Social Networks

Oded Netzer*

Columbia Business School

*Thanks to Professors Christophe Van De Bulte and Zsolt Katona for material

Interdependence in Choice

$$U_{ijt} = X_{ijt}^{'}\beta_i + \varepsilon_{ijt}$$ i.i.d

- How could we make these choices interdependent?

Spatial Autoregressive Model

- Direct Effect of others
  - Others' **choices** directly affect my utility

$$U_{ijt} = X'_{ijt}\beta_i + \boxed{\tau \mathbf{W}\,\mathbf{Y}_{(t,-i)}} + \varepsilon_{it}$$

W is a similarity\network matrix

Oded Netzer - Columbia Business School
Social Networks

Auto-Correlated Errors

- Indirect Effect
  - My choices are interdependent with others

$$\mathbf{U}_{jt} = \mathbf{X}'_{jt}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{jt} + \boxed{\boldsymbol{\theta}}$$

$$\boxed{\boldsymbol{\theta} = \boldsymbol{\rho}\ \mathbf{W}\boldsymbol{\theta} + \mathbf{u}}$$

References:

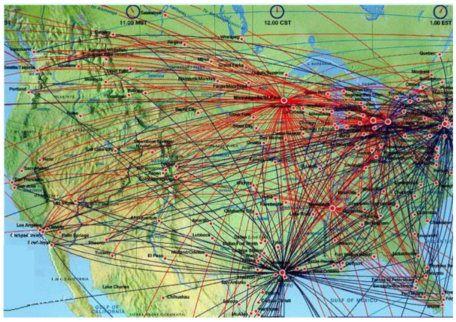
• Yang, Sha, Greg M. Allenby. (2003). "Modeling interdependent consumer preferences," *Journal of Marketing Research*. 40(3) 282–294.

•LeSage, James P. (2000), "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, 32 (1), 19-35.

Oded Netzer - Columbia Business School
Social Networks

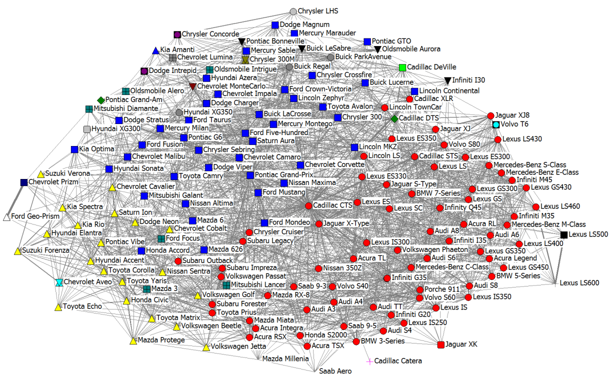
Where do we encounter networks?

Transportation



Where do we encounter networks?

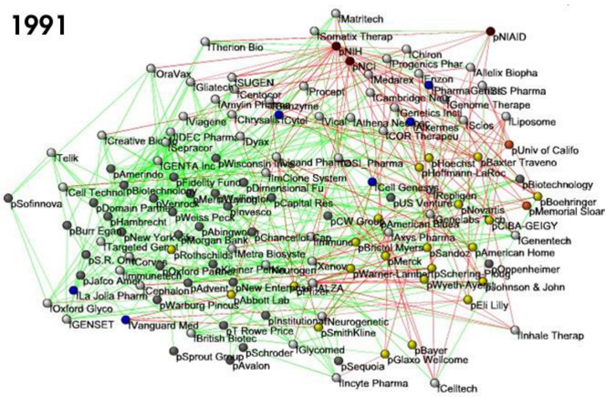
Word associations



Where do we encounter networks?

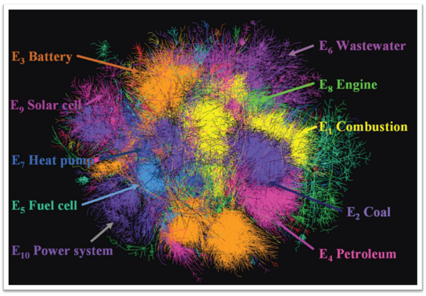
Business ties in biotech

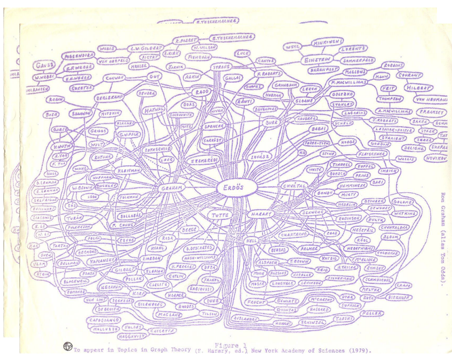
1991



Where do we encounter networks?

