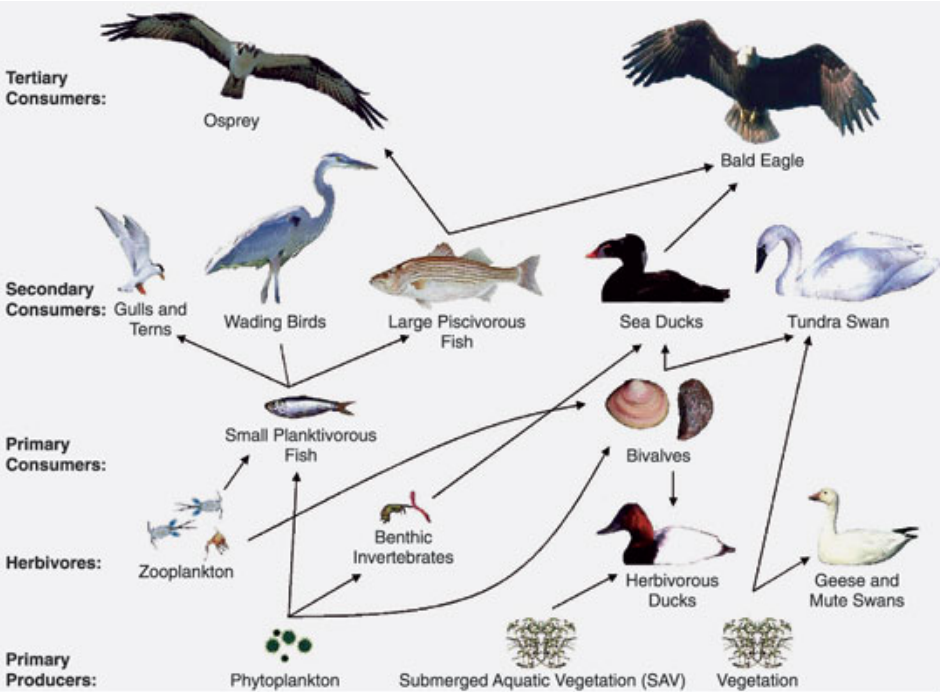


Protein-protein interactions (Jeong)

Biological Networks



Protein-protein interactions (Jeong)



Chesapeake Bay Waterbird Food Web

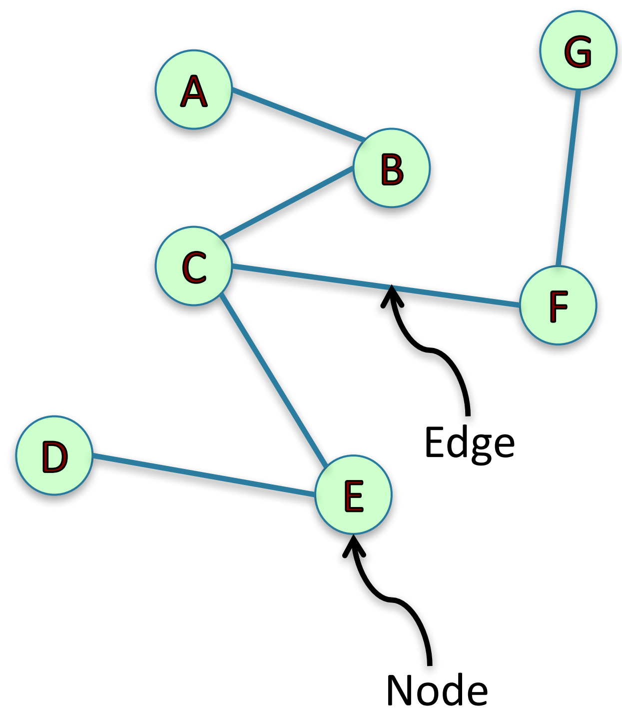
And More…

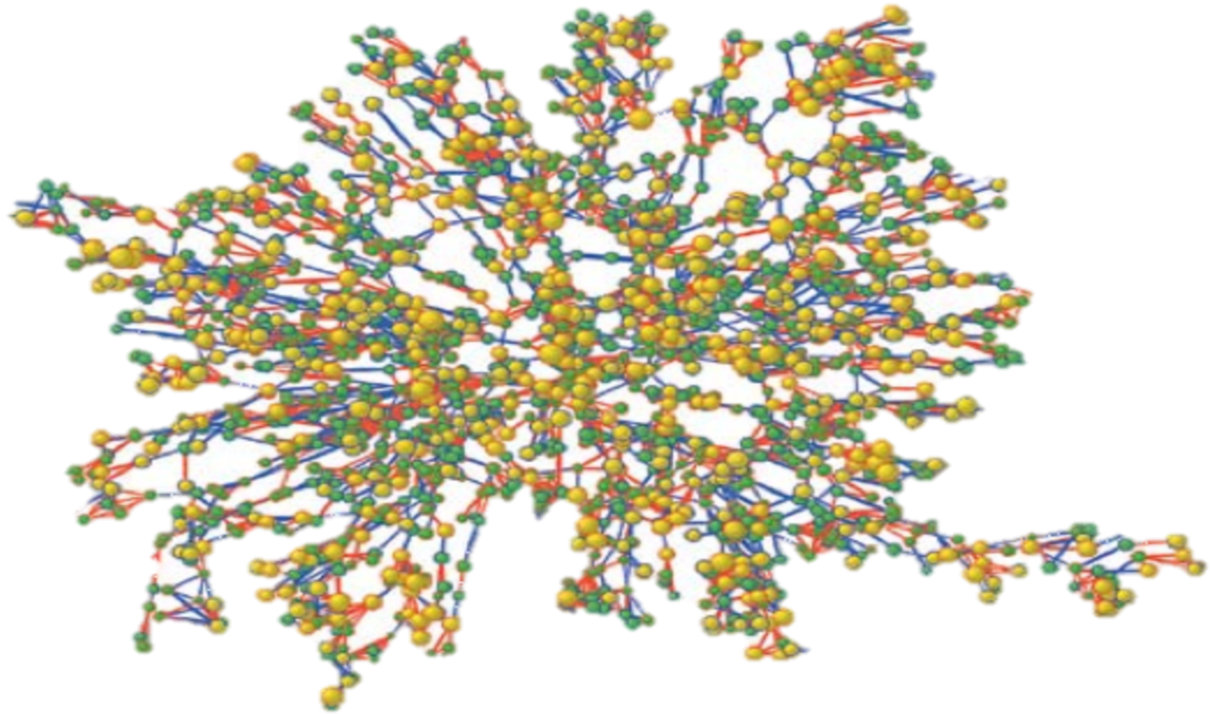
- Financial networks
- Co-authorship networks
- Trade networks
- Citation networks
- See Assignment 1

Network Definition and Vocabulary

Network (or Graph): A representation of **connections** among a set of **items**.

- Items are called **nodes** (or vertices)
- Connections are called **edges** (or link or ties)

