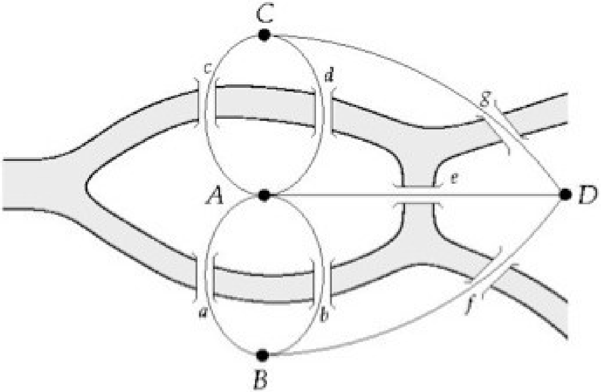
Representing Networks

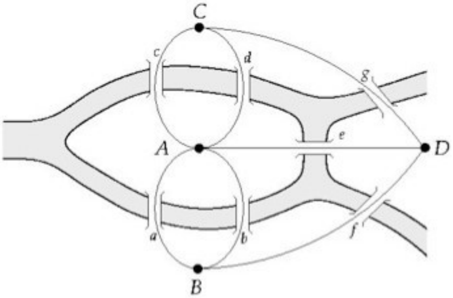
December 11, 2023

The Bridges of Königsberg



Can one walk across the seven bridges and never cross the same bridge twice?

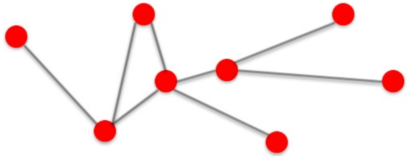
The Bridges of Königsberg



1735: Euler's theorem:

1. If a graph has more than two nodes of odd degree, there is no path.
2. If a graph is connected and has no odd-degree nodes, it has at least one path

A Graph (Network)



- Components: nodes, vertices N/V
- interactions: links, edges L/E
- system: network, graph $(N, L)/(V, E)$

Some examples

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Free-Trade Networks (Furusawa and Konishi 2002)

- Nodes: Countries
- a link between two countries is interpreted as a free-trade agreement, which means that the goods produced in either of the countries can be traded without any tariff to consumers in the other country.

The Connections Model (Jackson and Wolinsky 1996)

Links represent social relationships between nodes such as friendships. These relationships offer benefits in terms of favours, information, and so on. Moreover, these could be interpreted as sharing some common interests.

The Coauthor Model (Jackson and Wolinsky 1996)

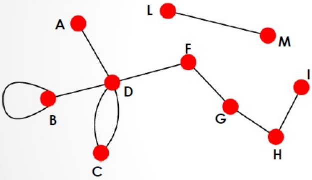
- Each vertex is a researcher who spends time working on research projects.
- If two researchers are connected, then they are working on a project together.

Directed & Undirected Graphs

Undirected

Links: undirected (*symmetrical*)

Graph:

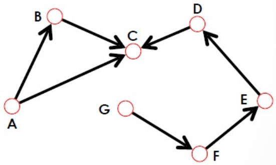


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

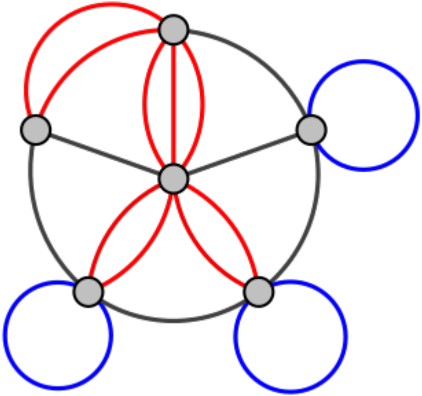
Links: directed (*arcs*).

Digraph = directed graph:



An undirected link is the superposition of two opposite directed links.

