RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

RISIS
Research infrastructure for research and innovation policy studies

Basic concepts of Social Network Analysis

Antonio Zinilli
National Research Council (CNR)

Rome, November 20-21, 2017

Consiglio Nazionale delle Ricerche
IRCrES
ISTITUTO di RICERCA sulla CRESCITA ECONOMICA SOSTENIBILE
INSTITUTE of RESEARCH on ECONOMIC SUSTAINABLE GROWTH

Objectives for Today

- Understand what network analysis is
- Introduce basic concepts
- Introduce main network measures
- Overview methodological approaches

What Are Networks?

- Networks are patterns of relationships that connect individuals, institutions, or objects (or leave them disconnected).

- EXAMPLES
- Individuals' co-memberships in organizations
- Relationship between countries

When to Study Networks?

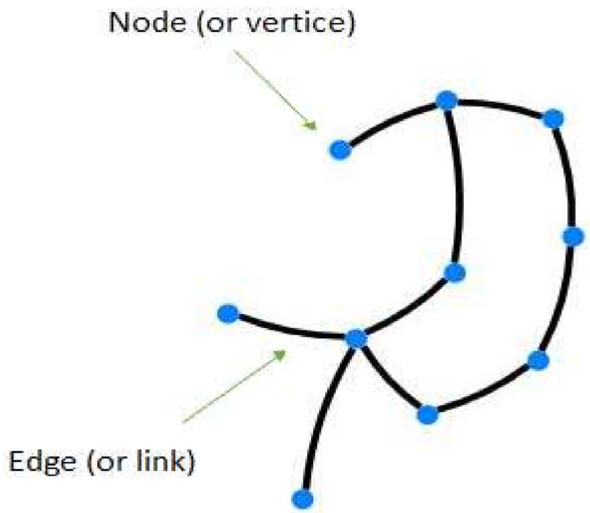
- It is possible applied network analysis tecniques in more contexts. But the question we want to ask is: when in the network aspect of phenomenon particularly pertinent to the social dynamics that matter to us?

- Network analysis tends to place a strong emphasis on the *relationship* (or “the dyad”) as a unit of analysis.

Graphs

Social networks can be represented as **graphs**

Graphs are made up of **nodes** (i.e., actors, cities, organizations, articles etc.) that are connected by **links** (i.e., relationships, membership, citations etc.).



Types of Links

Undirected vs. directed links

Undirected links, identified with a simple straight line**,** are used when there is a symmetric relationship:

E. g. Facebook has always used a symmetric model, if you add someone as a friend they have to add you as a friend as well.

Directed links, identified with arrows**,** are used when there is an asymmetry in a relationship:

E. g. On Twitter you can "follow" someone else without them following you back.

Dichotomous vs. Valued Links

Dichotomous vs Valued Links

Dichotomous Links: either a link exists or it doesn’t (e.g. if we are friends or we are not friends, there is a collaboration or not etc..) :

Valued Links: the links vary on the based of a weight (strength) (e.g. our friendship may be strong or weak, the number of times that each pair of countries cooperate):

