

# **Getting Started in Social Network Analysis with NETDRAW**

**Bruce Cronin**

**University of Greenwich Business School**

**Occasional Paper 01/15**

**January 2015**

Author contact: [b.cronin@greenwich.ac.uk](mailto:b.cronin@greenwich.ac.uk)

University of Greenwich, a charity and company limited by guarantee, registered in England (reg no. 986729). Registered Office: Old Royal Naval College, Park Row, Greenwich SE10 9LS.

## **Getting Started in Social Network Analysis with NETDRAW**

Bruce Cronin

NETDRAW is one of the most accessible and undoubtedly the most widely used software for social network visualisation. Bundled with UCINET, the most widely used social network analysis software (Borgatti, Everett and Freeman 2002), its popularity arises from its embrace of the Microsoft Windows environment and the regular inclusion of new analytical techniques in the field as they are developed. While UCINET provides extensive tools for comprehensive network analysis, NETDRAW has powerful analytical capacity in its own right and is freely available. For exploratory network analysis, NETDRAW is a perfectly appropriate tool on its own.

Despite its popularity, however, the software is also very frustrating at times for new users. It has very limited documentation of its features, assuming a grasp of mathematically challenging concepts as its frame of reference, has some idiosyncrasies that appear insurmountable without experience and its outputs are sometimes difficult to interpret.

This paper provides a brief step-by-step guide to allow a new user to undertake a simple social network analysis yet yield fairly sophisticated results. It commences with a brief introduction to social networks, moves to issues in data collection, preparation for analysis and data import. Simple visualisation is followed by extraction of the main component and the calculation of node centrality metrics.

Instructions on command sequence given in this guide assume the default options are unchanged. Defaults normally only need to be changed when there is a clear reason to do so.

The placement and availability of commands varies slightly from one version of NETDRAW to the next and it is frequently updated. The descriptions in this guide are drawn from Version 2.139.

## 1. Introduction to Social Networks

There has been a long-standing distinction in organisational studies between the formal, mandated, organisation and the informal, spontaneous, organisation necessary to get things done. Figures 1 and 2 contrast the formal organisational structure of a firm with the informal structure of relationships seen by staff of the same firm to be important to their work.

Figure 1. Formal Organisation

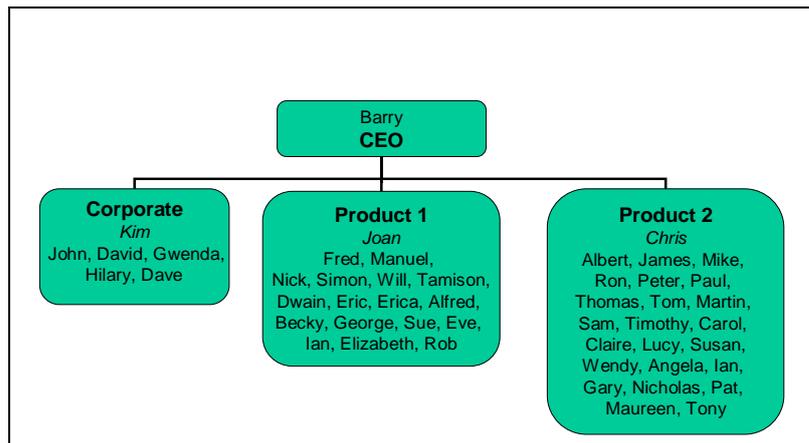
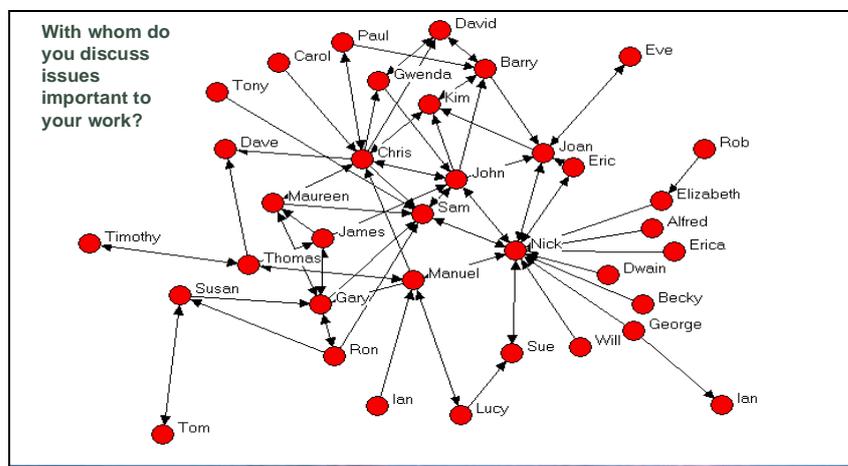


Figure 2. Informal Organisation



The representation of the informal relationships in the organisation is the result of a social network analysis of data collected from individuals. Social Network Analysis (SNA) is a set of techniques for identifying and representing patterns of interaction among social entities, be it individuals, groups, organisations or social artefacts. It provides precise and specific insight in place of intuition and general hunches.

Social network analysis is predominantly employs graphical techniques, an application of the mathematics of graph theory. The approach employs four principal tools, the first two illustrated in Figure 2:

- social entities are represented as points, each known as a 'node' or 'vertex';
- relationships are represented by lines, known as 'ties', 'edges' or 'arcs'.

It is also possible to represent:

- the strength of the relationship, for example, by line width;
- attributes of the nodes, for example, by different colours or size.

Relationships are distinguished in social network analysis principally by their directionality and value. Directional relationships, formally represented as ‘arcs’, involve a transfer from one node to another; providing information or advice, for example. Non-directional relationships, formally represented as ‘edges’, comprise a sharing; members of the same organisation, for example. Relationships may be valued in terms of frequency of contact or subjective evaluation, for example.

## *2. Data Collection*

The most common method for data collection is by questionnaire. Members of a group of interest are listed on a roster and then each member is asked to assess their relationship with each of the other members on some dimension. Such questionnaires may be administered by personal interview or by self-completed questionnaire, often postal or, increasingly, web-based. Unlike traditional probability based approaches to social research, however, network analysis is highly sensitive to the response rate, so intrusive methods of data collection are favoured. There is a corresponding overhead in terms of negotiating access to sufficient numbers of respondents and in terms of ethical considerations.

An alternative to questionnaire-based approaches is observation. These include ethnography, where a researcher joins a group and observes interactions; expert panel studies, where highly sensitive participants such as business managers are asked to systematically reflect on what are normally hunches in this regard; and experiment, where some information is introduced to one part of a group and the group is watched to see where the information subsequently emerges. A recent development in observational methodology has been the emergence of ‘smart’ email scanning techniques to identify areas of expertise within an organisation and internal or external contacts members may have.

