Business Analytics and Emerging Trends

Instructor: Dr. Murat Tunc

Social Network Analysis

Instructor: Dr. Murat Tunc

Module 6

December 1st, 2021

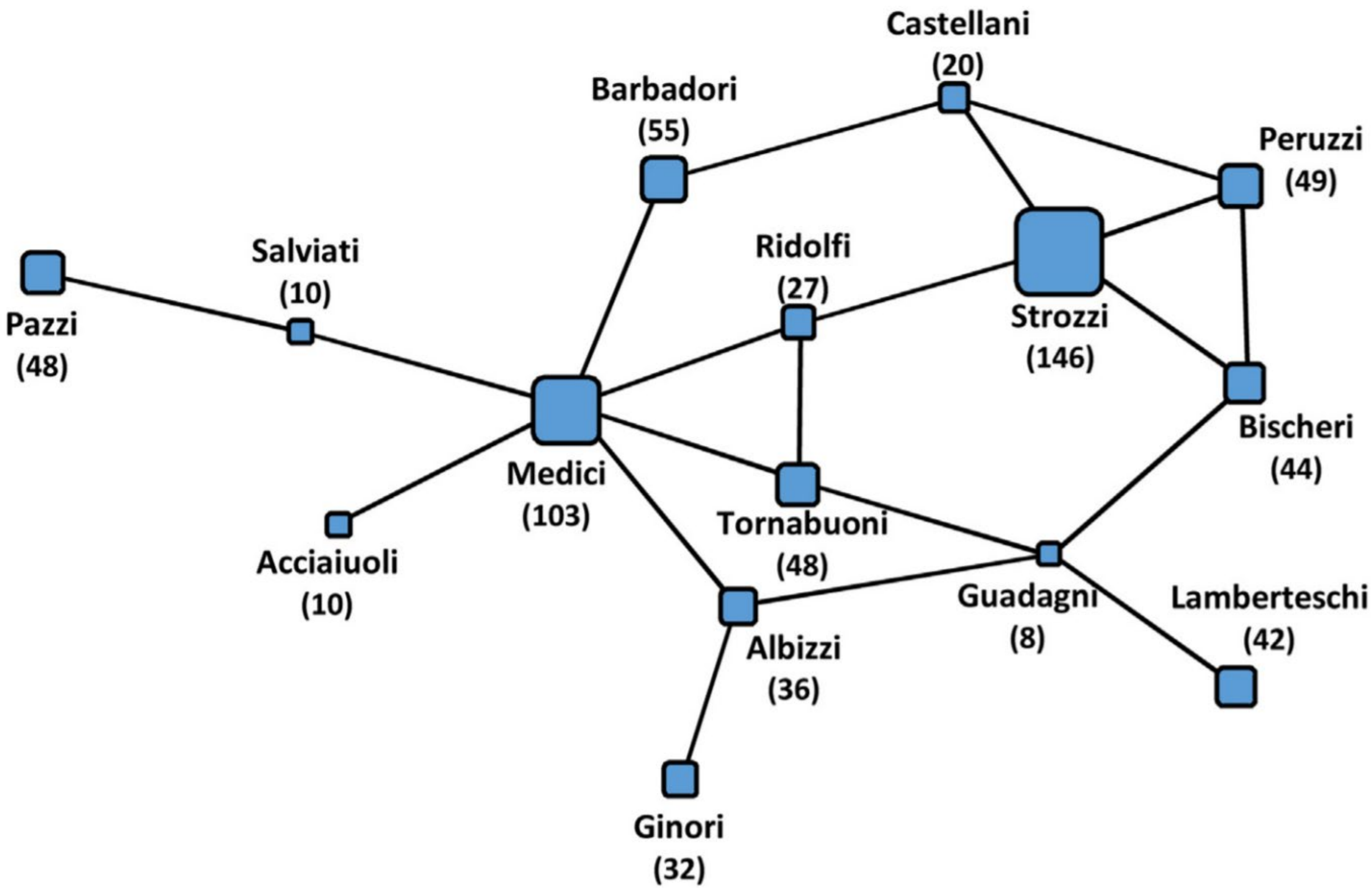
Why analyze social networks?

- Network influences **behavior**
- Labor markets
  - **Textile** workers
    - 62% found their **first job through a contact**
    - 23% by direct applications
  - Chicago labor market: First job through a contact?
    - 37% of typists
    - 65% of janitors
    - 23% of accountants
    - 57% of electricians
    - 73% of material handlers

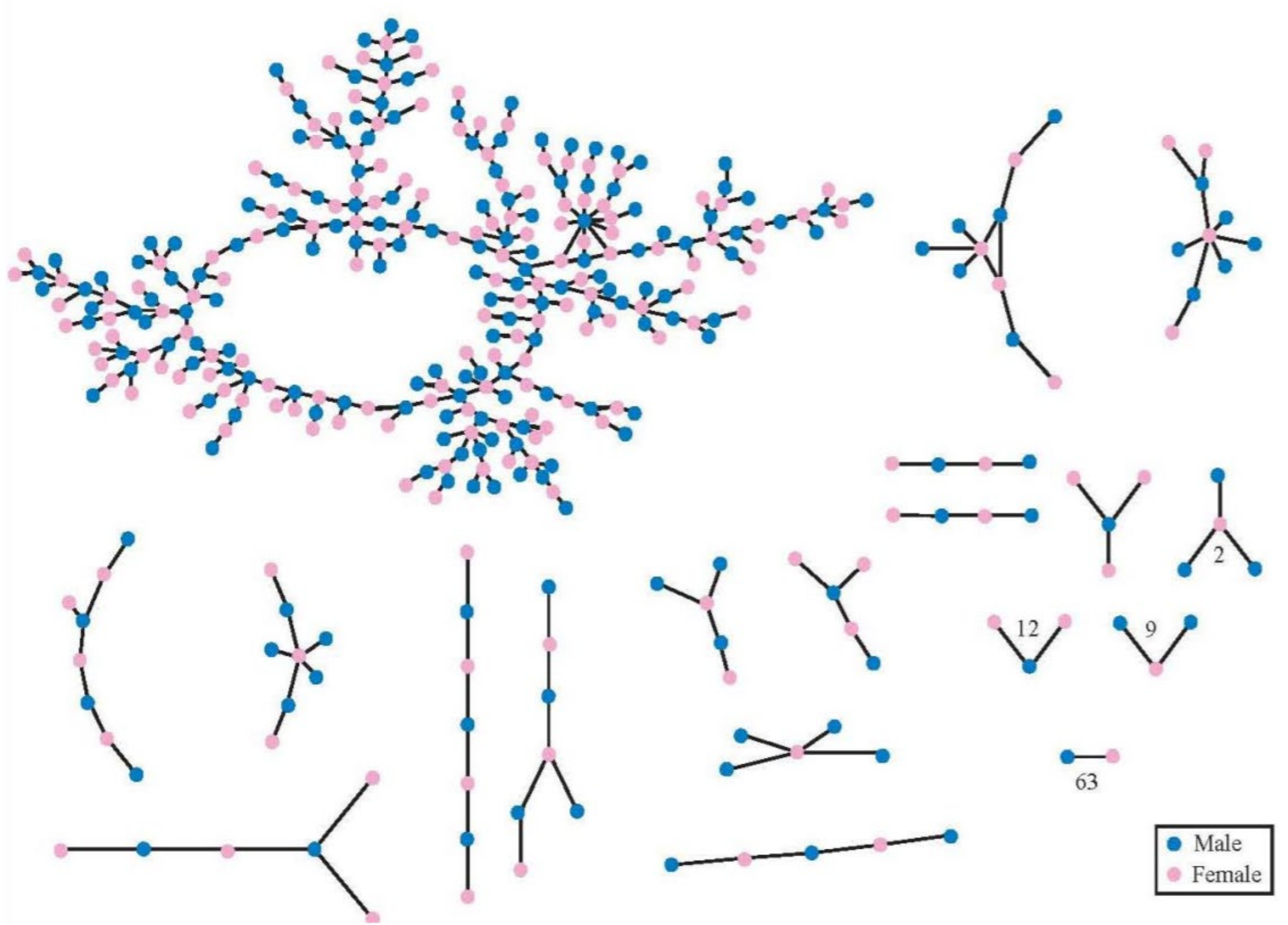
Why analyze social networks?

