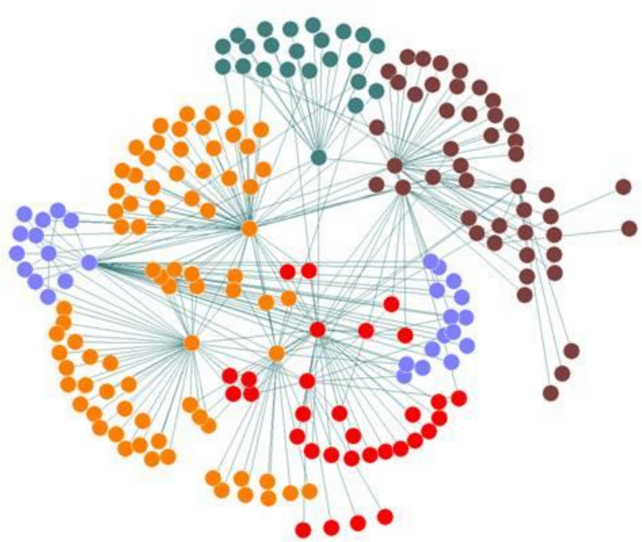
TEL AVIV UNIVERSITY

Algorithms and Applications in Social Networks



2019/2020, Semester B

Slava Novgorodov

Lesson #2

- Random network models
- Centrality measures

Random Graphs

Erdős–Rényi model

- Two variants of the model:
  - G(n, m) – a graph is chosen uniformly from a set of graphs with **n** nodes and **m** edges
  - G(n, p) – a graph is constructed on **n** nodes, with probability of edge equals to **p**
- We will focus on the second variant
- Expected number of edges and average degree:

$$\overline{m} = \frac{n(n-1)}{2} p \qquad \overline{k} = \frac{1}{n} \sum_i k_i = \frac{2\overline{m}}{n} = p(n-1)$$

Erdős–Rényi model

- Probability of node **i** having a degree **k**:

$$P(k_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

- Binomial distribution, which becomes Poisson when n infinity

$$\lambda = pn$$

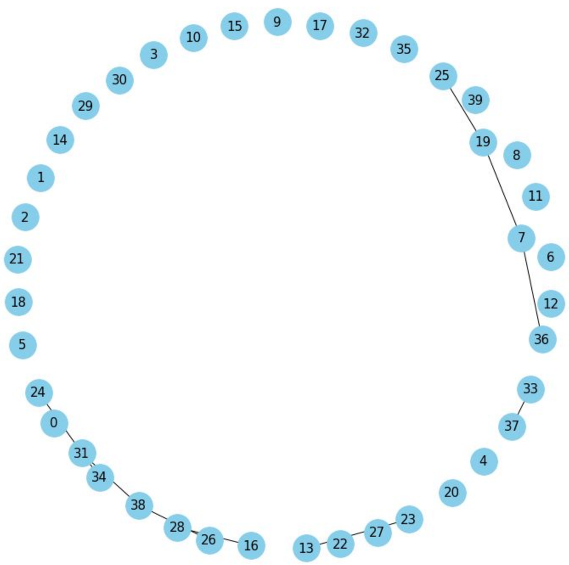
$$P(k_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- Phase transition at $\mathbf{p_c}$ (critical point) = 1/n

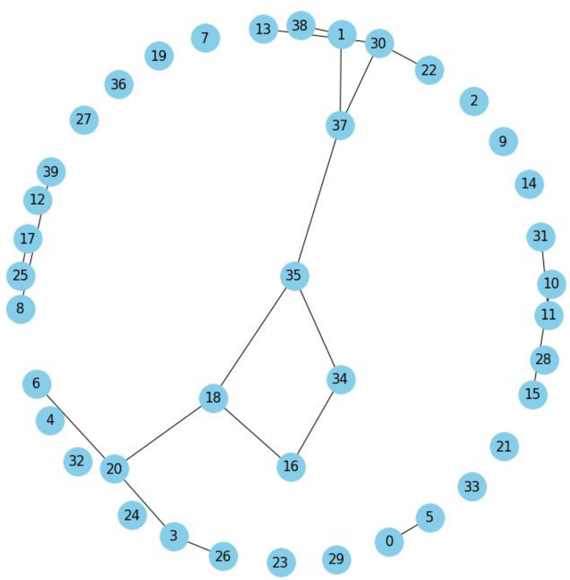
Erdős–Rényi model

- Example – 40 nodes, different **p**

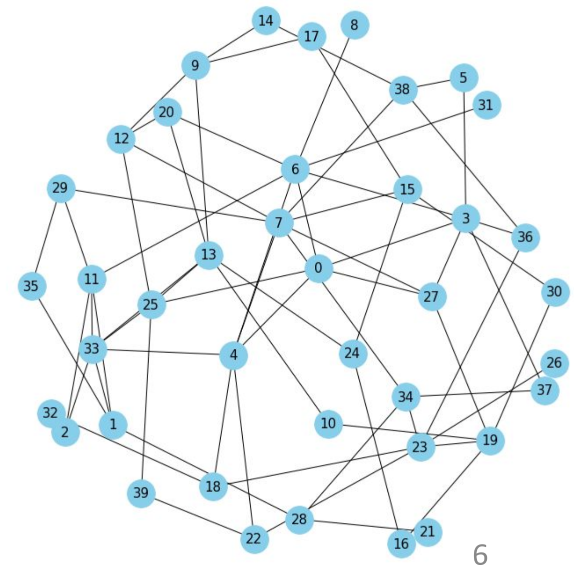
p = 0.01,   9 edges
Avg. degree = 0.45



p = 0.025, 19 edges
Avg. degree = 0.95



p = 0.1, 69 edges
Avg. degree = 3.45



Erdős–Rényi model

- Clustering coefficient = **p**

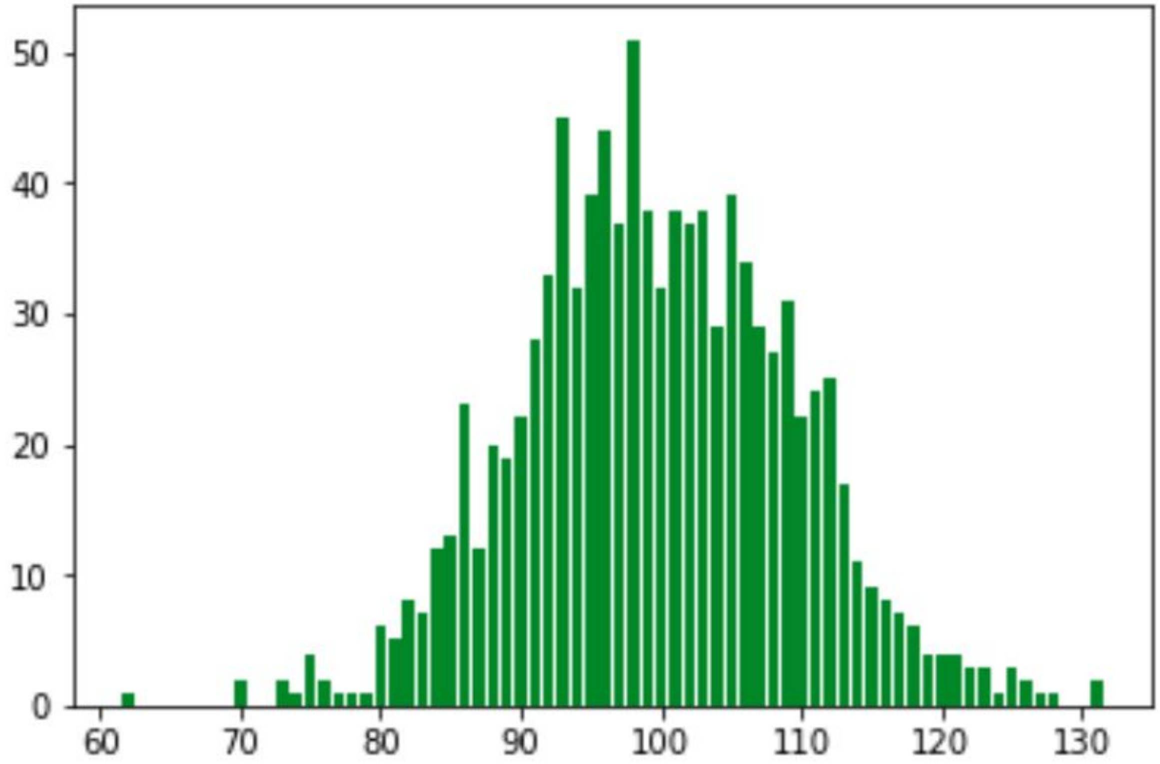
For a node with k neighbors:

#links between neighbors/#max links between neighbors =

$$= [\ p*(k(k-1)/2)\ ]\ /\ [k(k-1)/2] = p$$

Erdős–Rényi model

- Example – degree distribution for G(1000, 0.1)



“Small-world” model

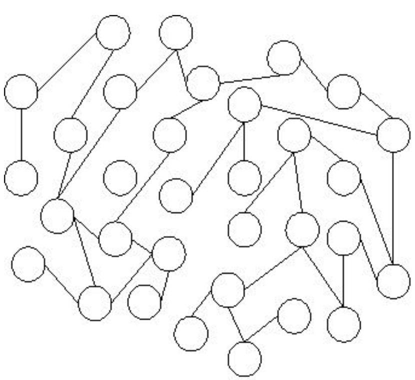
- Properties:
  - Small diameter (proportional to log N)
  - High clustering coefficient

- A class of random graphs by Watts and Strogatz

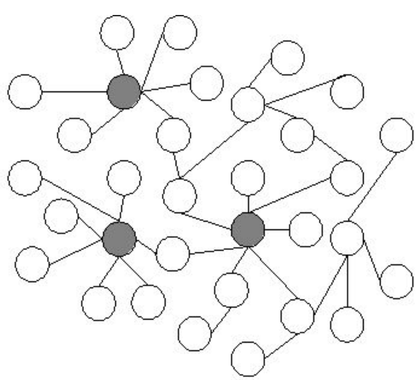
Scale-free networks

- A network whose degree distribution follows power law.

