Social Network Analysis: A Handbook; SAGE Publication
(http://www.analytictech.com/mb119/tableof.htm)

Chapter 4: Lines, Direction, and Density

In the previous chapter I looked at the ways in which relational data can be handled and managed in matrix form. Many fundamental features of social networks can be analysed through the direct manipulation of matrices - the transposing, adding and multiplying of matrices all yield information on their structure. Matrix algebra, however, is rather complex for most researchers (but see Meek and Bradley, 1986). Although matrices are useful for the organization and storage of relational data, specialist computer programs allow an easier and more direct approach to network analysis. The available packages implement a variety of analytical procedures, and any user of the programs must have some understanding of how they work.

A common framework for social network analysis programs is the mathematical approach of graph theory, which provides a formal language for describing networks and their features. Graph theory offers a translation of matrix data into formal concepts and theorems which can be directly related to the substantive features of social networks. If the sociogram is one way of representing relational matrix data, the language of graph theory is another, and more general, way of doing this. While it is not the only mathematical theory which has been used for modelling social networks, it is a starting point for many of the most fundamental ideas of social network analysis.

It is the concepts of graph theory which figure as the principal procedures in the UCINET and GRADAP programs, though the readily accessible computer programs endeavour to keep as much of the mathematics as possible hidden from the user. Data in matrix form can be read by the programs, and suitable graph theoretical concepts can be explored without the researcher needing to know anything at all about the mechanics of the theory or of matrix algebra. Nevertheless, an understanding of graph theory will significantly help to improve the sophistication of a researcher's analyses, by ensuring that he or she chooses appropriate procedures. Indeed, GRADAP's data structure and management procedures require an understanding of basic graph theoretical ideas.

Graph theory concerns sets of elements and the relations among these, the elements being termed points and the relations lines.' Thus, a matrix describing the relations among a group of people can be converted into a graph of points connected by lines. A sociograin, therefore, is a 'graph'. So far, this should be very familiar from what has already been discussed in Chapters 2 and 3. It is important to be clear about the difference between this idea of a 'graph' and the graphs of variables used in statistics and other branches of quantitative mathematics. These more familiar graphs we might term them 'graphs of variables' - plot, for example, frequency data on axes which represent the variables. The graphs of graph theory - 'graphs of networks' - express the qualitative patterns of connection among points. Indeed, graph diagrams themselves are of secondary importance in graph theory. As has already been suggested, it is often very difficult to draw a clear and comprehensible diagram for large sets of points with complex patterns of connection. By expressing the properties of the graph in a more abstract mathematical form, it is possible to dispense with the need to draw a sociogram and so make it easier to manipulate very large graphs.

Nevertheless, the drawing of graph diagrams has always been of great illustrative importance in graph theory, and many others will be used in this book. Because of the visual simplicity of small sociograms, I will begin with an introduction to the principles involved in drawing graph diagrams before going on to introduce the basic concepts of graph theory.

Sociograms and Graph Theory

A graph diagram aims to represent each row or column in an incidence matrix - each of the cases or affiliations under investigation - by a point on the paper. Once the appropriate adjacency matrix has been derived, the 'I' and 'O' entries in the cells of the matrix, representing the presence or absence of a relation, can be indicated by the presence or absence of lines between the points. In Figure 3.5, for example, the symmetrical 4-by-4 adjacency matrix of companies can be drawn as a four point graph containing six lines, which correspond to the non-zero entries in the matrix.

In a graph, it is the pattern of connections that is important, and not the actual positioning of the points on the page. The graph theorist has no interest in the relative position of two points on the page, the lengths of the lines which are drawn between them, or the size of character used to indicate the points. Graph theory does involve concepts of length and location, for example, but these do not correspond to those concepts of physical length and location with which we are most familiar. It is usual in a graph diagram to draw all the lines with the same physical length, wherever this is possible, but this is a purely aesthetic convention and a matter of practical convenience. Indeed, it is not always possible to maintain this convention if the graph is to be drawn with any clarity. For this reason, there is no one 'correct' way to draw a graph. The graph diagrams in Figure 4.1, for example, are equally valid ways of drawing the same graph - all convey exactly the same graph theoretical information.

The concepts of graph theory, then, are used to describe the pattern of connections among points. The simplest of graph theoretical concepts refer to the properties of the individual points and lines from which a graph is constructed, and these are the building blocks for more complex structural ideas. In this chapter I shall review these basic concepts and show how they can be used to give an overview of both the ego-centric and the global features of networks. Subsequent chapters will explore some of the more complex concepts.