- If a person had **a friend who became obese**
  - Chances of becoming obese increased by 57%
- Social interactions in youth
  - Determine who are **likely to commit a crime**
- Diffusion of new products
  - Prescription of a new drug
    - Depends on doctor networks
  - How long it takes to become a mainstream product?
- Aids transmission, immigration, voting behavior, epidemics…

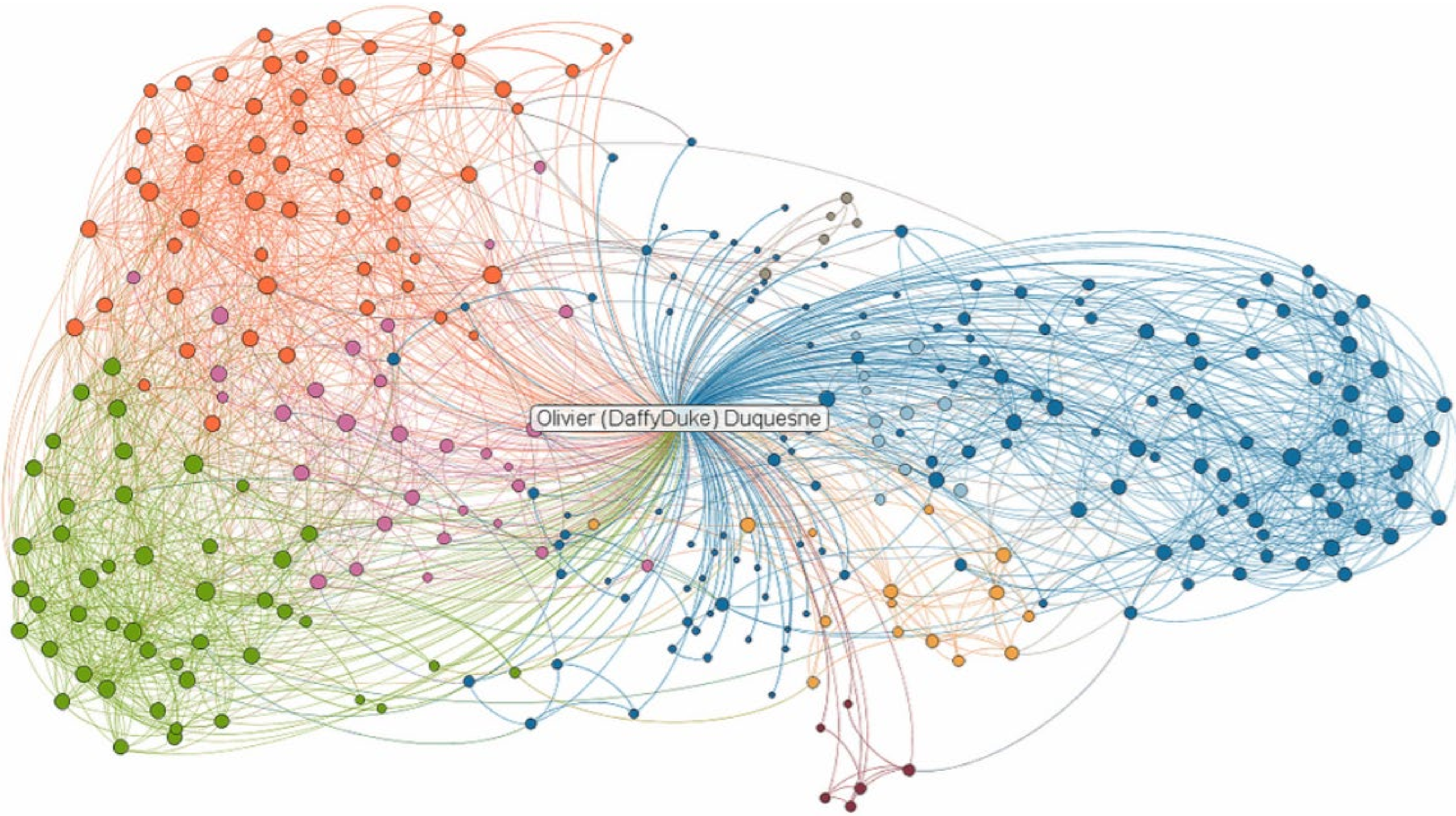
15th Century Florentine Marriages



Romantic relationships in a high school



Linkedin Friendship Maps



Analysis of Social Networks

Basic Definitions

Measures of Connectivity

Network Types

Information Flows

Basic Definitions

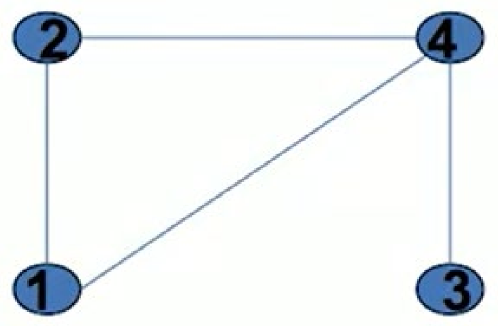
- **Nodes**, vertices
  - Objects, individuals, players
- **Edges**, links, ties
  - Connections between nodes
- **Weighted** edges
  - The intensity of a link
  - How many hours do two people spend together?
- **Unweighted** edges
  - 0 or 1
- **Directed** edges
  - One way relationship
- **Undirected** edges
  - Mutual relationships

Neighborhood and Degree

- **Degree** of a node
  - Number of edges connected to a node
- Neighborhood
  - Two nodes are **neighbors** if they share an edge in-between
- **Density** of a network
  - Average degree of all the nodes in the network
  - Only tells a partial story
- **Degree distribution** of a network
  - Explains more characteristics

Adjacency matrix

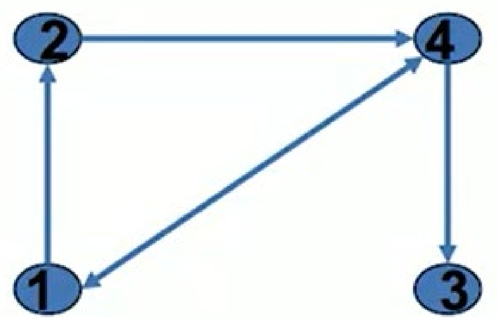
- An undirected network



Adjacency matrix

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

- A directed network



Adjacency matrix

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

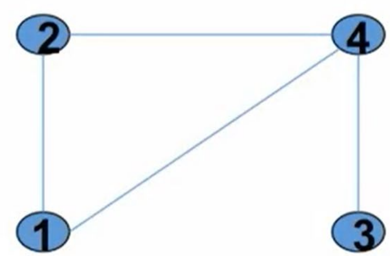
Walks and Paths

- Walk
  - Sequence of links connecting two nodes
- Cycle
  - A walk that starts and ends at the same node
- **Path**
  - A walk where a node appears at most once
- Geodesic
  - The **shortest path** between two nodes

Adjacency matrix and walks

- The n^{th} power of an adjacency matrix
  - Number of walks of length n

$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



$$g^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 3 \end{pmatrix}$$

number of walks of length 2 from i to j

Analysis of Social Networks

Basic Definitions

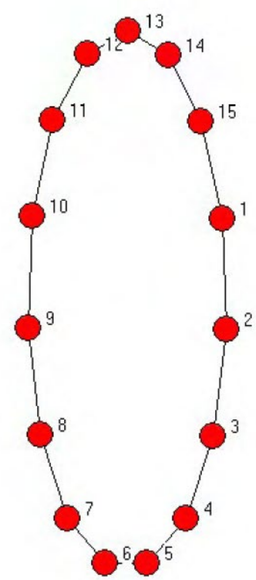
Measures of Connectivity

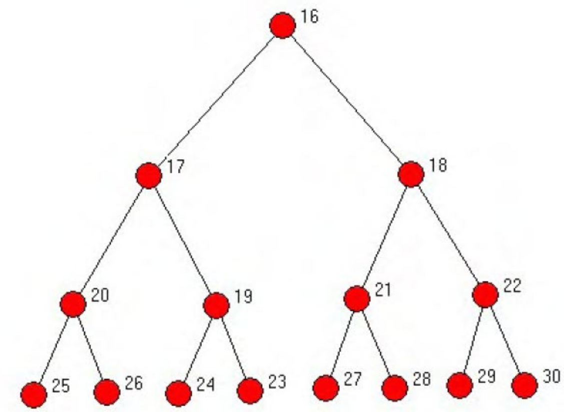
Network Types

Information Flows

Connectedness

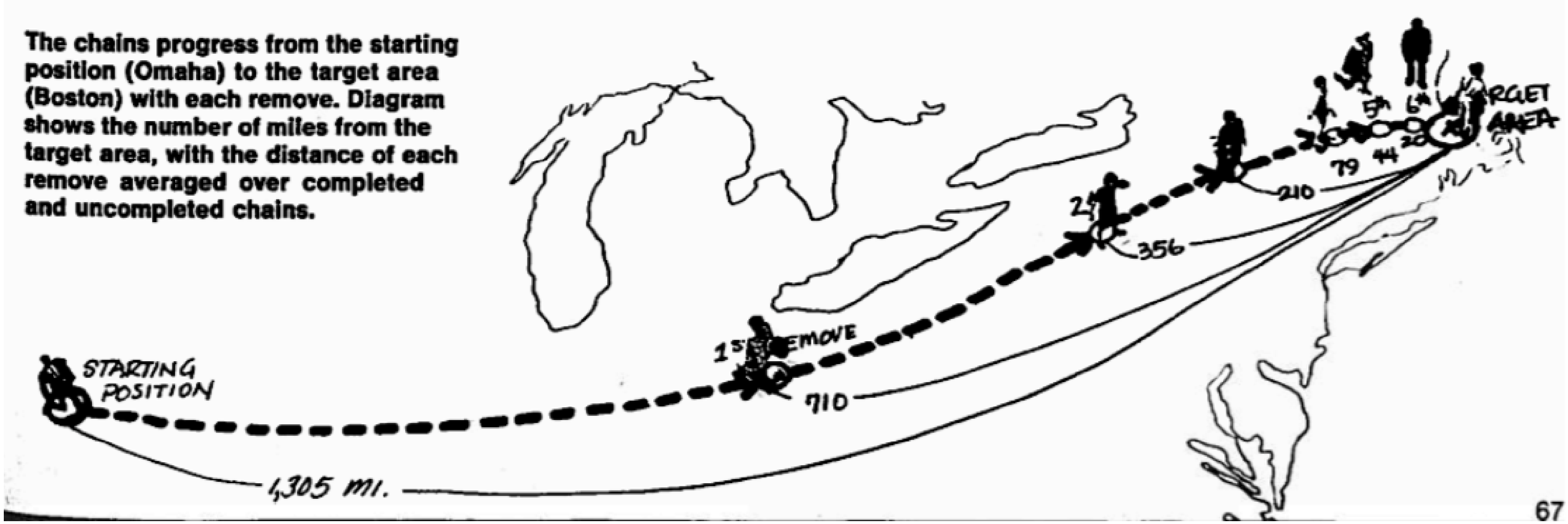
- A network is **connected** if
  - There is a path between **every** two nodes
- **Diameter** of a network
  - The largest geodesic (the maximum length of shortest paths)
- Average path length
  - What is the most likely distance between any nodes?