Main parts of a graph

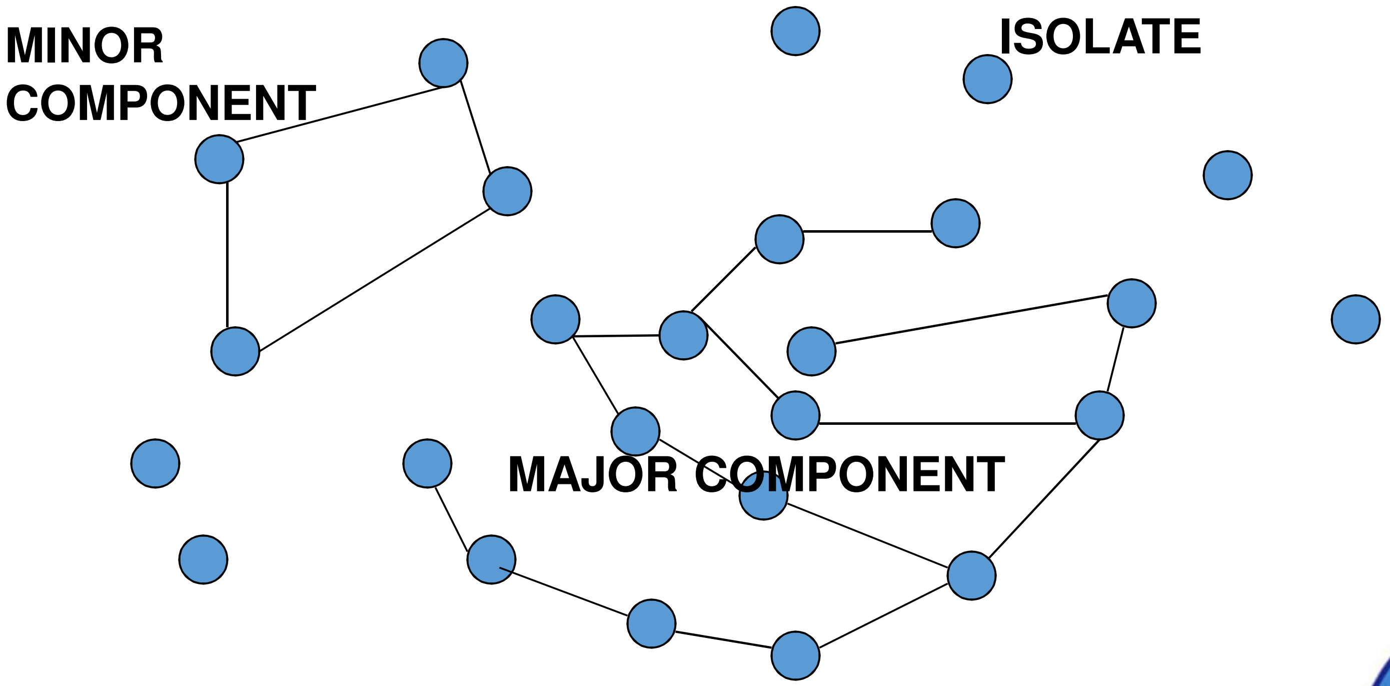
- **Component**: all nodes that assemble a connected subgraph within a network:

  main component is the largest component within a network;

  minor component is a component that is smaller than the main component. Usually there are more minor components.