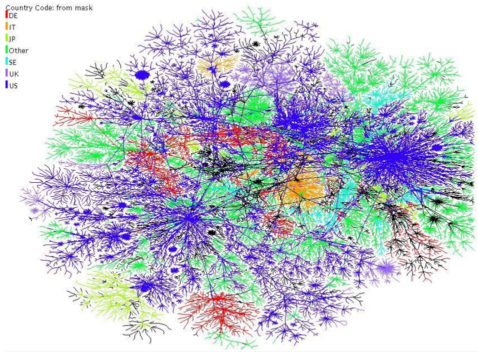
Academic citations





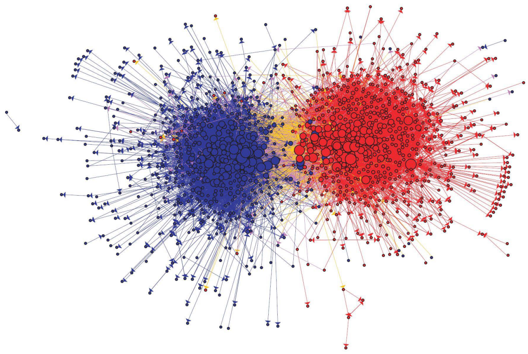
Where do we encounter networks?

Internet



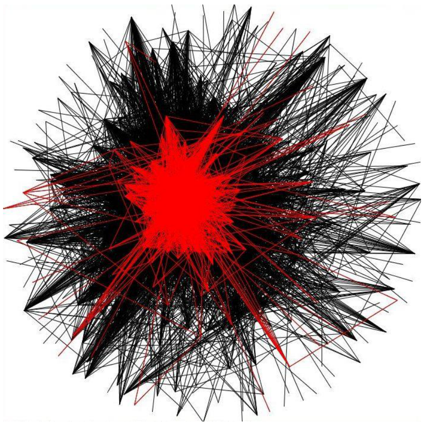
Where do we encounter networks?

Political blogsphere



Where do we encounter networks?

Social networks



Oded Netzer - Columbia Business School
Social Networks

Data Sources

- Direct observation (rare)
  - Anthropological field studies
  - Video recordings of gatherings
- Archival records
  - "Official" traces of ties: 10K filings, patent citations, patient referrals
  - Communications: telephone, email, web traffic, …
  - Very popular for organizational networks
  - Social media / Web 2.0 = archived self-reports of personal networks: "friends" on social networking sites like Facebook, trusted advisors on Epinions, Twitter followers, postings on threads, …
- Survey / self-reports
  - Most popular approach, esp. for personal networks (less for org. networks)
- Experiments
  - Create your own networks and see how people behave in them
  - Extremely high internal validity

Oded Netzer - Columbia Business School
Social Networks

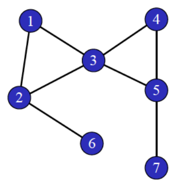
Analyzing Social Networks

- Definition of “network”:
  - a set of discrete entities (called nodes) and the collection of ties among them
- Definition of “social network”:
  - a set of discrete social entities (often called actors) and the collection of ties among them
- We are often interested in the structure of the ties among nodes
  - As explanatory variable
    - Information flow
    - Power, Resource mobilization
    - Trust and cooperation
    - Prestige and status
  - As dependent variable (statistically harder!)

