Social Network Analysis

INTRODUCTION

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

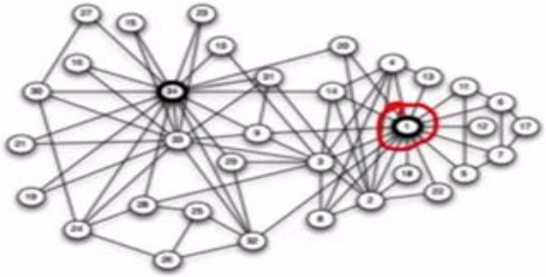
Networks are set of objects (nodes) with some relationship with each other called interconnections (edges)

Why study networks?

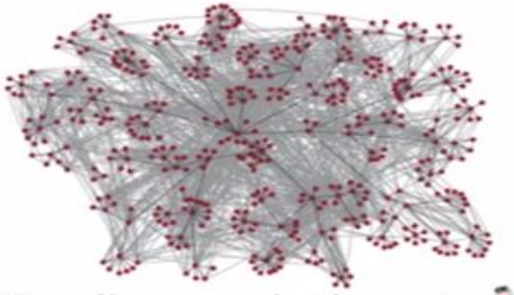
- Networks are everywhere
- How networks are in bunch of different settings

Social Network Analysis

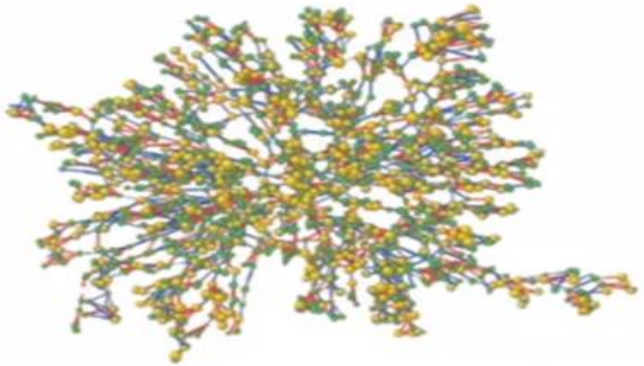
SOCIAL NETWORKS



Friendship network in a 34-person karate club
[Zachary 1977]



E-mail communication network
among 436 HP employees [Adamic & Adar 2005]



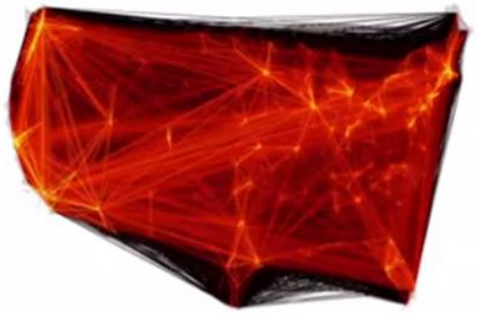
Network of friendship, marital tie, and
family tie among 2200 people
[Christakis & Fowler 2007]

Social Network Analysis

TRANSPORTATION AND MOBILITY NETWORKS

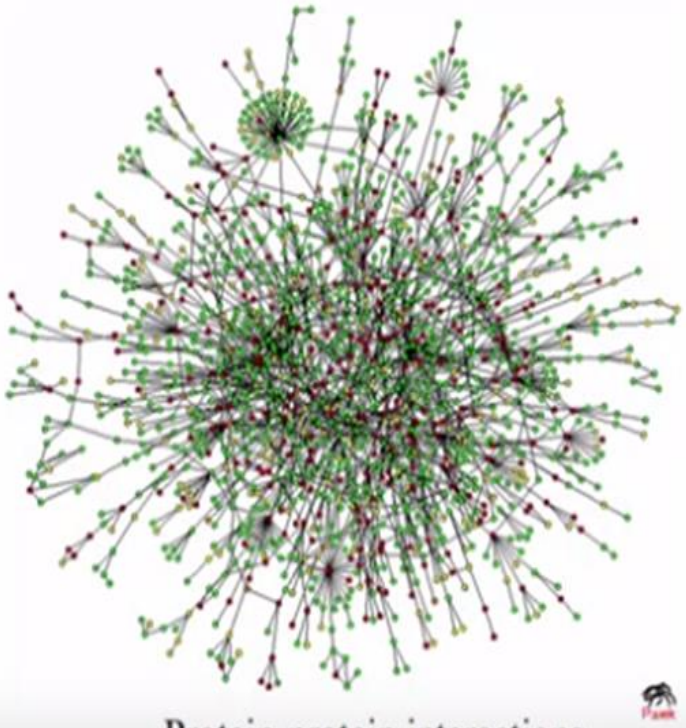
Network of direct flights around the world
[Bio.Diaspora]

Ann Arbor bus transportation network

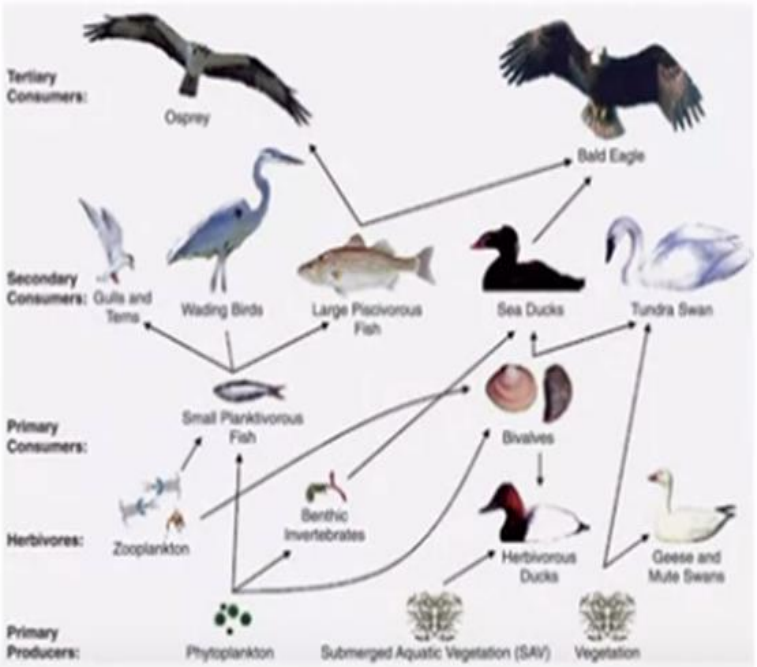


Human mobility network based
on location of dollar bills (Where's George)
[Thiemann et al. 2010]

Social Network Analysis



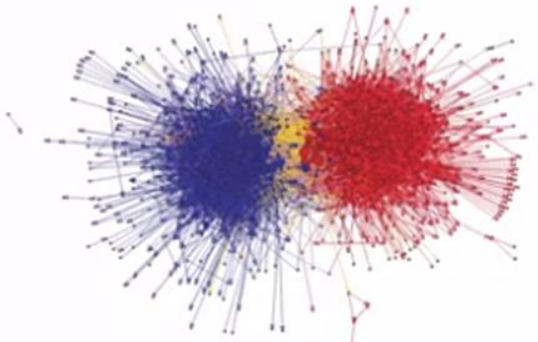
Protein-protein interactions
[Jeong et al. 2001]



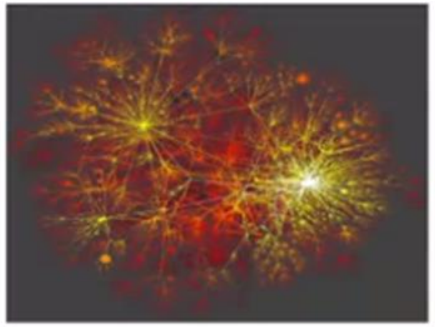
Chesapeake Bay Waterbird Food Web
[Perry et al. 2005]

Social Network Analysis

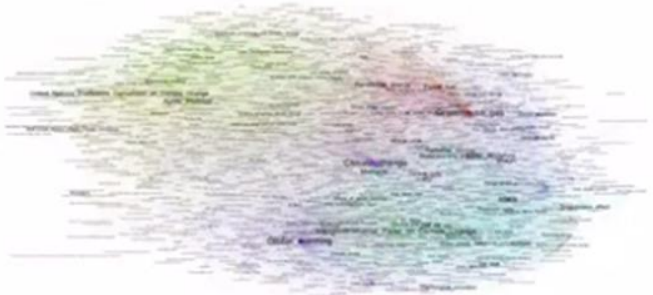
INFORMATION NETWORKS



Communication between left-wing and right-wing political blogs [Adamic & Glance 2005]



Internet Connectivity [K. C. Claffy]



Network of Wikipedia articles about climate change [EMAPS]

Social Network Analysis

AND MORE …

- Financial Networks
- Trade Networks
- Co-authorship Networks
- Citations Networks

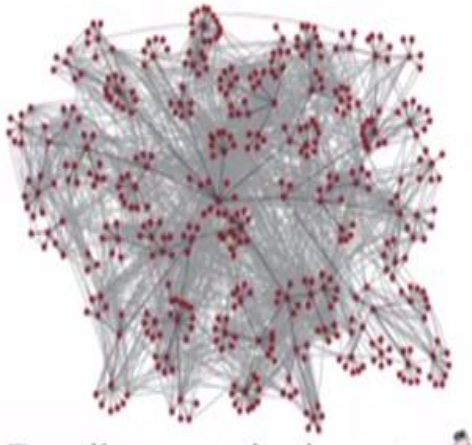
Social Network Analysis

NETWORK APPLICATIONS

Networks are everywhere, but what can we do with them?

