Basic SNA techniques and measures

Thomas Scherngell
AIT Innovation Systems Department

Course
SOCIAL NETWORK ANALYSIS. INTRODUCTION TO METHODS AND APPLICATIONS TO THE EUPRO DATABASE

LUGANO, FEBRUARY 16-18 2016

AUSTRIAN INSTITUTE OF TECHNOLOGY

Social Network Analysis (SNA)

- Focus on relationships among actors, and on the patterns and implications of these relationships (see, e.g., Scott 2000)

- Assumption that actors participate in systems connecting them to other actors, whose relations influence one another's behaviors (see, e.g., Knoke and Young 2008)

- Identification of structural forms, substructures and substantive contents of relations among actors is of central interest

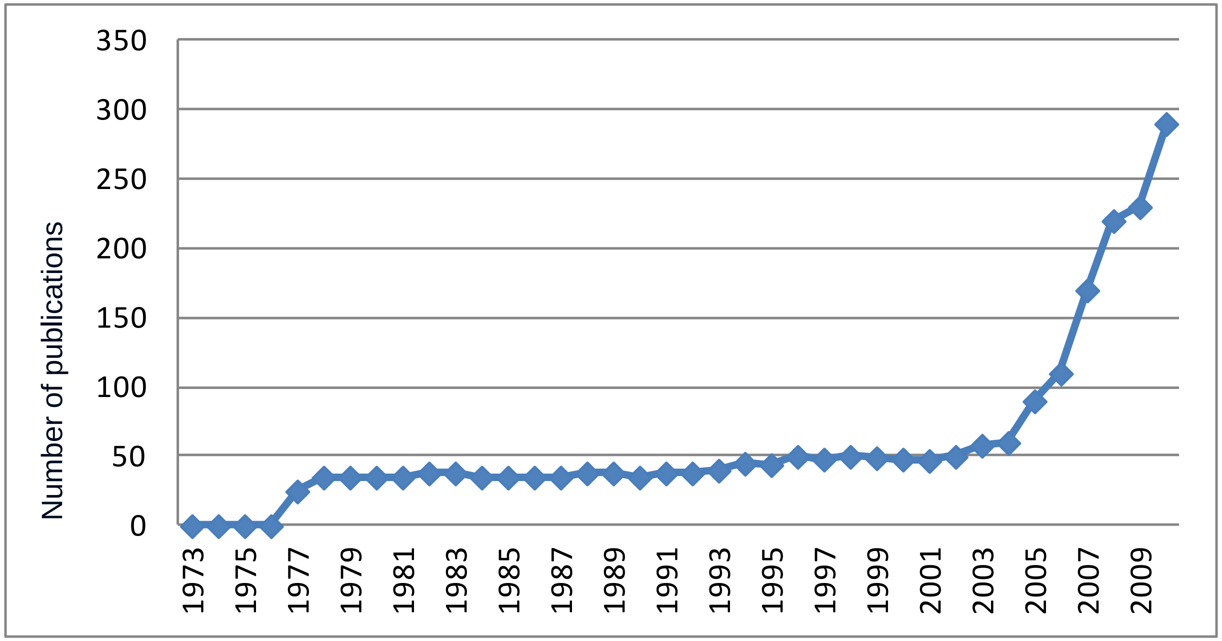
→ Requires a set of procedures and analytical concepts distinct from conventional statistics and data analysis, referred to as SNA (Wasserman and Faust 1994)

Outline

- A formal representation of a social network
- The sociomatrix
- Different formal network types
- Global and local network structural measures
- SNA tools

The increasing importance of SNA

- Networks and complex systems have gained increased importance in recent years in the scientific domain,
- In particular in sociology, economics, physics, geography and biology