- **Isolate:** a node that has no links to the other nodes within the network

Main parts of a graph



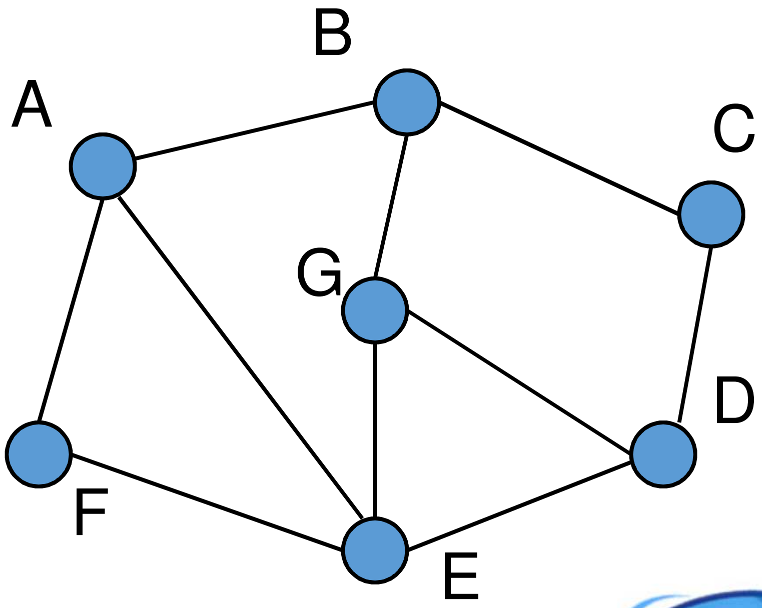
Main Graph Implementation Strategies

- Edge List
- Adjacency Matrix

Edge List

Edge list: an unordered list of all edges in the graph

A	A	A	B	B	C	E	E	E	G
B	E	F	G	C	D	F	G	D	D

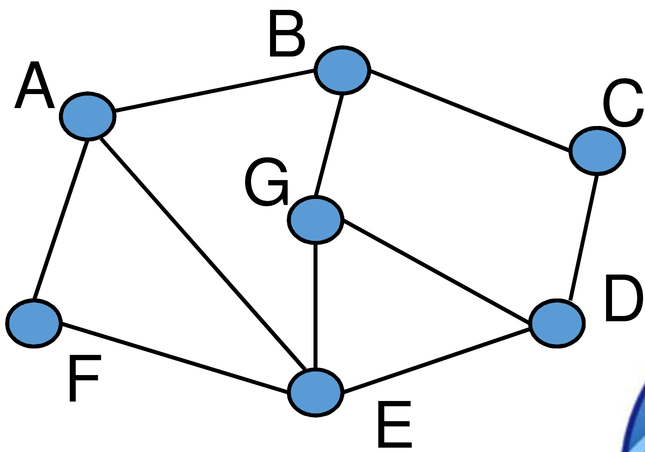


Matrices

The most basic matrix is an adjacency matrix: an *nxn* matrix where:

- the nondiagonal entry a_{ij} is the number of edges joining vertex *i* and vertex *j* (or the weight of the edge joining vertex *i* and vertex *j*). 1 indicates the presence of a link, while a 0 indicates the absence of a link.
- the diagonal entry a_{kk} corresponds to the number of loops (self-connecting edges) at vertex *k*. Usually loops are not counted.

$$
M = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \\ G \end{array}
\begin{array}{c} \begin{array}{ccccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} & \mathrm{G} \end{array} \\
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0
\end{bmatrix} \end{array}
$$



Matrices

	Austria	France	Germany	Italy
Austria	0	0	1	0
France	0	0	0	1
Germany	1	0	0	0
Italy	0	1	0	0

A 1 indicates the presence of a link, while a 0 indicates the absence of a link.

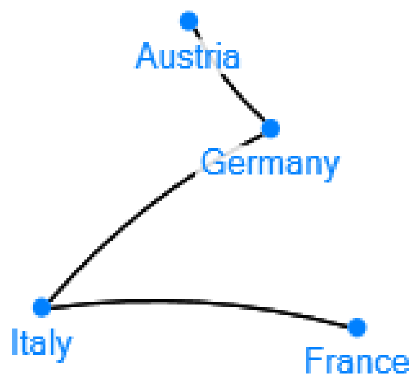
Symmetric Matrices

	Austria	France	Germany	Italy
Austria	0			
France	0	0		
Germany	1	0	0	
Italy	0	1	0	0

If matrices are symmetric, they may be represented by upper or lower triangle only.

One-Mode Network

- Network analysis typically involves only one mode. A mode is a class of nodes in a network.



Two-Mode Networks

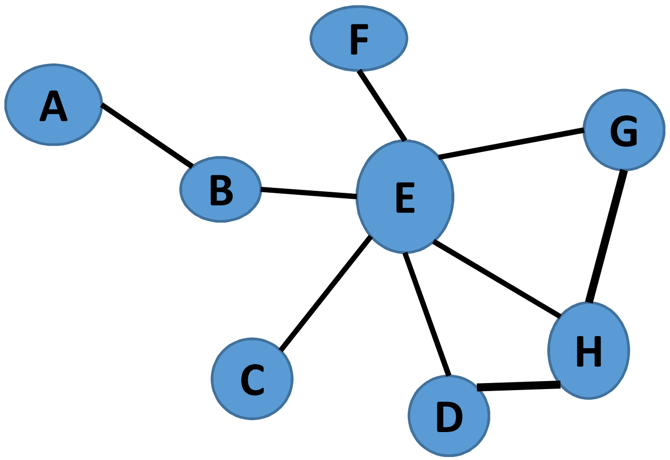
Mode 1	Mode 2
People	Events
Students	Universities
Countries	Projects or Programmes

Example: Two-mode data have countries and research programmes. France and Italy collaborate at the same programme.

Basic Network Statistics

- Path
- Geodesic distance
- Density
- Degree centrality
- Closeness centrality
- Betweenness centrality
- Clustering Coefficient
- Eigenvector centrality

Path



- ABEDHG is a path from A to G. There are multiple paths from A to G.
- Path length is the number of steps in a path. The path length of ABEDHG is equal to 5.