1. Isolated node and empty graph
2. Simple and Multigraph

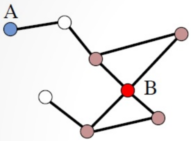


Some examples

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Degree

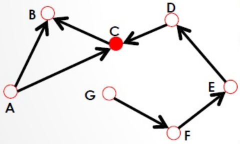
Undirected



Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Directed

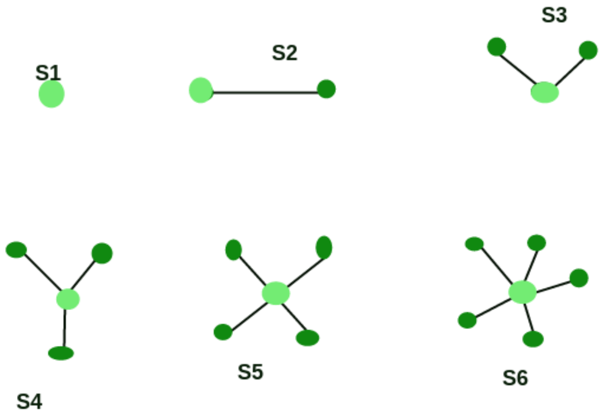


In *directed networks* we can define an in-degree and out-degree. The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \quad k_C^{out} = 1 \quad k_C = 3$$

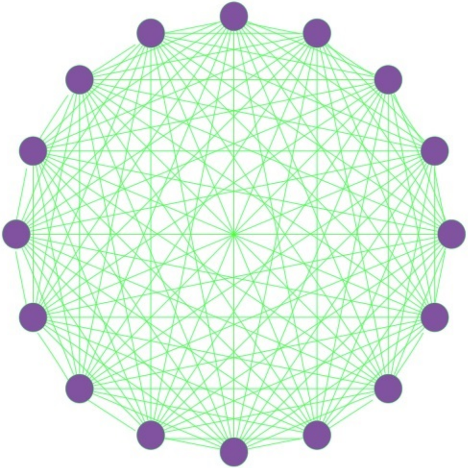
Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

Degree is a "local" property – belongs to a single vertex.



Star Graph of order-n (Sn)

Complete graph



$\#V = n$ and $\#E = \dfrac{n(n-1)}{2}$

Sub-graph

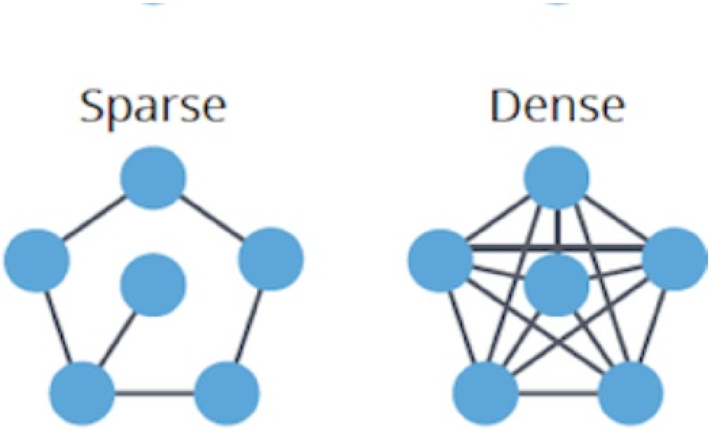
- A spanning subgraph contains all of the vertices from the parent graph and need not contain all of the edges.
- An induced subgraph contains a subset of the vertices of the parent graph along with all of the edges that connect the vertices that exist in both the parent graph and the subgraph.
  $g|S = g \cap g^S,\ S \subseteq V$

Regular Graph

- A graph (V, E) is said to be a regular graph if all its vertices have the same degree. A graph is said to be a k-regular graph if $d(v) = k, \forall v \in V$.
- Every complete graph is an $(n-1)$-regular graph.
- Every circle graph is a 2-regular graph.