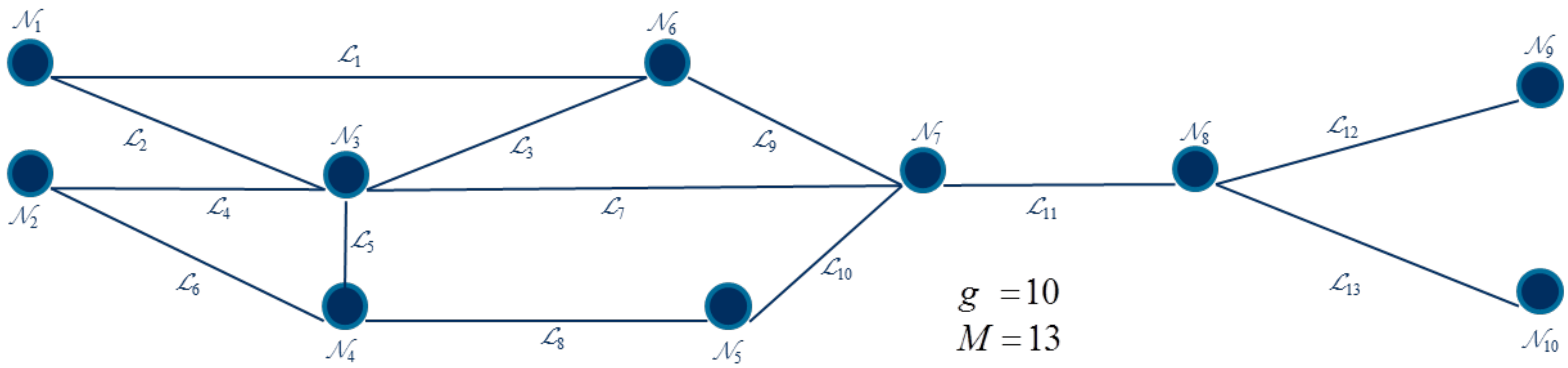
Increase of publications (SSCI and SCI) featuring the notion of SNA (1973-2009)

A formal network definition

Formally, a network is a set $\mathcal{N}$ of vertices, representing actors, and a set $\mathcal{L}$ of edges, representing interactions between the vertices;

$$\mathcal{N} = \{N_1, N_2, \ldots, N_g\}; \quad \mathcal{L} = \{L_1, L_2, \ldots, L_M\}$$

A hypothetical example



Graph theory as appropriate instrument to describe such a system (Wasserman und Faust 1994), with $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ referred to as simple graph

The sociomatrix

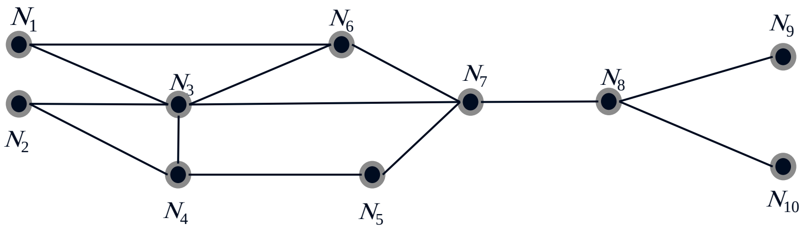
The graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ can also be described in form of a sociomatrix, defined as (Knoke und Yang 2008):

$$\boldsymbol{X} = (x_{ij}) \qquad i, j = 1, \ldots, g \tag{1}$$

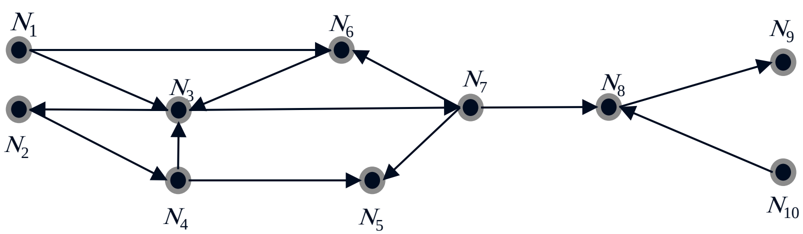
with the element x_{ij} denoting whehther two nodes n_i and n_j are connected to each other

