Scott, John. "Key concepts and measures." *What is Social Network Analysis?*. London: Bloomsbury Academic, 2012. 31–56. The 'What is?' Research Methods Series. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781849668187.ch-003>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 13:12 UTC.

Copyright © John Scott 2012. You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

3 Key concepts and measures

Chapter 2 traced the history of social network analysis through the three broad mathematical approaches that have dominated the field: graph theory, algebraic approaches and spatial approaches. In this chapter I will first consider the principal methods of data collection for social network analysis, and I will then outline and define, in sequence, the key concepts and measures associated with each mathematical approach considered in Chapter 2. I will suggest that graph theory provides the formal framework common to all these approaches. I will not present highly technical definitions, as these are more appropriate to the various handbooks of social network analysis (Scott 2012; Degenne and Forsé 1994; Prell 2012). Having done this, I will set out some of the statistical procedures used in assessing the validity of network measures in actual situations.

Collecting network data

Relational data for social network analysis can be collected through a variety of methods. These include asking questions about the choice of friends, observing patterns of interaction, and compiling information on organisational memberships from printed directories. All such methods require that an appropriate selection of cases be made from the total range of possible data sources. In many studies this will involve a complete, or near-complete, enumeration of a whole population. For example, an investigation into the interlocking directorships within the 'top 200' companies in an economy will typically involve the collection of data on all 200 companies. Similarly, an investigation into the friendship patterns of the children in a school class will wish to collect information on all the children in the chosen class.

This emphasis on complete enumeration reflects, in part, the small size of the groups involved and the relative ease with which complete data can be collected. More importantly, however, it reflects the difficulty of using sampling techniques with relational data. Many concepts and measures in social network analysis, as will be seen, require the use of complete data on all network linkages, and only a relatively few measures relating to individuals can be assessed from sampled data. In general it can be said that individual-level or ego-centric measures can be generated from sample data, but that macro-level, structural measures can be produced only with considerable difficulty and uncertainty, or perhaps not at all. While this difficulty is not absolute, it is a major limitation and much research is currently

being undertaken to find ways of making more reliable inferences from sample data on social networks.

Once relational data have been acquired, they can be prepared for use in social network analysis. In the case of very small networks, this will generally involve purely manual methods as the analyst can easily keep track of the various individuals and organisations involved. Even where the number of cases reaches 20 or 30, manual analysis of the data may still be possible. In the case of large networks, however, analysis by computer will generally be necessary and the data will need to be stored within the formatted files of the particular software that is to be used.

Data must be sorted on a case-by-case basis, as each case will be represented by a 'point' in the sociograms and the network analyses. Typically, each point and its connections are listed as a numbered row in a data file. For example, each individual may be listed with the organisations or meetings in which he or she is involved. Each such point can be given an identifying reference label, perhaps the name of the individual or an abbreviated form of this, and the label can be stored as part of the row or in a linked file. The various organisations or meetings in which the individuals are involved will also need to be given a label so that this can be married up with the individual data. In some analyses, an investigator will be concerned with the connections among the cases themselves: for example, with the connections among individuals who are involved with the same organisations. In these situations, the rows in the file will need to list the other individuals to which each individual is connected, and this kind of file can easily be generated from the original list by the standard software programs.

Such a data file might show, for example, eighteen women and their participation in a number of social events. Each woman would be represented by a row in the file, and the row would show the particular events that she has attended. From this initial file it is possible to produce a file listing the women and the others who they meet at the various events, and also a file listing the various events that are similar because the same women attend them. Similarly, a file for the analysis of interlocking directorships would show each company as a row, with the rows listing the directors of the company. From this can be produced listings of individuals and their meetings with each other and of companies and their board-level connections to each other. A file showing the friendship choices of schoolchildren would typically record the connections among the children directly, each row showing the other children that are chosen as friends.

Standard social network analysis packages can read these so-called linked lists and can convert them into the various other files that are required for specialist analyses. This normally occurs behind the scenes with there being little or no need for the investigator to directly manipulate the files. It is to the particular concepts and measures available in these packages that I now turn.

