Social and Information Networks

Module 1 - GraphsMatrices

Reference Book:

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

Directed Graph

Directed Walk

- A directed walk is a sequence of alternating nodes and arcs so that each arc has its origin at the previous node and its terminus at the subsequent node
- Length of a directed walk : Number of instances of arcs in it

Directed Trail:

- A directed walk in which no arc is included more than once

Directed path:

- A directed walk in which no node and no arc is included more than once
- Directed Path joining nodes n_i and n_j is a sequence of distinct nodes, where each arc has its origin at the previous node, and its terminus at the subsequent node
- Consists of all arcs "pointing" in the same direction

Directed Graph

SemiWalk

- A walk in which the arc between previous and subsequent nodes in the sequence may go in either direction
- A semiwalk joining nodes n_i, and n_j is a sequence of nodes and arcs in which successive pairs of nodes are incident with an arc from the first to the second, or by an arc from the second to the first
  - For all successive pairs of nodes, the arc between adjacent nodes may be either $< n_i, n_j >$ or $< n_j, n_i >$.
  - Direction of the arc is irrelevant
- Length of a semiwalk : Number of instances of arcs in it

Directed Graph

Semipath :

- A semipath joining nodes n_i and n_j is a sequence of distinct nodes, where all successive pairs of nodes are connected by an arc from the first to the second, or by an arc from the second to the first for all successive pairs of nodes
- Direction of the arc is irrelevant
- Length of a semipath : Number of arcs in it
- Every path is a semipath, but not every semipath is a path

Directed Graph

Closed walks:

- A walk that begins and ends at the same node

Cycle :

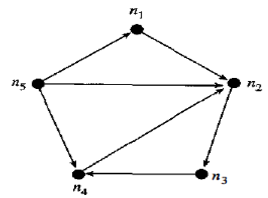
- A closed directed walk of at least three nodes in which all nodes except the first and last are distinct
- Arcs must all "point" in the same direction

Semicycle :

- A closed directed semiwalk of at least three nodes in which all nodes except the first and last are distinct
- Arcs may go in either direction
- Used to study structural balance and clusterability

Directed Graph

- Example



Directed walk	$n_5\ n_1\ n_2\ n_3\ n_4\ n_2\ n_3$
Directed path	$n_5\ n_4\ n_2\ n_3$
Semipath	$n_1\ n_2\ n_5\ n_4\ n_3$
Cycle	$n_2\ n_3\ n_4\ n_2$
Semicycle	$n_1\ n_2\ n_5\ n_1$

Directed walks, paths, semipaths, and semicycles

Directed Graph

Reachability and Connectivity in Digraphs:

•Pairs of Nodes: A pair of nodes is reachable if there is a path between them

•If there is a directed path from n_i to n_j, then node n_j is reachable from node n_i

•Considering both paths and semipaths between pairs of nodes, there are four different ways that two nodes can be connected by a path, or semipath. They are

(i) Weakly connected
(ii) Unilaterally connected
(iii) Strongly connected
(iv) Recursively connected

Directed Graph

Connectivity in Digraphs:

- A pair of nodes, n_i and n_j is:
  - (i) Weakly connected if they are joined by a semipath
  - (ii) Unilaterally connected if they are joined by a path from n_i to n_j or a path from n_j to n_i
  - (iii) Strongly connected if there is a path from n_i to n_j and a path from n_j to n_i; the path from n_i to n_j may contain different nodes and arcs than the path from n_j to n_i
  - (iv) Recursively connected if they are strongly connected, and the path from n_i to n_j uses the same nodes and arcs as the path from n_j to n_i, in reverse order
- These forms of connectivity are increasingly strict
- Any strict form implies connectivity of any less strict form
- Example: Any two nodes that are recursively connected are also strongly connected, unilaterally connected and weakly connected

Directed Graph -Connectivity

Digraph Connectedness:

- If a digraph is connected, then it is connected by one of the four different kinds of connectivity ; otherwise, it is not connected.