	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	Total
N_1	0	0	1	0	0	1	0	0	0	0	2
N_2	0	0	1	1	0	0	0	0	0	0	2
N_3	1	1	0	1	0	1	1	0	0	0	5
N_4	0	1	1	0	1	0	0	0	0	0	3
N_5	0	0	0	1	0	0	1	0	0	0	2
N_6	1	0	1	0	0	0	1	0	0	0	3
N_7	0	0	1	0	1	1	0	1	0	0	4
N_8	0	0	0	0	0	0	1	0	1	1	3
N_9	0	0	0	0	0	0	0	1	0	0	1
N_{10}	0	0	0	0	0	0	0	1	0	0	1
Total	2	2	5	3	2	3	4	3	1	1	26

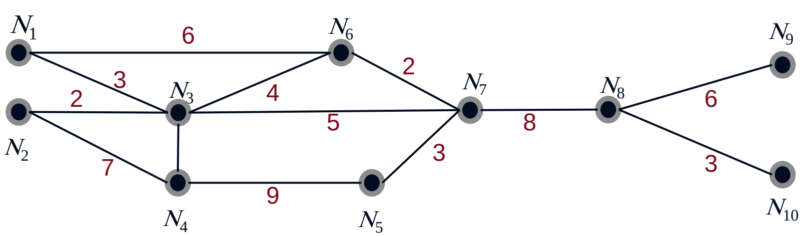
Different formal network types



Basic network
(symmetric: binary, no direction, no weighting)



Directed network
(asymmetric; binary)



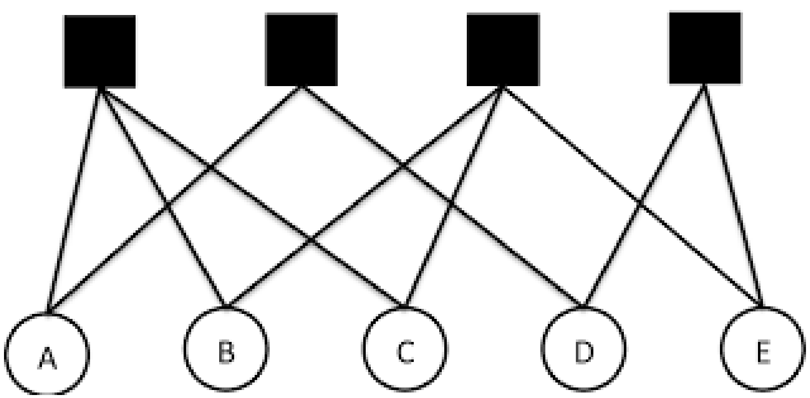
Weighted network:

$\mathcal{G} = (\mathcal{N}, \mathcal{L}, \mathcal{V})$ with

$\mathcal{V} = \{v_1, v_2, ..., v_M\}$

The notion of bipartite networks

- Bipartite networks have two distinct sets of nodes
- Links exist between the sets, but never within the sets



Node set 1 (e.g. persons, organisation, regions, countries)

Participation / Relation

Node set 2 (e.g. events, movies, publications, projects)

- Empirically very important approach for network definition; arguably the empirical default case

Fundamental concepts: Degree

The number of links of a node n_i is referred to as degree (k_i) of node i

The average degree of a graph is therefore given by

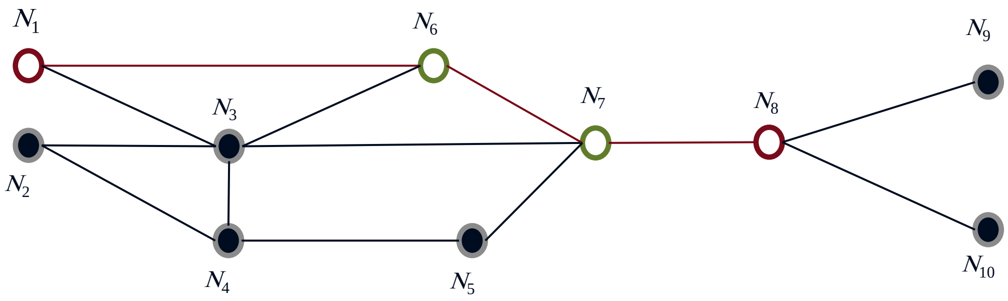
$$\overline{k} = \frac{1}{g}\sum_{i=1}^{g} k(n_i) \qquad (2)$$