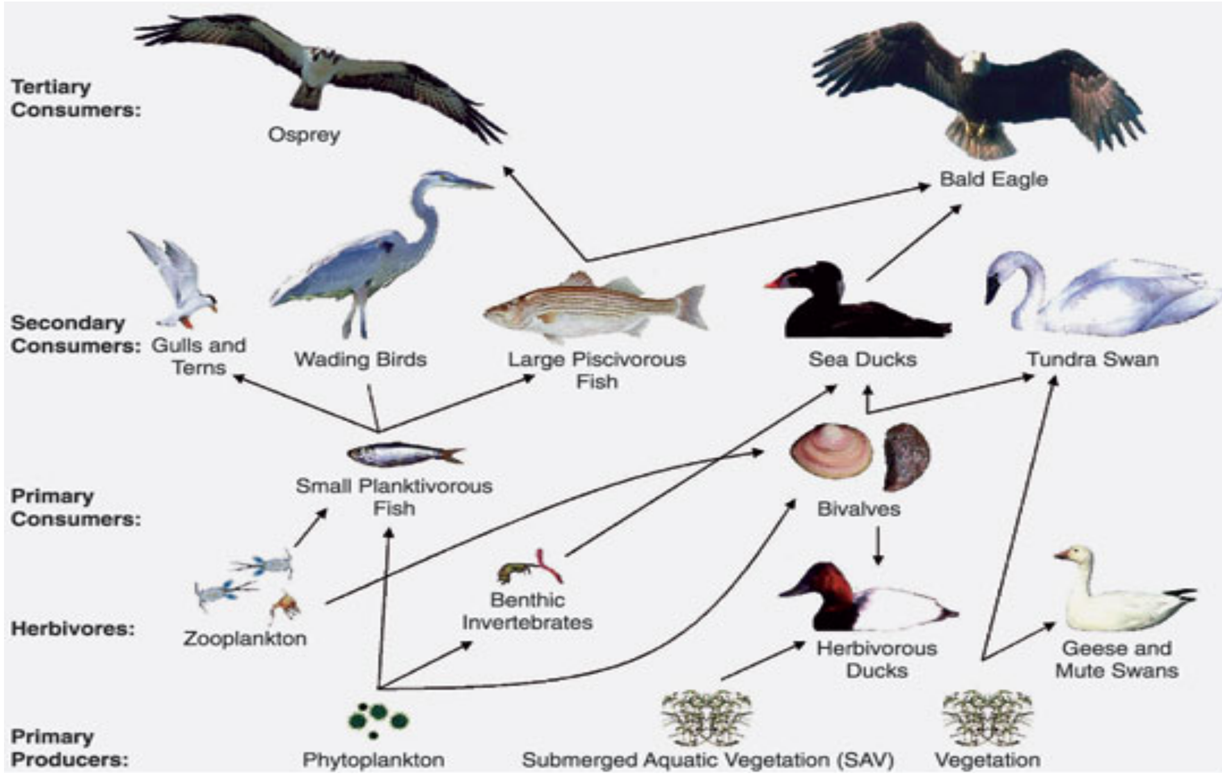




Network of friendship, marital tie, and family tie among 2200 people

Nodes: People
Edges: Friendship, marital or family tie

Symmetric relationships

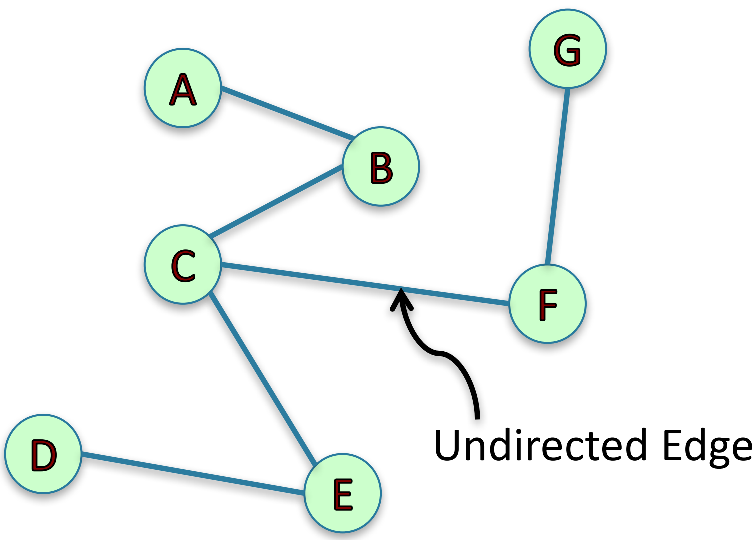


Chesapeake Bay Water bird Food Web

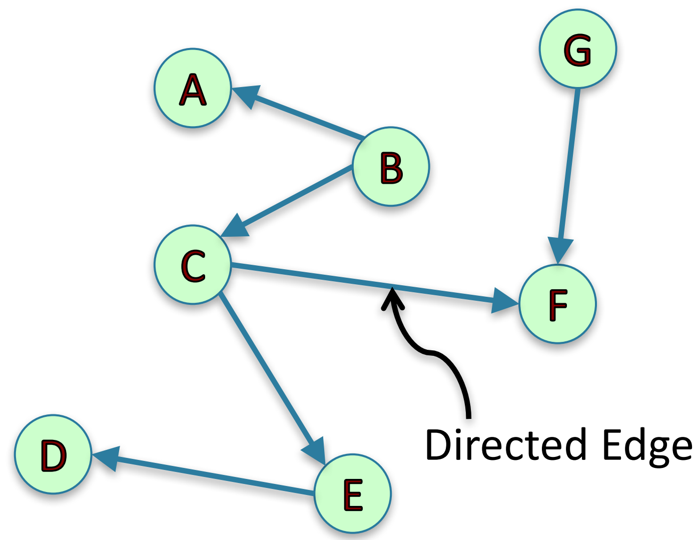
Nodes: Birds

Edges: What eats what

Asymmetric relationships



Undirected network



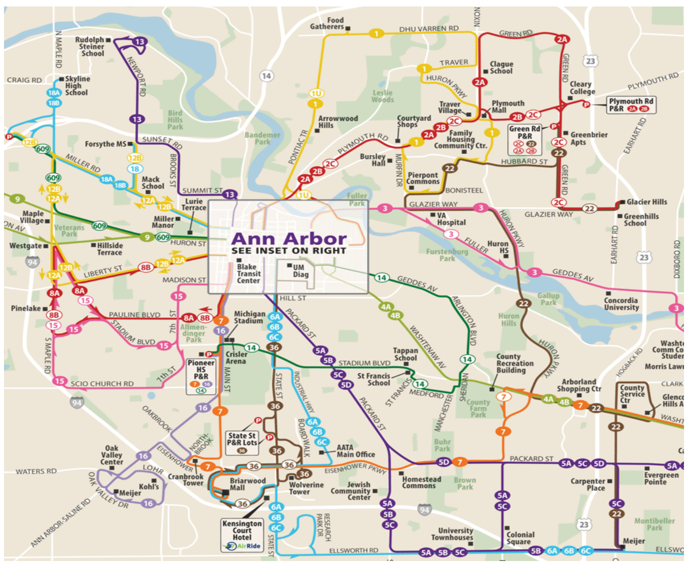
Directed network

Connectivity

Moving from node to node through edges

Connectivity

Moving from node to node through edges



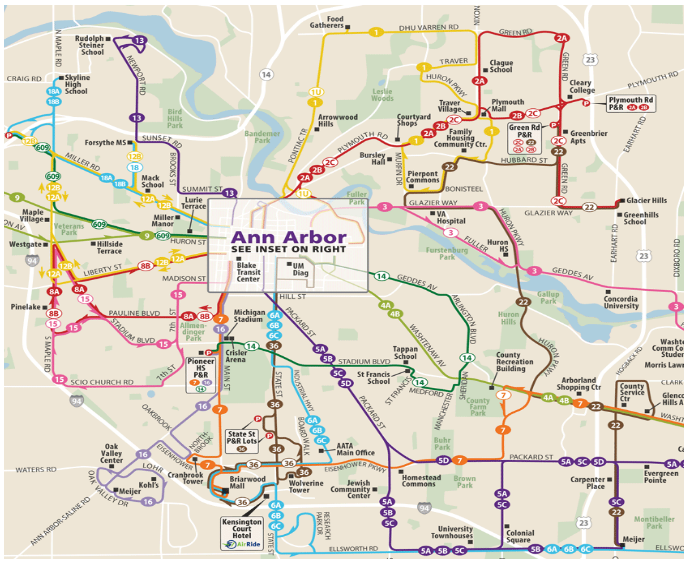
Ann Arbor bus transportation network

Nodes: Bus stops

Edges: Direct bus connections

Connectivity

Moving from node to node through edges



Ann Arbor bus transportation network

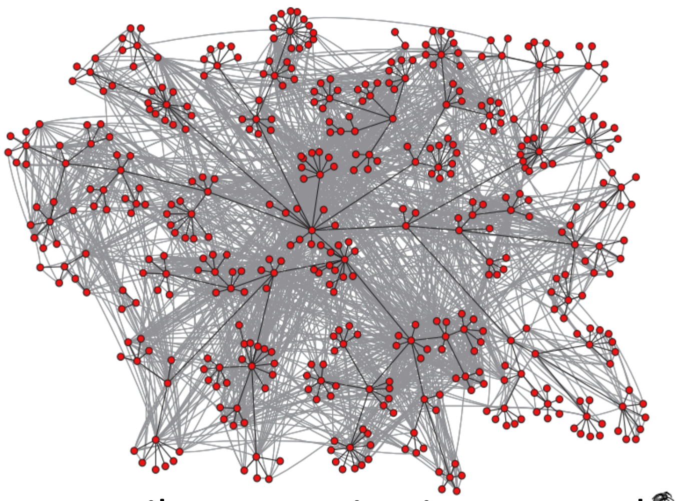
Nodes: Bus stops
Edges: Direct bus connections

How can I get from stop A to stop B?

Are there areas of AA that are disconnected from the rest?

Connectivity

Moving from node to node through edges

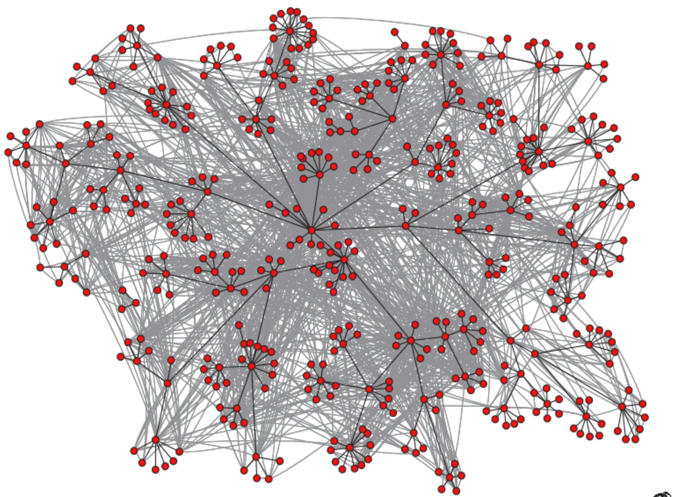


E-mail communication network among 436 HP employees

Nodes: HP employees
Edges: Email communication

Connectivity

Moving from node to node through edges



E-mail communication network among 436 HP employees

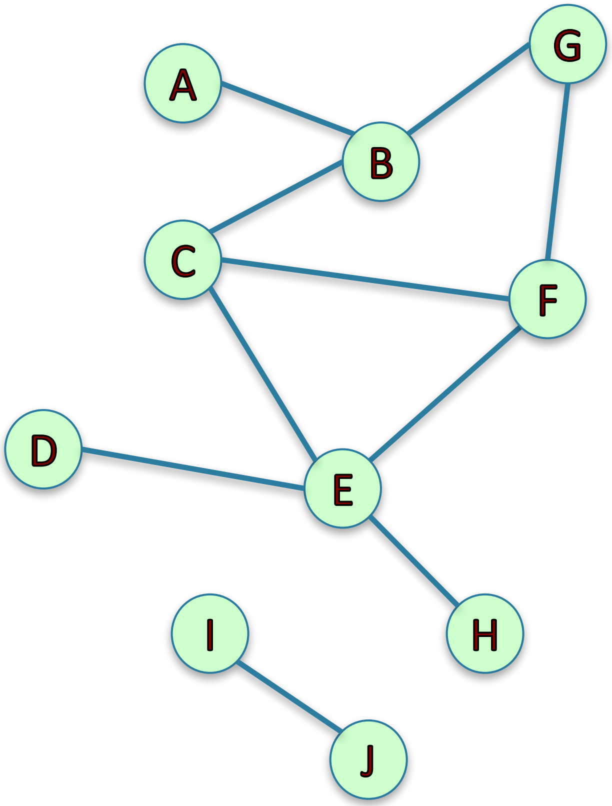
Nodes: HP employees
Edges: Email communication

How can a rumor originated from person A arrive at person B?

Is it possible for everyone to hear the rumor through email?

Paths

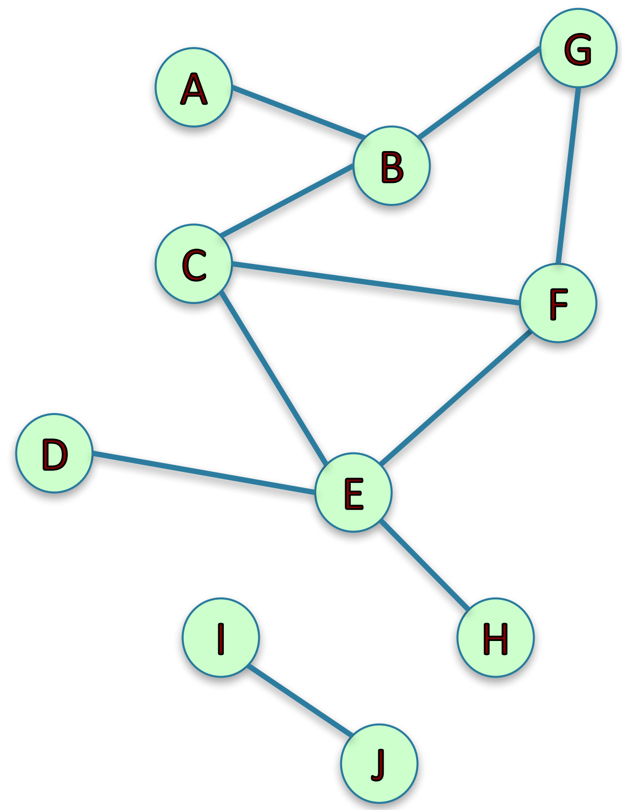
Path: A sequence of nodes connected by an edge.



Paths

Path: A sequence of nodes connected by an edge.

Find two paths from node G to node C:

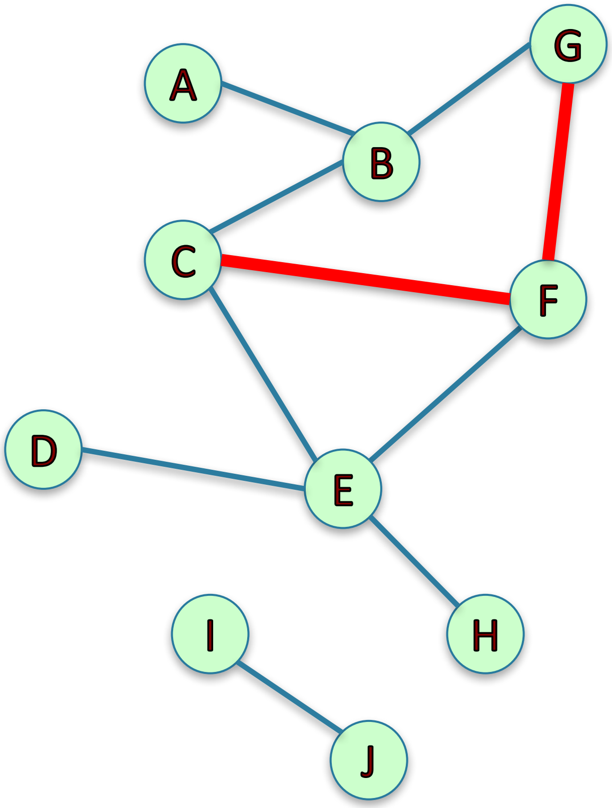


Paths

Path: A sequence of nodes connected by an edge.

Find two paths from node G to node C:

G – F – C

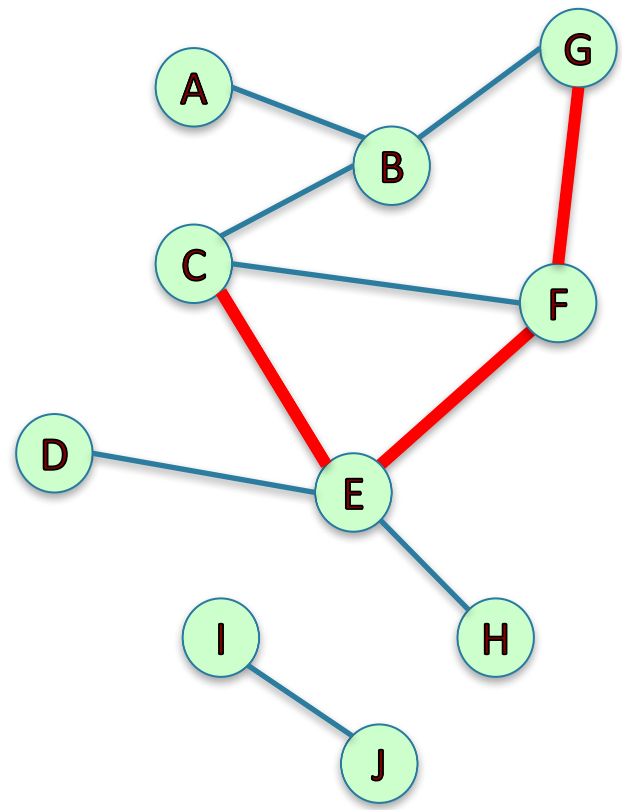


Paths

Path: A sequence of nodes connected by an edge.

Find two paths from node G to node C:

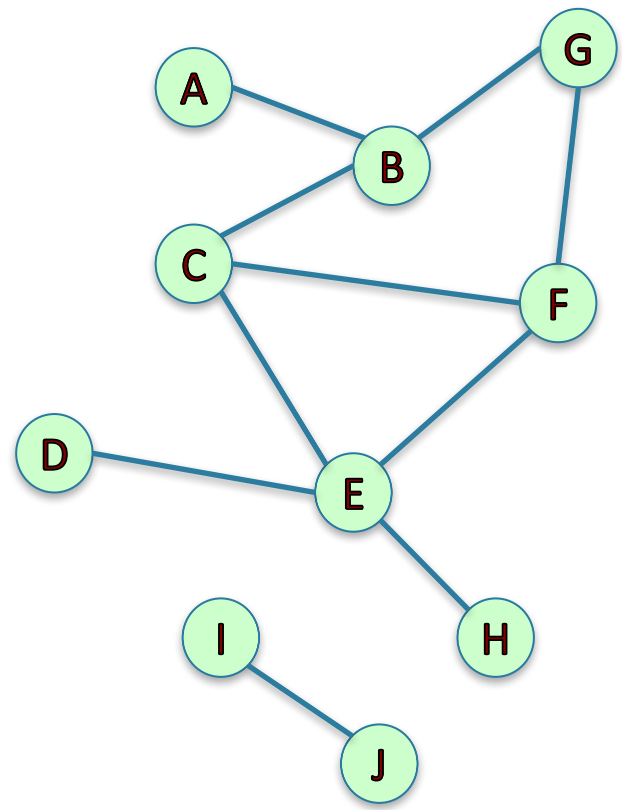
G – F – E – C



Cycles

Cycle: A path of at least 3 edges, in which the first and last nodes are the same, but all other nodes are distinct.

