# Week 5 Video 5

## Relationship Mining
## Network Analysis

# Today's Class

- Network Analysis

# Network Analysis

- Analysis of anything that can be seen as connections between nodes

- Most common – social networks
  - Connections between friends on the internet
  - Connections between students in a class
  - Connections between collaborators in a work project

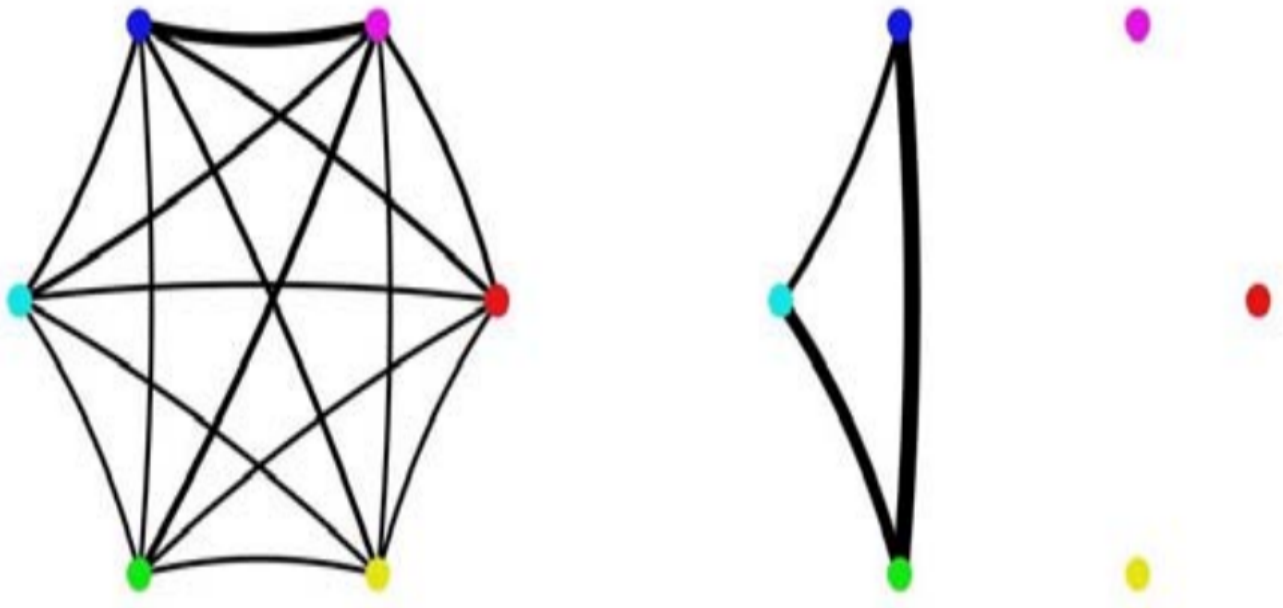
# Network Analysis

- Could also be considered structure discovery
- Placed here in the course because of how it's typically used in practice

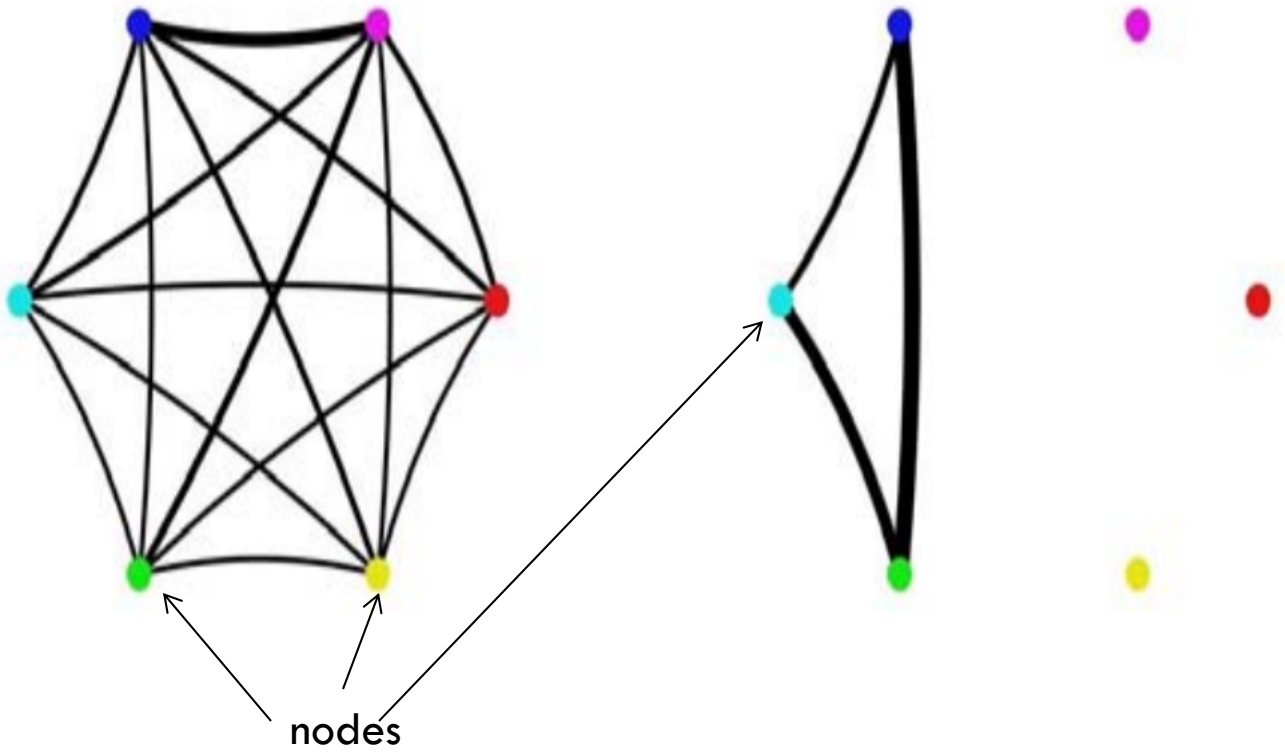
# General Postulates of Network Analysis

- There are things, referred to as nodes or vertices
- Nodes have connections to other nodes, referred to as ties or links
- Nodes can have different types or identities
- Links can have different types or identities
- Links can have different strengths

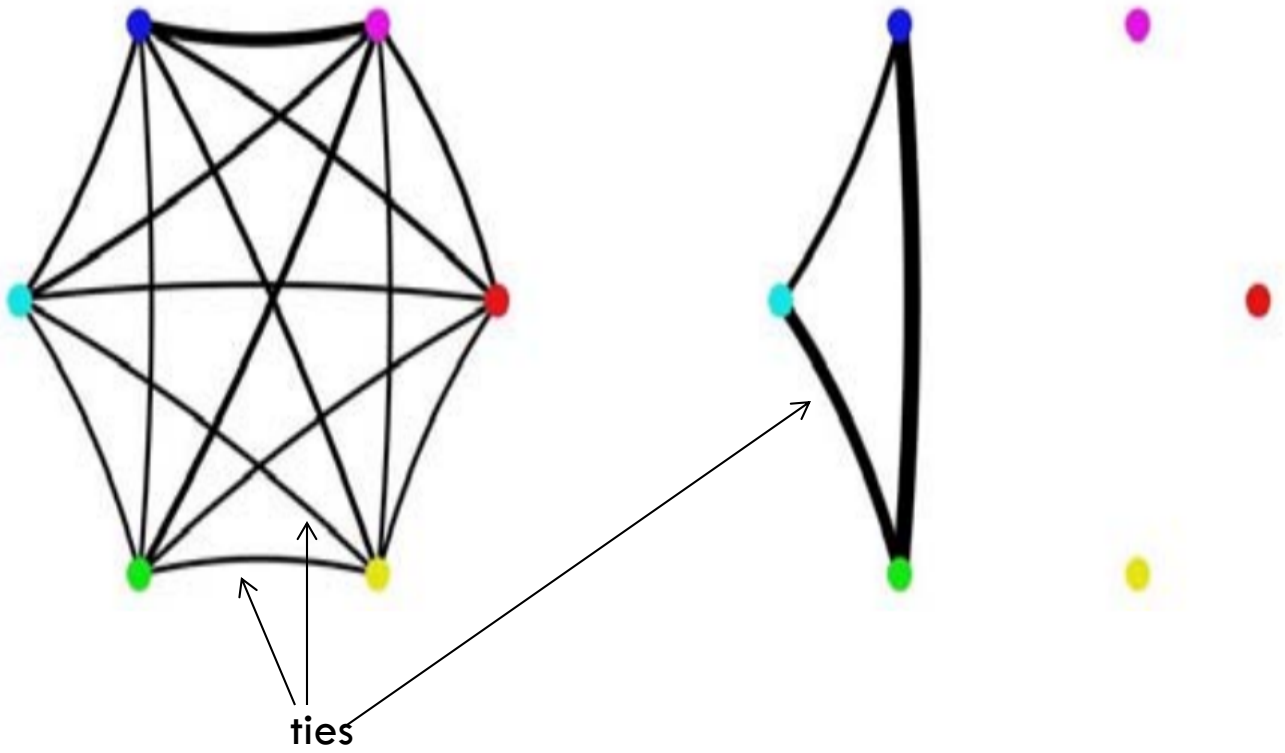
# Example: (Student work groups – Kay et al., 2006)



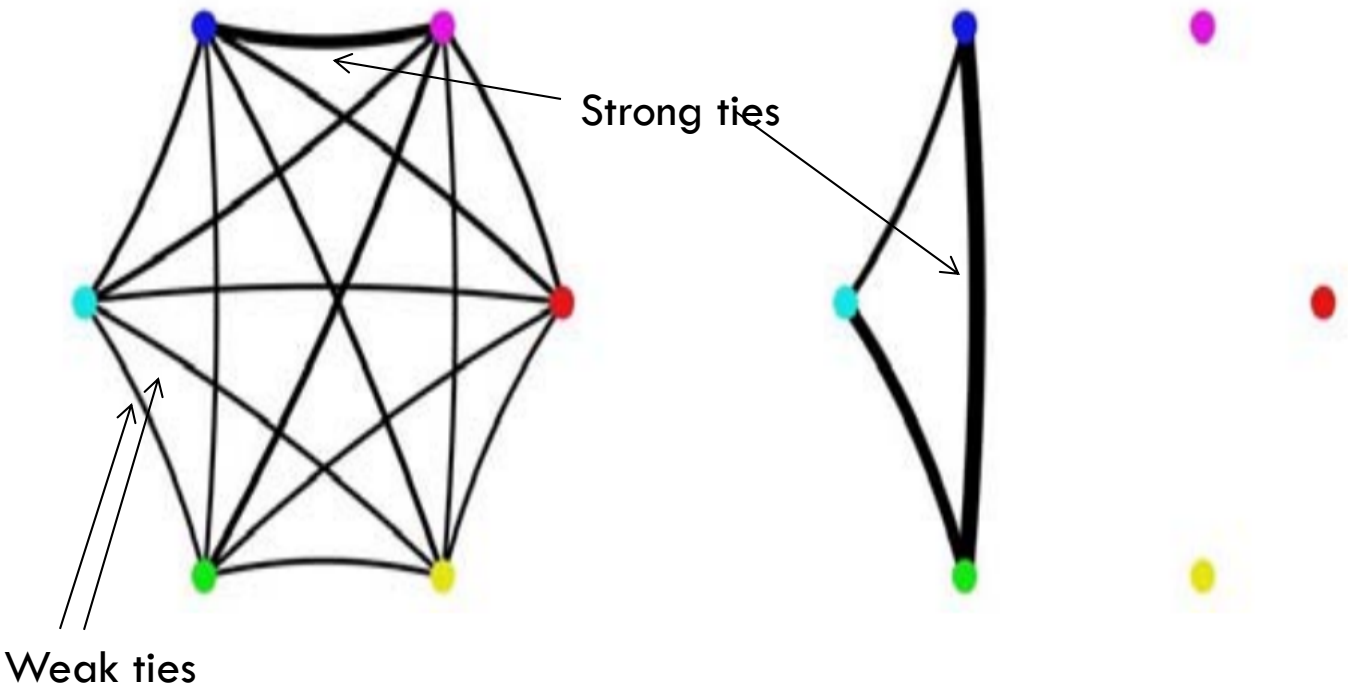
# Example: (Student work groups – Kay et al., 2006)



