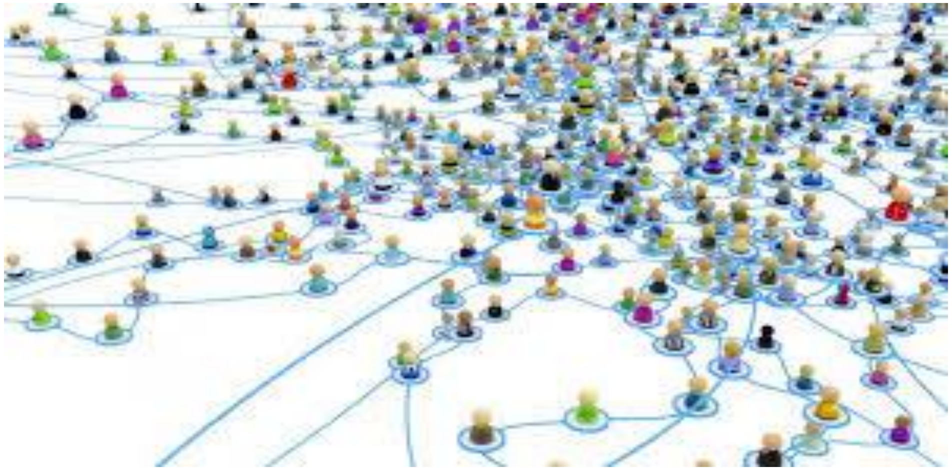
Social Network Analysis



NOTATIONS FOR SOCIAL NETWORK DATA

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

- Social network data can be represented mathematically in different ways

- Three different notational schemes

- Based on its appropriateness, clarity, or efficiency any one is used

- Graph theoretic
- Sociometric
- Algebraic

Social Network Analysis

Graph theoretic notation

It is most useful for centrality and prestige methods, cohesive subgroup ideas and as dyadic and triadic methods

Sociometric notation

It is often used for the study of structural equivalence and block models.

Algebraic notation

It is most appropriate for role and positional analyses and relational algebras.

$N = \{n_1, n_2, n_3 \dots n_g$

Social Network Analysis

Graph Notations

It is viewed as an elementary way to represent actors and relations

It views is as a graph, consisting of nodes joined by lines $N = \{n_1, n_2, n_3 \dots n_g\}$

Set N contains g actors

Example:

$$N = \{Alex, George, Alan, Bob, michael, Harris\}$$

$$Here\ we\ infer,\ \ n_1 = Alex, n2 = George, n3 = Alan \dots$$

Social Network Analysis

Single Relation

Single relation records whether each actor in *N* relates to every other actor on this relation

The relation be dichotomous and directional

If a tie is present between pair of actors n_i and n_j then the ordered pair belongs to set *L*

Maximum element in the L is g (g-1) and the minimum can be 0

Social Network Analysis

If the ordered pair $< n_i, n_j >$ *has tie between them it is represented by* $n_i \rightarrow n_j$

$L = \{l_1, l_2, l_3 \ldots l_l\}$, *here each l represents ordered pairs*

L can be represented graphically by drawing line from first actor in the element to second actor

The graph is called as directed graph and directed lines are called arcs

Graph g consists of set of nodes $\mathbf{N}$ *and set of lines* $\mathbf{L}$, *mathematically represented as* (N, L)

Social Network Analysis

$In\ some\ relation, individual\ actor\ don't\ relate\ to\ itself, here\ self\ choices\ are\ not\ considered$

$In\ non-directional\ no\ distinguish\ between\ the\ line\ n_i\ and\ n_j\ and\ n_j\ and\ n_i$

$Example{:}\ set\ of\ actors$ live near each other

$L\ contains\ \dfrac{g(g-1)}{2}\ pairs\ for\ undirected\ graph$

$L\ can\ be\ l1 = < Ross, Alan >, l2 = < Alex, michael >, \ldots .\, l8 = < Sarah, Drew >$

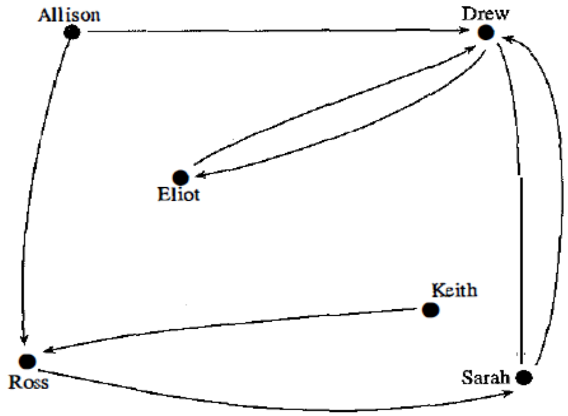
$Here\ friendship\ is\ not\ reciprocal, so\ it\ can\ be\ n_i \nrightarrow n_j$

Social Network Analysis

A graph represented as diagram where, nodes are represented as points in 2D space and arcs are represented by directed arrows between points

Location of points in two-dimensional space is irrelevant

Example:



Social Network Analysis

Multirelational

Graph theoretic notation can be generalized to multirelatiorial networks

It could include both directional and nondirectional relations

Example:

Between 2 persons, two types of relationship: friendship and marital tie

Each relation has a corresponding set of arcs L_r, in $\mathbf{L}_r$ which contains ordered pairs of actors as elements

Where r ranges from 1 to R, the total number of relations

each relation is defined on the same set of nodes, but each has a different set of arcs
Example: relation1: friendship, relation2: classmate, relation3: neighborhood

Social Network Analysis

For each relations, the directed graphs can be viewed in one or more figures

each relation is defined on the same set of nodes, but each has a different set of arcs

Example: relation1: friendship, relation2: classmate, relation3: neighborhood

Single Relation

single relation measured on one set of gactors in

Define x_{ij} as the value of the tie from the i^{th} actor to the j^{th} actor on the single relation

Place these measurements into a sociomatrix

Rows and columns of this sociomatrix index the individual actors.