It is necessary first to consider the types of lines that can be used in the construction of graphs. Lines can correspond to any of the types of relational data distinguished in Figure 3.6: undirected, directed, valued or both directed and valued. The graphs in Figure 4.1 consist of undirected lines. These graphs derive from a symmetrical data matrix where it is simply the presence or absence of a relation which is of importance. If the relations are directed from one agent to another, then they can be represented in a directed graph, sometimes termed a 'digraph'. A directed graph is represented in drawn form by attaching an arrow head to each line, the direction of the arrow indicating the direction of the relation. Figure 4.2 shows a simple directed graph.

If, on the other hand, the intensity of the relation is an important consideration and can be represented by a numerical value, the researcher can construct a valued graph in which

(i) Adjacency matrix

	A	B	C	D	E	Row sum
A		1	0	0	1	2
B	1		1	1	1	4
C	0	1	-	1	0	2
D	0	1	1	-	0	2
E	1	1	0	0	-	2
Column sum	2	4	2	2	2	

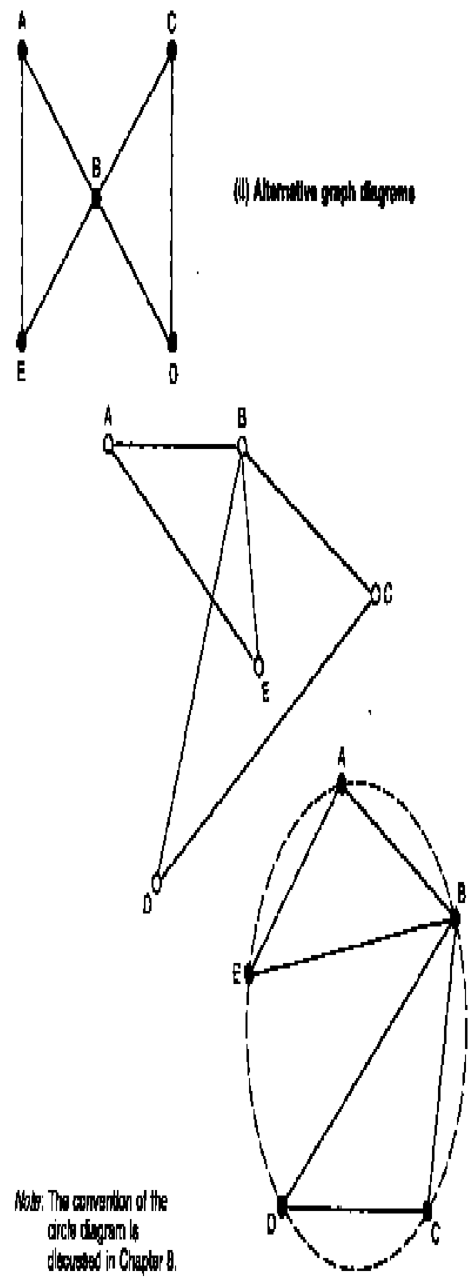
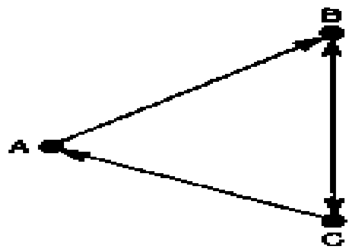


Figure 4.1 Alternate drawings of a graph

numerical values are attached to each of the lines. I have already shown that a matrix for a directed graph will not usually be symmetrical, as relations will not normally be reciprocated. A matrix for a valued graph may or may not be symmetrical, but it will contain values instead of simple binary entries. 3 An example of a valued graph is that in Figure 3.5. One of the simplest and most widely used measures of intensity is the multiplicity of a line. This is simply the number of separate contacts which make up the relationship. If, for example, two companies have two directors in common, the relation between the two companies can be represented by a line of multiplicity 2. If they have three directors in common, the interlocking directorship can be seen as a line of multiplicity 3. The values in a graph can, of course, relate to any other suitable measure of intensity, such as, for example, the frequency of the relation.



	A	B	C	Row sum
A	—	1	0	1
B	0	—	1	1
C	1	1	—	2
Column sum	1	2	1	

Figure 4.2 A directed graph and its matrix

The fundamental ideas of graph theory can most easily be understood in relation to simple undirected and un-valued graphs. A number of apparently straightforward words are used to refer to graph theoretical terms, and it may appear pedantic to define these at great length. But these definitional matters are important, as the apparently simple words are used in highly specific and technical ways. It is essential that their meanings are clarified if the power of graph theory is to be understood.

