

# SOCIAL NETWORK ANALYSIS for EGO-NETS

Nick Crossley • Elisa Bellotti • Gemma Edwards  
Martin G Everett • Johan Koskinen • Mark Tranmer



Los Angeles | London | New Delhi  
Singapore | Washington DC | Boston

# 1

## INTRODUCTION

### Learning Outcomes

By the end of this chapter you will:

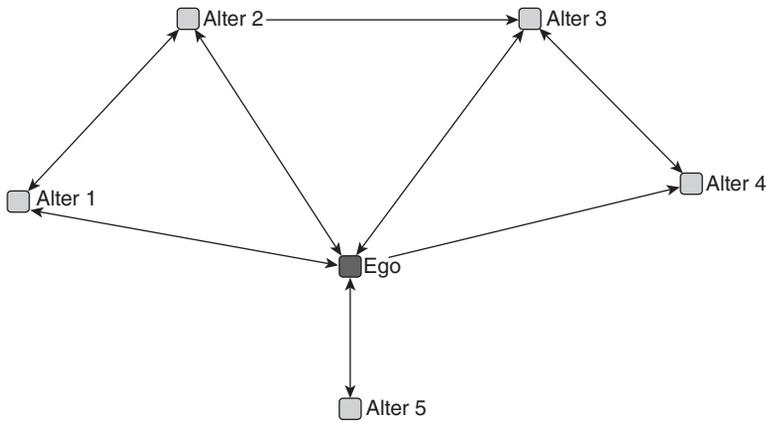
1. Know how 'network' is defined in social network analysis.
2. Be familiar with three different approaches to social network analysis: ego-net analysis, whole network analysis and two-mode analysis.
3. Know what is distinctive about ego-net analysis.
4. Understand the pros and cons of ego-net analysis, relative to whole network analysis, and where it is most appropriate to use each approach.
5. Understand some of the ways in which network data are stored and represented for purposes of network analysis, and also certain fundamental concepts and measures used by network analysts.
6. Be familiar with the basic plan for the book as a whole.

### Introduction

In this book we offer a comprehensive introduction to one of the most widely used forms of social network analysis (SNA): actor-centred or 'ego-net' analysis. An ego-net is the network which forms around a particular social actor, be that a human actor or a corporate actor, such as an economic firm or national government. In theory it involves all other actors (alters) with whom an ego enjoys a specific type or types of tie (e.g. emotional closeness, information sharing, economic exchange, etc.) and all relations

(of the same type or types) between those alters. Useful and important work can be conducted without information on ties between alters, however, and this aspect of the definition of an ego-net is therefore sometimes relaxed: an ego-net is then simply a list of alters with whom a target individual (ego) enjoys a particular type of relation.

Thus defined, ego-nets can be visualised, as in Figure 1.1, using coloured shapes ('vertices') to represent an ego and her alters (the nodes of the network) and connecting lines ('edges' or 'arcs') to represent ties between them. The 'ego' is coloured black in Figure 1.1 to distinguish her from her (grey) alters.



**Figure 1.1** Visualising an ego-net

Ego-net analysis is one of several approaches to SNA. Like each of the others and like any other research method, it has strengths and weaknesses and is more appropriate in some circumstances than others. Our decision to focus the book exclusively upon ego-net analysis is not an expression of preference on our part or an argument in favour of it over other forms. We have all used a variety of forms of SNA in the course of our work. Our decision to focus upon ego-net analysis here is based upon the observation that it tends to receive less coverage than other approaches in general texts on SNA, when we, as teachers of the range of SNA methods, find that many newcomers to the approach either wish to use an ego-net approach or probably should use it, given the nature of their research problem, and when a large number of papers published on networks, including many influential papers, use this approach. In short, we have written this book because there is no other book-length introduction to ego-net analysis and there should be.

Ego-net analysis is best understood in the context of a wider appreciation of SNA and of the concept and importance of social networks more

generally. We therefore begin this chapter with a brief review of the field and of the two key alternatives to ego-net analysis within SNA: whole network analysis and two-mode analysis. This will allow us to draw out the distinctiveness of ego-net analysis and its strengths and weaknesses, relative to the other approaches. Furthermore, it will allow us to explain when and where ego-net analysis is more (and less) appropriate as an approach compared to the other approaches. The chapter ends with a brief discussion of the plan for the rest of the book.