Since there are g actors, the matrix is of size g x g

Sociometric notation uses such matrices to denote measurements on ties

For the relation *X,* we define **X**as the associated sociomatrix. The entries are defined as:

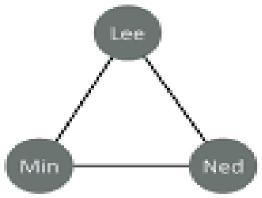
$$X_{ij} = \textit{the value of tie from } n_i \textit{ to } n_j \textit{where } i \textit{ and } j \ (i \neq j) \textit{range over all integers from } 1 \textit{ to } g$$

$$\textit{pairs listing same actor twice } (n_i, n_i), i = 1,2, \ldots g \textit{ are called self choices}$$

Self choices are usually undefined, lie along diagonal of sociomatrix

Sociomatrix

	Lee	Min	Ned
Lee	0	1	1
Min	1	0	1
Ned	1	1	0



Social Network Analysis

The possible values of the relation C, if it is dichotomous C=2 or if relation is valued and discrete can take no. of different values

Example, if the relation can take on the values -1, 0, 1,

Then map - 1 to 0,0to 1, and +1 to 2 (so that C=3)

Single relation is just a special ease of the multirelational

Social Network Analysis

Multiple relations

Suppose R relations $X_1, X_2, X_3, \ldots\ldots$ measured on a single set of actors where $r = 1,2,3, \ldots R$.

Relations are valued and come from the set $\{0,1,2, \ldots \ldots C-1\}$

$$X_{ijr}\ is\ the\ strength\ of\ the\ tie\ from\ ith\ actor\ to\ jth\ actor\ on\ rth\ relation$$

It is placed in the collection of sociomatrices, one for each relation

Example:

Friendship at Beginning of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Social Network Analysis

There are R, g x g sociomatrices, one for each relation defined for the actors in N

R sociomatrices viewed as the layers in a three-dimensional matrix of size g x g x R

The rows index the sending actors, the columns index the receiving actors, and the layers index the relations

Also referred to as a **super sociomatrix** as represents information in a multirelational network.

Example:

Consider of a collection of g = 6 children and R =. 3 relations:
1) Friendship at beginning of the school year
2) Friendship at end of the school year
3) Lives near

Social Network Analysis

Friendship at Beginning of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Friendship at End of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	1	1
Eliot	0	0	-	0	1	0
Keith	0	1	0	-	1	0
Ross	0	0	0	1	-	1
Sarah	0	1	0	0	0	-

Lives Near

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	0	0	0	1	1
Drew	0	-	1	0	0	0
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	1
Ross	1	0	0	1	-	1
Sarah	1	0	0	1	1	-

$X_{121} = the\ value\ of\ tie\ from\ n_1\ to\ n_2\ on\ the\ relation\ X_1$

$X_{211} = 0\ , no\ friendship\ between\ two$

Summary

- Social network data can be represented mathematically in 3 notations

- Graph theoretic
- Sociometric
- Algebraic

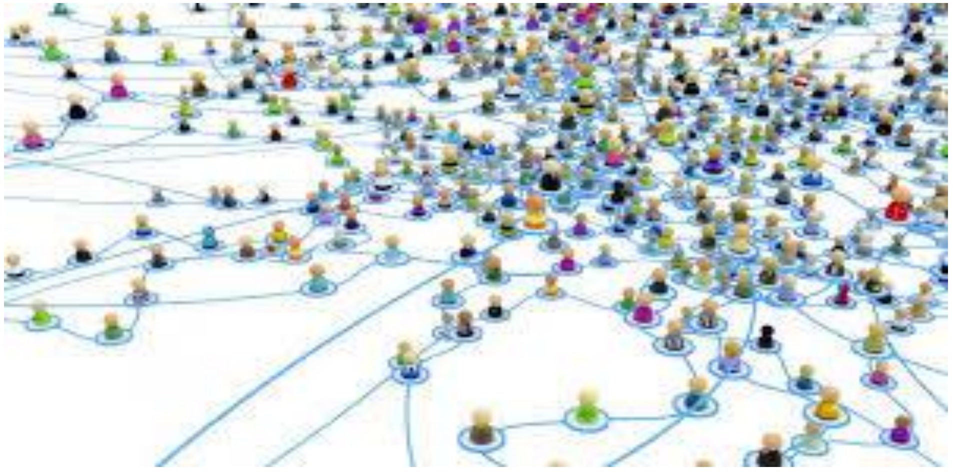
- Graph theoretic notation used in centrality and prestige methods, subgroups

- Sociometric notation used in structural equivalence and block models

- Algebraic notation used in role and positional analyses and relational algebras

Social Network Analysis

GRAPH - Part2



Course Instructor: Dr.V.S.Felix Enigo

Eccentricity of a Node

It is the largest geodesic distance between a node and any other node in a graph

Formally, the eccentricity of node n_i in a connected graph is equal to the maximum d(i,j), for all j, (or $\max_j d(i,j)$)

It shows how far a node is from the node most distant from it in the graph

Social Network Analysis

Diameter of a Graph

Largest geodesic distance between any pair of nodes in a graph

It is the largest eccentricity of any node

Diameter range from 1 to g-1

It quantifies how far apart the farthest two nodes in the graph are

Message takes shortest route over a path of length no greater than the diameter of the graph

Social Network Analysis

Diameter of a Graph

Largest geodesic distance between any pair of nodes in a graph

It is the largest eccentricity of any node

Diameter range from 1 to g-1

It quantifies how far apart the farthest two nodes in the graph are

Message takes shortest route over a path of length no greater than the diameter of the graph

Diameter of a subgraph

Geodesic of a subgraph is the length of the shortest path between the nodes n_i and n_j within the subgraph

Any path, and any geodesic, including nodes and lines outside the subgraph, is not considered.