Two points which are connected by a line are said to be adjacent to one another. Adjacency is the graph theoretical expression of the fact that two agents represented by points are directly related or connected with one another. Those points to which a particular point is adjacent are termed its neighborhood, and the total number of other points in its neighbourhood is termed its degree (strictly, its 'degree of connection'). Thus, the 'degree' of a point is a numerical measure of the size of its neighbourhood. The degree of a point is shown by the number of non-zero entries for that point in its row or column entry in the adjacency matrix. Where the data are binary, as in Figure 4.1, the degree is simply the row or column sum for that point. Because each line in a graph connects two points -- is 'incident' to two points -- the total sum of the degrees of all the points in a graph must equal twice the total number of lines in the graph. The reason for this is that each line is counted twice when calculating the degrees of the separate points. This can be confirmed by examining Figure 4. 1. In this graph, point B has a degree of 4 and all the other points have a degree of 2. Thus, the sum of the degrees is 12, which is equal to twice the number of lines (six).

Points may be directly connected by a line, or they may be indirectly connected through a sequence of lines. A sequence of lines in a graph is a 'walk', and a walk in which each point and each line are distinct is called a path. The concept of the path is, after those of the point and the line, one of the most basic of all graph theoretical concepts. The length of a path is measured by the number of lines which make it up. In Figure 4. 1, for example, points A and D are not directly connected by a line, but they are connected through the path ABD, which has a length of 2. A particularly important concept in graph theory is that of 'distance', but neither distance nor length correspond to their everyday physical meanings. The length of a path, I have said, is simply the number of lines which it contains - the number of 'steps' necessary to get from one point to another. The distance between two points is the length of the shortest path (the 'geodesic') which connects them.

Social Network Analysis: A Handbook; SAGE Publication
(http://www.analytictech.com/mb119/tableof.htm)

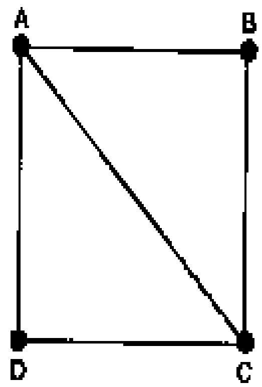


Figure 4.3 Lines and paths

Consider the simple graph in Figure 4.3. In this graph, AD is a path of length 1 (it is a line), while ABCD is a path of length 3. The walk ABCAD is not a path, as it passes twice through point A. It can be seen that points A and D are connected by three distinct paths: AD at length 1, ACD at length 2, and ABCD at length 3.' The distance between A and D, however, is the length of the shortest path between them, which, in this case, is 1. The distance between points B and D, on the other hand, is 2. Many of the more complex graph theoretical measures take account only of geodesics, shortest paths, while others consider all the paths in a graph.

These same concepts can be used with directed graphs, though some modifications must be made to them. In a directed graph lines are directed to or from the various points. Each line must be considered along with its direction, and there will not be the symmetry that exists in simple, undirected relational data. The fact that, for example, A chooses B as a friend does not mean that there will be a matching friendship choice from B to A. For this reason, the 'degree' of a point in a directed graph comprises two distinct elements, called the 'indegree' and the 'outdegree'. These are defined by the direction of the lines which represent the social relations. The indegree of a point is the total number of other points which have lines directed towards it; and its outdegree is the total number of other points to which it directs lines. The indegree of a point, therefore, is shown by its column sum in the matrix of the directed graph, while its outdegree is shown by its row sum. The column sum of point B in Figure 4.2, for example, is 2, as it ‚receives' two lines (from A and from C). The corresponding sociogram shows clearly that its indegree is 2. The row sum for B, on the other hand, is 1, reflecting the fact that it directs just one line, to point C.

A path in a directed graph is a sequence of lines in which all the arrows point in the same direction. The sequence CAB in Figure 4.2, for example, is a path, but CBA is not: the changing direction of the arrows means that it is not possible to 'reach' A from C by 6 passing through B. It can be seen that the criteria for connection are much stricter in a directed graph, as the researcher must take account of the direction of the lines rather than simply the presence or absence of a line. The distance between two points in a directed graph, for example, must be measured only along the paths that can be identified when direction is taken into account. When agents are regarded as either 'sources' or 'sinks' for the 'flow' of resources or information through a network, for example, it is sensible to take serious account of this directional information in analysing the graph of the network. Sometimes, however, the direction of the lines can legitimately be ignored. If it is the mere presence or absence of a line which is important, its direction being a relatively unimportant factor, it is possible to relax the usual strict criteria of connection and to regard any two points as connected if there is a sequence of lines between them, regardless of the directions of the arrows. In such an analysis it is usual to speak of the presence of a ‚semi-path' rather than a path. CBA in Figure 4.2 is a semi-path. Treating directed data as if they were undirected, therefore, means that all the usual measures for undirected data may then be used.