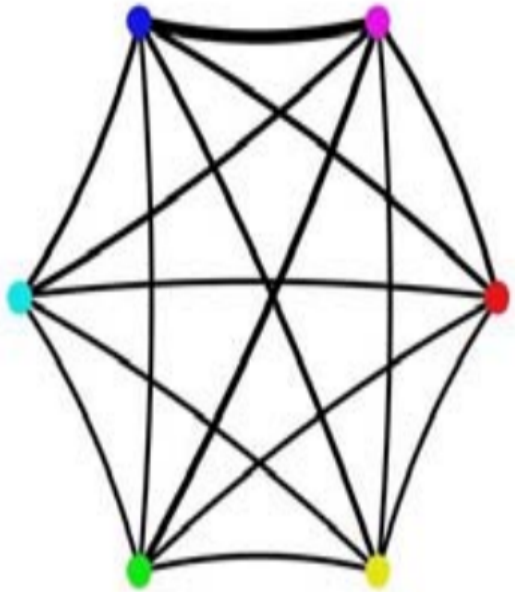
# Example: (Student work groups – Kay et al., 2006)

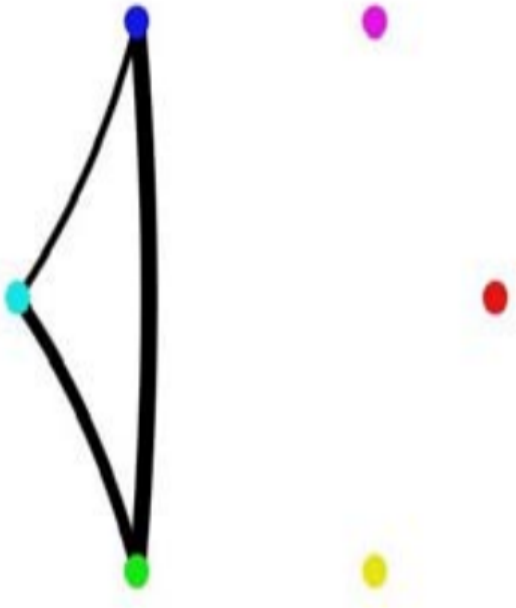


# Example: (Student work groups – Kay et al., 2006)

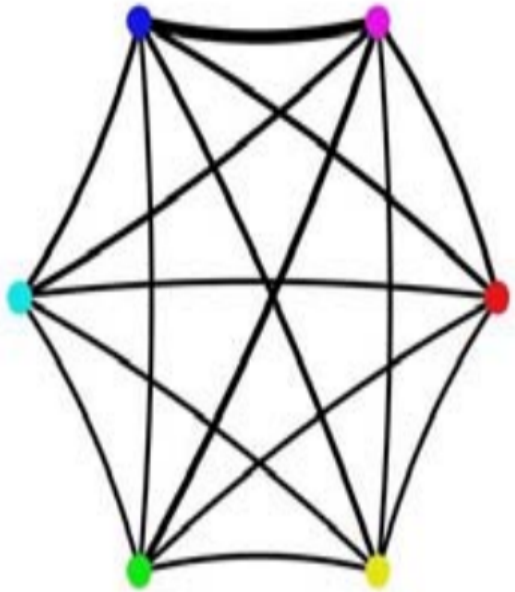


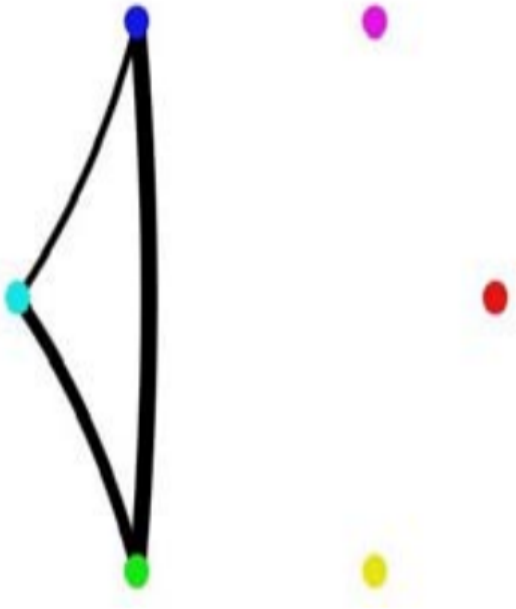
# Which student group works together better?



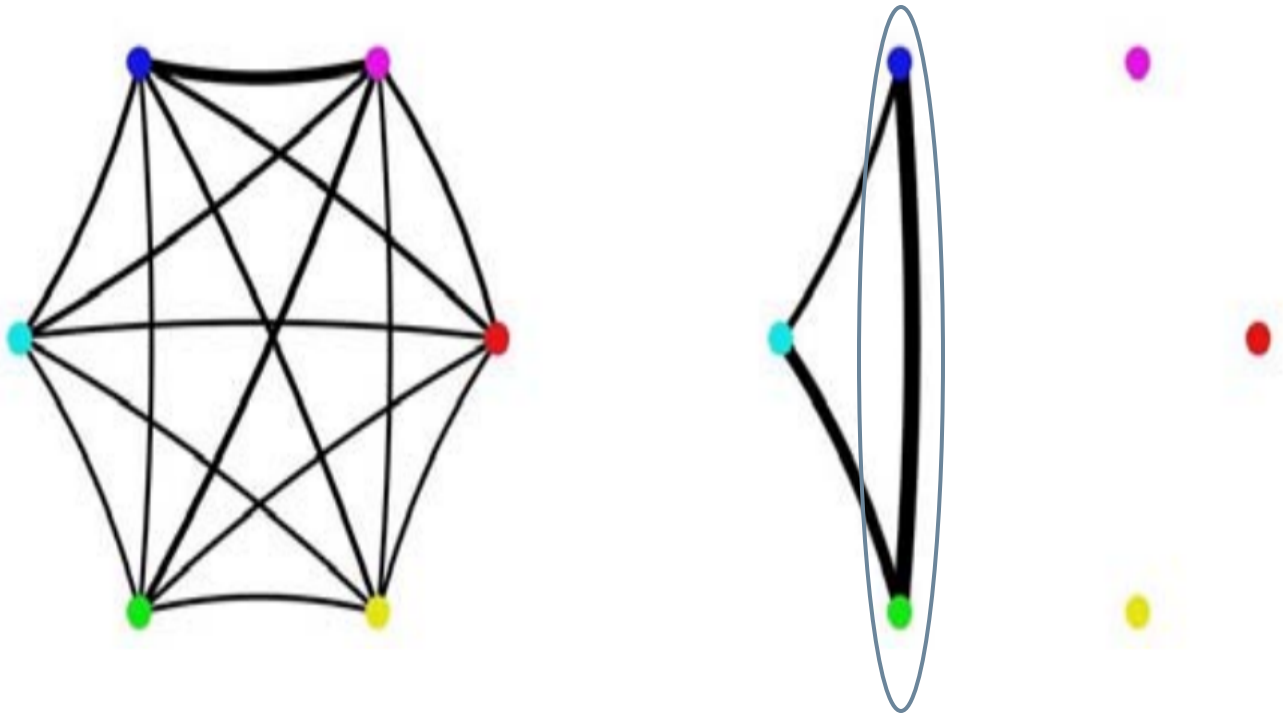


# Which is the most collaborative pair?

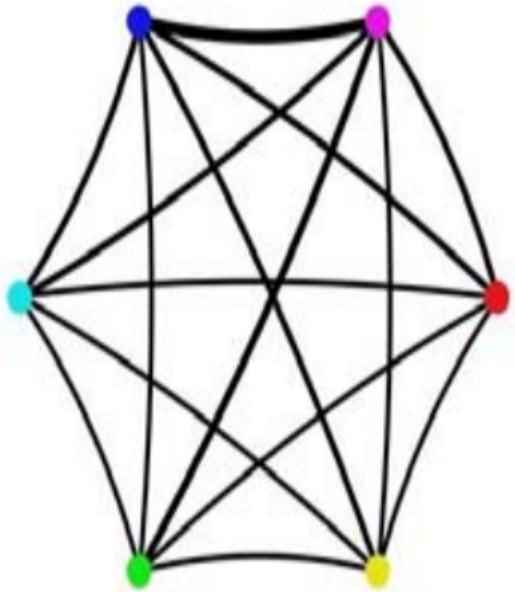


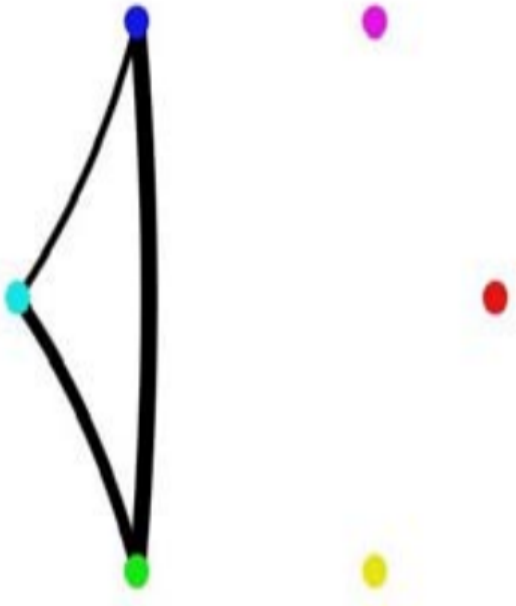


# Which is the most collaborative pair?

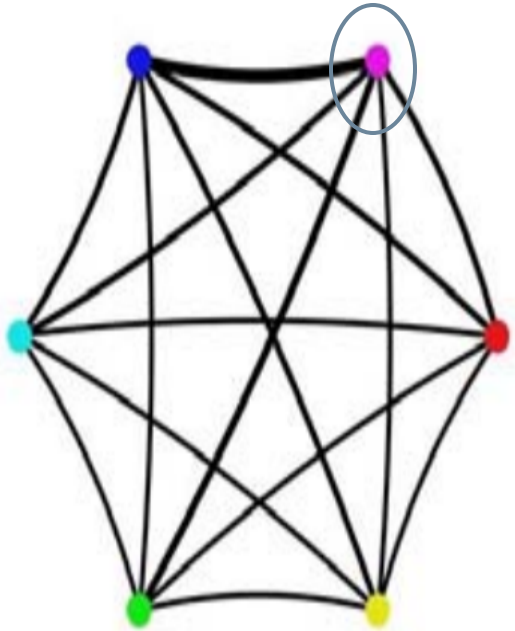


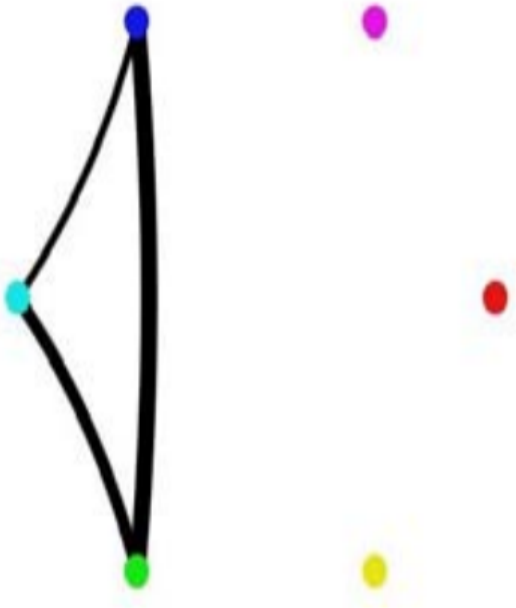
# Who is the most collaborative student?