The diameter of a sub graph is the length of the largest geodesic within the subgraph.

Social Network Analysis

Connectivity of Graphs

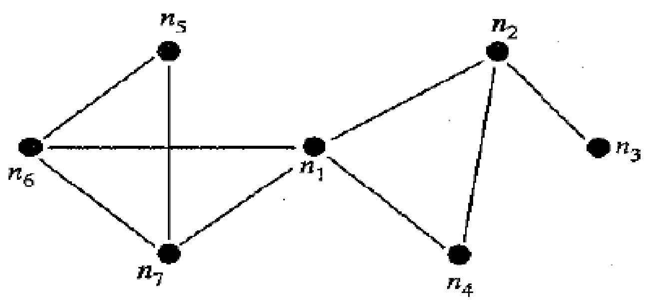
Connectivity of a graph is a function of whether a graph remains connected when nodes and/or lines are deleted

Cutpoints

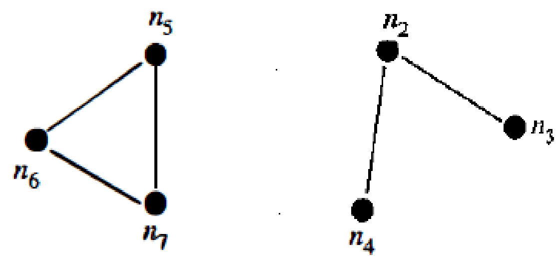
A node n_i is a cutpoint if the number of components in the graph g that contains n_i is fewer than the number of components in the subgraph g_s that results from deleting n_i from the graph.

Social Network Analysis

Example Cutpoint



Node n_1 is a node cut, or cutpoint



The graph without node n_1

Typical Scenario:

In a communications network, an actor who is a cutpoint is critical, if removed, in the remaining network that has two subsets of actors, between whom no communication can travel

Cutset

Cutpoint can be extended from a single node to a set of nodes necessary to keep the graph connected

Cutset is the set of nodes is necessary to maintain the connectedness of a graph

If the set is of size k, then it is called a k-node cut

Bridges

A bridge is a line that is critical to the connectedness of the graph.

A bridge is a line such that the graph containing the line has fewer components than the subgraph that is obtained after the line is removed

The removal of a bridge leaves more components than when the bridge is included

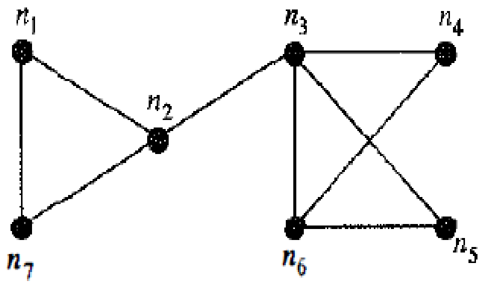
l-line cut is a set of l lines that, if deleted, disconnects the graph

A bridge is a 1-line cut

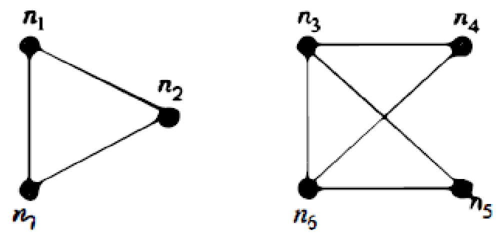
Social Network Analysis

In social networks, a bridge is a critical tie, or a critical interaction between two actors

Example:



Line (n_2 n_3) is a bridge



Graph without line (n_2 n_3)

The line (n2, n3) is a bridge. If the line (n2, n3) is removed from the graph, there is no path between nodes n1 and n5 and the graph becomes disconnected

If the line (n2, n3) were nonexistent, nodes n1, n2, and n7 would not be reachable from nodes n3, n4, n5 and n6

Social Network Analysis

Node and Line-Connectivity

- One way to measure the cohesiveness of a graph is by its connectivity

A graph is cohesive if:

- There are relatively frequent lines
- Many nodes with relatively large degrees
- Relatively short or numerous paths between pairs of nodes

- Cohesive graphs have many short geodesics, and small diameters, relative to their sizes

Social Network Analysis

If a graph is not cohesive then it is "vulnerable" to the removal of a few nodes or lines

A vulnerable graph is more likely to become disconnected if a few nodes or lines are removed

Point-connectivity or node-connectivity of a graph

It is the minimum number of nodes that must be removed to make the graph disconnected, or to leave a trivial graph

K(W), is the minimum number K for which the graph has a K-node cut.