In directed networks,the indegree, denotes the number of links *coming into* node n_i, the outdegree the number of links *going out* from node n_i

The *degree* provides a first impression on the prominence of an actor; prominent actors are assumed to influence the flow of information in a network, and can therefore more easily access and mobilise network resources

Fundamental concepts: Path

A path S is defined as a sequence of distinct nodes and edges between two nodes n_i and n_j (Knoke und Yang, 2008; Scott, 2000)



$S(N_1, N_6, N_7, N_8)$

A path between two nodes n_i and n_j exists, if the two nodes are connected by a sequence of edges (without touching one node more than once)

The length of a path $S(n_i, n_j)$ is defined as the number of entailed edges

Note that not all node pairs in the network are necessarily connected by a path. Groups of nodes that are reachable by a path in the network are called components

Fundamental concepts: Geodesic distance

The geodetic distance d_{ij} is defined as the length of the shortest path between two nodes n_i and n_j (Knoke und Yang, 2008; Scott, 2000)

Thus, the average geodesic distance in a graph is defined as

$$\bar{d} = \frac{1}{g(g-1)} \sum_{i,j=1}^{g} d_{ij} \tag{3}$$

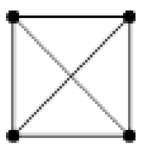
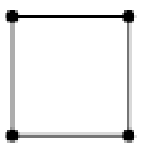
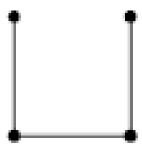
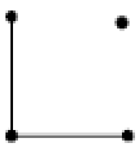
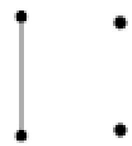
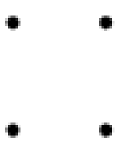
In relation, the diameter of a graph is defined as the longest geodetic distance between any two nodes in a graph

Both measures are central indicators for the connectedness of a social system (Brandes und Erlebach 2005)

Fundamental concepts: Density

The network density is defined as the ratio between the number of observed edges and possible edges (fully connected), given by

$$H = M \Big/ \left(\frac{g(g-1)}{2} \right) \tag{4}$$

						
Connected Points	4	4	4	3	2	0
Inclusiveness	1.0	1.0	1.0	0.7	0.5	0
Sum of degrees	12	8	6	4	2	0
No. of lines	6	4	3	2	1	0
Density	1.0	0.7	0.5	0.3	0.1	0

Positioning of nodes in the network: The concept of centrality

- Capturing the embeddedness of single nodes in a social system is of essential importance
  - to describe specific roles and functions of single nodes, and
  - to understand important structural characteristics of the network as a whole

- Centrality measures are used to characterise the positioning of nodes in terms of their prominence and/or prestige in a social system

- More prominent nodes in a network feature a higher number of interactions with a high number of nodes (Freeman 1978)

Different SNA centrality measures (I)

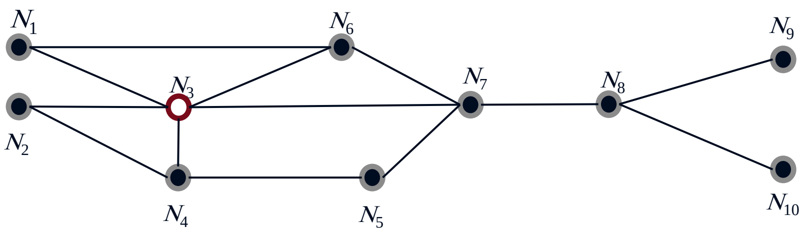
- The simplest measure is degree-based centrality of a node n_i, that is the normalised degree (k_i) of node n_i

- Closeness (or global) centrality of node n_i captures also indirect links as the total distance from n_i to all other nodes