Analysis

The Key Concept for Analysis

The network can be represented as a matrix

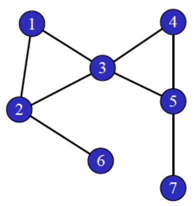


Nodes	1	2	3	4	5	6	7
1	0	1	1	0	0	0	0
2	1	0	1	0	0	1	0
3	1	1	0	1	1	0	0
4	0	0	1	0	1	0	0
5	0	0	1	1	0	0	1
6	0	1	0	0	0	0	0
7	0	0	0	0	1	0	0

If ties are valued, then entries in matrix will be valued as well

The Key Concept for Analysis

Simple graph

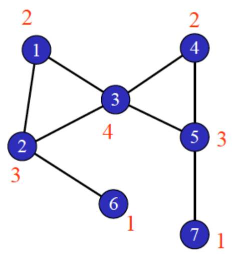


Simple Graph:
Undirected edges
No multiple edges
No loops

Nodes	1	2	3	4	5	6	7
1	~~0~~	1	1	0	0	0	0
2	1	~~0~~	1	0	0	1	0
3	1	1	~~0~~	1	1	0	0
4	0	0	1	~~0~~	1	0	0
5	0	0	1	1	~~0~~	0	1
6	0	1	0	0	0	~~0~~	0
7	0	0	0	0	1	0	~~0~~

Some basic concepts

Degree

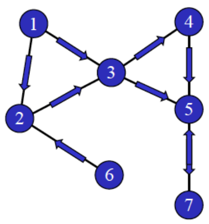


Nodes	1	2	3	4	5	6	7	
1	0	1	1	0	0	0	0	
2	1	0	1	0	0	1	0	=3
3	1	1	0	1	1	0	0	
4	0	0	1	0	1	0	0	
5	0	0	1	1	0	0	1	
6	0	1	0	0	0	0	0	=1
7	0	0	0	0	1	0	0	

Degree:
Number of edges incident to a vertex

Some basic concepts

Directed graph

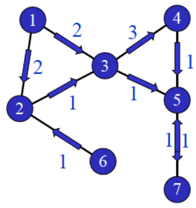


Out-degree

To / From	1	2	3	4	5	6	7	
1	0	1	1	0	0	0	0	
2	0	0	1	0	0	0	0	
3	0	0	0	1	1	0	0	2
4	0	0	0	0	1	0	0	
5	0	0	0	0	0	0	1	
6	0	1	0	0	0	0	0	
7	0	0	0	0	1	0	0	
In-degree		2						

Some basic concepts

Weighted directed graph



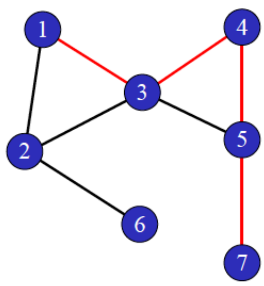
To / From	1	2	3	4	5	6	7	Out-degree
1	0	2	2	0	0	0	0	
2	0	0	1	0	0	0	0	
3	0	0	0	3	1	0	0	4
4	0	0	0	0	1	0	0	
5	0	0	0	0	0	0	1	
6	0	1	0	0	0	0	0	
7	0	0	0	0	1	0	0	
In-degree		3						

Some basic concepts

Paths

A path from Node 1 to Node 7

Length=4



Some basic concepts

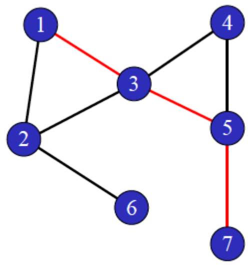
Shortest path

The shortest path from Node 1 to Node 7

Length=3

Distance: length of the shortest path

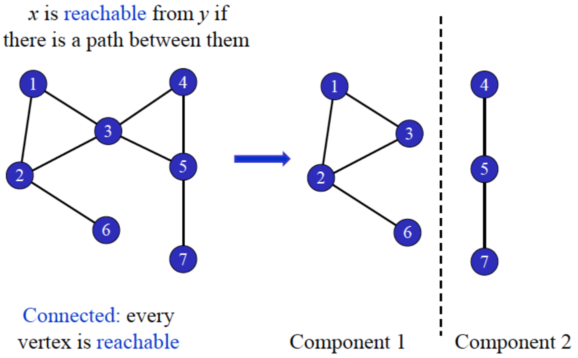
Diameter: The longest shortest path between any two nodes



Some basic concepts

Connectedness, Reachability

x is reachable from y if there is a path between them



Connected: every vertex is reachable from every other

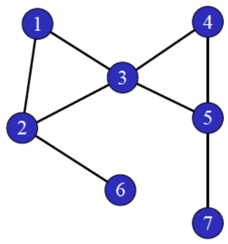
Storing and accessing network data

- Adjacency Matrix (huge)
- Node List:

  1: 2, 3
  2: 3, 6
  3: 4, 5
  4: 5
  5: 7

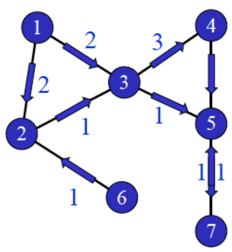
- Edge List:

  1, 2
  1, 3
  2, 3
  2, 6
  3, 4
  3, 5
  4, 5
  5, 7



Storing and accessing network data

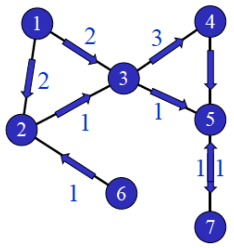
Directed edge list with with weights



1 2 2
1 3 2
2 3 1
3 4 1
3 5 1
4 5 1
5 7 1
6 2 1
7 5 1

Storing and accessing network data

Directed edge list with with weights



1	2	2
1	3	2
2	3	1
3	4	1
3	5	1
4	5	1
5	7	1
6	2	1
7	5	1

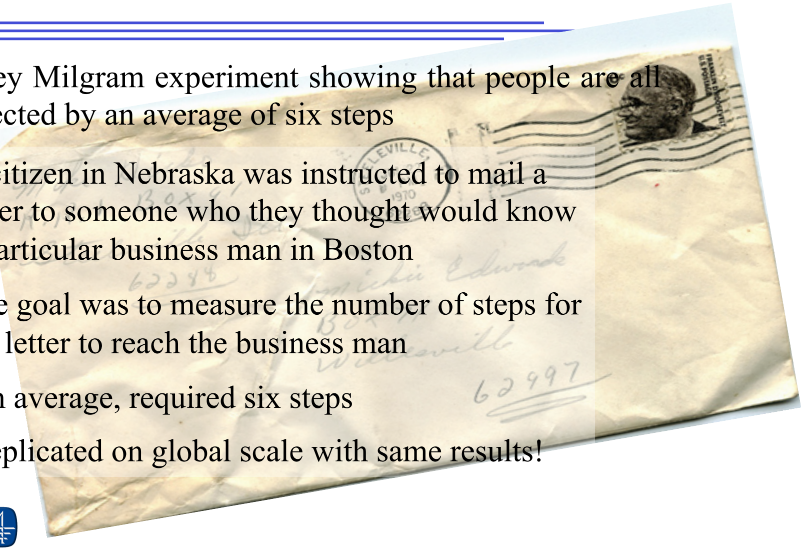
Basic operations on ties and matrices

Ties		**Matrix**
Reversing direction	→	Transposing
From directed to non-directed	→	Symmetrizing
From valued to binary	→	Dichotomizing

Six Degrees of Separation

Stanley Milgram experiment showing that people are all connected by an average of six steps

- A citizen in Nebraska was instructed to mail a letter to someone who they thought would know a particular business man in Boston
- The goal was to measure the number of steps for the letter to reach the business man
- On average, required six steps
- Replicated on global scale with same results!



It's a small world after all

The Kevin Bacon Game

Six Degrees of Kevin Bacon Game

Invented by Albright College students in 1994:
- Craig Fass, Brian Turtle, Mike Ginelly

Goal: Connect any actor to Kevin Bacon, by linking actors who have acted in the same movie.

Oracle of Bacon website uses Internet Movie Database (IMDB.com) to find shortest link between any two actors:

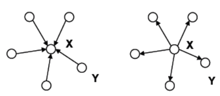
http://oracleofbacon.org/

Centrality and Centralization

- Centrality
  - A characteristic of a node's position
  - Importance or prominence of a node in the network
- Centralization
  - A characteristic of the network
  - Extent to which the network is structured around central nodes
  - Operationalized, in essence, as the amount of variation or heterogeneity in centrality across the nodes in the network

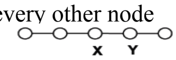
Freeman's Three Concepts / Measures

- Degree
  - Concept = Activity
  - Measure = Number of ties
  - In directed graph: indegree and vs. outdegree

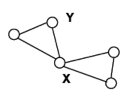


indegree outdegree

- Closeness
  - Concept = Access or reach
  - Measure = Inverse of the sum of the geodesic distance to every other node in the network