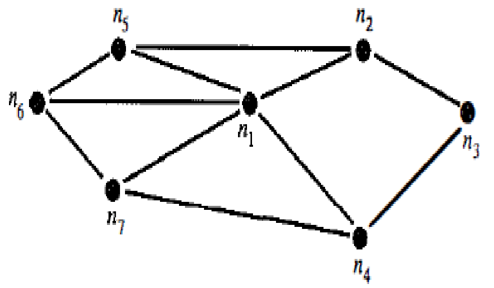
If the graph is disconnected, then k=0, since no node must be removed

If the graph contains a cutpoint, then K = 1 since the removal of the single node leaves the graph disconnected

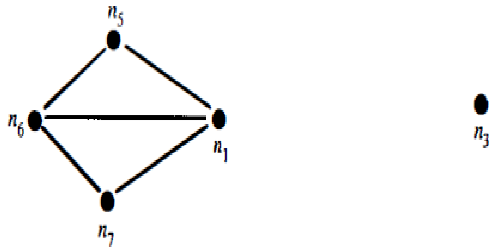
If a graph contains a pair of nodes whose removal together would disconnect the graph, then K = 2

Higher values of K indicate higher levels of connectivity of the graph

Social Network Analysis



n_2 and n_4 comprise a 2-node cut



The graph without n_2 and n_4

The 2-node cut consists of n2 and n4, because without them n3 would not be connected to the remainder of the graph

The value K is the minimum number of nodes that must be removed to make the graph disconnected

For any value k less than K the graph is said to be k-node connected.

Social Network Analysis

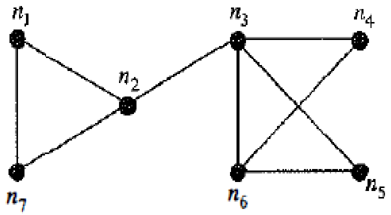
A complete graph has no cutpoint as all nodes are adjacent to all others

To disconnect a complete graph, one would need to remove g - 1 nodes resulting in a trivial graph (g = 1)

Social Network Analysis

The line-connectivity or edge-connectivity of a graph, λ (G), is the minimum number λ for which the graph has a λ -line cut

The value, λ, is the minimum number of lines that must be removed to disconnect the graph or leave a trivial graph



Line (n_2 n_3) is a bridge

l1=(n2,n3) is a bridge, λ(G)=1, the minimum number of lines whose removal disconnects the graph is 1

The graph is said to be l-line connected, since l is the minimum number of lines that must be removed to make the graph disconnected.

Summary

Diameter is the largest geodesic distance between any pair of nodes

Connectivity of a graph is a function of whether a graph remains connected when nodes and/or lines are deleted

Cutpoints are critical nodes when removed graph gets disconnected

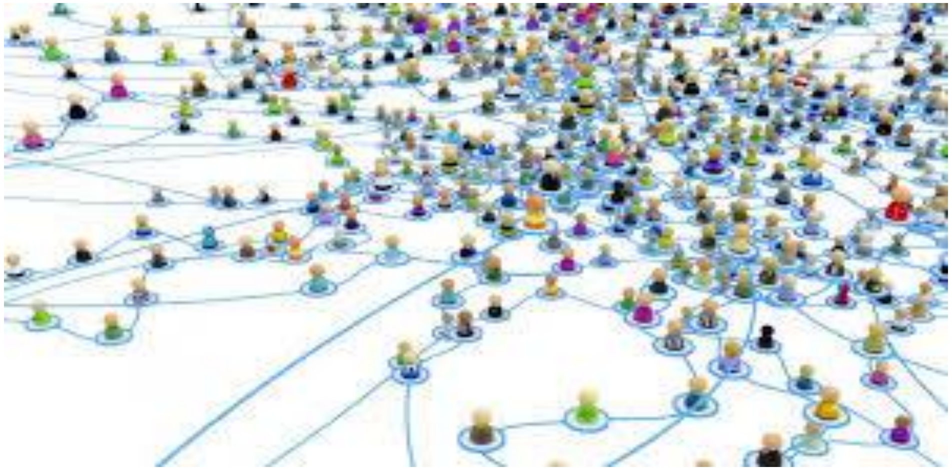
Bridge is a critical line when removed graph splits into components or subgraph

Node connectivity - minimum number of nodes that must be removed to make the graph disconnected

Line Connectivity - minimum number of lines that must be removed to make the graph disconnected

Social Network Analysis

GRAPHS



Course Instructor: Dr.V.S.Felix Enigo

Introduction

Graph theory has been useful in social network analysis as:

- Provides vocabulary used to label and denote social structural properties
- Provides mathematical operations for the properties to be quantified and measured
- Provides ability to prove theorems about representations of social structure

Social Network Analysis

- Barnes and Harary (1983) noted "Network analysts makes too little use of the theory of graphs”
- Graphs is used in SNA to formally represent social relations, quantify social structural properties
- Visual representation of data by sociogram allows researchers to uncover patterns which otherwise is missed

Social Network Analysis

Graph

- A graph is a model for a social network with an undirected dichotomous relation
- A tie is either present or absent between each pair of actors
- Non-directional relations Examples:
  - co-membership -in formal organizations or informal groups
  - kinship relations - "is married to," "is a blood relative of”
  - proximity relations - "lives near”
  - interactions -"works with"
- In a graph, nodes represent actors and lines represent ties between actors

Social Network Analysis

- A graph g consists of two sets of information : a set of nodes $N = \{n_1, n_2, n_3 \ldots n_g\}$ and a set of lines, $L = \{l_1, l_2, l_3 \ldots l_l\}$ between pairs of nodes.