Density: Ego-centric and Socio-centric

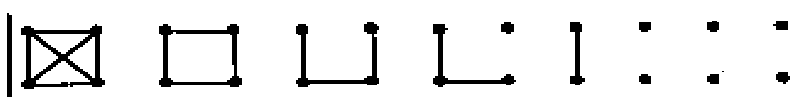
One of the most widely used, and perhaps over-used, concepts in graph theory is that of 'density', which describes the general level of linkage among the points in a graph. A 'complete' graph is one in which all the points are adjacent to one another: each point is connected directly to every other point. Such completion is very rare, even in very small networks, and the concept of density is an attempt to summarize the overall distribution of lines in order to measure how far from this state of completion the graph is. The more points that are connected to one another, the more dense will the graph be.

Social Network Analysis: A Handbook; SAGE Publication
(http://www.analytictech.com/mb119/tableof.htm)

Density, then, depends upon two other parameters of network structure: these are the 'inclusiveness' of the graph and the sum of the degrees of its points. Inclusiveness refers to the number of points which are included within the Various connected parts of the graph. Put in another way, the inclusiveness of a graph is the total number of points minus the number of isolated points. The most useful measure of inclusiveness for comparing various graphs is the number of connected points expressed as a proportion of the total number of points. Thus, a 20-point graph with five isolated points would have an inclusiveness of 0.75. An isolated point is incident with no lines and so can contribute nothing to the density of the graph. Thus, the more inclusive is the graph, the more dense will it be. Those points which are connected to one another, however, will vary in their degree of connection. Some points will be connected to many other points, while others will be less well connected. The higher the degrees of the points in a graph, the more dense will it be. In order to measure density, then, it is necessary to use a formula which incorporates these two parameters. This involves comparing the actual number of lines which are present in a graph with the total number of lines which would be present if the graph were complete.

The actual number of lines in a graph is a direct reflection of its inclusiveness and the degrees of its points. This may be calculated directly in small graphs, but in larger graphs it must be calculated from the adjacency matrix. The number of lines in any graph is equal to half the sum of the degrees. In Figure 4. 1, as I have already shown, half the sum of the row or column totals is six. The maximum number of lines which could be present in this graph can be easily calculated from the number of points that it contains. Each point may be connected to all except one other point (itself), and so an undirected graph with n points can contain a maximum of n(n-1)12 distinct lines. Calculating n(n-1) would give the total number of pairs of points in the graph, but the number of lines which could connect these points is half this total, as the line connecting the pair A and B is the same as that connecting the pair B and A. Thus, a graph with three points can have a maximum of three lines connecting its points; one with four points can have a maximum of six lines; one with five points can have a maximum of ten lines; and so on. It can be seen that the number of lines increases at a much faster rate than the number of points. Indeed, this is one of the biggest obstacles to computing measures for large networks. A graph with 250 points, for example, can contain up to 31,125 lines.

The density of a graph is defined as the number of lines in a graph, expressed as a proportion of the maximum possible number of lines. The formula for the density is $\frac{1}{n(n-1)/_2}$ where 1 is the number of lines present. This measure can vary from 0 to 1, the density of a complete graph being 1. The densities of various graphs can be seen in Figure 4.4: each graph contains four points and so could contain a maximum of six lines. It can be seen how the density varies with the inclusiveness and the sum of the degrees.



No. of 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

Figure 4.4 Density comparisons

Social Network Analysis: A Handbook; SAGE Publication
(http://www.analytictech.com/mb119/tableof.htm)

In directed graphs the calculation of the density must be slightly different. The matrix for directed data is asymmetrical, as a directed line from A to B will not necessarily involve a reciprocated line directed from B to A. For this reason, the maximum number of lines which could be present in a directed graph is equal to the total number of pairs that it contains. This is simply calculated as n(n - 1). The density formula for a directed graph, therefore, is lln(n-1).

Barnes (1974) has contrasted two approaches to social network analysis. On the one hand is the approach of those who seek to anchor social networks around particular points of reference (e.g., Mitchell, 1969) and which, therefore, advocates the investigation of ego-centric' networks. From such a standpoint, the analysis of density would be concerned with the density of links surrounding particular agents. On the other hand, Barnes sees the 'socio-centric' approach, which focuses on the pattern of connections in the network as a whole, as being the distinctive contribution of social network analysis. From this standpoint, the density is that of the overall network, and not simply the 'personal networks' of focal agents. Barnes holds that the socio-centric approach is of central importance as the constraining power of a network on its members is not mediated only through their direct links. It is the concatenation of indirect linkages, through a configuration of relations with properties that exist independently of particular agents, that should be at the centre of attention.

