SOCIAL NETWORK ANALYTICS

Graph Theory

Prakash C O
Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Graph Theory

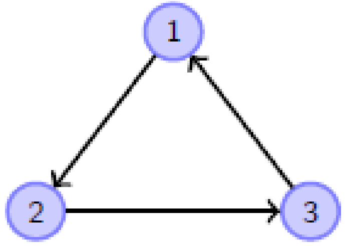
Prakash C O

Department of Computer Science and Engineering

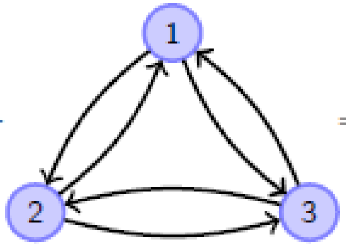
Graph Theory

- We represent a network by a graph (N, g), which consists of a set of nodes $N = \{1, \dots, n\}$ and an $n \times n$ matrix $g = [g_{ij}]_{i,j \in N}$ (referred to as an adjacency matrix), where $g_{ij} \in \{0, 1\}$ represents the availability of an edge from node i to node j.
    - The edge weight $g_{ij} > 0$ can also take on non-binary values, representing the intensity of the interaction, in which case we refer to (N, g) as a weighted graph.
- We refer to a graph as a directed graph (or digraph) if $g_{ij} \neq g_{ji}$ and an undirected graph if $g_{ij} = g_{ji}$ for all $i, j \in N$.

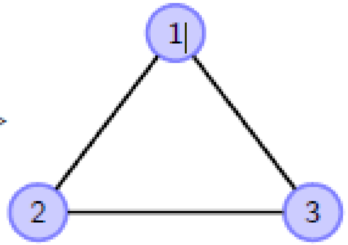
Example 1: $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \Rightarrow$



Example 2: $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \Rightarrow$

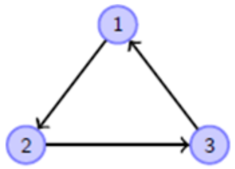


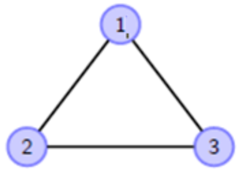
$\Rightarrow$



Graph Theory

- Another representation of a graph is given by (N, E), where E is the set of edges in the network.
  - *For directed graphs:* E is the set of "directed" edges, i.e., $(i, j) \in E$.
  - *For undirected graphs:* E is the set of "undirected" edges, i.e., $\{i, j\} \in E$.
- In Example 1, $E_d = \{(1, 2), (2, 3), (3, 1)\}$
- In Example 2, $E_u = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$



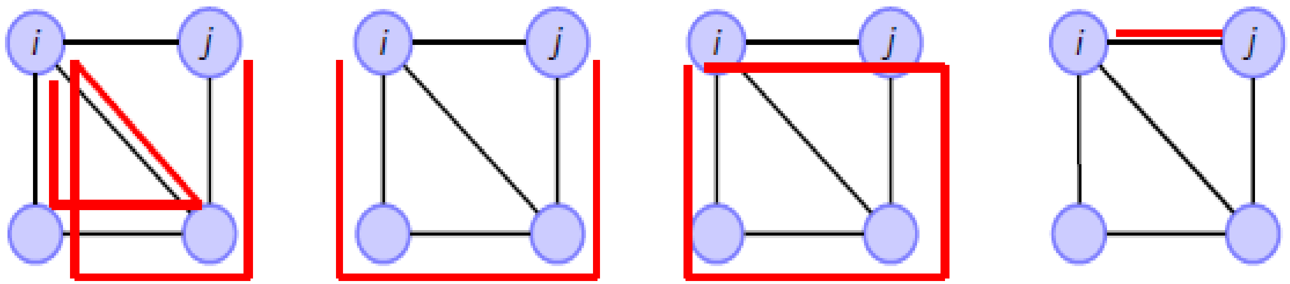


- When are directed/undirected graphs applicable?
  - Citation networks: directed
  - Friendship networks: undirected
- We will use the terms network and graph interchangeably.
- We will sometimes use the notation $(i, j) \in g$ (or $\{i, j\} \in g$) to denote $g_{ij} = 1$.

Graph Theory

- We consider "sequences of edges" to capture indirect interactions.
- For an undirected graph (N, g):
  - A walk is a sequence of edges $\{i_1, i_2\}, \{i_2, i_3\}, \ldots, \{i_{K-1}, i_K\}$.
  - A path between nodes i and j is a sequence of edges $\{i_1, i_2\}, \{i_2, i_3\}, \ldots, \{i_{K-1}, i_K\}$ such that $i_1 = i$ and $i_K = j$, *and each node in the sequence* $i_1, \ldots, i_K$ *is distinct.*
  - A cycle is a path with a final edge to the initial node.
  - A geodesic between nodes i and j is a "shortest path" (i.e., with minimum number of edges) between these nodes.
- A path is a walk where there are no repeated nodes.
- The length of a walk (or a path) is the number of edges on that walk (or path).
- For directed graphs, the same definitions hold with directed edges (in which case we say "a path from node i to node j").

Walks, Paths, and Cycles



walk — path between i and j — cycle — shortest path

- *Note:* Under the convention $g_{ii} = 0$, the matrix g^2 tells us number of walks of length 2 between any two nodes:
  - $(g \times g)_{ij} = \sum_k g_{ik} g_{kj}$
  - Similarly, g^k tells us number of walks of length k.

Graph Theory – Connectivity and Components

- An undirected graph is connected if every two nodes in the network are connected by some path in the network.
- Components of a graph (or network) are the distinct maximally connected subgraphs.
- A directed graph is
  - connected if the underlying undirected graph is connected (i.e., ignoring the directions of edges).
  - strongly connected if each node can reach every other node by a "directed path".