Is rumor likely spread in this network?

Who are the most influential people in this organization?



E-mail communication network among 436 HP employees [Adamic & Adar 2005]

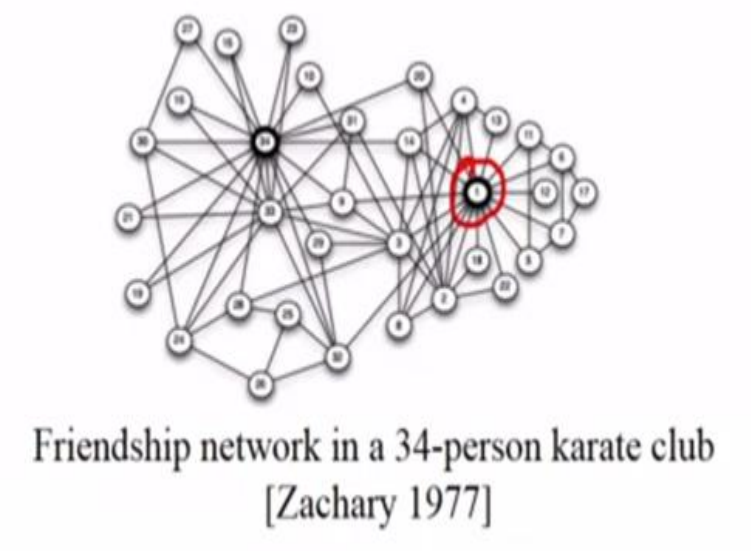
Social Network Analysis

NETWORK APPLICATIONS

Networks are everywhere, but what can we do with them?

Is this club likely to split into two groups?

If so, which node will go to which group?



Friendship network in a 34-person karate club
[Zachary 1977]

NETWORK APPLICATIONS

Networks are everywhere, but what can we do with them?

Which airports are at highest risk for virus spreading ?

Are some parts of the world more difficult to reach?

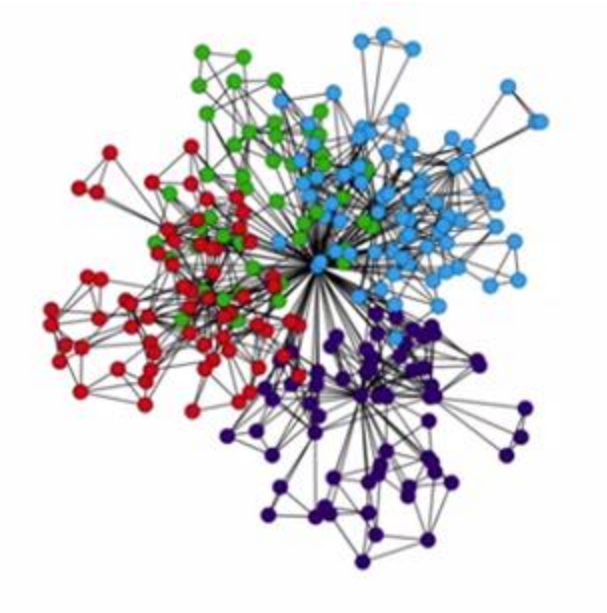
Network of direct flights around the world [Bio.Diaspora]

Social Network Analysis

SUMMARY

More complex structures can be modeled by networks

Studying the structure of a network can allow us to answer questions about complex phenomena



Social Network Analysis

INTRODUCTION

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

GENISIS

- Social Network Analysis convergence of applied research in sociology + social psychology + anthropology
- Developed independently by various researchers through empirical studies of various social settings
- Formal description of social groups found by many social psychologists of the 1940s
- Anthropologists found network representations useful in generalizing actual field observations

Social Network Analysis

- Some of the concepts of network analysis have come naturally from social studies
- 1950s network analysis began converge to a view that distinguishes network analysis from other approaches to sociological research
- Convergence facilitated by the adoption of a graph representation of social networks by Moreno

Social Network Analysis

- Moreno called a *sociogram* was a visual representation of social networks as a set of nodes connected by directed links
- Explored how relations are beneficial and limitations for psychological behavior
- His journal *sociometry* investigated the relationship between psychological well-being and "social configurations”

Social Network Analysis

- John Barnes first introduced the term “Social network” (Barnes, 1954), did formal development of the analysis of social structures
- Barnes was the first to use the concept of social networks in a scientific context – political system, pattern of life, system of kinship and affinity
- In 1954, he published the article "Class and Committees in a Norwegian Island Parish"
- Tremendous increase in the capabilities of network analysis mostly through new applications and the vocabulary, models and methods of network analysis

Social Network Analysis

Two developments led to an explosion of SNA

- First, advances in information technology brought a wealth of electronic data
- Second, Methods of SNA applied to networks other than social networks E.g. Hyperlink structure on the Web or the electric grid
- The reason for adoption in other networks, share a number of commonalities with social networks

Social Network Analysis

What is Social Network?

- It is the study of social relations among a set of actors
- Focus is on relationships between actors than actors itself

Example: Performance of a researcher assessed by count of publications than attributes like grant possessed, age, size of team…

Social Network Analysis

- A statistical analysis relates attributes to outcome variable (the number of publications)

- Analysing social structures identifies presence / absence of relationship patterns

- Example: For researcher, arrangement of relationship in certain pattern is good to his performance than pattern not present

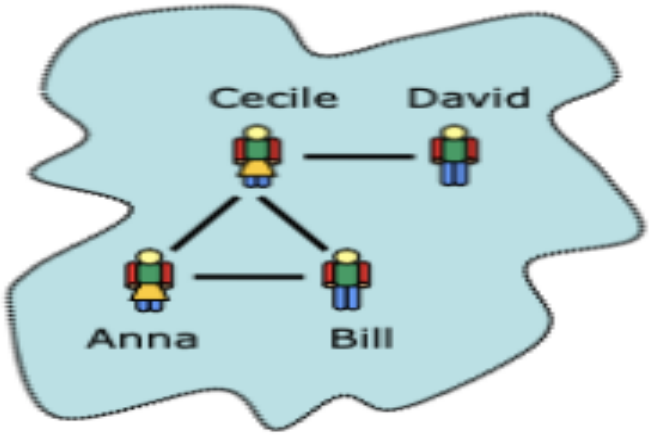
$M \coloneqq (m_{i,j})_{n*n}\ m_{i,j} = \begin{cases} 1 \mid (v_i,\ v_j) \in E \\ 0 \mid otherwise \end{cases}$

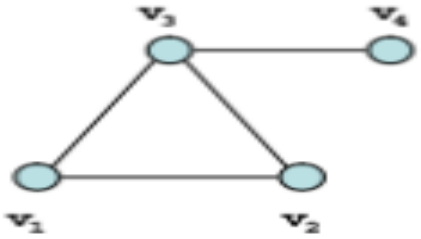
Social Network Analysis

- SNA needs new concepts and methods for data collection and analysis
- Methods of analysis originate from graph theory as social network data represented as graph
- A social network can be represented as a graph G = (V, E), V denotes the finite set of vertices and E denotes a finite set of edges such that $E \subseteq V \times V$

Characteristic matrix of the graph $M \coloneqq (m_{i,j})_{n*n}\ m_{i,j} = \begin{cases} 1 \mid (v_i,\ v_j) \in E \\ 0 \mid otherwise \end{cases}$

Social Network Analysis





	v_1	v_2	v_3	v_4
v_1	0	1	1	0
v_2	1	0	1	0
v_3	1	1	0	1
v_4	0	0	1	0

- *Loops* are not excluded rarely in occur in practical social network data sets (diagonal of matrix is empty)

- In case of multiple component in a graph, one component is chosen for analysis

The global structure of network

Most real world network are represented as abstract graph representation

Questions:

Are there any interesting feature in social network could impose on existing graph models?