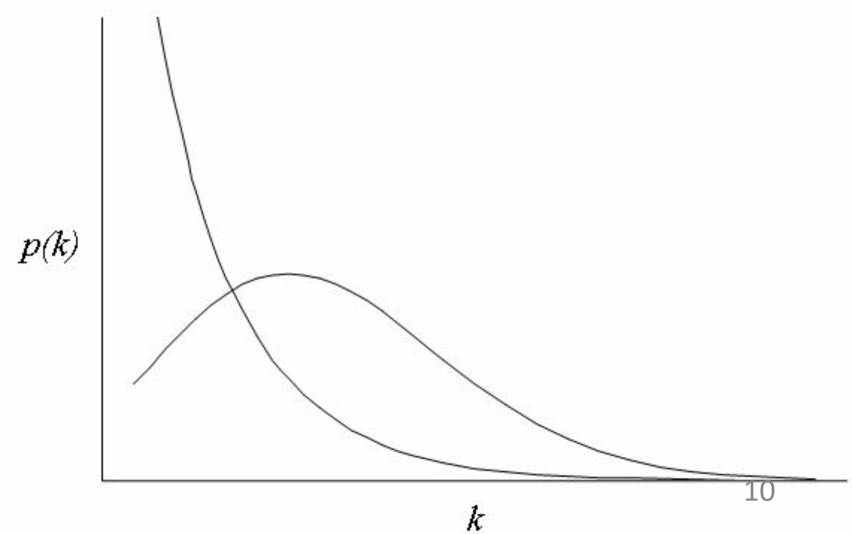
$$P(k) \sim k^{-\gamma}$$



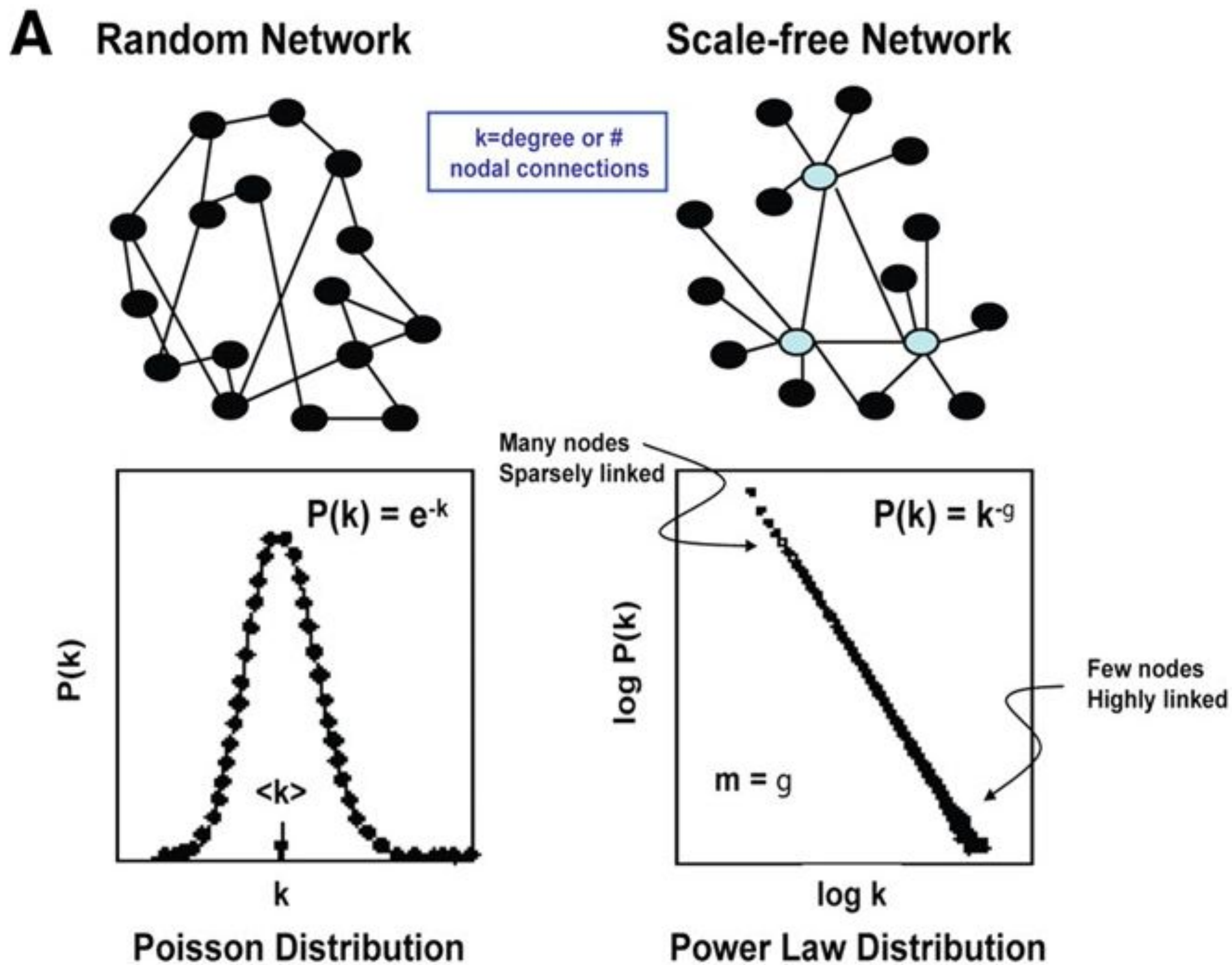
(a) Random network



(b) Scale-free network



Random vs Scale-free networks



“Small-world” model

- Small-world examples:
  - Co-authors in the same domain
  - Colleagues
  - Classmates

- Non small-world examples:
  - “went-to-same-school” people

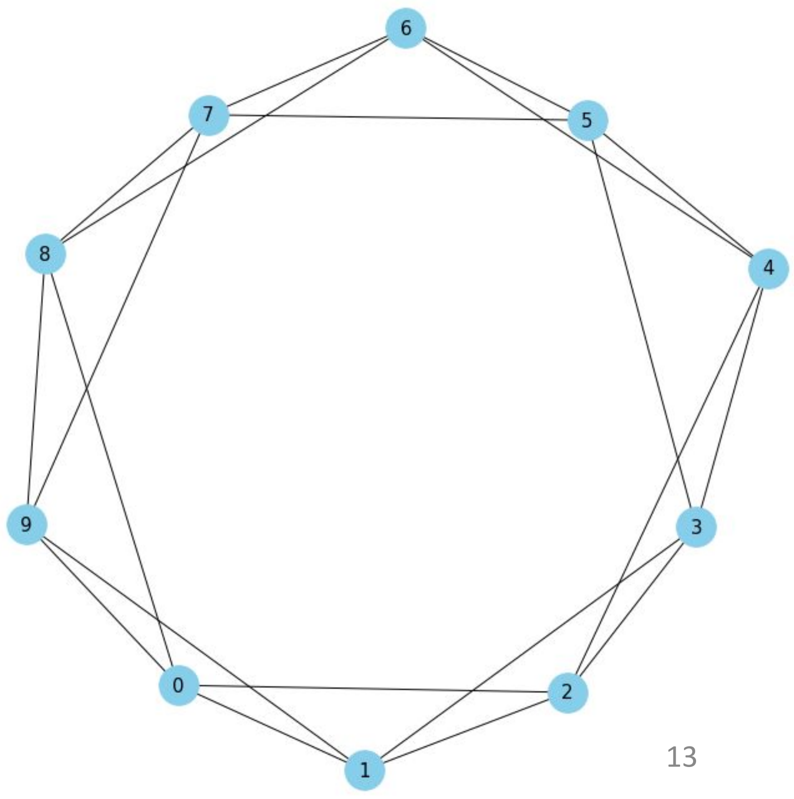
Watts-Strogatz model

- <u>Input</u>: **N** nodes, with average degree **K** and probability **p** of "recreating" the edge.

Step 1:

Create N nodes, connect each node to K/2 neighbors on the left and right (by IDs)

Result: High clustering coefficient, but also big diameter

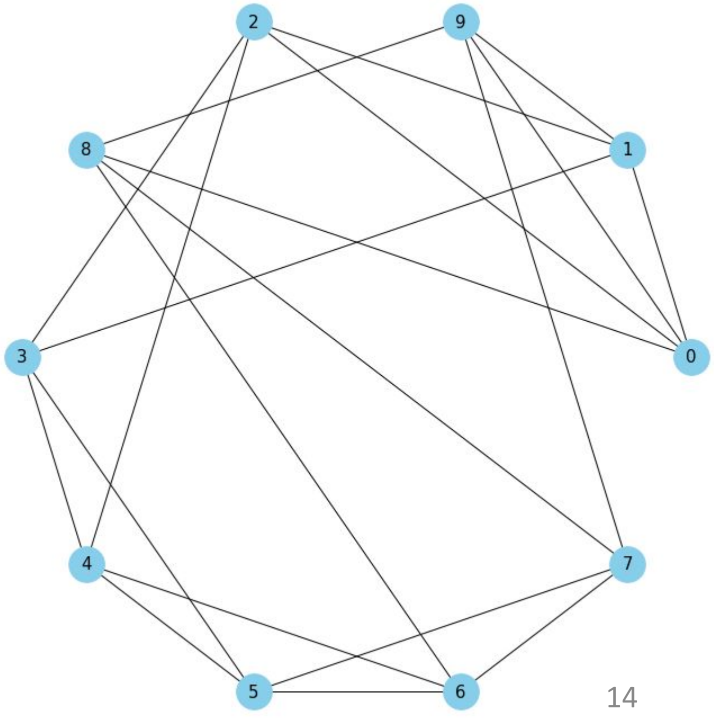


Watts-Strogatz model

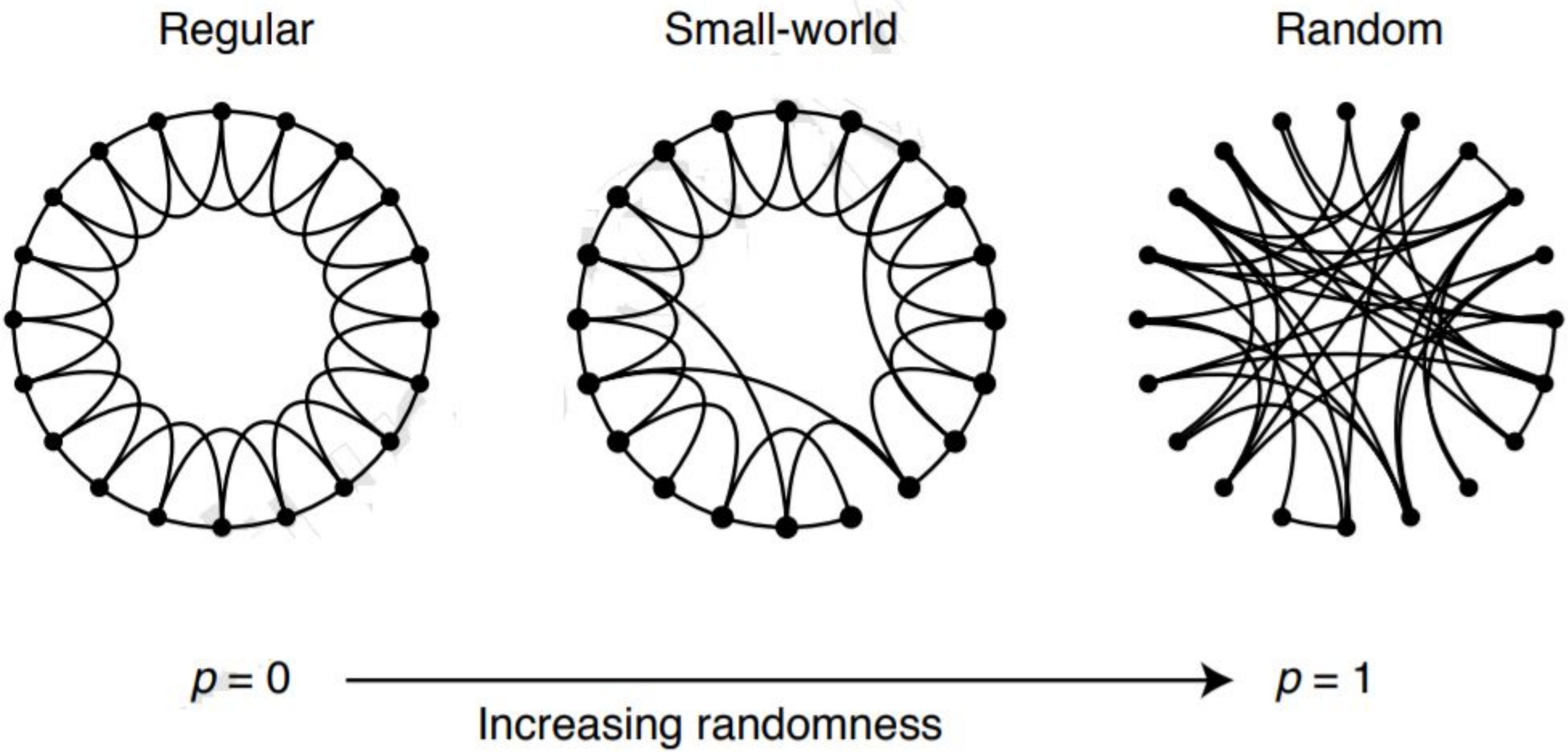
Step 2:

For each edge (i, j), decide if it should be recreated with probability p

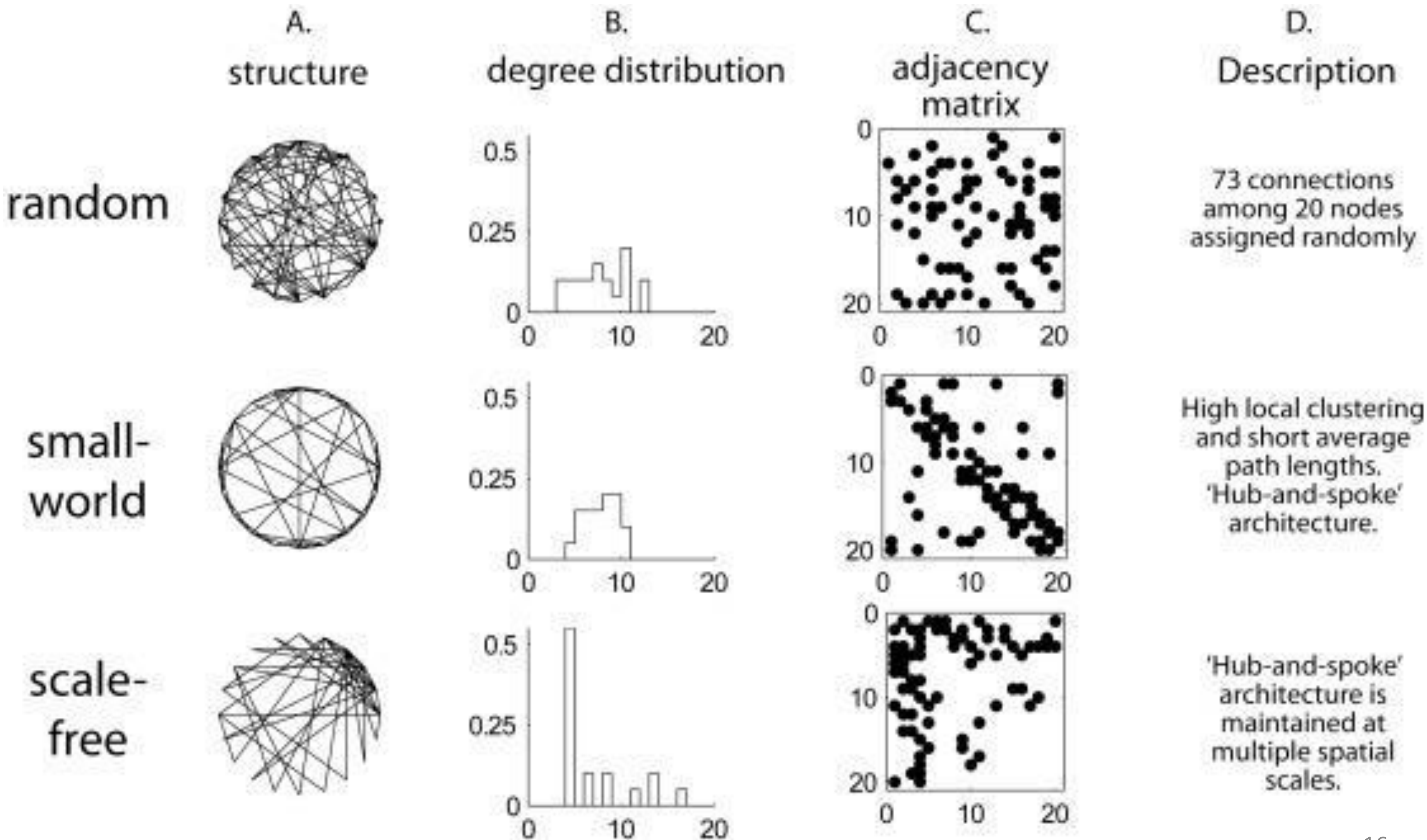
Result: High clustering coefficient, and smaller diameter



Watts-Strogatz model



Summary


A.
structure
B.
degree distribution
C.
adjacency matrix
D.
Description
random
73 connections among 20 nodes assigned randomly
small-world
High local clustering and short average path lengths. 'Hub-and-spoke' architecture.
scale-free
'Hub-and-spoke' architecture is maintained at multiple spatial scales.
0.5
0.25
0
0
10
20

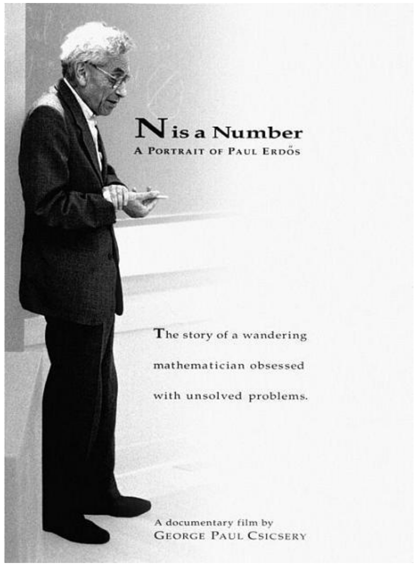
Real World examples

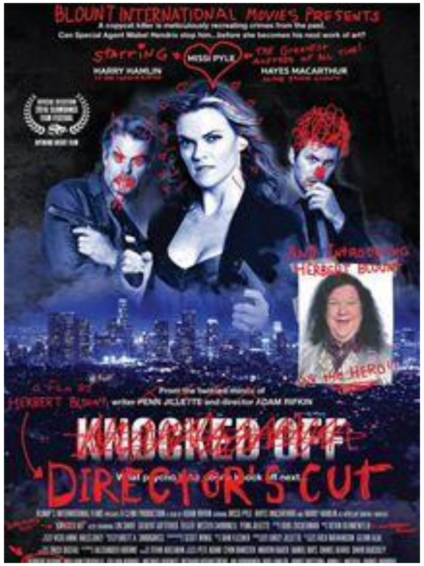
- Erdős number – collaboration distance to Erdős

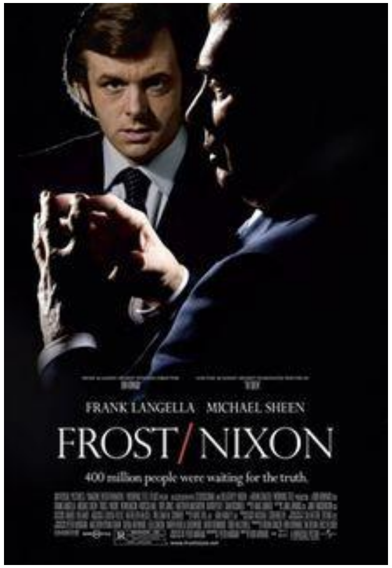
- Kevin Bacon number:
  - Kevin Bacon himself has a Bacon number of 0.
  - Those actors who have worked directly with Kevin Bacon have a Bacon number of 1.
  - If the lowest Bacon number of any actor with whom X has appeared in any movie is N, X's Bacon number is N+1.

Erdős–Bacon number

- Paul Erdős has Erdős–Bacon number 3
  - Erdős number 0
  - Bacon number 3







Ronald Graham Dave Johnson

Erdős–Bacon number

- Natalie Portman has Erdős–Bacon number 7
  - Erdős number 5
  - Bacon number 2

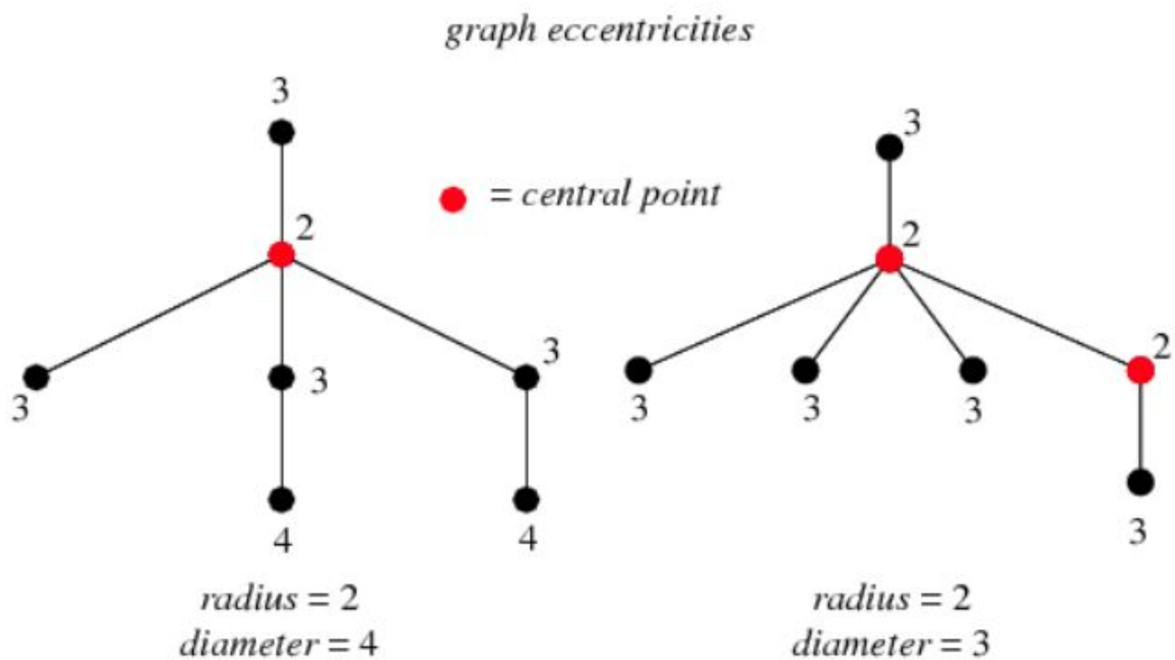
Centrality Measures

Centrality

- Identify the most important vertices in a graph
- Applications:
  - Most influential people
  - Key infrastructure nodes
  - Information spread points
- The measure we chose is often depends on the application

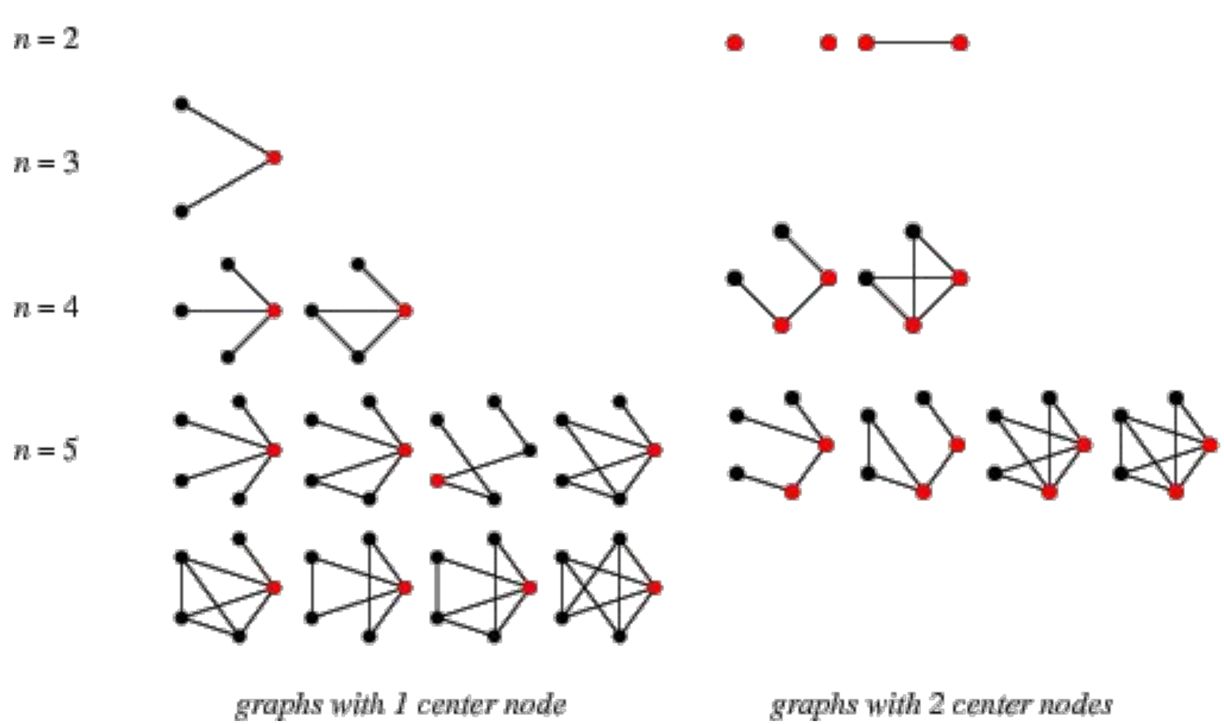
Preliminaries

- Eccentricity (of node **v**) – maximal distance between **v** and any other node.
- **Diameter** – ***maximum*** eccentricity in graph
- **Radius** – ***minimum*** eccentricity in graph