A high response rate is important because the research centres on particular relationships among a series of social entities. Whereas in traditional quantitative approaches, individual respondents are often representative of broad trends and so relatively unimportant in themselves, in network research single relationships may have particular importance that would greatly distort findings if missed. There are, however, emerging statistical techniques to test for missing data, involving the comparison of network data against what would be expected to be found randomly.

Because social network analysis is a reasonably novel research technique (despite its 70-year pedigree), and because data collection is often necessarily intrusive, negotiating access to respondents is typically more involved than in traditional social science research. In particular, comprehensive buy-in is needed from the principal decision-makers. There have been numerous cases where data collection has been disrupted by the late intervention by a senior decision-maker who was not sufficiently briefed during the establishment of a project.

Social network analysis also raises a particular range of ethical challenges, as Borgatti & Molina (2003) discuss. Firstly, unlike much of traditional social science research, it is not possible to respondent anonymity because specific personal relationships are the object of study; we are interested in the relationship between ‘Tom’ and ‘Joan’, rather than the relationship between any ‘X’ and ‘Y’. Secondly, respondents may cite relationships with third parties who may not wish to participate in the study. While respondents are reporting their own perception, they are also reporting shared information – the fact that a relationship exists. Thirdly, it is difficult to offer confidentiality as participants may deduce the identity of individuals from a network position, even when presented anonymously in a sociogram. Lastly, it is obtaining informed consent is difficult because the technique is new and participants probably will not fully anticipate the implications of the information they provide.

Having addressed these challenges, the resulting data comprises a list of nodes, which of these are related and normally some attributes about the nodes and relations. Small amounts of data can be stored in a matrix or spreadsheet table, as in Figure 3. But larger amounts are more usefully stored as a series of related node pairs, or ‘edge list’, as in Figure 4. Both forms can be conceived as a matrix of node-to-node relationships.

In the first row of Figure 3 and the first two lines of Figure 4, Barry is reporting a relationship with Kim and with David. Note that John is reporting a relationship with Kim, while Kim is not reporting a relationship with John, indicating a directional relationship.

Figure 3. Adjacency Matrix

		<i>With whom</i>				
		Barry	Kim	John	David	Gwenda
<i>Who</i>	Barry	0	1	0	1	0
	Kim	1	0	0	0	0
	John	1	1	0	0	0
	David	1	0	0	0	1
	Gwenda	0	0	1	1	0

Figure 4. Edge List

```
Barry Kim
Barry David
Kim Barry
John Barry
John Kim
David Barry
David Gwenda
Gwenda John
Gwenda David
```

### 3. Importing Data

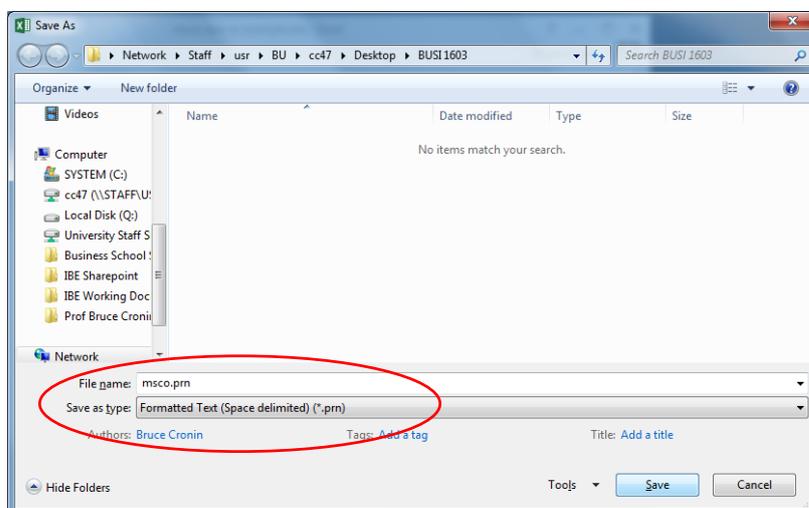
The most convenient way to store network data is as an edgelist, or list of node pairs, in a spreadsheet. A spreadsheet can hold a great deal of data; the 32-bit version of Excel 2013 can hold 1048576 rows in a single worksheet; the 64-bit version is limited only by RAM capacity. For larger datasets, Access 2013 can hold an edgelist in a table of 2GB, or 10GB in association with the SQL Server or Sharepoint extension.

An edgelist can be imported to NETDRAW as a .dl formatted file. This involves the insertion of header information (as in the first 5 rows of Figure 7) then saving the file as a plain (space delimited) text file (as in Figure 8). This creates a file with a .prn extension; it is clearer to rename this in Internet Explorer with a .dl extension (by default, Windows Explorer hides known extensions; this can be changed via the Options menu in Windows Explorer).

Figure 7. Preparing Edge List in Spreadsheet for Export in DL File Format

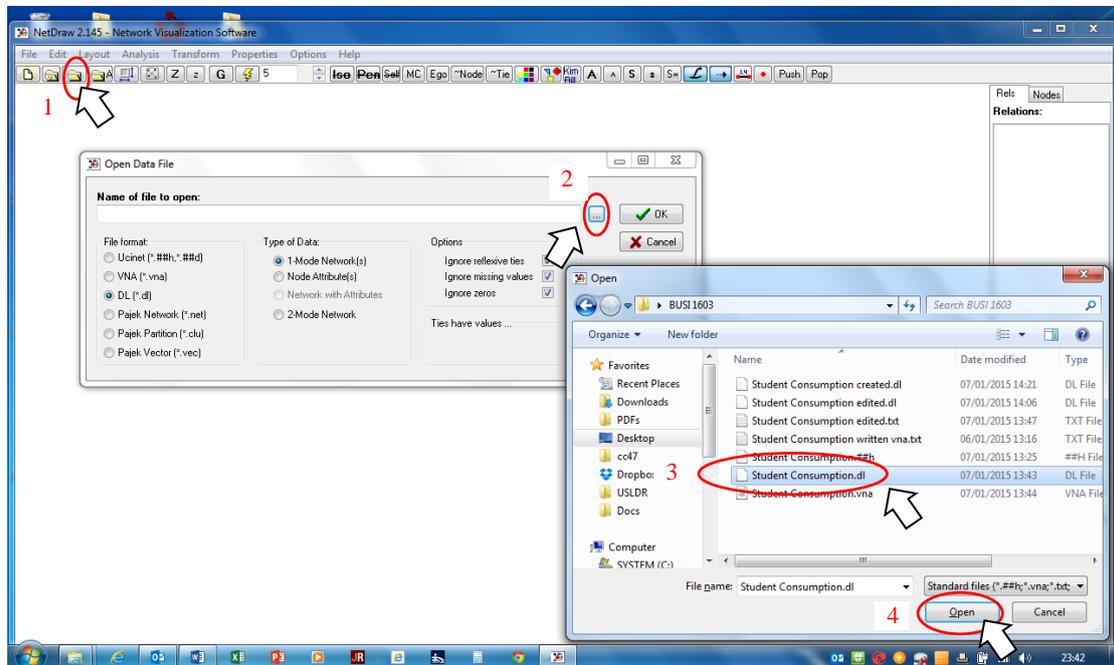
	A	B	C	D	E	F	G	H	I	J	K	L
1	dl											
2	n = 5											
3	format = edgelist1											
4	labels embedded											
5	data:											
6	Barry	Kim										
7	Barry	David										
8	Kim	Barry										
9	John	Barry										
10	John	Kim										
11	David	Barry										
12	David	Gwenda										
13	Gwenda	John										
14	Gwenda	David										
15												
16												