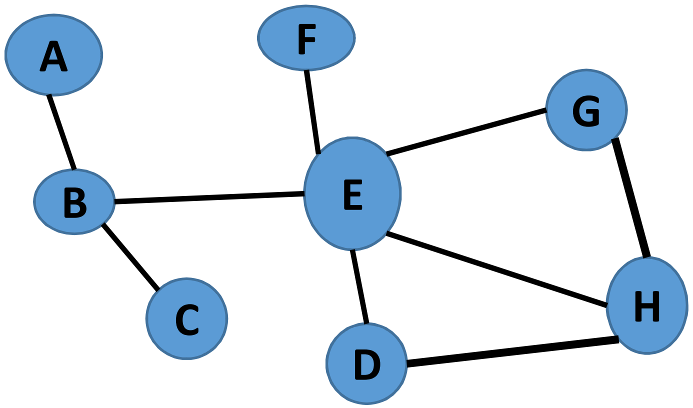
Geodesic distance

- Geodesic distance is the shortest path from one node to another node

- The shortest path is the path that achieves that distance.

- The average network diameter is the average of shortest path lengths over all pairs of nodes in a network.

Geodesic distance: an example



ABEG is the geodesic from A to G

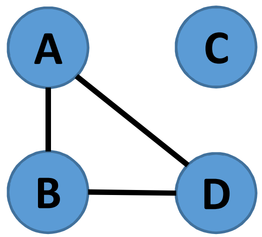
Density

- Density is a property of a network.

- The proportion of links in a network relative to the total number possible.

Density: an example

In this example we have 3 links and 4 vertices



Network Density:

$$\frac{\textit{Actual connections}}{\textit{Potential connections}}$$

Potential connections:

$$PC = \frac{n(n-1)}{2}$$

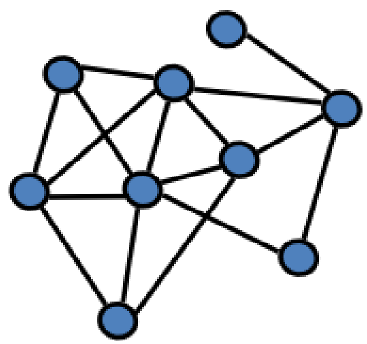
Potential connections = [(4 (4-1))/2] = 12/2 = 6

Density: 3/6=0,5

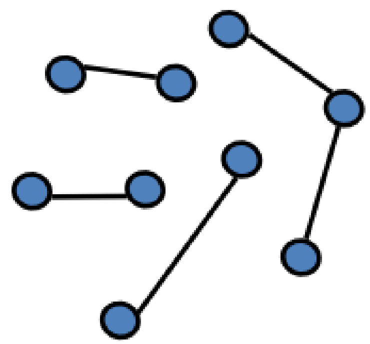
This graph has one half of all possible links.

Density

- Proportion of ties in a graph



High density (44%)

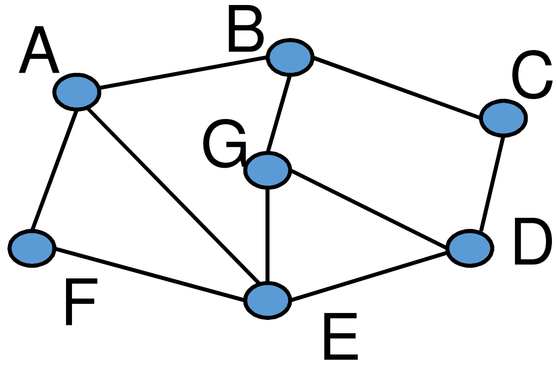


Low density (14%)

Degree

- Degree is a property of a node. The degree of a node is equal to the number of links that it has.

- B has a degree of 3.



The average degree of a graph is given by

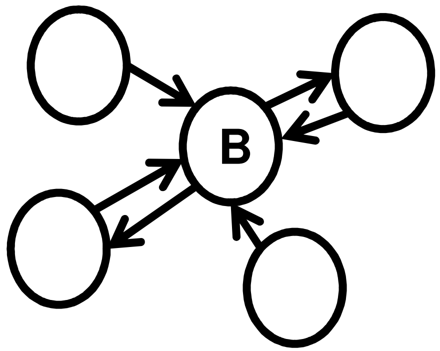
$$\bar{k} = \frac{1}{g}\sum_{i=1}^{g} k(n_i)$$

Degree distribution

- A degree distribution a property of a network.
- A degree distribution is the number of nodes of a network that have each degree level.
- A degree distribution may be a good way of summarizing the activity of nodes in a network. This measure provides a first indication on the importance of a node; important nodes are those that have a greater influence to the flow of information in a network.
- May be a good way of comparing networks to one another.