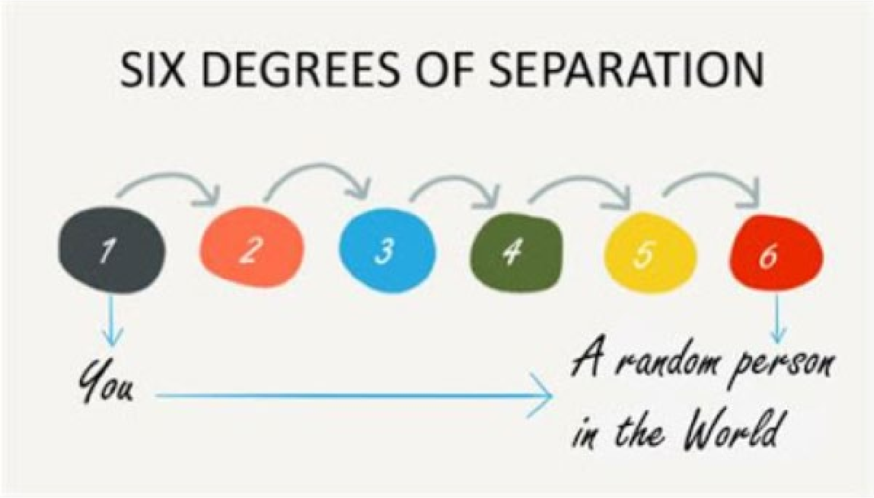
Average Path Length in real world

- Milgram (1967) – **letter experiment**
  - **Please send this letter to someone you know**
    - Starts from the Midwest (Nebraska)
    - Destination: An address in the Northeast (Massachusetts)
  - Median number of steps is 6 out of 25% of letters made it



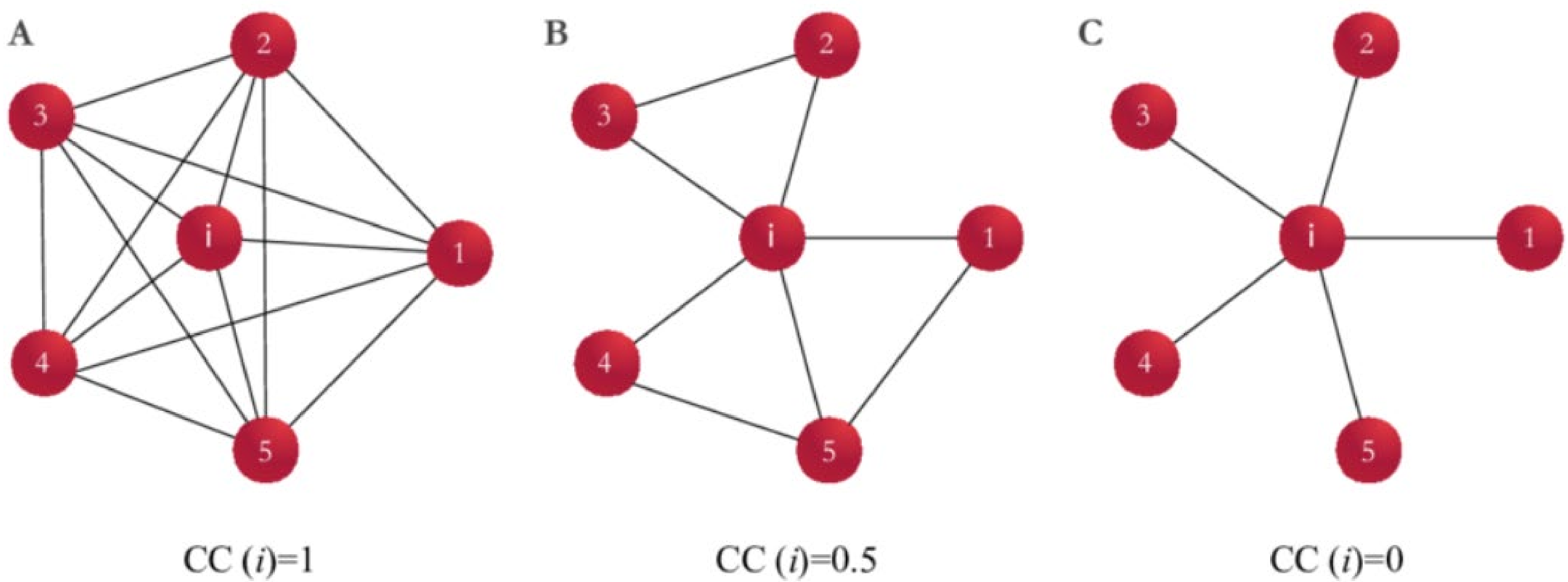
Six degrees of separation

- Academic **co-authorship** networks
  - Math: mean 7.6
  - Physics: mean 5.9
  - Economics: mean 9.5
- **Facebook friendship** network
  - Mean 4.74 (721 million users)



Clustering Coefficient

- How many of your friends know each other?