Are there any kind of special graph occur in social network?

Social Network Analysis

- American psychologist Stanley Milgram experimented on the structure of social network

- Milgram found that the world around us to be small

"we routinely encounter persons not known to us who turn out to be the friends of our friends"

- Milgram's ***Six degrees of separation*** gives *Average distance between any two individuals in the world*

Graph measures of Milgram

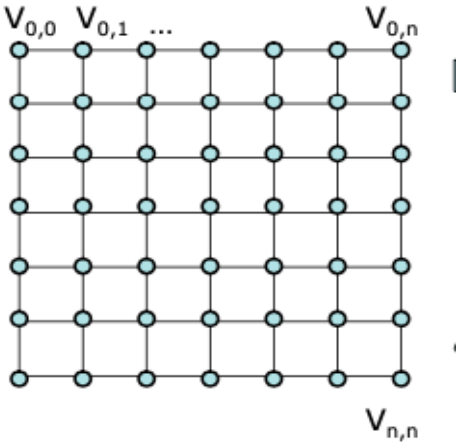
characteristic path length - average shortest path of the network

Geodesic - The shortest path between two vertices is also called a *geodesic*

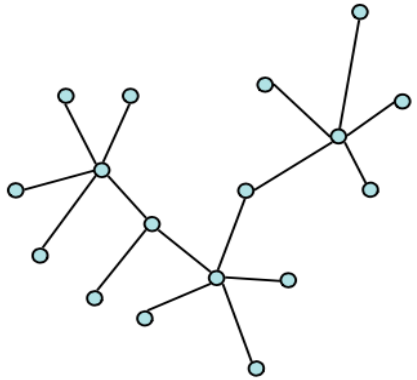
Diameter of the graph - The longest geodesic in the graph

Average shortest path not possible for weakly connected graph

Social Network Analysis



2D Lattice



Tree Graph

Milligram excluded certain structures as models for social graph E.g. 2D Lattice, tree graph

In tree graph all nodes have exactly one edge connected and no loops

In real life people socialize in group but Tree graph shows no clustering

Social Network Analysis

$$\text{Clustering of single vertex} = \frac{\text{actual no. of edges between the neighbors of a vertex}}{\text{possible number of edges between the neighbors}}$$

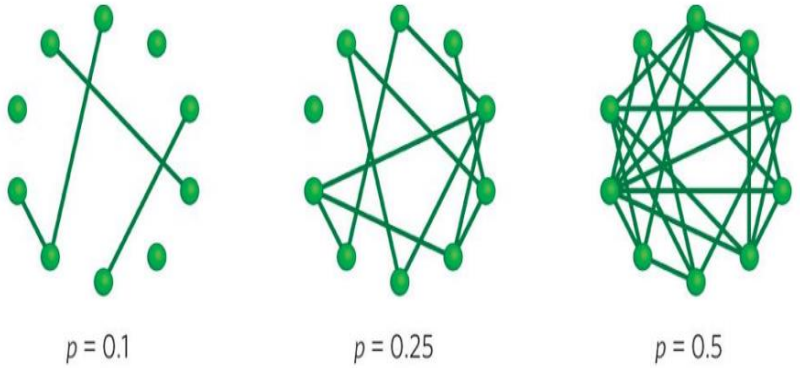
- Clustering coefficient is the average of clustering of all vertex
- Clustering coefficient of a tree is zero (no triads in a graph)

1. **Random Graph Model**:
   - Proposed by Erdős and Rényi.
   - Starts with isolated vertices and gradually adds edges with a certain probability.
   - Increasing the probability results in networks with small characteristic path lengths and some clustering.
   - Although it has limitations in accurately representing real-world networks, it's valuable for studying basic network properties and emergence.
2. **Alpha Model**:
   - Introduced by Strogatz and Watts.
   - The likelihood of forming a new connection between two nodes depends on the number of mutual friends they share.
   - The parameter "alpha" fine-tunes the influence of mutual friends on connection probability.
   - Successfully generates networks with small path lengths and substantial clustering coefficients, capturing characteristics of real-world social networks.
3. **Beta Model**:
   - A generative model that starts with a one-dimensional lattice.
   - Nodes are connected to neighbors and neighbors of neighbors.
   - Rewires links with a probability parameter "beta," leading to small path lengths and relatively high clustering coefficients.
   - Useful for creating networks with both local and global structure.
4. **Scale-Free Graphs**:
   - Proposed by Barabási.
   - Alpha and beta models fail to replicate the scale-free property of degree distributions.
   - Scale-free networks exhibit a power-law degree distribution, where a few nodes have many connections while most have only a few.
   - Common in real networks like co-authorship networks and the World Wide Web.
   - The "rich get richer" principle explains the emergence of scale-free structures.
   - Analyzing these models helps researchers understand network vulnerabilities ar
     spread of information or failures within them.

Social Network Analysis

Random Graph Model

- Generative graph proposed by Hungarian mathematicians Erdő́s and Ré nyi
- Initially start with set of vertices with no edges connected
- Subsequently add edges picking pair of nodes with certain probability
- Increasing the probability, results in small characteristics path and exhibits some clustering



Social Network Analysis

- Random graph model failed as unlikely to make friends completely in random from anywhere in the world
- Yet random graphs are used to study complex structures emerges through elementary Interactions

Alpha Model

- Proposed by Steven Strogatz and Duncan Watts in 1999 in a seminal paper

- No. of mutual friends shared by any two nodes determines the likelihood of a new tie

- Parameter alpha fine-tunes the influence of the number of friends on the probability of a tie

- It successfully generates graphs with small path lengths and large clustering coefficients

Social Network Analysis

Beta Model

The beta-model is also generative model

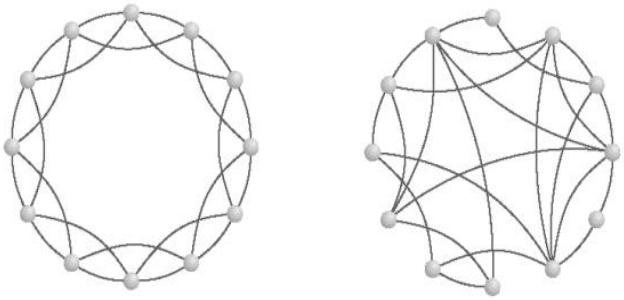
It starts with a one-dimensional torroidal lattice

Every node is connected to its neighbors and the neighbors of its neighbors

One end edge fixed and the other end reassigned randomly selected node

Process continued and rewires every link with a probability of a parameter beta

Choosing appropriate beta, model generates networks with small path lengths and relatively large clustering coefficients



Scale-free graphs

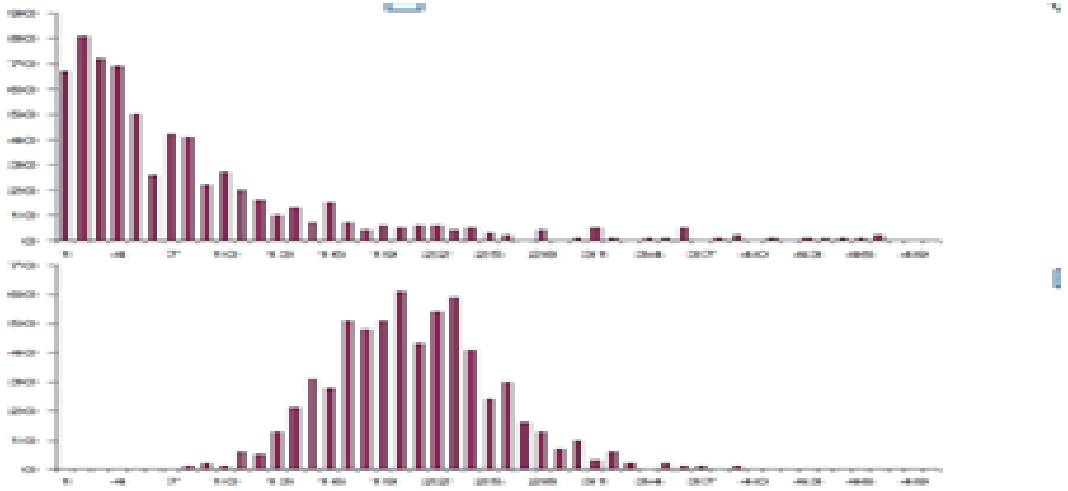
- Proposed by Hungarian physicist Albert-L´aszl´o Barab´asi
- Alpha and beta model failed to recreate scale-free characteristic of the degree distribution
- Distribution is normal in alpha and beta models and random graphs
- Real social networks, nodes with highest degree is less than nodes with small degree

