Social Network Analysis

GRAPHS

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

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 $N = \{n_1, n_2, n_3 \dots n_g\}$ and a set of lines, $L = \{l_1, l_2, l_3 \dots l_l\}$ 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

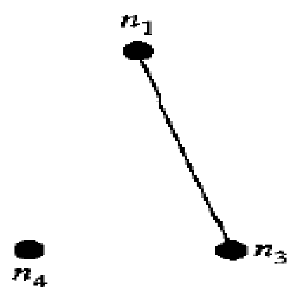
Social Network Analysis

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

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

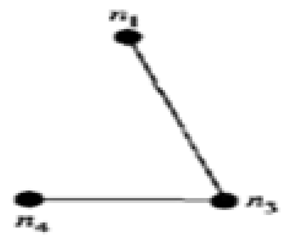
$\mathscr{L}_s = \{l_2\}$

Social Network Analysis

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:

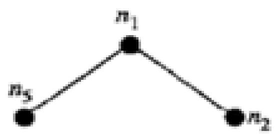
- if the researcher considers only a subset of the g members of the network.
- Relational data might be missing for some of the network members
- In a longitudinal study in which a network is studied over time, some actor, or subset of actors, might leave the network
- Analyses of the network might have 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

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



d. subgraph generated by lines $l_1\ l_3$

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

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

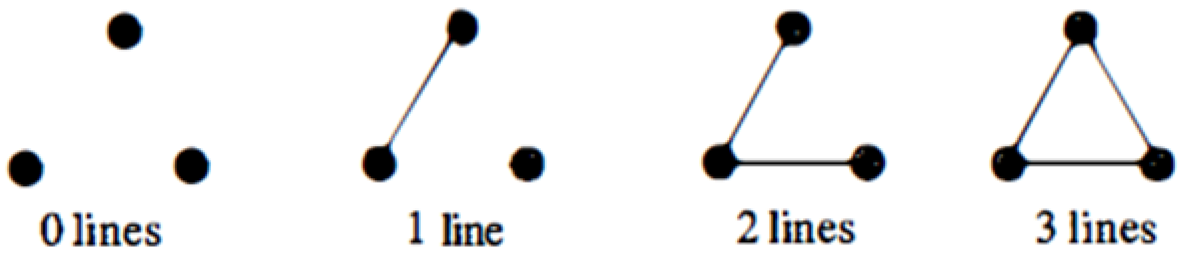
- Unordered pair of nodes can be in only one of two dyadic states: two nodes are adjacent or they are not adjacent

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

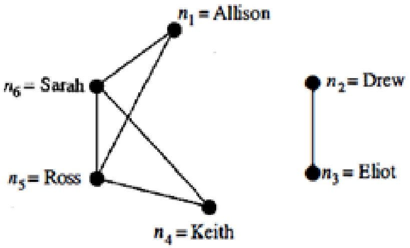
The degree of a node, denoted by d(ni) is the number of lines that are incident with it

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

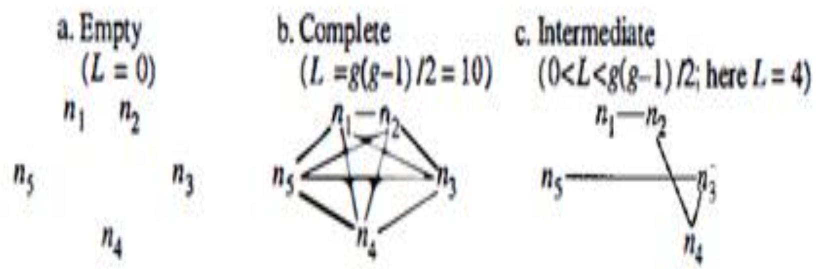
The density of a graph is the ratio of the number of lines present, L, to the maximum possible.

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

Social Network Analysis

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

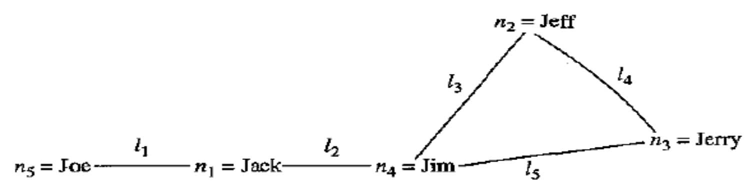
The length of a walk is the number of occurrences of lines in it

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

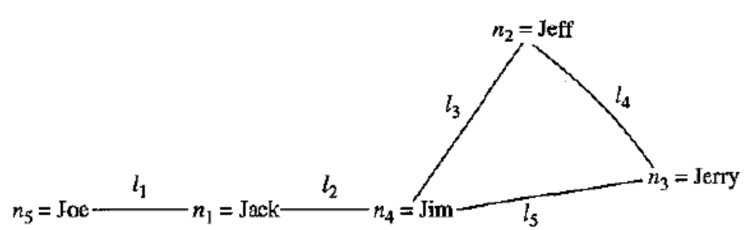
Properties to be noted in walk

- 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

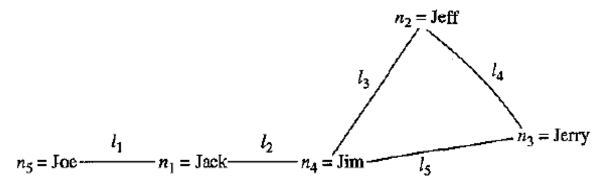
Example: a path through a communication network means no actor is informed more than once

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.

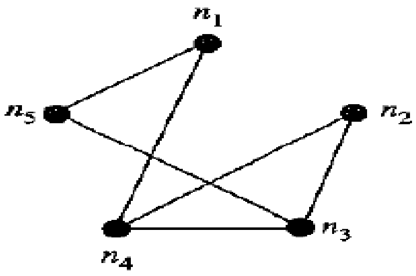
Example, there are two paths between n1 and n2 : n1n4n2 and n1n4n3n2

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