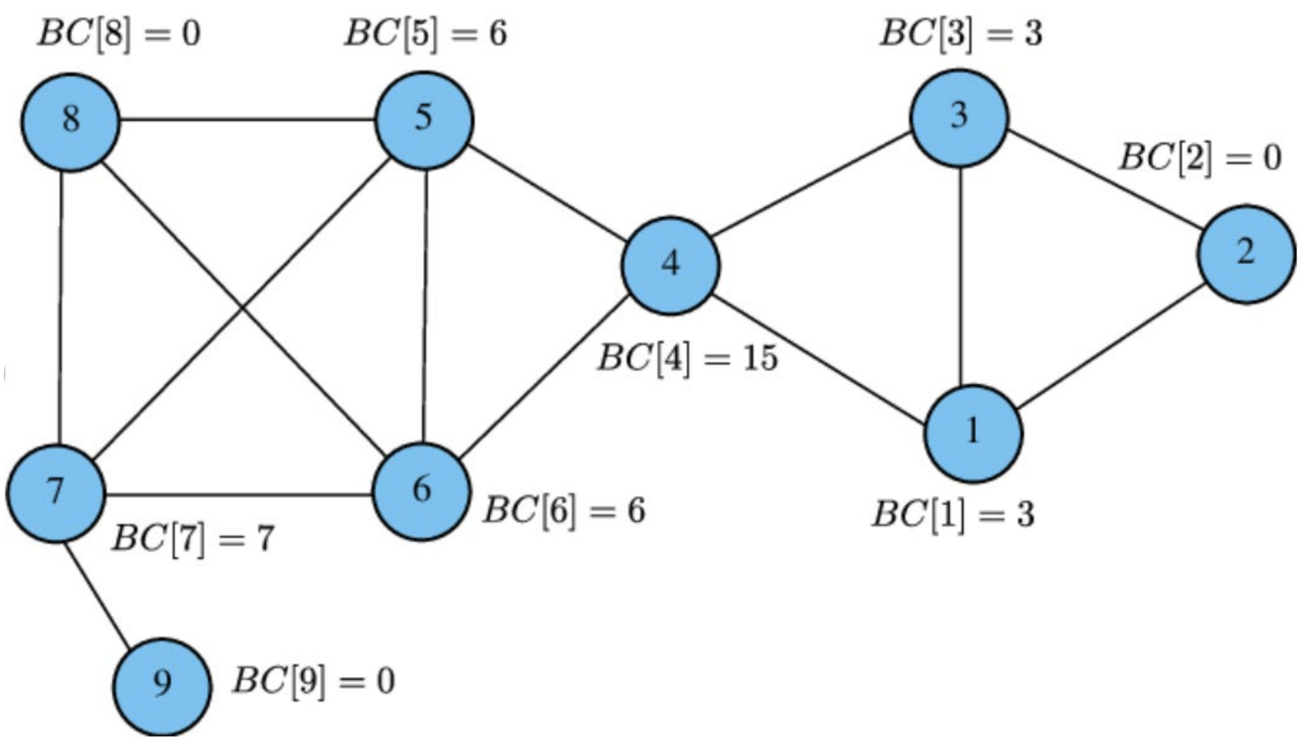


Clustering Coefficient in real world

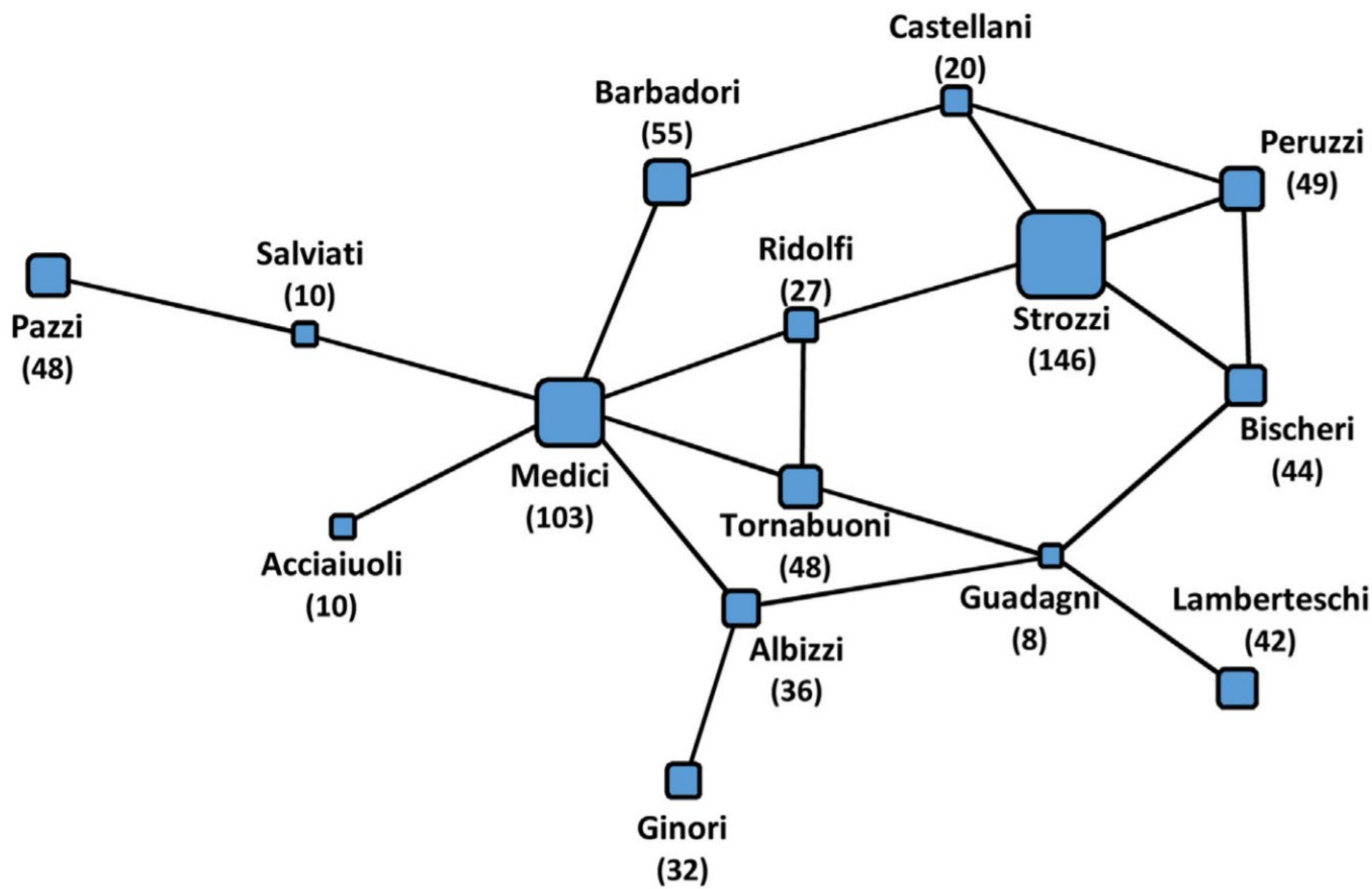
- **Prison friendship**
  - 0.31 (MacRae 1960)
- Academic **co-authorship networks**
  - 0.15 (Math)
  - 0.09 (Biology)
  - 0.19 (Econ)
- www
  - 0.11 (Web links)
- Real world networks are **highly clustered**

Betweenness Centrality

- Number of shortest path that **passes through** a node
  - Represents the influence of a node for information flows



Medici family became the wealthiest



Analysis of Social Networks

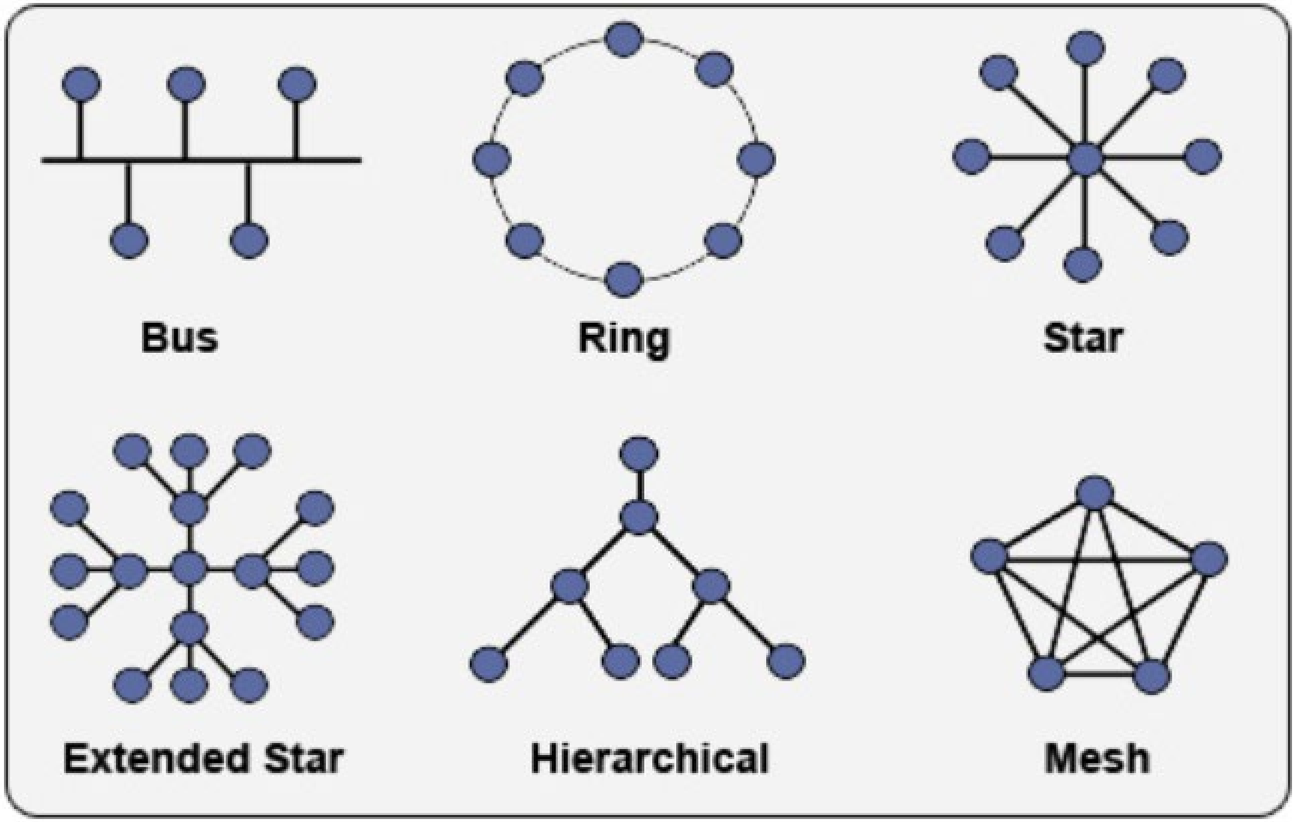
Basic Definitions

Measures of Connectivity

Network Types

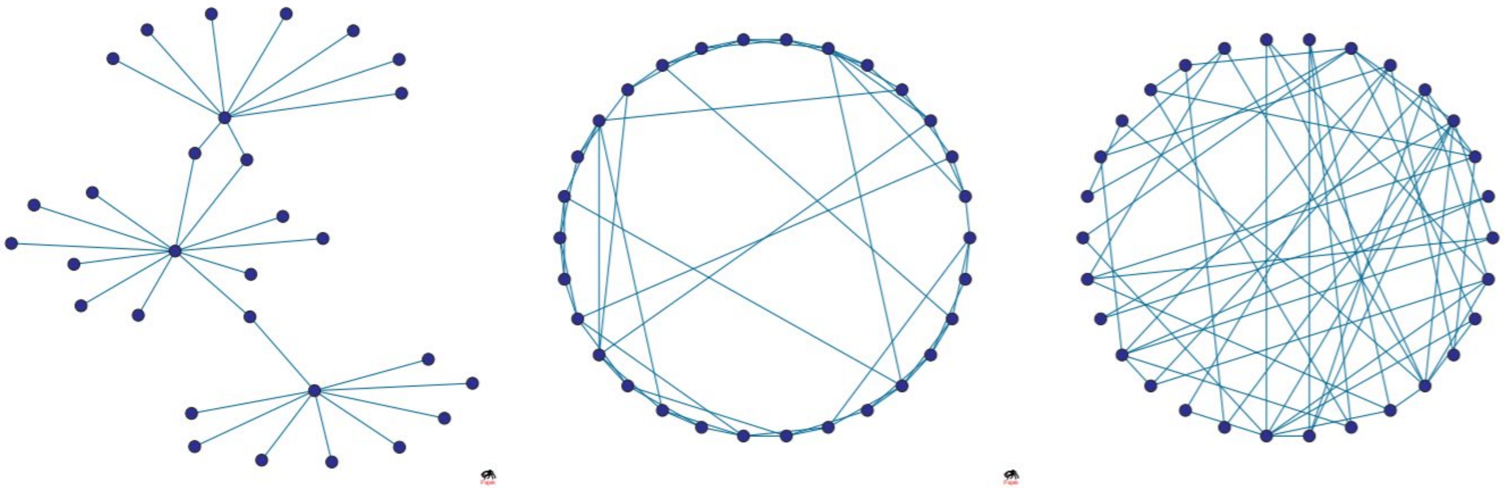
Information Flows

Network Topologies



Network Types to model real world cases

- Random networks
  - Erdos & Renyi 1959
- Small world networks
  - Watts & Strogatz 1998
- Scale free networks
  - Barabasi & Alert 1999