Preliminaries

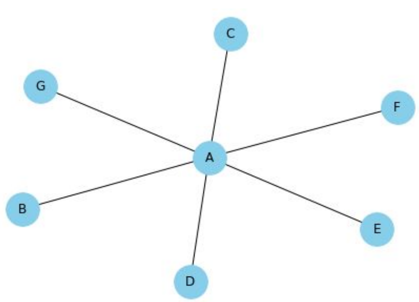
- **Central point** – node with eccentricity = radius
- **Graph center** – set of central points
- **Periphery** – set of nodes with eccentricity = diameter

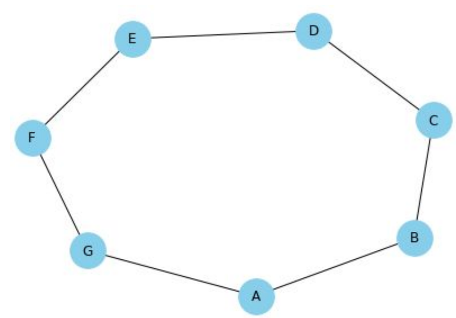


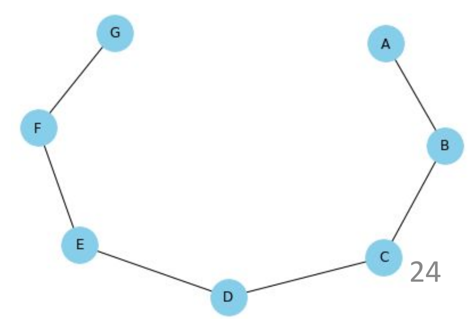
Types of Centrality

- There are many types of centrality measures:
  - Degree Centrality
  - Closeness Centrality
  - Betweenness Centrality
  - Eigenvector Centrality
- To demonstrate, we use 3 types of graphs:

Star graph, Circle graph, Line graph







Things to measure

- Degree Centrality:
  – Connectedness
- Closeness Centrality:
  – Ease of reaching other nodes
- Betweenness Centrality:
  – Role as an intermediary, connector
- Eigenvector Centrality
  – "Whom you know…"

Degree Centrality

- How “connected” is a node?

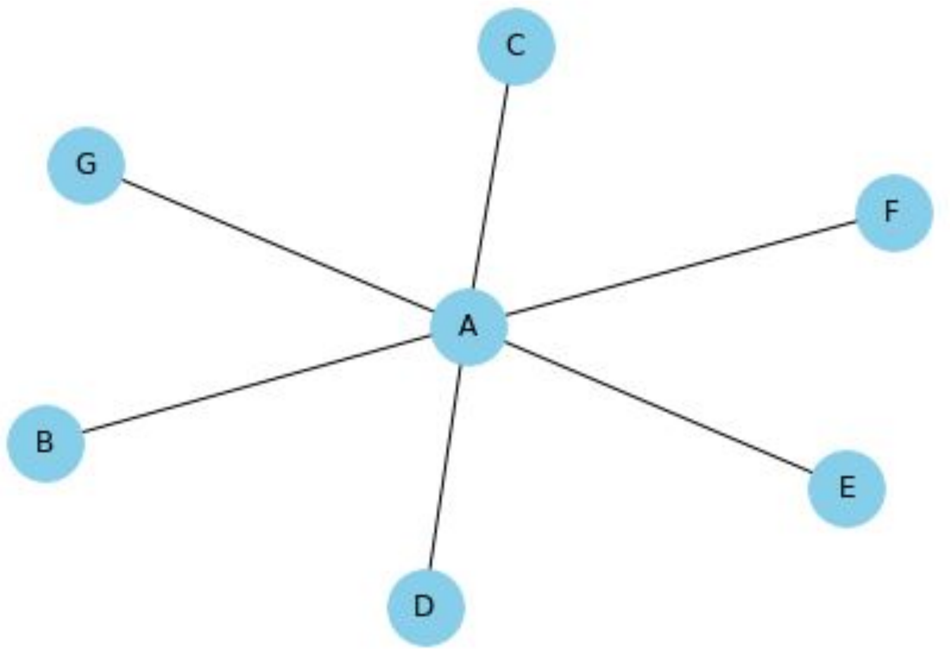
$$C_D(i) = k(i) = \sum_j A_{ij}$$

- Normalized: Divide by (n-1)

$$C_D^*(i) = \frac{1}{n-1} C_D(i)$$

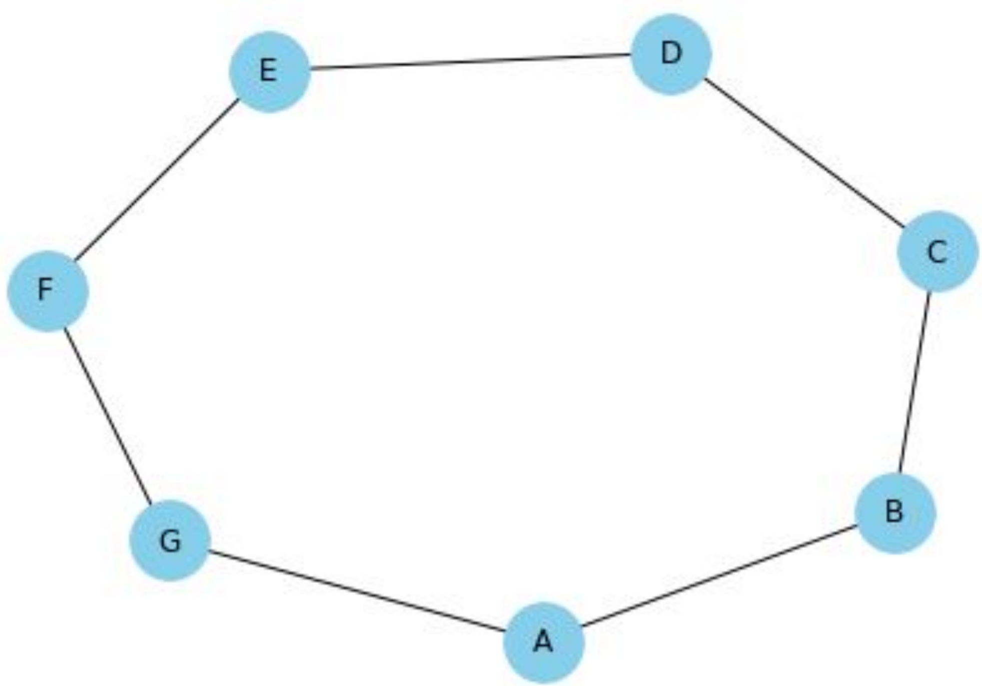
- High centrality – direct contact with many others
- Low centrality – not active

Degree Centrality



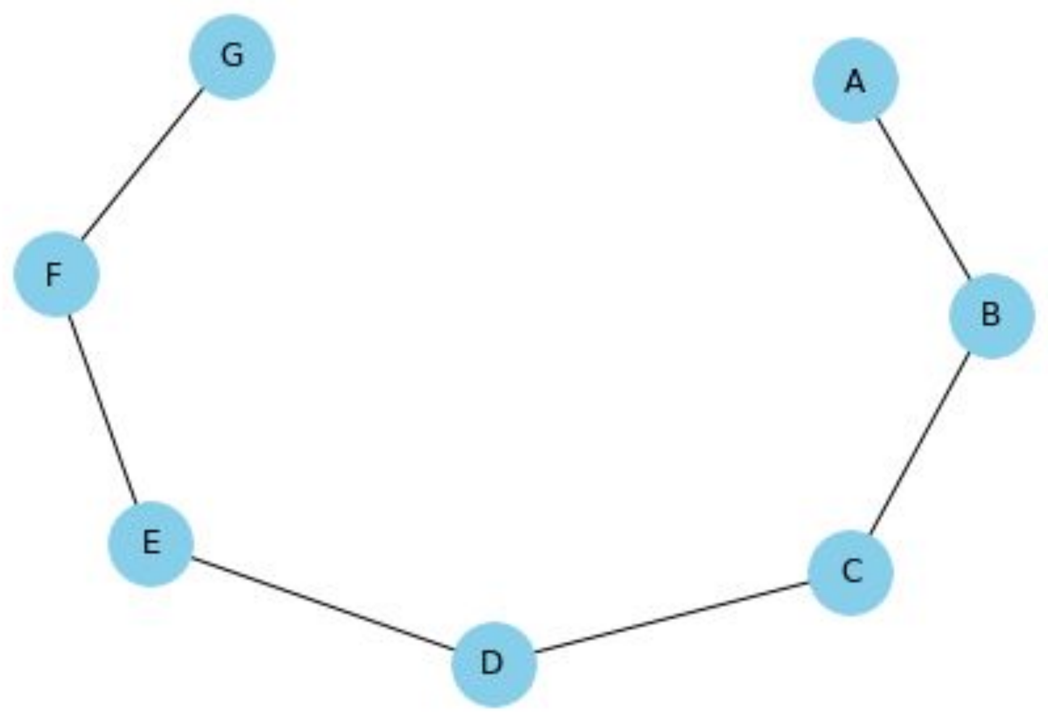
$C_D(A) = 6$
$C_D(B) = 1$
$C_D(C) = 1$
$C_D(D) = 1$
$C_D(E) = 1$
$C_D(F) = 1$
$C_D(G) = 1$

Degree Centrality



$C_D(A) = 2$
$C_D(B) = 2$
$C_D(C) = 2$
$C_D(D) = 2$
$C_D(E) = 2$
$C_D(F) = 2$
$C_D(G) = 2$

Degree Centrality



$C_D(A) = 1$

$C_D(B) = 2$

$C_D(C) = 2$

$C_D(D) = 2$

$C_D(E) = 2$

$C_D(F) = 2$

$C_D(G) = 1$

Closeness Centrality

- How close the node to other nodes in a graph

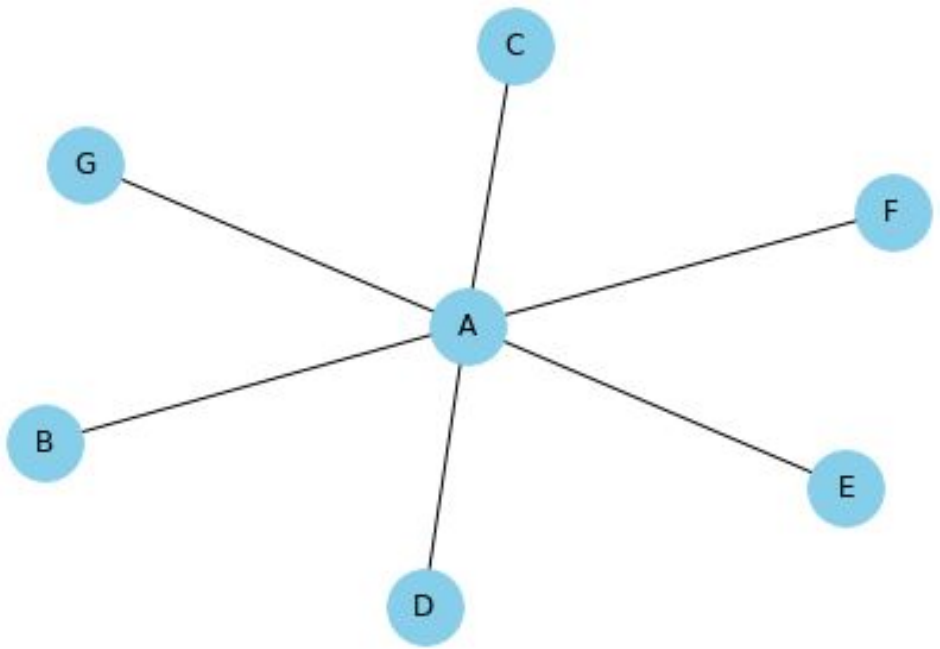
$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