Find a cycle in the network:

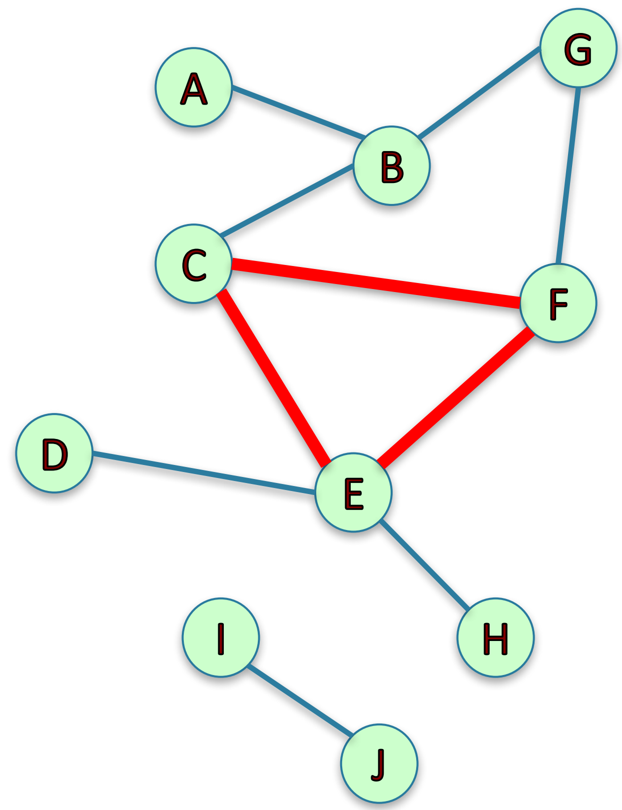


Cycles

Cycle: A path of at least 3 edges, in which the first and last nodes are the same, but all other nodes are distinct.

Find a cycle in the network:

C – F – E – C

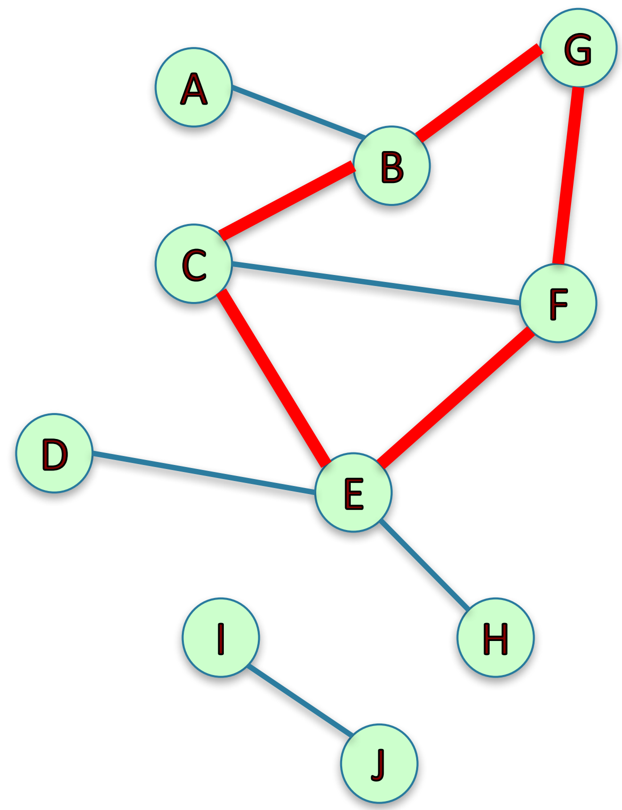


Cycles

Cycle: A path of at least 3 edges, in which the first and last nodes are the same, but all other nodes are distinct.

Find a cycle in the network:

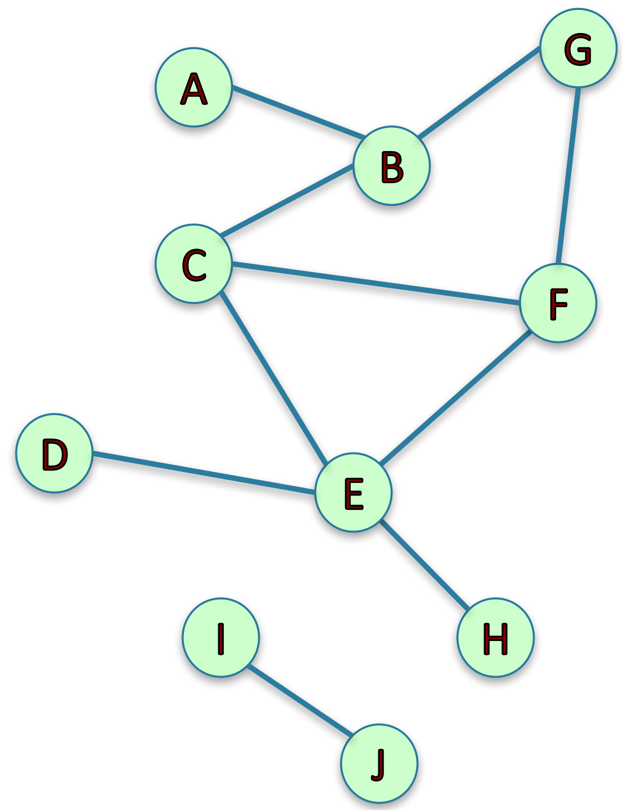
G – F – E – C – B – G



Connectivity

Connected graph: A graph is connected if, for every pair nodes, there is a path between them

Is this graph connected?

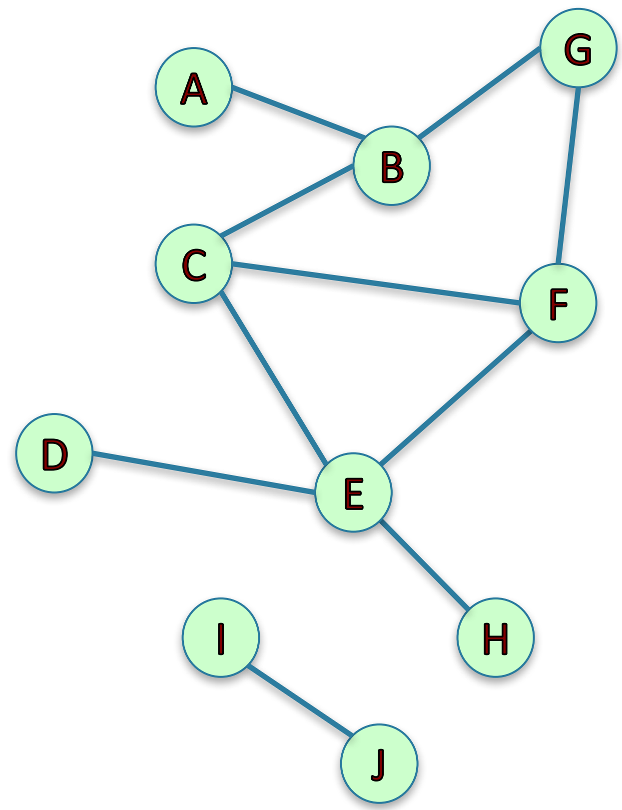


Connectivity

Connected graph: A graph is connected if, for every pair nodes, there is a path between them

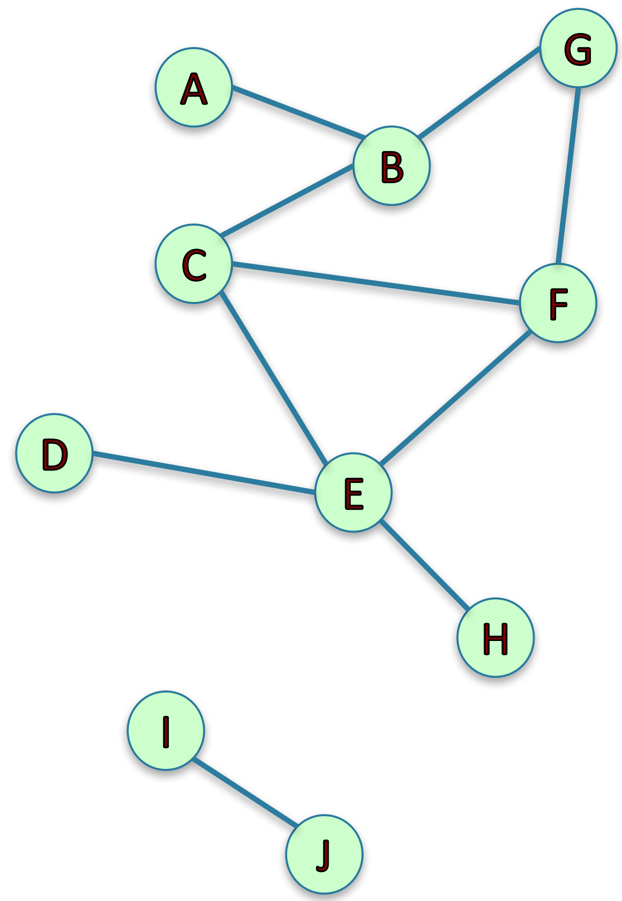
Is this graph connected?

No, some paths are missing. For example, there is no path from J to D.



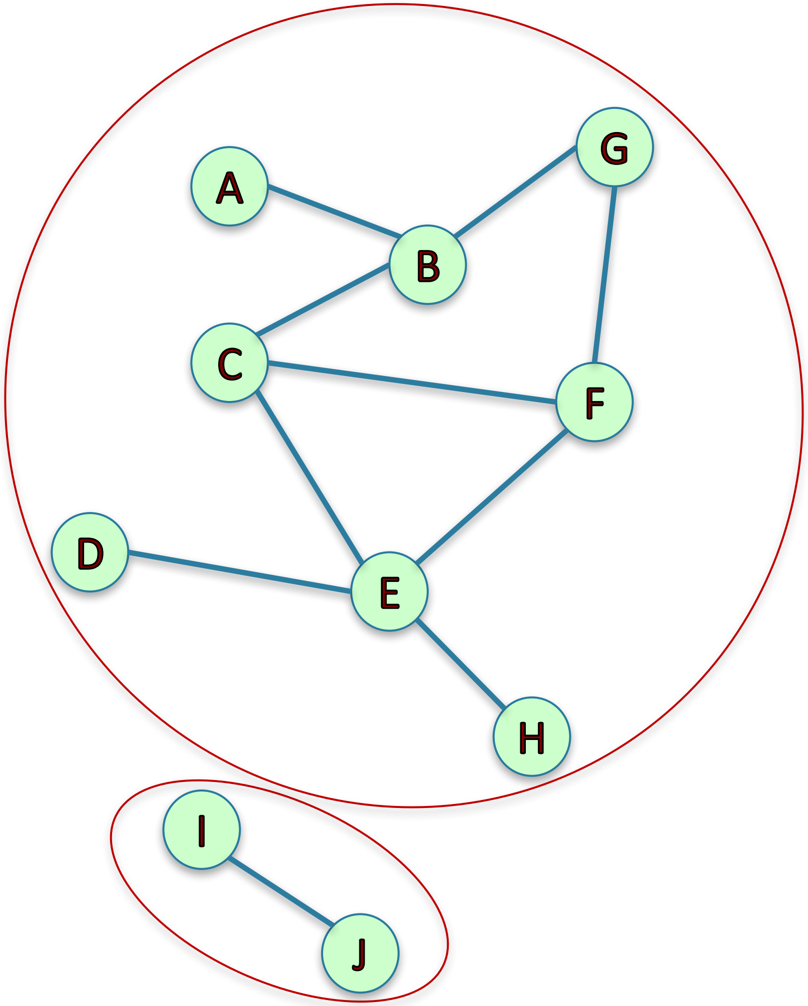
Connected Components

Connected component:



Connected Components

Connected component:

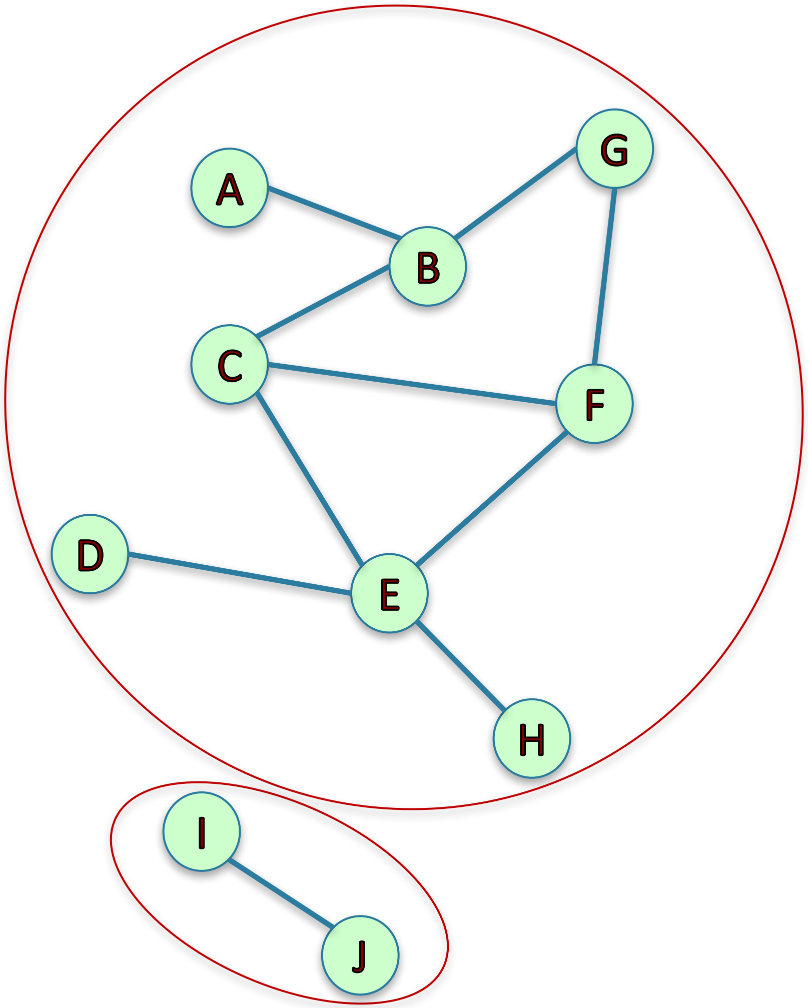


Connected Components

Connected component:

A subset of nodes such as:

i. *Every node in the subset has a path to every other node.*



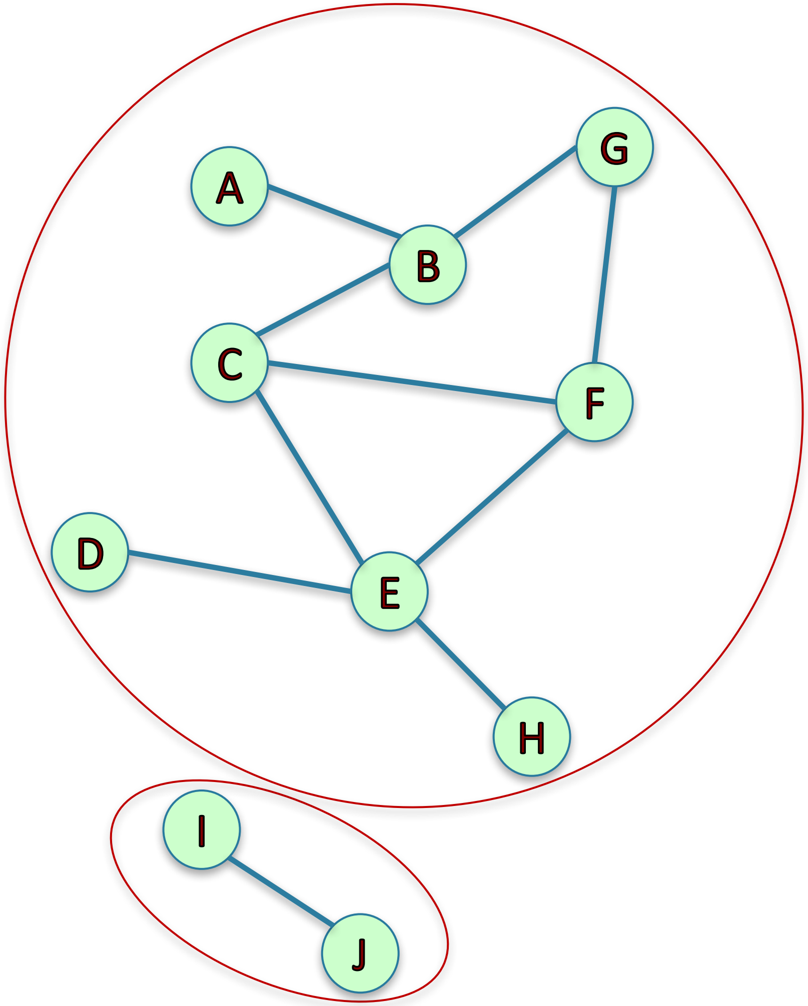
Connected Components

Connected component:

A subset of nodes such as:

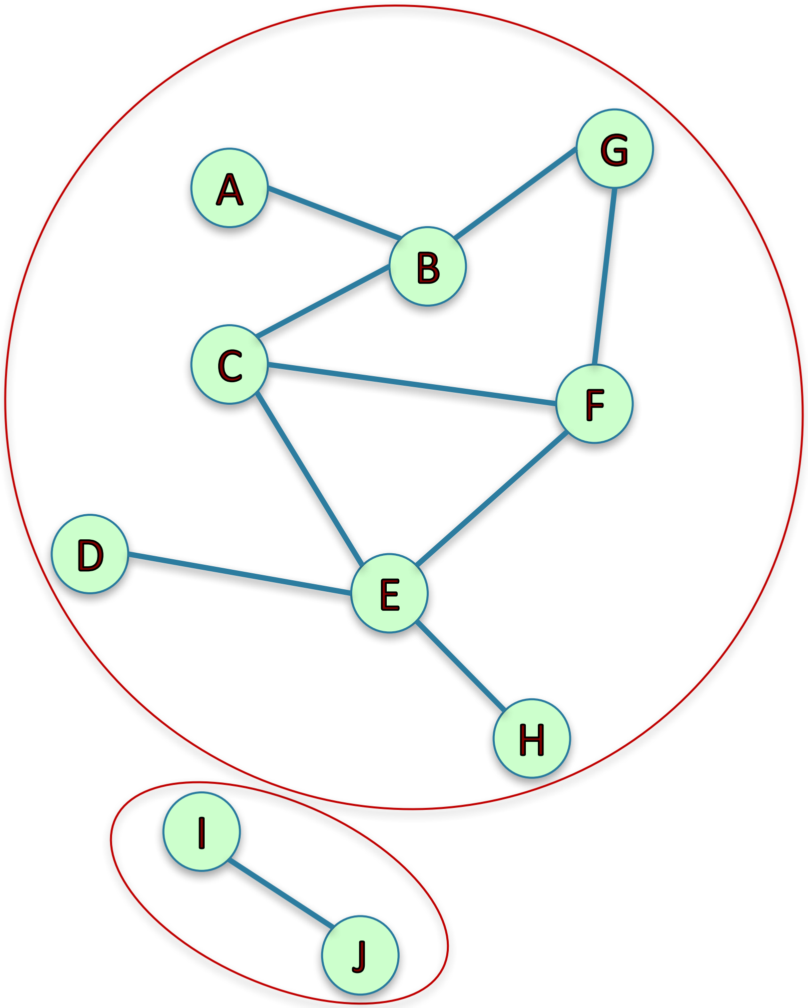
i. *Every node in the subset has a path to every other node.*

ii. *No other node is connected to every node in the subset.*



Connected Components

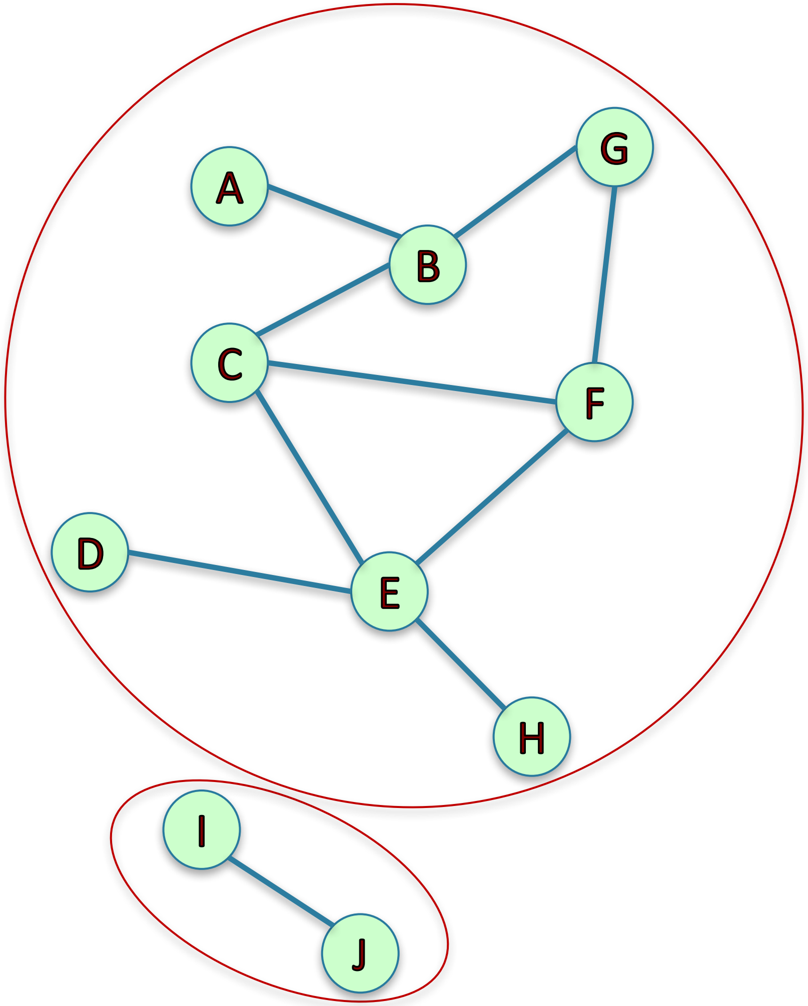
Is the subset {A, B, C} a connected component?



Connected Components

Is the subset {A, B, C} a connected component?

No, there are other nodes that have paths to A,B, and C. For example, node F.

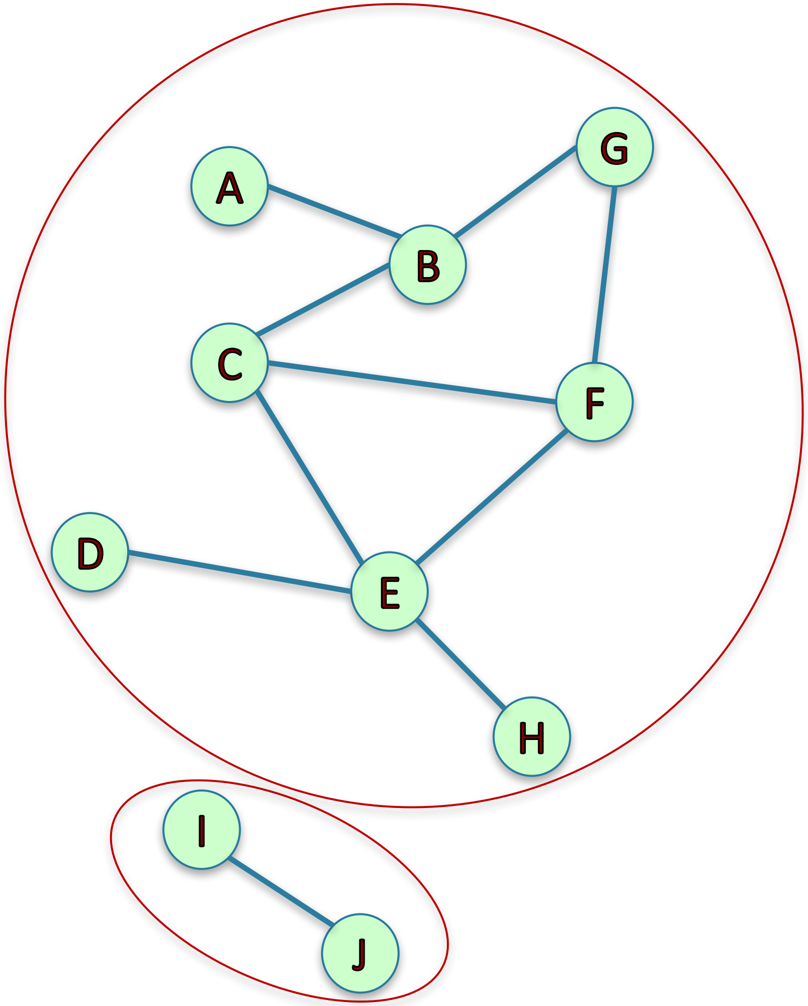


Connected Components

Is the subset {A, B, C} a connected component?

No, there are other nodes that have paths to A,B, and C. For example, node F.

Is the subset {I, J, H} a connected component?



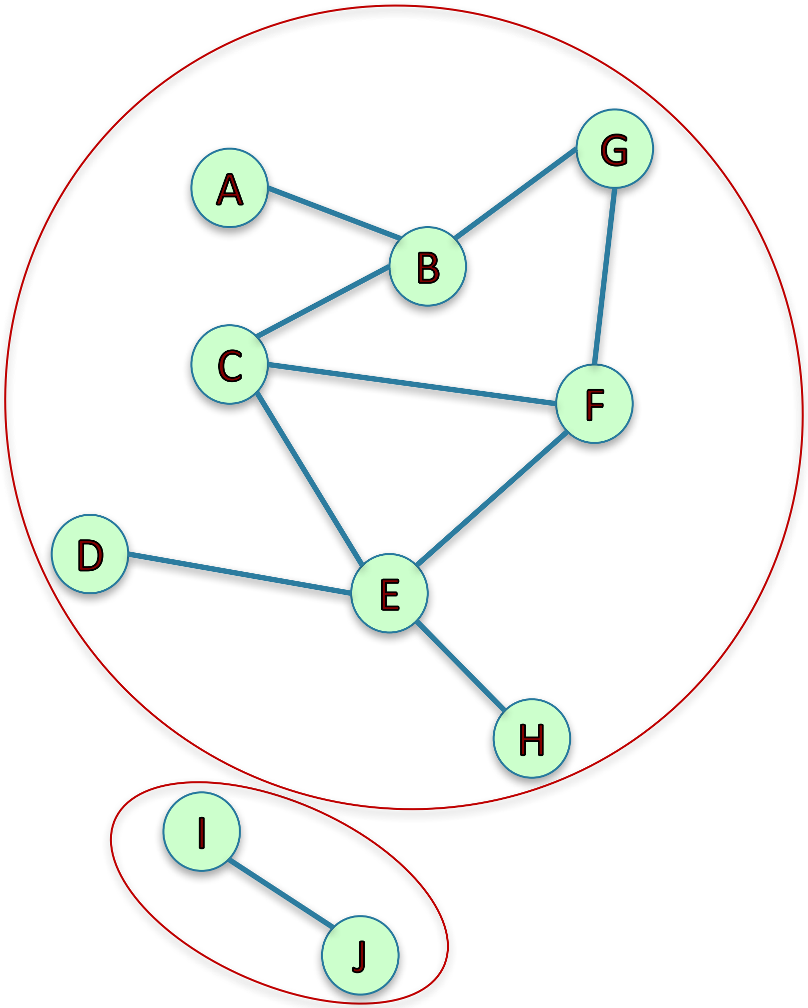
Connected Components

Is the subset {A, B, C} a connected component?

