Social and Information Networks

Module 1 - GraphsMatrices

Reference Book:

Wasserman Stanley, and Katherine Faust. (2009). Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences.

Graph

Graph

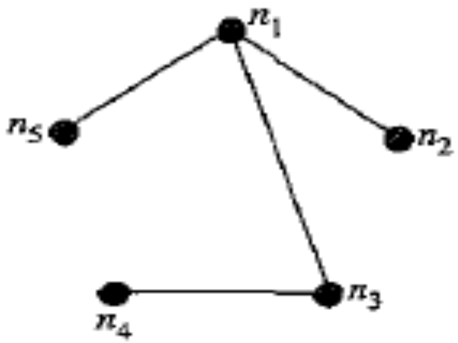
- A graph is a model for a social network
  - G ={N, L} ; N is set of nodes and L is set of lines

Subgraph:

- A graph, G_s is a subgraph of a graph G, if the set of nodes of G_s is a subset of the set of nodes of G, and the set of lines in G_s, is a subset of the lines in the graph G
- There may be lines in the graph between pairs of nodes in the subgraph that are not included in the set of lines in the subgraph.
- Any generic subgraph may not include all lines between the nodes in the subgraph

Graph

- Subgraphs of a Graph



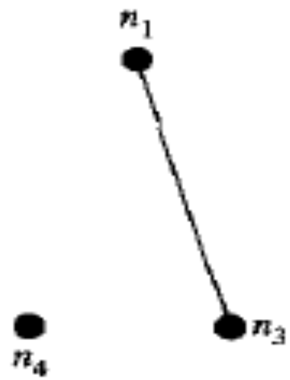
a.

$\mathscr{N} = \{n_1\ n_2\ n_3\ n_4\ n_5\}$

$\mathscr{L} = \{l_1\ l_2\ l_3\ l_4\}$

$l_1 = (n_1\ n_2) \qquad l_3 = (n_1\ n_5)$

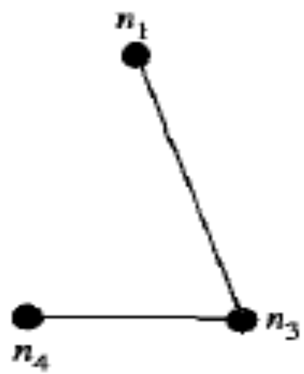
$l_2 = (n_1\ n_3) \qquad l_4 = (n_3\ n_4)$



b. subgraph

$\mathscr{N}_s = \{n_1\ n_3\ n_4\}$

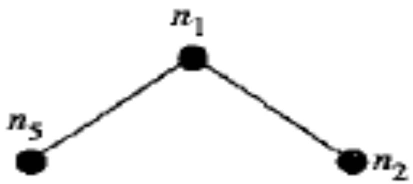
$\mathscr{L}_s = \{l_2\}$



c. subgraph generated by nodes $n_1\ n_3\ n_4$

$\mathscr{N}_s = \{n_1\ n_3\ n_4\}$

$\mathscr{L}_s = \{l_2\ l_4\}$



d. subgraph generated by lines $l_1\ l_3$

$\mathscr{N}_s = \{n_1\ n_2\ n_5\}$

$\mathscr{L}_s = \{l_1\ l_3\}$

Graph

Subgraph:

- Two special kinds of subgraphs
  - Node-generated subgraph : Take a subset of nodes and consider all lines that are between the nodes in the subset
  - Line-generated subgraph : Take a set of subset of lines and consider all nodes that are incident with the lines in the subset

Graph

Node-generated **Subgraph**:

- A subgraph, G_s is generated by a set of nodes, N_s, and line set L_s , where L_s includes all lines from L that are between pairs of nodes in N_s
- In a longitudinal study, some actors or subset of actors, might leave the network
- Analyses of the network might have to be restricted to the subset of actors for whom data are available for all time instances
- Widely used in the analysis of cohesive subgroups in networks (Subsets of actors among whom the ties are relatively strong, numerous)

Graph

Line-generated **Subgraph**:

- A subgraph, G_s is generated by a line set L_s and set of nodes, $N_{s,}$, where set of nodes N_s includes all nodes from N that are incident with line in L_s