Figure 8. Save as Formatted Text (Space delimited) (\*.prn)



To open a .dl file in NETDRAW, click on the third icon, a plain open folder, as illustrated in Figure 9. Then open a Windows browser using the  icon to the right of the dialogue box. Select the .dl file to open and click 'OK'.

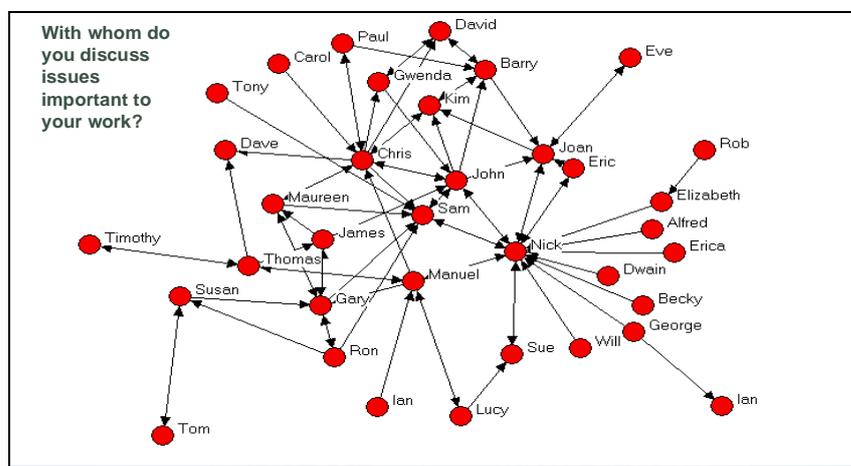
Figure 9. Opening a .DL File in Netdraw



#### 4. Network Visualisation

On opening the file, NETDRAW automatically generates a visualisation of the data using default options. This uses a standard algorithm to push the most connected nodes to the centre of the screen and the least connected nodes to the periphery, as in Figure 10.

Figure 10. Network visualisation



The visualisation can be saved for further editing as a .vna file, the native format of NETDRAW, using the command:

**File\Save Data as**

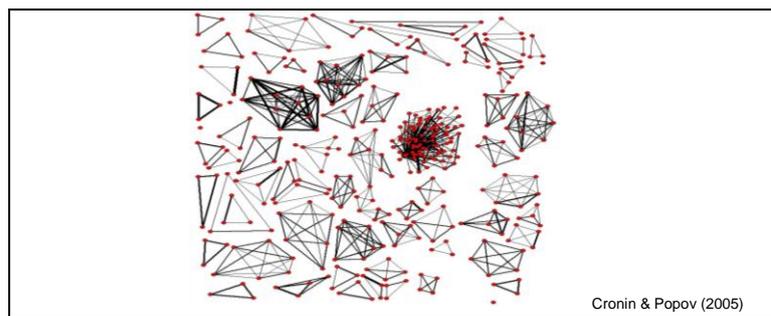
The image itself can be saved as an image file for import into Word or other applications. Available formats are jpeg, bitmap or Windows metafile, the latter more versatile for rescaling. The image can be created using the command:

**File\Save Diagram as**

### 5. *Extracting the Main Component*

Extensive datasets typically involve many small independent clusters around larger ones and one particularly large and dense cluster, as in Figure 11. Each of these separate, internally connected clusters is known as a ‘component’. The largest cluster is the ‘main’ or ‘giant’ component.

Figure 11. Multiple Components



To identify the distinct components, use the command:

**Analysis/Components**

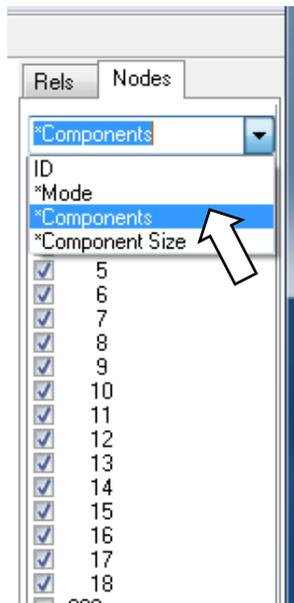
To select a single component, in the command panel on the right of the screen, select the Nodes tab and then ‘Components’ from the drop-down list, as illustrated in Figure 12. Individual or multiple components can be selected by checking the checkboxes.

To select the main component, select ‘Component Size’ from the drop-down list and click on the largest number. Checking the -999 box will select isolated nodes.

Once a component is selected, this subset of nodes can be saved as a separate network file, using the command.

**File/Save As**

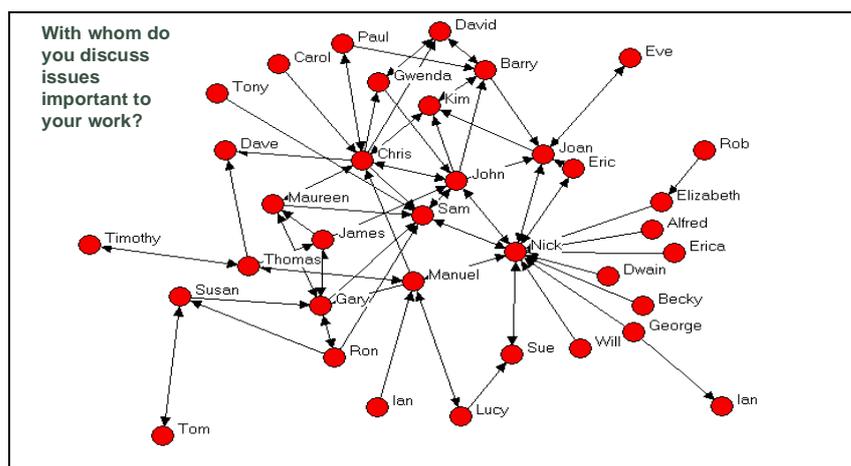
Figure 12. Selecting Components



## 6. Centrality of Nodes

A network visualisation invites the question of which are the most important nodes in the network. Standard algorithms locate the most central nodes in a network in the centre of the visualisation. But the mathematics of graph theory provides the means to answer such questions more precisely.