- Betweenness
  - Concept = Being in the middle and hence Control
  - Measure = Extent to which one is on the shortest path between any other two nodes in the network
  - Burt's brokerage idea



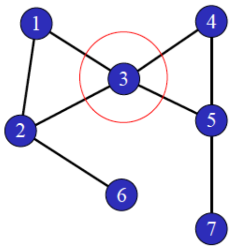
betweenness

Centrality - Degree

Degree centrality of a node = number of neighbors $=d_x$

Normalized degree centrality $=d_x/(n\text{-}1)$

Example (Node 3):
Degree centrality=4
Normalized=4/6=0.6667

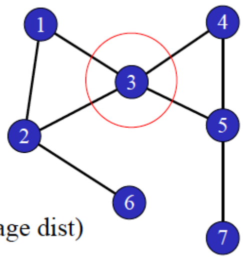


Centrality - Closeness

Closeness centrality: Inverse sum of distances to other nodes

Closeness centrality of node x = $\dfrac{1}{\sum_y dist(x,y)}$

Normalized: $\dfrac{n-1}{\sum_y dist(x,y)}$



Example (node 3)

1/(1+1+1+1+2+2)=0.125

Normalized=6/8=0.75 (inverse of average dist)

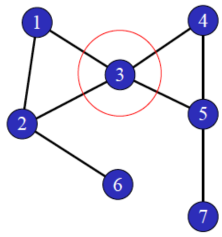
Centrality - Betweeness

Betweenness centrality: Potential for control of communication

Betweenness centrality of node x = $\sum_{y<z} \frac{g_{yz}(a)}{g_{yz}}$

Where g_{yz} is the number of geodesics between y and z, and $g_{yz}(a)$ is the number of geodesics between y and z that contain x

Normalized: $\frac{2}{(n-1)(n-2)} \sum_{y<z} \frac{g_{yz}(a)}{g_{yz}}$



Example (node 3)
1+1+1+1+1+1+1+1+1=9
Normalized=9/15=0.6

Centrality - Eigenvector

Eigenvector centrality: An iterative process that takes into account: the number of friends you have, the number of friends these friends have, ….

Sum of a nodes connection to other nodes weighted by their centrality

Google's initial Page Rank measure is based on the same idea.

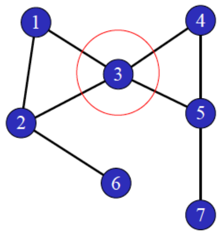
Clustering

Clustering coefficient: Density of connections between a node's neighbors

Clustering coefficient of node x = the proportion of existing edges to all possible edges between your neighbors.

Example (node 3)

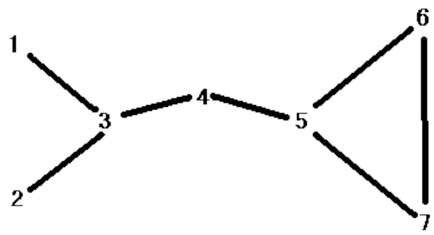
2/6=0.33



Some Additional Concepts / Measures

- Information centrality
  - A variant of betweenness
  - Instead of looking only at shortest paths (geodesics), one looks at all paths, with weights depending on their length (weight = inverse of length)
  - Advantage: this metric is more sensitive to the impact of connections to "peripheral" nodes who rarely lie on geodesics

Network Example



Oded Netzer - Columbia Business School
Hidden Markov Models

Source: Motoki Watabe 1998

Network Example

Node	Degree	Closness	Betweeness
1			
2			
3			
4			
5			
6			
7			

Oded Netzer - Columbia Business School
Social Networks

Source: Motoki Watabe 1998

Power Law and Scale-Free Networks

- The distribution of node degree often approximates a power law distribution

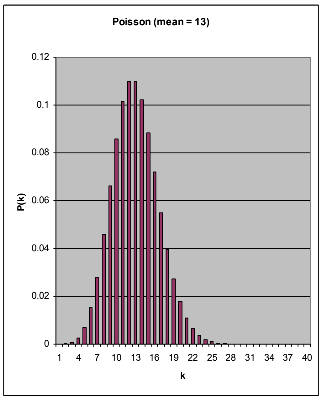
$$\boldsymbol{K} = \textbf{node degree}$$