## Networks and Network Analysis

Connection is a constitutive fact of social life. A social world comprises not only a plurality of social actors, both human and corporate (e.g. firms or governments), but also interaction and enduring ties between those actors. Actors influence one another and exchange resources, becoming interdependent. They cooperate, compete, conflict, support and seduce one another. And these interactions and ties make a difference. For example, where ties cluster, generating a dense nexus of mutual influence, we often find greater homogeneity in attitudes, preferences and practices (Coleman 1988). To give another example, pathways of ties through networks provide channels for the diffusion of culture, resources, information and often viruses too. Finally, where specific patterns of ties give rise to trust and norms of cooperation ('social capital') this can facilitate forms of action, both individual and collective, that would not be possible in the absence of that particular configuration of ties – although this is usually at the cost of certain constraints (Coleman 1990). Networks are social structures which, as Durkheim (1964) said of social structures more generally, afford both opportunities and constraints for those entangled within them.

Some of the effects just mentioned can be generalised across a network. Everybody within the network is affected to a similar degree. Some apply to certain sub-groups within a network more than others, however, and some may apply specifically to particular actors, on account of the position they occupy within the network. This might be a matter of who they know, to invoke everyday wisdom, or, more generally, of the types of people they know. However, it may be a matter of network structure; where they fit within a pattern of relations: for example, which parts of a network they uniquely bridge (Burt 1992, 2005) or the pattern of connection in their immediate network neighbourhood.

These observations raise important methodological questions. How do we capture and analyse relational phenomena? With a certain amount of tweaking, which we discuss in Chapter Three, many of the standard methods of data gathering within social science can be used to generate relational,

network data. Nodes and their ties must be systematically surveyed but we can do that with a questionnaire, a structured or semi-structured interview, through direct observation (participant or non-participant), by trawling archives and texts, and perhaps by other means too. Furthermore, in the ‘information age’ and more especially the age of Web 2.0, a great deal of relational data is routinely generated in the course of everyday life, prompting some to ask if social scientists should not be taking more advantage of these sources too (Savage and Burrows 2007). Of course many social scientists are now taking advantage of them.

What we do with relational data when we have them, how we store and analyse them, poses more problems for conventional social scientific approaches, however. Relational data differ from the data usually analysed in social science and require dedicated techniques for their storage, representation and analysis. This is where SNA comes in. SNA is the collective label for a set of interconnected concepts, theories and techniques, developed for the most part within a relatively cohesive, interdisciplinary research ‘network’, devoted to the gathering and analysis of relational data (for a comprehensive introduction see Borgatti et al. 2013, Scott 2000 or Wasserman and Faust 1994).

SNA has a long history, stretching back to the 1930s (see Freeman 2006, Scott 2000) and its development has involved seminal contributions from sociologists, anthropologists, social psychologists, business analysts and increasingly also political scientists and economists. The distinctiveness of the approach owes at least as much to a wider interdisciplinary reach, into a branch of mathematics known as graph theory, however, and to collaboration between social scientists, mathematicians and increasingly also statisticians. It is not an exclusively quantitative approach and in this book we will stress the gains to be made from adopting a mixed method, qualitative and quantitative, approach to it (see also Bellotti 2014, Crossley 2010, Edwards 2010, Edwards and Crossley 2009). However, it is the interplay between social science and graph theory, in large part, which facilitates relational analysis and marks SNA out as a distinct research methodology.

## What Are Networks?

All networks comprise two essential elements:

- A set of nodes.
- A set or sets of ties.

Optionally, they may also include:

- A set of node attributes.

## Nodes

What counts as a node will vary between research projects and is at the discretion of the researcher. Anything might be defined as a node for purposes of SNA if it is meaningful to define it thus in the context of a particular study; that is, if a researcher has good reasons to want to regard it as a node, and if it is capable of the type of tie of interest to the researcher. Nodes might be: human individuals, chimpanzees, organisations, cities, nation-states, etc. Network analysis is a formal analytic approach, focused upon patterns of connection. It can be applied to any type of connection between any type of object. However, most analytic routines and algorithms assume that all nodes are, in principle, equally capable of engaging in the type of connection under consideration and this is therefore a constraint upon node choice. Each of the nodes in a friendship network must be capable of forming a friendship with any and every other, for example, at least in principle.