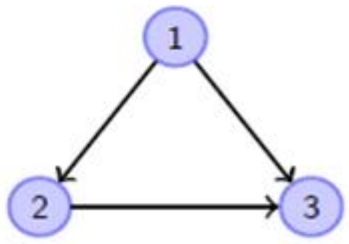


Figure: A directed graph that is connected but not strongly connected

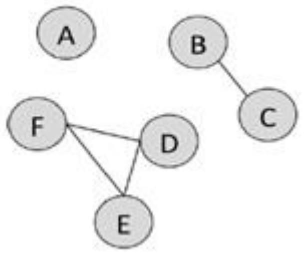
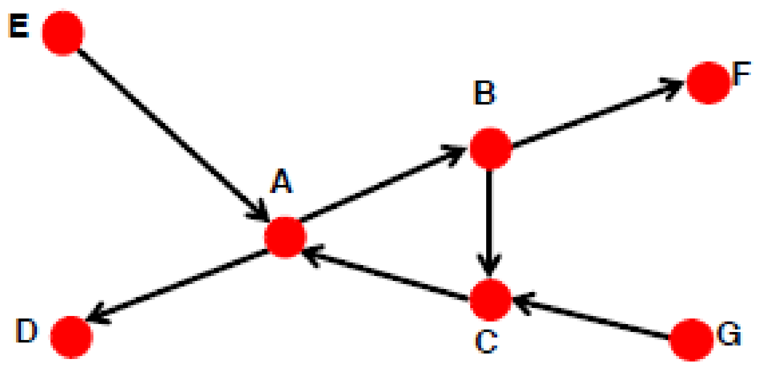


Figure: Containing 3 Components (A, BC, and DEF)

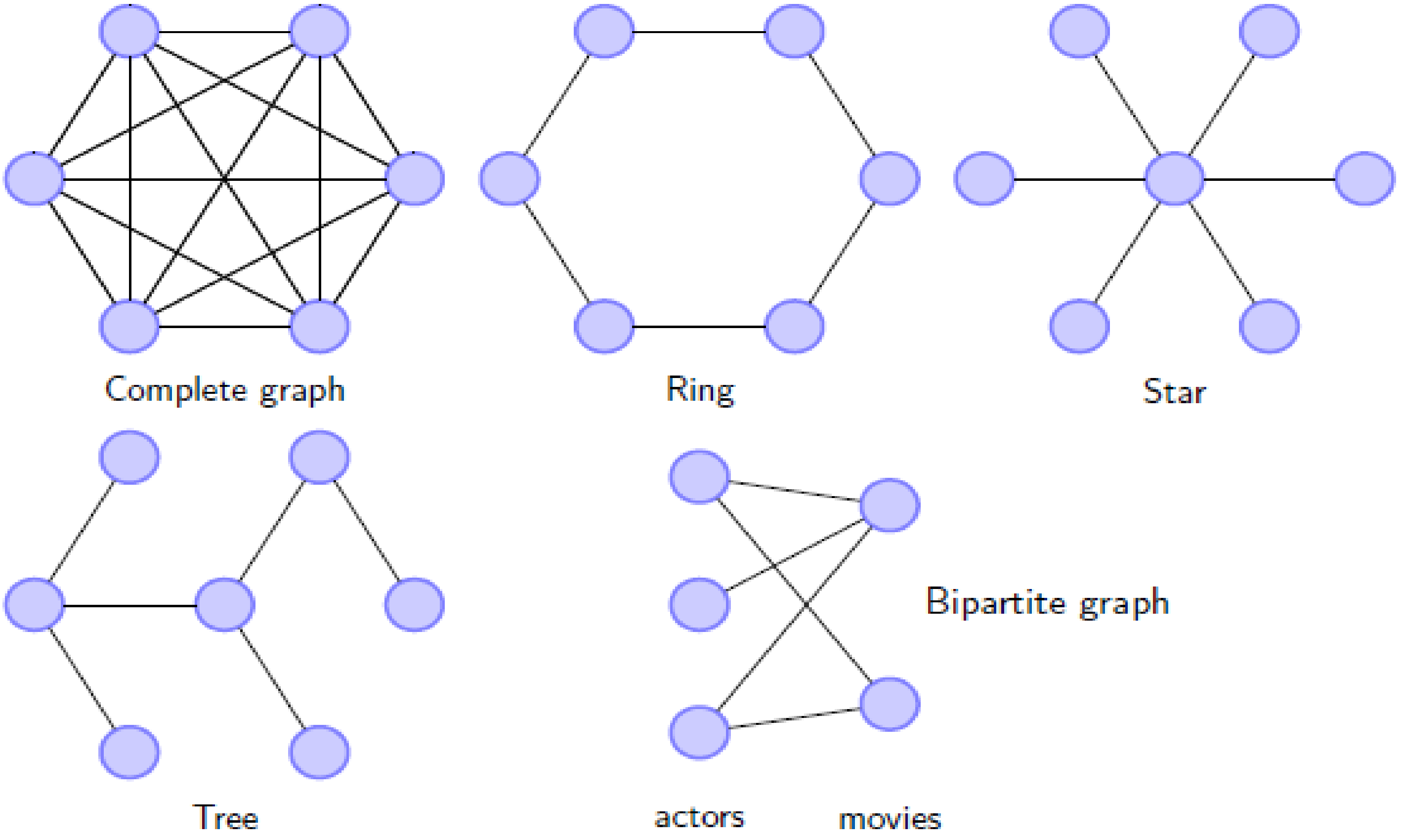
- **Strongly connected directed graph**
  - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
  - is connected if we disregard the edge directions



Graph on the left is connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions).

Graph Theory

- A tree is a connected (undirected) graph with no cycles.
  - A connected graph is a tree if and only if it has $n - 1$ edges.
  - In a tree, there is a unique path between any two nodes.

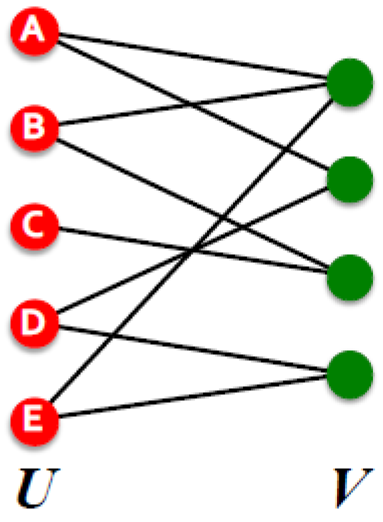


Bipartite graph

➢Bipartite graph 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 node in V*;** that is, *U and V are independent sets.*

➢**Examples:**

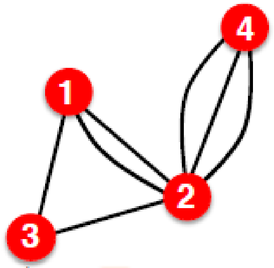
- Authors-to-papers (they authored)
- Actors-to-Movies (they appeared in)
- Users-to-Movies (they rated)



Multigraph

- A multigraph is a graph which is **permitted to have multiple edges** (**also called parallel edges**), that is, edges that have the same end nodes. Thus **two vertices may be connected by more than one edge.**

- In social networks parallel edges between any two nodes is called as multiplexity that is having more than one relationship.