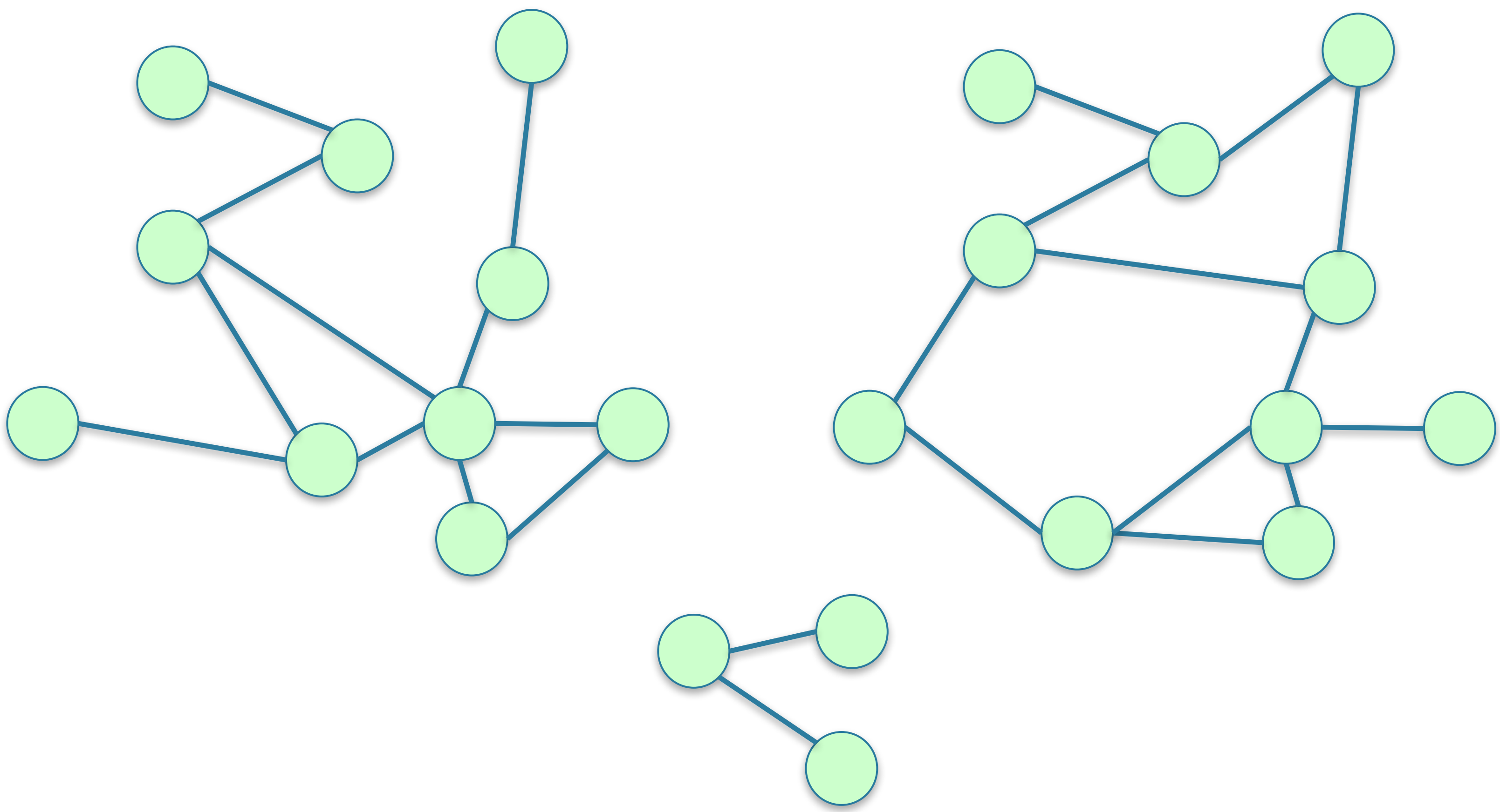
No, there are other nodes that have paths to A,B, and C. For example, node F.

Is the subset {I, J, H} a connected component?

No, there is no path from node J to node H.

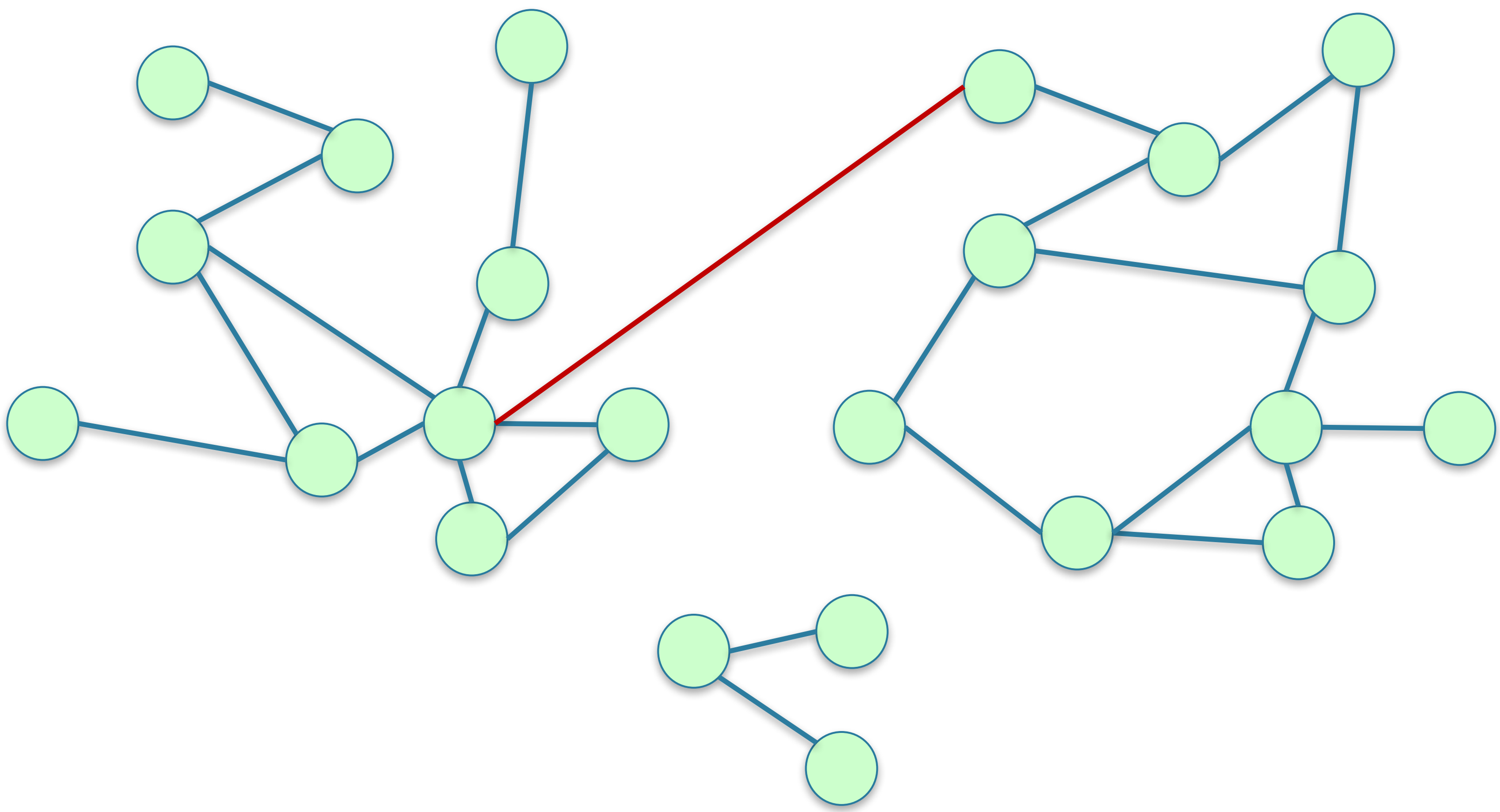


Friendship Network in a Classroom



What's unusual with this network?

Friendship Network in a Classroom

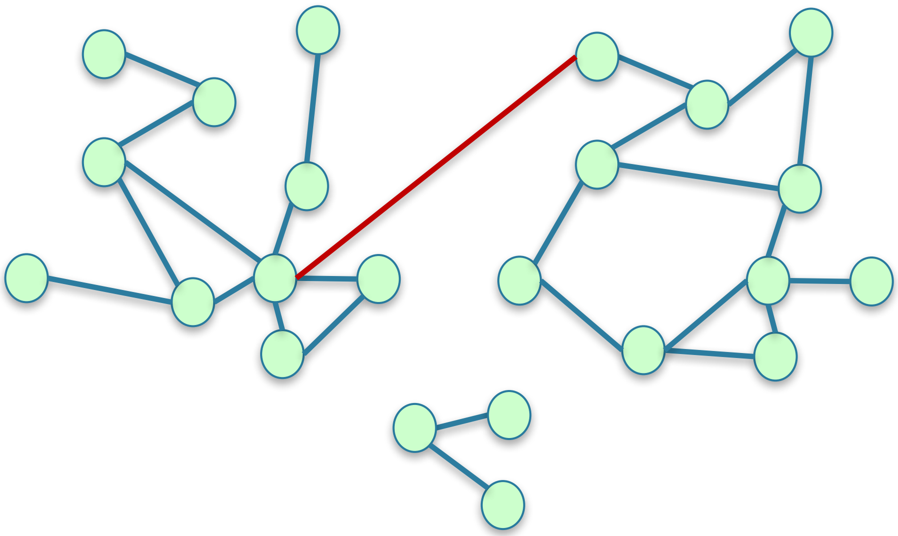


Connections among two large connected components are likely to happen

Giant Components

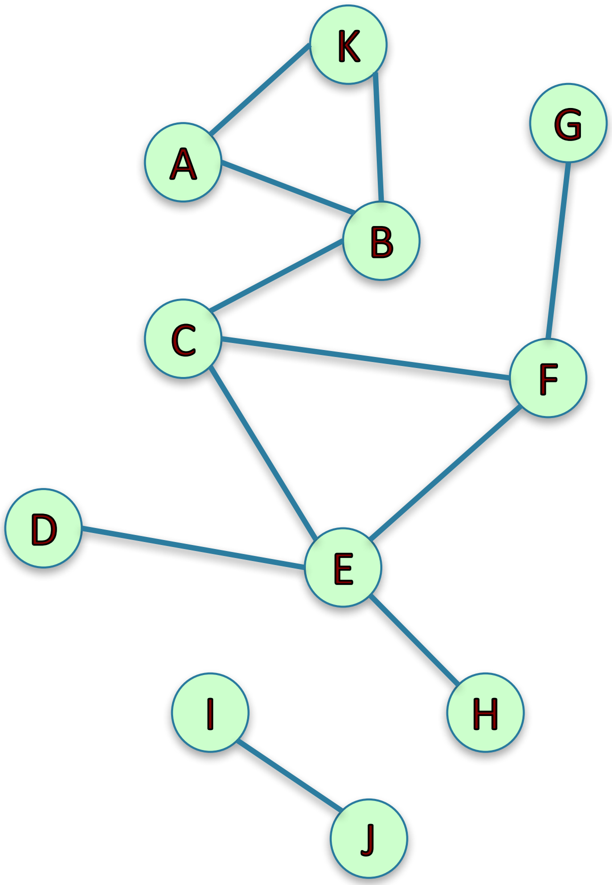
Giant component: A component that contains a large portion of the nodes of a graph. **Informal term**.

Usually networks with a giant component usually contain only one.



Distance

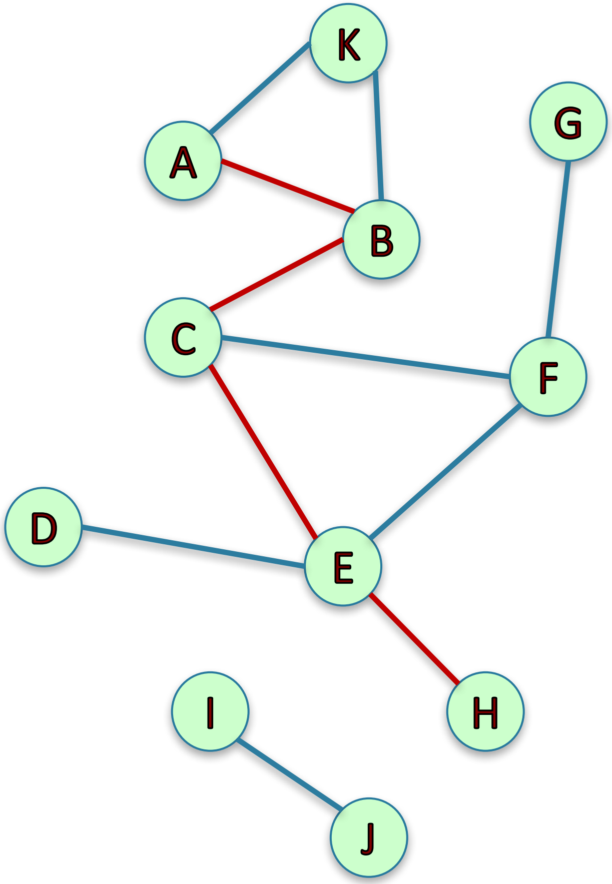
How far is node A from node H?



Distance

How far is node A from node H?

Path 1: A – B – C – E – H (4 “hops”)

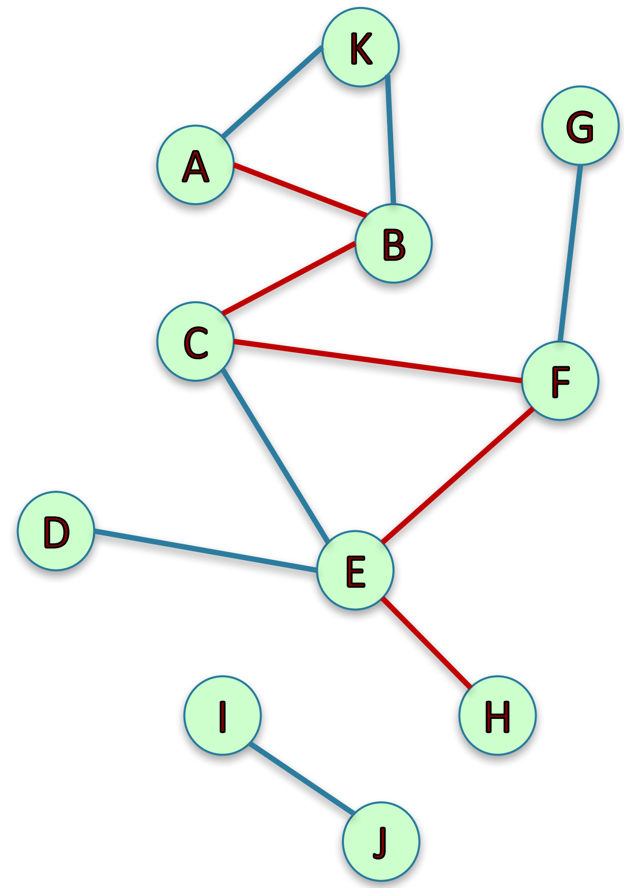


Distance

How far is node A from node H?