A directed graph is:

(i) Weakly connected if all pairs of nodes are weakly connected
- All pairs of nodes are connected by a semipath

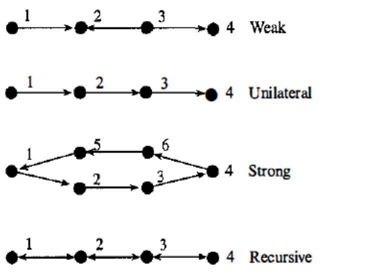
(ii) Unilaterally connected if all pairs of nodes are unilaterally connected
- Between each pair of nodes there is a directed path from one node to the other; in other words at least one node is reachable from the other in the pair.

(iii) Strongly connected if all pairs of nodes are strongly connected
- Each node, in each pair, is reachable from the other ; there is a directed path from each node to each other node.

(iv) Recursively connected if all pairs of nodes are recursively connected
- Each node, in each pair, is reachable from the other, and the directed paths contain the same nodes and arcs, but in reverse order

- Every strongly connected digraph is unilaterally connected, but the reverse is not true

Directed Graph -Connectivity



Different kinds of connectivity in a directed graph

Directed Graph

Geodesics, Distance and Diameter:

- Geodesic: Shortest path between two nodes
- Geodesic Distance: Length of the shortest path between the two nodes
  - Represented by $d(i,j)$ [length of geodesic from node n_i to node n_j]
- Path from node n_i to node n_j may be different from the path from node n_j to node n_i
- Geodesic is useful if there is a path from each node to each other node in the graph
- Diameter: Length of the longest geodesic between any pair of nodes
  - Diameter of a weakly or unilaterally connected directed graph is undefined

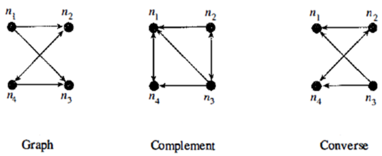
Directed Graph

Special Kinds of Directed Graphs

- Complement of a Digraph (G_d^c)
  - Complement, G_d^c, has the same set of nodes as G_d
  - if the arc < n_i , n_j > is in G_d then the arc < n_i , n_j > is not in G_d^c
  - if the arc < n_i , n_j > is not in G_d then the arc < n_i , n_j > is in G_d^c
- Converse of a Digraph (G_d')
  - Converse, G_d', has the same set of nodes as G_d
  - Obtained from G_d by reversing the direction of all arcs
  - arcs in the converse connect the same pairs of nodes as the arcs in the digraph with direction of arc reversed
  - An arc in the digraph from n_i , n_j becomes an arc in the converse from n_j to n_i

Directed Graph

- Examples: Converse may be used to represent relations that have "opposites"
  - A digraph represents a dominance relation (example, n_i "wins over" n_j)
  - Its converse would represent the submissive relation (n_j "loses to" n_i)
- Complement of a digraph might be used to represent the absence of a tie (example: Complement of "likes" is "does not like"



Converse and complement of a directed graph

Directed Graph

Tournament:

- Represents a set of actors competing in some event(s) and a relation indicating superior performances or "beats" in competition
- If team n_i beats team n_j, an arc is directed from n_i to n_j
- Round-robin tournament : Each team plays with each other team exactly once

Signed graph

Signed graph :

- A relational tie can be interpreted as being either positive or negative in affect, evaluation, or meaning
  - Example: Relations "is allied with" and "is at war with" among countries.
- Such relations can be represented as a signed graph or as a signed directed graph
- Lines carry the additional information of a valence : a positive or negative sign
- A signed graph, $G_{\pm}(N,L,V)$, consists of three sets of information :
  - a set of nodes, $N = \{n_1, n_2, n_3, \ldots, n_g\}$
  - a set of lines, $L = \{l_1, l_2, l_3, \ldots, l_L\}$
  - a set of valences, $V = \{v_1, v_2, v_3, \ldots, v_L\}$ attached to the lines
- Each line is associated with a valence, v_k
- A line, $l_k = (n_i, n_j)$ is assigned the valence
  - $v_k = +$, if the tie between actors i and j is positive in meaning
    (or)
  - $v_k = -$ if the tie between the actors I and j is negative