(a) Scale-free network (b) Small-world network (c) Erdős–Rényi network

Random Networks - Generation

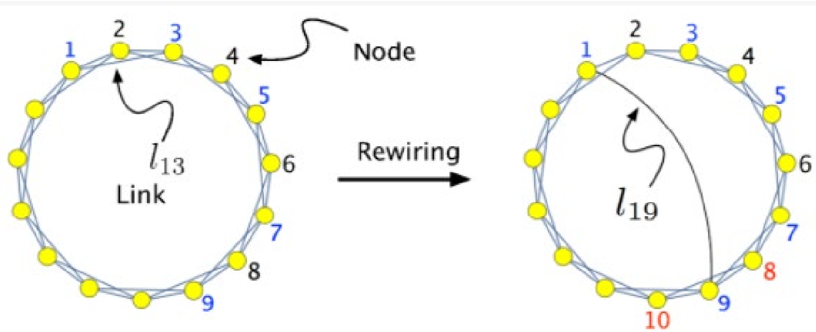
- Links between each node is random
  - With equal probability
- Start with **n** nodes
  - **Connect each pair** of nodes with a probability **p**
  - All nodes have approximately the **same degree**: **k**
- Random networks
  - # of nodes: n
  - Probability of an edge between any two nodes: p
  - **Notation:** G(n, p)

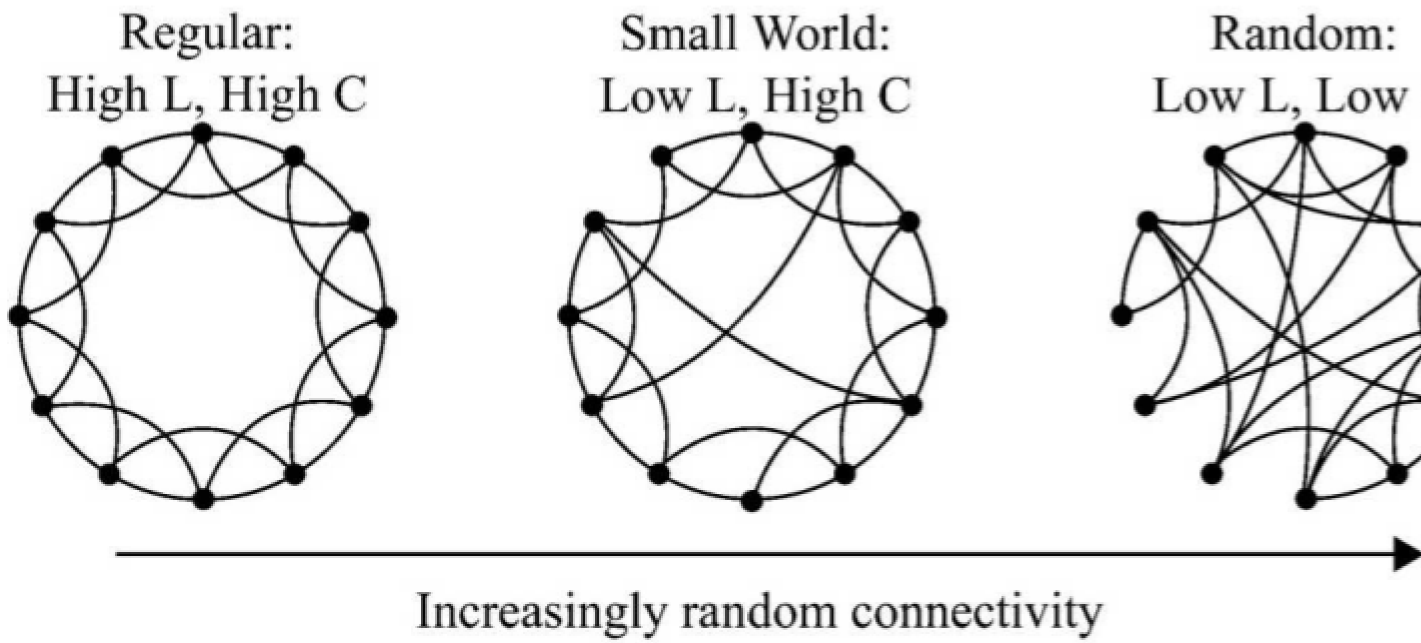
Random Networks - Properties

- Average path length: $\frac{\ln(n)}{\ln(k)}$
- Clustering coefficient: $\frac{k}{n} = p$
- Degree distribution:
    - Binomial distribution for small n
    - Poisson distribution for large n

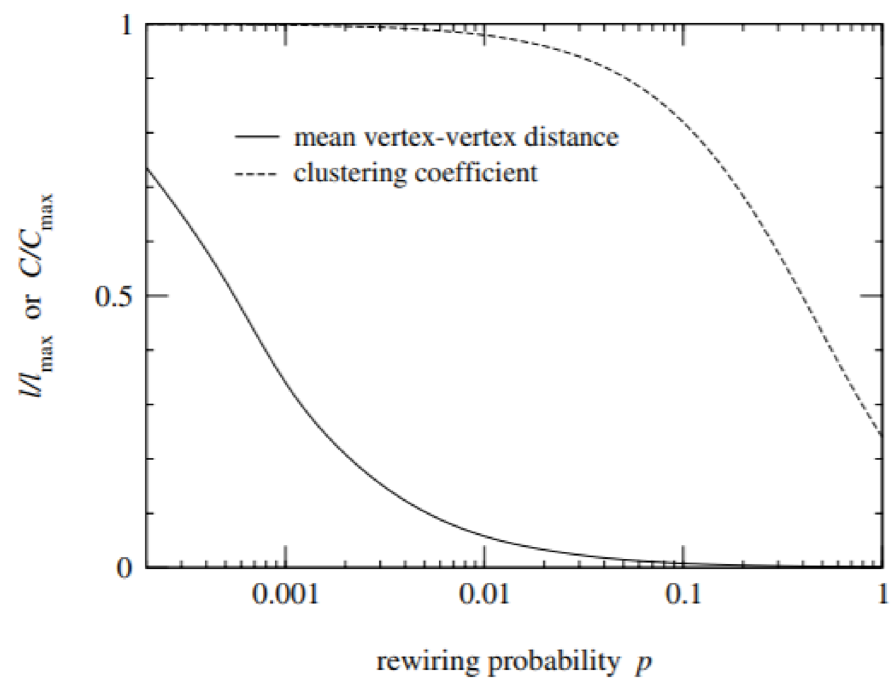
Small World Networks - Generation

- Similar to social networks in real world
  - A group of people are **closely related**
  - A few people have **far reaching connections**
- Generation
  - Start with a ring lattice of **n** nodes
  - Each node connected to its closest **k** neighbors
  - **Rewire the edges**
    - Delete edges with probability **p**
    - Create a random edge such that the number of links remain the same





Regular:
High L, High C
Small World:
Low L, High C
Random:
Low L, Low C
Increasingly random connectivity



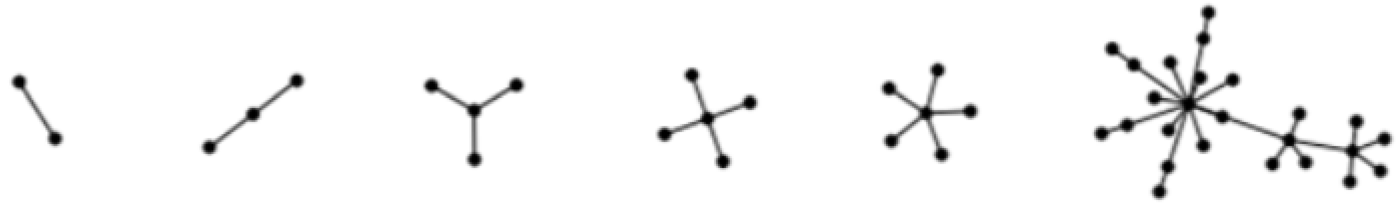
mean vertex-vertex distance
clustering coefficient
l/l_{max} or C/C_{max}
rewiring probability p
1
0.5
0
0.001
0.01
0.1
1