This doesn't mean that every node will be a friend with every other. That wouldn't be a very interesting network to analyse! Nor does it preclude the possibility that certain conditions might make friendship between some nodes more likely than others. Indeed, one of the questions we might be interested in is whether certain properties, either of the network or the nodes (e.g. beliefs or identities), affect the likelihood of connection between them. Such patterns and properties are only of interest, however, where we believe that, in principle, any node could form a tie (e.g. a friendship) with any other. It may be interesting if we find that members of one ethnic group less often form business ties with members of another ethnic group, for example, or if one ethnic group is found to be marginal in the network of a particular business community but only because we believe that, in principle, any member of the node set could form a tie with any of the others.

The relative absence of constraints upon node choice imposed by the theories and procedures of SNA does not mean that anything goes with regard to node selection. To reiterate our above point, nodes and node sets must be defined and selected carefully, with reference to the ideas and theories driving a particular research project. As in statistics, a network analysis is only as good as the data upon which it is based and it is the responsibility of the researcher to ensure that their data are meaningful and of a high quality. SNA packages will generate impressive visualisations and numerical arrays out of any old rubbish but it will still be rubbish. 'Garbage in' leads to 'garbage out' (the GIGO principle) and we must be careful to ensure that the nodes/node set that we select for analysis will allow us to answer the scientific questions that we have set for ourselves.

The question of which nodes to focus upon for a social network analysis is often a matter of where to draw the boundaries around a node set. Some networks are already bounded for us. If we are interested in friendship

patterns between children in a school or shop-floor workers in a factory, for example, then the boundaries of the formal organisation itself suggest obvious boundaries for our node set, and there will usually be a register of some sort that we can use, listing all members of that set. Many of the networks that we want to analyse have no neat boundary, however. When Saunders (2007) elected to survey the network of environmental organisations in London, for example, she confronted a range of problems. In particular she had to decide which of the organisations known to her counted as environmental organisations (there are plenty of obvious inclusions and exclusions but inevitably also a high number of more ambiguous cases) and she had to tackle the problem of accessing those which were not, at the start of the project, known to her. Many potential populations of interest have no clear and unequivocal criteria of inclusion and nothing approximating a membership list or register that we can draw upon to define them. To quote a well-known American statesman, they involve both known-unknowns and unknown-unknowns, and we have no option but to try to work around this. Such problems are not unique to SNA. They pose a problem for all types of social science research. But they are no less of a problem either.

## Ties

As with nodes, the formalism of SNA means that any type of tie can be focused upon, as long as all potential pairs of nodes are capable of entering into them and they are meaningful and appropriate to both the research questions being asked and the theories and conjectures which are driving them. If we are interested in the spread of sexually transmitted diseases, for example, then we need to know who engages in risky sexual practices (i.e. practices which facilitate disease transmission) with whom. Any other relation between the members of our node set is irrelevant because it does not facilitate transmission of a pathogen. Unless, that is, we want to track the diffusion of safe sex messages too, in which case we might also be interested in who talks to whom about intimate matters. If, by contrast, we want to predict the manner in which an economic crisis may cascade from one country to others then we need to know which countries, within the relevant set of countries, trade heavily with which of the others. And if we are interested in social capital and the potential for certain sorts of collective action within a community we may want to know the pattern and extent of relations of trust (or cooperation) between its various members. There is no type of tie which is correct for all research purposes. It always depends and is in many cases highly specific.

In some cases, of course, we may be interested in multiple types of relationship amongst the same population of nodes. Salient ties in many networks of interest are ‘multiplex’ (they have many strands, incorporating

multiple types of relation). Studying students in a college, for example, we might want to know who studies together, who socialises together between lectures, who socialises together in evenings and at weekends, and who lives together. We might expect some overlap between these relations and SNA affords various ways of exploring such overlaps, but each is a distinct type of relation and it is reasonable to expect that some pairs of nodes may be linked in one of these ways but not the others.