Important feature of a subgraph: Maximal or not with respect to some property

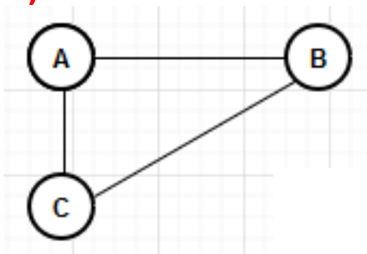
- – A subgraph is maximal with respect to a given property if that property holds for the subgraph but does not hold if any node or nodes are added to the subgraph

Graph

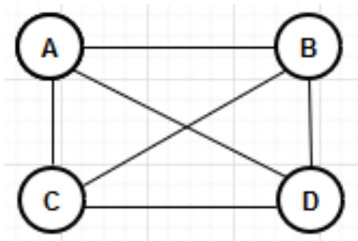
Important feature of a subgraph: Maximal or not with respect to some property

- A subgraph is maximal with respect to a given property if that property holds for the subgraph but does not hold if any node or nodes are added to the subgraph
- Graphs in a) and b) are complete graphs (i.e. degree of all nodes is n-1)
- Subgraph with nodes A,B,C is maximal complete in graph c) and d) but not in graph b)

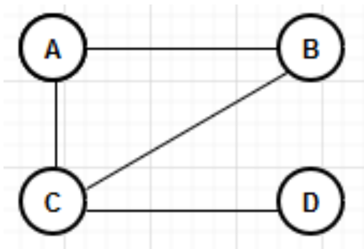
• Examples: a)



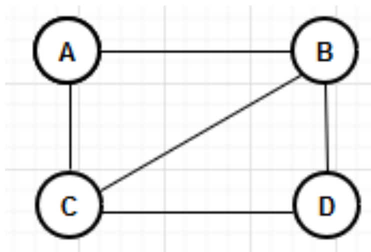
b)



c)



d)



Graph

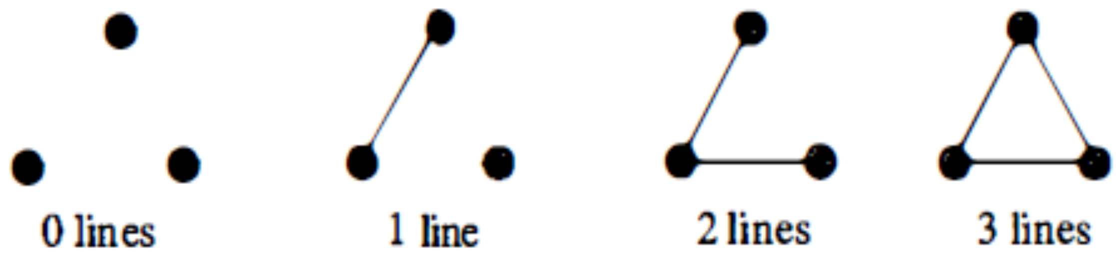
Dyad:

- Represents a pair of actors and the possible tie between them
- Node-generated subgraph consisting of a pair of nodes and the possible line between the nodes
- Two dyadic states for an undirected relation; the actors have a tie present, or they do not

Graph

Triad:

- Subgraph consisting of three nodes and the possible lines among them
- A triad may be in one of four possible states, depending on whether, zero, one, two, or three lines are present among the three nodes in the triad



Four possible triadic states in a graph

Graph

Triad:

- According to Granovetter's model,
  - Triad with two lines present and one line absent is a forbidden triad
  - If lines represent strong ties between actors, then actor i has a strong tie with actor j, and actor j in turn has a strong tie with actor k, it is unlikely that the tie between actor i and actor k will be absent

Graph

Nodal Degree:

- Degree of a node
  - denoted by $d(n_i)$ is the number of lines that are incident with it
  - number of nodes adjacent to it
  - A count that ranges from a minimum of 0 and to a maximum of g – 1 (a node is adjacent to all other nodes)
- A node with degree equal to 0 is called an isolate
- Degree of a node is a measure of the "activity" of the actor it represents

Graph

Nodal Degree:

- Mean nodal degree $$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g}$$
- d-regular graph : Degrees of all of the nodes are equal, where d is the constant value;
  - For all the degrees $d(n_i)$, $d(n_i) = d$, for all i and some value d
  - d-regularity -> a measure of uniformity

Graph

Nodal Degree:

- Variance of the degrees, $$S_D^2 = \frac{\sum_{i=1}^{g}(d(n_i) - \bar{d})^2}{g}$$

- d-regular graph has zero variance

- Variability : Actors represented by the nodes differ in "activity," as measured by the number of ties they have to others

Graph

- Order of a graph is number of vertices contained in the graph.
- Size/Length of a graph is number of edges contained in the graph.
- Eccentricity: The maximum distance between a vertex to all other vertices is considered as the eccentricity of the vertex.
- Radius of a graph is the minimum eccentricity of any vertex in the graph.
  - Smallest distance between any pair of vertices
- Diameter of a graph is the maximum eccentricity of any vertex in the graph.
  - Largest distance between any pair of vertices

Graph

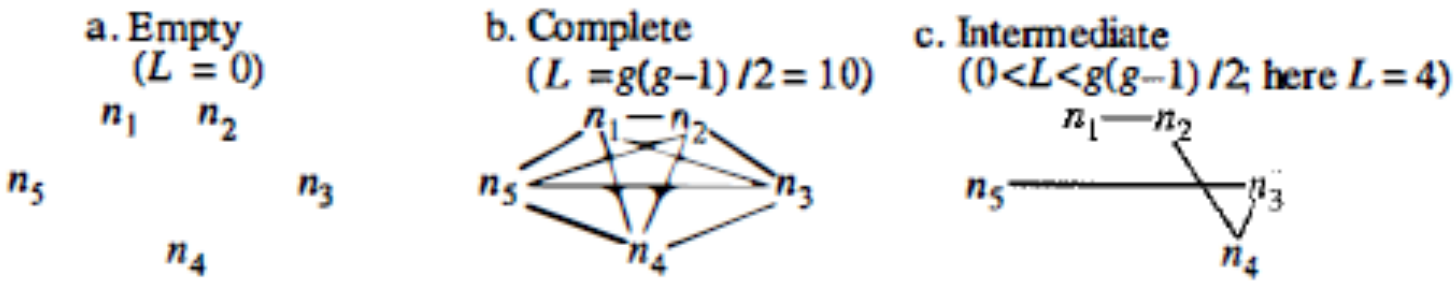
Density of a graph:

- Proportion of possible lines that are actually present in the graph
- Maximum possible number of lines in a graph with g nodes is $m = g(g - 1)/2$
- Ratio of the number of lines present, L, to the maximum possible
  - Density of a graph, $$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)}$$
- Density ranges from 0 (no lines present) to 1 (all lines are present)

Graph

Density of a graph:

- A Complete graph contains all g(g - 1)/2 possible lines, density is equal to 1, and all nodal degrees are equal to g – 1



Complete and empty graphs

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g}$$

- Alternatively, density of a graph, $\Delta = \dfrac{\bar{d}}{(g-1)}$
  - Average proportion of lines incident with nodes in the graph

Graph

Density of a subgraph:

- Number of lines present in the subgraph, divided by the number of lines that could be present in the subgraph
  - density of a subgraph, $$\Delta_s = \frac{2L_s}{g_s(g_s - 1)}$$
- Expresses the proportion of ties that are present among a subset of the actors in a network
- Used to evaluate the cohesiveness of subgroups

Graph (Walks, Trails and Paths)

- Example: A study to understand the communication of information among employees in an organization. An important consideration is
  - whether information originating with one employee could eventually reach all other employees, and if so,
  - how many lines it must traverse in order to get there
  - whether there are multiple routes that a message might take to go from one employee to another, and
  - whether some of these paths are more or less "efficient"

Graph (Walks, Trails and Paths)

Walk in a graph:

- A sequence of nodes and lines, starting and ending with nodes, in which each node is incident with the lines following and preceding it in the sequence
- An alternating sequence of incident nodes and lines beginning and ending with nodes
- Some nodes may be included more than once
- Some lines may be included more than once
- Length of a walk: Number of occurrences of lines in it
- If a line is included more than once in the walk, it is counted each time it occurs
- A walk, W, may be described by listing the nodes involved and excluding the lines
  - Walk, $W = n_1\ l_2\ n_4\ l_3\ n_2\ l_3\ n_4$
    - may be written briefly as