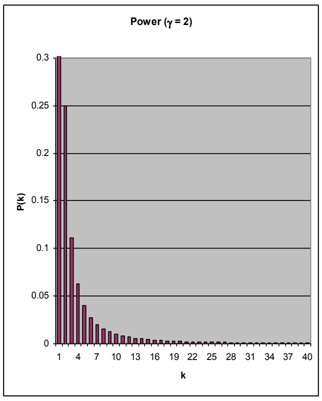
$$\mathbf{P}(\boldsymbol{K} = \boldsymbol{k}) \sim \boldsymbol{k}^{-\mathbf{g}}$$

$$\mathbf{lnP}(\boldsymbol{k}) \sim \mathbf{-g\ ln}\boldsymbol{k}$$

$$P(\text{linking to node i}) = \frac{k_i}{\sum_j k_j}$$

Poisson vs. Power Distribution





Power Law in Networks

Albert and Barabasi (1999)

Power-law distributions are straight lines in log-log space.

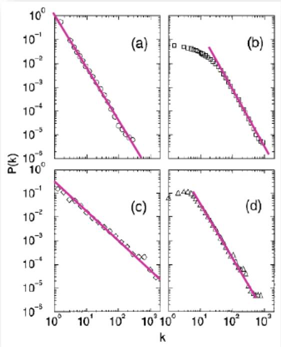
- slope being r

$y=k^{-r} \rightarrow \log y = -r \log k \rightarrow ly = -r\ lk$

Power laws in real networks:
(a) WWW hyperlinks
(b) co-starring in movies
(c) co-authorship of physicists
(d) co-authorship of neuroscientists

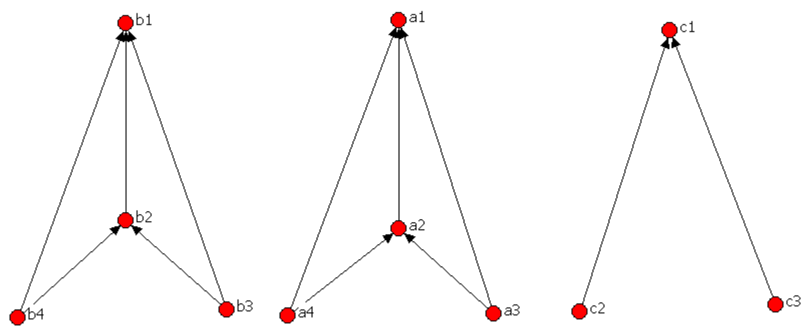
Why? As Number of nodes grow edges are added in proportion to the number of edges a node already has.

Alternative: Copy model—where the new node copies a random subset of the links of an existing node



Identifying Subgroups

How Would You Group these Nodes?



Grouping based on Network Structure

- This requires a measure of “distance” or “similarity”
  - Standard approach outside SNA = based on actor attributes
  - In network analysis, we typically want to group actors not based on their attributes, but based on their pattern of ties

- Relational versus positional approaches
  - Relational approach
    - Grouping actors into sub-graphs based on connectedness (direct, indirect, interconnectedness/density)
    - Often associated with a logic of “Cohesion” (access, support, …)
  - Positional approach
    - Grouping actors based on similarity in their position in the network (their portfolio of ties)
    - Often associated with a logic of “Equivalence” (status, substitutability and competition, …)

Contagion vs. Homophily

- **Contagion** – Social influence - One person performing an action can **cause** her contacts to do the same.
- **Homophily** - Birds of a feather flock together - people interact more often with people who are "like them" than with people who are dissimilar.

Contagion vs. Homophily

"If your friend Joey jumped off a bridge, would you jump too?"

1.yes: Joey inspires you (social contagion or influence)
2.yes: Joey infects you with a parasite which suppresses fear of falling (actual contagion)
3.yes: you're friends because you both like to jump off bridges (manifest homophily)
4.yes: you're friends because you both like roller-coasters, and have a common risk-seeking propensity (latent homophily)
5.yes: because you're both on it when it starts collapsing and that's the only way off (external causation)

Network Analysis Software

- Gehpi — Gephi makes graphs handy
- UCINET — UCINET Software
- Pajek — **Networks / Pajek**
- SIENA