Path 1: A – B – C – E – H (4 “hops”)

Path 2: A – B – C – F – E – H (5 “hops”)



Distance

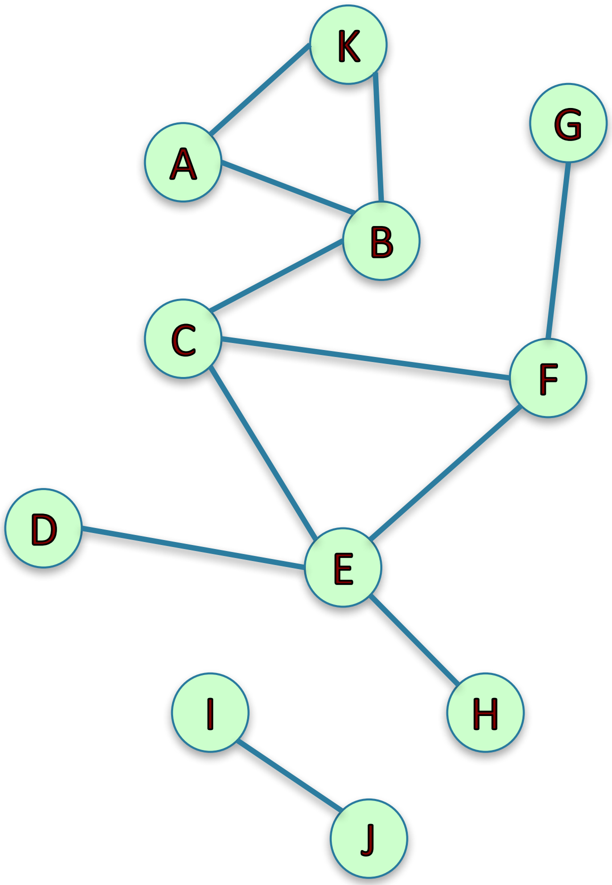
How far is node A from node H?

Path 1: A – B – C – E – H (4 “hops”)

Path 2: A – B – C – F – E – H (5 “hops”)

Path length: Number of steps it contains from beginning to end.

Path 1 has length 4, Path 2 has length 5



Distance

How far is node A from node H?

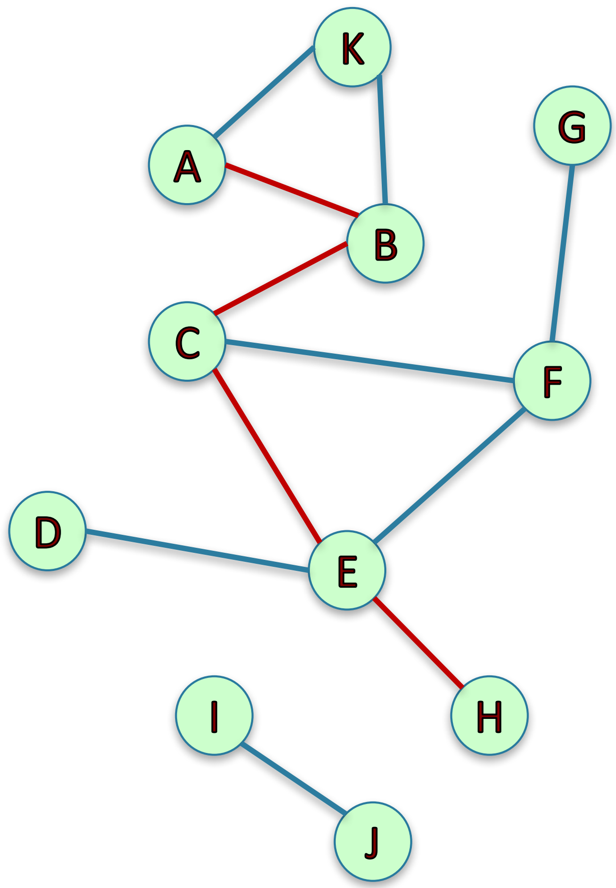
Path 1: A – B – C – E – H (4 “hops”)

Path 2: A – B – C – F – E – H (5 “hops”)

Path length: Number of steps it contains from beginning to end.

Distance between two nodes: The length of the shortest path between them.

The distance between node A and H is 4

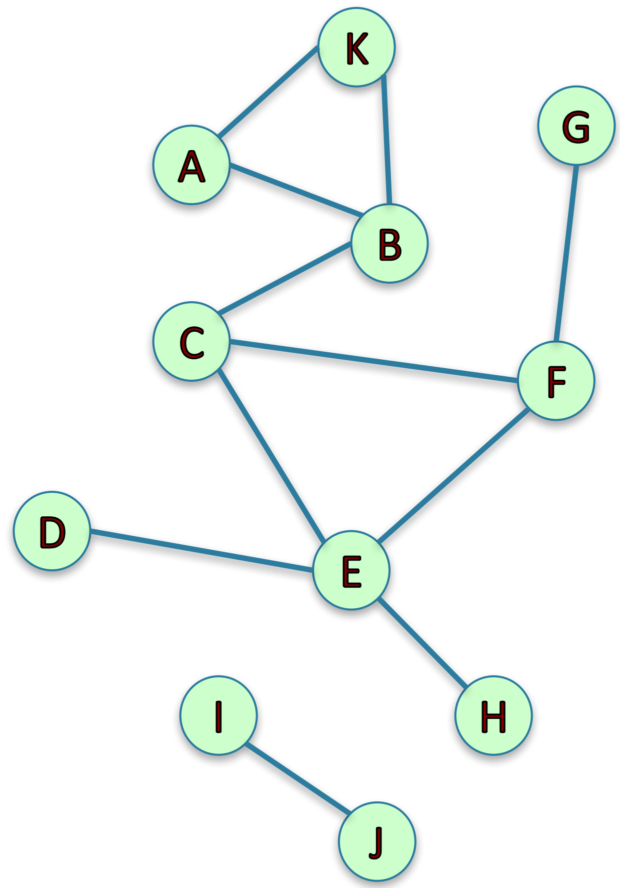


Distance

How to summarize the distances between nodes in a graph?

1. **Average distance** between every pair of nodes.

2. **Diameter**: maximum distance between any pair of nodes.



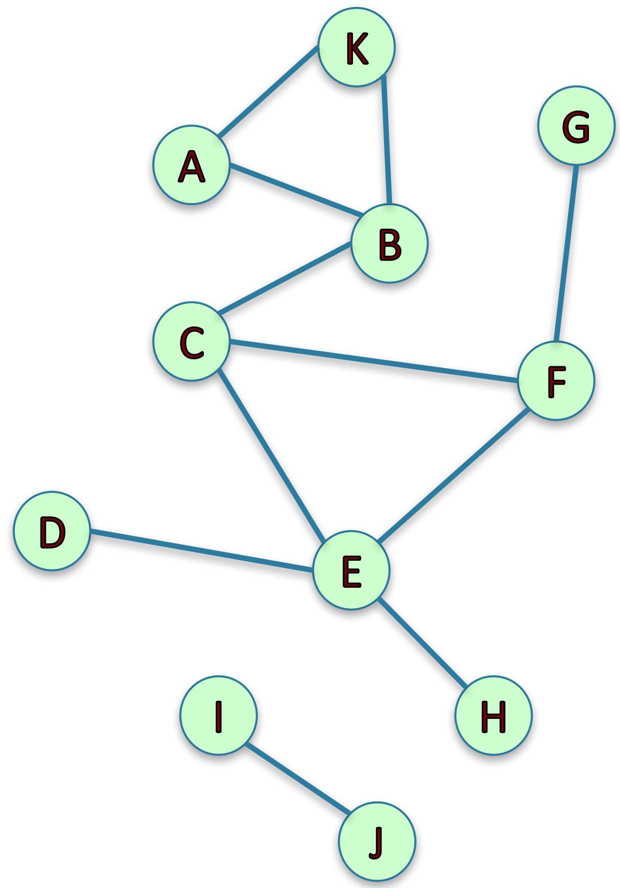
Distance

How to summarize the distances between nodes in a graph?

1. **Average distance** between every pair of nodes.

2. **Diameter**: maximum distance between any pair of nodes.

What is the diameter of this network?



Distance

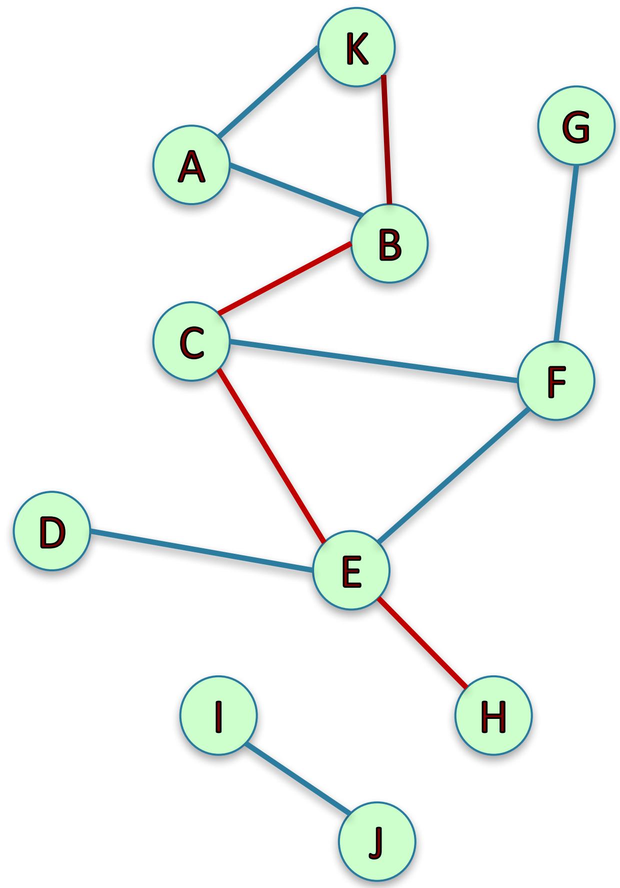
How to summarize the distances between nodes in a graph?

1. **Average distance** between every pair of nodes.

2. **Diameter**: maximum distance between any pair of nodes.

What is the diameter of this network?

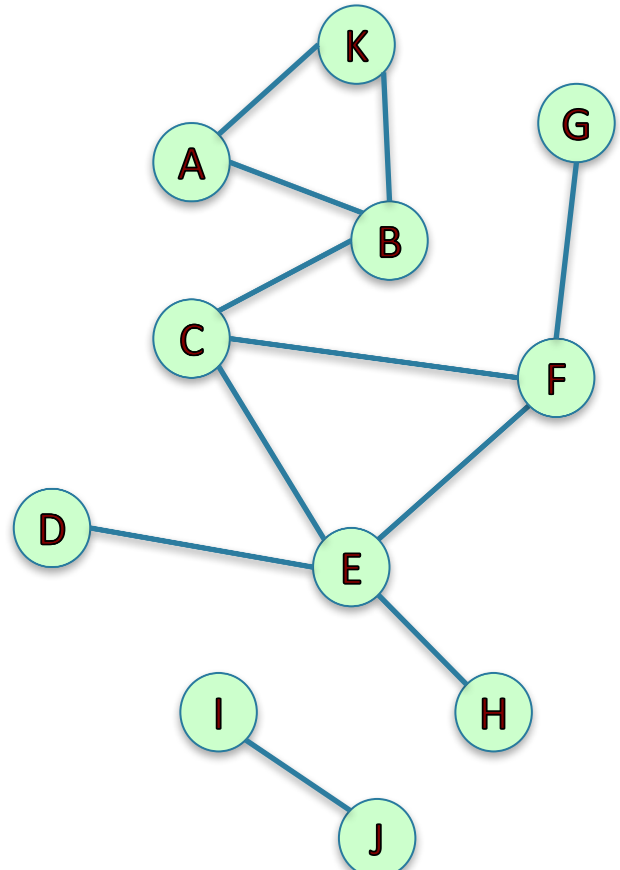
4: K-B-C-E-H



Distance

Finding the distance from node A to *every other node*.

Easy in small networks but tedious in large (real) networks.