Figure 13. Network Visualisation Example



Attributing 'importance' to the position of a node in a social network implicitly employs a theory of social interaction, normally a variant of the theory of social capital. Coleman (1990) and Putnam (1993, 2001), for example, argue that the social resources available to an individual, such as information and status, are strongly

determined by the extent to which that individual is at the centre of a cohesive network of relationships. In Figure 13, nodes such as Nick, John, Sam and Chris would appear to be central in terms of cohesiveness. Social network analysis provides three principal ways of measuring the extent to which a node is at the centre of a cohesive network: degree centrality, eigenvector centrality and closeness.

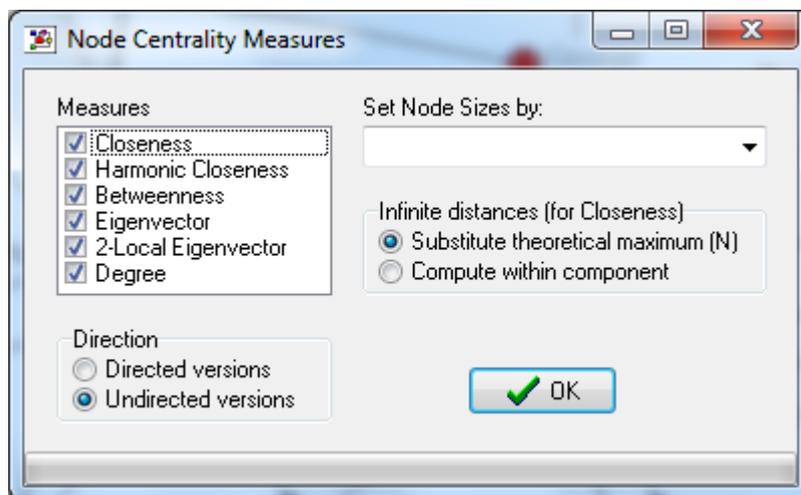
Another view of social capital, however, challenges the notion that individuals at the centre of cohesive groups have the greatest access to valuable resources. Burt (1993) argues that the most important position in a network is one that bridges otherwise less connected parts of the network, a position of brokerage. In Figure 13, Nick, John and Sam are still important in these terms but Manuel also enters the picture. Again, social network analysis provides a means of measuring the position of nodes in these terms: betweenness centrality.

NETDRAW provides a simple command to calculate these metrics for each node. Note. These measures are only meaningful when carried out on a single component (typically the main component); it makes little sense to talk about the centrality of a node if the nodes are drawn from unconnected components.

To calculate the centrality of each node on a range of measures, use the command: **Analysis/Centrality measures**

A dialogue box (See Figure 14) invites you to select which centrality measures to calculate but in reality all are calculated. Leave the defaults unchanged unless you have a particular reason to do so.

Figure 14. Analyse / Centrality Measures



The calculated metrics are stored against each node as 'Attributes'. To view the metrics for a particular node, right click on the node and select attributes (See Figure 15). The attributes of the node, including the metrics, are displayed as in Figure 16.

Figure 15. Menu Displayed from Right-click on a Node

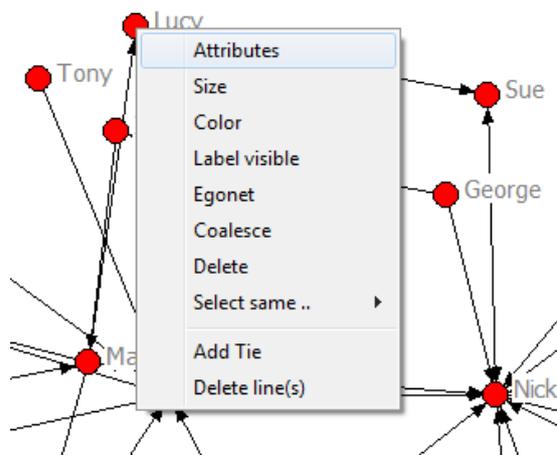


Figure 16. Node Attributes

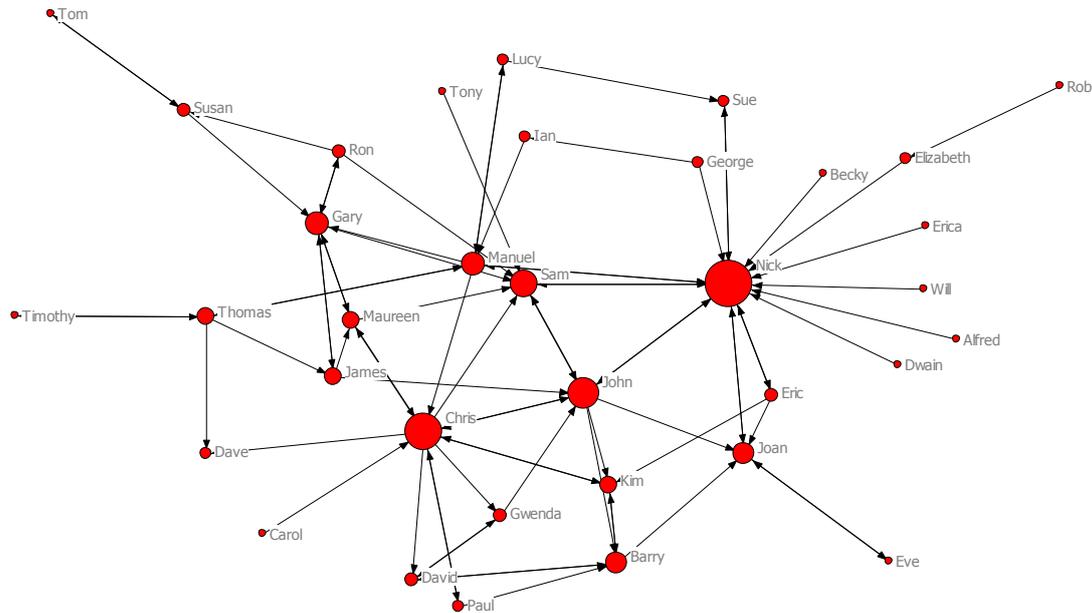
Key	Value
ID	Lucy
Degree	2
Betweenness	3.810
Closeness	98.000
HarmonicClos	13.083
Eigenvector	0.057
2-Local Eigen	8.000