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

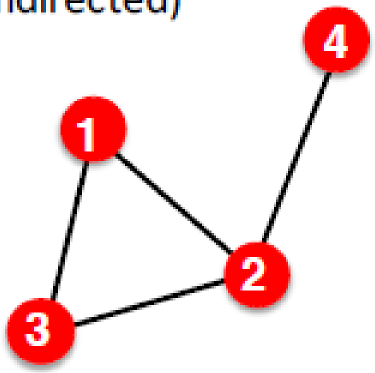
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

Examples: Communication, Collaboration

Graph Theory

- **Unweighted**
  (undirected)



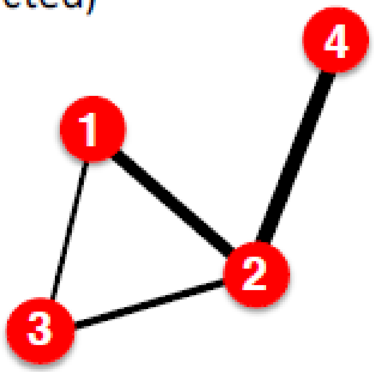
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad \bar{k} = \frac{2E}{N}$$

Examples: Friendship, Hyperlink

- **Weighted**
  (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

Examples: Collaboration, Internet, Roads

Edge Attribute - Possible options:

- **Weight** (e.g. frequency of communication)
- **Ranking** (best friend, second best friend…)
- **Type** (friend, relative, co-worker)
- **Sign**: Friend vs. Foe, Trust vs. Distrust
- …

Network - Graph Representations

- WWW →
- Facebook friendships →
- Citation networks →
- Collaboration networks →
- Mobile phone calls →
- Protein Interactions →

Network - Graph Representations

- WWW → directed multigraph with self-edges
- Facebook friendships → undirected, unweighted
- Citation networks → unweighted, directed, acyclic
- Collaboration networks → undirected multigraph or weighted graph
- Mobile phone calls → directed, (weighted?) multigraph
- Protein Interactions → undirected, unweighted with self-interactions

Neighborhood and Degree of a Node

- The neighborhood of node i is the set of nodes that i is connected to.
- For undirected graphs:
  - The degree of node i is the number of edges that involve i (i.e., cardinality of his neighborhood).
- For directed graphs:
  - Node i's in-degree is $\sum_j g_{ji}$.
  - Node i's out-degree is $\sum_j g_{ij}$.

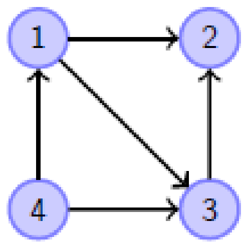


Figure: Node 1 has in-degree 1 and out-degree 2

SOCIAL NETWORK ANALYTICS

Adjacency Lists and Matrices, & Connectivity and Cohesion

Prakash C O
Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

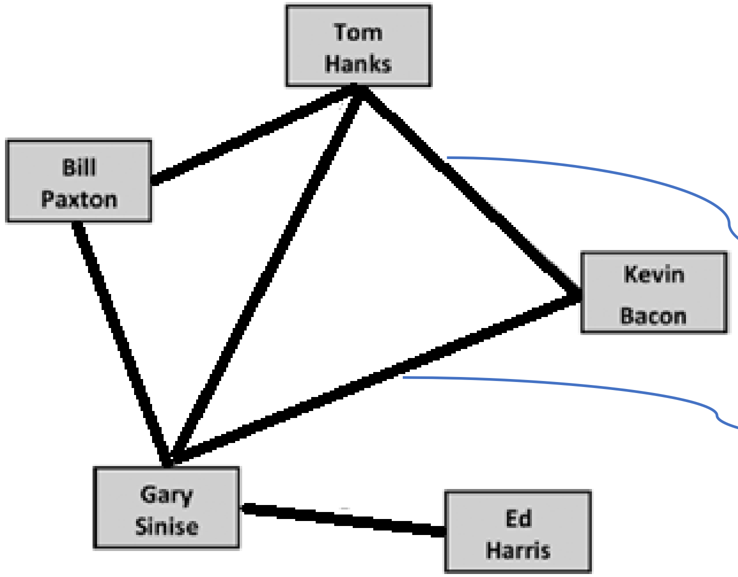
Adjacency Lists and Matrices, & Connectivity and Cohesion

Prakash C O

Department of Computer Science and Engineering

Adjacency List

- An adjacency list, **also called an edge list**, is one of the most basic and frequently used representations of a network.
- **Each edge in the network** is indicated by **listing the pair of nodes that are connected**.
- The adjacency list for the below Actors network is as follows:



Adjacency list :

(Tom Hanks, Bill Paxton)

(Tom Hanks, Gary Sinise)

(Tom Hanks, Kevin Bacon)

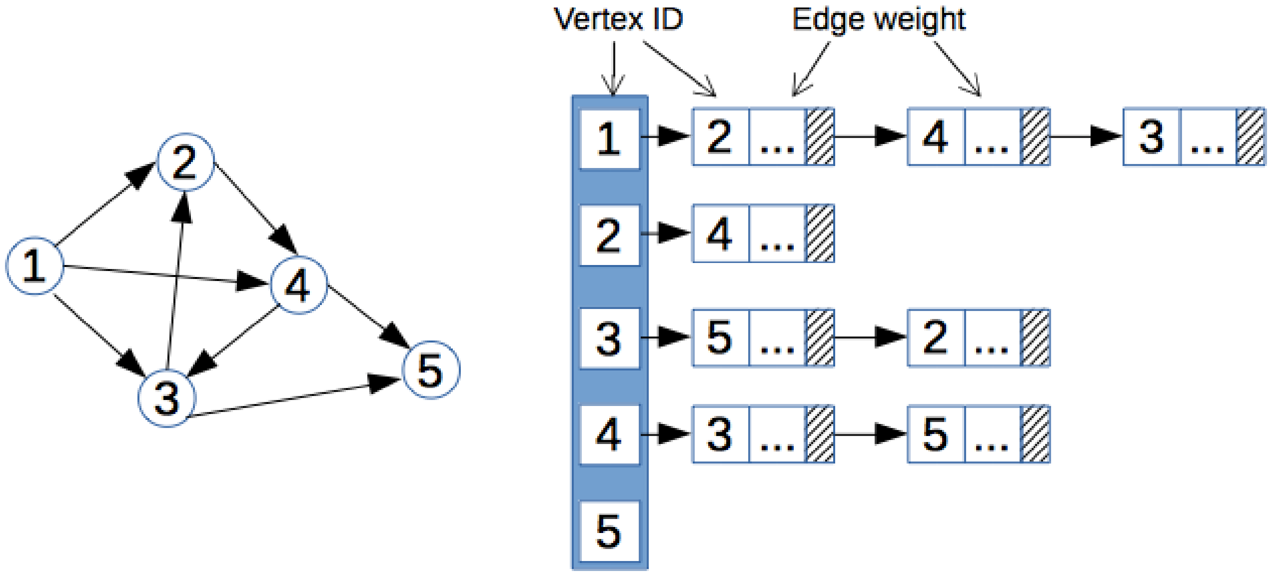
(Bill Paxton, Gary Sinise)