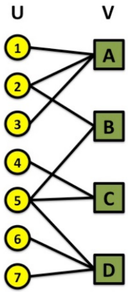
- A dense graph is a graph in which there is a large number of edges. Typically in a dense graph, the number of edges is close to the maximum number of edges.
- The opposite, a graph with only a few edges, is a sparse graph
- A complete graph has the maximum number of edges for its number of vertices. A dense graph is one where there are many edges, but not necessarily as many as in a complete graph.

Sparse: Star, line, circle etc
Dense: complete graph



Bipartite Graph

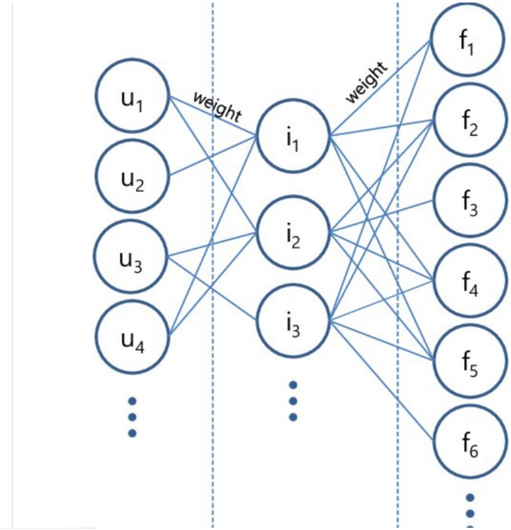
bipartite graph (or **bigraph**) is a graph whose nodes can be divided into two disjoint sets *U* and *V* such that every link connects a node in *U* to one in *V*; that is, *U* and *V* are independent sets.



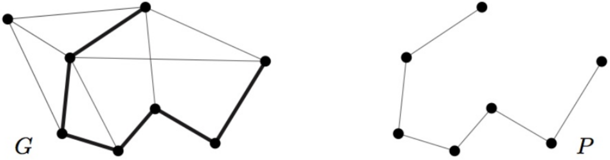
Examples:

U – People, V – Hobbies
U – Recipies, V – Ingredients
U – Documents, V – Keywords

Tripartite Graph: a set of graph vertices decomposed into three disjoint sets such that no two graph vertices within the same set are adjacent



Paths

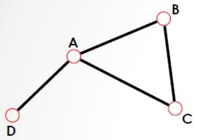


A *path* is a non-empty graph $P = (V, E)$ of the form

$$V = \{ x_0, x_1, \ldots, x_k \} \qquad E = \{ x_0x_1, x_1x_2, \ldots, x_{k-1}x_k \},$$

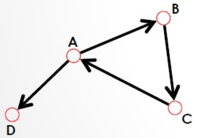
where the x_i are all distinct.

Distance



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

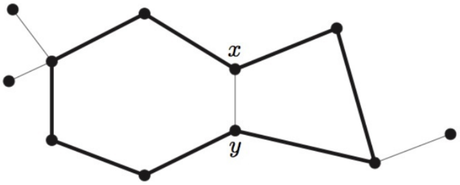
*If the two nodes are disconnected, the distance is infinity.



In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

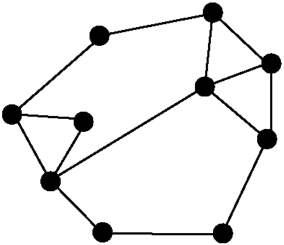
Cycles



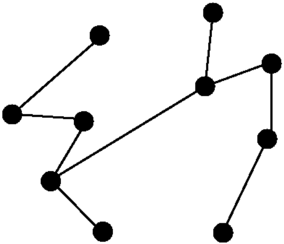
If $P = x_0 \dots x_{k-1}$ is a path and $k \geqslant 3$,

then the graph $C := P + x_{k-1}x_0$ is called a *cycle*.