Signed graph

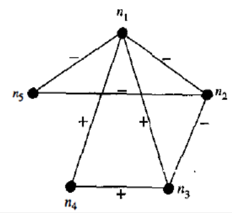
- Complete signed graph: All unordered pairs of nodes are included in the set of lines.
  - All lines have a valence either "+" or "-"
- Dyad: Dyad between a pair of actors is in one of three states
  - a positive line between them
  - a negative line between them
  - no line between them
- In a complete signed graph
  - each dyad is in one of two states, either "+" or "-"
- A triad
  - may be in one of four possible states, depending on whether zero, one, two, or three positive (or negative) lines are present among the three nodes

Signed graph

Cycle:

- A closed walk in which all nodes except the beginning and ending node are distinct.
- Each line of a cycle in a signed graph is either "+" or "-"
- Sign of a cycle: Defined as the product of the signs of the lines included in the cycle.

Signed Graph - Example



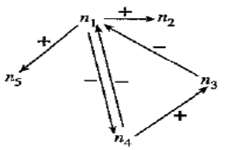
$l_1 = (n_1\ n_2)$	$v_1 = -$
$l_2 = (n_1\ n_3)$	$v_2 = +$
$l_3 = (n_1\ n_4)$	$v_3 = +$
$l_4 = (n_1\ n_5)$	$v_4 = -$
$l_5 = (n_2\ n_3)$	$v_5 = -$
$l_6 = (n_2\ n_5)$	$v_6 = -$
$l_7 = (n_3\ n_4)$	$v_7 = +$
Cycle	Sign of cycle
$n_1\ n_2\ n_5\ n_1$	$- \times - \times - = -$
$n_3\ n_4\ n_1\ n_3$	$+ \times + \times + = +$
$n_1\ n_2\ n_3\ n_1$	$- \times - \times + = +$

Example of a signed graph

Signed Directed Graph

- A signed directed graph is $G_{d\pm}(N,L,V)$
- Example: Claims of friendship and enmity among people

$g=5$ children's friends (+) and enemies (–)



Example of a signed directed graph

Signed Directed Graph

Semicycles: A closed sequence of distinct nodes and arcs in which each node is either adjacent to or adjacent from the previous node in the sequence.

- Arcs may point in either direction

- Sign of semicycle: Product of the signs of the arcs
- Signed graphs and signed directed graphs generalize graphs and directed graphs by allowing the lines or arcs to have valences

Valued Graph

- Social network data may consists of valued relations in which the strength or intensity of each tie is recorded.
- Examples :
  - Frequency of interaction among pairs of people
  - Dollar amount of trade between nations
  - Rating of friendship between people in a group
- Each line carries a value.

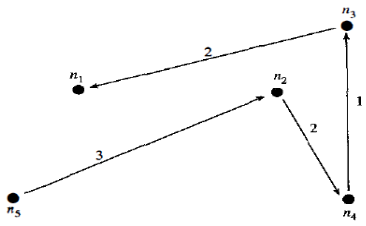
Valued Graph

- A Valued Graph, $G_v(N,L,V)$, consists of three sets of information :
  - a set of nodes, $N = \{n_1, n_2, n_3, \ldots, n_g\}$
  - a set of lines, $L = \{l_1, l_2, l_3, \ldots, l_L\}$
  - a set of values, $V = \{v_1, v_2, v_3, \ldots, v_L\}$ attached to the lines
- Integer weighted digraph : A valued digraph in which all values are from the set of integers
- A **signed graph is a special case of a valued graph** in which the values are only + 1 and -1
- A graph is a special case of a valued graph in which each and every line has a value 1

Valued Graph

- Example: If nominations of three best friends and three worst enemies were requested, ties might be labeled +3 for a best friend,+2, + 1, -1, -2, and - 3 for a worst enemy
- Application of valued graphs - Markov chain : Set of graphs whose values are probabilities. Their sociomatrices are referred to as transition matrices or stochastic matrices
  - In a Markov chain, the values of all arcs incident from each node are constrained to sum to 1

Valued Directed Graph - Example



$l_1 = < n_3, n_1 >$ $v_1 = 2$
$l_2 = < n_2, n_4 >$ $v_2 = 2$
$l_3 = < n_4, n_3 >$ $v_3 = 1$
$l_4 = < n_5, n_2 >$ $v_4 = 3$