(Gary Sinise, Kevin Bacon)

(Gary Sinise, Ed Harris)

Adjacency List

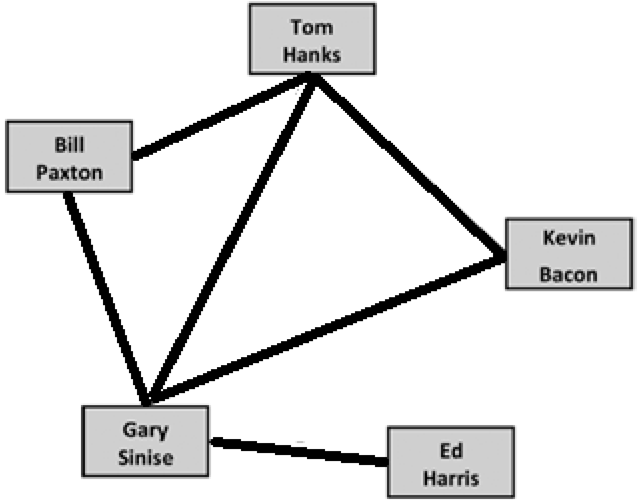
➢ Memory-efficient adjacency list implementation technique



- Easier to work with if network is large and sparse.
- Allows us to quickly retrieve all neighbors of a given node
- Adjacency lists are not well suited for parallelism since the lists require that we traverse the neighbors of a vertex sequentially.

Adjacency Matrix

- **For an unweighted graph containing n nodes**, construct an ***n × n*** **matrix of binary values** in which **location (i, j)** in the matrix is
  - **1** if there is a link from node i to node j
  - **0** otherwise
- **For an undirected graph the matrix is symmetric** and 0 along the diagonal. **For directed graphs the 1s can be in arbitrary positions.**

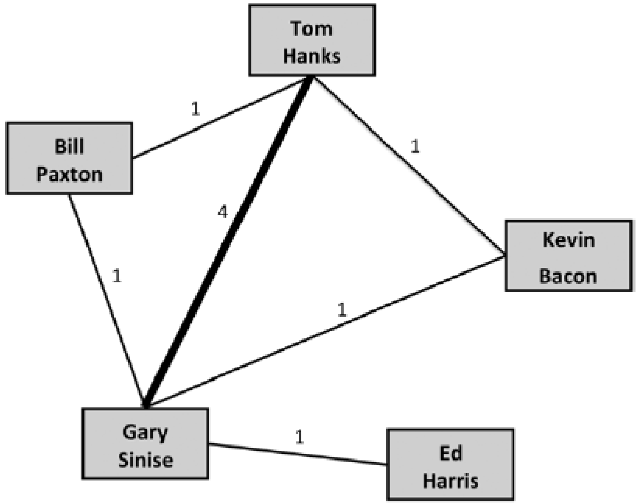


The Adjacency Matrix for the Apollo 13 Network

	Tom Hanks	Bill Paxton	Gary Sinise	Kevin Bacon	Ed Harris
Tom Hanks	0	1	1	1	0
Bill Paxton	1	0	1	0	0
Gary Sinise	1	1	0	1	1
Kevin Bacon	1	0	0	0	0
Ed Harris	0	0	1	0	0

Adjacency Matrix

➢The Adjacency Matrix for the **Apollo-13 Network with Edge Weights**



The Adjacency Matrix for the Apollo 13 Network with Edge Weights

	Tom Hanks	Bill Paxton	Gary Sinise	Kevin Bacon	Ed Harris
Tom Hanks	0	1	4	1	0
Bill Paxton	1	0	1	0	0
Gary Sinise	4	1	0	1	1
Kevin Bacon	1	0	0	0	0
Ed Harris	0	0	1	0	0

The disadvantage of adjacency matrices is their **space demand of $\Theta(n^2)$**.

Graphs are often sparse, with far fewer edges than $\Theta(n^2)$.

Connectivity and Cohesion

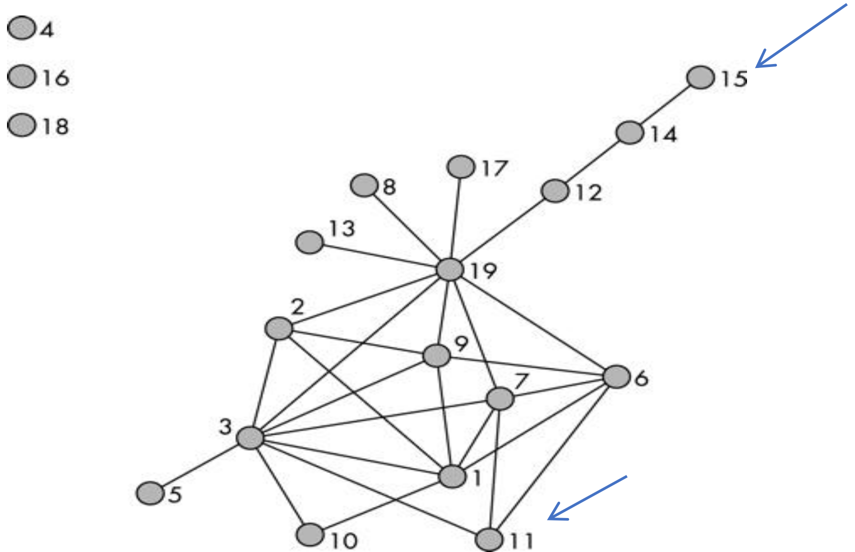
- **Cohesion describes the interconnectedness of actors in a network.**

- **There are 3-common measures of cohesion:**

  1. **Distance,**
  2. **Reachability** and
  3. **Density.**

Connectivity and Cohesion

- **Distance between two actors in a network** (or nodes in a graph) is calculated by **summing the number of distinct ties (lines) that exist along the shortest route between them.**
- In figure below actor 15 is a distance of 5 from actor 11. This is the notion of "degrees of separation" made familiar to many by a popular play.