- Normalized: Multiply by (n-1)

$$C_C^*(i) = (n-1)C_C(i)$$

- High centrality – quick interaction with others, short communication path, low number of steps

Closeness Centrality



$C_c(A) = 1/6$

$C_c(B) = 1/11$

$C_c(C) = 1/11$

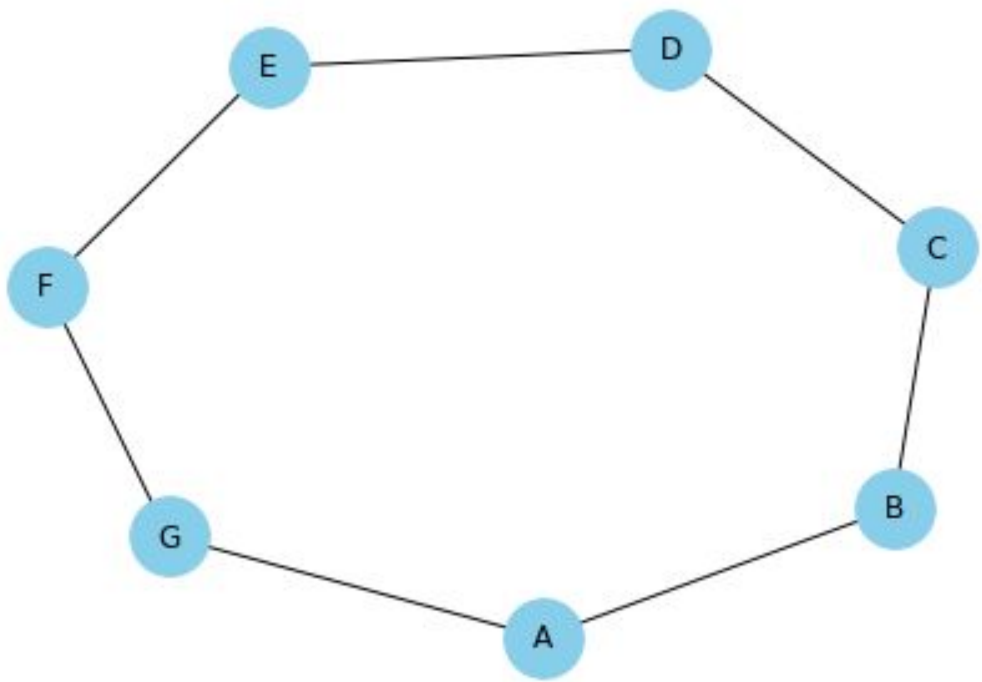
$C_c(D) = 1/11$

$C_c(E) = 1/11$

$C_c(F) = 1/11$

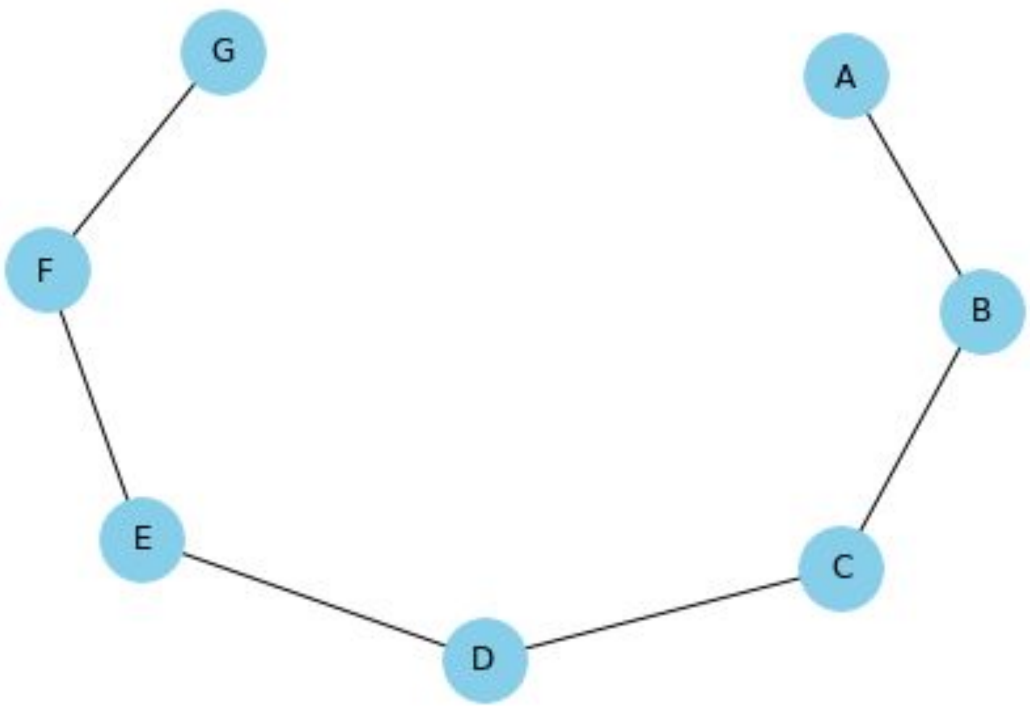
$C_c(G) = 1/11$

Closeness Centrality



$C_c(A) = 1/12$
$C_c(B) = 1/12$
$C_c(C) = 1/12$
$C_c(D) = 1/12$
$C_c(E) = 1/12$
$C_c(F) = 1/12$
$C_c(G) = 1/12$

Closeness Centrality



$C_c(A) = 1/21$
$C_c(B) = 1/16$
$C_c(C) = 1/13$
$C_c(D) = 1/12$
$C_c(E) = 1/13$
$C_c(F) = 1/16$
$C_c(G) = 1/21$

Betweenness Centrality

- Number of shortest pathes going through **i**

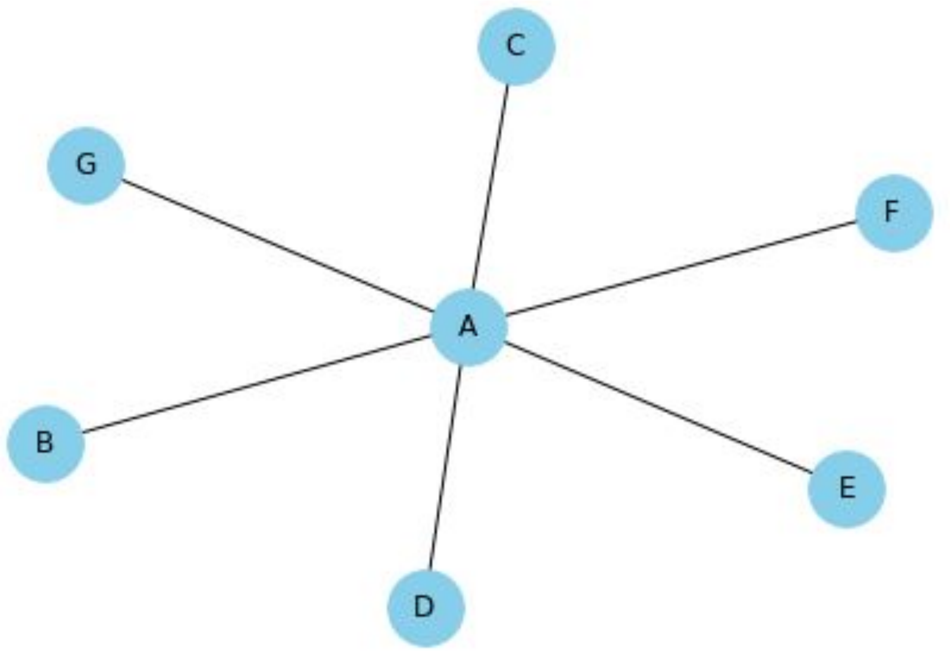
$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

- Normalized: Divide by (n-1)(n-2)/2

$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i)$$

- High centrality – probability of communication between **s** and **t** going through **i**