- Lines are unordered pairs of nodes, so $L_k = (n_i, n_j) = (n_j, n_i)$

- Loops or reflexive ties are excluded

- $Two\ nodes\ n_i\ and\ n_j\ are\ adjacent\ if\ the\ L_k = (n_i, n_j)\ is\ in\ the\ set\ of\ the\ line\ L$

- A node is incident with a line, and the line is incident with the node, if the node is one of the unordered pair of nodes defining the line

- A graph that contains only one node is trivial and all other graphs are nontrivial

Social Network Analysis

- A graph that contains g nodes and no lines L=0 iscalled empty graph
- A social network consist of one actor (the trivial graph)
- Network consisting of more than one actor, but no ties between the actors (the empty graph)
- Sociogram only represents set of nodes and presence or absence of lines between the nodes
- The location of the points on the page is arbitrary, and the length of the lines between points is meaningless

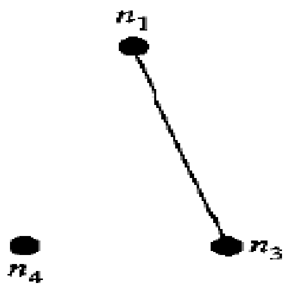
Social Network Analysis

Subgraphs

A graph g_s is a subgraph of g if the set of nodes and lines of g_s is a subset of the set of nodes and the set of lines in g

$In\ G\ set\ of\ nodes\ N = \{n_1, n_2, n_3 \ldots n_g\}\ L = \{l_{1,} l_2, l_3, l_4\}, here\ the\ subgraph\ g_s$ $contains\ N_s = \{n_1, n_3, n_4\}\ and\ set\ of\ lines\ is\ L_s = \{l_2\}$

Example:



b. subgraph

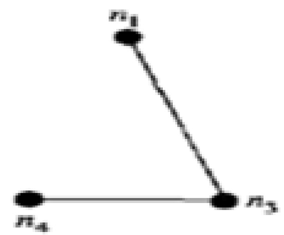
$\mathscr{N}_s = \{n_1\ n_3\ n_4\}$

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

Node generated subgraph

It is a subgraph g_s generated by a set of nodes Ns and line set Ls, where set of lines Ls includes all lines from L that are between pair of node in Ns

$$\mathcal{N} = \{n_1\ n_2\ n_3\ n_4\ n_5\}$$
$$\mathcal{L} = \{l_1\ l_2\ l_3\ l_4\}$$
$$l_1 = (n_1\ n_2) \qquad l_3 = (n_1\ n_5)$$
$$l_2 = (n_1\ n_3) \qquad l_4 = (n_3\ n_4)$$



c. subgraph generated by nodes n_1 n_3 n_4

$$\mathcal{N}_s = \{n_1\ n_3\ n_4\}$$
$$\mathcal{L}_s = \{l_2\ l_4\}$$

Social Network Analysis

Scenario where node-generated subgraph used:

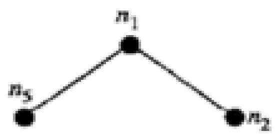
- if the researcher considers only a subset of the g members of the network.
- Relational data might be missing for some of the network members
- In a longitudinal study in which a network is studied over time, some actor,
or subset of actors, might leave the network
- Analyses of the network might have to be restricted to the subset of actors for whom data are available for all time points
- Node-generated subgraphs are widely used in the analysis of cohesive subgroups in networks
- Focus on subsets of actors among whom the ties are relatively strong, numerous, or close.

Social Network Analysis

Line generated subgraph

A subgraph g_s is generated by a set of lines Ls, where the set of nodes Ns includes all the nodes that are incident with the lines Ls

$\mathscr{N} = \{n_1\ n_2\ n_3\ n_4\ n_5\}$
$\mathscr{L} = \{l_1\ l_2\ l_3\ l_4\}$
$l_1 = (n_1\ n_2) \qquad l_3 = (n_1\ n_5)$
$l_2 = (n_1\ n_3) \qquad l_4 = (n_3\ n_4)$



d. subgraph generated by lines $l_1\ l_3$

$\mathscr{N}_s = \{n_1\ n_2\ n_5\}$
$\mathscr{L}_s = \{l_1\ l_3\}$

Social Network Analysis

- An important feature of a subgraph is maximal with respect to some property
- A subgraph is maximal with respect to a given property if that property holds for the subgraph gs but does not hold if any node or nodes are added to the subgraph

Dyads

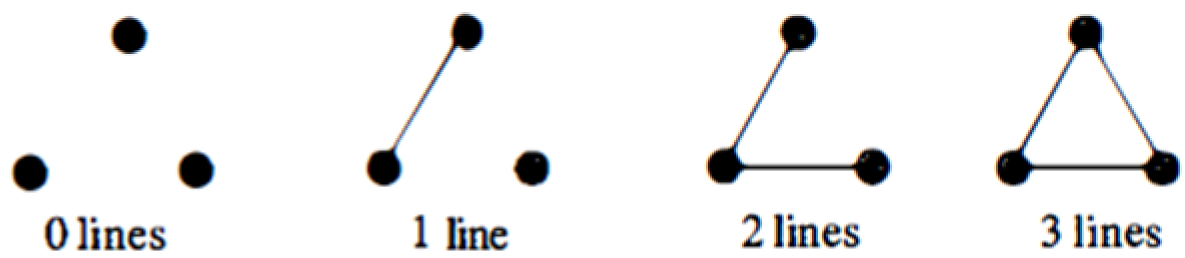
- A dyad, represents subgraph consisting of a pair of nodes and the possible line between the nodes
- Unordered pair of nodes can be in only one of two dyadic states: two nodes are adjacent or they are not adjacent

Social Network Analysis

Triads

A triad is a 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



Social Network Analysis

Granovetter (1973) refers to the triad with two lines present and one line absent as the **forbidden triad**