### 6.1 Degree Centrality

The most intuitive measure of centrality is the number of direct connections a node has to other adjacent nodes. The number of connections is known as the ‘degree’ of a node. In Figure 16, the node Lucy has a degree centrality of 2, that is, Lucy is connected to 2 other nodes (Sue and Manuel). The most central node in a network in these terms is the node with the highest degree, that is, the greatest number of connections. This is an indicator of ‘popularity’. In Figure 17, the size of each node varies by its degree centrality. This can be visualised using the command:

**Properties/Nodes/Symbols/Size/Attribute-based** (Set the Select Attribute field to ‘Degree’)

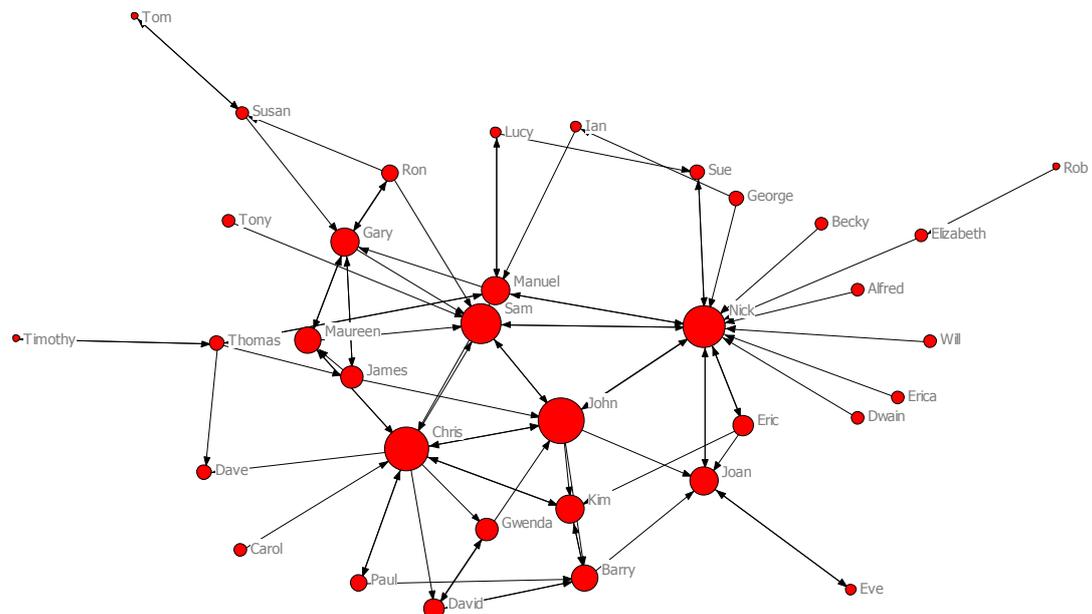
Figure 17. Node Size by Degree Centrality



### 6.2 Eigenvector Centrality

Degree centrality is a limited interpretation of social importance. Consider John and Sam in Figure 17. They have potential social influence or power because they are closely related to nodes with high degree centrality (Nick and Chris). A method of identifying such nodes is Eigenvector Centrality (Bonacich 1972), which weights degree centrality by the degree centrality of the nodes a node is connected to (see Figure 18).

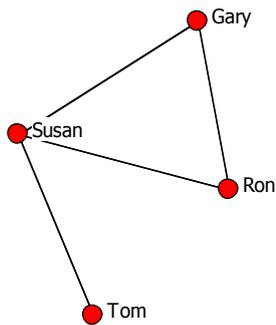
Figure 18. Node Size by Eigenvector Centrality



### 6.3 Closeness Centrality

An alternative measure of centrality in a cohesive group to degree or eigenvector centrality is closeness, which does take into account the position of all other nodes in the network. Closeness can be best conceived in terms of its reciprocal, farness, which is the sum of the shortest path distance, that is, the number of steps, from one node to each other. Figure 19, extracted from Figure 13, presents the farness for each node. Tom has the greatest farness. From the reciprocal, Susan has the greatest closeness.

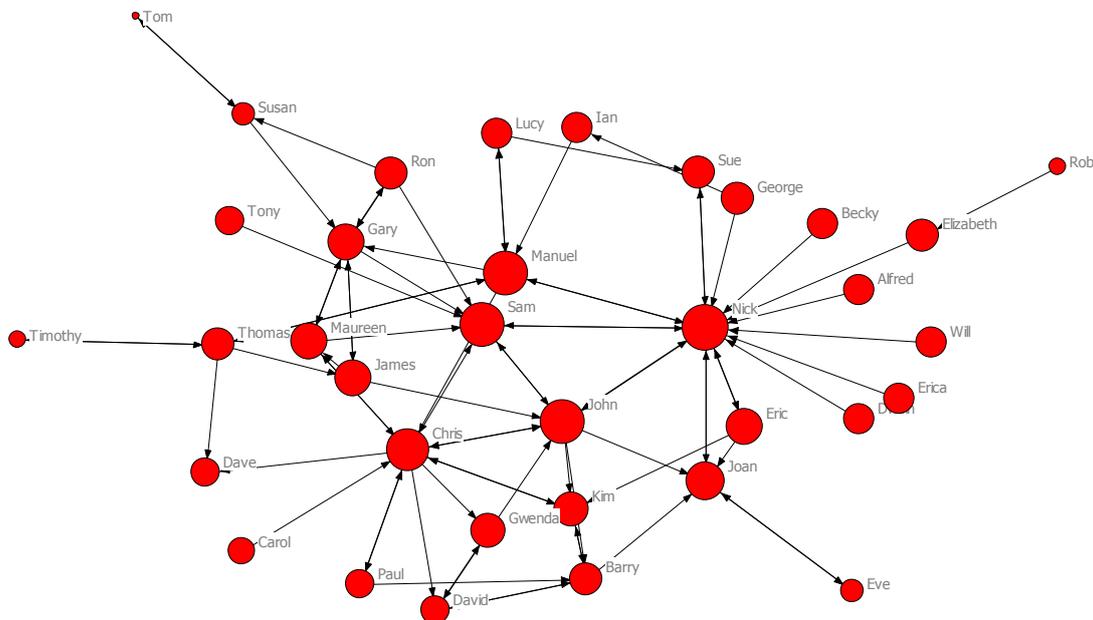
Figure 19. Farness / Closeness



Tom	$1 + 2 + 2 = 5$
Susan	$1 + 1 + 1 = 3$
Gary	$1 + 1 + 2 = 4$
Ron	$1 + 1 + 2 = 4$

The closeness centrality of each node is also captured by the Analysis/Centrality measures command in NETDRAW. **Note** that a programme error shows Farness rather than centrality. To show Closeness, check the 'Reverse values' option in the dialogue box when selecting the attribute to set node size by.