Social Network Analysis

- Exactly correlates power law, i.e. $p(d) = d^{-k}$ where $k > 0$ is a parameter of the distribution

- i.e. a small amount of occurrences is common, while larger occurrences are rare

Example: Co-author network, few researchers have more co-authors and most have one or two co-authors



Social Network Analysis

• Small world phenomenon and scale-free characteristics found in many networks in nature

• Wikipedia, gene network, internet, hyperlink structure of the WWW or the electrical transport network exhibit scale-free characteristics

• *The rich get richer logic creates scale-free structures*

• By analyzing the model instead of particular instances of it allows scientists to formulate general claims about these networks

Example: vulnerability to specific kind of attacks or the possibility for the spread of viruses or failures across them

Social Network Analysis

Summary

- Social Network is amalgamation of sociology + social psychology + anthropology
- Breakthrough in Social Network adoption of a graph representation of social networks by Moreno
- Social Network Analysis is the study of social relations among a set of actors
- Various models developed to represent the behaviour of social network - Random, Alpha, Beta models
- Scale free graphs by Albert-L´aszl´o Barab´asi able to explain exact nature of social network (power law)

Social Network Analysis

WEB BASED NETWORK

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

- Web is a vast, diverse and free to access nearly up to date

Downside:

- Quality of information varies significantly

- Reusing for network analysis (web mining) requires efficient search provided by only commercial search engines

1. **Web Data for Network Analysis**:
   - The web offers a vast and freely accessible source of data for social network analysis.
   - However, the quality of information on the web can vary significantly.
2. **Efficient Search**:
   - To perform network analysis on web data (web mining), efficient search capabilities are essential.
   - Commercial search engines often provide the most effective means for retrieving web data.

Social Network Analysis

Two features of web pages considered as basis of extracting social relations:

- Links and co-occurrences are chosen because
  - linking structure represents real world relationships
  - links are authoritative and relevant as it is chosen by author

Social Network Analysis

Drawback :

- Direct links between personal pages are sparse
- Automating searching personal pages for network analysis results in home page search problem
- Linking structure at higher level are studied for network analysis
- Example:
- Heimeriks et al. studied communication and collaboration networks across different fields of research using a multi-layered approach

CO-OCCURENCES

Co-occurrences of names in web pages serve as evidence of relationships

- Extracting relationships based on co-occurrence of names requires web mining
- Requires statistical methods to analyze the contents of web pages
- Web mining first tested for social network extraction by Kautz el al. on ReferralWeb project for *referral chaining*
- *Referral chaining looks for experts with a given expertise close to the user of the system*

Social Network Analysis

- Referralweb extracted through co-occurrence analysis and counts pages using the search engine, AltaVista
- It collected page counts for individual names and number of pages where the names co-occurred
- Disadvantage: very shallow parsing of the web page as indirect references are not counted

Example:

"the president of the United States" will not be associated with George Bush

Social Network Analysis

Jaccard-coefficient (Tie strength) = number of co-occurrences / number of pages returned for the two names individually

- Tie strength ranges 0 – 1
- Jaccard value exceeds certain fixed threshold concluded as a tie
- Jaccard takes relative measure of co-occurrence and not absolute sizes of the sets
- Expertise of individual are extracted using proper name extraction, NLP technique the result is used to extract new names (repeated 2 or 3 times) [snowballing technique]

Social Network Analysis

- Kautz did not evaluated his system for accuracy, but indicated the level of confidence in its decisions
- He proved it is better than official records, as personal pages are more up to date
- Extraction of names and finding tie between names by Search Engine (SE) is a quadratic problem
- Matsuo et al. to reduce the queries for SE first extracted possible contacts from results of SE

Social Network Analysis

- Jaccard-coefficient (JC) penalizes popular ties, but less popular individuals
- To address this, variant of JC is used for confirming a tie
- Variant JC = number of pages for the individual / number of pages for both names

Social Network Analysis

- To compute the strength of association between the name of a given person and a certain topic

$$\text{Tie strength} = \frac{\text{No. of pages found Cooccurences of interest and name of a person}}{\text{Total number of pages about the person}}$$

- Mutschke and Quan Haase, clustered keywords on publications to themes, assign documents to themes, found themes relevant for researcher

Disambiguating Names

- Biggest technical challenge in social network mining is disambiguating person names
- Problem due to polysemy and synonymy
- Polysemy - SE returns partial set of records different variations of name and names with international characters
- Synonymy - Common names return all pages of all names
- Coverage of the Web is very skewed (over-represented) [web pages are largely ranked by popularity]

Social Network Analysis

- Bekkerman and McCallum dealt ambiguity problem using limited background knowledge
- Clustered list of names related to each other, disambiguated based on hyperlinks between the pages, common links or similarity in content

Social Network Analysis

- Weighted directed link between two persons computed as given below:
- Relevant set constitute top ‘n’ pages of ordered list of pages for the first person and a set of pages for the second
- *rel(n), the relevance at position n, is 1 if the relevant document is at position n and zero otherwise (1 ≤ n ≤ N)*

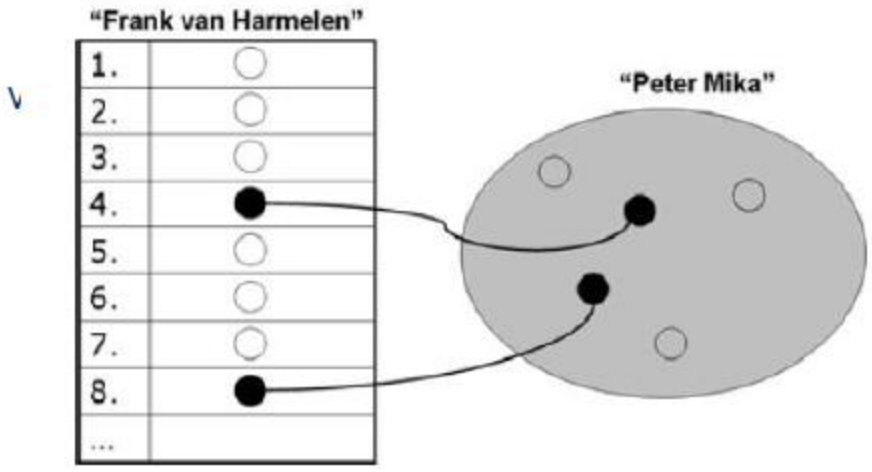
Let *P(n) denote the precision at position n (p@n):*

$$P(n) = \frac{\sum_{r=1}^{n} rel(r)}{n}$$

Social Network Analysis

Average precision is defined as the average of the precision at all relevant positions

$$P_{ave} = \frac{\sum_{r=1}^{N} P(r) * rel(r)}{N}$$



Social Network Analysis

Summary

- Internet provides vast, free resource of data for analysis
- Links and Co-occurrences are treated as tie between actors
- Direct link between personal pages are too scarce, so indirect methods are sought
- Jaccard Coefficient skewed towards popular ties, so variant used to compute the probability of tie
- Disambiguating names problem is solved by additional knowledge
- Average precision at all position in co-occurences is used as strength of tie between two persons

Social Network Analysis

ELECTRONIC DATA SOURCES FOR NETWORK ANALYSIS

Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

DATA COLLECTION COMPLEXITY

- Social networks studied by observation requires close watch
- Standardized surveys minimize influence of observer but need active engagement of population to be studied
- Doubts whether responses are spontaneous and genuine
- To learn Dynamic Networks need multiple surveys
- Manual methods are labor intensive, 50% time spend on data collection
- Force to reuse the same dataset for research due to its complexity

Social Network Analysis

ALTERNATIVES

Reuse existing electronic records of social interactions not created for network analysis

Examples:

- Publication or project databases showing collaborations among authors or institutes
- DB on corporate technology agreements to study networks of innovation
- News paper archives – study of social-cognitive n/w's, terror organization structure

▪Supports dynamic studies through historical analysis at significant price tag

E.g. access to publication & patent DBs, media archives, legal and financial records

Social Network Analysis

ALTERNATIVES

- Internet is vast, diverse, dynamic and free for all resource
- Contributes to rapidly emerging field of *e-social science*
- Data collection automated as all rely on electronic networks, online information sources
- Allows to exploit dynamics of electronic data to perform longitudinal analysis

Limits to e-social science:

- social settings studied only by offline methods
- Technological limits to accuracy

ELECTRONIC DISCUSSION NETWORKS

First study on electronic data is from Information Dynamics Labs of Hewlett-Packard

- Tyler et al studied communication among employees using corporate email archive
- Recreated network assigning tie between employee
- Assigned tie if exchanged certain threshold of emails
- Able to identify leaders, formal and informal communities
- Authors verified through employees of lab feedback by series of interviews