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

This type of triad, with only two lines is forbidden in Granovetter's model

Both dyads and triads are node-generated subgraphs, since they are defined as a subset of nodes and all lines between pairs of nodes in the subset

Social Network Analysis

Nodal Degree

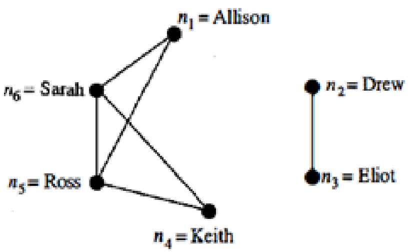
The degree of a node, denoted by d(ni) is the number of lines that are incident with it

Equivalently, the degree of a node is the number of nodes adjacent to it

The minimum degree is 0, if no nodes are adjacent to a given node, also called isolate

The maximum degree is g - 1, if a given node is adjacent to all other nodes in the graph

Example:



The degrees of the nodes are: d(n1) = 2, d(n2) = l, d(n3) = l, d(n4) = 2,d(n5) = 3, and d(n6) = 3

Social Network Analysis

Computing degree is informative in many applications

Example: Children playing together smaller degree infers children played with few children, large degree infers children played with many children

Summarizing the degrees of all the actors in the network is useful in many applications

The mean nodal degree is a statistic that reports the average degree of the nodes in the graph

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

Social Network Analysis

- If all the degrees of all of the nodes are equal, the graph is said to be d-regular

  where d is the constant value for all the degrees (d(ni) = d,for all i and some value d)

- d-regularity can be thought of as a measure of uniformity If a graph is not d-regular, the nodes differ in degree

- The variance of the degrees, which we denote by S_D^2 it is calculated as:

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

- A graph that is d-regular has $S_D^2 = 0$

Social Network Analysis

- Variability in nodal degrees means that the actors represented by the nodes differ in "activity," as measured by the number of ties they have to others
- The variability of nodal degrees is one measure of graph centralization
- Nodal degrees are important to study higher-order network properties (such as reciprocity) to control for or condition on the set of nodal degrees in a graph

Density of Graphs and Subgraphs

In a graph of g excluding loops, there are g(g - 1)/2 possible unordered pairs of nodes, and thus g(g - 1)/2 possible lines or maximum number of lines

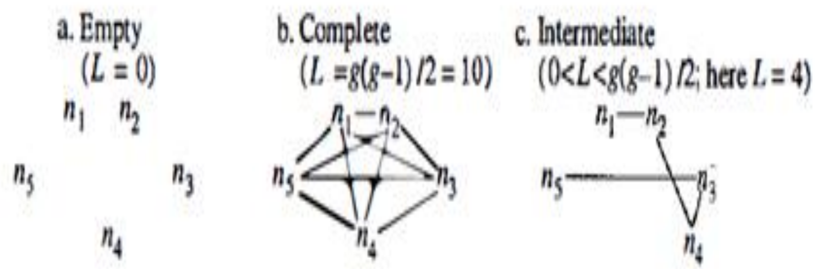
The density of a graph is the ratio of the number of lines present, L, to the maximum possible.

It is denoted by Δ

$$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)}.$$

Social Network Analysis

The density of a graph goes from 0, if there are no lines present (L= 0), to 1 , if all possible lines are present



If all lines are present, then all nodes are adjacent, and the graph is said to be complete

It is standard to denote a complete graph with g nodes as Kg

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

Example

Relation such as "communicates with," where all g actors communicated with all other actors

Relationship between the density of a graph and the mean degree of the nodes in the graph $\Delta = \frac{\bar{d}}{(g-1)}$

Density of the subgraph is $\Delta_s = \frac{2L_s}{g_s(g_s - 1)}$

The density of a subgraph expresses the proportion of ties that are present among a subset of the actors in a network

Social Network Analysis

A walk is 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.

The beginning and ending nodes may be different

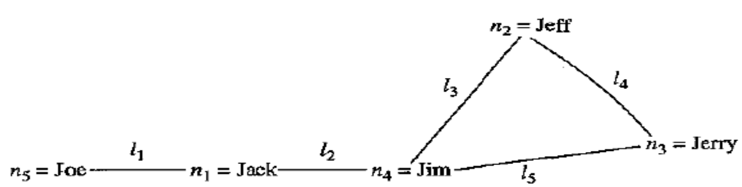
The length of a walk is the number of occurrences of lines in it

A line is included more than once in the walk, it is counted each time it occurs

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 $W = n_1\ l_2\ n_4\ l_3\ n_2$



The inverse of a walk, denoted by W^{-1}, is the walk W listed in exactly the opposite order, using the same nodes and lines

Example: Communication ties among g = 5 employees

Social Network Analysis

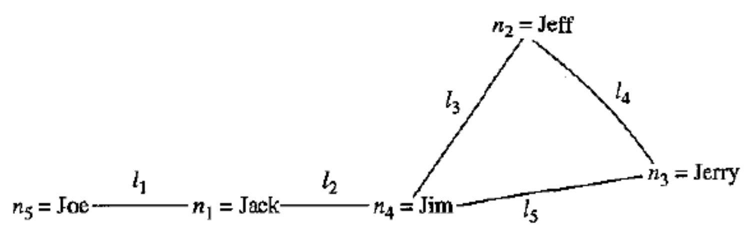
Properties to be noted in walk

- First, not all nodes are involved : the message never reached Joe (n5) or Jerry (n3)
- Second, some nodes were used more than once: Jim (n4) was included in the walk twice.
- Third, some lines were not used (l1,l4,l5)
- Some lines were used more than once (that is, 13)
- The walk W = n1l2,l3n4,l3n2,n4, be written more briefly as W=n1,n4,n2,n4.
- The origin and terminus in this walk are n1, and n4
- The length of the walk is 3, since there are three lines: l2,l3,l3
- The length is 3 even though there are only 2 distinct lines as one of the line repeated twice