# Who is the most collaborative student?





# Types

- In a graph of classroom interactions, there could be several different types of nodes
  - Teacher
  - TA
  - Student
  - Project Leader
  - Project Scribe

# Types

- In a graph of classroom interactions, there could be several types of links
  - Leadership role (X leads Y)
  - Working on same learning resource
  - Helping act
  - Criticism act
  - Insult

  - Note that links can be directed or undirected

# Strength

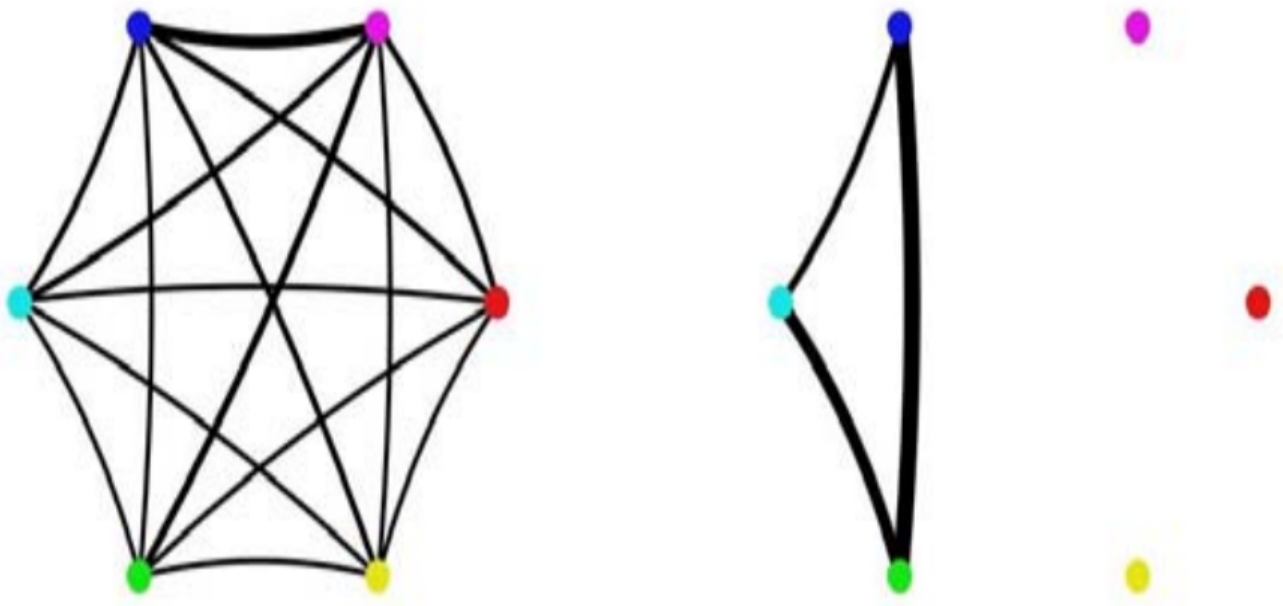
- In a graph of classroom interactions, links could be stronger or weaker due to
  - Intensity of act
  - Frequency of act

# Network Analysis

- Use network graphs to study the patterns and regularities of the relationships between the nodes

# Density

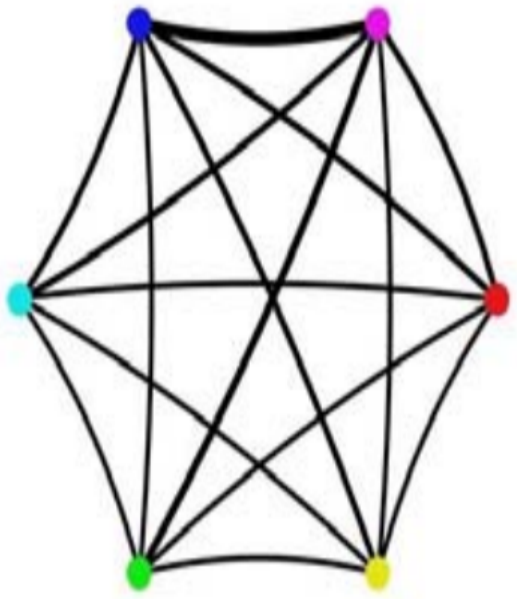
- Proportion of possible lines that are actually present in graph
- What is the density of these graphs?



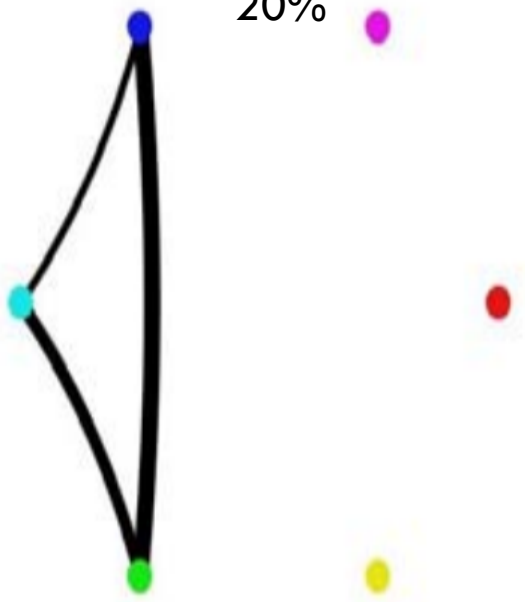
# Density

- Proportion of possible lines that are actually present in graph
- What is the density of these graphs?

100%



3/15= 20%

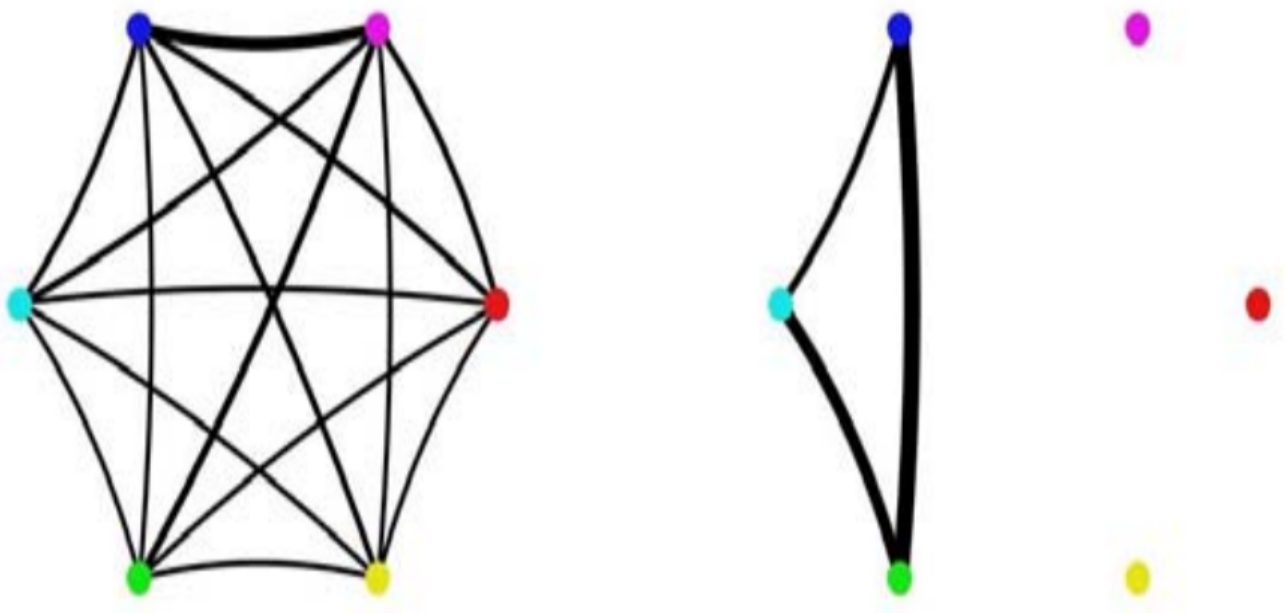


# Density

- Could be used to figure out how collaborative a class is overall

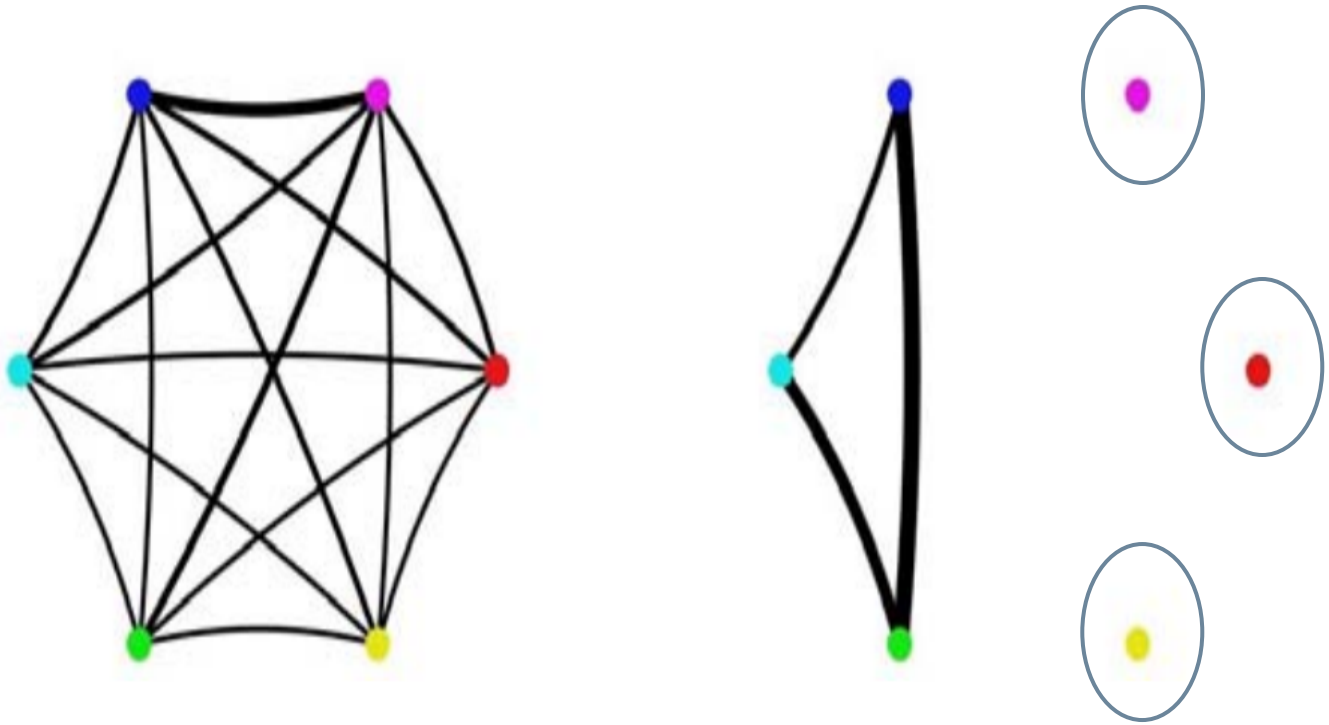
# Reachability

- A node is “reachable” if a path goes from any other node to it
- Which nodes are unreachable?



# Reachability

- A node is “reachable” if a path goes from any other node to it
- Which nodes are unreachable?