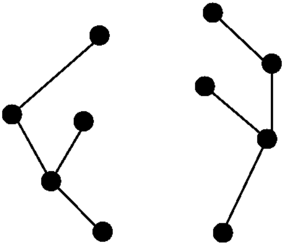
Graph
(with cycles)



Tree
(no cycles, connected)

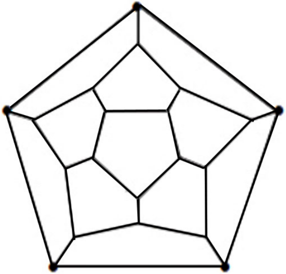


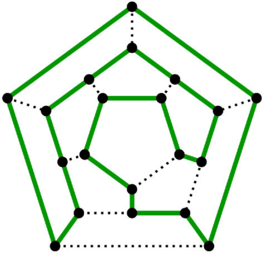
Forest
(no cycles, not connected)



Eulerian Vs Hamiltonian

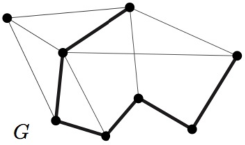
- An Eulerian path is a path that uses every edge of a graph exactly once. An Eulerian cycle is a cycle that uses every edge of a graph exactly once.
- A Hamiltonian path is a path that uses every vertex of a graph exactly once. A Hamiltonian cycle is a cycle that uses every vertex of a graph exactly once.

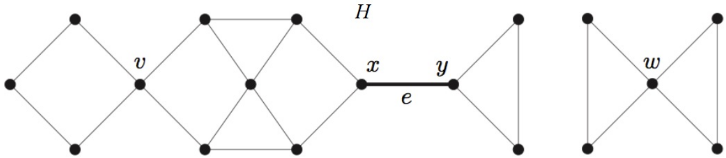




Hamiltonian path

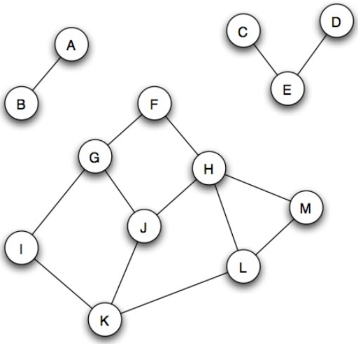
Connectivity





A non-empty graph G is called *connected* if any two of its vertices are linked by a path in G. If $U \subseteq V(G)$ and $G[U]$ is connected, we also call U itself connected (in G).

Connected Components

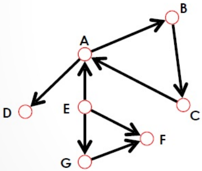


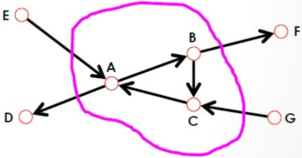
Can you please try to define a connected component?

Strongly-Connected Component

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.



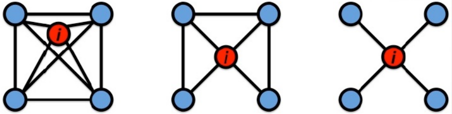


Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.

Clustering Coefficient

- **What fraction of your neighbours are connected?**
  - Node i with degree k_i
  - Clustering Coefficient C_i for a vertex i is in [0,1]

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



$C_i = 1$ $\quad$ $C_i = 1/2$ $\quad$ $C_i = 0$

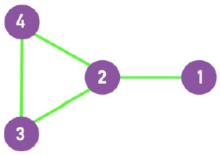
Clustering coefficient is a “local” property – each vertex has one.

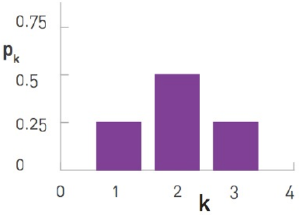
Watts & Strogatz, Nature 1998.

Clustering Coefficients

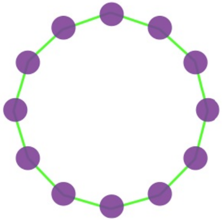
For each $i \in V$ in a complete graph, $C_i = 1$
For a start graph, $C_i = 0$, if i is the central node.
Circle graph: For all $i \in V$, $C_i = 0$
Line graph: $C_i = 0$, if i is not an ending node.