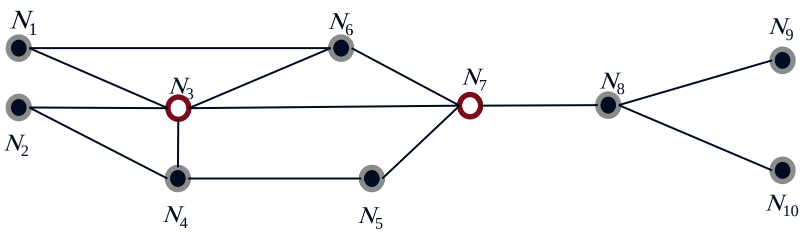
- Betweenness centrality of node n_i is the ratio of all geodesics of a graph and the geodesic going through n_i

- Prestige centrality of node n_i – based on the extraction of eigenvectors from the sociomatrix (Bonacic 1987) – denotes the degree of node n_i but accounting for the degree of the neighboring nodes

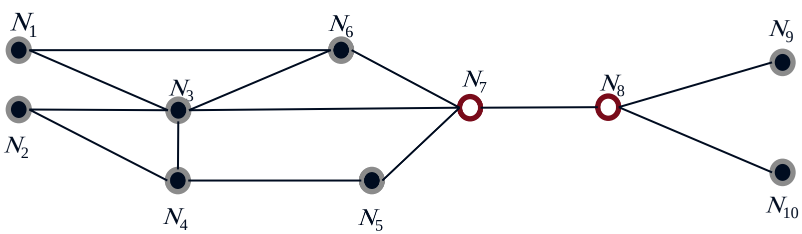
Different SNA centrality measures (II)



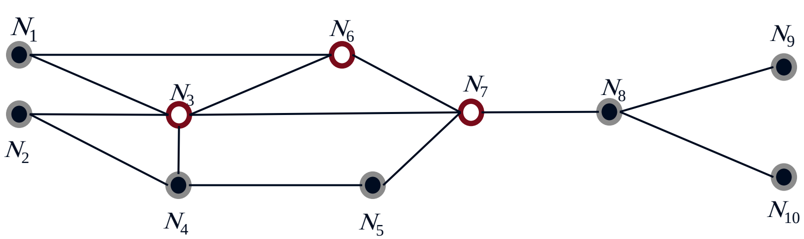
- High degree centrality



- High closeness centrality



- High betweenness centrality



- High prestige centrality

Different SNA centrality measures (III)

$$C_D(n_i) = k(n_i) = x_{i.} = \sum_j x_{ij} = \sum_j x_{ji}$$

Degree centrality

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}$$

Closeness centrality

$$C_B(n_i) = \sum_{\substack{j=1 \\ j<k}}^{g} \left[d_{jk}(n_i) / d_{jk} \right]$$

Betweenness centrality

$$C_E(n_i) = \frac{1}{\lambda} \sum_{j=1}^{g} x_{ij} C_E(n_j) \text{ mit } \quad \lambda \boldsymbol{e} = \boldsymbol{X e}$$

Prestige centrality

Centralisation of the network as a whole

- Computed unit centralities in a network can have large or small variance
  - Networks where some nodes have much higher centrality than other nodes is highly centralised (e.g. star network);
  - small differences between degree centralities indicate a low centralisation of a network (e.g., fully connected, or circle)
- Freeman (1979) centralisation index: number between 0 and 1
  - the index is 0 if all units have equal centrality value (cycle), and
  - 1, when one unit completely dominates all other units (star)

Transitivity and Clustering

- Transitivity describes a situation, that, when node A is connected to node B, and node B to node C, it follows that A is also connected to C (closure), also referred to the notion of cliques in the network
- The local clustering coefficient is defined as the ratio of the observed connections between all neighbours, and all possible connections between the neighbours
- The clustering of a graph as a whole is then given by the average of all local clustering coefficients

Clustering: formal definition

$$C_i = \frac{2b_i}{k_i(k_i - 1)}$$

Local Clustering coefficient

with b_i denoting the number of observed links between all neighbours of node $\mathcal{N}_i$