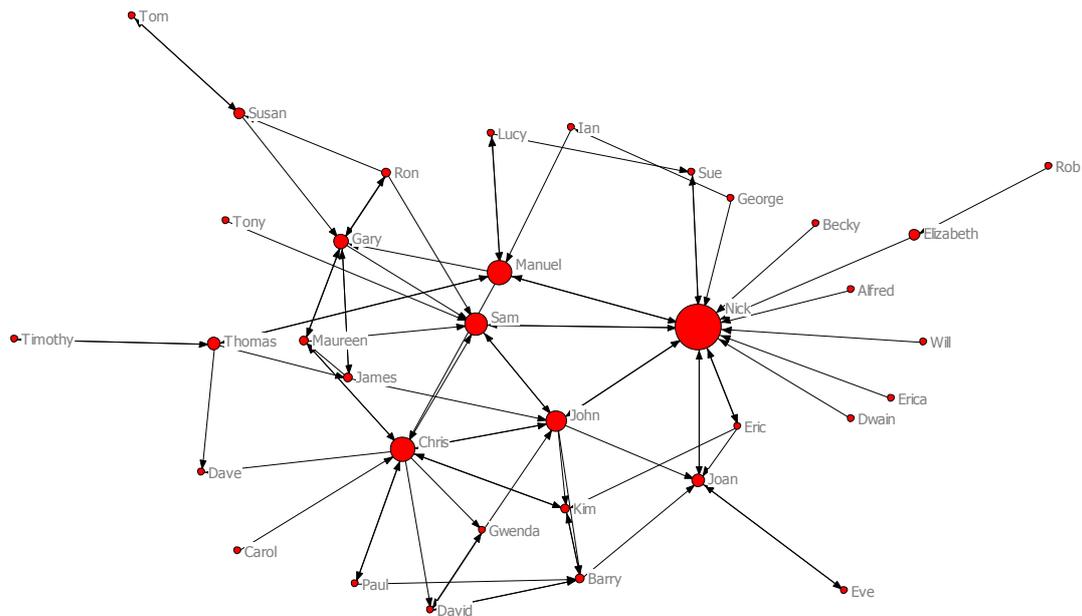
Figure 20. Node Size by Closeness Centrality



## 6.4 Betweenness Centrality

In contrast to the measures of centrality within a cohesive group discussed above, betweenness centrality identifies nodes that bridge less connected parts of the network. The betweenness of a node is the proportion of times a node appears on the shortest paths between each other pair of nodes. In Figure 21, Nick has the highest betweenness.

Figure 21. Node Size by Betweenness Centrality



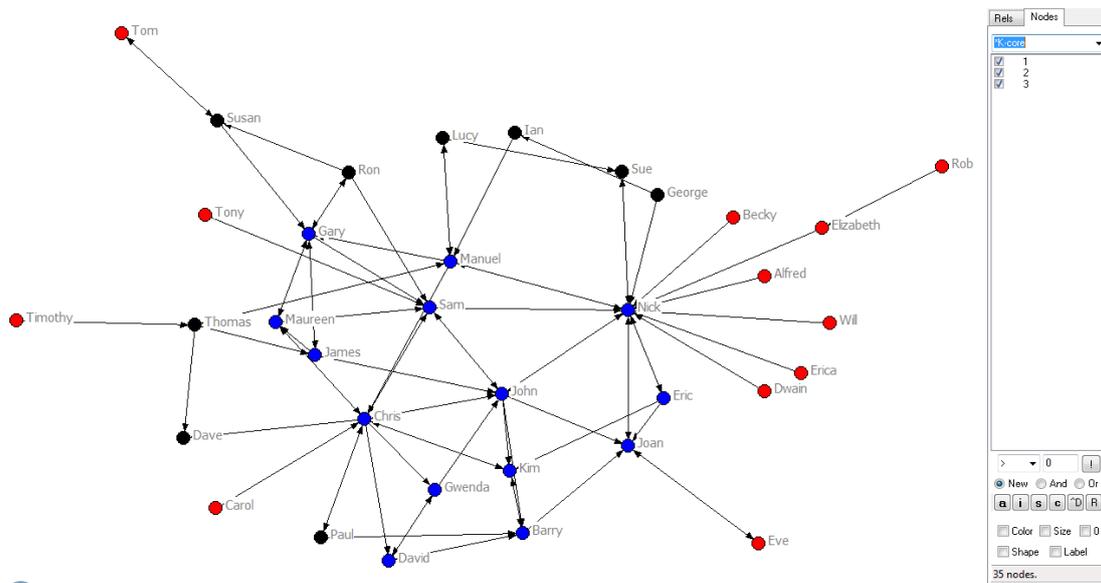
## 7. Network Cohesiveness

While individual nodes in a network may have positions of particular influence, the network structure as a whole also provides particular opportunities and limitations for all nodes. Some parts of the network are more connected than others and these regions may be more influential than others or in other respects perhaps more constrained than others.

### 7.1 *k*-core Analysis

*k*-core analysis identifies parts of a network that are more connected than others. *k* refers to the number of immediate connections a node has. A *k*-core of 1 refers to all nodes that have a degree of 1 or more, ie. all nodes in the network. A *k*-core of 2 refers to the subset of all nodes that have a degree of 2 or more, etc. Figure 22 presents three *k*-cores. Checking and unchecking the checkboxes in the control panel on the right allows the identification of different concentrations of connectivity in the network.