Similarly, ties may have different strengths and we may wish to record and take account of these in our analyses. This might be captured by using a Likert Scale on a questionnaire; for example, by asking respondents how much, on a scale of 1–5 (or whatever), they like each of the people whom they have nominated as friends, how well they know them or how often they see them, etc. Alternatively, it might be captured through observation. Ethnologists observing animal interaction in the wild, for example, will often count how often any two animals interact in a particular way, weighting their ties accordingly. Such detail is not always necessary or even helpful. Often it will suffice to ascertain whether two nodes enjoy a tie or not. But weighting is an option.

Finally, ties can be directed or undirected. We say that a tie is directed when it is meaningful to ask whether or not it is reciprocal. Liking is directed, for example, because knowing that John likes Jane does not tell us whether Jane likes John. She might but she might not. Living with someone, by contrast, is necessarily reciprocal and therefore ‘undirected’. If we know that John lives with Jane then we know that Jane lives with John, or rather we know that they live together.

## Node Attributes

Node attributes are not necessary to the definition of networks and play no role in many network analytic routines, even when they are known. However, they can be included and may be very important in some cases. We may wish to know whether nodes in a network disproportionately form ties with others who are similar to them in some respect, for example – an effect referred to as ‘homophily’. Alternatively, we may wish to know whether particular node attributes are correlated with certain network positions. Are men more central than women within a particular network, for example? Do particular ethnic groups disproportionately find themselves in a particular position? These are categorical node attributes but in other cases nodes might have ordinal or interval level properties. We might wish to determine whether income is correlated with popularity, for example, or whether individuals are disproportionately likely to form ties with others of the same or a similar age as themselves.

## Whole Networks

Beyond the choices we make about node and tie sets, SNA offers a range of possibilities about the way in which we capture and analyse networks. This book is focused upon one very specific way: ego-net analysis. Before we narrow down on ego-nets, however, it is important to introduce the two main alternatives within the SNA toolbox: whole network analysis and two-mode analysis.

When we analyse a whole network we identify a relevant population of nodes and, as far as possible, conduct a census survey of all members of that population, seeking to establish the existence or not of a relevant tie between each pair of nodes in that population. In a population of 20, for example, there are potentially 190 undirected ties or 380 directed ties (see Box 1.1 for an explanation of this) and whole net analysis requires that we know about the existence or not of each one of them.

### BOX 1.1

#### Calculating the Potential Number of Ties in a Network:

##### A Worked Example

- In a population of 20 nodes, assuming it is not meaningful to refer to a node's relationship with itself (it is meaningful in some cases but often not), each has a potential 19 ties (the figure is 20 if nodes can enjoy ties with themselves – 'reflexive ties').
- So the maximum potential number of ties in the network is  $20 \times 19 = 380$ .
- This calculation assumes that our network is directed, however. It treats node number 1's tie to node number 2 as distinct from node number 2's tie to node number 1. Each of the 20 nodes potentially 'sends' a tie to each of the 19 others ( $20 \times 19$ ) and each potentially 'receives' a tie from each of the 19 others.
- If our network is undirected this calculation is problematic because it counts each tie twice, giving us double the number of (undirected) ties in the network. We therefore halve our original answer:  $380/2 = 190$ .

This information is stored within an adjacency matrix (see Figure 1.2). This is a matrix whose first column and top row each list all of the nodes in the network, in the same order, with ties between nodes being indicated in the cell where the row of one meets the column of the other. In Figure 1.2, for example, there is a number 1 in the cell where Paul's row intersects with

Ed's column. That indicates that they have a tie. The 0 at the intersection of Ed's row and Jo's column, conversely, suggests that they have no tie. This is a basic, binary network. If our ties were weighted the numbers populating the cells would reflect the weighting. If Paul had rated his relationship with Ed as '5' on a Likert scale or we had observed that he telephoned Ed five times during the period covered by our survey then we would have put a '5' in the cell where his row intersects with Ed's column.