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

Global Clustering coefficient

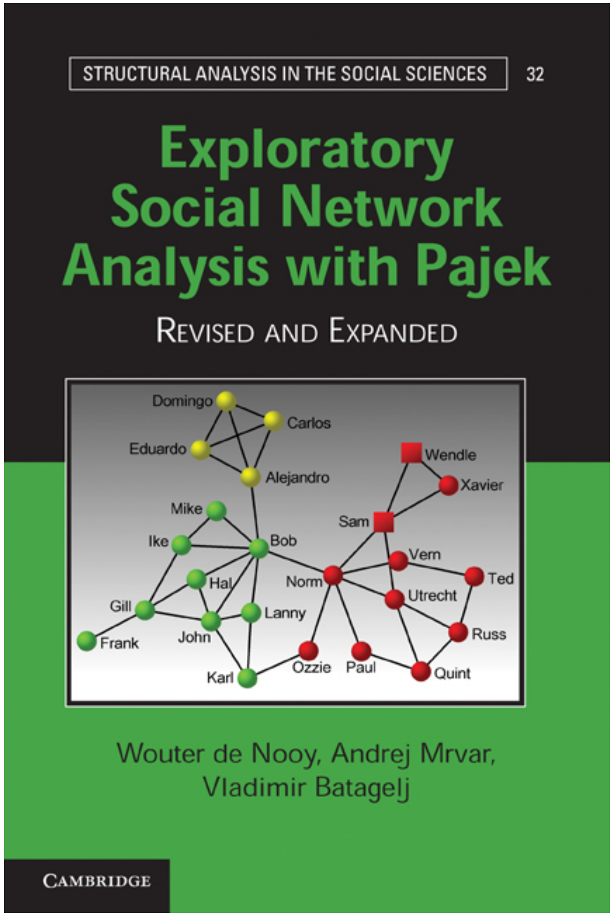
Milgram’s experiment (1967)

- 60 random selected persons from Omaha (Nebraska) and Wichita (Kansas) are motivated to send a message to a pre-defined person in Boston
- The message is not allowed to be sent directly to that person,
- but via another person, they know personally and they think the probability is higher that this person knows the target person
- “Six degrees of separation” principle

The small world effect

- Small world networks (Watts und Strogatz 1998) show a high clustering, and short global path lengths (short average path length, diameter),
- i.e., each node can reach each other node in the network via a relatively short path length
- a large number of nodes has relatively few links, while some nodes (hubs) have extremely many links

Pajek



Source: http://pajek.imfm.si/

UCINET

UCINET Software

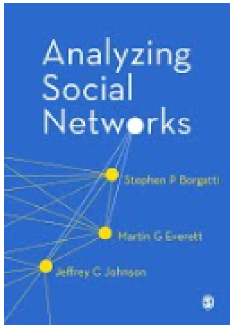
Home | Download | Versions | FAQs & Tips | Buy!

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.