Indegree and Outdegree

- **Directed networks only**
- **Indegree**: The number of links that a node **receives** from other nodes
- **Outdegree:** The number of links that a node **sends** to other nodes



- What is B’s indegree? Answer: 4
- What is B’s outdegree? Answer: 2

Closeness centrality

Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

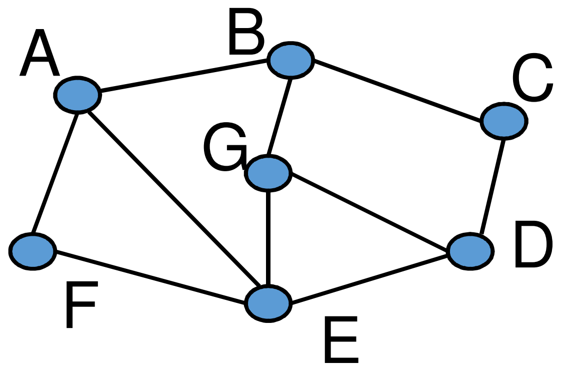
Closeness Centrality:

$$C_c(i) = \left[\sum_{j=1}^{N} d(i,j) \right]^{-1}$$

Normalized Closeness Centrality

$$C_C^{'}(i) = (C_C(i)) / (N-1)$$

Closeness centrality: an example



$$C_c^{'}(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[\frac{\mathrm{d(AB)+d(AC)+d(AD)+d(AE)+d(AF)+d(AG)}}{N-1} \right]^{-1} = \left[\frac{1+2+2+1+1+2}{6} \right]^{-1} = \left[\frac{9}{6} \right]^{-1} = 0.66$$

Betweenness centrality

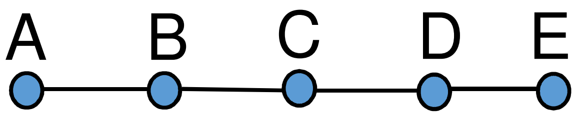
Betweenness centrality indicates the extent to which a vertex lies on paths between other vertices.

paths between j and k that pass through *i*

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

all paths between j and k

Betweenness centrality: an example



In this simple example there are no alternate paths.

A lies between no two other vertices

B lies between A and 3 other vertices: C, D, and E - (AC), (AD), (AE)

C lies between 4 pairs of vertices: A, B, D, E - (A,D),(A,E),(B,D),(B,E)

D lies between E and 3 other vertices: A, B, and C - (AE), (BE), (CE)

E lies between no two other vertices

We can conclude that C gets full credit.

Clustering coefficient

- **Clustering coefficient** is a measure of the degree to which nodes in a graph tend to cluster together.
- We can speak about:
- **Local Clustering coefficient**
- **Global Clustering coefficient**

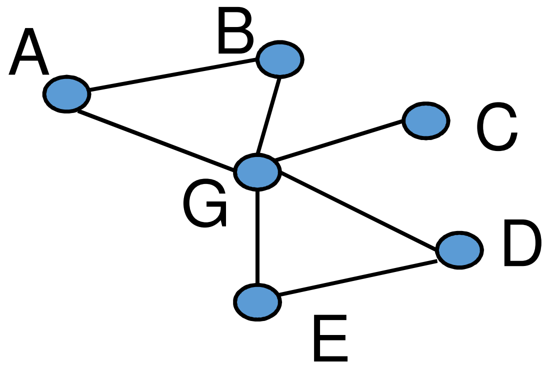
Local Clustering coefficient

The local clustering coefficient is defined as the ratio of the observed connections between all neighbours, and all possible connections between the neighbours

$$C_i = \frac{b_i}{\dfrac{n!}{k!(n-k)!}}$$

with *bi* indicates the number of observed links between all neighbours of a specific node; *Ki* indicates all possible connections between the neighbours.