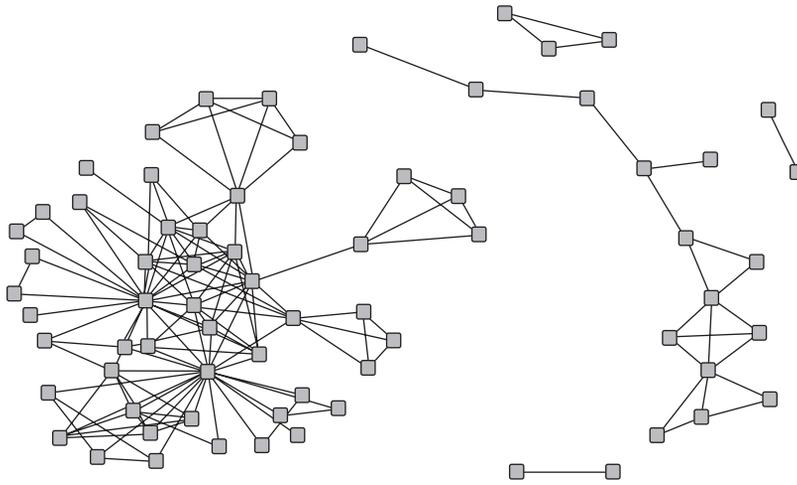
	John	Paul	Ed	Jo	Mick	Keith
John	1	1	0	1	1	0
Paul	1	1	1	1	0	1
Ed	0	1	1	0	0	1
Jo	1	1	0	1	0	0
Mick	1	0	0	0	1	1
Keith	0	1	1	0	1	1

**Figure 1.2** An adjacency matrix

Note that the diagonal running from the top left to the bottom right of Figure 1.2 comprises the cells where each node's row coincides with its column, potentially recording the node's relation with itself. As noted above, it is often meaningless to ask if a node has a relation with itself. This is reflected in the main software packages, such as UCINET, whose default setting for many analytic routines is to ignore the diagonal. The ties from an actor to their self ('reflexive ties') may be relevant, however, and can and should be included in computations where this is so. If our network involves ties of 'esteem', for example, then we may wish to measure and record each node's self-esteem, as well as their valuation of others, particularly if we believe that the former influences the latter or is affected by the opinions of others.

Note also that each pair of nodes ('dyad') in the network is represented twice in the matrix, once on either side of the diagonal. There is a cell where John's row meets Keith's column and a cell where Keith's row meets John's column. In the matrix for an undirected network each of the two cells will contain the same information, thereby giving an element of redundancy. The same tie will be recorded twice. For a directed network, however, this doubling up allows us to capture the direction of ties and any asymmetry in a relation. The intersection of John's row and Keith's column records whether John 'sends' a tie to Keith, whilst the intersection of Keith's row and John's column records whether Keith sends a tie to John. If a tie only exists in one direction, we can capture this.

A whole network can be visualised in a graph, in the manner shown in Figure 1.3. Nodes (also referred to as 'vertices' in this context) are represented by small grey squares. Ties are represented by lines which connect them



**Figure 1.3** A whole network

(also referred to as ‘edges’). If this was a directed network then the lines would have arrow-heads indicating direction (connecting lines in graphs of directed ties (‘digraphs’) are sometimes referred to as ‘arcs’) and if the ties were weighted their weighting might be represented either by giving edges differing thicknesses or by way of numerical labels at the side of each edge.

Similarly, node attributes might be represented in a graph. Categorical attributes can be represented by giving vertices different colours and shapes, for example, and ordinal or interval level attributes can be represented by varying the size of vertices. If Figure 1.3 was a network of trade relations between countries, for example, then we might indicate the continent to which each country belongs by way of a colour code, their system of government (e.g. democratic or not) by reference to different shapes, and their GDP by way of size – the bigger the GDP, the bigger the node.