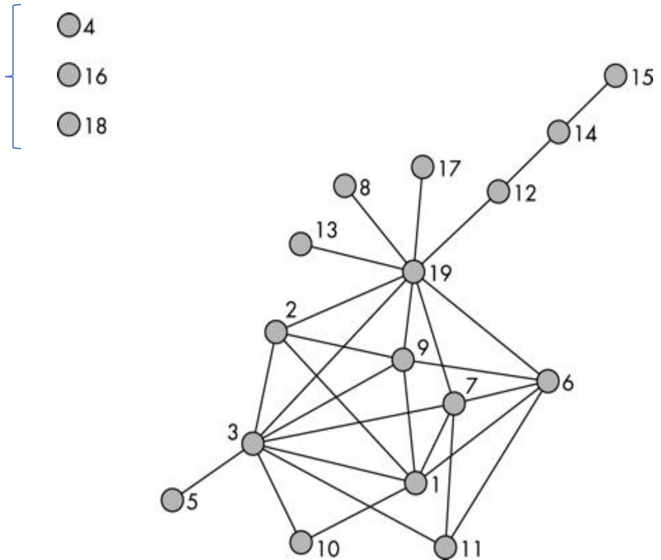
Connectivity and Cohesion

- **Reachability** measures **whether actors within a network are related, either directly or indirectly, to all other actors.**

  Actors who are not connected to any other actors are called ***isolates*.**

  With the exception of the three isolates (**actors 4, 16, and 18**), **all of the remaining actors in below figure can reach one another**.

Isolates



Connectivity and Cohesion

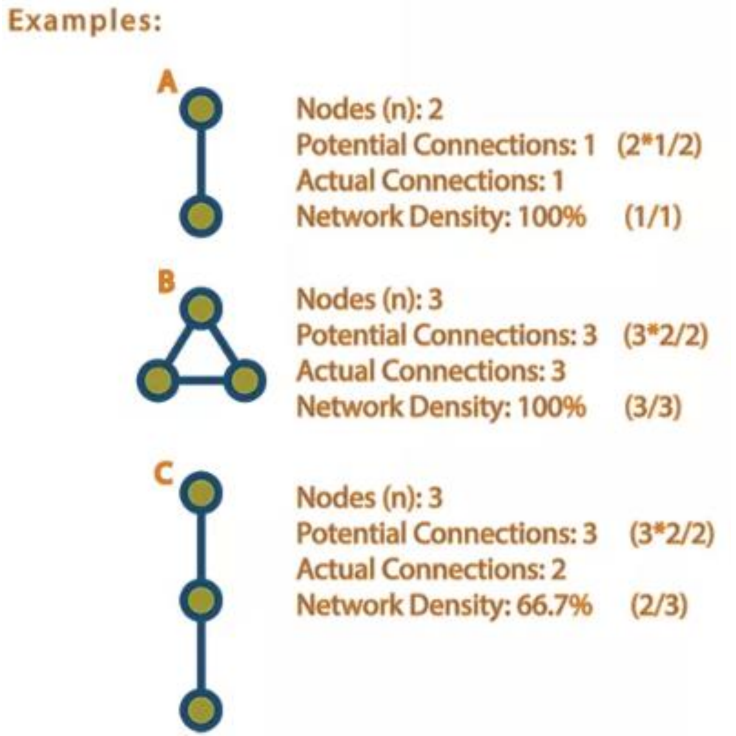
- **Density of a network** is **the total number of relational ties divided by the total possible number of relational ties.**

Network Density:

$$\frac{\text{Actual Connections}}{\text{Potential Connections}}$$

Potential Connections:

$$PC = \frac{n * (n-1)}{2}$$



- Density is one of the most basic measures in network analysis and one of the most commonly used notions in social epidemiology.

SOCIAL NETWORK ANALYTICS

Basic network structures and properties:

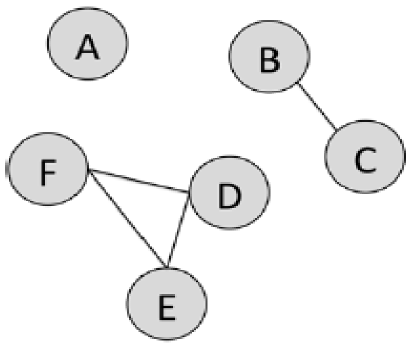
Clique, Core and Subnetwork

Prakash C O

Department of Computer Science and Engineering

Clique

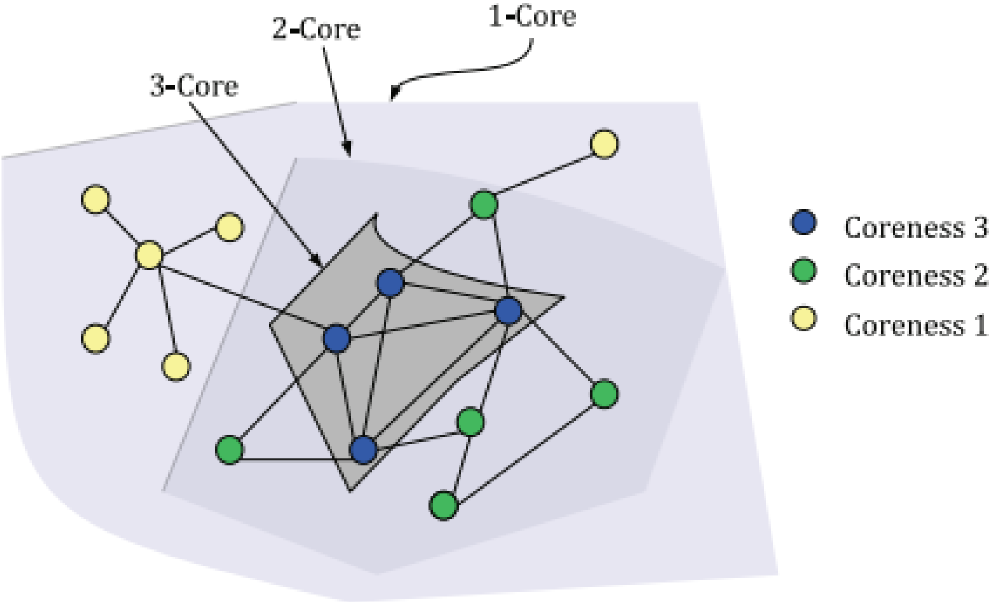
- **For a graph(or subgraph) to be a clique, every node must be connected to every other.**
- A clique is defined as **a maximal complete subgraph** of a given graph i.e., a group of people where everybody is connected directly to everyone else.
- In below Figure nodes D, E, and F form a clique.