Example of a valued directed graph

Valued Graph

Nodal degree :

- Equal to the number of lines incident with the node
- Equal to the number of arcs incident to it or from it (Valued directed graph)
- Values attached to the lines must be considered
- Average the values over all lines incident with a node or all arcs incident to or from a node.
  - Average value of the lines incident with the node or of the arcs to or from the node

Valued and Valued Directed Graph

Density:

- Ratio of the number of lines or arcs present to the maximum possible number of lines/arcs
- Each line or arc is given a value of 1, and pairs of nodes for which lines are absent are given a value of 0
  - Sum of these values is equal to the number of lines or arcs
- Alternatively,
  - Density = Average of the values assigned to the lines/arcs across all lines(arcs)
- Density , $\Delta = \sum v_k / g(g-1)$
  - where the sum is taken over all k.
- Measures the average strength of the lines/arcs in the valued graph/digraph

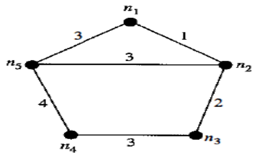
Valued Graph

Path:

- Nodes n_i and n_j are reachable if there is a path between them, considering "strengths" or "values" of reachability
- Value of a Path(semipath): Equal to the smallest value attached to any line (arc) in it
  - Value of a path is thus the "weakest link" in the path.
- Path Length: Sum of the values of the lines in it
- A high value for a path can result either if the values of the lines in the path are high, or if the path is long
- Example: If the lines represent the amount of communication between each pair of people in a group, then the value of a path between two people represents the most "restricted" amount of communication between any pair of people in the path

Valued Graph

:



$l_1 = (n_1\ n_2)$ $v_1 = 1$
$l_2 = (n_1\ n_5)$ $v_2 = 3$
$l_3 = (n_2\ n_3)$ $v_3 = 2$
$l_4 = (n_2\ n_5)$ $v_4 = 3$
$l_5 = (n_3\ n_4)$ $v_5 = 3$
$l_6 = (n_4\ n_5)$ $v_6 = 4$

Path	*Length*	*Value*
$n_1\ n_5\ n_4$	7	3
$n_1\ n_2\ n_3\ n_4$	6	1
$n_1\ n_5\ n_2\ n_3\ n_4$	11	2

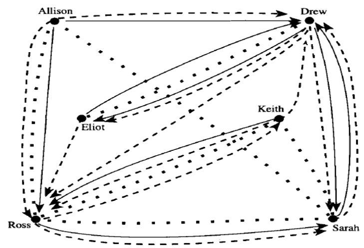
Paths in a valued graph

Multigraph

- A simple graph is used to represent a single relation in a social network
- Multiple relations are represented in a multigraph
- A multigraph, or a multivariate (directed) graph allows more than one set of lines
- More than one relation can be measured on the same set of actors
- A multi graph, G, consists of a set of nodes, N = $\{n_1, n_2, n_3, \ldots, n_g\}$ and two or more sets of lines, $L^+ = \{L_1, L_2, \ldots, L_R\}$, R is the number of sets of lines in the multigraph

Multigraph

- Graph with Multiple Relations
  - Example:
    G (N$\{n_1, n_2, \ldots n_n\}$, $L_1\{l_1, l_2, \ldots l_n\}$, $L_1\{l_1, l_2, \ldots l_n\}$, $L_1\{l_1, l_2, \ldots l_n\}$)



Hypergraph

- Affiliation networks or membership networks, require considering subsets of nodes in a graph
  - Subsets can be of any size
- Consider ties among subsets of actors in a network
  - Tie among people who belong to the same club or civic organization.
- Hypergraphs are the appropriate representations for such networks.
- Affiliation network is a two-mode network consisting of a set of actors and a set of events
- Each event is a subset of the actors from N

Hypergraph

- Affiliation network data cannot be fully represented in terms of pairwise ties
  - since the subsets can include more than two actors
- Hypergraph :
  - consists of a set of objects, A, and a collection of subsets of objects, B, in which each object belongs to at least one subset and no subset is empty, H(A,B)
  - Objects are called nodes/points and the collections of objects are called edges