Betweenness Centrality



$C_B(A) = 15$

$C_B(B) = 0$

$C_B(C) = 0$

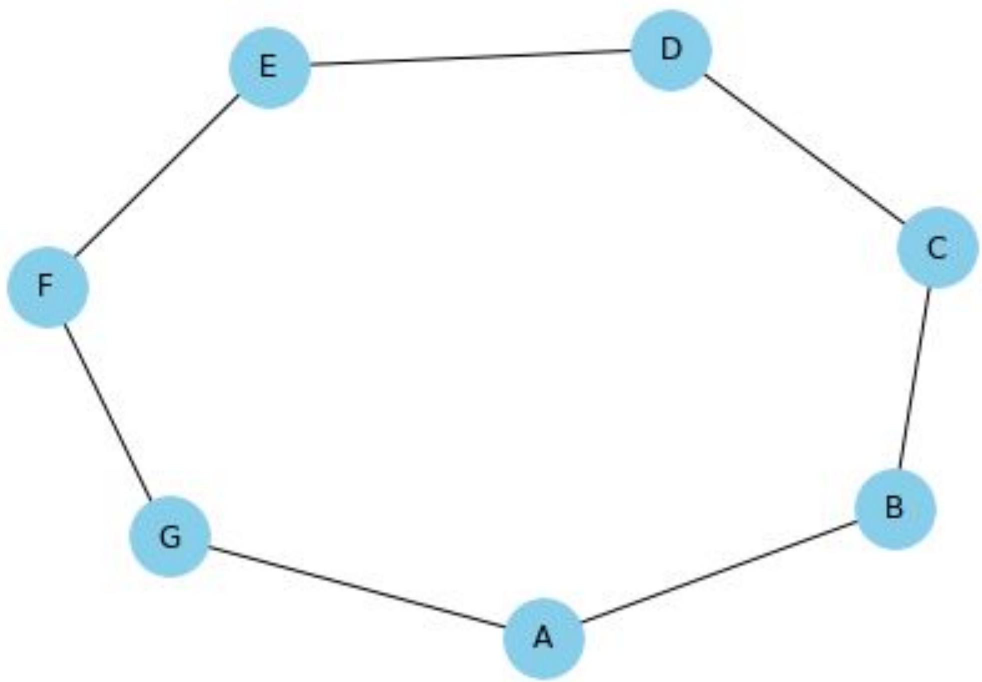
$C_B(D) = 0$

$C_B(E) = 0$

$C_B(F) = 0$

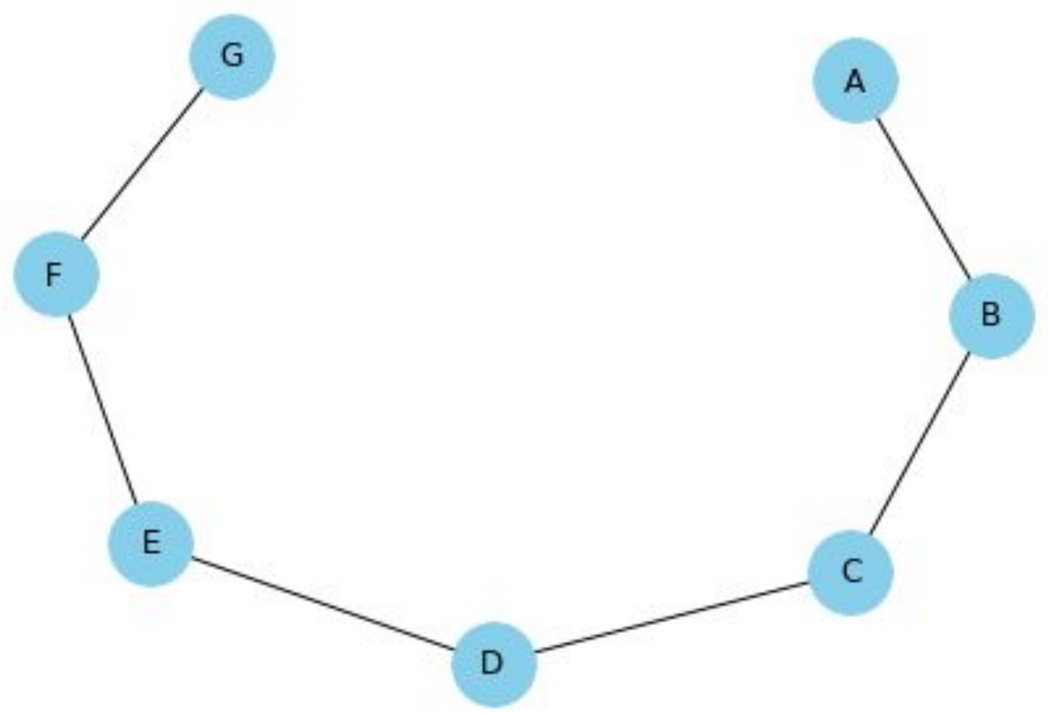
$C_B(G) = 0$

Betweenness Centrality



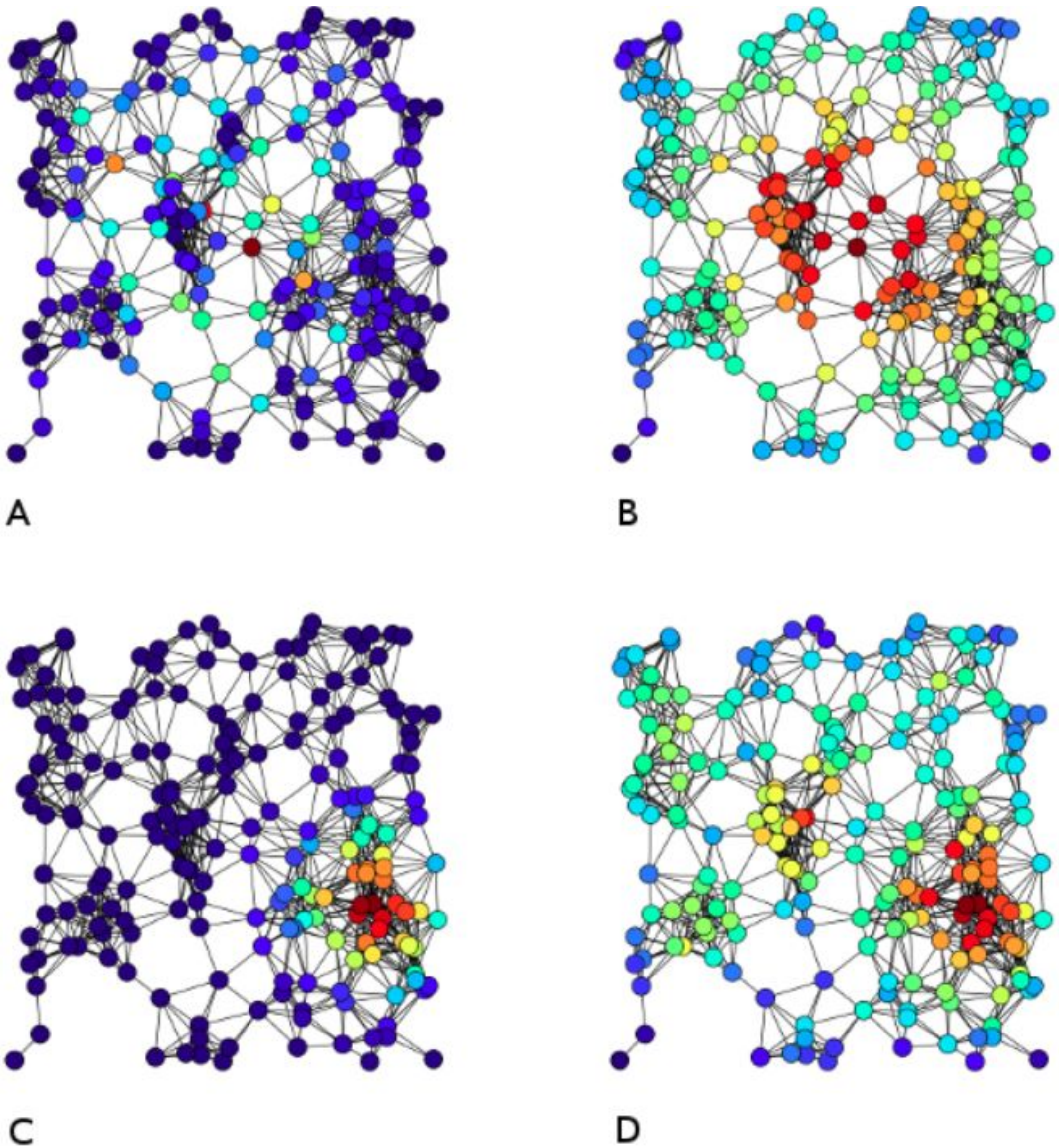
$C_B(A) = 3$
$C_B(B) = 3$
$C_B(C) = 3$
$C_B(D) = 3$
$C_B(E) = 3$
$C_B(F) = 3$
$C_B(G) = 3$

Betweenness Centrality



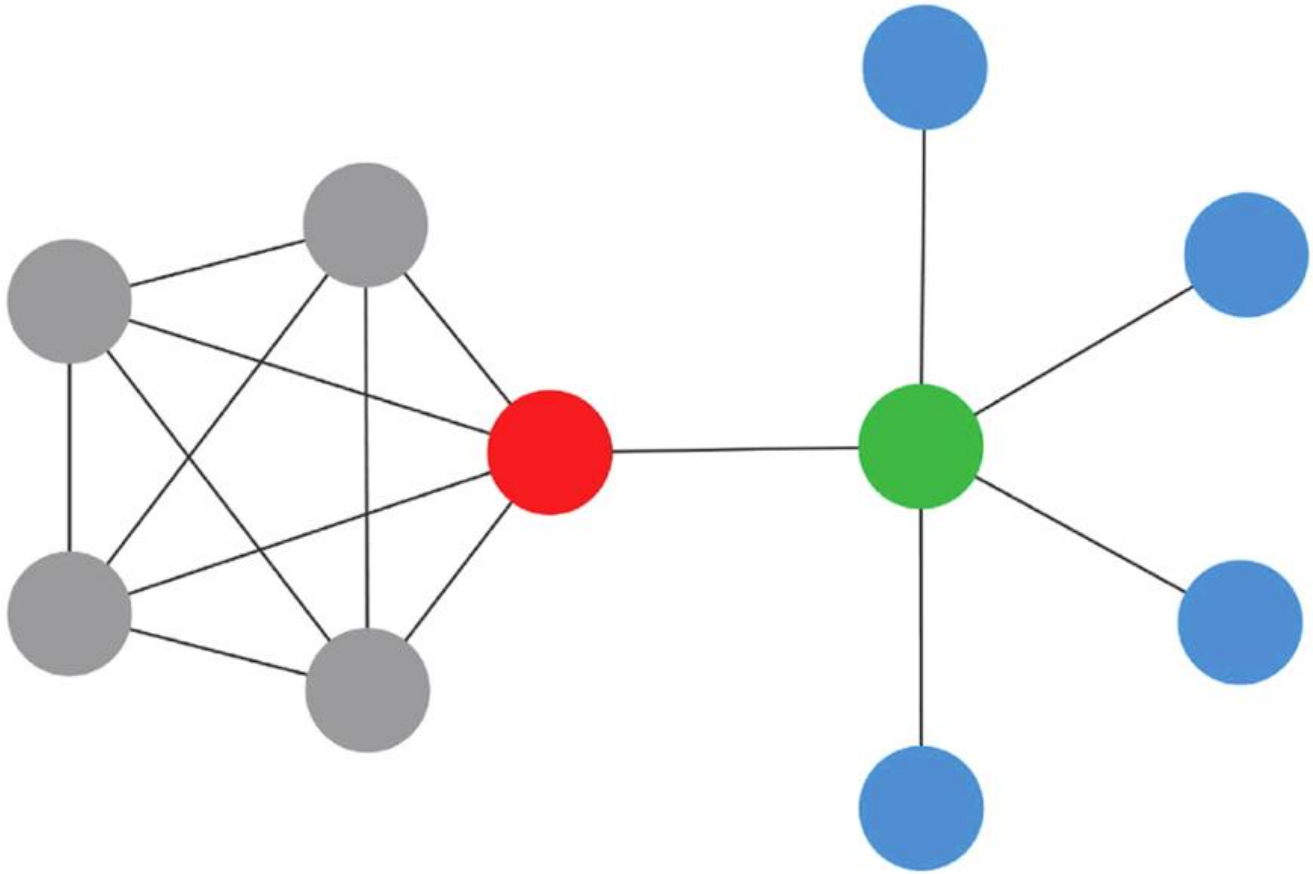
$C_B(A) = 0$
$C_B(B) = 5$
$C_B(C) = 8$
$\mathbf{C_B(D) = 9}$
$C_B(E) = 8$
$C_B(F) = 5$
$C_B(G) = 0$

Centralities



A) Betweenness
B) Closeness
C) Eigenvector
D) Degree

HW question example



Compute and explain 3 types of centrality

Families of Florence

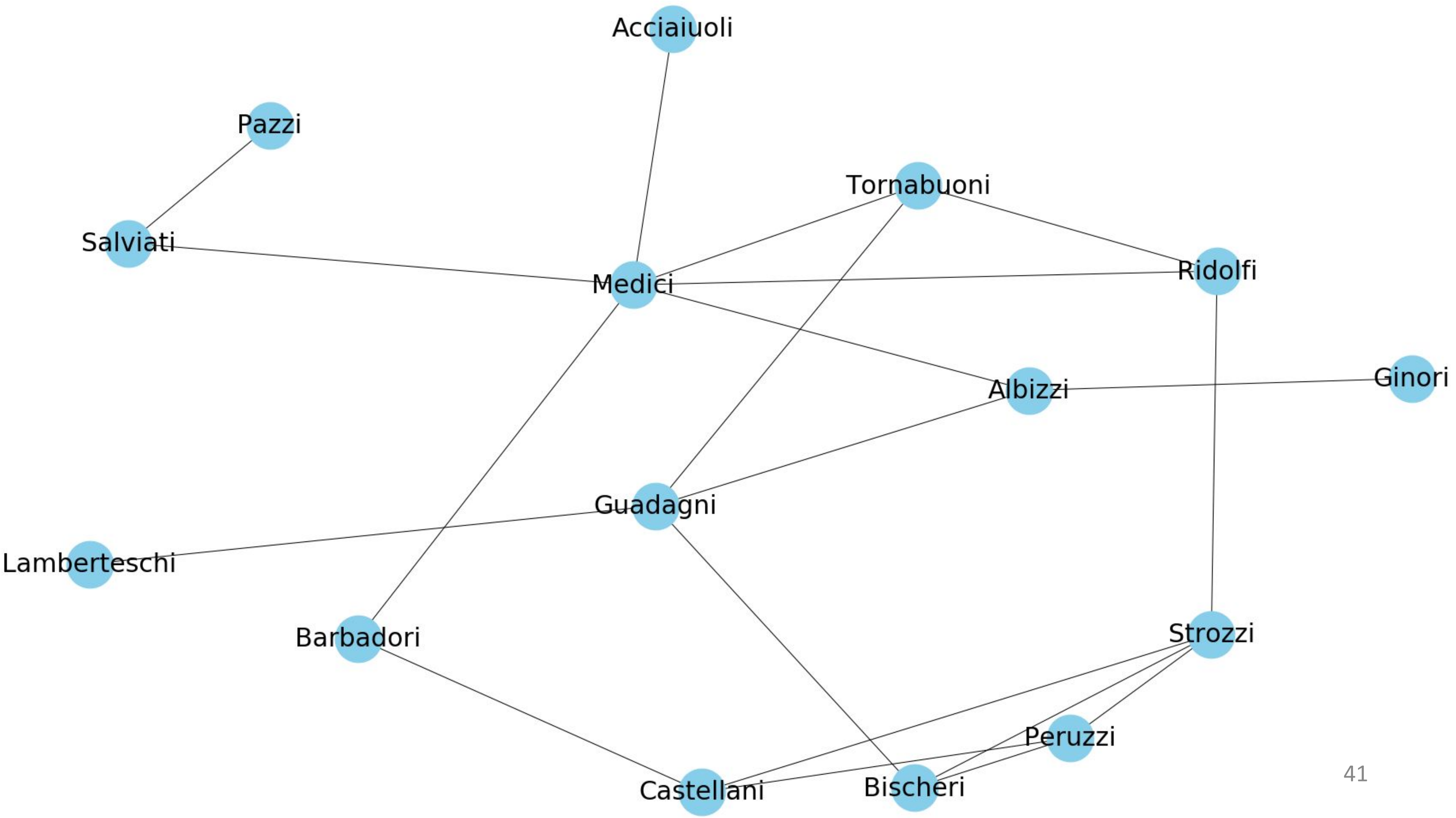
- Marriage and relationships of 16 families in Florence in middle ages