- In social network, a clique represents a subset of people who all know each other, and algorithms for finding cliques can be used to discover these groups of mutual friends.

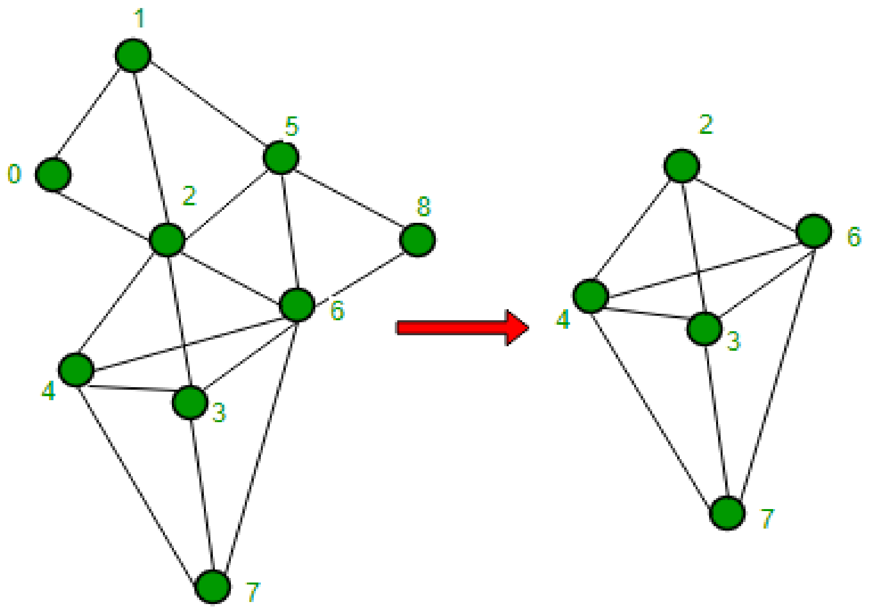
Core (K-core)

- The **k-core of a network** is the **maximal subgraph** in which **each node has at least k-neighbors.**
- The **k-core decomposition** is a process to find the k-core of a network .
- The most commonly used algorithm to perform k-core decomposition is a **Pruning process.**
  Pruning process recursively removes the nodes that have degrees less than k.



Core (K-core)

- **Given a graph G** and **an integer K**, K-cores of the graph are connected components that are left after all vertices of degree less than k have been removed.



Input Graph

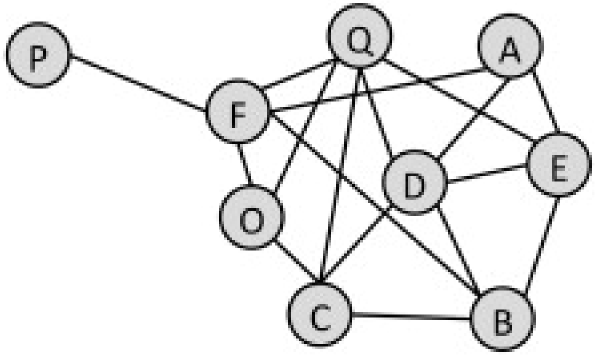
3-Cores

Core (K-core)

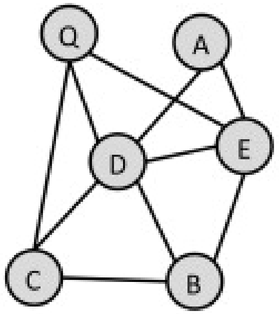
- **The k-core decomposition is widely accepted to reveal the network structure**.
- Many researchers have proposed a number of methods that can
  - identify the key nodes in the network
  - measure the influence of users in OSNs.
- Many researchers discovered that the **best communicators/ influential spreaders have been on k-core** on different social platforms, such as Twitter, Facebook, Livejournal and the American Physical Society.

Subnetwork(Subgraph)

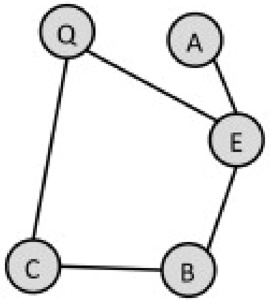
- **Subnetwork:** Its a subset of the nodes and edges in a graph.



Network/Graph



Subnet - 1



Subnet - 2

Subnetwork(Subgraph)

- **Singletons**(simplest subnetworks) **are nodes that have no edges.** In social networks, singletons are nodes are not very "social", but they are still part of a social network. In fact, it is very common to find singletons in online social networks.

- **Components** of a graph **are also sub-graphs** that are connected within, but disconnected between sub-graphs.