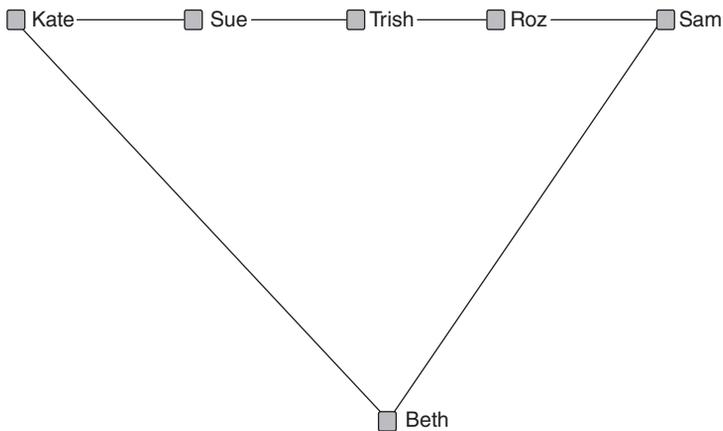
Graphs are a great way of representing network data and can be very useful. They can be misleading, however, especially if we try to read them as we might read a scatterplot, imputing vertical and horizontal axes to them and assigning significance to a node’s location along these axes. Nodes are often assigned a location in the graph space, by the main software packages, using algorithms which locate them close to others which have a similar profile of ties to them. There are different algorithms, however, based upon different principles. All only ever approximate a layout which operationalises their chosen principle, often with many ‘errors’. Analysts routinely change layouts, manually, either for aesthetic reasons or in order to better illustrate an observation that they have made regarding the network.

This is permissible because SNA operates with a different conception of space to the Cartesian conception employed in scatterplots. Network

space is defined exclusively by patterns of connection. A node's position in a network refers to its pattern of ties and bears no relation to its location (high or low, left or right) on the graph plot. Similarly, we commonly refer to the centrality of nodes (see below), deeming some more central than others, but again the various definitions of centrality that we work with all refer to patterns of connection rather than location on a graph plot. The least central node in a network may well be positioned towards the middle of a graph plot. Finally, 'distances', in network analysis, are measured in ties (or 'degrees') rather than centimetres or scales represented along graph axes.

### BOX 1.2

#### Paths and Geodesic Distance



- In the above network there are two paths connecting Kate and Sam.
- One path goes via Sue, Trish and Roz. It has a length of four degrees.
- The other path goes via Beth. It has a length of two degrees.
- The geodesic distance between Kate and Sam is the shortest path length between them; in this case, two degrees.
- Note that although Kate appears closer to Roz than to Sam on the plot she is closer to Sam in network terms because her geodesic distance from Sam is only two degrees, whereas her geodesic distance from Roz is three degrees.
- Note finally that Kate has two paths to Roz (one via Beth and Sam, the other via Sue and Trish). In this case they are both the same length: three degrees.

If two nodes are directly tied then they are at a distance of one degree. If they are not directly tied but each have a tie to a common, third node and are, in this respect, indirectly tied, then they are at a distance of two degrees. If their indirect connection involves two intermediaries, and therefore three ties, then they are connected by three degrees, and so on (see Box 1.2). These chains of connection are referred to as paths. Any two nodes may be connected by multiple paths but it is usually the shortest of these paths that we are interested in. The distance of the shortest path between any two nodes, measured in degrees, is referred to as the geodesic distance between these nodes. Geodesic distance will not usually correspond to the physical distance between nodes as represented on a graph (see Box 1.2).

Whole networks have a large number of properties, which can be defined at various levels. It would be useful to briefly outline these levels and introduce one or two properties at each level.

## The Whole Network Level

These are properties which exist at the level of the whole network. There are many of them. Simple examples include: order, which is the number of nodes in the network; and density, which is the number of ties in the network expressed as a proportion of the number of ties there could be, given the number of nodes. There are six nodes in the network in Box 1.2, for example, and six ties. To work out the density of this network we would calculate the number of ties that there could be, using the method explained in Box 1.1. Assuming that ties are undirected and that it makes no sense to ask if a node has a ('reflexive') tie to itself, that gives us  $(6 \times 5) / 2 = 15$ . We then express the six ties that we have found to exist as a proportion of the 15 that could exist: i.e.  $6 / 15 = 0.4$ . Our network has a density of 0.4. Note that density always varies between 0 (no ties in the network) and 1 (every possible tie is present).

Another whole network property is number of components. A component is a subset of nodes, each of which has a path connecting it to each of the others. There are five components in Figure 1.3, for example: a big one to the left of the plot, a long stringy one to the right, two dumbbell shaped dyadic nodes and a triangular shaped triadic component. We discuss components further below.