Graph theory and egocentric measures

True to its origins in sociometry, uses of graph theory tend to employ the language of points and lines to describe the patterns of connection in a social structure. Its various concepts and

measures aim to convey the properties of the visual image of a sociogram. Visualisation is difficult for large and complex networks, but the visual imagery behind the basic concepts of graph theory can help us to imagine the more complex structures of which they are the building blocks. Sociograms are arbitrary aesthetic representations that maintain only the patterns of connections and not any information about physical distance and relative location. The mathematical language of graph theory is, similarly, an attempt to grasp and summarize merely the pattern of connections.

In the simplest cases, any particular point can be connected to any or all of the other points by single lines. The total number of lines connecting an individual is termed the 'degree' of the corresponding point. Where the lines represent, say, relations that result from the presence of two people at the same event or their shared membership of an organisation, the lines are said to be undirected. However, not all social relations are of this simple type. Where friendship choices are being analysed, for example, choices may not be reciprocated and so the 'direction' of the line needs to be considered. Thus, choosing a friend can be seen as a line directed outwardly from the individual making the choice to the person chosen. Similarly, the receiving of a friendship choice from another can be seen as an inwardly directed line. A graph containing directed lines is technically termed a digraph.

Of course, social relations may not all be equally salient, and it is often necessary to have some measure of the intensity or value of the lines. Thus, identifying someone as a 'close friend' involves greater subjective feelings of emotional intimacy than is the case for a 'distant friend' or an 'acquaintance'. It is conventional

to assign a numerical value to a line in order to represent the intensity of the relationship, producing a valued graph. Any such numerical value is likely to be arbitrary, being based on the analyst's best guess at the strength of the relationship. Typically, intensity has been measured on an arbitrary 1 to 9 scale, with a value of 9 representing the most intense relationship. One particularly important measure of value is the 'multiplicity' of the line. This is a measure that recognises that two people who are each members of three organisations may need to be represented by a line with multiplicity 3, meaning that they have a closer or more intense relationship than two people with only one organisational membership in common.

The discussion so far has assumed that social relations are positive: that people like each other or find their common organisational memberships significant. However, many social relations are negative or involve hostile attitudes: people dislike each other, reject friendship choices, and so on. Where it is important to distinguish positive and negative relations in a network, a mathematical sign (+ or -) may be attached to the line and the resulting graph is called a signed graph.

These basic concepts of direction, value, and sign allow a mapping of the myriad ways in which individuals and organisations relate to each other. They grasp the basic qualities of interpersonal relations and the more abstract features of macro-structural relations. The most fundamental measures are the so-called egocentric measures that are based on a particular focal individual or organisation, referred to as 'ego'. An individual's immediate friendship choices can be represented as a number of outgoing lines, the size of a person's friendship set

being measured by the number of lines—technically referred to as the ‘outdegree’ of the point. In the same way, friendship choices received are represented and measured by the ‘indegree’. Relations of hostility can be measured as the negative outdegree, and all these relations can be regarded as varying in the strength or intensity of the measured lines. Thus, an individual’s immediate field of interaction—his or her ‘neighbourhood’—can be fully charted with the basic concepts of graph theory.

Not all of the relationships of an individual or organisation are immediate or direct. Some social relations are indirect, involving intermediary points. For example, an individual may not know a particular other, but may ‘know a man that does’. Thus two individuals may be indirectly connected through one or more common neighbours. In the words of an old song, a young debutante may sing that she has danced with a man, who danced with a girl, who danced with the Prince of Wales. Such relationships are more distant than are direct relations, and graph theory uses measures of ‘distance’ to represent this. As I have already noted, this measure departs from the assumptions of spatial distance and is calculated by the number of lines needed to connect two points. Ego is connected to others at varying distances according to the number of lines that run between them. A friend of a friend is connected by two lines and so is said to be at distance 2 from ego. Thus, the singer of the song was connected to the Prince of Wales at distance 3. Distances greater than 2 typically lose their significance for the focal individual very rapidly, and such egocentric measures need to be used with care. Nevertheless, it can be important to know, for example, the size of a person’s neighbourhood at distance 2,

as this will include a number of potential intermediaries who can connect them to new contacts or sources of information. This was the basis of Granovetter's (1973) important argument about the strength or importance of relations through relatively distant acquaintances.