Small World Network - Properties

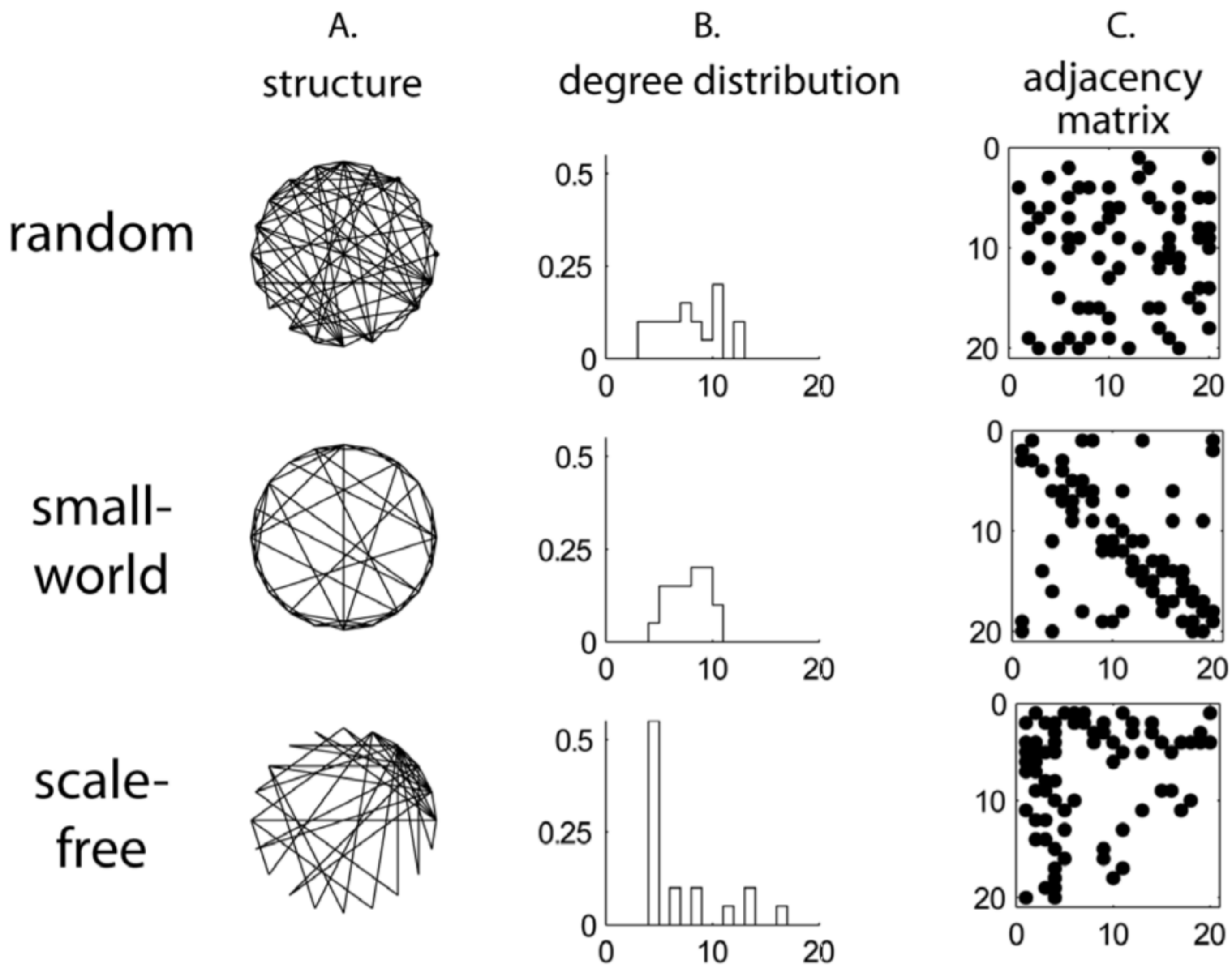
- Average path length:
  - Proportional to $\ln(n)$
- Clustering coefficient:
  - **Highly clustered**
  - Compared to random networks, $CC_{SW} \gg CC_{RN}$
- Degree distribution
  - Similar to random networks
  - Binomial distribution for small n
  - Poisson distribution for large n

Scale Free Networks - Generation

- Power-law network
  - Hub networks
- Preferential attachment
  - Start with 1 node
  - Add a new node via a link
  - **Which node** will be linked in the existing network?
  - **Depends on the degree** of the node
  - Rich get richer phenomenon



Comparison of the 3 Network Types



D.
Description

73 connections among 20 nodes assigned randomly

High local clustering and short average path lengths. 'Hub-and-spoke' architecture.

'Hub-and-spoke' architecture is maintained at multiple spatial scales.