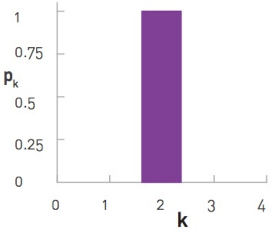
Degree Distribution





$$p_k = \frac{N_k}{N}$$ denotes the probability of a vertex having degree k





Degree Distribution

Star Graph:

$$P_k = \begin{cases} 1 - \dfrac{1}{n} & \text{if } k = 1; \\ \dfrac{1}{n} & \text{if } k = n-1. \end{cases}$$

Expected value$= 2(1 - \dfrac{1}{n})$

Line Graph:

$$P_k = \begin{cases} \dfrac{2}{n} & \text{if } k = 1; \\ 1 - \dfrac{2}{n} & \text{if } k = 2. \end{cases}$$

Expected value$= 2(1 - \dfrac{1}{n})$

Circle graph: $P_2 = 1$, Expected value$= 2$

Network Diameter & Average Distance

Diameter: $\mathbf{d}_{max}$ the maximum distance between any pair of nodes in the graph.

Average path length (Average distance), <d>, for a directed graph:

$$\langle d \rangle \equiv \frac{1}{N(N-1)} \sum_{i,j \neq i} d_{ij}$$

where d_{ij} is the distance from node *i* to node j

In an *undirected graph* $d_{ij} = d_{ji}$, *so* we only need to count them once:

$$\langle d \rangle \equiv \frac{2}{N(N-1)} \sum_{i,j>i} d_{ij}$$

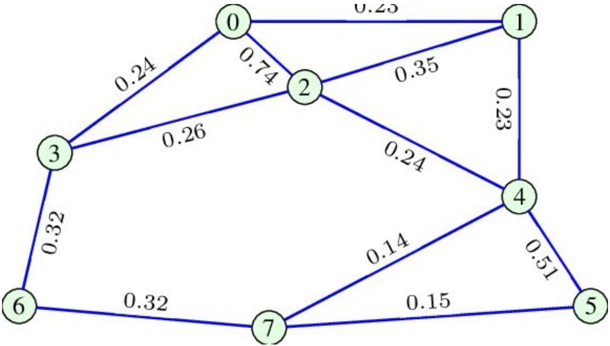
Star graph: $< d >= 2(1 - \frac{1}{n})$

Complete graph: $< d >= 1$

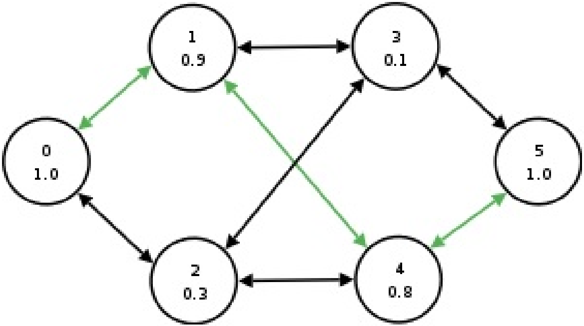
Line graph: $< d >= \frac{1}{n(n-1)} \sum_{i=1}^{n} i(i+1)$

Circle graph: $< d >= \frac{2(n-2)}{n-1}$

Weighted and Probabilistic graph

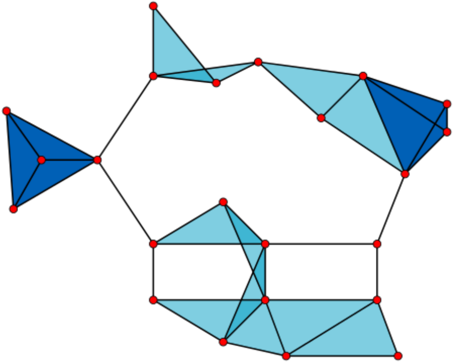


Graph with weighted vertices



Clique

A clique, C, in an undirected graph $G = (V, E)$ is a subset of the vertices, $C \subseteq V$, such that every two distinct vertices are adjacent.