    Walk, $W = n_1\ \ n_4\ \ n_2\ \ n_4$

Graph (Walks, Trails and Paths)

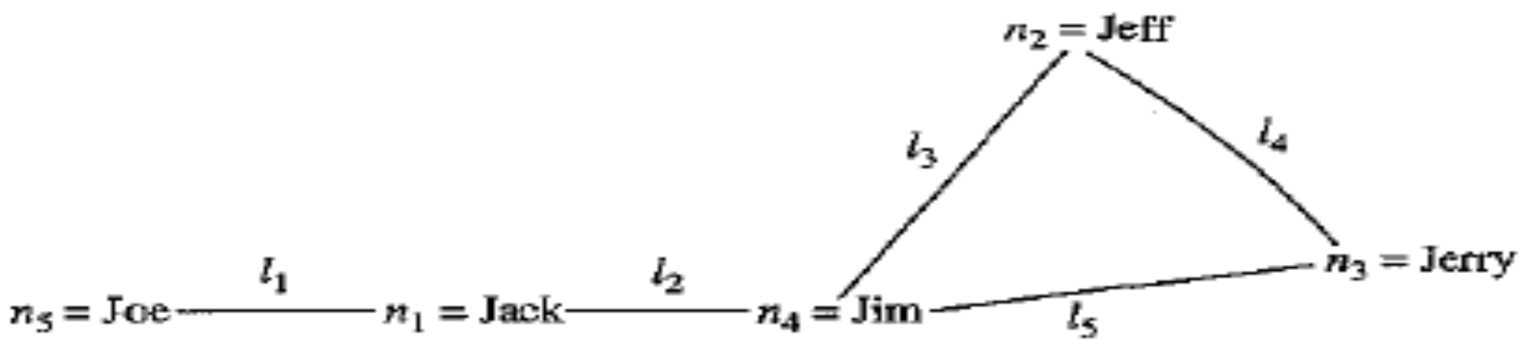
Walks in a graph:

- The starting node of W is the origin and ending node in the terminus of W
- Inverse of a walk, W^{-1}, is the walk W listed in exactly the opposite order, using the same nodes and lines

A walk would be $W = n_1\ l_2\ n_4\ l_3\ n_2\ l_3\ n_4$

A trail would be $W = n_4\ l_3\ n_2\ l_4\ n_3\ l_5\ n_4\ l_2\ n_1$

A path would be $W = n_1\ l_2\ n_4\ l_3\ n_2$



Walks, trails, and paths in a graph

Graph (Walks, Trails and Paths)

Trails:

- Walks with special characteristics
- A walk in which all of the lines are distinct, though some node(s) may be included more than once
- Length : Number of lines in it

Paths:

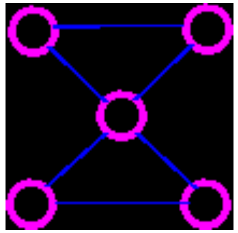
- Walks with special characteristics
- A walk in which all nodes and all lines are distinct
- Length : Number of lines in it
- Every path is a trail and every trail is a walk
- Path is likely shorter compared to a walk or a trail

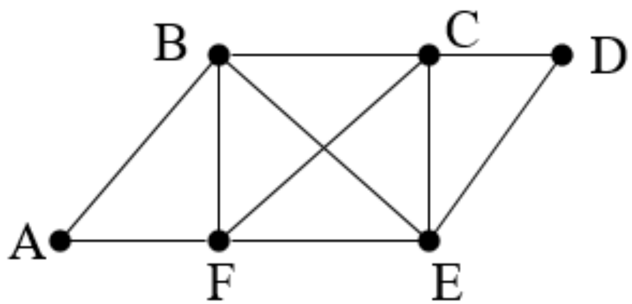
Graph (Walks, Trails and Paths)