Example



$$C_i = \frac{2}{\frac{n!}{k!(n-k)!}} = \frac{2}{\frac{5!}{2!(5-2)!}} = \frac{2}{\frac{5!}{2!(5-2)!}} = \frac{2}{\frac{120}{12}} = 0,2$$

Global Clustering coefficient

The global clustering of a graph is given by the average of all local clustering coefficients

$$\bar{C} = \frac{1}{g} \sum_{i=1}^{g} C_i$$

Eigenvector centrality

Eigenvector: is a measure of centrality that takes into account the centrality of other nodes to which a node is connected.

$$E_i = \frac{1}{\lambda} \sum_{j<k} a_n n_i$$

i is a node that is distinct from *j* and *k*
a_n is an adjacency matrix with *n* nodes,
n_i is the realized value of a link in the network,
λ is an eigenvector solved through an iterative algorithm.

Other Measures of centrality

There are a large number of other possible measures of centrality.

K-star, bridge, Transitivity etc.

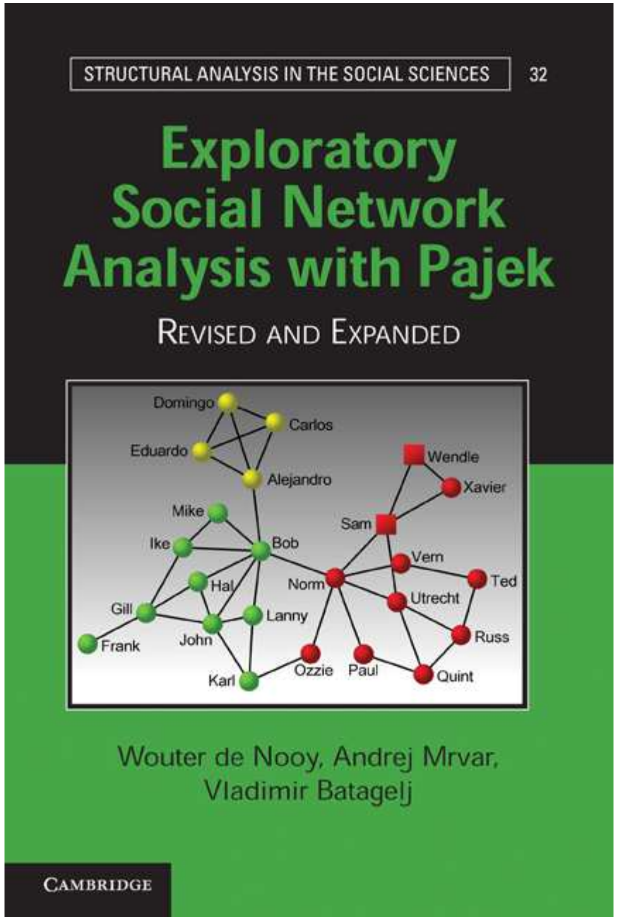
Usually, these different measures are highly correlated