Analysis of Social Networks

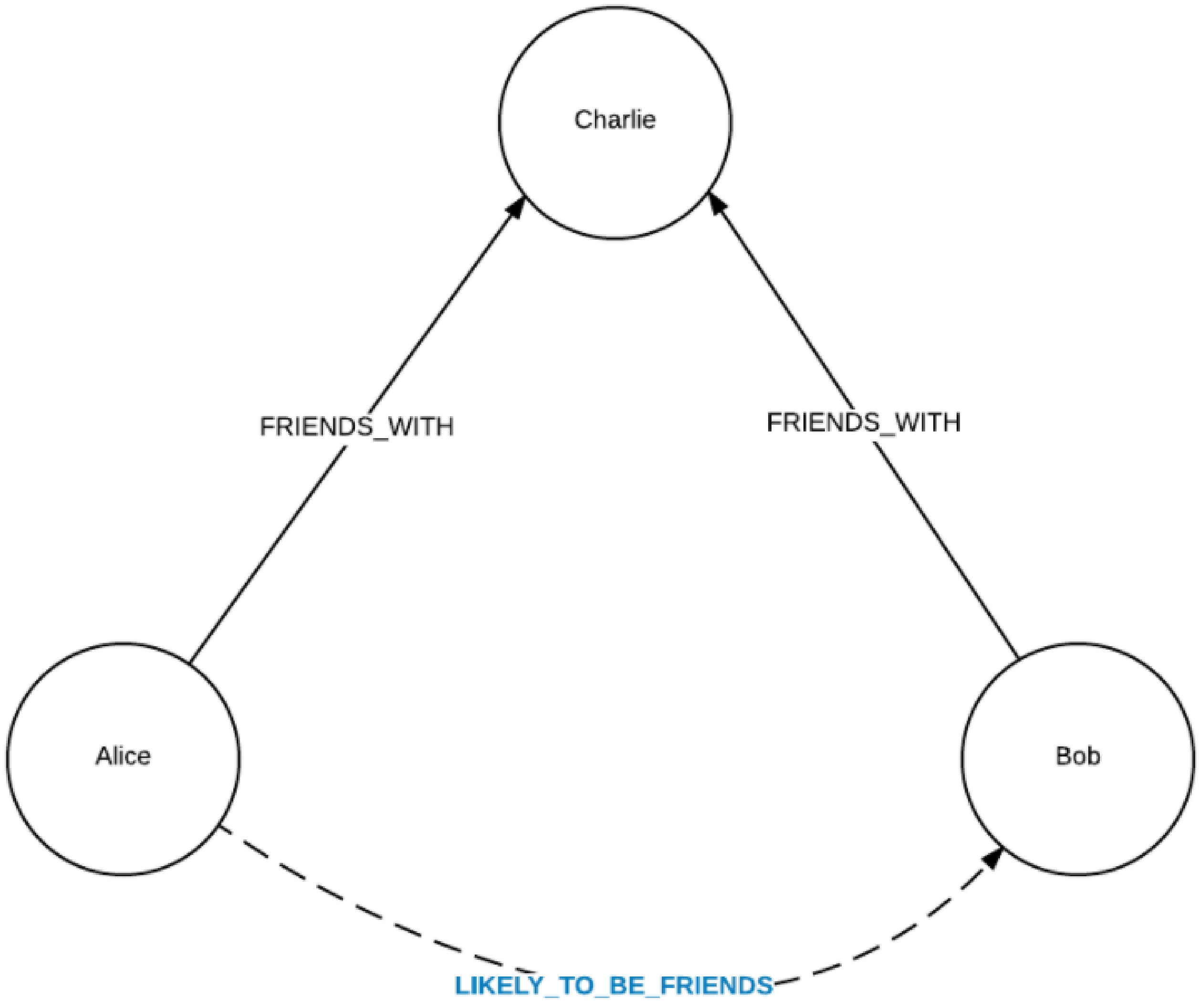
Basic Definitions

Measures of Connectivity

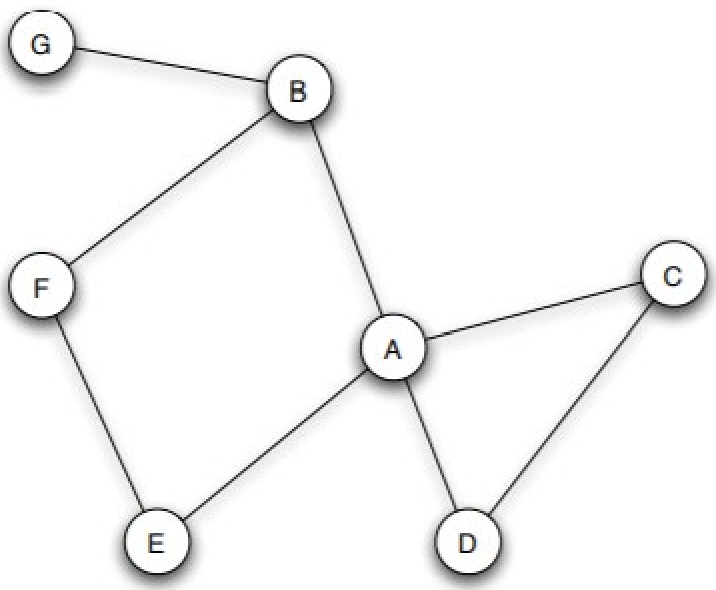
Network Types

Information Flows

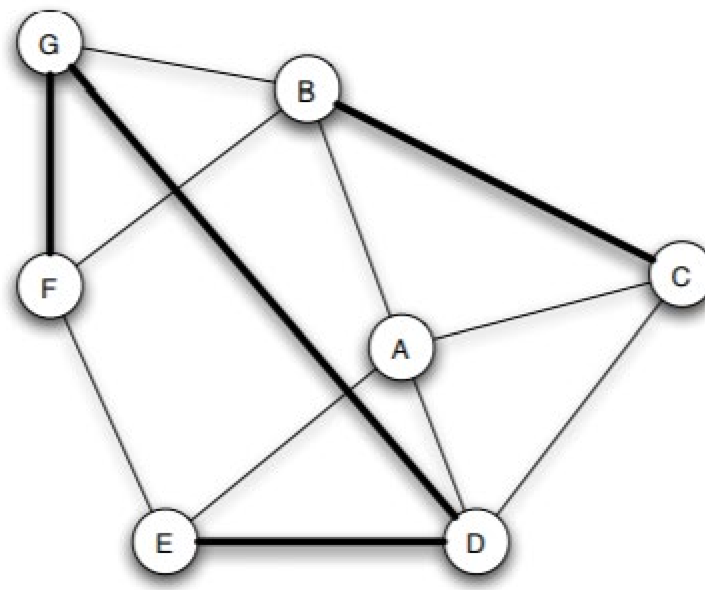
Triadic Closure



Evolution of Networks via Triadic Closure

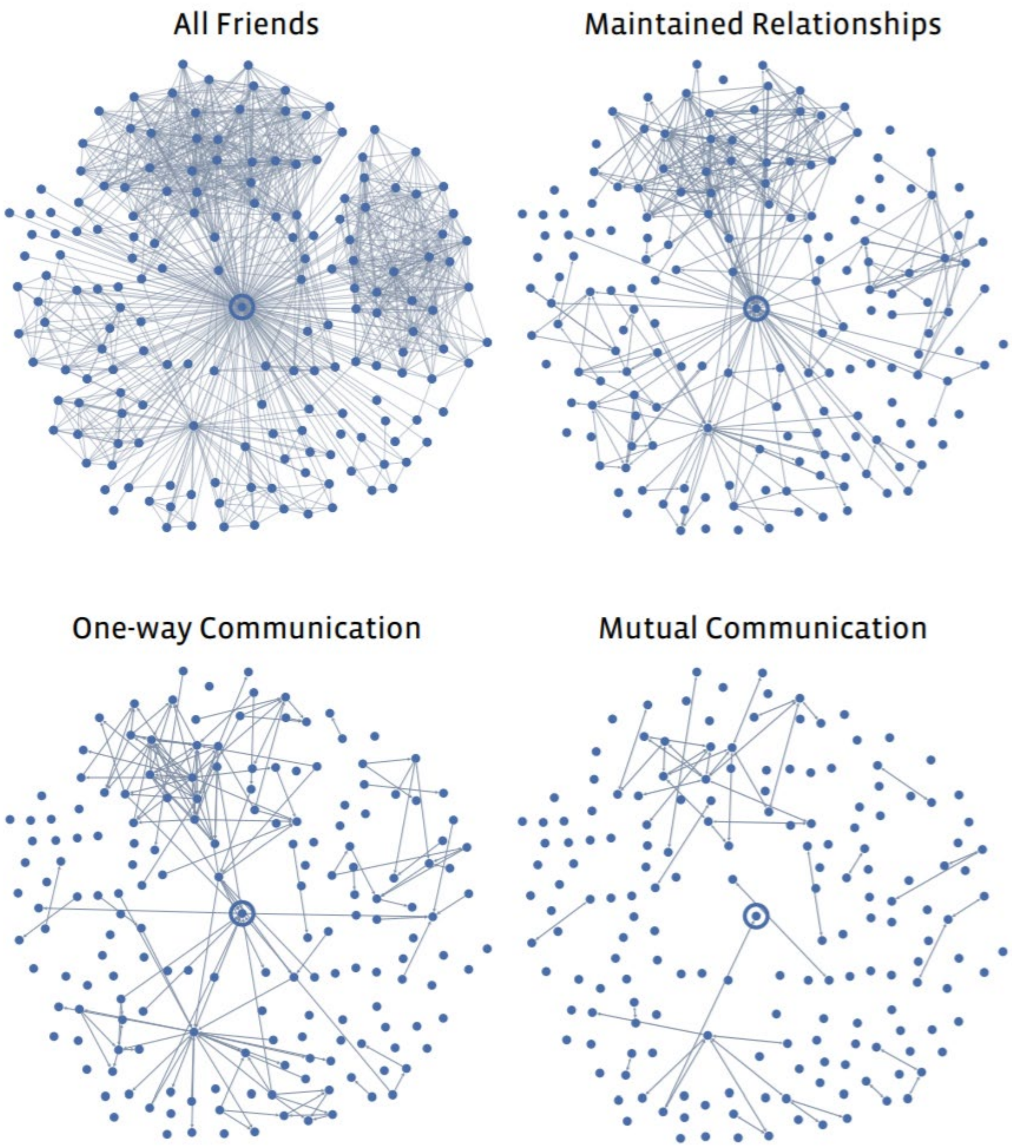


(a) *Before new edges form.*

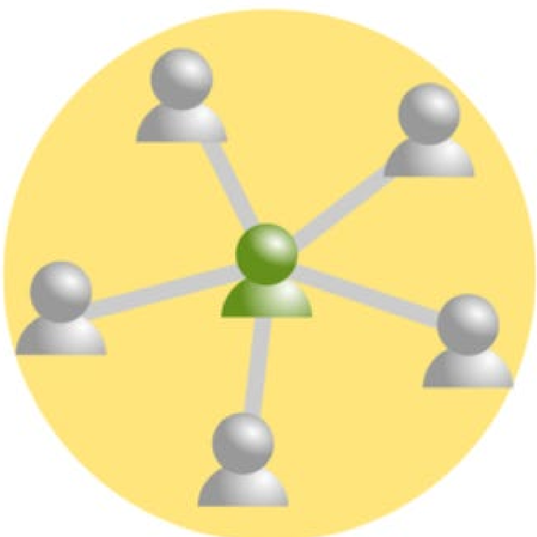


(b) *After new edges form.*

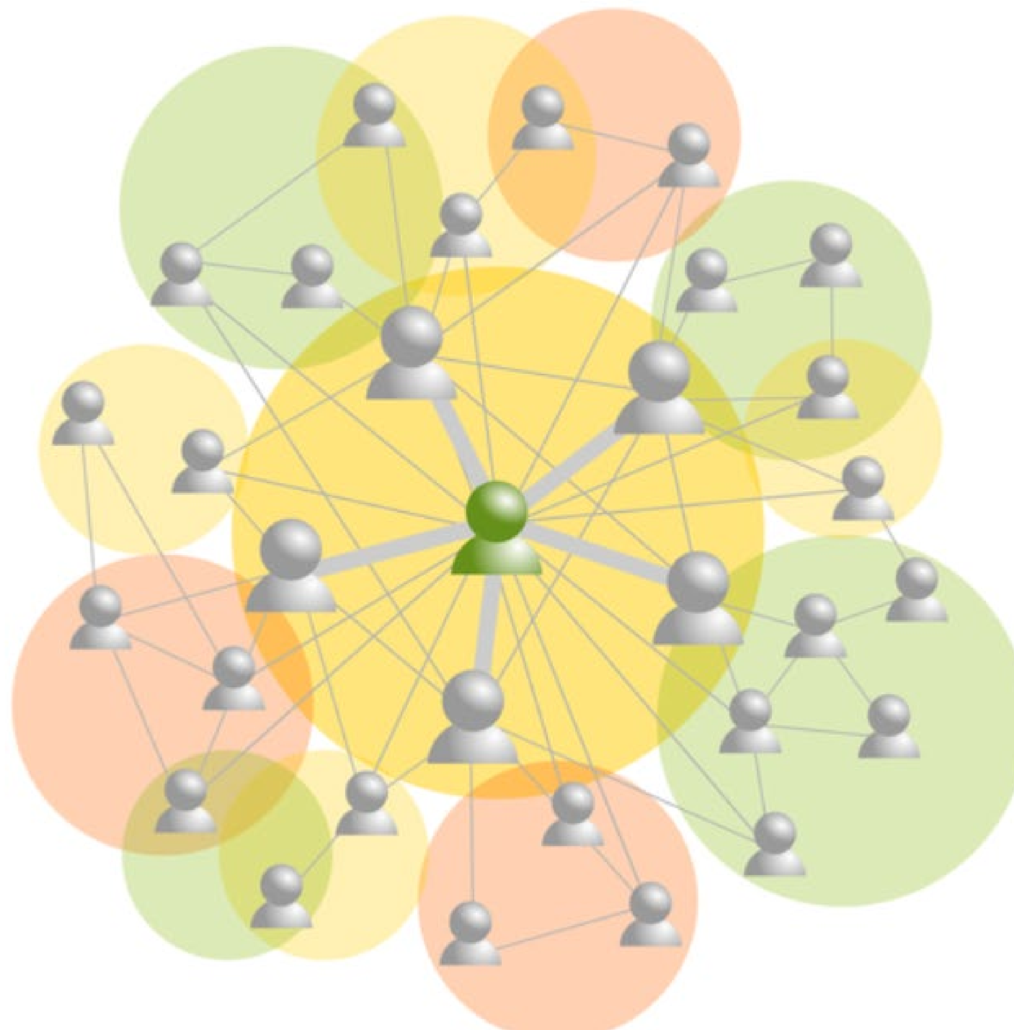
Tie Strength – Facebook users network



Strength of weak ties



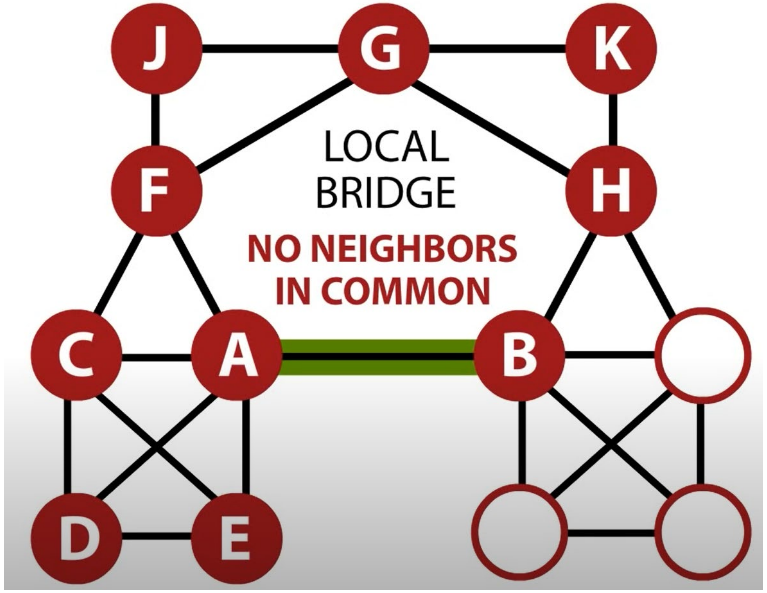
CONNECTIONS THROUGH STRONG TIES



CONNECTIONS THROUGH WEAK TIES

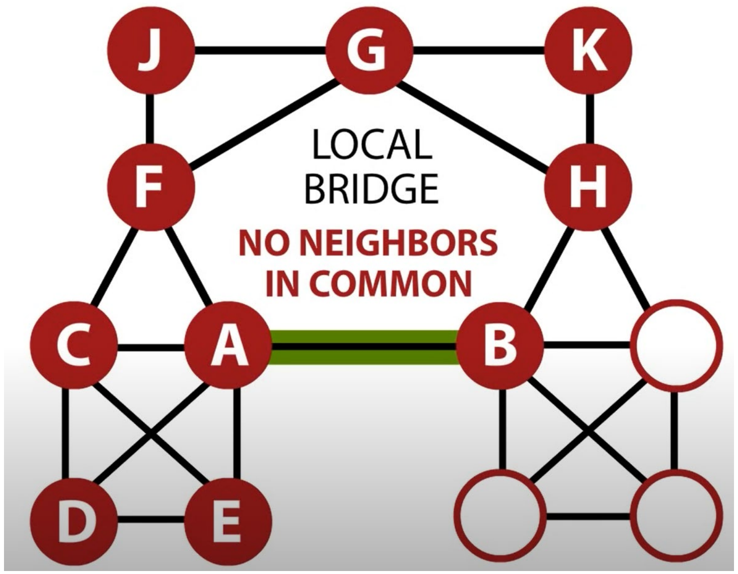
Strong Ties and Weak Ties

- Strong ties
  - Close friendship
  - Your **friends are also friends** with each other
- Weak ties
  - A distant friend
  - Your **friends do not know each other**

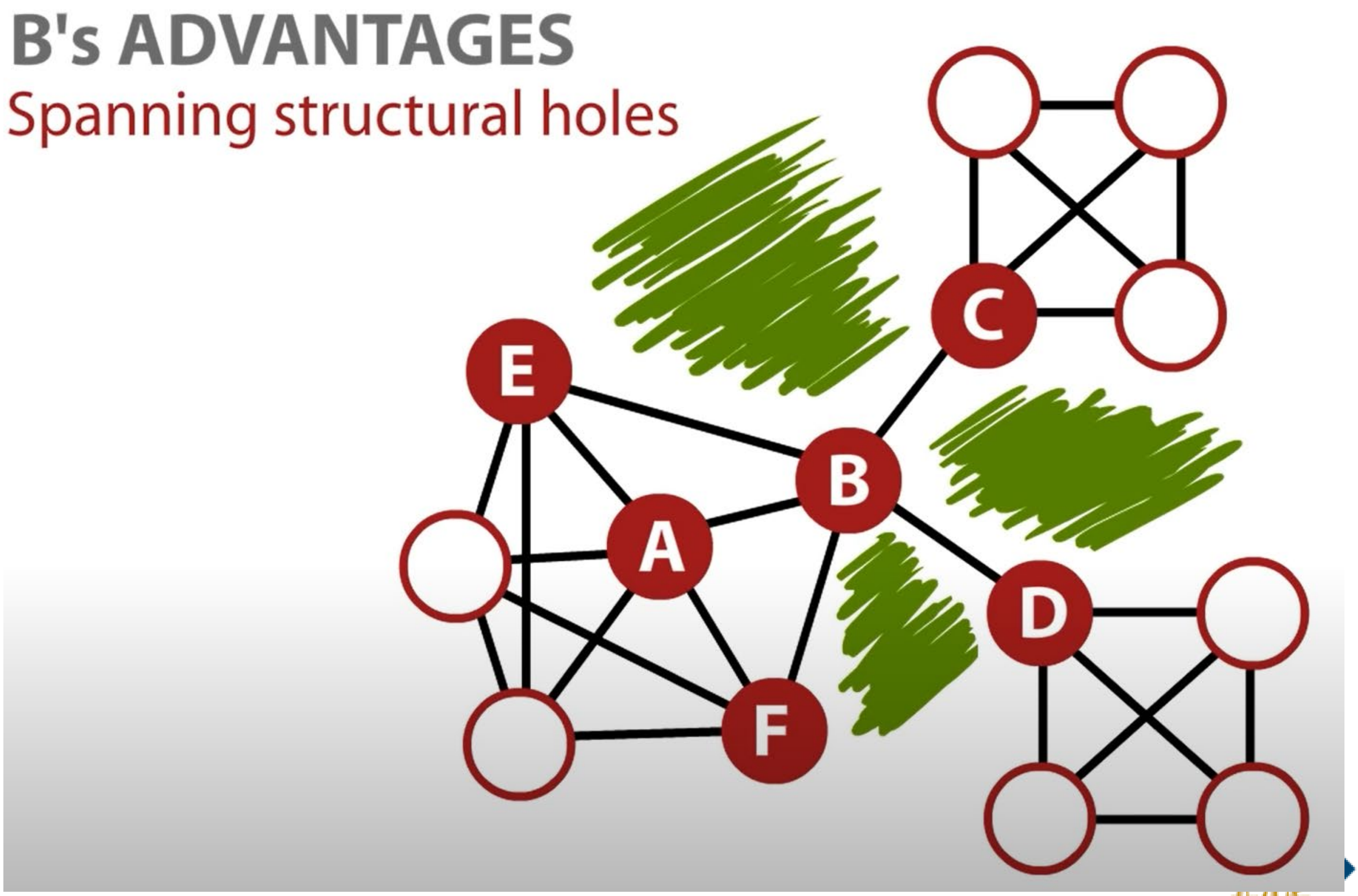


Strength of Weak Ties

- A is looking for a job
  - C, D, E and F are close friends and want to help A
  - But **what they know is similar to what A knows**
- B has access to a bunch of information
  - That A cannot directly perceive
- Job leads, novel information, etc.



Structural Holes



Structural Holes and Good Ideas

- Burt (2004)
  - People **connected across groups**
    - More familiar with **alternative ways** of thinking
    - More options to select and **synthesize**
- Nodes spanning structural holes
  - Information advantage
    - Information across groups are **more additive than overlapping**
  - Control advantage
    - Third-party opportunities
    - Brokerage
    - Entrepreneurship

Readings

- Granovetter, M. S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360-1380.