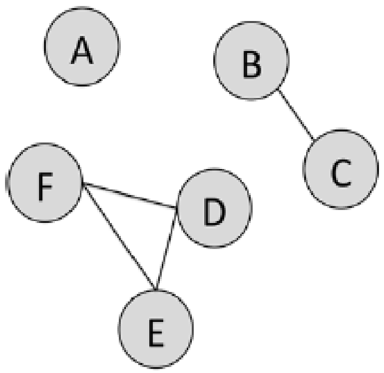
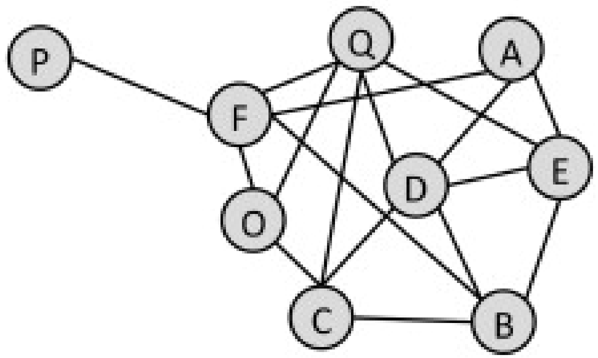


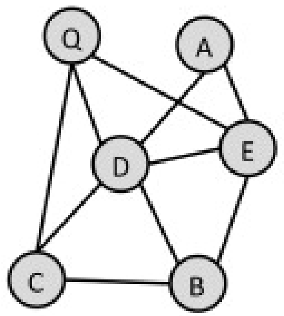
FIG: A social network with a singleton, dyad, and closed triad

Subnetwork(Subgraph)

- One of **the most important types of subnetwork** we study is the **egocentric network**.
- On Facebook and LinkedIn, Ego networks are simply described as "your network".



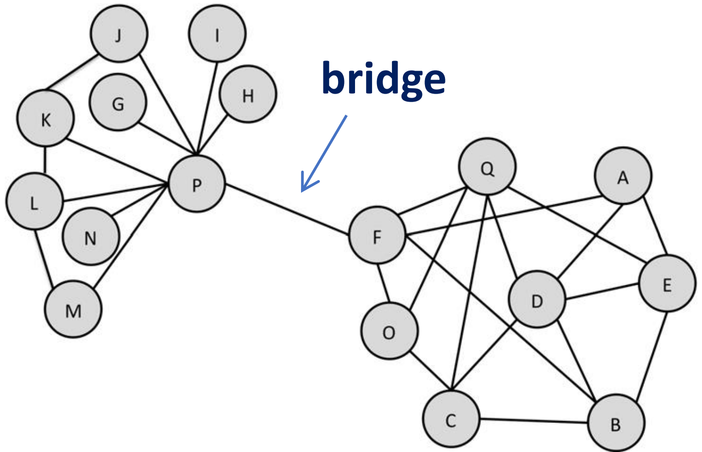
Whole network

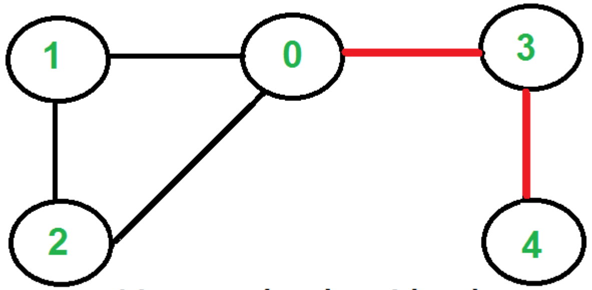


Ego network of node D

Bridges

- A bridge is an edge that connects two otherwise separate groups of nodes in the network.
- **A bridge is an edge that, if removed, will increase the number of connected components in a graph.**
- In Figure below, the edge between nodes P and F is a bridge.

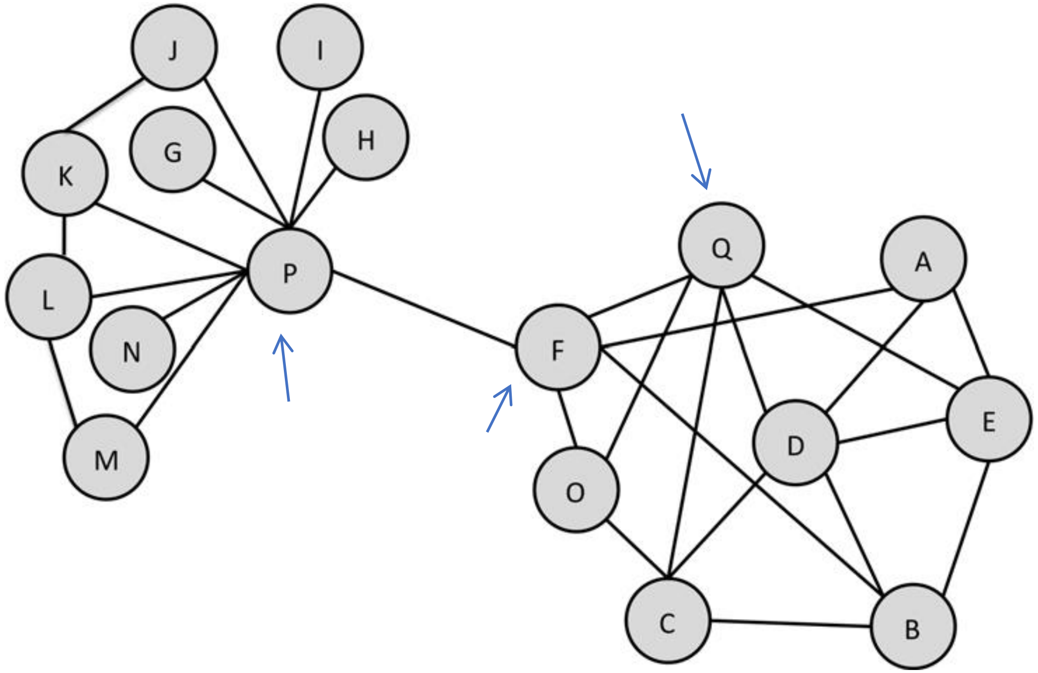




Bridges are (0, 3) and (3, 4)

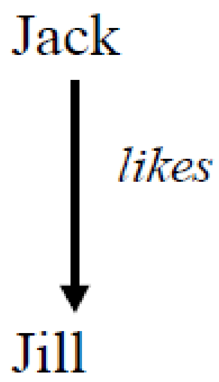
Hubs

- **Hubs are important nodes rather than edges**.
- **In networks, large-degree nodes are often referred to as *hubs*.**
- In Figure below node P would be a hub because it has many connections to other nodes.



TYPES OF NETWORKS : **One mode networks**

- **One mode networks** involve **relations among a single set of similar actors**, such as information exchange among physicians within a hospital.
- If a network contains **nodes of only one type**, it's called a *1-mode* network.
- **One mode networks:** Person to Person, URL to URL, Organization to organization ….