- Very interesting, “classic” network to analyze

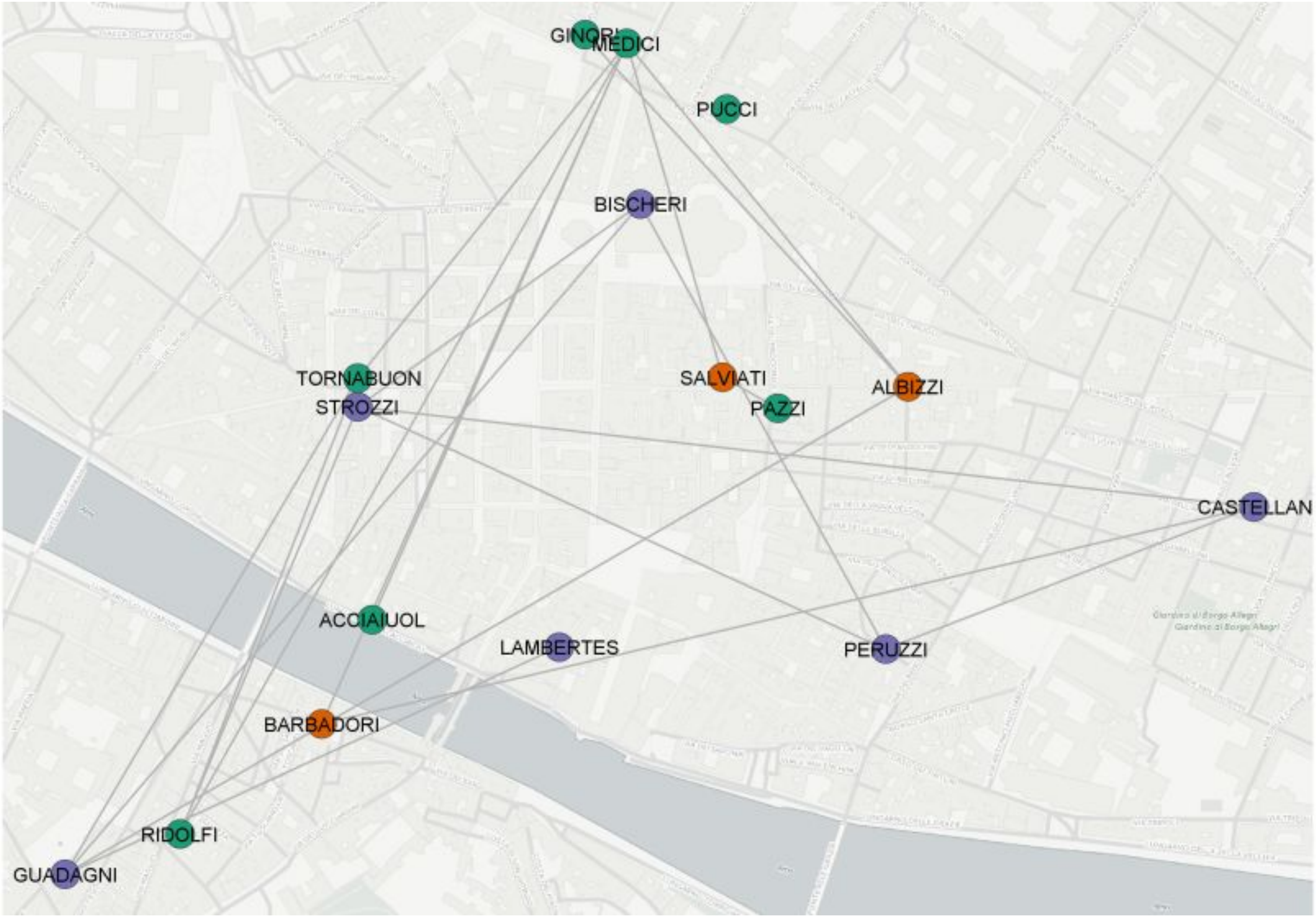
- The rise of Medici family

(https://www2.bc.edu/candace-jones/mb851/Mar12/PadgettAnsell_AJS_1993.pdf)

Families of Florence



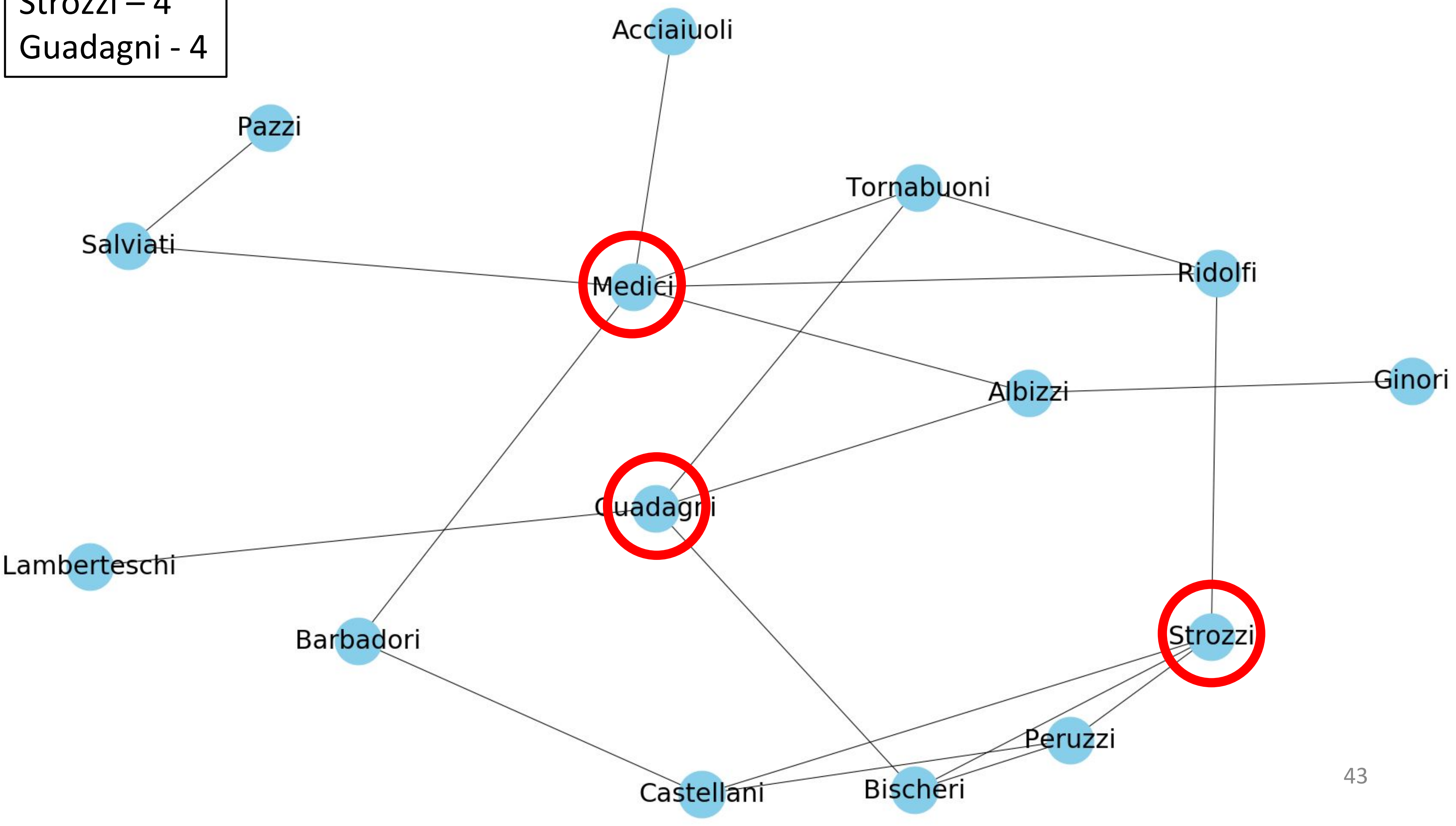
Families of Florence



From: https://simonfink.wordpress.com/2016/05/11/the-medici-marriage-network-in-florence/