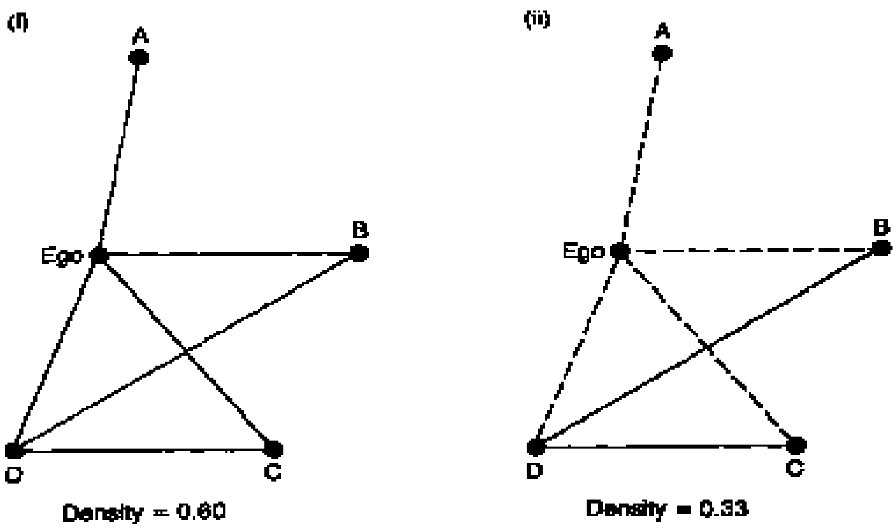


Figure 4.5 Ego-centric measures of density

In the case of an ego-centric approach, an important qualification must be made to the way in which density is measured. In an egocentric network it is usual to disregard the focal agent and his or her direct contacts, concentrating only on the links which exist among these contacts. Figure 4.5 shows the consequences of this. Sociogram (i) shows a network of five individuals anchored around 'ego'. The sociogram shows ego's direct contacts and the relations which exist among these contacts. There is a total of six lines, and the density of the sociogram is 0.60. But the density is at this relatively high level principally because of the four lines which connect ego to A, B, C and D. These relations will exist almost by definition, and should usually be ignored. If these data had, for example, been obtained through a questionnaire which asked respondents to name their four best friends, the high density would be an artifact of the question wording. The relations to the four nominated contacts of each respondent will swamp any information about the relations among those who are named by each respondent. The significant fact about sociogram (i) is that there are relatively few connections among ego's own contacts. In sociogram (ii), where ego's direct contacts are shown as dotted lines, there are two relations among A, B, C and D (shown as solid lines), and the four person network has a density of 0.33. It should be clear that this is a more useful measure of the density of the ego-centric network.

Social Network Analysis: A Handbook; SAGE Publication
(http://www.analytictech.com/mb119/tableof.htm)

It is also possible to use the density measure with valued graphs, though there is very little agreement about how this should be done. The simplest solution, of course, would be to disregard the values of the lines and to treat the graph as a simple directed or undirected graph. But this involves a considerable loss of information. It might be reasonable, for example, to see lines with a high multiplicity as contributing more to the density of the graph than lines with a low multiplicity. This would suggest that the number of lines in a valued graph might be weighted by their multiplicities: a line with multiplicity 3 might be counted as being the equivalent of three lines. Simple multiplication, then, would give a weighted total for the actual number of lines in a graph. But the denominator of the density formula is not so easy to calculate for valued graphs. The denominator, it will be recalled, is the maximum possible number of lines which a graph could contain. This figure would need to be based on some assumption about the maximum possible value which could be taken by the multiplicity in the network in question. If the maximum multiplicity is assumed to be 4, then the weighted maximum number of lines would be equal to four times the figure that would apply for a similar unvalued graph. But how might a researcher decide on an estimate of what the maximum multiplicity for a particular relation might be? One solution would be to take the highest multiplicity actually found in the network and to use this as the weighting (Barnes, 1969). There is, however, no particular reason why the highest multiplicity actually found should correspond to the theoretically possible maximum. In fact, a maximum value for the multiplicity can be estimated only when the researcher has some independent information about the nature of the relationships under investigation. In the case of company interlocks, for example, average board size and the number of directorships might be taken as weightings. If the mean board size was five, for example, and it is assumed that no person can hold more than two directorships, then the mean multiplicity would be 5 in a complete and fully connected graph.