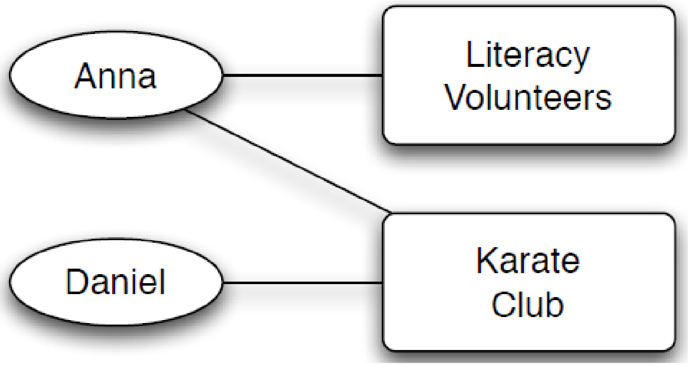


TYPES OF NETWORKS : **Two mode networks**

- **Two mode networks** involve **relations among two different sets of actors**.

  An example would be the analysis of a network consisting of private profit organizations and their links to non-profit agencies in a community.

- **Two mode networks are also used to investigate the relationship between a set of actors and a series of events.**



References

- Jure Leskovec, Stanford CS224W: Social and Information Network Analysis.
- http://snap.stanford.edu/class/cs224w-2016/handouts.html
- Lecture 2: Graph Theory and Social Networks, Daron Acemoglu and Asu Ozdaglar, MIT
- *"Analyzing the social web"*, Jennifer Golbeck, Morgan Kaufmann, 2013.

THANK YOU

Prakash C O
Department of Computer Science and Engineering

coprakasha@pes.edu

+91 98 8059 1946

PES UNIVERSITY ONLINE

Weighted and Unweighted Networks

- **In real-world networks, ties are often associated with weights** that differentiate them in terms of their **strength, intensity, or capacity**.
- In **infrastructure and information networks**, variations in the **strength of a tie** depend on the **flow of information, energy, people, and goods along that tie** (Barrat et al., 2004).
- In **Social Networks**, the **strength of social relationships** is a function of their **duration, frequency, emotional intensity, intimacy, and exchange of services.**

- ***Graphs* are visual representations of networks, displaying actors as *nodes* and the relational ties connecting actors as *lines*.**

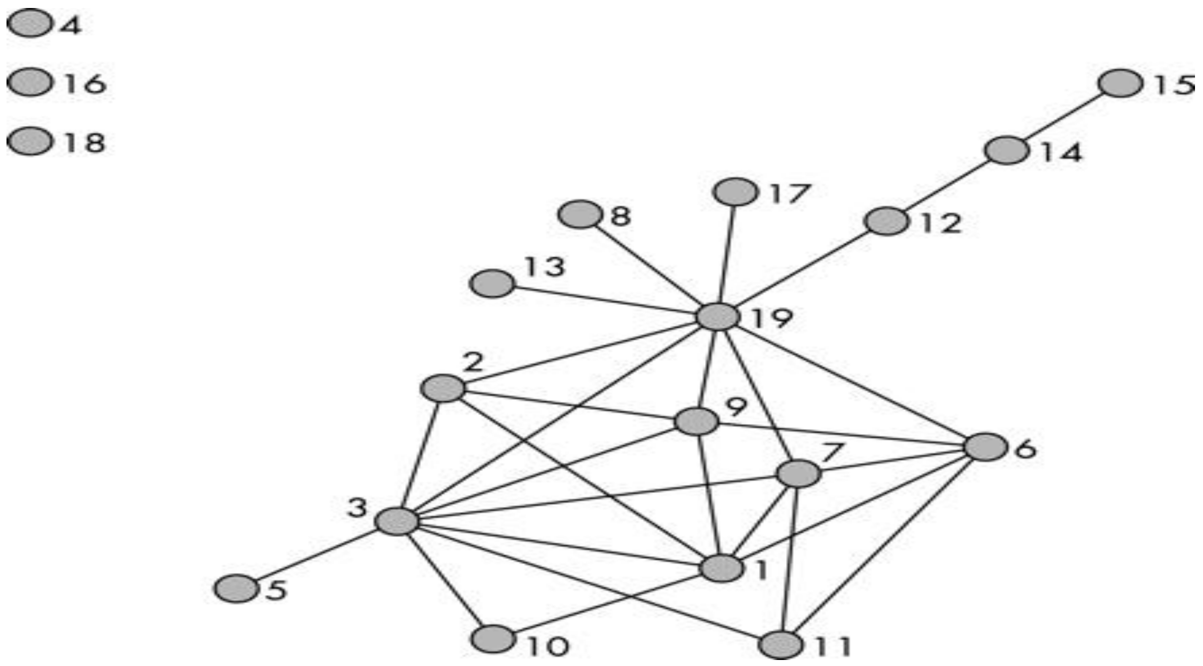


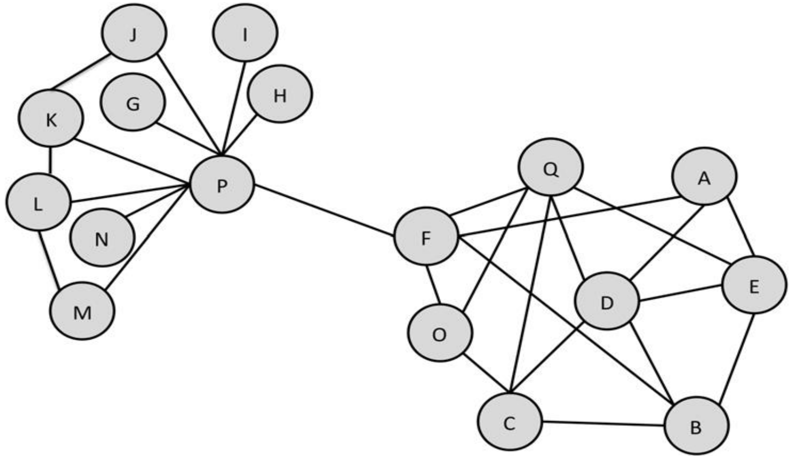
Figure 1: Graphical display of an inter organisational network with 19 actors.

Cluster

- We are also interested in clusters of nodes.

  In below Figure, we see a group of nodes to the lower right that have many connections between them.

- **The group is clearly more connected to one another than the graph is as a whole or compared to other subgraphs**. While **there is no strict definition of a cluster like there is for a clique**.



Cluster