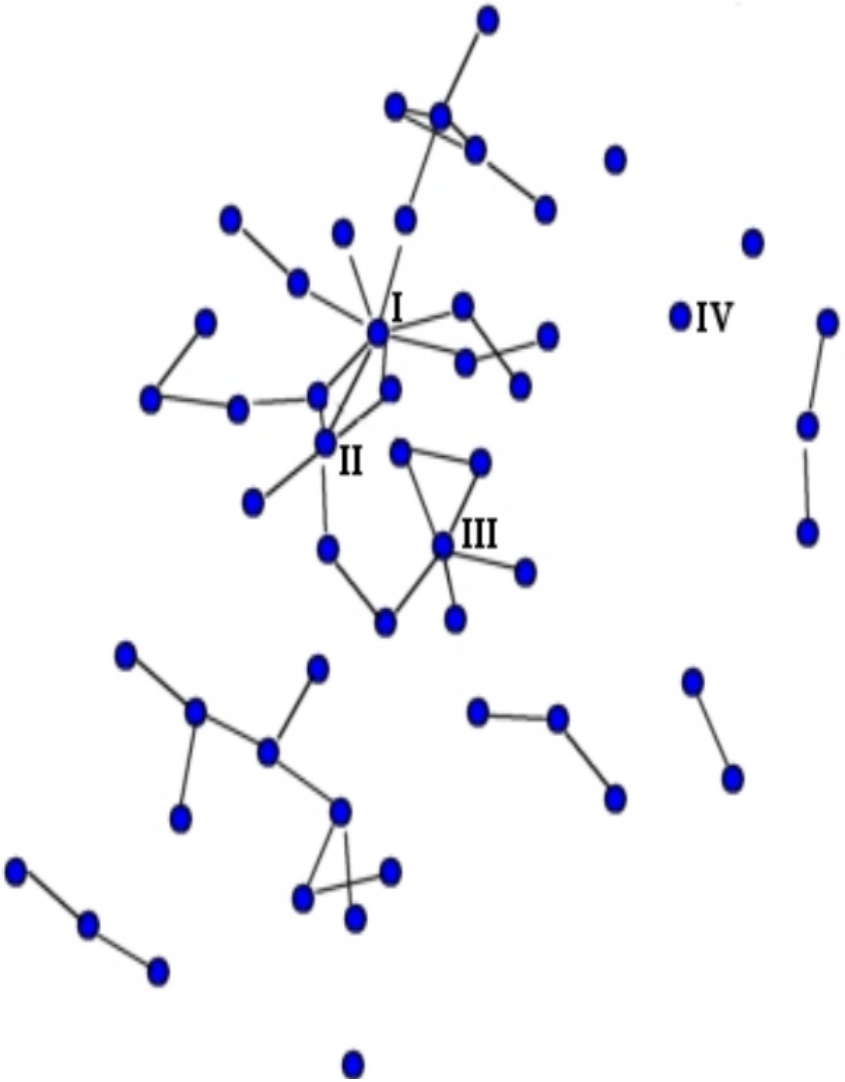
# Reachability

- Are there any students who don't collaborate with anybody?

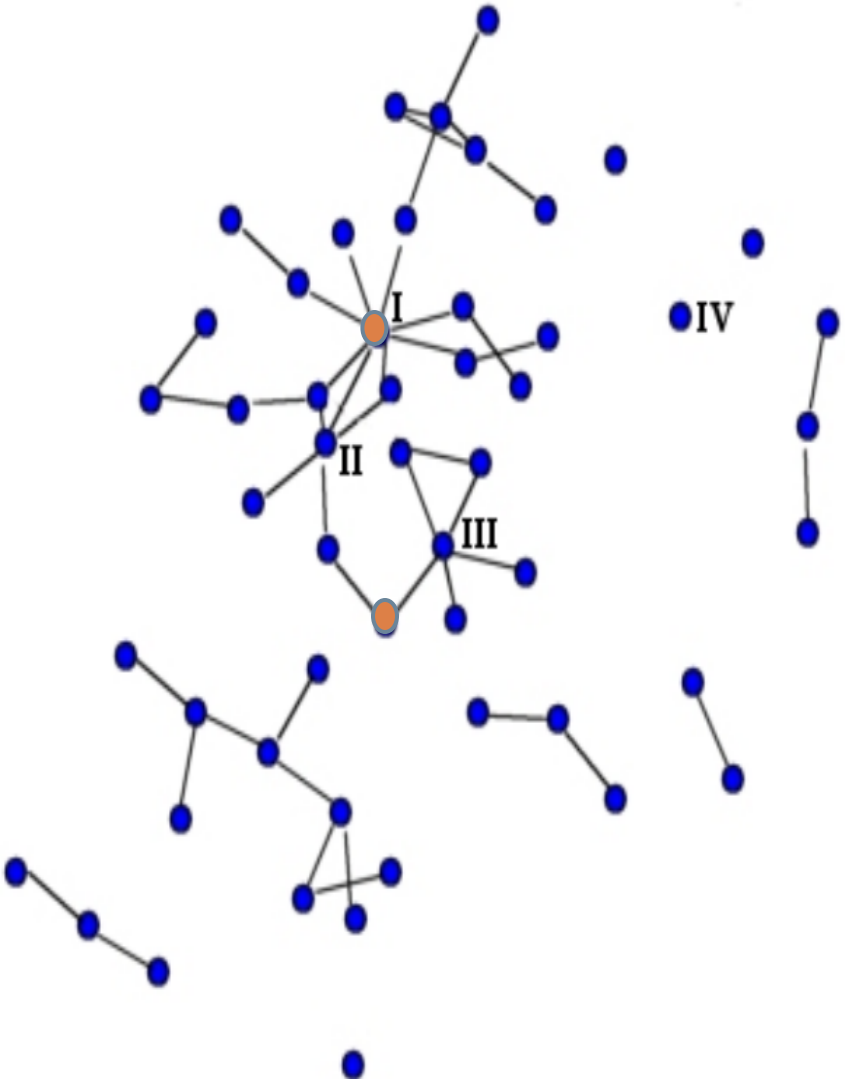
# Geodesic Distance

- The number of edges between one node N and another node M, in the shortest path connecting them

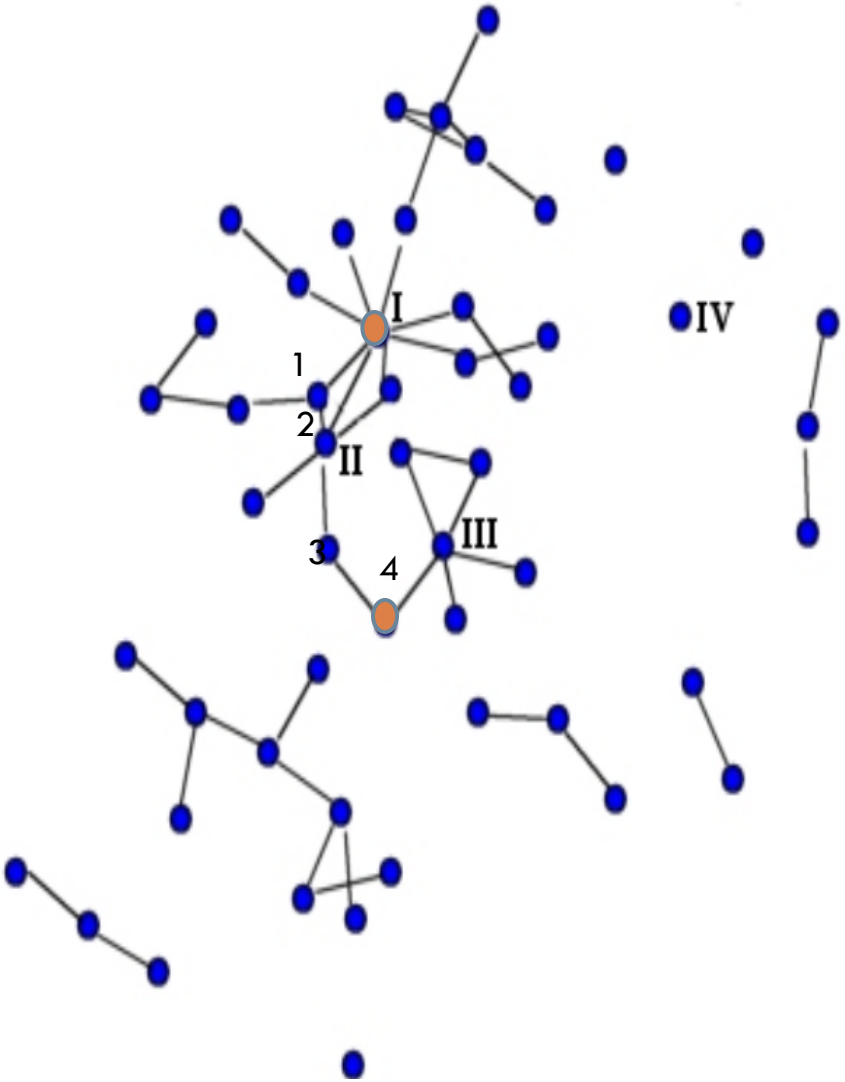
# Student social network: (Dawson, 2008)



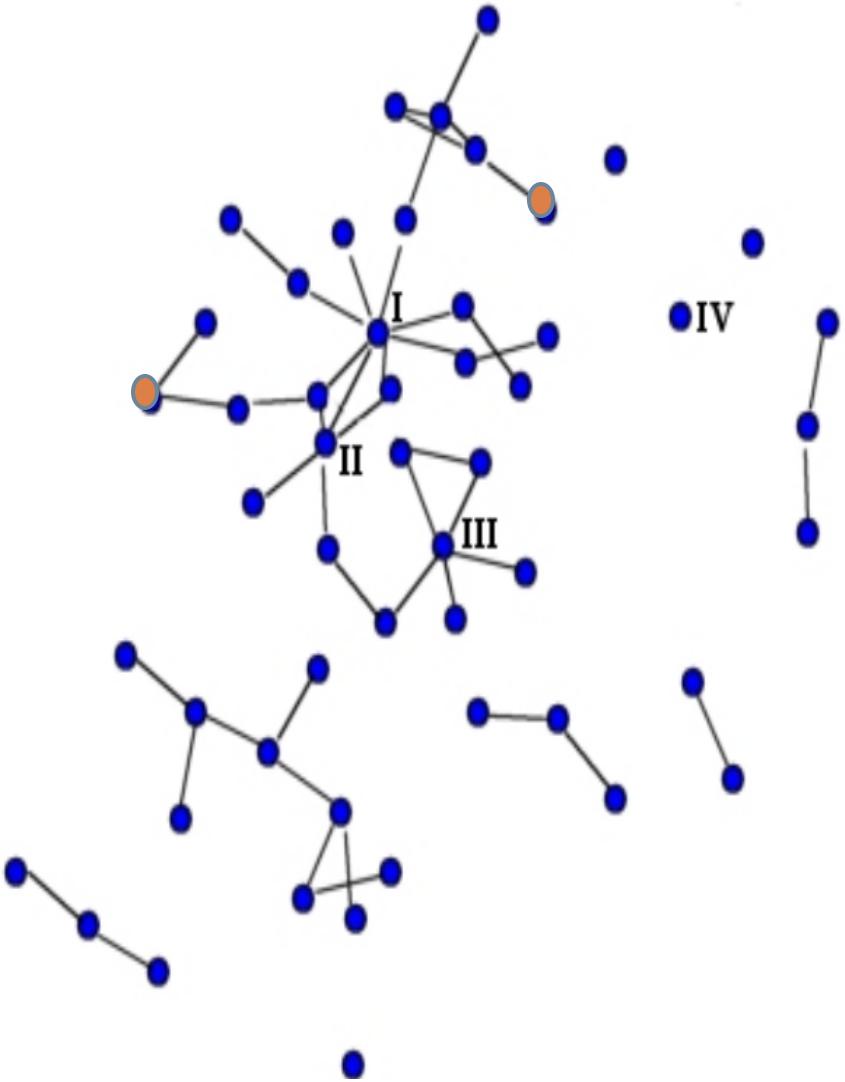
# What is the geodesic distance?