In the case of the company sociogram in Figure 3.5, for example, the weighted total of lines measured on this basis would be 5 times 6, or 30. The actual total of weighted lines in the same Sociogram, produced by adding the values of all the lines, is 12, and so the multiplicity-based density would be 12/30, or 0.4. This compares with a density of 1.0 which would be calculated if the data were treated as if they were unvalued. It must be remembered, however, that the multiplicity-based calculation is based on an argument about the assumed maximum number of directorships that a person can hold. If it were assumed that a person could hold a maximum of three directorships, for example, then the density of the company sociogram would fall from 0.4 to 0.2. For other measures of intensity, there is no obvious way of weighting lines.

The density measure for valued graphs, therefore, is highly sensitive to those assumptions which a researcher makes about the data. A measure of density calculated in this way, however, is totally incommensurable with a measure of density for unvalued data. For this reason, it is important that a researcher does not simply use a measure because it is available in a standard r)rolram. A researcher must always be perfectly clear about the assumptions that are involved in any particular procedure, and must report these along with the density measures calculated. The problem in handling valued data may be even more complex if the values do not refer to multiplicities.

A far more fundamental problem which affects all measures of density must now be considered. This is the problem of the dependence of the density on the size of a graph, which prevents density measures being compared across networks of different sizes (see Niemeijer, 1973; Friedkin, 1981; Snijders, 1981). Density, it will be recalled, varies with the number of lines which are present in a graph, this being compared with the number of lines which would be present in a complete graph. There are verv good reasons to believe that the maximum number of lines achievable in any real graph may be well below the theoretically possible maximum. If there is an upper limit to the number of relations that each agent can sustain, the total number of lines in the graph will be limited bv the number of agents. This limit on the total number of lines means that larger graphs will, other things being equal, have lower densities than small graphs. This is linked, in particular, to the time constraints under which agents operate. Mayhew and Levinger (1976) argue that there are limits on the amount of time that people can invest in making and maintaining relations. The time that can be allocated to any particular relation, they argue, is limited, and it will decline as the number of contacts increases. Agents will, therefore, decide to stop making new relations, new investments of time, when the rewards decline and it becomes too costly. The number of contacts that they can sustain, therefore, declines as the size of the network increases. Time constraints, therefore, produce a limit to the number of contacts and, therefore, to the density of the network. Mayhew and Levinger have used models of random choice to suggest that the maximum value for density that is likely to be found in actual graphs is 0.5.1'

The ability of agents to sustain relations is also limited by the particular kind of relation that is involved. A 'loving' relation, for example, generally involves more emotional commitment than an ‘awareness' relation, and it is likely that people can be aware of many more people than they could love. This means that any network of loving relations is likely to have a lower density than any network of awareness relations.

I suggested in Chapter 3 that density was one of the network measures that might reasonably be estimated from sample data. Now that the measurement of density has been more fully discussed, it is possible to look at this suggestion in greater detail. The simplest and most straightforward way to measure the density of a large network from sample data would be to estimate it from the mean degree of the cases included in the sample. With a representative sample of a sufficient size, a measure of the mean degree would be as reliable as any measure of population attributes derived from sample data, though I have suggested in the previous chapter some of the reasons why sample data may fail to reflect the full range of relations. If the estimate was, indeed, felt to be reliable, it can be used to calculate the number of lines in the network. The degree sum - the sum of the degrees of all the points in the graph - is equal to the estimated mean degree multiplied by the total number of cases in the population. Once this sum is calculated, the number of lines is easily calculated as half this figure. As the maximum possible number of lines can always be calculated directly from the total number of points (it is always equal to n(n - 1)/2 in an undirected graph), the density of the graph can be estimated by calculating

$$\frac{(n * mean\ degree)/2}{n(n-1)/2}$$

which reduces to

$$\frac{n * mean\ degree}{n(n-1)}$$

Granovetter (1976) has gone further than this and has attempted to provide a method of density estimation that can be used when the researcher is uncertain about the reliability of the initial estimate of the mean degree. In some situations there will be a high reliability to this estimate. With company interlock data, for example, the available directories of company information allow researchers to obtain complete information on the connections of the sample companies to all companies in the population, within the limits of accuracy achieved by the directories. In such circumstances, an estimate of mean degree would be reliable. In studies of acquaintance, on the other hand, such reliability is not normally the case, especially when the population is very large. Granovetter's solution is to reject a single large sample in favour of a number of smaller samples. The graphs of acquaintance in each of the sub-samples (the ‘random sub-graphs') can be examined for their densities, and Granovetter shows that an average of the random sub-graph densities results in a reliable estimate of the population network density. Using standard statistical theory, Granovetter has shown that, for a population of 100,000, samples of between 100 and 200 cases will allow reliable estimates to be made. With a sample size of 100, five such samples would be needed; with a sample size of 200, only two samples would be needed. 12 These points have been further explored in field research, which has confirmed the general strategy (Erickson et al., 1981; Erickson and Nosanchuck, 1983).