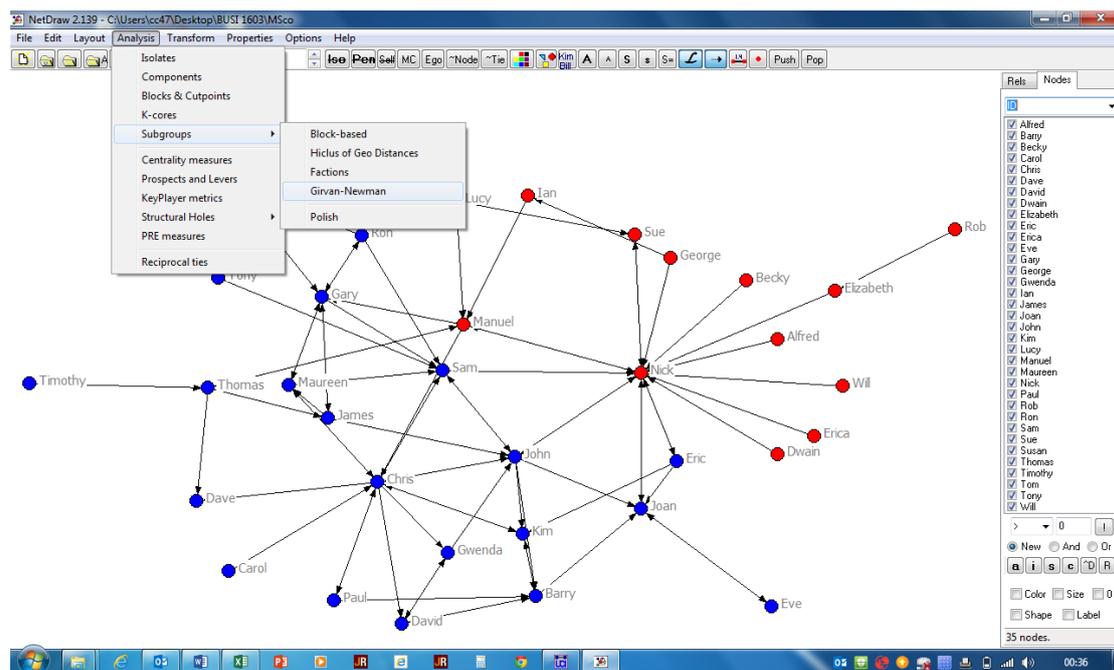
Figure 22. *k*-core analysis



## 7.2 Sub-group Analysis

Network structure can also be analysed in terms of nodes that are more closely related to each other than other nodes, ie. as clusters or communities. A popular algorithm for the demarcation of community structures is the Girvan-Newman (2002) algorithm. Figure 23 illustrates the application of this algorithm to distinguish 2 distinctive groups. Accept the defaults unless there is a specific reason not to.

Figure 23. Sub-Group Detection using the Girvan-Newman Algorithm



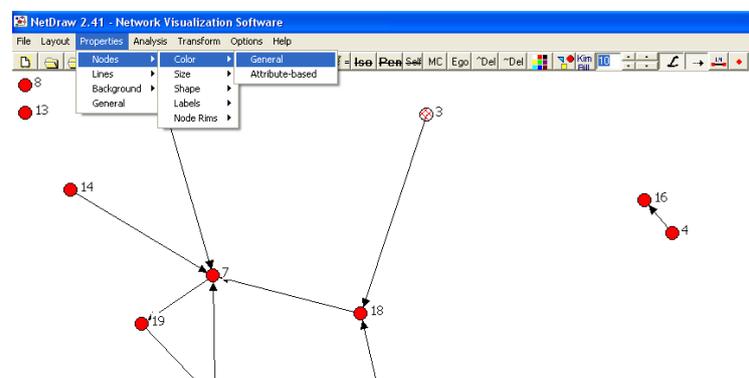
## 8. Enhancing Visualisations

Network visualisation aims to provide a meaningful visual representation of a network dataset. How this is represented depends on the particular algorithm used and the output media. The visualisation will differ with different screen sizes and resolutions, for example.

NETDRAW has a wide range of editing tools to allow the user to configure the visualisation to better convey the meaning of the data. Any node can be moved to any position by clicking on it and dragging it to another location. The size, colour, shape and labelling of any individual node or groups of nodes can be adjusted. Nodes and links can even be added or deleted. To adjust the properties of nodes, click on each node of interest and use the command:

**Properties\Nodes**

Figure 24. Changing the Colour of a Selected Node



In Figure 24, the selected node 3 is highlighted. A new colour for the node can be applied with the command:

**Properties\Nodes\Colour\General**

To deselect a node or group of nodes, click anywhere on the main drawing area. Where no single or group of nodes is selected, the **Properties** command will make changes to all nodes. Where attributes have been associated with nodes, changes can be made to nodes according to their attribute.

Of course, it is the ethical responsibility of researchers to ensure that this editing is only undertaken for the purpose of improving the communication of the meaning of the data. The best way to ensure this is to begin with standard algorithms to generate the network visualisation and then to minimise editing strictly to the goal of clarifying meaning. For example, it may help to separate two nodes that are drawn in the same location so that both are visible, or to move a node slightly so that a label becomes visible.

## 8.1 NETDRAW Filters

NETDRAW has some tools to allow the selection or isolation of nodes on the basis of general properties as below.



- Iso** deletes isolated nodes from the visualisation
- Pen** deletes pendants (nodes with a degree of 1) from the visualisation
- Self** deletes self-loops from the visualisation
- MC** displays only the main component
- Ego** displays the ego net control box.
- ^Del** deletes non-active nodes and lines
- ~Del** restores all deleted nodes

## 8.2 NETDRAW Graphic Controls



The first icon opens the attribute node colour control, allowing different colours to be associated with different node attributes.

The second icon opens the attribute node shape control, allowing different shapes to be associated with different node attributes.

The third icon opens the Label editor, allowing the label of each node to be edited.

The numeric box reports the label size in points, which can be adjusted with the adjacent up and down arrows. The adjacent up and down arrows change the size of the nodes,

- L** toggles labels off or on
- toggles arrowheads on or off
- 1.4** toggles line width on or off

The last icon allows the insertion of an additional node by clicking on the visualisation.