It is important, of course, to take account of the direction, sign, and value of lines in measuring distance. If the lines connecting two individuals vary in direction, it may make no sense at all to measure the overall distance by the number of lines. Similarly, if the sign of a second line is the opposite of the sign of the first, then the intermediate point may be a barrier to communication rather than a facilitator of it. Finally, lines of high value may, other things being equal, bring points 'closer' together than lines of low value. These remarks highlight the fact that the properties of a social network can never be inferred simply from the mathematical measures that it is possible to make. It is always necessary to ensure that the measure can be given a substantive sociological significance. A connection at distance 2 can easily be given such an interpretation, but a connection at, say, distance 10 may not be at all significant for the individuals or organisations concerned. Such judgements of significance will vary from one type of relationship to another. Distance in a network of friendship may be more significant than, say, distance in a network of credit relations, though this may not always be the case. Thus, a distant and apparently trivial link held in the computer of a credit-rating agency may affect an individual's credit rating quite significantly. The interpretation of any distance measure depends on the expert judgement of the analyst and the purposes of the research.

The last of the egocentric measures to be reviewed here is the neighbourhood density. This is a measure of the extent to which the immediate contacts of a point are mutually connected with each other. Density is what Bott (1957) referred to as the 'connectedness' of a person's network and is high whenever, for example, many of a person's friends are also friends of each other. In an egocentric network with lower density, a person's various friendships are more segmented. For example, if four people are all directly connected to each other, rather than being connected only through the focal ego, then ego's neighbourhood has a density of 100 per cent. This measure of density is more typically represented on a scale of 0 to 1, with complete 100 per cent density being represented as a value of 1. Many of the qualifications entered about line distance apply to the calculations of neighbourhood density, especially where neighbourhoods are defined at line distances of two or greater. Even in the simpler case, however, it is important to take account of the direction, sign, and value of lines in measuring and interpreting density. Other issues in the measurement of density will be discussed later in this chapter.

Graph theory and global measures

While egocentric measures are always focused on a particular point—ego—the overall, or global properties of a network can also be assessed. The connections of particular individuals concatenate to produce a more or less cohesive structure of relations that is unlikely to have been intended by any of the

participants. Graph theory provides a set of concepts for analysing these macro-level properties of social networks.

In order to measure global properties, however, it is not generally possible to rely on sample data. The egocentric network properties of a large number of individuals can be measured from a sample survey so long as all the normal issues of representativeness are considered. In the case of most global properties, however, the very act of sampling will result in a loss of the information that is required. Global properties are not mere aggregates or averages, as is usually the case in sample data, but are themselves structural features that are apparent only if a whole network is studied and not simply a sample drawn from it. This is not true for all global measures, and work is currently going on to find a solution to this problem. However, this is an issue that always needs to be borne in mind when seeking to measure global properties.

The most fundamental global properties of graphs are degree-based measures. The idea of line distance, for example, is an important global measure and may be more useful at the global level than it is in egocentric networks. The various lines connecting two points form a 'path' whose length can be defined by the number of constituent lines that chain together to form it. Movement through a network depends importantly on the path distances involved. Thus, the communication of ideas or innovations may occur through a whole chain of intermediaries with little or no reliance on the participants all having direct personal knowledge of each other. This was the key to Milgram's discussion of the 'small world'. Some studies of the diffusion of information have suggested that information degrades with distance simply

because each intermediary has less contextual knowledge about the original act of communication and so may misinterpret or distort the information. This is the problem alluded to in the game of Chinese whispers.

Measures of path distance must, of course, take account of the directionality, sign, and value of the various lines. The latter is particularly important as it has been seen as a crucial determinant of the efficiency of the network flow and movement and so as reinforcing any degradation or attenuation effects. A network in which the strength of lines is constant or increasing across paths might be expected, other things being equal, to be an efficient network for communicating ideas or transmitting financial resources. A network in which path distances involve a decline in value at successive steps might be expected to be far less efficient and to result in a significant attenuation of meanings and ideas with distance.