Social Network Analysis

Adamic and Adar revisits question of local search problem:

- ***how do people find short paths in*** **social networks based on only local information about their immediate contacts?**

- Found that additional knowledge (E.g. physical location and position) helps to find path quickly

- More efficient than simple strategy of passing to most connected neighbor

Social Network Analysis

- Study using email electronic data limited by privacy issue

- **E.g.** In case of HP, content of messages is ignored and data set cannot be shared with the community

Alternatives

- Public forums, mailing lists has no privacy issues E.g. Analysis of USENET groups, WWW mailing lists
- Gloor recreated discussion networks from limited information (headers of messages)
- Gloor main contribution is dynamic visualization
- Able to quickly identify activation of entire group during key discussion not just few members
- Gloor also studied the structures that emerge over time

BLOGS AND ONLINE COMMUNITIES

- Primary purpose of analyzing Blogs (Web logs) is trend analysis for the purposes of marketing
- Blogs allow to easily comment and react to the comments resulted communication web among bloggers
- Discussion networks in blogs formed through syndicate blogs, blog rolls and Blog walk series of meeting
- Discussion networks leads to formation of dynamic communities
- Communication and discussion network in blogs enabled for network analysis

Social Network Analysis

Sunday, August 20, 2006

Thinking and berries in Umea

I has not been blogging much last week, but this is only because I has been writing :) And, the best thing of it is where and how I has been writing.

I'm in Umea, Sweden, for DIRN workshop, presentation and work/fun with Stephanie. I'm happy I was able to come a few days earlier.

So far it has been almost perfect work-life balance environment. I worked on my own stuff (more productively than in my own office), discussed tons of things with Stephanie (mainly on weblog research, life and baking), enjoyed culture and nature, and all of that with picking and eating lots of berries.

Some time back Aldo wrote about thinking locations - places where you can get away from the pressures of thr urgent to think your big deep thoughts - I was thinking of it while I enjoyed work and fun here in Umea.

> The social component is very important, and perhaps one of the unique aspect of such a Deep Thought-network: thinkers need on the one hand to be able to concentrate, focus, and withdraw from the world. On the other hand, they very much need to be able to talk with kindred spirits, preferably people working on their own creative projects.

More on http://thinkingcommunities.wikispaces.com

Continued: 1 comments | TrackBacks | Links from other weblogs
More on: life PhD

Link to another blog

Link to another blog post

Quote

Links from other blogs

Comments

Social Network Analysis

Blogs attracted research due to structured electronic data in RSS (Rich Site Summary)

- RSS aids dynamic analysis due to the metadata (timestamp) of the content
- Kumar et al. and Gruhl et al. studied information diffusion in blogs based on this information
- Efimova and Anjewierden are the first to study blogs from a communication perspective
- Adar and Adamic offered visualization of communication in blogs

Social Network Analysis

- Blogs exploited in 2004 US electoral campaign to build networks among individual activists and supporters
- Blogs used by marketers interested in understanding product choices of young demographics
- Separate blog research track is conducted in the yearly Sunbelt social networks conference

Social Network Analysis

- Social networking services cater to socialization more directly than blogs
- Appealing features of social network services are social networking , messaging and photo sharing E.g. LiveJournal
- Paolillo et al studied about the interests and social networks of LiveJournal community
- Backstrom et al. studied influence of structural properties on community formation and growth, changes in membership of communities related to certain topic

Social Network Analysis

- Most social network safeguard their and users data
- Alternative: centralized services like FOAF network
- FOAF profiles are stored on the web site of the users and linked together using hyperlinks
- Studies say FOAF network exhibits similar characteristics to other online social networks
- Drawback of FOAF is lack of tools for maintaining profiles and exploiting the network

Social Network Analysis

Summary

- Manual method of data collection for social network analysis is laborious
- Alternative is online electronic data sources
- Email electronic data limited by privacy issue
- Public forums, mailing lists has no privacy issues can be used
- Blogs and online communities are alternate sources of network analysis as they exhibit network of communication
- Social networking services are more appealing than blogs but has privacy issues
- FOAF centralized services mimic social network but they lack tools to analyze

Social Network Analysis

SOCIAL NETWORK DATA – II

Course Instructor: Dr.V.S.Felix Enigo

Types of Network

One-Mode Networks

One-mode involves measurements on just a single set of actors

Actors can be of a variety of types

- Subgroups
- Organizations
- Collectives/Aggregates:
- Communities
- Nation-states

Social Network Analysis

Relations

The relations measured on the single set of actors in a one-mode network

- Individual evaluations: friendship, liking, respect, and so forth
- Transactions or transfer of material resources: lending or borrowing; buying or selling
- Transfer of non-material resources: communications, sending/receiving information
- Interactions
- Movement: physical (migration from place-to-place), social (movement between occupations or statuses)
- Formal roles
- Kinship: marriage, descent

Social Network Analysis

Actor Attributes

In addition to relational information social network data sets holds actor attributes

People attributes their age, gender, race, socioeconomic status, place of residence, grade in school, and so on

For corporate actors, profitability, revenues, geographical location, purpose of business

Social Network Analysis

Two Mode Networks

- Two-mode involves measurements on two sets of actors, or on a set of actors and a set of events

- At least one relation must be defined between the two sets of actors

Example:
A collection of corporations, and the non-profit organizations (such as the Red Cross) rely on contributions from the public sector for their operating budgets

- Relation is unidirectional

- Additionally number of relations defined for each mode

Example: for the corporations (such as shared country club memberships among the chief executive officers)

- For non-profits (such as interlocking boards of directors)

Social Network Analysis

One Set of Actors and One Set of Events

- Referred as affiliation network or membership network
- First mode is a set of actors, and the second is a set of events which affiliates the actors.

Example: A set of women attended a variety of social functions

  - Each social function can be viewed as a affiliation variable, binary measurement made as to whether a specific actor attended the specific function

- Affiliations are measured on subsets of actors, such networks are non-dyadic, two-mode networks

Social Network Analysis

Actors

Actors same as in one-mode and two-mode networks - only requirement is that the actors must be affiliated with one or more events

Events

Actors (the first mode) are related to each other through their joint affiliation with events

Example:

The events are often defined on the basis of:

- membership in clubs or voluntary organizations
- attendance at social events
- sitting on a board of directors
- socializing in a small group etc…….

Social Network Analysis

Attributes:

Actor attribute variables the same as one-mode and two-mode networks

Two sets of attribute variables in an affiliation network data set: Actor attributes and Event attributes

Social Network Analysis

Special Dyadic Networks and Ego-centered

Special Dyadic Networks

In two-mode networks with two sets of actors, all actors in the first mode can relate to all in the second

some data collection designs gather structural information on some pairs but not others

An example of such data arises in studies of couples

Each partner in the couple can interact with the each other but with no other person during counseling sessions

When interest centers on a collection of pairs (husband-wife, father-son, and so forth), these **non-network relational data referred as special dyadic designs**

Social Network Analysis

- An actor may also relate to a limited number of "special" other actors
- Example, mothers interacting with their own children and children only interact with their own mother
- Partners for one person (either mother or child) are different from the partners for another
- It constrains the interactions among the set of people so that all people cannot interact with all others

Social Network Analysis

Ego centered network

- It is also called as personal network data
- It consists of a focal actor, termed ego, as set of alters who have ties to ego
- Measurements on the ties among these alters
- Example:
- Here one samples respondents, and each respondent reports on a set of alters to whom they are tied

Social Network Analysis

Ego Networks

These data are relational, but limited, since ties from each actor are measured only to some (usually only a few) alters

Ego-centered networks are used in the study of social support

"social support" has been used to refer to social relationships that aid the health or well-being of an individual

It helps to study relationships in clinical and community psychology and sociology

Social Network Analysis

Network Data, Measurement and Collection

Measurement

- Social network data consist of one (or more) relations measured among a set of actors
- Presence of relations has implications for a number of measurement issues
- Unit of observation (actor, pair of actors, relational tie, or event)
- Modeling unit (the actor, dyad, triad, subset of actors, or network)
- Quantification of the relations (directional vs nondirectional; dichotomous vs. valued)

Social Network Analysis

Unit of Observation

- The unit of observation is an actor, from whom we get information about ties.
- The dyad is the unit of observation when one measures ties among pairs of actors directly
- Example, one could record instances of aggression among pairs of children on a playground
- For affiliation network data the unit of observation is often the event

Social Network Analysis

Modeling Unit

There are several levels at which network data can be modeled or summarized

- Actor
- Dyad
- Triad
- Subgroup
- Set of actors or network

Relational Quantification

There are two properties of relations for measurement and for categorizing the methods:

- Directional or non-directional
- Dichotomous or valued

Social Network Analysis

Directional Relations

In a directional relation, the relational tie between a pair of actors has an origin and a destination

Example: one country exports manufactured goods to a second country

non-directional relation the tie between a pair of actors does not have a direction

Example: tie between two countries if they share a border

Social Network Analysis

Second important property of a relation is whether it is **dichotomous or valued**

Dichotomous relations are coded as either present or absent, for each pair of actors

Example: an ambassador send to a country or not taking two values "send" or "not send."