- Burt, R. S. (2004). Structural holes and good ideas. American journal of sociology, 110(2), 349-399.

References

- Charles Andrew Myers and George P Shultz. The Dynamics of a Labor Market: A Study of the Impact of Employment Changes on Labor Mobility, Job Satisfactions, and Company and Union Policies. Prentice-Hall, New York, 1951.
- Albert Rees and George Schultz. Workers in an Urban Labor Market. University Of Chicago Press, Chicago, 1970.
- Perera, S., Bell, M. G., & Bliemer, M. C. (2017). Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. Applied network science, 2(1), 33.
- Agneessens, F., Borgatti, S. P., & Everett, M. G. (2017). Geodesic based centrality: Unifying the local and the global. Social Networks, 49, 12-26.
- Stobb, M., Peterson, J. M., Mazzag, B., & Gahtan, E. (2012). Graph theoretical model of a sensorimotor connectome in zebrafish. PLoS One, 7(5), e37292.
- Cuadra, L., Pino, M. D., Nieto-Borge, J. C., & Salcedo-Sanz, S. (2017). Optimizing the structure of distribution smart grids with renewable generation against abnormal conditions: A complex networks approach with evolutionary algorithms. Energies, 10(8), 1097.

References

- David Easley and Jon Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010.
- Matthew O Jackson and Leeat Yariv. Diffusion, strategic interaction, and social structure. Handbook of Social Economics, 1:645–678, 2011.
- Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. Nature, 393(6684):440–442, 1998.
- Paul Erdos and Alfréd Rényi. On random graphs. Publicationes Mathematicae Debrecen, 6:290–297, 1959.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. Science, 286(5439):509–512, 1999.
- Tunç, M. M. (2015). Diffusion of innovation and collective action in complex networks (Master dissertation).