A whole collection of global measures cluster around the idea of the centrality of points within their graphs. The degree of a point—its total of incoming and outgoing lines—is the most basic measure and has been termed local centrality. Calculating the degrees of all points in a network and ranking them from highest to lowest gives a rank order of local centrality. This centrality is 'local' because it highlights points that are well-connected in their immediate neighbourhoods. Such points may not, however, be central in the more global sense that a circle or sphere has a unique centre that can be understood in quasi-spatial terms. Locally central points are well-connected within particular parts of the network but may not be at all well-connected in a global sense.

It is for this reason that a distance-based measure of 'closeness' has generally been preferred as a measure of global centrality. The most commonly used measure of closeness involves calculating the aggregate distances from each point to all others. The score for any particular point is a measure of how close or distant the point is likely to be to any randomly chosen point. The point with the lowest aggregate distance is the most central point in the network. It will be apparent that all the usual qualifications apply and that closeness must be calculated with respect to appropriate indicators of direction, sign, and value.

The converse of centrality is peripherality and it can be useful to know those points that are least close to other members of their networks. Such points are not isolated but are poorly integrated into their network. They are likely to have little influence and to be uninvolved in significant communication flows through the network.

A final measure of centrality is one that measures the extent to which a point is able to act as an intermediary in a large number of network flows. This measure has been called 'betweenness' and refers to the extent to which a particular point is able to serve as an intermediate point of contact between any two other points. The person represented by such a point may be able to control access or the flow of information to others because of the 'structural hole' (Burt 1992) that exists between the two others that he or she connects. A point has a high betweenness centrality when it fills a large number of structural holes in the network. A point with high betweenness centrality may not be at all central in terms of its local degree or global closeness, yet it may play an important part in the structure of the network.

A further set of global measures involve a move away from the properties of points to the properties of the network as a whole. These measures comprise the inclusiveness, density, and centralisation of the network.

Inclusiveness simply measures the number or proportion of the whole set of points that are actually connected into one or more parts of the graph. Some points may be isolates, having no ties to other points, while others will be connected, to a greater or lesser extent, into larger structures. The inclusiveness of a graph is simply the total number of non-isolated points, generally expressed as a percentage of the total number of points. Inclusiveness is a rough and ready approximation to cohesion, but it is usually more informative to measure the actual density of the graph. The idea of neighbourhood density has already been introduced and so the global density of a network should be easy to understand. Global density is the actual number of lines connecting the points of a graph (directly related to inclusiveness) expressed as a proportion of the total number of lines that could possibly exist among the points. Thus, the existence of a large number of isolates reduces the density of a graph. Density can be calculated for valued and directed lines as well as for lines in an undirected graph.

The density of a graph is a very useful and direct measure of its cohesion, but it has one major limitation as a comparative measure of social structure. In real situations, density varies with the size of a network and this limits the possibilities of using the measure to compare different types of network. It is highly unlikely that agents are able to sustain more than a certain number of relationships: our ability to be 'friends' with people,

for example, has its limits. Other things being equal, then, an increase in the size of a network is likely to mean a reduction in its density. This is exacerbated by the fact that the ability to sustain relationships varies with the type of relationship. We may, for example, be able to 'recognise' far more people than we would acknowledge as 'friends', and we are likely to 'love' even fewer. A comparison of the density of the recognition, friendship, and love networks, even if they are of a similar size, will therefore be difficult. Nevertheless, for networks of a similar type, the global density can be a useful comparative measure if reported alongside measures of the size and inclusiveness of the graph.

A further measure of global cohesion is the centralisation of a network. Where centrality relates to the position of particular points, centralisation relates to the overall structure of a network. Centralisation measures the extent to which the cohesion of a network is organised around a specific point or set of connected points. The spokes on a bicycle wheel, for example, form a highly centralised network around its hub. Measures of centralisation can be based on the degree, distance, or betweenness of points, and extensions of these concepts have involved the idea that it is possible to identify the sets of points that comprise the centre, margin, and periphery of the network as a whole.

Graph theory and network differentiation

The global measures considered so far are attempts to measure different aspects of the cohesion of a graph understood as its compactness. An alternative strategy is to focus directly on the

differentiation or fragmentation of a graph in order to uncover the plurality and diversity of its structure. Such a strategy recognises that connected points in a graph may not be connected as a single whole but may instead be connected into a number of distinct structural parts of the whole network. Two broad approaches to the differentiation of networks into 'subgraphs' can be identified: the measurement of 'components' and the measurement of 'cliques'. Each of these types of subgraph has distinct properties.