More information

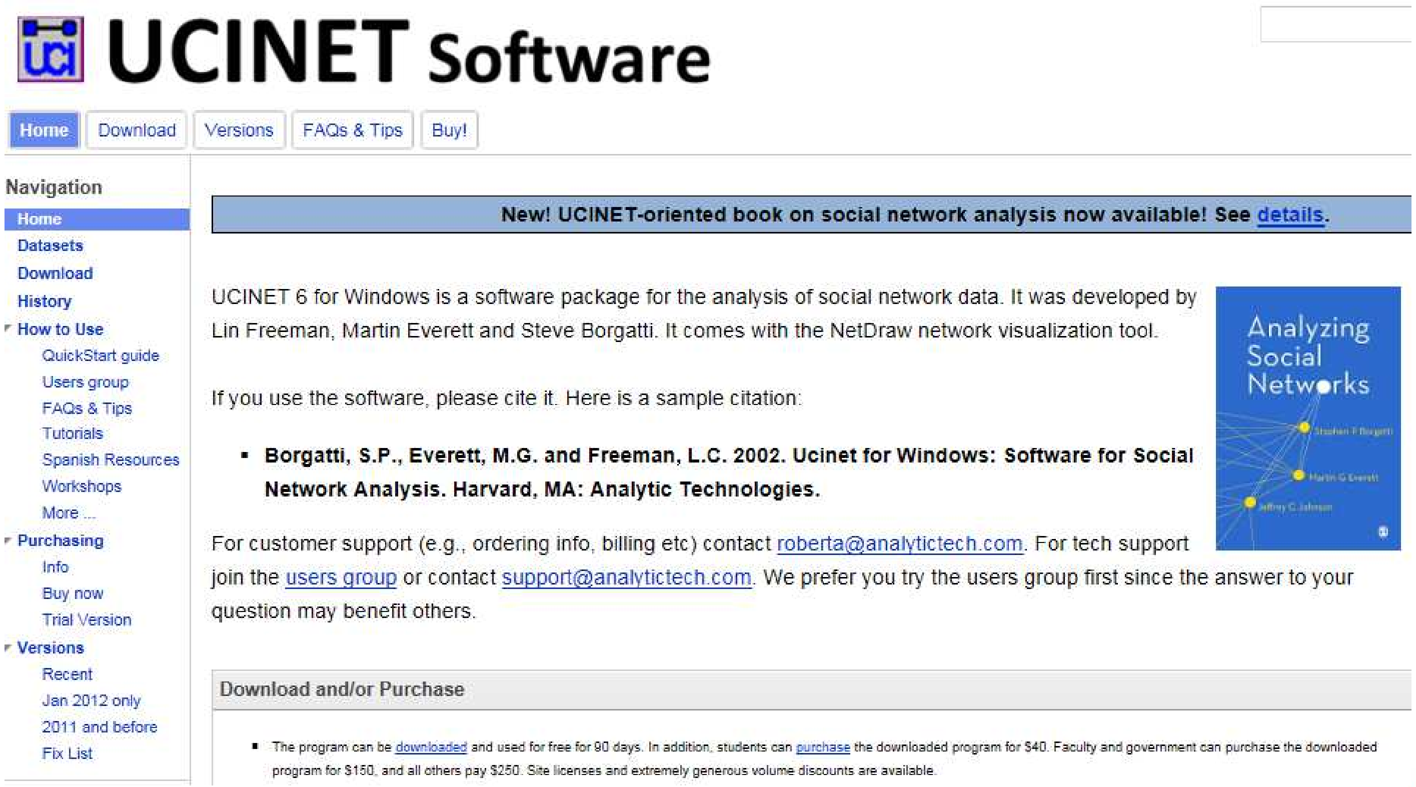
- Usually, measures of centrality are used as independent variable.

- Usually the network ties are used as dependent variable.

PAJEK


STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES 32
Exploratory Social Network Analysis with Pajek
Revised and Expanded
Domingo
Carlos
Eduardo
Wendle
Alejandro
Xavier
Mike
Sam
Ike
Bob
Vern
Norm
Hal
Ted
Utrecht
Gill
Lanny
Russ
Frank
John
Ozzie
Paul
Karl
Quint
Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj
Cambridge

UCINET



UCINET Software

Home | Download | Versions | FAQs & Tips | Buy!

Navigation

- Home
- Datasets
- Download
- History
- How to Use
  - QuickStart guide
  - Users group
  - FAQs & Tips
  - Tutorials
  - Spanish Resources
  - Workshops
  - More ...
- Purchasing
  - Info
  - Buy now
  - Trial Version
- Versions
  - Recent
  - Jan 2012 only
  - 2011 and before
  - Fix List

New! UCINET-oriented book on social network analysis now available! See details.

UCINET 6 for Windows is a software package for the analysis of social network data. It was developed by Lin Freeman, Martin Everett and Steve Borgatti. It comes with the NetDraw network visualization tool.