Centrality measures

- Centrality is a very important concept in identifying important/central nodes in a graph.
- By identifying central nodes, we can better understand the flow of information, resources, or influence within the graph.
- Measures of centrality can be categorized into four main groups depending on the types of statistics on which they are based.
  1. degree - how connected a node is,
  2. closeness - how easily a node can reach other nodes,
  3. betweenness - how important a node is in terms of connecting other nodes,
  4. neighbours' characteristics - how important, central, or influential a node's neighbours are.

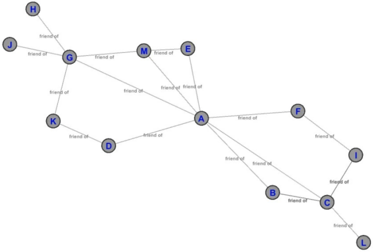
Mathematically, $d(a,b)$= No. of edges between a and b on the shortest path (geodesic) from a to b

$d(a,b) = 0$, if a = b

$d(a,b) = \infty$, if no path exists from a to b

Mathematically, Closeness Centrality C(i) of a node i:

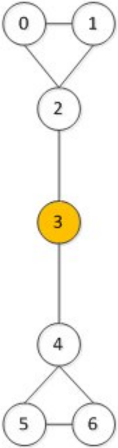
$C(i) = \sum_{j \in V} d(i,j)$



$C(A) = 17, C(L) = 33$.

- Betweenness centrality: $B(i) = \sum_{j \in V} \sum_{k \in V} \frac{\#d_i(j,k)}{\#d(j,k)}$ where $\#d_i(j,k)$ is the number of the shortest paths between j and k through 'i'
- In a telecommunications network, a node with higher betweenness centrality would have more control over the network because more information will pass through that node.

$$D(2) = D(4) = 4 \ D(3) = 2$$



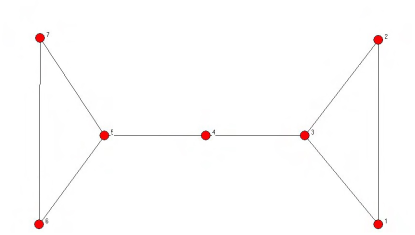
Betweenness of node 2:
Node 3→ 0: 1
Node 3→ 1: 1
Node 4→ 0:1
Node 4→1:1
Node 5→0:1
Node 5→1:1
Node 6→0:1
Node 6→1:1
Betweenness Centrality of node 2:8

Betweenness of node 3:
Node 4→ 0: 1
Node 4→ 1: 1
Node 4→ 2:1
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Betweenness Centrality of node 3:9

Betweenness of node 4:
Node 5→0:1
Node 5→1:1
Node 5→2:1
Node 5→3:1
Node 6→0:1
Node 6→1:1
Node 6→2:1
Node 6→3:1
Betweenness Centrality of node 4:8

Betweenness Centrality of node 0: 0
Betweenness Centrality of node 1: 0
Betweenness Centrality of node 5: 0
Betweenness Centrality of node 6: 0

Prestige, Power and Eigenvector related Centrality



	Nodes 1,2,6,7	Nodes 3 and 5	Node 4
Degree (and Katz Prestige P^K)	.33	.50	.33
Closeness	.40	.55	.60
Decay Centrality ($\delta = .5$)	1.5	2.0	2.0
Decay Centrality ($\delta = .75$)	3.1	3.7	3.8
Decay Centrality ($\delta = .25$)	.59	.84	.75
Betweenness	0.0	.53	.60
Eigenvector Centrality	.47	.63	.54
Katz Prestige-2 P^{K2}, $a = 1/3$	3.1	4.3	3.5
Bonacich Centrality $b = 1/3$, $a = 1$	9.4	13	11
Bonacich Centrality $b = 1/4$, $a = 1$	4.9	6.8	5.4

THANK YOU