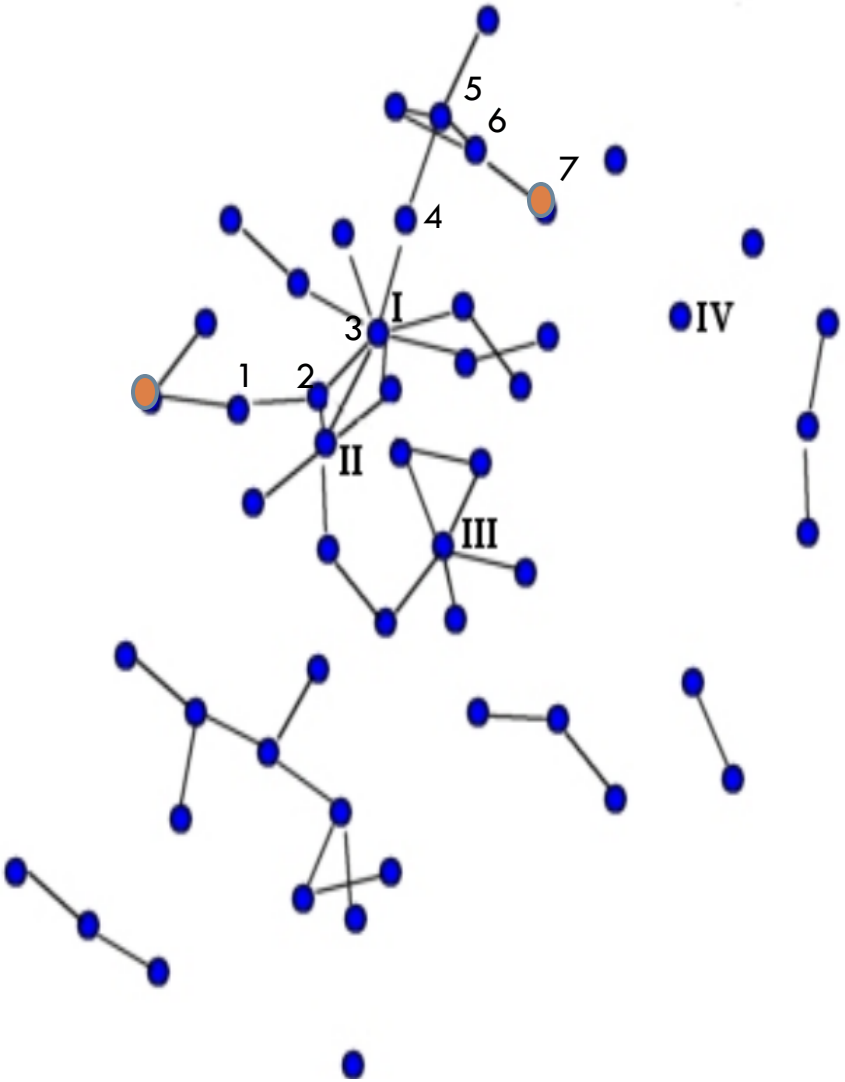
# Geodesic distance = 4



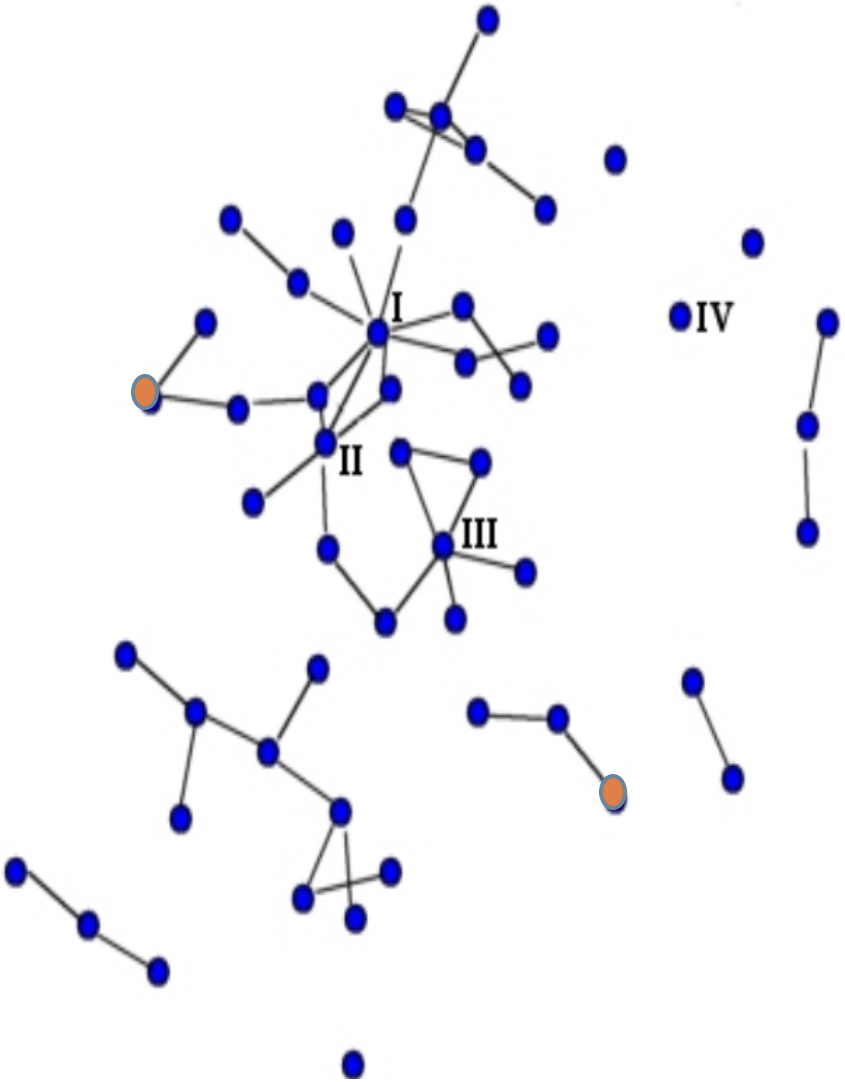
# What is the geodesic distance?