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 $W = n_1\ l_2\ n_4\ l_3\ n_2$



Social Network Analysis

Trials and path

A trail is a walk in which all of the lines are distinct, though some node(s) may be included more than once

The length of a trail is the number of lines in it

Example: no communication tie is used more than once

Social Network Analysis

A path is a walk in which all nodes and all lines are distinct

The length of a path is the number of lines in it

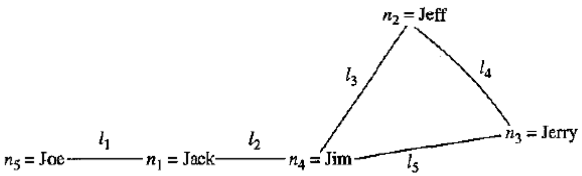
Example: a path through a communication network means no actor is informed more than once

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 $W = n_1\ l_2\ n_4\ l_3\ n_2$

One of the trails is n4n2n3n4 (no line is repeated)



One of the paths is nln4n2 (no line or node is repeated)

There may be more than one path between a given pair of nodes.

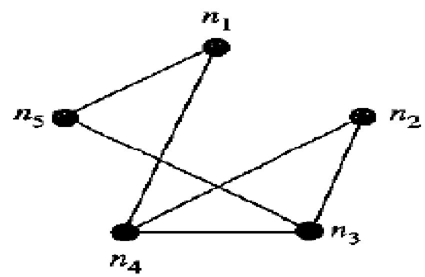
Example, there are two paths between n1 and n2 : n1n4n2 and n1n4n3n2

Social Network Analysis

If there is a path between nodes ni, and nj, then ni and nj are said to be reachable

Example: two actors should be reachable to pass message from one actor to the other through intermediaries

A walk that begins and ends at the same node is called a closed walk



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

Social Network Analysis

A cycle is closed walk of at least three nodes in which all lines are distinct, and all nodes except the beginning and ending node are distinct.

Special case: A cycle is labeled Hamiltonian if every node in the graph is included exactly once.

A graph that contains no cycles is called acyclic

Cycles are important in the study of balance and clusterability in signed graphs

A tour is a closed walk in which each line in the graph is used at least once

$$W = n_3 n_2 n_4 n_3 n_5 n_1 n_4 n_3.$$

Special case:

Eulerian trails are special closed trails that include every line exactly once

Social Network Analysis

Summary

- Social network uses graph theory for notations, vocabularies and proving theorems
- A graph is a model for a social network with either an undirected or directed relationships
- Subgraph is the subset of a graph which can be node generated or line generated
- Various measures can be computed by applying graph theory in social networks:
  - Density, nodal degrees, walk, trial, path

Graph and Matrices

Course Instructor: Dr.V.S.Felix Enigo

Introduction

- The information in a graph expressed in a variety of ways in matrix form

E.g. Sociomatrix, incidence matrix

- Sociomatrix – adjacency matrix represents whether two nodes are adjacent or not
- The sociomatrix for a graph (for a nondirectional relation) is symmetric
- Example of a sociomatrix : friendship relation between 4 friends

Incidence Matrix

	n_1	n_2	n_3	n_4
n_1	-	0	0	0
n_2	0	-	1	0
n_3	0	1	-	0
n_4	0	0	0	-

- Incidence matrix, I - records which lines are incident with which nodes
- Incidence matrix has nodes indexing the rows, and lines indexing the columns
- Incidence matrix is binary, as it records whether a line incident with a node or it is not
- It is not necessarily square matrix
- For digraph the entries are choice-based (if i node in row chooses j node in column the entry in the cell is 1 otherwise 0)
- So the entry for i and j may be different from j and i

Matrix Operations

- Permutations
- A permutation of a set of objects is any reordering of the objects (possible reordering)
- It is used in the study of cohesive subgroups
- Important in constructing block-models and in evaluating the goodness-of-fit of blockmodels

	n_1	n_2	n_3	n_4	n_5
n_1	-	0	1	0	1
n_2	0	-	0	1	0
n_3	1	0	-	0	1
n_4	0	1	0	-	0
n_5	1	0	1	0	-

X permuted

	n_5	n_1	n_3	n_2	n_4
n_5	-	1	1	0	0
n_1	1	-	1	0	0
n_3	1	1	-	0	0
n_2	0	0	0	-	1
n_4	0	0	0	1	-

- Reordering rows and cols of the matrix, helps to discover patterns, identify subsets

Transpose

- Transpose of a sociomatrix is analogous to reversing the direction of the ties between the actors
- For non-directional relations, transpose is identical to original matrix
- Matrix multiplication is a very important operation in social network analysis
- It can be used to study walks and reachability in a graph

Powers of the Matrix

- Power of the matrix and Boolean matrix multiplication is also used in social network analysis
- Studying powers of the matrix X used to find walk of specific length
- For example, elements of X^3 counts the number of walks of length 3 between each pair of nodes
- Also used to find walks of longer lengths
- It is also used to find the reachability matrix $\mathbf{X}^{[R]} = \{x_{ij}^{[R]}\}$, says that each pair of nodes whether they are reachable, or not
- The entries tell us total number of directed walks from row node n^i, to column node n^j