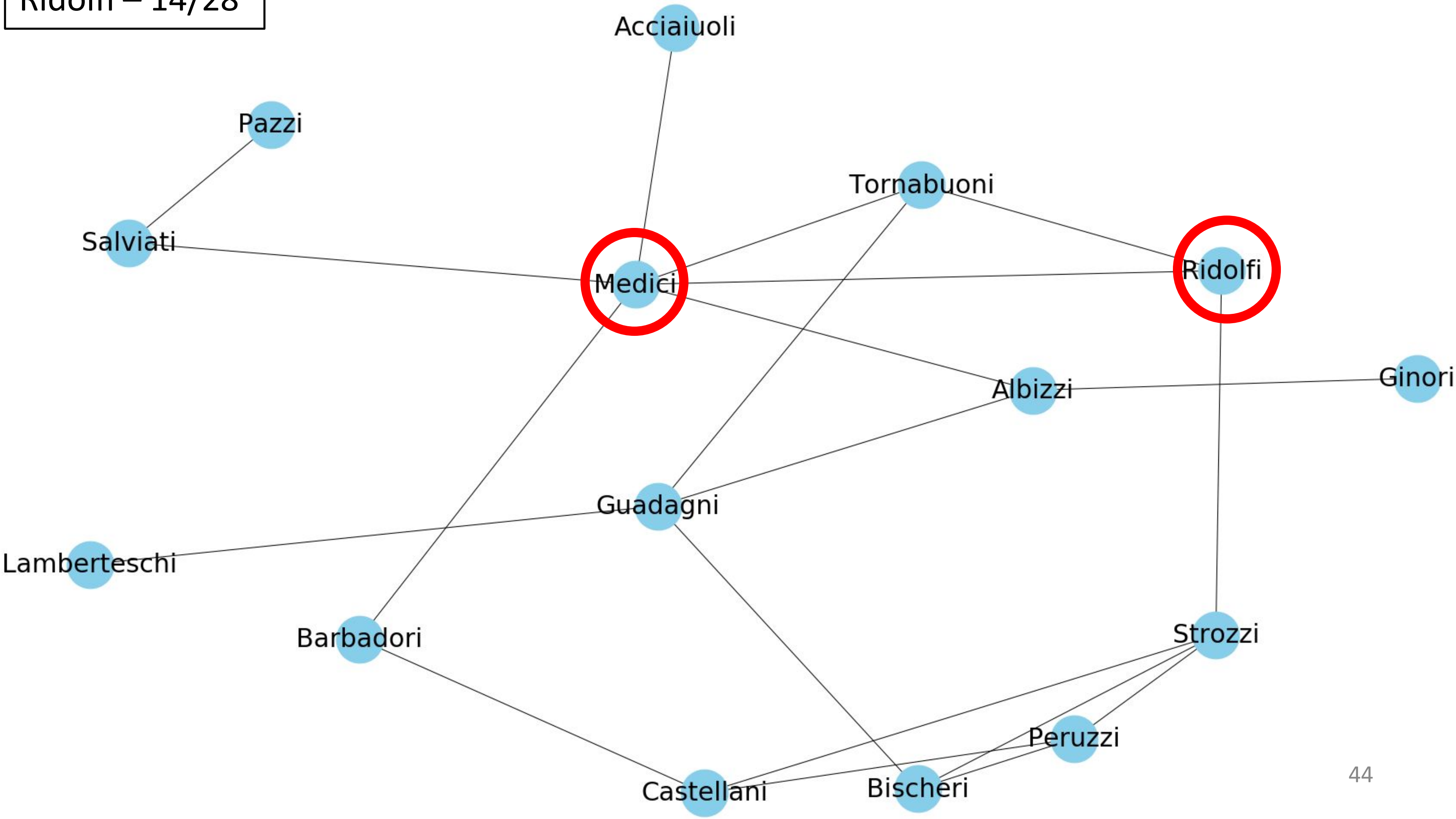
Degree Centrality

Medici – 6
Strozzi – 4
Guadagni - 4



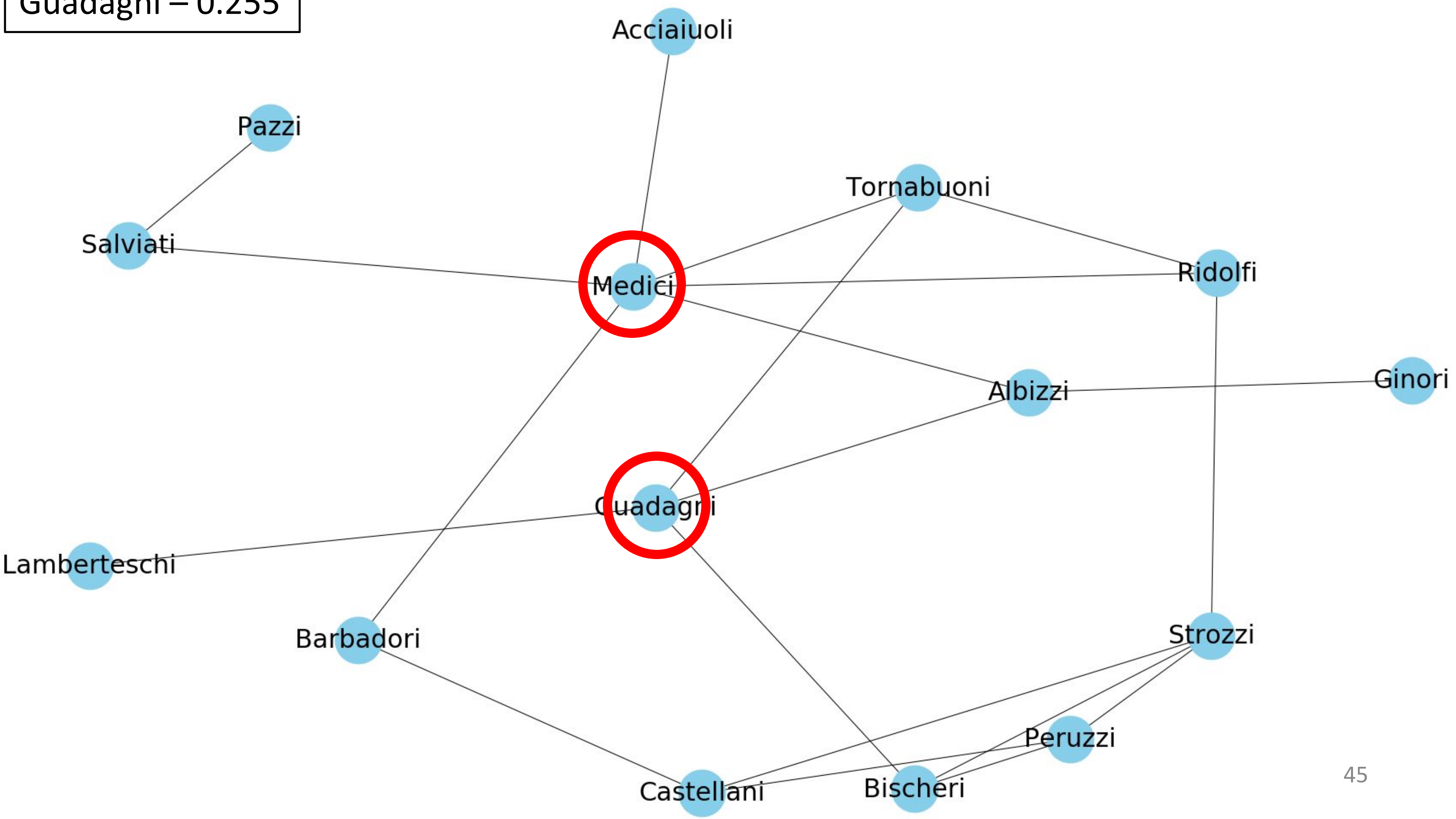
Closeness Centrality

Medici – 14/25
Ridolfi – 14/28

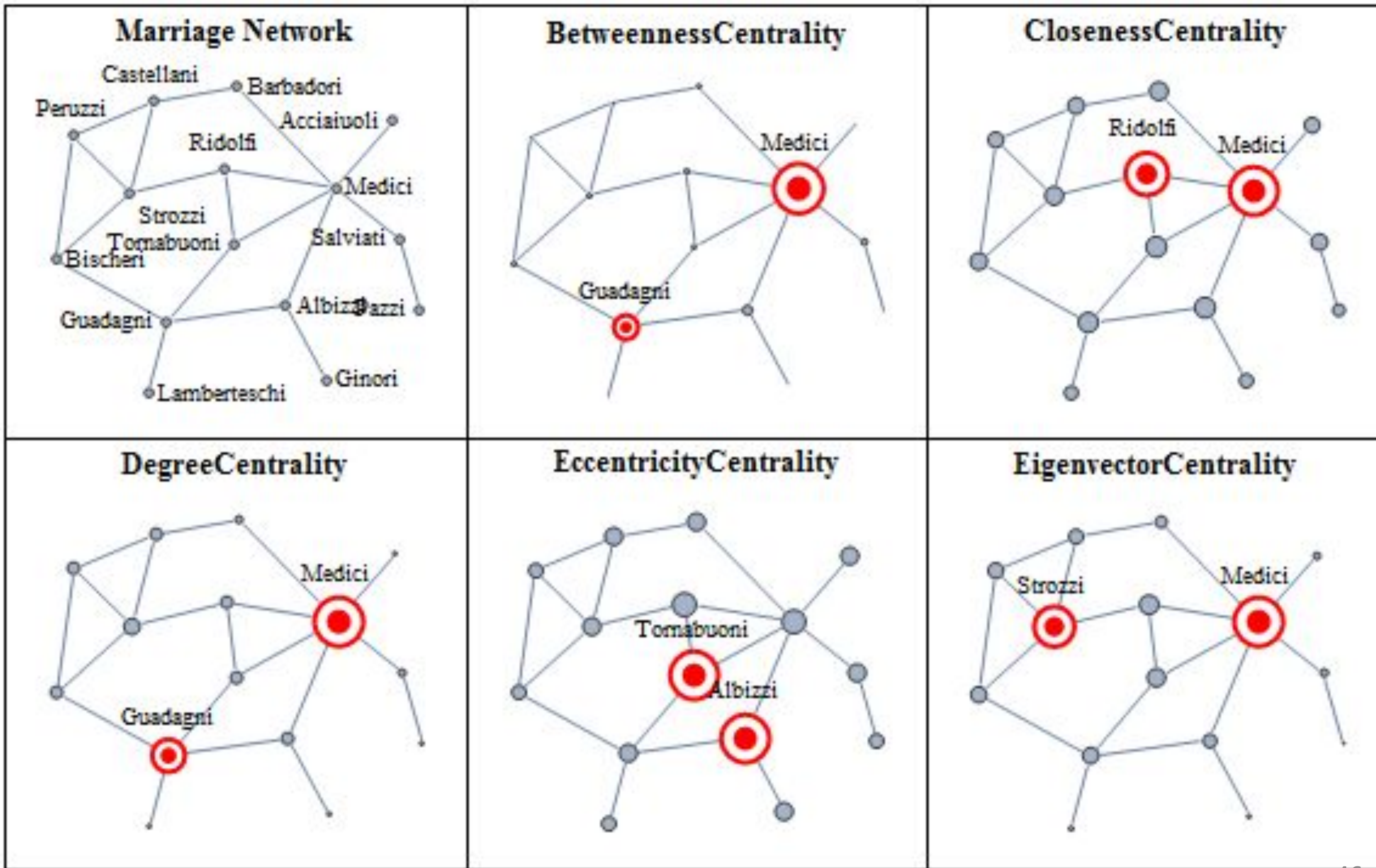


Betweeness Centrality

Medici – 0.522
Guadagni – 0.255

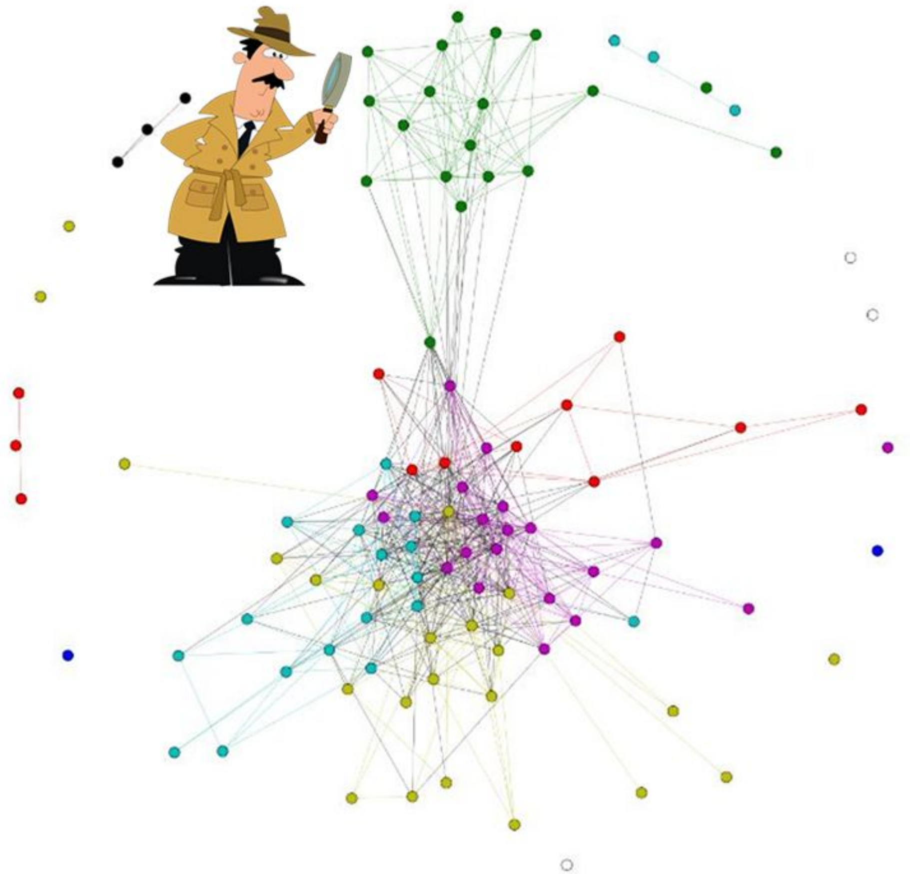


Families of Florence



Another usecase

- Sister found her “lost” brother by analyzing his (online)social network connections



Another usecase

- Sister found her “lost” brother by analyzing his (online)social network connections

Degree centrality	Betweenness centrality	Closeness centrality
A	A	A
B	E	E
C	F	B
D	G	C
E	B	D

Another usecase

Happy End – brother was found through the connection E!

Degree centrality	Betweenness centrality	Closeness centrality
A	A	A
B	E	E
C	F	B
D	G	C
E	B	D

Thank you!

Questions?