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.**

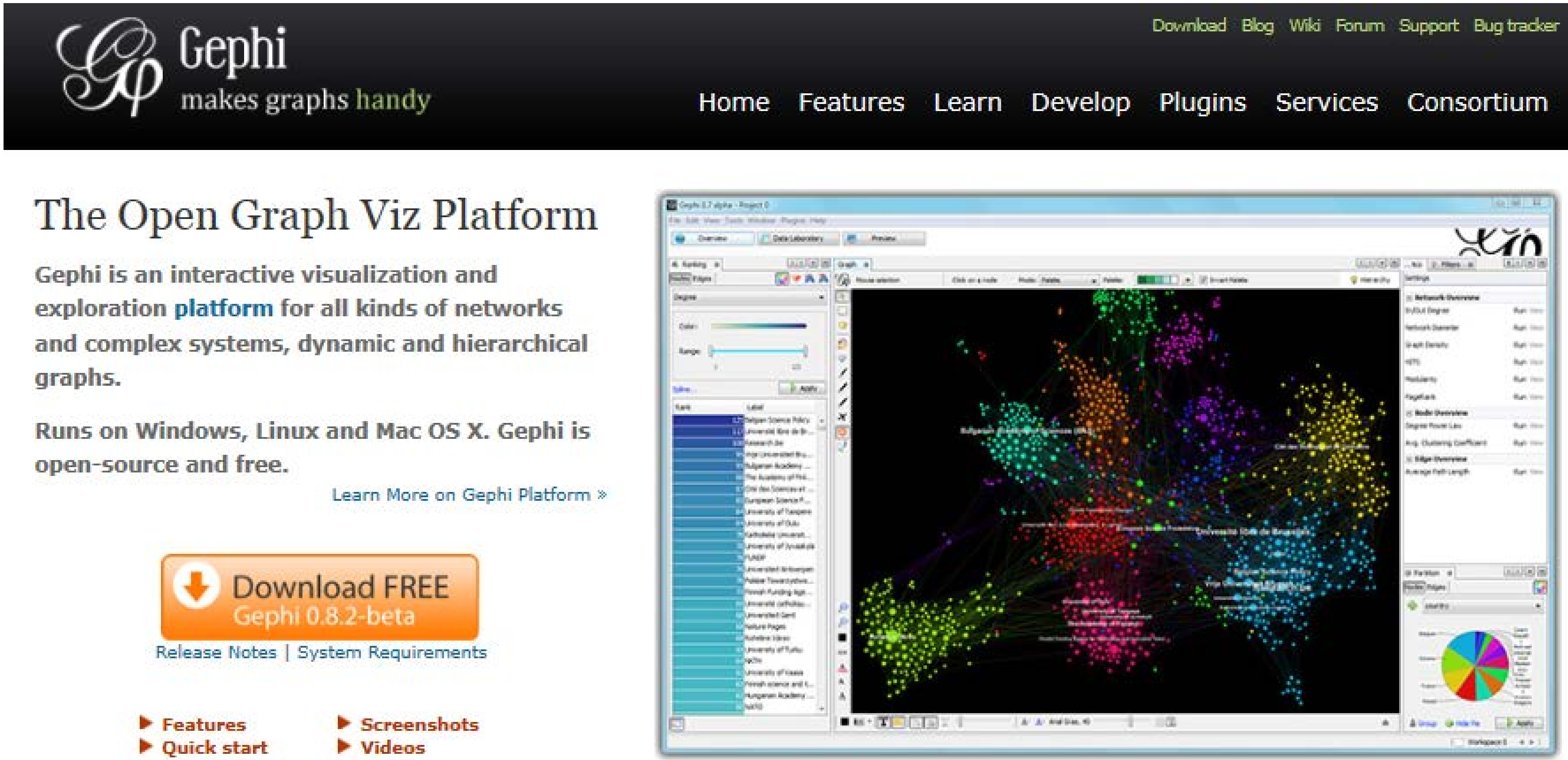
For customer support (e.g., ordering info, billing etc) contact roberta@analytictech.com. For tech support join the users group or contact support@analytictech.com. We prefer you try the users group first since the answer to your question may benefit others.

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.

Quelle: https://sites.google.com/site/ucinetsoftware/home

Gephi



Quelle: http://gephi.github.io/

R-CRAN: SNA or i-graph package

- Packages ‘SNA‘ or ‘igraph‘
- Functions available for a range of SNA tools, including
  - node and graph-level indices,
  - structural distance and covariance methods,
  - structural equivalence detection, network regression
  - random graph generation, and 2D/3D network visualization
- Advantage: high flexibility, very rapidly applicable to a great number of networks (not necessary to click and drag all kinds of figures to be calculated)

Summary

- SNA techniques as a valuable tool to describe structural properties and dynamics of social systems
- Rich toolboxes available for different software platforms
- Increasing provision of freeware for analysis and visualisation (R-cran, Gephi)
- Appropriate for explorative and descriptive analysis, not for explanatory analyses