Analyzing Social Networks

If you use the software, please cite it. Here is a sample citation:

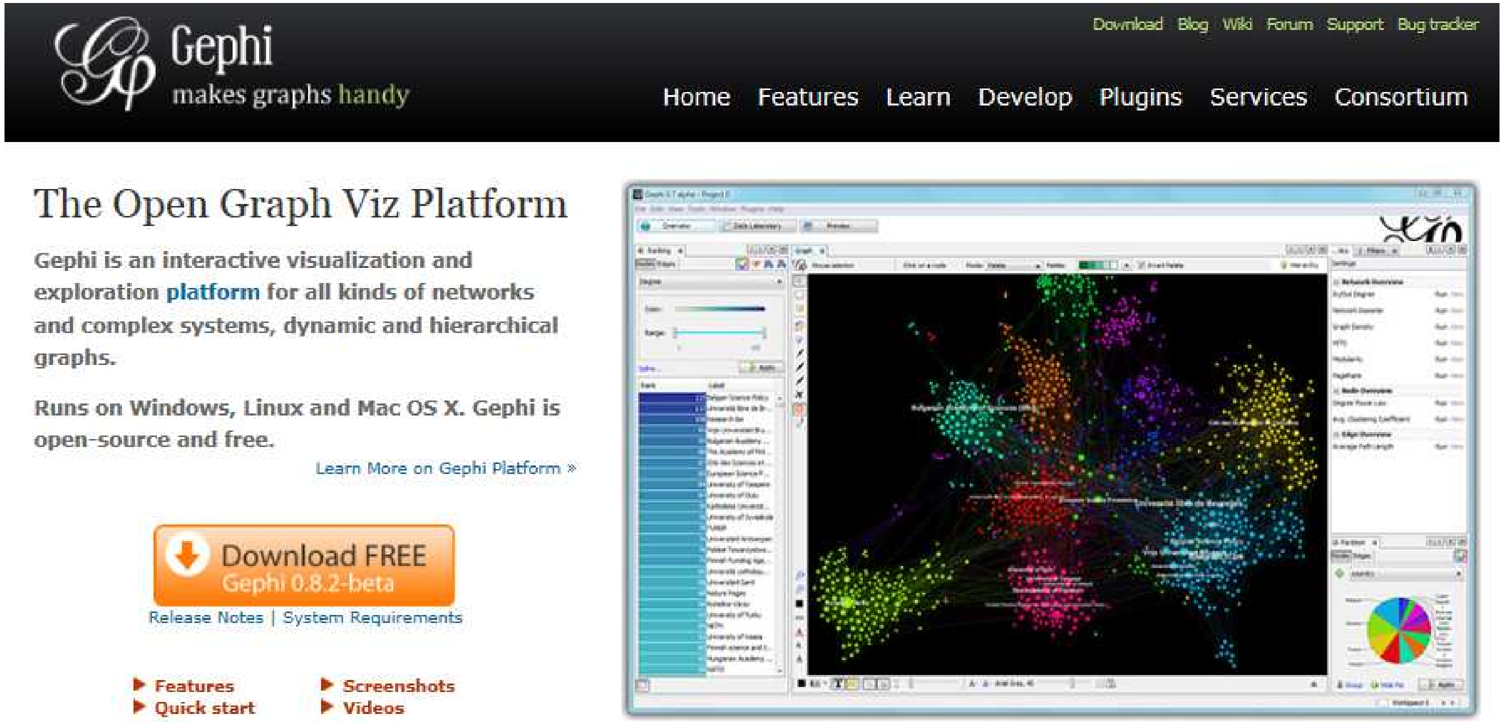
- **Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.**

For customer support (e.g., ordering info, billing etc) contact roberta@analytictech.com. For tech support join the users group or contact support@analytictech.com. We prefer you try the users group first since the answer to your question may benefit others.

Download and/or Purchase

- The program can be downloaded and used for free for 90 days. In addition, students can purchase the downloaded program for $40. Faculty and government can purchase the downloaded program for $150, and all others pay $250. Site licenses and extremely generous volume discounts are available.

GEPHI


Gephi
makes graphs handy
Download Blog Wiki Forum Support Bug tracker
Home Features Learn Develop Plugins Services Consortium
The Open Graph Viz Platform
Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.
Runs on Windows, Linux and Mac OS X. Gephi is open-source and free.
Learn More on Gephi Platform »
Download FREE
Gephi 0.8.2-beta
Release Notes | System Requirements
Features
Quick start
Screenshots
Videos

Textbooks

- Hanneman & Riddle (2005) Introduction to Social Network Methods, available at http://faculty.ucr.edu/~hanneman/nettext/
- Wasserman & Faust (1994): Social Network Analysis – Methods and Applications, Cambridge: Cambridge University Press.