Valued relations can take on a range of values, indicating the strength, intensity, or frequency of the tie between each pair of actors

Example: Record the dollar value of manufactured goods exported from one country to other country

Social Network Analysis

Collection

Variety of ways in which social network data can be gathered

These techniques are:

- Questionnaires
- Interviews
- Observations
- Archival records
- Experiments
- Other techniques : ego-centered, small world, and diaries

Social Network Analysis

Questionnaire

- This data collection most commonly used
- The questionnaire usually contains questions about the respondent's ties to the other actors.
- Questionnaires are used when relation(s) that are being studied are ones that the respondent can report on.
- There are three different question formats that can be used in a questionnaire :
  - Roster vs. free recall
  - Free vs. fixed choice
  - Ratings VB. complete rankings

Social Network Analysis

Roster

- when questionnaire gather network data each actor should be presented with a complete list, or roster
- Rosters can be constructed only when the members known in the set prior to data gathering
- Example: Each person rate their friendship with every member of the class on a five point scale, Corresponding questionnaire is
- "Please place a check in the space that best describes your relationship with each person on the list."
- Five categories one could choose: "trust as a friend", "know welt", "acquaintance", "associate name with face", and "do not know"

Social Network Analysis

Free Call

In Free Call, the researcher does not present a complete list of the actors to the respondent on the questionnaire

Example:

"name those people with whom you (fill in specific tie)“

Each pupil in both schools was asked to write his name, age, grade, and home room number on a card and to fill in the blanks in the statements:

"My best friend in (name of school) Junior High School is . . ."
"My second best friend is . . ."
"My eighth best friend is. . . ."

Here though the network membership is known beforehand but students listed their friends using free recall

Social Network Analysis

Free Call

If not even have a list advance sampling or enumeration techniques are used

Example:

In studies of community elites - selected actors are asked to name other actors they believe to be influential in the community

Social Network Analysis

Free vs, Fixed Choice

Fixed Choice

In a fixed choice design each actor has a fixed maximum number of ties to the other actors in the set of actors.

Example:
Each doctor interviewed was asked three sociometric questions:
(i) "To whom did he most often turn for advice and information?"
(ii) "With whom did he most often discuss his cases in the course of an ordinary week?"
(iii) "Who were the friends, among his colleagues, whom he saw most often socially?"

In response to each of these questions, **the names of three doctors** were requested

Social Network Analysis

Free Choice

If actors are not given any such constraints on how many nominations to make, the data are free choice

Example:

Each individual was asked to denote for each member of the user group whether or not they:

- Had an office next to each other
- Attended the same school at the same time
- Shared an office

Here there is no constraint on the number of people that an individual respondent can choose on these 3 relations

Social Network Analysis

Ratings VS. Complete Ranking

Ratings

- Actors are asked to rate or rank order all the other actors in the set for each measured relation
- Measurements reflect the intensity of strength of ties.
- Ratings require each respondent to assign a value or rating to each tie
- Complete rankings require each respondent to rank their ties to all other actors

Example:

Each of forty members of a social science research office to report the amount of communication with each other member of the office

Asked to rank the cards from most to least on how often they talked to others in the office during a normal working day

Social Network Analysis

- Alternatively, ratings can be dichotomous study (ties are either present or absent)
- Valued - the study where ratings were made by choosing one of five possible categories for the strength of each tie
- In either case, when "choices" can be directional or non-directional

Social Network Analysis

Interviews

Interviews can be of either face-to-face or over the telephone

Used occasionally where questionnaires are not feasible

For example: interviewing the CEOs of the largest corporations , prefer face-to-face interviews than via an impersonal questionnaire

Interviews have been used to gather data from respondents in ego-centered networks

Social Network Analysis

Observation

Observing interactions among actors is another way to collect network data

Used to study relatively small groups of people who have face to-face interactions

Observing interactions does not require verbal responses from the people

Method is quite useful with people who are not able to respond to questionnaires or interviews

Social Network Analysis

Observation

Widely used in the study of interactions among non-human primates

It is useful for collecting affiliation network data

The researcher can record who attends each of a number of social events

Used to observe a set of actors for an extended period of time, summarizes the impressions of ties among all pairs of actors in the set

Social Network Analysis

Archival Records

Measures ties by examining measurements taken from records of interactions

Example:

journal articles, newspapers, court records, minutes of executive meetings

Give rise to longitudinal relations and can be used to reconstruct ties that existed in the past

Affiliation data can be compiled from archival sources

Used in study of sociology of science, specifically, patterns of citations among scholars

One can examine "who cites whom" in order to understand diffusion of a scientific innovation

Social Network Analysis

Others

Other designs for collecting relational data:

- Cognitive Social Structure
- Experimental
- Ego-centered
- Small-world
- Diary

Social Network Analysis

Cognitive Social Structure

Ask respondents to give information on their perceptions of other actors' network ties

Example1:

Employees perceptions of friendships among all other employees in the restaurant

Gives info about own friendships and their perceptions of the friendships among all other pairs of employees

Example2:

Report subgroups of people within the larger collection of people

Gives more information than the usual sociometric design

Social Network Analysis

Experimental

Social network data can be collected using experimental designs in 2 ways:

- Choose a set of actors and observe their interactions in an experimentally controlled situation
- Records interactions or communications between pairs of actors

  - choose actors and specify which pairs of actors are permitted to communicate with each other
  - Records frequency or content of communications between those pairs of actors

Group problem-solving experiments

- Actors assigned to positions defined by the experimenter and allowed to communicate only with specific others
- Experimenter manipulates both group members and their ties

Social Network Analysis

Ego-centered

An ego-centered, or local, network consists of a focal person or respondent (ego), a set of alters who have ties to ego

Measurements are taken on the ties from ego to alters and on the ties between alters

Collection of respondents are asked about their ties to other people to elicit the set of alters

Social Network Analysis

Small World

- Special network designs are also used in small world and reverse small world studies
- A small world study is an attempt to determine how many actors a respondent is removed from a target individual based on acquaintanceship
- Focus of interest is how long these "chains" are
- Also the characteristics of the intermediate actors in the chain
- A reverse small world study focuses on the ties from a specific respondent to a variety of hypothetical targets

Social Network Analysis

Diary

- Each respondent to keep a continuous record of the other people with whom they interact
- Used in the study of personal networks among people
- Data sets frequently include information on the type of relation and characteristics of the alters in each ego-centered network

Social Network Analysis

Longitudinal Data Collection

Study focused on how ties in a network change over time

One measures one or more relations at fixed intervals of time

Such designs allow one to study how stable ties are and whether ties reach an equilibrium state

It deals with 2 questions

- How the process has changed over time
- Whether the past, or the history of the process, can predict the future

Longitudinal social network data can be collected using any of the methods (questionnaire, interview, observation etc.)

Mostly studied in friendship and communications throughout a network over time

Social Network Analysis

Measurement Validity, Reliability, Accuracy, Error

Freeman and Romney (1987):

"social structure refers to a relatively prolonged and stable pattern of interpersonal relations“

Holland and Leinhardt found observed structure of measured network data contains error compare to true structure

Bernard et al concluded 50% of people reported on their own interactions is incorrect – people are not very good at reporting on their interactions

Social Network Analysis

Freeman, Romney, and Freeman (1987)

Argued that verbal reports (recall of interactions) needs memory and cognition

Study of long-range social structure, rather than to particular instances is good

Information on ties is collected from individuals as representatives – so interviewed person should be knowledgable of the information

Social Network Analysis

Valid

A measure of a concept is valid to the extent that it actually measures what it is intended to measure

Example: Which people in this group 'are your friends ?"