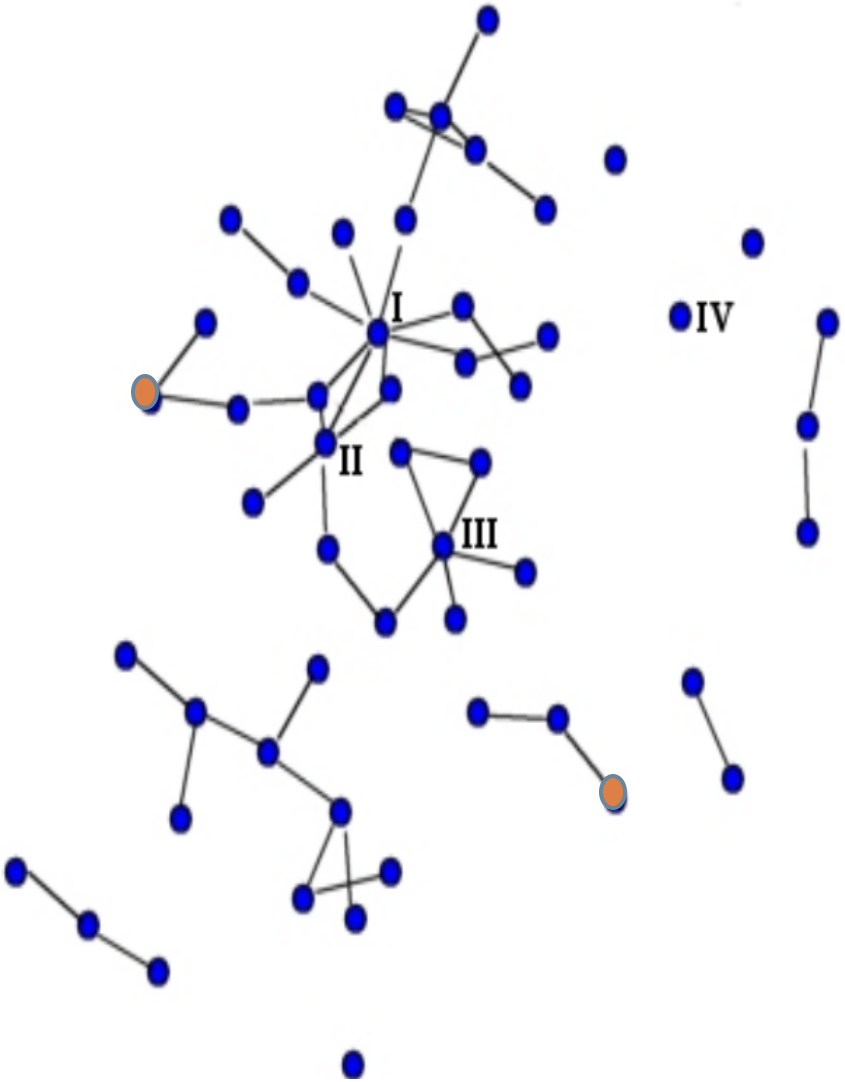
# Geodesic Distance = 7



# What is the geodesic distance?

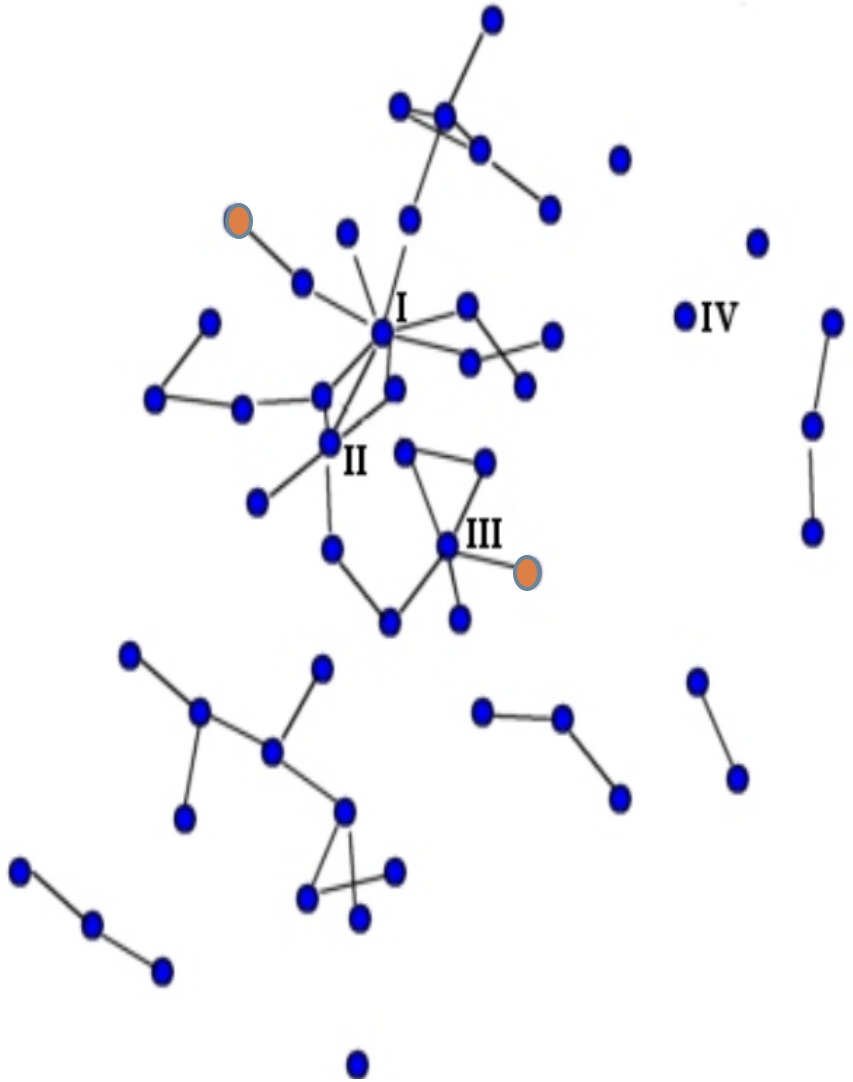


# Geodesic Distance = Infinite



# Quiz

## What is the geodesic distance?




A) 6

B) 7

C) 8

D) 9

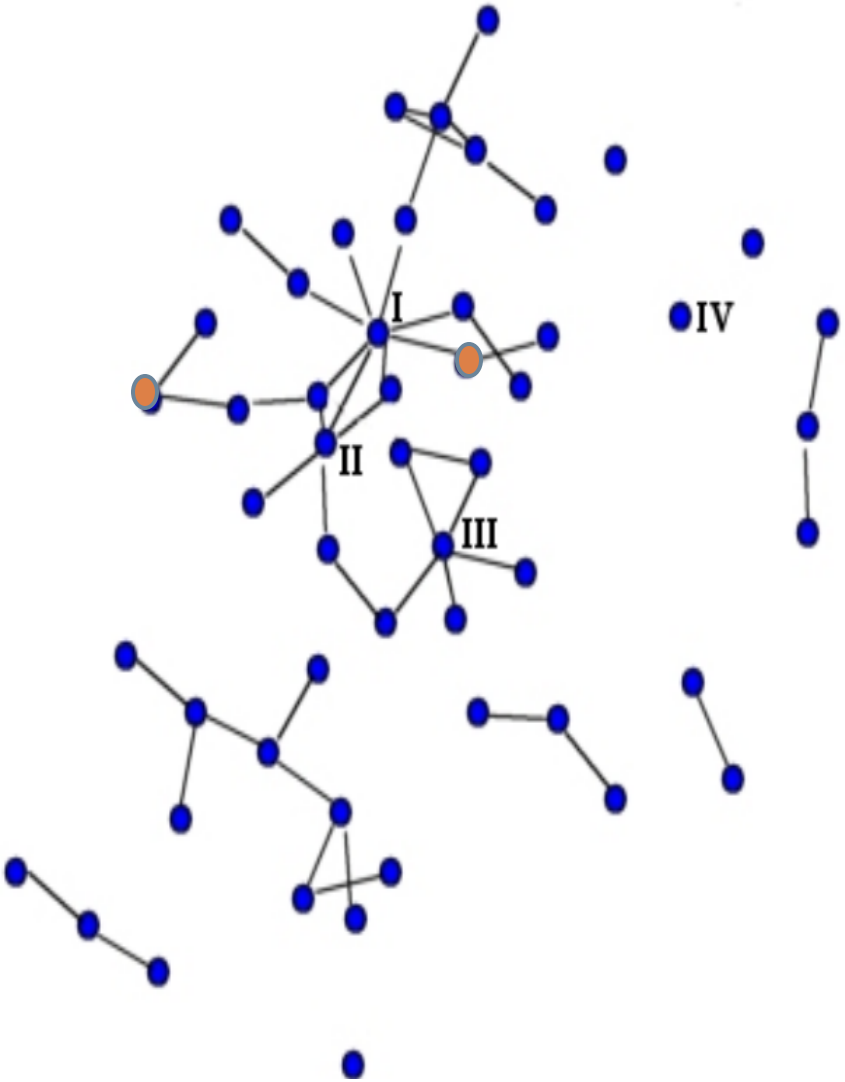
# Geodesic Distance

- How many people does an idea need to go through to get between people?

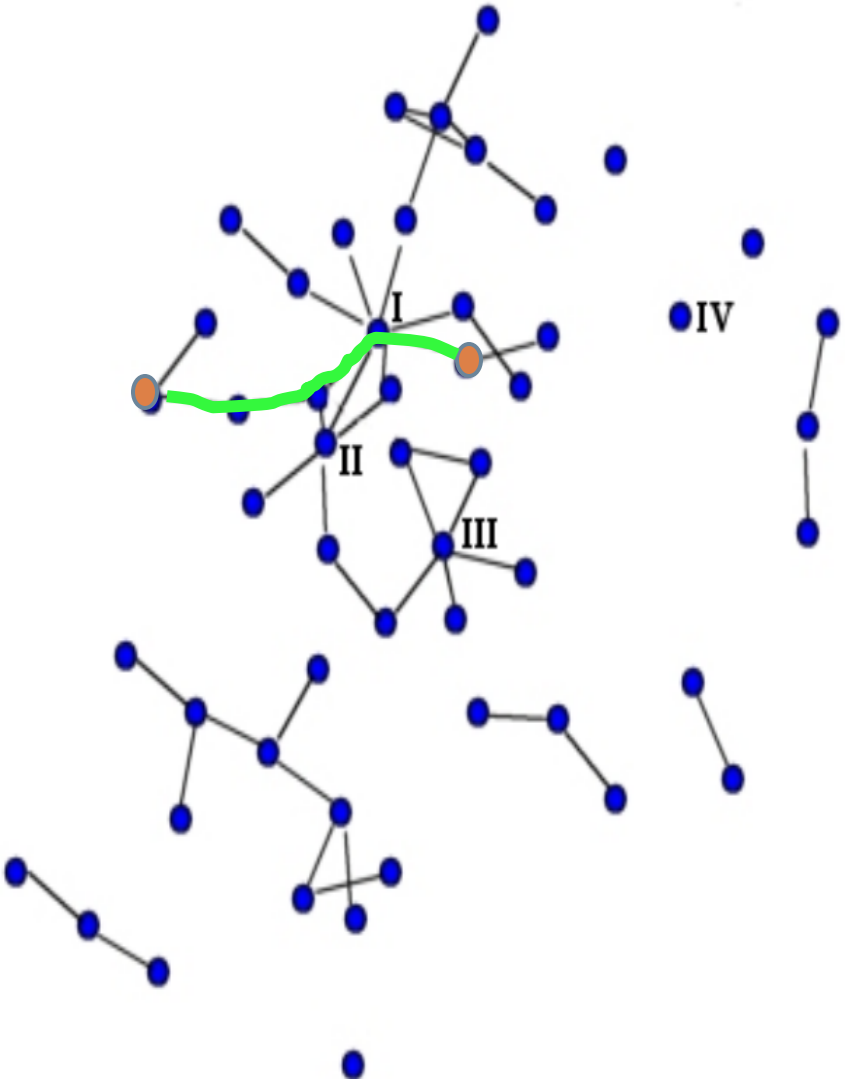
# Flow

- How many possible paths are there between node N and node M, that do not repeat a node?

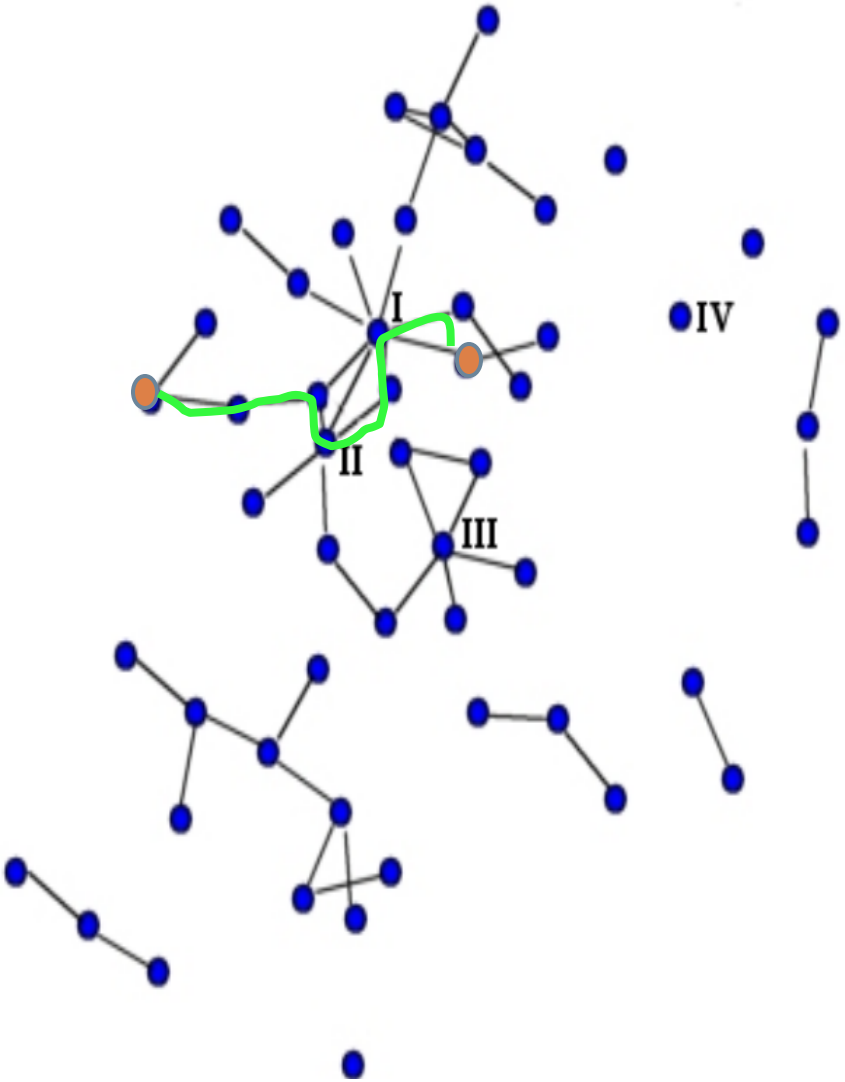
# What is the flow?



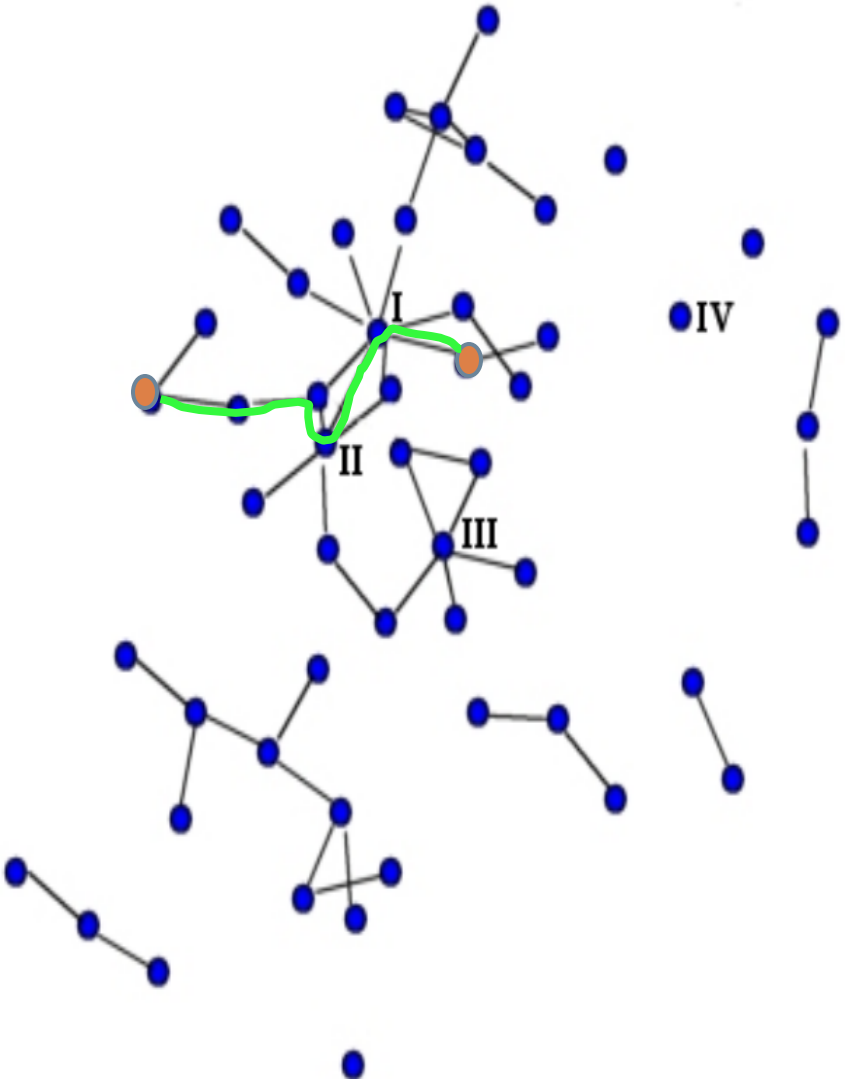
# 1



# 2



# 3



# Flow

- How many possible paths are there for an idea to go between people?