## Endogenously Defined Sub-Groups

A network's node set can often be divided into various subsets on the basis of patterns of ties. Components are subsets, for example, each distinguished by the paths connecting their constituent members. And they may be important. We would expect diffusion, contagion and/or cascades to happen within components, for example, depending upon the type of tie we are

looking at, but not across components because distinct components are not connected to one another. Similarly, we would not expect all members of a network to become involved in collective action if that network involved distinct components because the lack of connection between components would prevent coordination between them. We would only expect concerted action within components.

Another example of an endogenous sub-group is a clique. This is a subset of three or more nodes, each member of which is connected to every other. The density of a clique is always 1 because all possible ties are actualised. Cliques are important because their membership is highly cohesive, making the diffusion of information within them very quick and the potential for collective action, where triggered by an external event, much greater.

Components and cliques are defined by their cohesion. Members are more connected to one another than to others outside of the group. Not all sub-groups are defined by their cohesion, however. Sub-groups might be defined where their members occupy equivalent positions in a network, irrespective of their cohesion, and SNA offers a number of distinct definitions of equivalence. The most straightforward is structural equivalence. Two nodes are structurally equivalent if they have ties to the same alters, irrespective of whether they are tied to one another. Nodes that have no ties within a network ('isolates') are a special case of this and nicely illustrate the distinction between structurally equivalent sub-groups and cohesive sub-groups. All isolates within a network are structurally equivalent to one another because they have exactly the same pattern of ties (i.e. no ties at all). They are clearly not a cohesive group, however, because they have no ties to one another. A whole branch of SNA, referred to as blockmodelling, is devoted to modelling networks on the basis of such equivalently positioned groups ('blocks').

Finally, a very popular form of sub-group analysis focuses upon the often observed division within networks between 'core' nodes, which are all relatively well-connected to one another and apparently dominant, and more peripheral nodes with a greater density of ties to the core than to one another but only a low density of ties in both cases. Core-periphery analysis might assume a categorical form, in which case we seek to partition our nodes into two categories: core and periphery. Alternatively it may be continuous, in which case we calculate a 'coreness' score for each node. The core-periphery divide is important because the core is often the dominant sub-group within a network.

## Exogenously Defined Sub-Groups

Sub-groups may also be defined by factors external to the network, especially node attributes. All of the women in a network involving both men and women constitute an exogenously defined sub-group, for example, as do

different ethnic groups in a multi-ethnic network. Analysis of exogenously defined sub-groups typically centres upon the effect (or lack of effect) of node attributes on patterns of connection, or, alternatively, where attributes are potentially changeable (e.g. political identity or musical taste), it centres on the effect of connection upon attributes: e.g. processes of social influence, which persuade actors to change their behaviour or tastes.

Many networks, for example, are characterised by homophily; that is, nodes are disproportionately linked to other nodes with whom they share a salient social status ('status homophily') or an interest, taste or value ('value homophily') (Lazarsfeld and Merton 1964). This suggests either that people like to form ties with others similar to themselves or that they are influenced by others to whom they are connected (or both). Having established the existence of homophily in a network, our analysis might turn to exploring which is the more likely possibility. Alternatively, we might be interested to see whether particular node attributes (e.g. ethnicity) are associated with particular positions in the network: e.g. membership of the core or a given structurally equivalent block.

## The Node Level

In addition to their exogenously defined attributes, nodes have properties in virtue of their pattern of connection. In particular they can be more or less central to the network, as defined by one or more of the very different types of centrality identified in SNA. At a very basic level, for example, nodes vary in their number of ties within the network; that is, their 'degree centrality'. Some inevitably have more ties than others; that is, a higher degree or degree centrality. They are more degree central. Degree is only one measure of centrality, however. There are many others, including closeness, betweenness and eigenvector centrality. These other forms are explained in texts devoted to whole network analysis (e.g. Borgatti et al. 2013, Scott 2000 or Wasserman and Faust 1994).

## Dyads and Triads

Recent advances in the statistic modelling of networks have focused upon dyads and more especially triads as units of analysis. For example, early statistical approaches focused upon issues of reciprocity. They hypothesised that, in certain types of directed networks, involving certain types of tie, a node was more likely to 'send a tie' to an alter if the alter sent a tie to them. This hypothesis was tested (and usually confirmed) by looking at the number of reciprocated ties in a network and comparing it against the number one would expect by chance, given a particular density of ties.