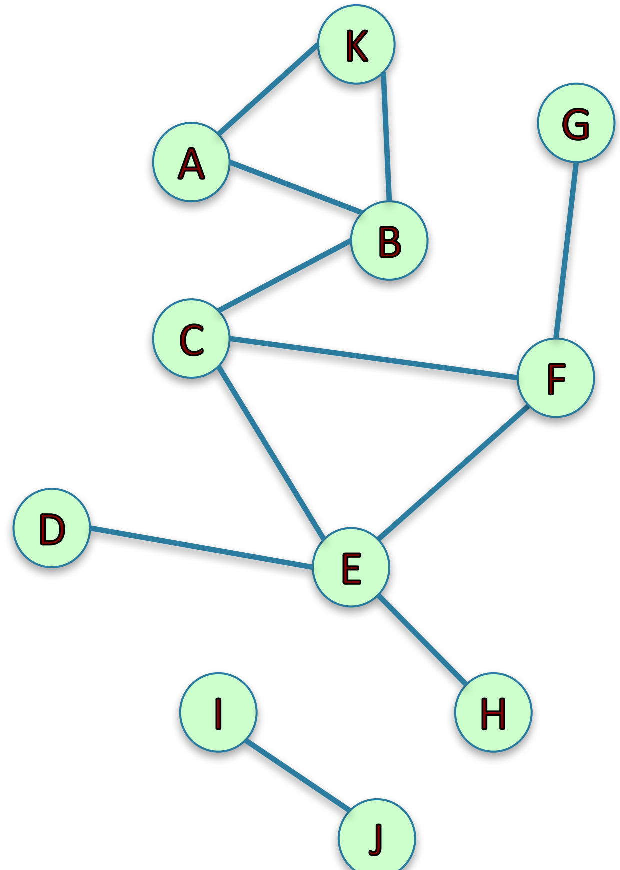


Distance

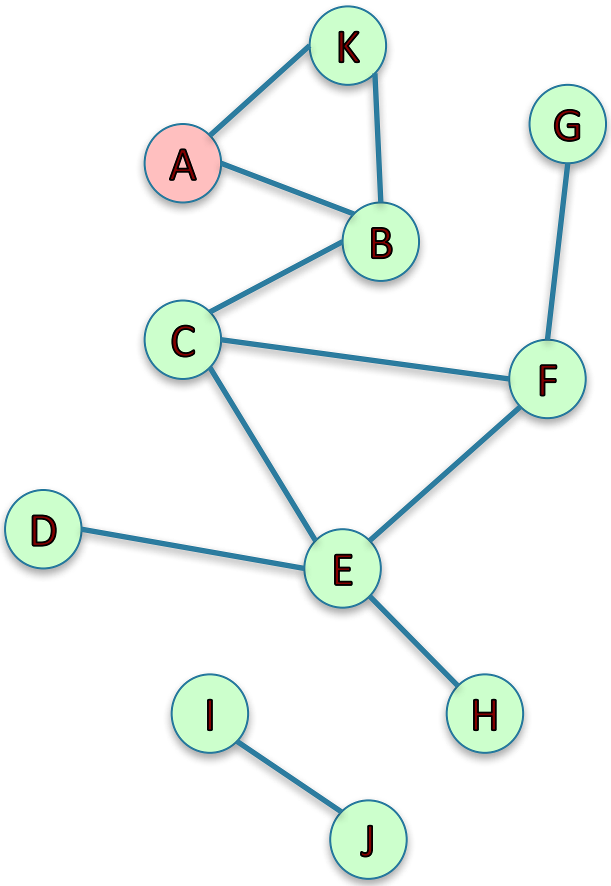
Finding the distance from node A to *every other node*.

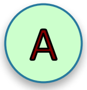
Easy in small networks but tedious in large (real) networks.

Breadth-first search: a systematic and efficient procedure for computing distances in a large network.