Geodesic and distances

- The first power *p* for which the (i, j) element is non-zero gives the length of the shortest path and is equal to d(i,j)

$$d(i,j) = min_p x_{ij}^{[p]} > 0.$$

- If the graph is connected or if the digraph is at least strongly connected, the diameter of the graph is then the largest entry in the distance matrix; otherwise, the diameter is infinite or undefined

Example

X

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	-	1	0	0	1	0
n_2	0	-	1	0	0	1
n_3	0	1	-	0	0	0
n_4	0	0	0	-	1	0
n_5	0	0	0	0	-	1
n_6	0	1	0	0	0	-

$\mathbf{X}^2$

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	0	0	1	0	0	2
n_2	0	2	0	0	0	0
n_3	0	0	1	0	0	1
n_4	0	0	0	0	0	1
n_5	0	1	0	0	0	0
n_6	0	0	1	0	0	1

$\mathbf{X}^3$

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	0	3	0	0	0	0
n_2	0	0	2	0	0	2
n_3	0	2	0	0	0	0
n_4	0	1	0	0	0	0
n_5	0	0	1	0	0	1
n_6	0	2	0	0	0	0

$\mathbf{X}^4$

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	0	0	3	0	0	3
n_2	0	4	0	0	0	0
n_3	0	0	2	0	0	2
n_4	0	0	1	0	0	1
n_5	0	2	0	0	0	0
n_6	0	0	2	0	0	2

$\mathbf{X}^5$

	n_1	n_2	n_3	n_4	n_5	n_6
n_1	0	6	0	0	0	0
n_2	0	0	4	0	0	4
n_3	0	4	0	0	0	0
n_4	0	2	0	0	0	0
n_5	0	0	2	0	0	2
n_6	0	4	0	0	0	0

- For Example : From N1 to N6, the shortest path is 1, though there exist 2 and 3.
- The diameter (largest entry in the distance matrix) is 6

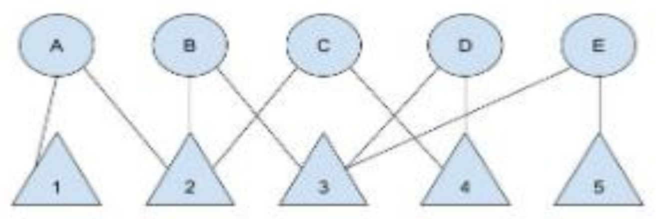
Computing Nodal Degrees

- For a nondirectional relation, the nodal degrees are equal to either the row sums or the column sums.

	1	2	3	4	5
A	1	1	0	0	0
B	0	1	1	0	0
C	0	1	0	1	0
D	0	0	1	1	0
E	0	0	1	0	0

Indegree & Outdegree of directed graph:

- Indegree - Summing over columns (that is, lines) gives the number of lines incident with the node
- Outdegree - Summing over rows (that is, lines) gives the number of lines originate from the node

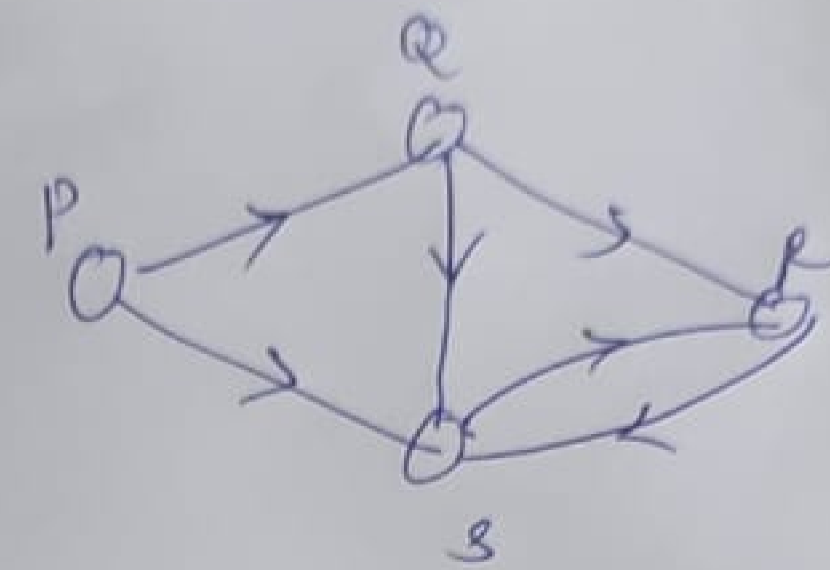


Computing Density

- The density of a graph, digraph, or valued (di)graph can be calculated as the sum of all entries in the matrix, divided by the possible number of entries

$$\Delta = \frac{\sum_{i=1}^{g}\sum_{j=1}^{g} x_{ij}}{g(g-1)}$$

Reachability Matrix



(1 step) $\begin{array}{c|c} & \begin{matrix} P & Q & R & S \end{matrix} \\ \begin{matrix} P \\ Q \\ R \\ S \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$

(2 step) $C^2 = \begin{bmatrix} 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(3 step) $C^3 = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

$$C + C^2 + C^3 = \begin{bmatrix} 0 & 1 & 3 & 4 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

Adding up colums of P, Q R, S we get

$\begin{array}{c} P \\ Q \\ R \\ S \end{array} \begin{bmatrix} 0 \\ 1 \\ 9 \\ 10 \end{bmatrix}$ Whether in 1, 2 or 3 steps to reach end destination S. So 'S' is the most reachable out of all 4 Vertices

(10) infer that, there are 10 pathways