### 8.3 The NETDRAW Control Panel



#### Relations Tab

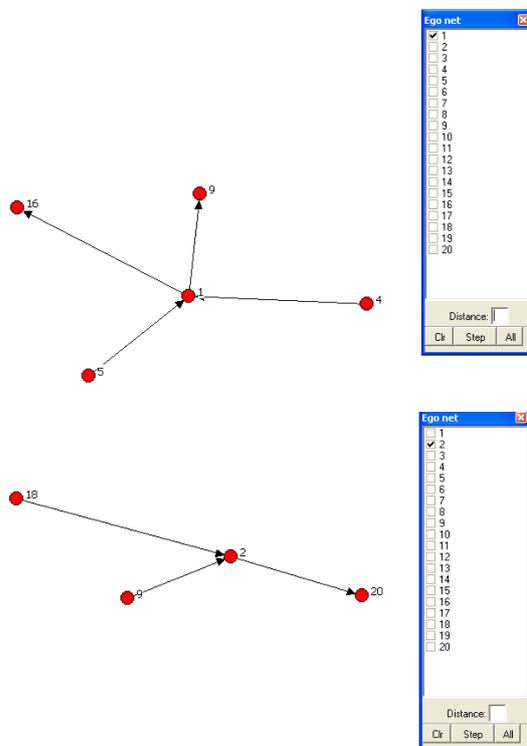
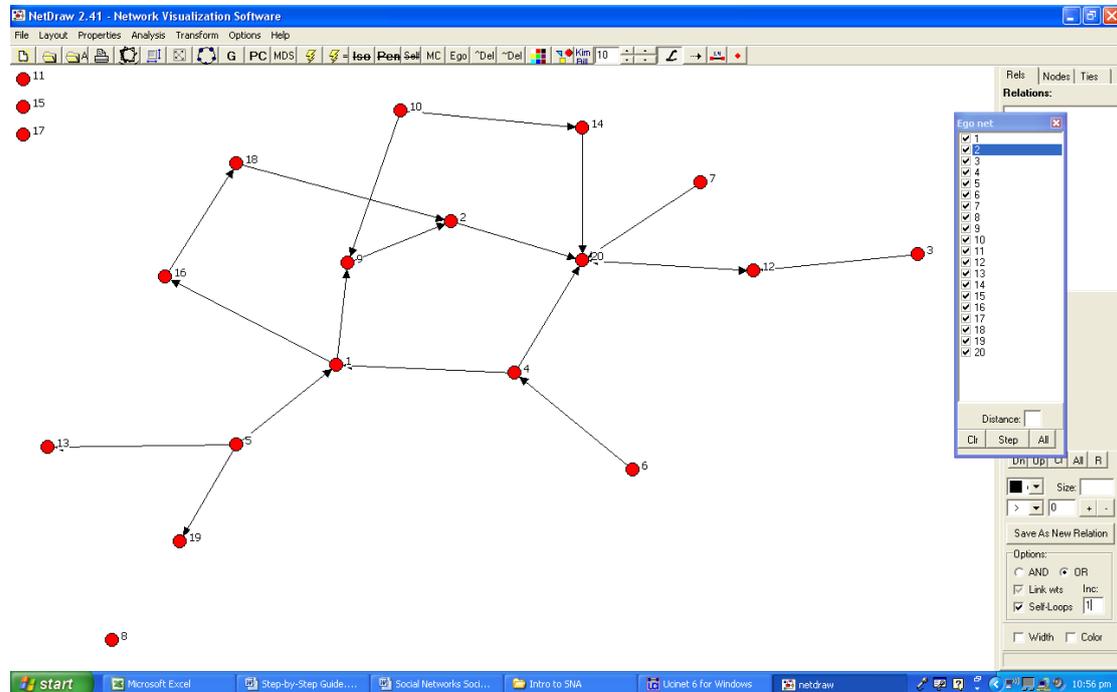
Where the relations have been transformed within NETDRAW, each visualisation is listed in the Relations box on the Rels tab.

Clicking on the Dn or Up button displays each visualisation in turn. It is thus possible to cycle through a series of related networks for comparative purposes. The All button selects all visualisations and Cl deselects these.

The colour and width of all displayed ties between nodes can be adjusted with the colour drop-down box beneath the Dn button and the Size input box to the right of this.

## 8.4 The NETDRAW Egonet Control

The Egonet control is activated by the Ego icon. This displays a control window listing all nodes in the network. Clicking on the step button in this control window displays each node in turn, together with all nodes adjacent to this. The distance input box allows the setting of the path distance (number of steps) to include in the ego net.



### 8.5. Adding Attributes to a Visualisation in NETDRAW

A common application of a network visualisation is to identify different nodes within a network on the basis of some attributes. For example, male and female employees or different work-groups may be distinguished.

Attribute information is imported to an existing visualisation, that is, a network must already be visualised before the attribute information is added. With the network visualisation open, the attribute information is added simply by opening a file with this information in it.

The attribute file has as many rows as the network has nodes and at least two columns, the second being the attribute associated with the node. Multiple attributes can be included as multiple columns. Note. Attribute information must be numeric.

For small datasets, attribute data can be pasted into the Node Attribute Editor via the command:

#### **Transform/Node attribute editor**

For larger datasets, attributes can be imported via a dl file. This needs to be in the edgelist2 format. Any number of attributes can be included for each node but each attribute needs to be listed on a separate line.

```
dl
nr = 3, nc = 2
format = edgelist2
labels embedded
data:
FirmAmarket      1
FirmAregion      0
FirmB      market      1
FirmBregion      0
FirmCmarket      1
FirmCregion      2
```

In this example, there are three nodes ( $nr = 3$ ) and two attributes ( $nc=2$ ). The labels for the nodes and attributes are drawn from the data. Each row lists the node, then the attribute, then the value of the attribute.

## References

- Bonacich P. 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* 2, 113-120.
- Borgatti, S. P. 2002. *NetDraw: Graph visualization software*. Harvard: Analytic Technologies.
- Borgatti, S. P., Everett, M.G. and Freeman, L.C. 2002. *Ucinet for Windows: software for social network analysis*. Harvard, MA: Analytic Technologies.
- Borgatti, S.P. and Molina, J-L. 2003. Ethical and strategic issues in organizational network analysis. *Journal of Applied Behavioral Science* 39(3): 337-350.
- Burt, R.S. 1992. *Structural holes: The social structure of competition*. Cambridge: Harvard University Press.
- Coleman, J. S. 1990. *Foundations of social theory*. Cambridge, MA, Belknap Press.
- Cronin, B. and V. Popov, V. 2005. Director networks and UK corporate performance. *International Journal of Knowledge, Culture and Change Management* 4, 1195-1205.
- Girvan M. and Newman M. E. J., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 7821–7826.
- Gould, J. and Fernandez, J. 1989. Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology* :89-126.
- Krackhardt, D. 1999. The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations* 16, 183-210.
- Putnam, R. D. 1993. *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ, Princeton University Press.
- Putnam, R. D. 2001. *Bowling alone: the collapse and revival of American community*. London, Simon & Schuster.
- Seidman, S. 1983. Network structures and minimum degree. *Social Networks* 5(2), 269-287.