A component is a subgraph in which all points are directly or indirectly connected to each other and there are no connections to points outside the subgraph. Information or resources can, therefore, flow along a path through all the members of the component but cannot reach any other points in the graph. A graph may comprise one or more components of varying size, and the number and size distribution of components is a fundamental measure of network differentiation and of the existence of boundaries to the flow of information and resources.

The identification of components must go beyond the simple components of undirected graphs and must take account of any direction attached to lines as this will affect the ability of information or resources to flow from one point to another. In some cases it may be legitimate to disregard direction—for example, in order to measure the 'awareness' of agents choosing friendship partners—and this is what are referred to as 'weak components' in contrast to the 'strong components' in which all points are connected through a path of uniformly directed lines.

A further refinement of the simple component idea is that of the cyclic component built from intersecting cycles of

connection. A cycle is a directed path that returns to its starting point. The overlapping of such cycles produces a cyclic component in which all points are connected by one or more cycles and no points have cyclic connections outside the component. A cyclic component is a structural element within a strong component and may be connected to other members of the strong component through 'bridges' that do not lie on the cycle itself.

The various types of component can be further analysed by taking account of degree-based measures such as the intensity or value of the constituent lines. Such an analysis shows the arrangement of components of varying tightness, one inside another. For example, if a measure of value has been attached to the lines, it is possible to slice through the network at various levels of intensity of connection. Such an analysis discloses the 'contour' structure of the network. Components of the kind described so far have taken account of all lines, regardless of intensity and so are connected at the lowest level of intensity. If lines below a chosen level of intensity are disregarded, then the higher level of connection is specified and the analysis moves from the 'valley' of the network to its lower slopes. Progressively raising the chosen level of intensity allows the analysis to approach the 'peaks' of the structure, with intermediate levels marking the gradient of the network. Where points are connected at varying levels of multiplicity, this multiplicity can be used to generate a picture of the 'nested' structure. Nested components, then, stack inside each other like Russian dolls.

The other approach to differentiation that I referred to involves the identification of cliques. A clique differs from a component

in that it is a set of points in which each point is directly and reciprocally connected to all other points. It represents, for example, a tight friendship group or gang. The definition of a clique is, therefore, far more restrictive than that of a component, and cliques tend to be smaller subsets within components. By taking account of the direction of the lines in a clique, it will be possible to identify strong and weak cliques, and if the values of the lines are taken into account it may even be possible to identify a nested structure of cliques.

This idea of the clique is clear and readily interpretable, but it is relatively unusual to find such a structure in real social networks, except in those with exceptionally high density. For this reason, social network analysts have generalised the idea into that of the '*n*-clique' in which less restrictive conditions apply. This '*n*' is the criterion of connectedness in the clique and so altering the value of '*n*' allows the identification of looser cliques. In a 2-clique, for example, all members are connected at distance 2 or less: each point is either directly connected to all others or is connected through an intermediary point. Such a concept can be seen as representing social groups such as those in which individuals are either friends or friends of friends. Although values of '*n*' greater than 2 can be used in clique analysis, they can less readily be given a substantive sociological interpretation and so may be less useful.

Various extensions of the clique idea have been suggested, though they have, so far, been less widely used in sociological work. The most important of these extensions are the 'clan', in which all points are connected by a path of a specified maximum length, and the *k*-plex, in which all points are adjacent to a

specified number of other points. All such measures are available within the principal network analysis programs, but great care must be taken to show that they designate meaningful sociological concepts before they are used.

A final and more useful extension of the clique idea is that of the social circle, seen as a set of overlapping and intersecting cliques (Alba and Kadushin 1976). This concept points to the importance of chains of indirect connections in tying individuals into larger structural elements. The idea of the social circle comes close to the informal idea of the clique originally identified by Warner and his colleagues in their community studies.

Algebraic measures of network structure

Algebraic methods represent the points of a graph in sets that correspond to social 'positions' and that can be handled in equations or analysed in matrices. I have traced the origins of this approach to social network analysis in the works of Warner and Homans. The most important recent developments in algebraic network analysis have been those inspired by Harrison White and his work on 'structural equivalence'.