- Closed Walk: A walk that begins and ends at the same node
- Cycle: A closed walk of at least three nodes in which all lines are distinct, and all nodes except the beginning and ending nodes are distinct
- A graph that contains no cycles is called acyclic
- Tour: A closed walk in which each line in the graph is used at least once
- Eulerian trails are special closed trails that include every line exactly once
- Hamiltonian cycle : Every node in the graph is included exactly once

Examples

- Examples

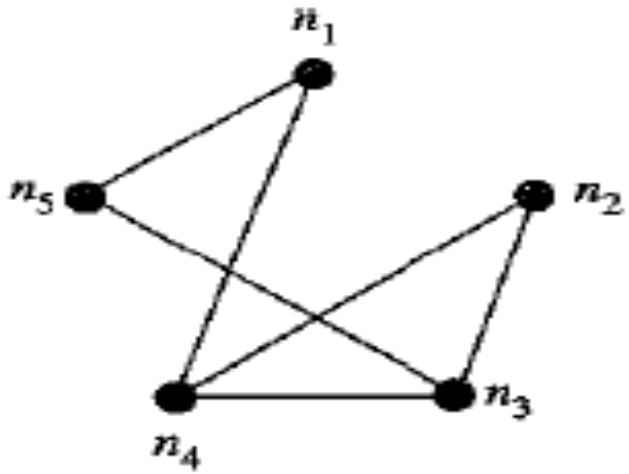




a) A B E D C F A

Graph (Walks, Trails and Paths)

- Example:



Tour $n_3\ n_2\ n_4\ n_3\ n_5\ n_1\ n_4\ n_3$

Cycles $n_5\ n_1\ n_4\ n_3\ n_5$

$n_2\ n_3\ n_4\ n_2$

$n_2\ n_4\ n_1\ n_5\ n_3\ n_2$

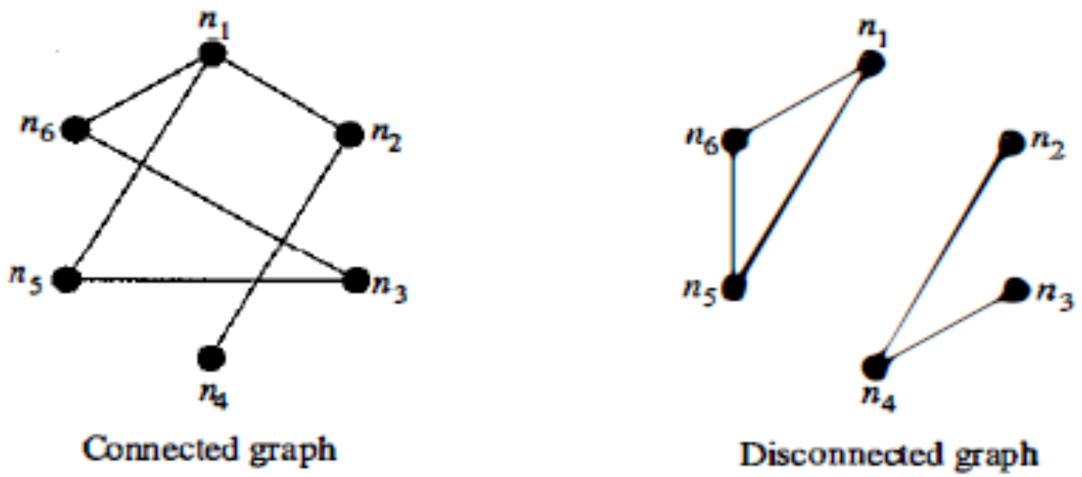
Closed walk $n_5\ n_1\ n_4\ n_3\ n_2\ n_4\ n_1\ n_5$

Closed walks and cycles in a graph

Graph

Connected Graphs :

- A graph is connected if there is a path between every pair of nodes in the graph.
- All pairs of nodes are reachable.



A connected graph and a graph with components

Graph

Components :

- Nodes in a disconnected graph may be partitioned into two or more subsets in which there are no paths between the nodes in different subsets.
- A component is a maximal connected subgraph [cannot be made larger and still retain its property]
- Connected graph contains one component.
- Disconnected graph contains more than one component.