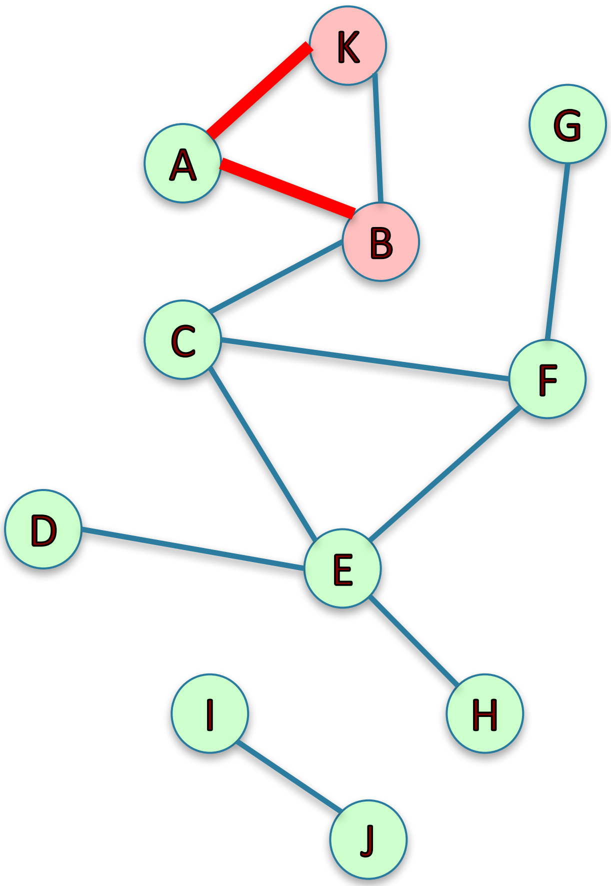


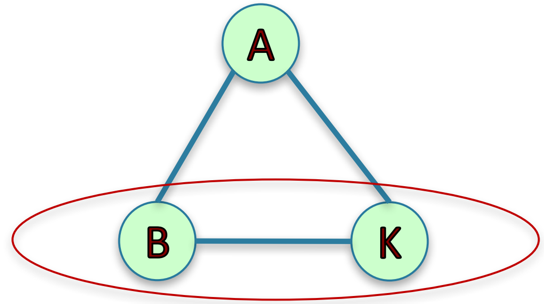
Breadth-First Search





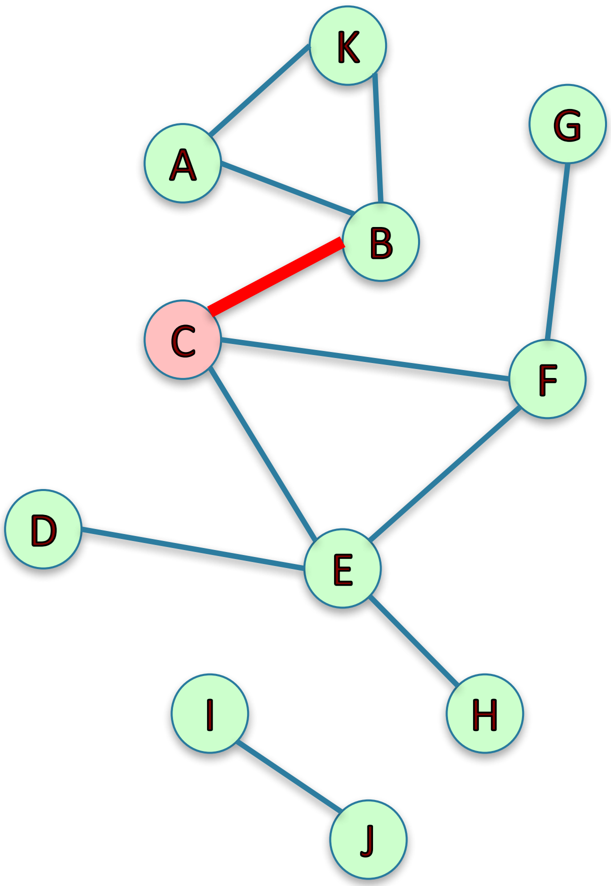
Breadth-First Search

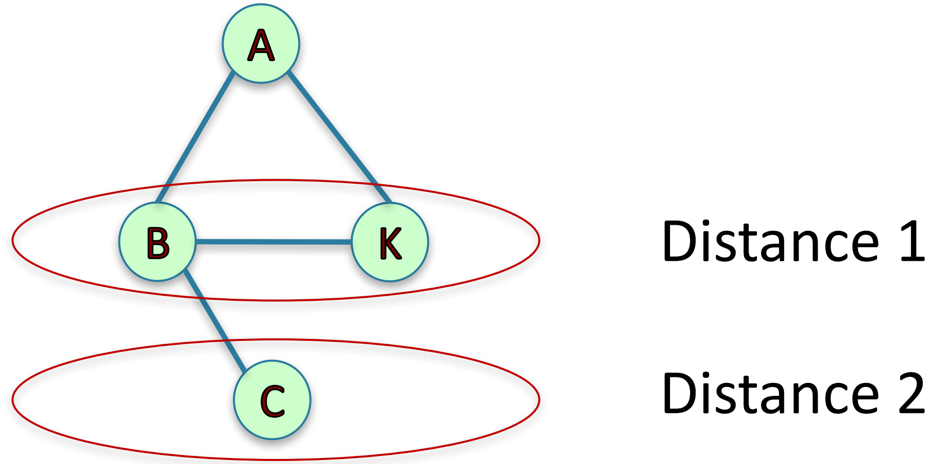




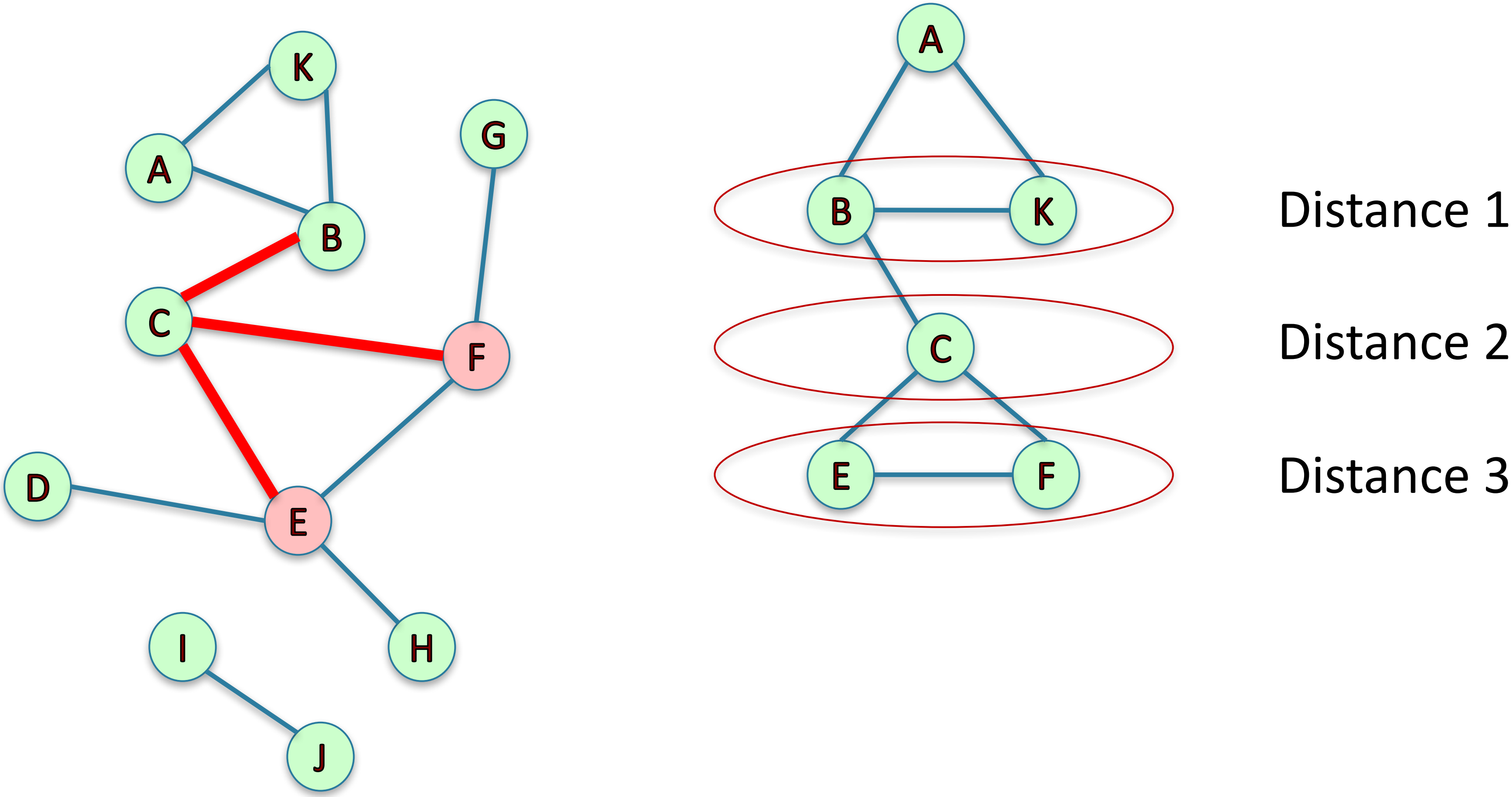
Distance 1

Breadth-First Search

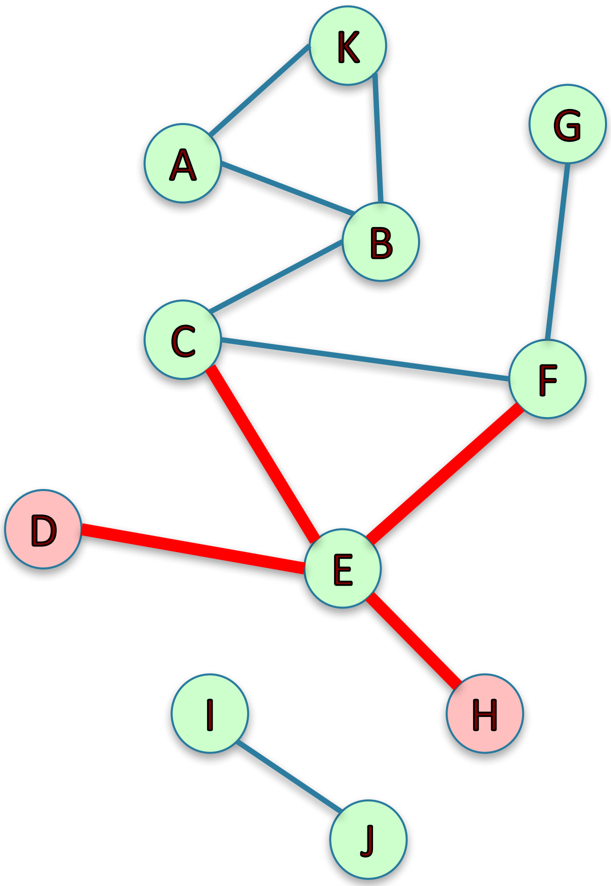


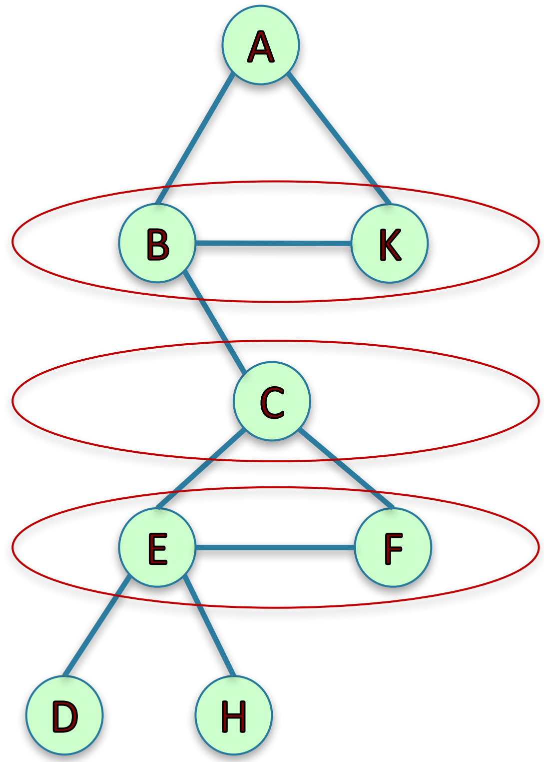


Breadth-First Search



Breadth-First Search





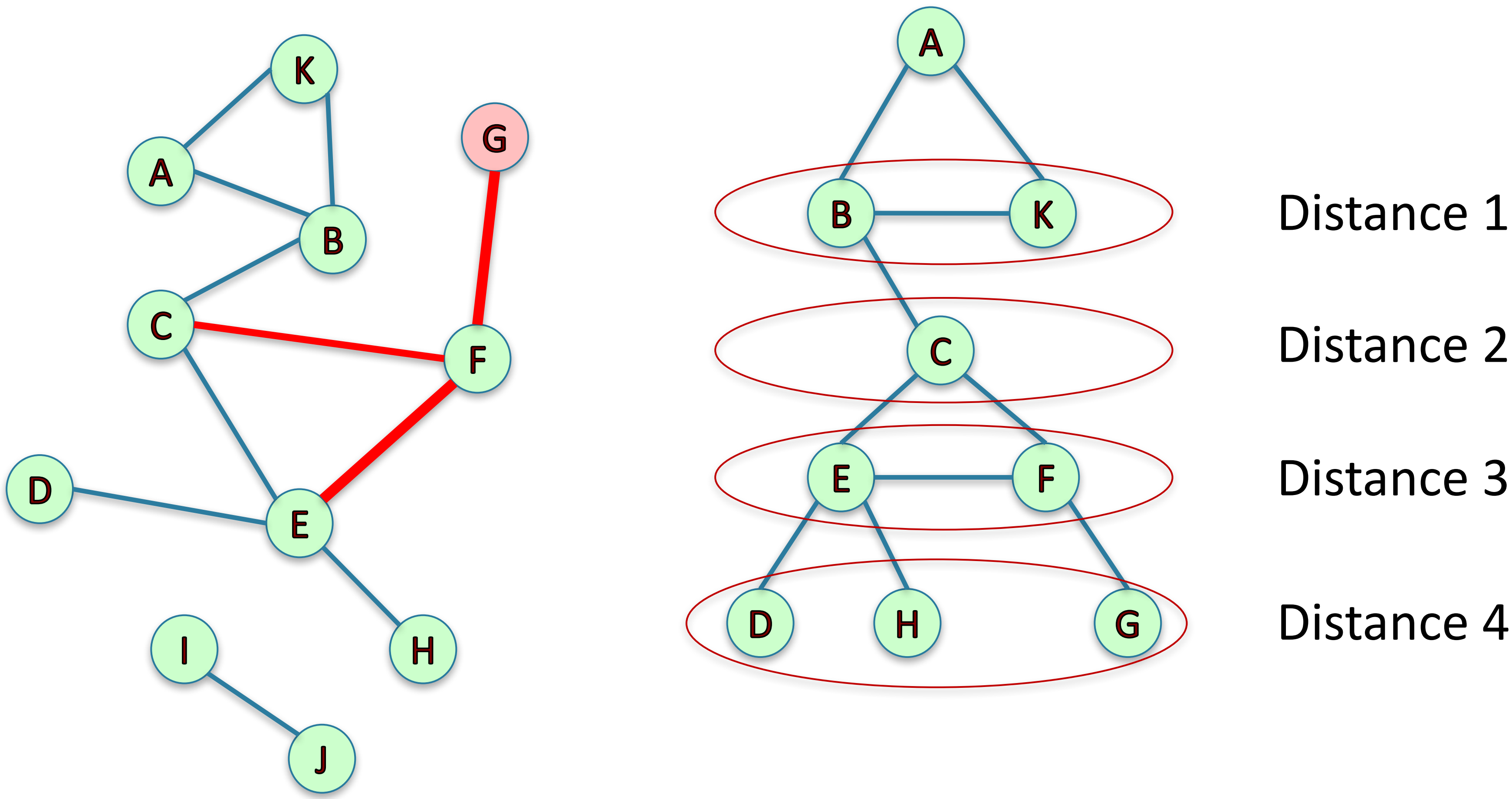
Distance 1

Distance 2

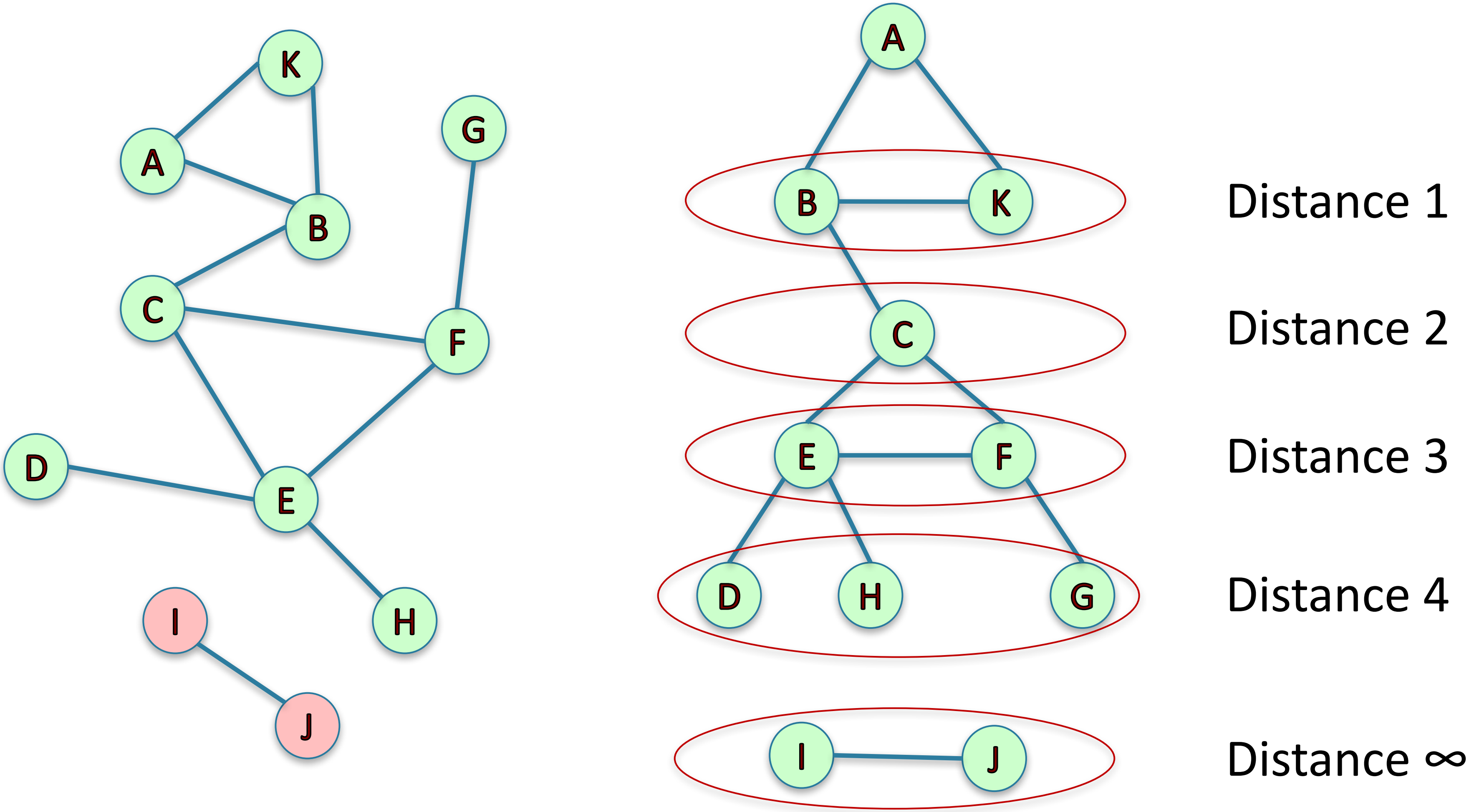
Distance 3

Distance 4

Breadth-First Search



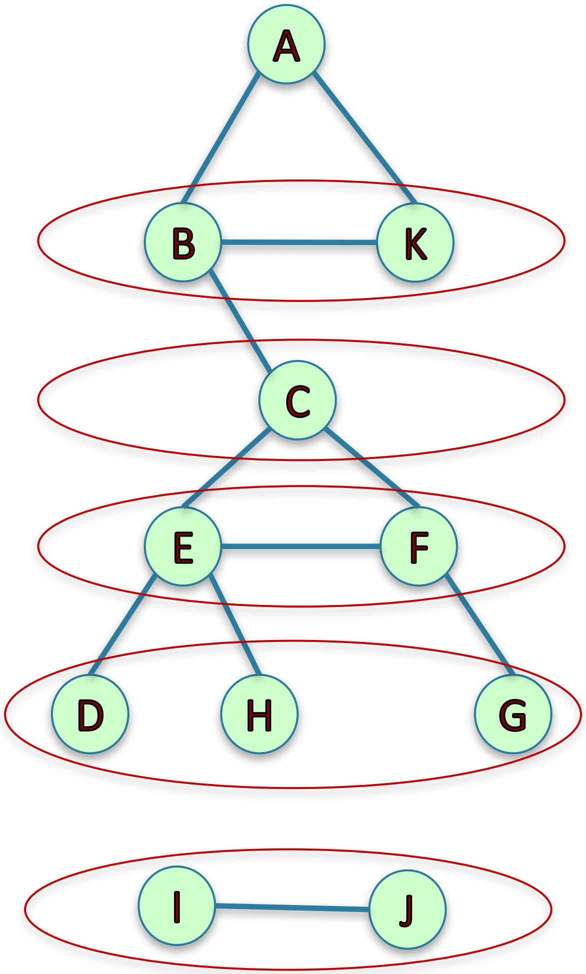
Breadth-First Search



Breadth-First Search

Finding all distance to a seed node.

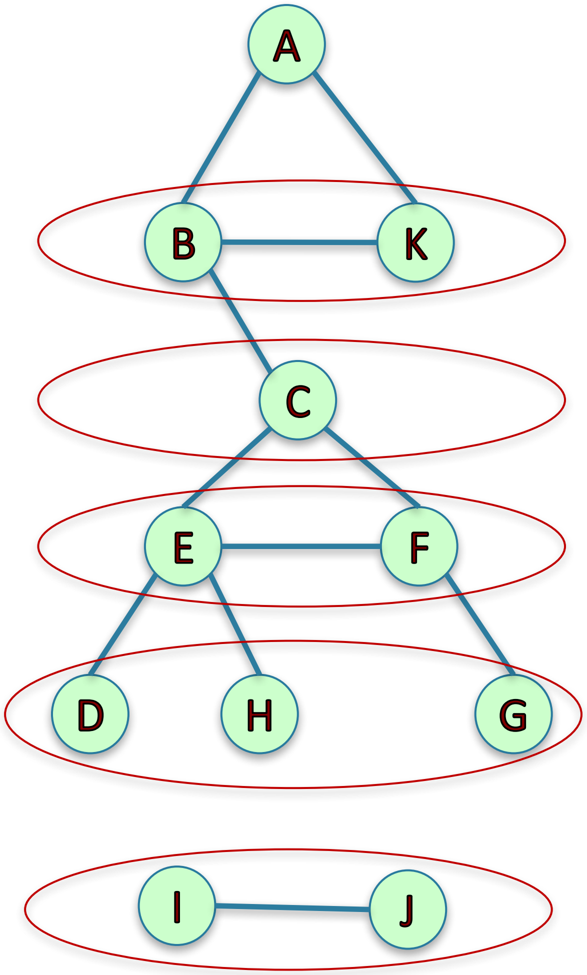
1. Find all connections to the seed node and declare them to be at distance 1



Breadth-First Search

Finding all distance to a seed node.

1. Find all connections to the seed node and declare them to be at distance 1
2. For each node found in step 1, find all their **connections which were not previously found**, and declare them to be at distance 2.



Breadth-First Search

Finding all distance to a seed node.

1. Find all connections to the seed node and declare them to be at distance 1
2. For each node found in step 1, find all their **connections which were not previously found**, and declare them to be at distance 2.
3. Continue discovering new nodes in layers, with each new layer adding one to the distance. Each layer consists of all **undiscovered** nodes connected to some node in the previous layer.