Validity based on answer to the question gives a set of actors who are related to the respondent through friendship ties

Very few construct validity of social network measures was studied

Social Network Analysis

Reliable

A measure of a variable or concept is reliable if repeated measurements give the same estimates of the variable

It is assessed by comparing measurements taken at two points in time (test-retest reliability), or by comparing measurements based on subsets of test items

Test-retest assessment of reliability to be appropriate, one must ensure that the "true" value of a variable has not changed over time

Social Network Analysis

Three approaches that have been used to assess the reliability of social network data are:

a) test-retest comparison
b) comparison of alternative question formats
c) reciprocity of sociometric choices

Reliability of sociometric data can also be assessed at different levels.

a) One can study the reliability of the "choices" made by individual actors,

b) One can study the reliability of measures aggregated over a number of individual responses

Example: the popularity of an actor measured as the total number of choices it received

Social Network Analysis

Sociometric questions using ratings or full rank orders are more reliable (have higher test-restest reliability)

Fixed choice designs in which just a few responses are allowed

sociometric questions about more intense or intimate relations have higher rates of reciprocation than less intense or intimate relations questions

Reliability of aggregate measures (such as popularity) is higher than the reliability of "choices" made by individual actors

Social Network Analysis

Measurement Error

Measurement error occurs when there is a discrepancy between the "true" score or value of a concept and the observed (measured) value of that concept

Need to understand the implications of measurement error at different levels.

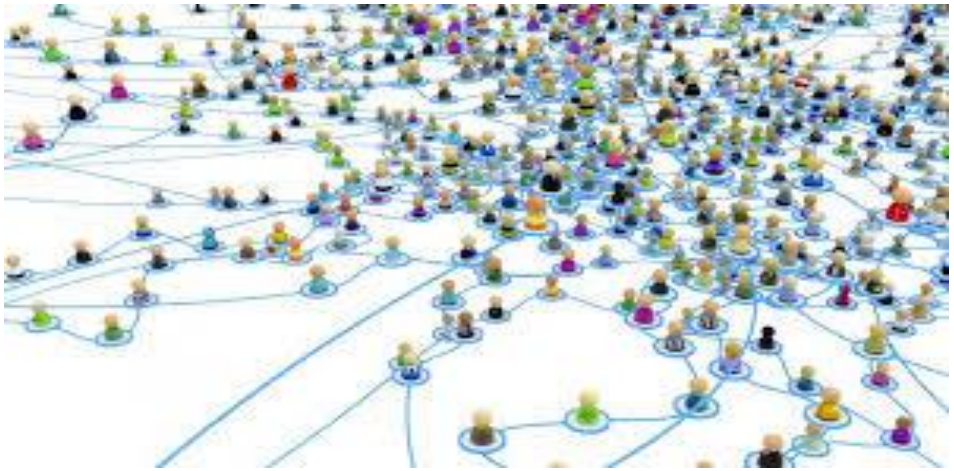
Error arises in fixed choice data collection designs due to restrictions

Example: “List your three best friends." It is quite unlikely that all people have exactly three best friends

The restriction also introduces error into the measurement of network properties, such as of triads and subgroups

Social Network Analysis

MACRO STRUCTURE OF SOCIAL NETWORK

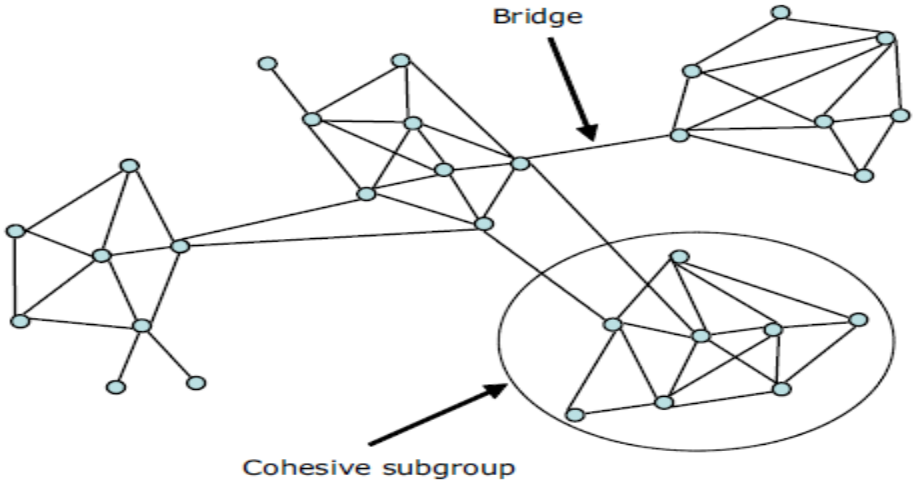


Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

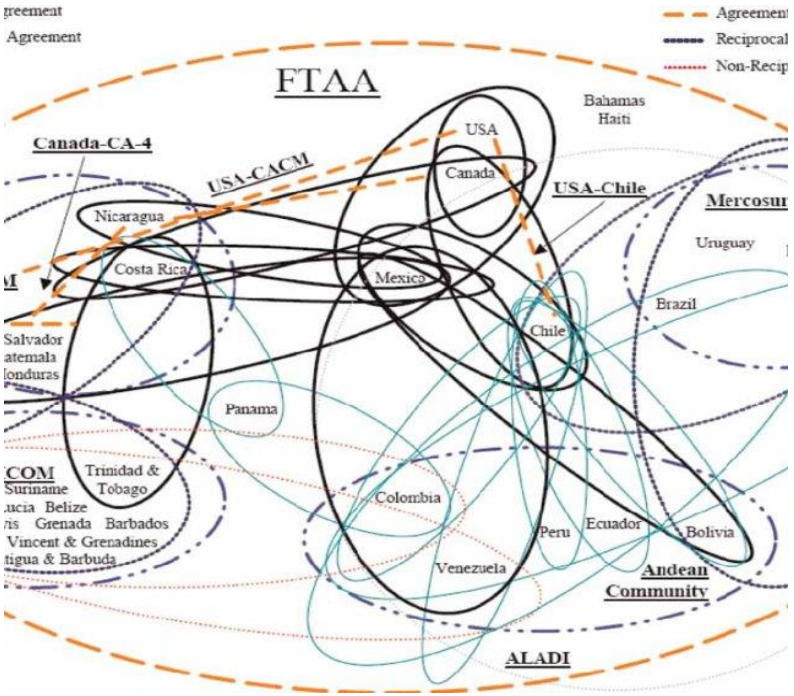
Global structure reveals dense clusters sparsely connected to each other by few ties

Example: Co-authorship network publishes colleagues within institute, rarely do projects with researchers abroad



Social Network Analysis

- Dense the graph with fewer the dimensions of the visualization results in meaningless “spaghetti bowl” tangle of nodes
- Clustering algorithms helps to uncover subgroup structure
- Identifies subgroups of disjoint or overlapping subset of nodes



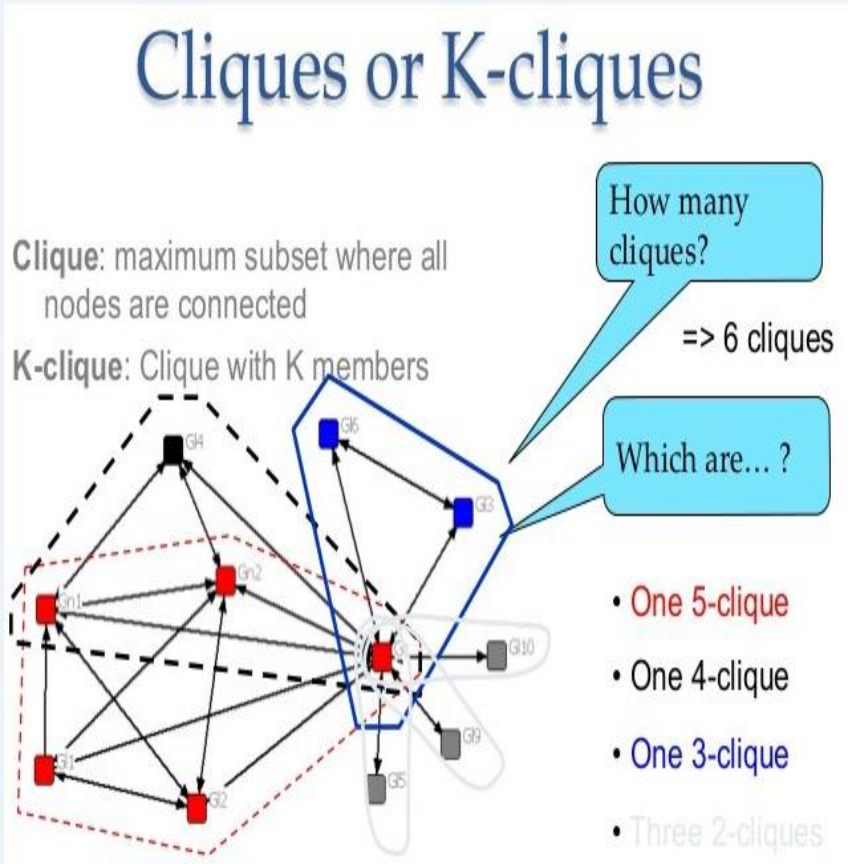
Spaghetti Bowl

Social Network Analysis

- A *clique* in a graph is maximal complete subgraph of three or more nodes

- *k-plex* is a maximal subgraph in which each node is adjacent to no fewer than gs − k nodes in the subgraph