Density is, then, an easily calculated measure for both undirected and directed graphs, it can be used in both ego-centric and sociocentric studies, and it can reliably be estimated from sample data. It is hardly surprising that it has become one of the commonest measures in social network analysis. I hope that I have suggested, however, some of the limits on its usefulness. It is a problematic measure to use with valued data, it varies with the type of relation and with the size of the graph, and, for this reason, it cannot be used for comparisons across networks which vary significantly in size. Despite these limitations, the measurement of density will, rightly, retain its importance in social network analysis. If it is reported along with such other measures as the inclusiveness and the network size, it can continue to play a powerful role in the comparative study of social networks.

Community Structure and Density

The power and utility of density analysis can be illustrated through some concrete studies. Barry Wellman (1979, 1982), a member of Harrison White's original cohort of network analysts at Harvard, has supervised a large study of community structure, in which density plays a key role. He took as his starting point the longstanding tradition of community studies, in which writers on 'community' were generally concerned to investigate whether the communal solidarities associated with small-scale, rural villages had been able to withstand the modernizing forces of industrialization and urbanization. Wellman wanted to use social network analysis to see whether the development of modern society had resulted in the disappearance of community and the emergence of urban anomie. It had been pointed out by some critics of community studies that social relations of all kinds had become detached from specific localities, with relations having an increasingly national or international scope (see the discussion in Bulmer, 1985). Wellman's research aimed to investigate this issue for a particular urban area in Toronto, East York, and, like Fischer (1977, 1982), he focused on the question of whether 'personal communities' had stretched beyond the bounds of the local neighbourhood itself.

East York is an inner city suburb of private houses and apartment blocks and was, at the time of the research in 1968, occupied mainly by skilled manual workers and routine white collar workers. The fieldwork involved interviews with a random sample of 845 adults, and a central question in the interview asked people to name their six closest associates. They were then asked to say whether those named were themselves close to one another (see also McCallister and Fischer, 1978). The responses to these questions could be used to construct ego-centric networks of intimate association for each respondent. By asking about the connections among the persons who were named by each respondent, Wellman was able to measure the density of each personal network. The calculation of density followed the strategy outlined earlier, and ignored links between respondents and their intimates. That is, data were collected on ego and his or her six intimate associates, but the densities of the egocentric networks were calculated for the links among the six associates only.

Wellman discovered that many of the intimate associates (about a half) were relatives of the respondents, but kin and non-kin associates were all to be found across a wide geographical area. The majority of all links were with people who lived in the city itself, though very few of these links were based in the immediate locality of East York. A quarter of all the intimate associates who were named lived outside the city, some living overseas. Having made a number of these summary statements about the broad framework of people's social networks, Wellman turned to the densities of these networks. The mean density of the ego-centric personal networks of the respondents was 0.33, 1 3 only one-fifth of networks having a density greater than 0.50 (Wellman, 1979: 1215). A density of 0.33 meant that five out of the 15 possible links among intimate associates were actually present. 14 Wellman discovered that the densest networks tended to be those that were composed mainly of kin, owing to the fact that it was more likely that the kin of the respondents would maintain mutual contacts. Where kinship obligations were absent, such contacts were less likely to be maintained.

Density	% of networks	% of network members who are kin
0–0.25	47.1	36.4
0.26–0.50	31.7	58.9
0.51–0.75	7.9	56.9
0.76–1.00	13.3	73.7
	100.0	
	(n = 824)	

Figure 4.6 Density of personal networks

Wellman's principal findings on personal networks are summarized in Figure 4.6. He interprets these data as indicating that people were involved in networks which were 'sparsely knit' - i.e., networks of low density. 'Communal' links were neither solidaristic nor localized. People had others that they could rely on, but the low density of their personal networks, their lack of mutual crosslinkages, meant that such help was limited. These personal networks were, nevertheless, important sources of help and support, on both an everyday basis and in emergencies: 'East Yorkers can almost always count on help from at least one of their intimates, but they cannot count on such help from most of them' (Wellman, 1979: 1217). Those intimate associates who were less likely to give help and support were more likely to be significant for sociability. Helpers were more likely to be kin, while those who were most important for sociability were more likely to be co-residents or co-workers.