Graph

Geodesics, Distance, and Diameter

- Several paths exists between a given pair of nodes, paths differ in length
- Geodesic : A shortest path between two nodes
- If there is more than one shortest path between a pair of nodes, then there are two (or more) geodesics between the pair
- Geodesic distance between two nodes : Length of a geodesic between them
- Geodesic between n_i and n_j is equal to the Geodesic between n_j and n_i ; $d(i, j) = d(j, i)$

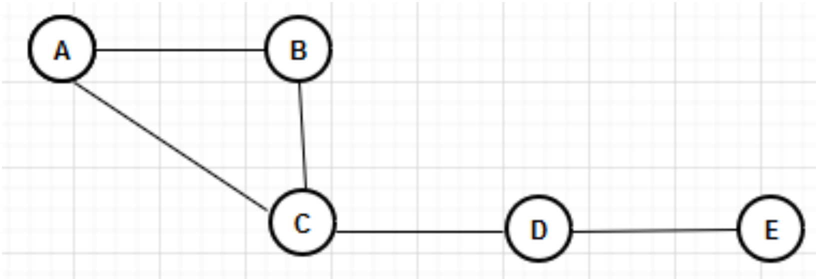
Graph

Geodesics, Distance, and Diameter

- Eccentricity of a Node: Consider the geodesic distances between a given node and the other g - 1 nodes in a connected graph
- Eccentricity or association number of a node is the largest geodesic distance between that node and any other node
- Formally, eccentricity of node n_i in a connected graph is equal to the maximum d(i,j), for all j
- Eccentricity of a node can range from a minimum of 1 (if a node is adjacent to all other nodes in the graph) to a maximum of g - 1

Graph

- Example:



  - Radius is 2
  - Diameter is 3

  - Eccentricity of
    - A is 3
    - B is 3
    - C is 2
    - D is 2
    - E is 3

Geodesic Matrix

	A	B	C	D	E
A	-	1	1	2	3
B	1	-	1	2	3
C	1	1	-	1	2
D	2	2	1	-	1
E	3	3	2	1	-

Graph

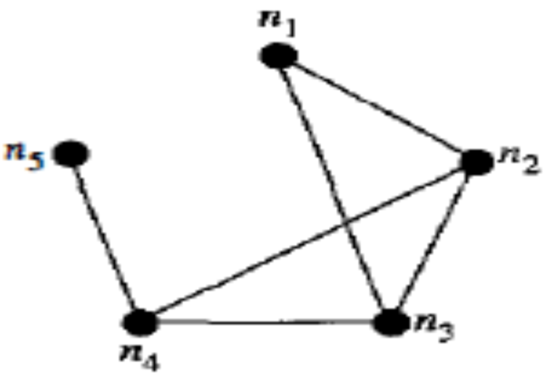
Geodesics, Distance, and Diameter

Diameter of a Graph:

- Consider the largest eccentricity of any node
- The diameter of a connected graph is the length of the largest geodesic between any pair of nodes (equivalently, the largest nodal eccentricity).
- Formally, the diameter of a connected graph is equal to the maximum d(i,j), for all i and j
- The diameter of a graph can range from a minimum of 1 (if the graph is complete) to a maximum of g - 1.
- If a graph is not connected, its diameter is infinite (or undefined) since the geodesic distance between one or more pairs of nodes in a disconnected graph is infinite.
- Diameter quantifies how far apart the farthest two nodes in the graph are.

Graph

- Example:



Geodesic distances

$d(1,2) = 1$
$d(1,3) = 1$
$d(1,4) = 2$
$d(1,5) = 3$
$d(2,3) = 1$
$d(2,4) = 1$
$d(2,5) = 2$
$d(3,4) = 1$
$d(3,5) = 2$
$d(4,5) = 1$

Diameter of graph = max $d(i,j) = d(1,5) = 3$

Graph showing geodesics and diameter

Graph

Diameter of a Subgraph:

- Consider a (node-generated) subgraph with node set Ns and line set Ls, containing all lines from Ls between pairs of nodes in Ns.
- Distance between a pair of nodes within the subgraph is defined for paths containing nodes from Ns, and lines from Ls
- Distance between nodes nj and nj in the subgraph is the length of the shortest path between the nodes within the subgraph
- Diameter of a sub graph is the length of the largest geodesic within the subgraph