where gs is the number of nodes in the subgraph



Social Network Analysis

Cohesive subgroup defined by lambda-set analysis method

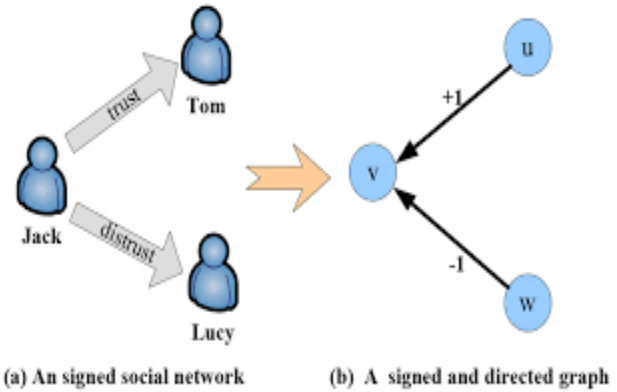
The lambda-set analysis method is based on the definition of edge connectivity

Edge connectivity of two vertices *vi* and *vj* is the minimum number of lines should be removed in a graph to leave no path between the two vertices

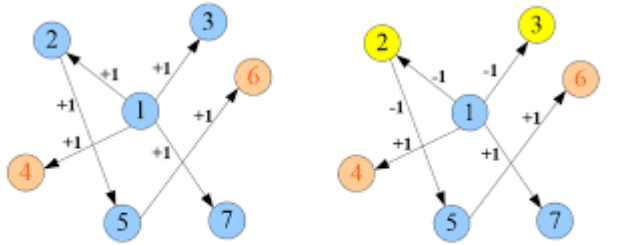
Social Network Analysis

A lambda-set defined for a pair of nodes has a larger edge connectivity than any pair of nodes, when one node is from within the set and the other node is from outside the set

Example. Signed network Relations with positive affections one subgroup and negative affections in other subgroup



(a) An signed social network (b) A signed and directed graph

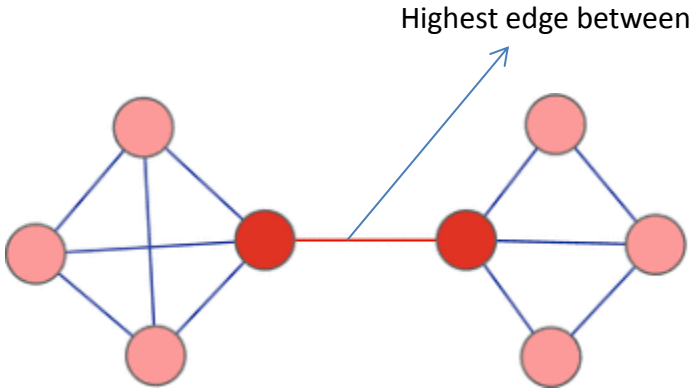


(a) A graph of an unsigned social network (b) A graph of a signed social network

Social Network Analysis

- Clustering method focusing on ties connect them than focusing on subgroups
- Ties between groups are spotted by betweenness

$$Edge\ Betweenness = \frac{fraction\ of\ edge\ that\ contains\ them}{set\ of\ all\ shortest\ paths\ in\ the\ graph}$$

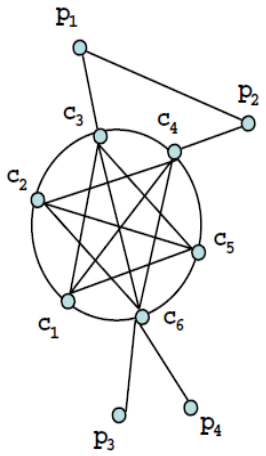


Social Network Analysis

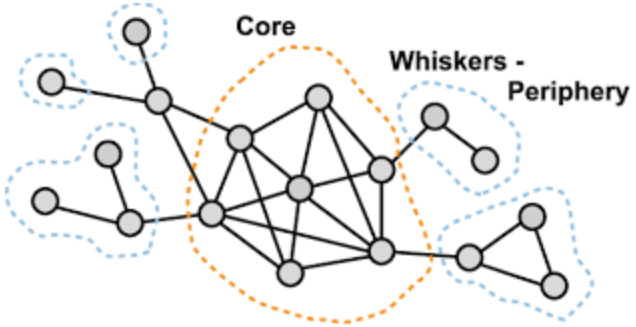
Another typical pattern in social network is *Core-Periphery (C/P) structure*

C/P has two subgroups:

- *nodes in the core densely connected with each other*
- peripheral nodes are not connected with each other , but with the core nodes
- Algorithms divides the set of nodes in a way that the error the between the actual image and the “perfect” image is minimal



c_1	c_2	c_3	c_4	c_5	c_6	p_1	p_2	p_3	p_4
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	1	0	0	0
1	1	1	1	1	1	0	1	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	1	1
0	0	1	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	0	0	0	0

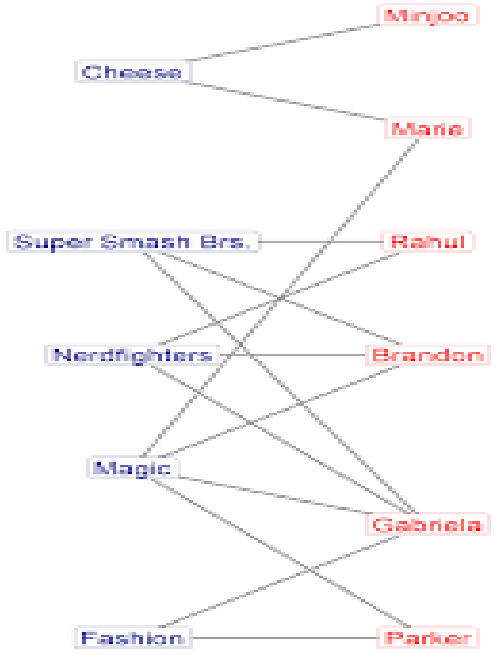


Social Network Analysis

- Attributes allows us to group the nodes into subgroups
  E.g. Affiliation Networks - network into clusters based on shared interests or affiliations

- Affiliation networks contain information about the relationships between two sets of nodes: a set of subjects and a set of affiliations

- Represented by *bipartite graph*, also known as a *two-mode network*



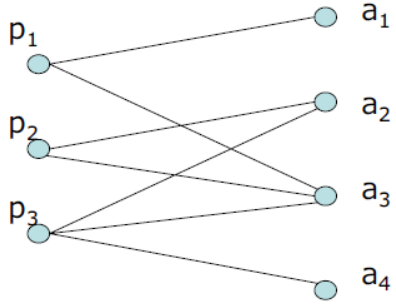
An n-partite graph or n-mode network is a graph $G = < V, E >$ where there exists a partitioning $V = \bigcup_{i=1}^{n} V_i$ such that $\bigcap_{i=1}^{n} V_i = 0$ and $(V_i \times V_i) \cap E = 0$

Social Network Analysis

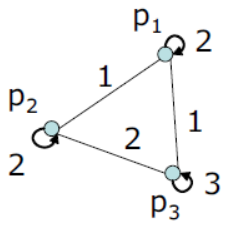
Affiliation networks are transformed directly to a regular, one-mode network

Transformation considers the overlaps between the affiliations as a measure of tie strength between the actors

E.g. Interlocking Directorates – overlaps in board membership of companies



	p_1	p_2	p_3	a_1	a_2	a_3	a_4
p_1	0	0	0	1	0	1	0
p_2	0	0	0	0	1	1	0
p_3	0	0	0	0	1	1	1
a_1	1	0	0	0	0	0	0
a_2	0	1	1	0	0	0	0
a_3	1	1	1	0	0	0	0
a_4	0	0	1	0	0	0	0



	p_1	p_2	p_3
p_1	2	1	1
p_2	1	2	2
p_3	1	2	3

Social Network Analysis

Summary

Some social structures revealed in Social Networks from global view

- Densely connected subgroups with sparse inter subgroup connection
- Core – Periphery structure with core node densely connected and periphery node may / may not connected with themselves, but with core nodes
- Affiliation network – Two / n-mode network, actors on one side, interest / affiliation shared on other sides, attributes may overlap