# Centrality

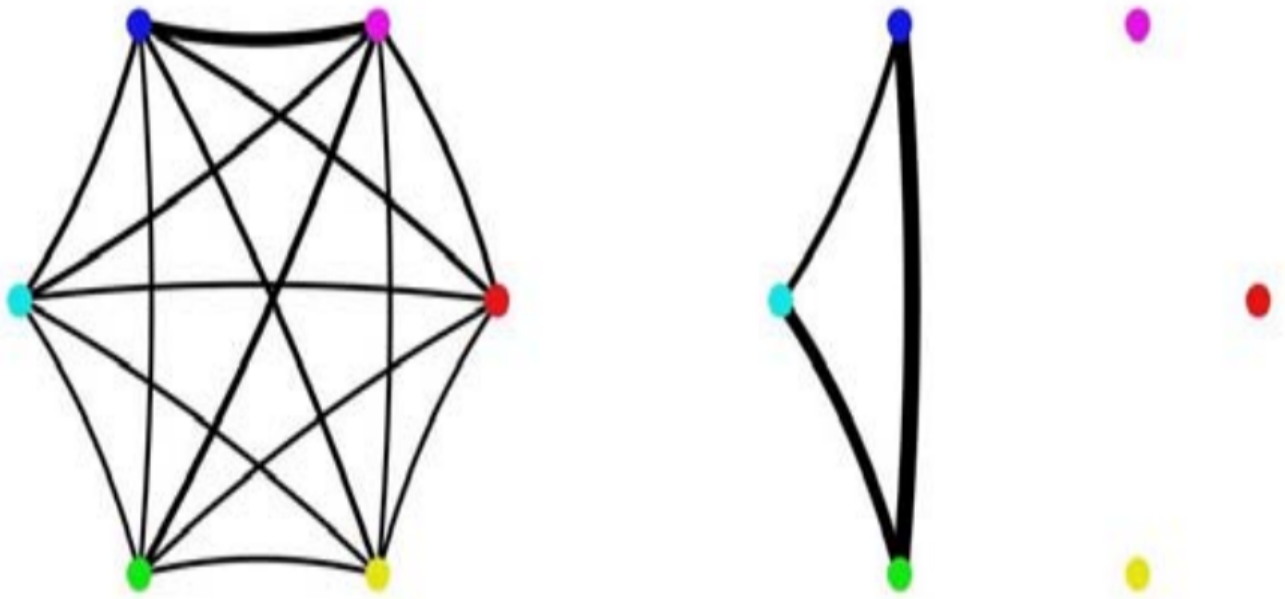
- How important is a node within the graph?
- Which kids are the popular or influential kids?

# Centrality

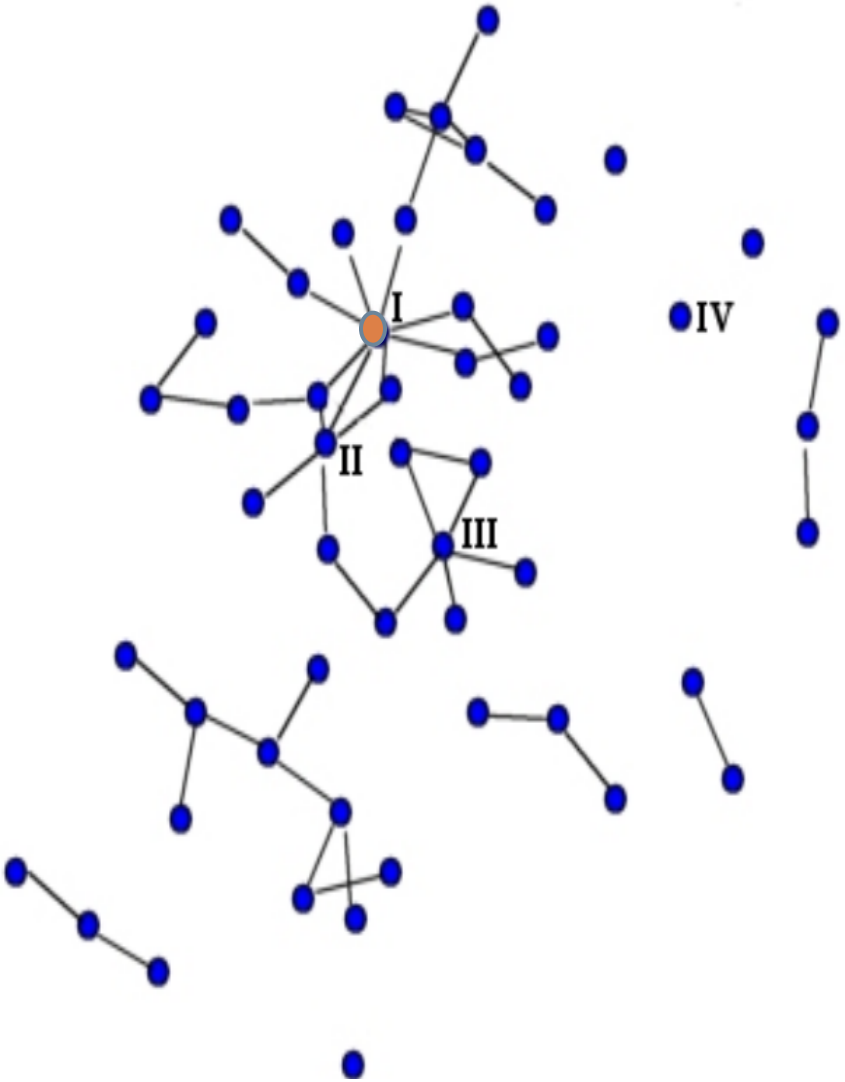
- Four common measures
  - Degree centrality
  - Closeness centrality
  - Betweeness centrality
  - Eigenvector centrality

# Nodal Degree

- Number of lines that connect to a node



# The node with the highest nodal degree



# Nodal Degree

- Indegree: number of lines that come into a node
- Outdegree: number of lines that come out of a node

# Closeness

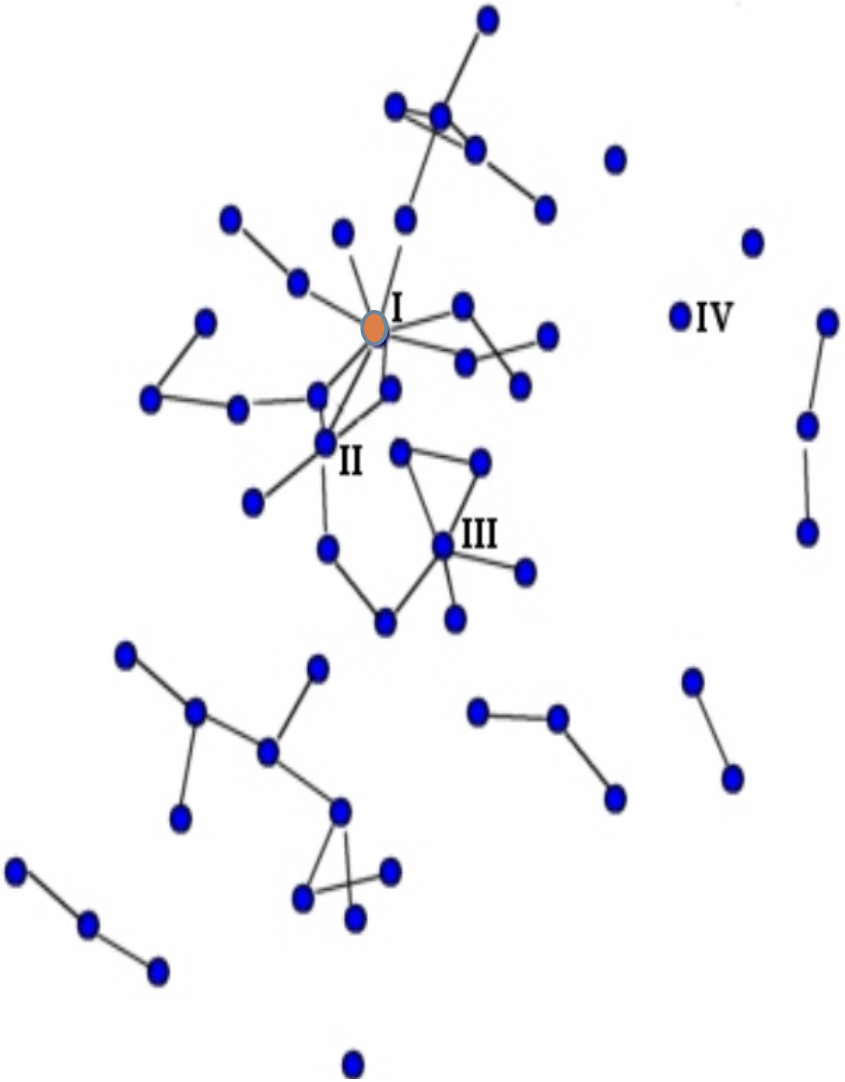
- A node N's closeness is defined as the sum of its distance to other nodes
- The most central node in terms of closeness is the node with the lowest value for this metric
- Note that strengths can be used as a distance measure for calculating closeness
  - Higher strength = closer nodes

# Betweenness

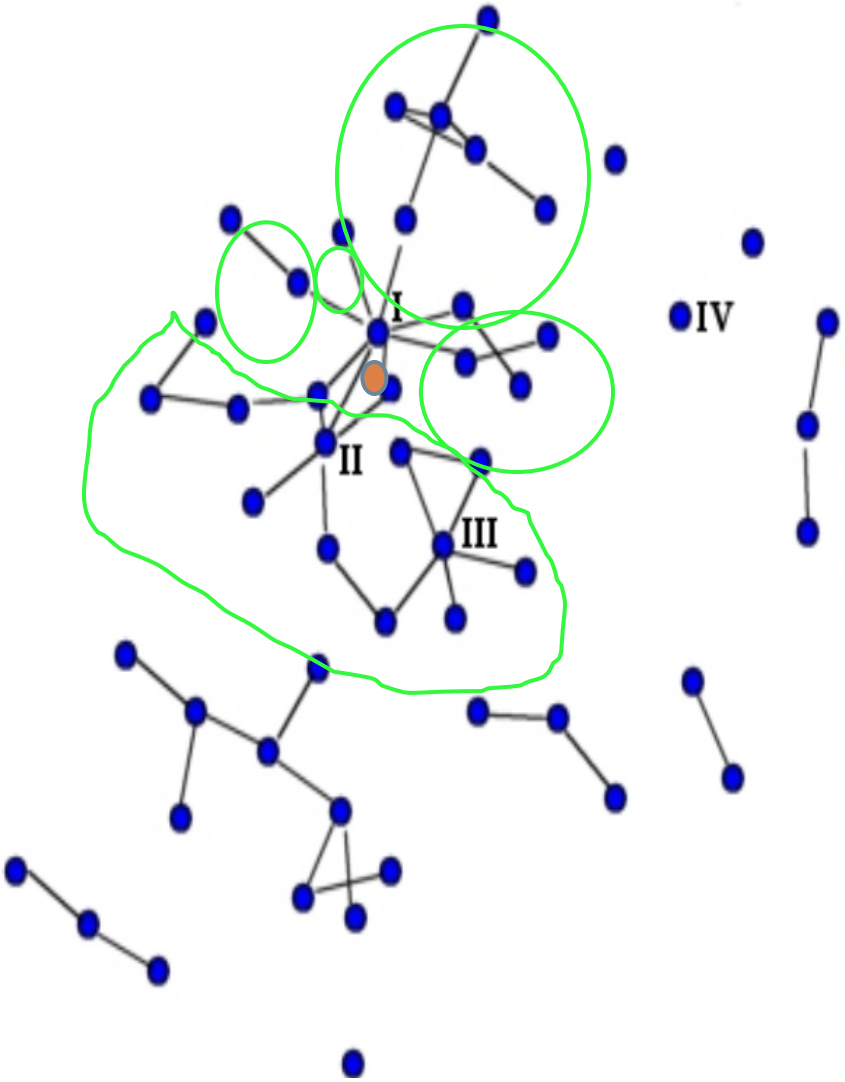
- Betweeness centrality for node N is computed as:

- The percent of cases where
- For each pair of nodes M and P (which are not N)
  - The shortest path from M to P passes through N