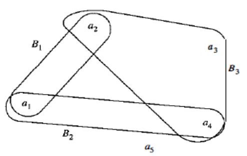
Hypergraph

- Dual hypergraph, denoted H* : Obtained by reversing the roles of the nodes/points and the edges
- If the hypergraph H = (A, B) has node/point set A and edge set B, then the dual hypergraph H* = (B, A) has node set B and edge set A
- A graph is a special case of a hypergraph in which the number of nodes in each edge is exactly equal to two
- Any graph can be represented as a hypergraph, by letting the nodes in the graph be the points in the hypergraph, and letting each line in the graph be an edge in the hypergraph.

Hypergraph

- Objects are called nodes/points and the collections of objects are called edges
- Example: Four actors attending three social events

$B_1 = \{a_1, a_2\}$
$B_2 = \{a_1, a_4\}$
$B_3 = \{a_2, a_3, a_4\}$



Example of a hypergraph

- Alternatively,

$A_1 = \{b_1, b_2\}$
$A_2 = \{b_1, b_3\}$
$A_3 = \{b_3\}$
$A_4 = \{b_2, b_3\}$

Relation

- A mathematical relation focuses on the ordered pairs of actors in a network between whom a substantive tie is present.
- In a social network, ties link pairs of actors
- Cartesian product of two sets (or of a set with itself) is a useful mathematical entity for studying relations.
- Cartesian product of two sets, M of size h, and N of size g, is the collection of all ordered pairs in which the first element in the pair belongs to set M and the second element belongs to set N
  - Denoted by M x N and it contains h * g elements

Relation

- A relation, R, on the set N is defined as a subset of the Cartesian product N x N
- Relation R consists of all ordered pairs $< n_i , n_j >$ for whom the substantive tie from i to j is present
- If the ordered pair $< n_i , n_j > \in R$ then we write iRj

Properties of Relations:

- A relation is reflexive if all possible $< n_i , n_i >$ ties are present in R; that is, iRi for all i.
- If no < ni, ni > ties are present in R, then the relation is irreflexive
- If a relation is neither reflexive nor irreflexive, then it is not reflexive
- A relation that is not reflexive is one on which iRi for some but not all is present

Relation - Properties

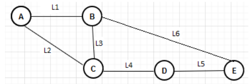
- A relation is symmetric if
  - it has the property that iRj if and only if jRi, for all i and j.
- A symmetric relation is one in which all dyads are either mutual or null.
- On some relations the presence of the < ni, nj > tie implies the absence of the < nj, ni > tie. Such a relation is antisymmetric.
- An antisymmetric relation is one on which iRj implies that not jRi.
  - An example of an antisymmetric relation is the relation "beats" in a sporting tournament

Relations

- A relation that is neither symmetric nor antisymmetric is called not symmetric, non-symmetric, or asymmetric.
- A relation that is not symmetric is one for which iRj and jRi exists, for some but not all i and j.
- A relation is transitive if whenever iRj and jRk, then iRk exists for all i, j, and k.
  - Substantively, transitivity captures the notion that "a friend of a friend is a friend."

Matrices for Graphs

- Adjacency Matrix/Sociomatrix
- Incidence Matrix
- Hypergraphs are represented through incidence matrix



Adjacency Matrix

	A	B	C	D	E
A	-	1	1	0	0
B	1	-	1	0	1
C	1	1	-	1	0
D	0	0	1	-	1
E	0	1	0	1	-

Incidence Matrix

	L1	L2	L3	L4	L5	L6
A	1	1	0	0	0	0
B	1	0	1	0	0	1
C	0	1	1	1	0	0
D	0	0	0	1	1	0
E	0	0	0	0	1	1

Matrices for Graphs

- Sociomatrix: The primary matrix used in social network analysis is called the adjacency matrix, or sociomatrix.
- Graph theorists refer to this matrix as an adjacency matrix because the entries in the matrix indicate whether two nodes are adjacent or not.
- In the study of social networks, the adjacency matrix is usually referred to as a sociomatrix.