To pursue some further issues, a follow-up study was undertaken in which in-depth interviews were carried out, during 1977-8, with 34 of the original respondents. The aim was to get more 'qualitative' contextual data for the structural data of the earlier study. Although the detailed results of this stage of the inquiry go beyond the immediate concerns of this chapter, some of the directions pursued can usefully be outlined. Wellman discovered that the interpersonal networks of households were differentiated by gender divisions and by the involvement of household members in paid work. The research discovered, for example, a number of differences between households where women were involved in paid employment and those where they were involved only in domestic work. He discovered that the social relations of a household and their access to interpersonal support from their kin, friends, neighbours and coworkers were most likely to be maintained by women rather than by men. This was, in particular, true of households where women were engaged solely in domestic work. Households where women were involved in both domestic work and in paid employment had far less dense networks of relations and were, therefore, able to obtain less support and fewer services from their contacts (Wellman, 1985).

Wellman's investigations used survey analysis to generate the relational data which he used in the study, but similar ideas can be used on other forms of relational data. Smith (1979), for example, used historical data derived from documentary sources to investigate communal patterns in an English village in the thirteenth century. Smith's data came from the records of the manor court of Redgrave in Suffolk, these records showing such things as patterns of landholding, property transactions and financial disputes among the villagers. In total, he considered 13,592 relations among 575 individuals over the period 1259 to 1293. Initially, he analysed the different types of relations and their frequency, which showed that about two-thirds of the relations were 'pledging' relations. These were relations in which one person gave a specific legal commitment in support of another person in relation to debt re-payments and other financial arrangements.

Smith's concern was with the role of kinship and other local ties in organising these relations and in structuring communal relations. Homans (1941) had previously undertaken a similar historical study of communal solidarity, but had not applied any social network concepts in his study. By contrast, Smith used the idea of the egocentric network as his principal orientating concept. The 425 Redgrave landholders of the year 1289 were divided into four categories according to the size of their landholdings, and equalsized random samples were drawn from each category. This gave 112 individuals for analysis, and their documented relations with all other people over the ten-year period from 1283 to 1292 were extracted from the database. The personal, ego-centric networks of the 112 people, taking account of the distance 1 relations, were then analysed in terms of their social bases and geographical spread. The distribution of the densities of the personal networks showed a curvilinear relation to landholding. Density increased steadily with size of landholding among those with four acres of land or less, and it decreased steadily with the size of holding for those with more than four acres. Those with three or four acres, therefore, had the densest personal networks, median density among these households being between 0.2 and 0.4. They were also the most involved in multiplex relations. It was, therefore, the middling landholders who were best integrated into their village community. In the light of the earlier discussion of the relation between network size and density, it is interesting to note that Smith discovered a correlation of just 0.012 between the two measures. He concluded, therefore, that the variations in network density which were observed were not a mere artefact of network size, but reflected real variations in the quality of interpersonal relations.

Taking a-.count of all his network data, Smith rejects the idea of a tightly knit organic community organized around kin and neighhours. The network structure of the medieval village, at least so far as Redgrave was concerned, was much looser than this image. Neither were distant kin an especially important source of social support: those individuals who interacted most frequently with near neighbours also interacted most frequently with kin, although probably on most occasions residing apart from them. These kin, however, tended to be close: siblings, uncles, nephews, nieces, fathers and mothers, sons and daughters. (Smith, 1979: 244)

Wellman recognized that the ego-centric networks that he studied in East York were linked into chains of connection through overlapping associations: there was, he held, a 'concatenation of networks' with personal networks being 'strands in the larger metropolitan web' (Wellman, 1979: 1227). But he does not directly investigate these global features of the socio-centric networks of East York. Some pointers to this 'concatenation' are provided in Grieco's (1987) extension of the work of White (1970) and Granovetter (1974). Grieco's research concerned the giving and receiving of information about job opportunities, and she showed that the flow of help from particular individuals to their network contacts produces an alteration in the global structure of the network. Where information is received indirectly, from contacts at a distance of 2 or more, there is a tendency for a new direct link, albeit a weak one, to be established between the originator of the information and those who received it (Grieco, 1987: 108ff.). The overall density of the network, therefore, increases, and some of these links may be solidified and strengthened through feelings of solidarity and obligation. Thus, some of the initial increase in density will persist.

When others in the network acquire the ability to reciprocate for the help that they have received they will, in turn, tend to create new direct links and a further alteration in the density of the network. In this way, changes at the individual level of ego-centric contacts result in a continual transformation of the density and the other socio-centric, global features of the network.