Structurally equivalent points are those that can be regarded as substitutable for each other: they occupy similar structural positions in relation to others, though they may be connected to *different* others. Culturally defined and institutionalised roles—such as father, mother, child, teacher, and so on—are the clearest examples, but the idea also involves purely relational categories such as classes or political factions that generate structural

uniformities of experience and action. The approach taken by Lorrain and White (1971) was to represent agents and their affiliations or participations in a matrix and then to simultaneously rearrange both the rows and the columns until blocks of similarly connected points appear. This 'blockmodelling' reduces the structure of the whole graph to a structure of social positions, such as a role set or a class structure.

The rearrangement of a matrix involves one or another form of cluster analysis (Everitt 1974). Methods of cluster analysis combine or separate points according to their similarity, and in White's approach the similarity between two points is measured by the correlation found in their patterns of connection within a graph (White *et al.* 1976; Boorman and White 1976). The method used by White and his colleagues to identify structurally equivalent positions was termed CONCOR and is available within the standard social network analysis software programs. The boundaries of clusters can sometimes be identified visually—as was the case in Homans's reanalysis of the Old City data—but in larger and more complex networks they must be identified from the patterns of high and low density found in the various blocks. Once a repeated cluster analysis has produced a simplified pattern of zero and non-zero blocks of density, a reduced network can be drawn in which non-zero blocks are represented as super-points and zero blocks are represented by the absence of connections. This is referred to as an 'image graph' and has been seen as a delineation of the positional structure of a social network.

An alternative matrix method is that of Doug White and Karl Reitz (1983), which they call regular equivalence. Structural

equivalence takes account of direct connections, but regular equivalence takes account of the wider patterns of connection. It looks at similarity of connections at varying specified path lengths. The resulting matrix of similarities can be analysed through similar cluster analytic methods to yield image graphs. This looser measure of substitutability is, arguably, more useful for much social network analysis.

Algebraic methods offer powerful measures for the analysis of social structure (see Pattison 1993), but they are difficult to apply in highly fragmented and differentiated networks. They may be usefully applied *within* components or fully connected graphs, but are impossible to apply *between* components where no connections exist.

Spatial and cartographical approaches

Spatial approaches such as smallest space analysis and multidimensional scaling are attempts to plot points in a spatial field through geometrical methods that allow the construction of sociograms as true 'maps' showing a configuration of points that depicts their relative physical distance and spatial direction as well as retaining information on the familiar path distances and line directions. As with blockmodelling, spatial approaches start out from measures of similarity (or dissimilarity) but use cartographical techniques to produce a best-fit configuration in which the relative similarities among points correspond directly to the relative physical distances that can be shown on paper or a computer screen.

Central to these methods is the search for an appropriate number of dimensions in which to embed the configuration. A maximum of two dimensions can be directly represented on paper, as is the case with an atlas of physical maps showing places in relation to the north/south and east/west dimensions. Computer software, however, allows the depiction of three-dimensional configurations on screen and allows a configuration to be rotated to show the shape and solidity of the networks from various angles. Such representations give clear visual impressions of the density, centralisation, and fragmentation of the networks. It can be difficult to represent more than three dimensions in physical form, and it may be necessary to resort to simplifications such as, on paper, presenting the successive two-dimensional views of a three-dimensional structure.

However many dimensions are used, there is always the question of providing a sociological interpretation of each dimension. It is rarely the case that a configuration will take a directly geographical form and place points in a north/south and east/west two-dimensional space. In social space, dimensions will tend to show scales of economic or political inequality, civic participation, religious affiliation and so on. The key task in spatial analysis of any given configuration is to rotate it in its space in order to yield the most meaningful interpretation of the arrangement of points. Such an interpretation is never given by the method used but must be made and supplied by the analyst.

The configurations produced by these geometrical methods retain the ideas of relative distance and spatial direction that were highlighted in Lewin's work, and they avoid the arbitrariness that is involved in the classic attempts to draw sociograms.