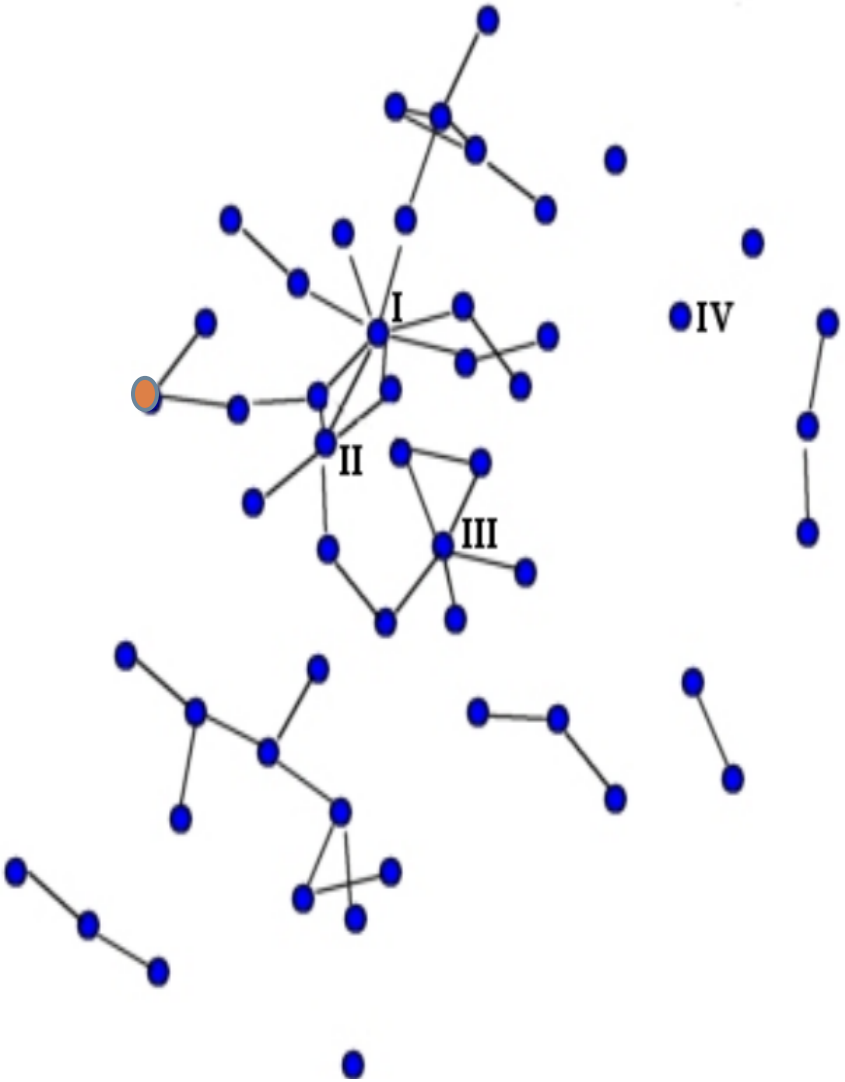
# What is this node's betweenness



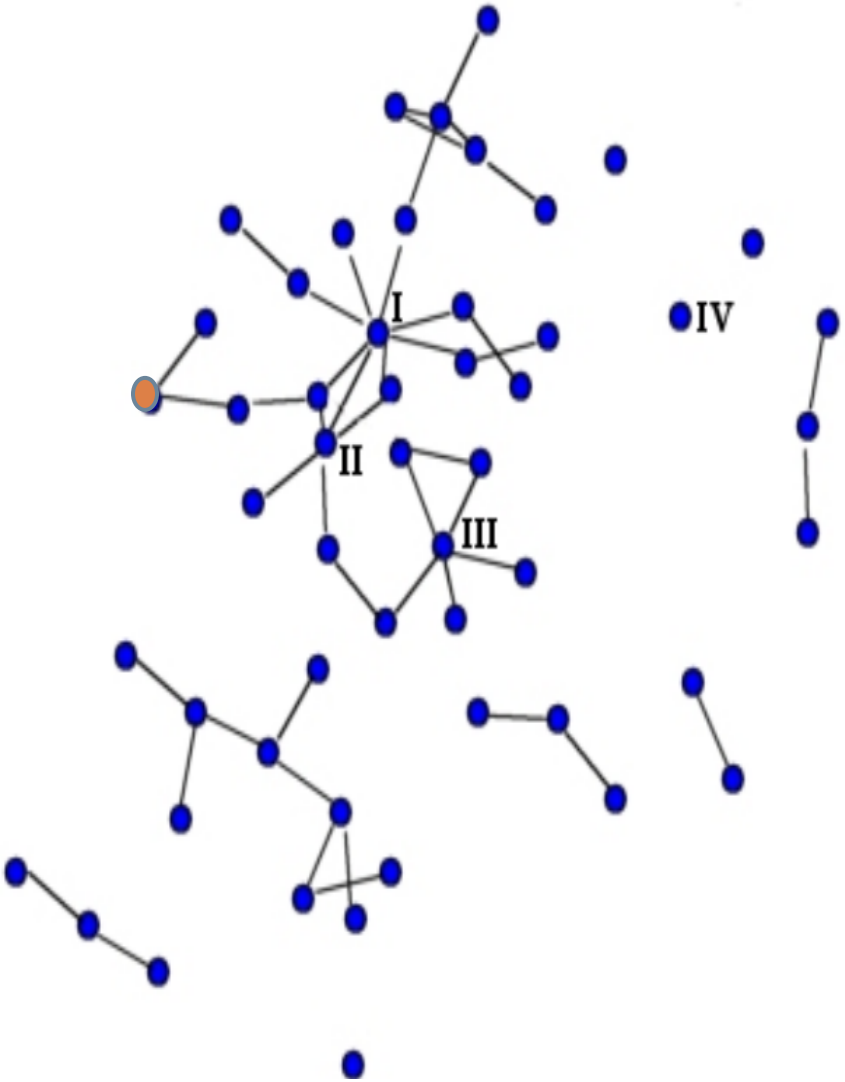
Betweenness is high;
each group can only get to other groups through this point



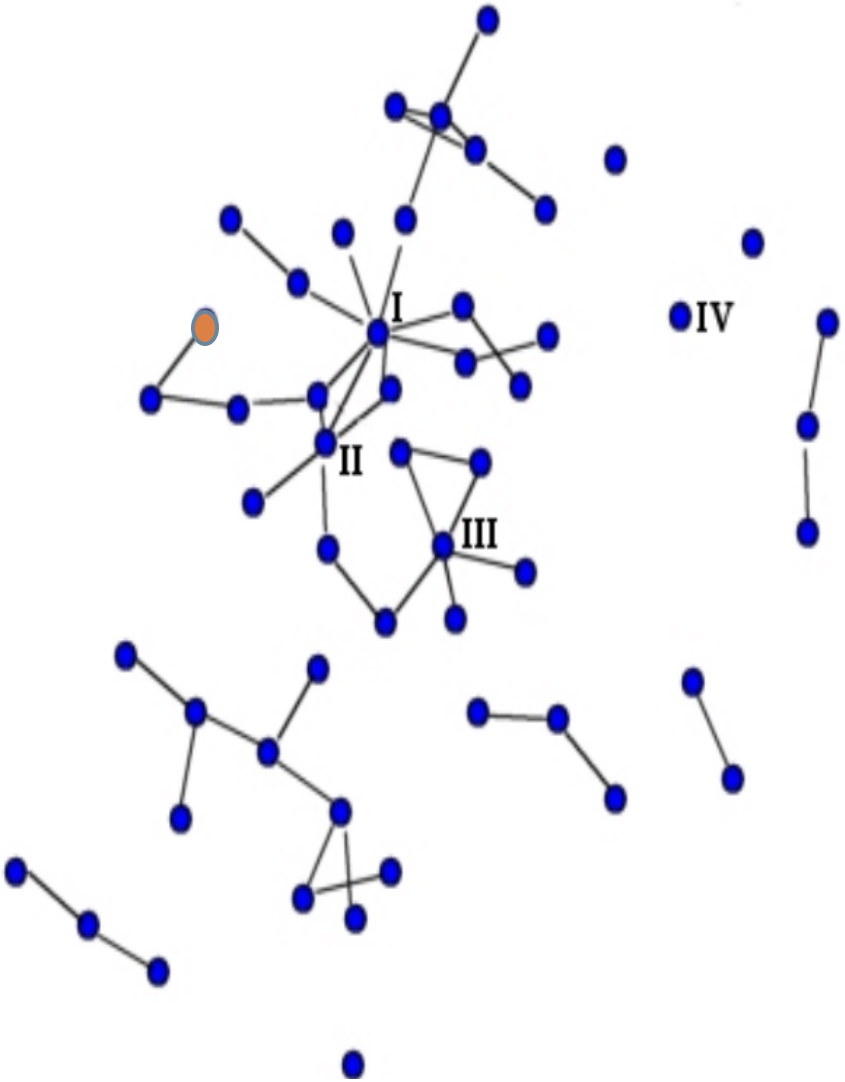
# What is this node's betweenness?



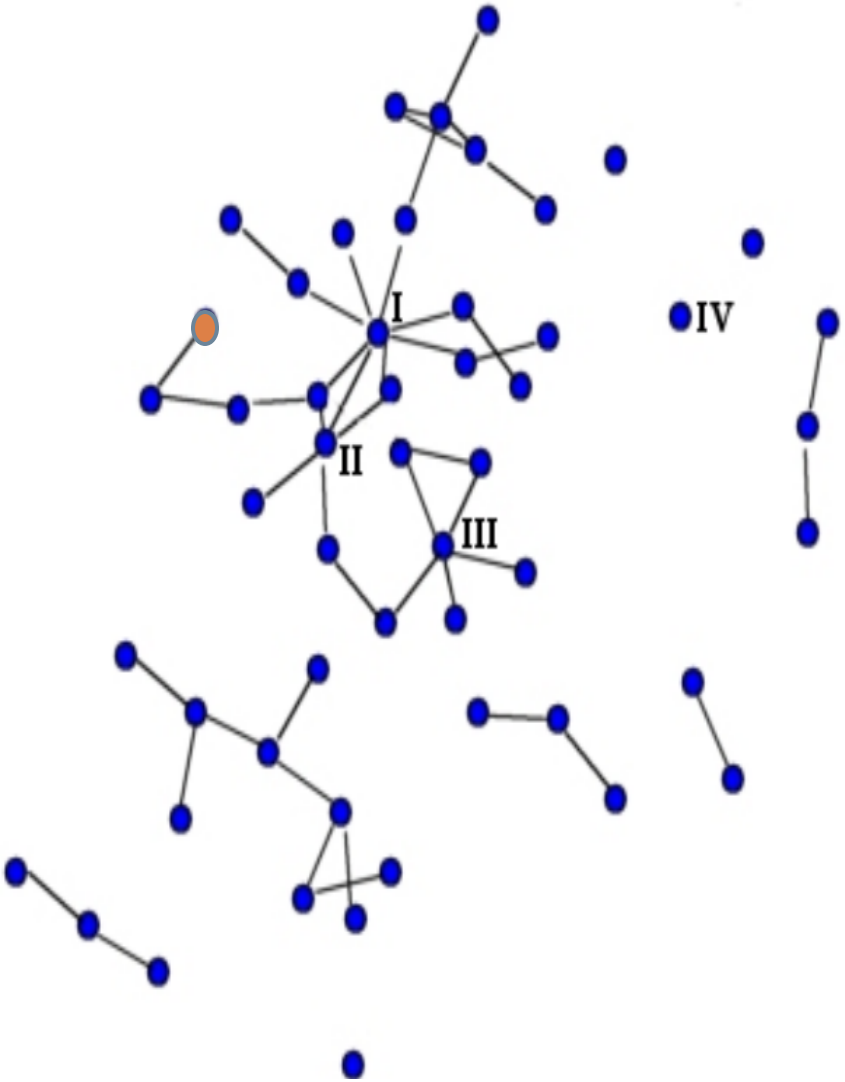
# Low, only one point connects through it



# What is this node's betweenness?



# Betweenness = 0



# Reciprocity

- What percentage of ties are bi-directional?
  - Can be computed as number of bi-directional ties over total number of connected pairs

# Eigenvector Centrality

- Complex math, but assigns centrality to nodes through recursive process where
- More and stronger connections are positive
- Connections to nodes with higher eigenvector centrality contribute more than connections to nodes with lower eigenvector centrality

# Eigenvector Centrality

- A key part of the original PageRank in Google

# Lots of uses

- There are lots of uses for network analysis

- But particularly useful for studying collaboration
  - Group-based learning
  - Teacher collaboration
  - Networks of influence
    - Why do some educational interventions seem to be dominant in specific regions?

# Next lecture: Epistemic Networks