It is also possible to show the lines that connect points within a spatial map, with the length of each line reflecting the 'as-the-crow-flies' distances between points. Path distances, which in basic graph theory simply represent the number of links, become measures of actual spatial distance along the paths that connect the points. Thus, spatial models combine spatial distance with accurate representations of the 'routes' that must be taken in the communication of information or resources through a network. Notions of centrality may also be clear from the patterns of lines, though globally central points may not appear as spatially central to the whole configuration.

It is also possible to represent components, nesting, and structurally equivalent positions within a spatial configuration. Component boundaries can be drawn around its connected points and measures of the value of the constituent lines can be used to draw the nested boundaries within this. Such a diagram corresponds closely to the depictions of hills and valleys on a map by the drawing of contour lines. Such an approach gives a ready depiction of the topography of a social space and can give a useful way of interpreting its structure. Positional ideas may be more difficult to represent with such clarity, as the members of the various blocks may not be spatially contiguous. Nevertheless, the use of different colours to represent the various blocks will bring out the equivalence patterns and so help to identify the relationships between block membership and the membership of cliques and components.

These uses of spatial analysis have been made possible by advances in computing. These have allowed the handling of large data sets and the easy and quick production of sociograms, but

they have also allowed the use of powerful on-screen methods for visualising the social networks generated. Standard software allows the inspection of three-dimensional configurations and allows the points to be coloured to indicate their properties. More specialised programs allow moving pictures to be constructed so as to depict patterns of change over time.

Statistical inference

Much work in social network analysis has been both static and descriptive. Recent work, however, has shown how it is possible to move towards dynamic understandings of social change and to construct explanations of the patterns described. This has led to much greater attention being given to statistical methods for assessing the significance of results and the validity of explanations. This work follows the general and well-established principles of statistical inference and hypothesis testing, adapting these to the specific requirements of using relational data.

A hypothesis to explain observations must be produced by expert use of the sociological imagination, but any hypothesis proffered must be tested before it can be accepted with any confidence. Statistical methods of hypothesis testing involve comparing the observed results from particular patterns of change to the results that might be expected to occur as a result of random variations alone. That is, the methods attempt to show the likelihood that any result could have occurred simply by chance. If this probability is low, then some confidence can be taken in the mechanisms that have been hypothesised.

A number of approaches to the significance testing of relational data have now been proposed, the most influential and important being the so-called exponential random graph models (ERGM) of Wasserman and his colleagues (see, for example, Wasserman and Pattison 1996; Wasserman 1980). This approach uses regression techniques to compare an observed pattern with a range of possible graphs produced through random simulations from the same data. Confidence in the significance of the observed results is greatest when it corresponds with a randomly generated pattern that has a very low probability of occurrence by chance.

Further discussion of statistical inference is beyond the scope of this book. The method will, however, be of increasing importance in social network analysis, which has constantly been bedevilled by the critical comment: 'So what?'. Critics have pointed to the fact that simply describing an observed configuration of points and lines means nothing unless a plausible sociological interpretation can be given to it. Statistical methods will be an essential element in avoiding this critical response in future.

Further reading

An appropriate working strategy for building on the ideas outlined in this chapter would be to work through the following sources in sequence:

Scott, John 2012. *Social Network Analysis*. Third Edition. London: Sage. The principal chapters in this book cover all the concepts and measures in greater detail and are designed to be read by those with little or no technical background in social network analysis.

Scott, John and Carrington, Peter C. (eds.). 2011. *The Sage Handbook of Social Network Analysis*. London: Sage. A comprehensive collection of introductory material that elaborates on basic concepts with greater background. Section 3 covers key methodological ideas.

Wasserman, Stan and Faust, Katie 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press. The best and most advanced text that gives a thorough and comprehensive discussion of techniques of social network analysis.

Carrington, Peter J., Scott, John and Wasserman, Stanley (eds.). 2005. *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press. A collection of specialised papers reviewing recent advances and leading-edge developments in network methodology.

Background sources on all key areas of network analysis have been brought together in various compilations and source books:

Scott, John (ed.). 2002. *Social Networks*, Four Volumes. London: Routledge. Volumes 1 and 2 include key sources on concepts and techniques.

Holland, P. and Leinhardt, S. (eds.). 1979. *Perspectives on Social Networks*. New York: Academic Press. A shorter collection with a number of important source papers.