## BOX 1.3

## Some Key Concepts

**Clique:** A subset of nodes all of whose members are directly tied to the others, giving them a maximum density (i.e. 1).

**Component:** A subset of nodes, all of whose members are linked by a path.

**Degree Centrality:** There are many ways of comparing nodes' levels of centrality within a network. The most straightforward – degree centrality – is to compare their respective numbers of ties.

**Density:** The number of ties in a network expressed as a proportion of the total number that are possible.

**Path:** A chain of connections and intermediaries linking two nodes in a network, such that information, viruses and other such things can pass between them.

**Reciprocity:** In a directed network any node A might 'send' a tie to a node B without necessarily receiving a tie back (A might like B (a tie of liking) whilst B does not like A or perhaps does not even know who they are). Reciprocity refers to a situation within a pair of nodes where each does send a tie to the other. It was of interest in early statistical approaches to network analysis as it was hypothesised that a tie from A to B is more likely when B sends a tie to A. Obviously certain types of tie are unlikely to be reciprocated. If B bullies A (a tie of bullying) it is unlikely that A will also bully B.

**Status Homophily:** A tendency within a network for nodes to be disproportionately tied to others who share one or more salient statuses with them (e.g. gender, ethnicity or age).

**Structural Equivalence:** Any two or more nodes are perfectly structurally equivalent where they have exactly the same pattern of ties: i.e. they are tied to exactly the same alters in exactly the same way.

**Transitivity:** The idea of transitivity suggests that any two nodes are more likely to enjoy a tie if each is tied to a common, third party. If Paul and Pete each have a tie to Frank, for example, then the idea of transitivity suggests that they are more likely to have a tie to one another compared to a situation in which they have no friends in common.

**Value Homophily:** A tendency within a network for nodes to be disproportionately tied to others with whom they share particular values and/or tastes.

Turning next to triads, statisticians were keen to test the thesis of transitivity associated with the work of Mark Granovetter (1973, 1982); that is, the claim that two nodes are more likely to have a tie when they each have a tie

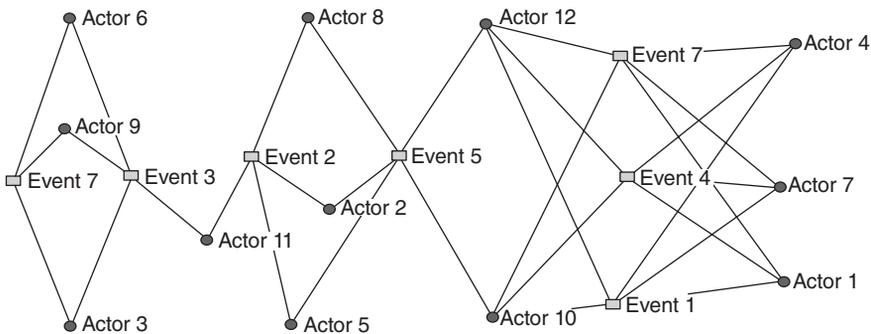
to a common third party. This can be tested by comparing the actual number of transitive triads in a network against the number expected by chance, controlling for both density and reciprocity. Complex statistical methods of modelling networks (Exponential Random Graph Models or ERGMs) have been devised from these relatively simple beginnings in recent years (Lusher et al. 2013).

We have barely scratched at the surface of whole network analysis here but we have hopefully said enough for scene-setting purposes (for a comprehensive introduction see Borgatti et al. 2013, Scott 2000 or Wasserman and Faust 1994). With this said we will briefly turn to two-mode networks.

## Two-Mode Networks

In addition to whole networks, network analysts sometimes analyse two-mode networks. In a two-mode network we have two different types of node and the type of tie that we are interested in exists only across these two types, not within them. A common example is a network of people (first mode) and events (second mode), with ‘attendance’ as the observed tie. People are tied to events where they attend those events but they are not, at least in the first case, tied to one another by a relation of attendance (people attend events but they do not attend one another) nor are events tied in this way (events do not attend events).

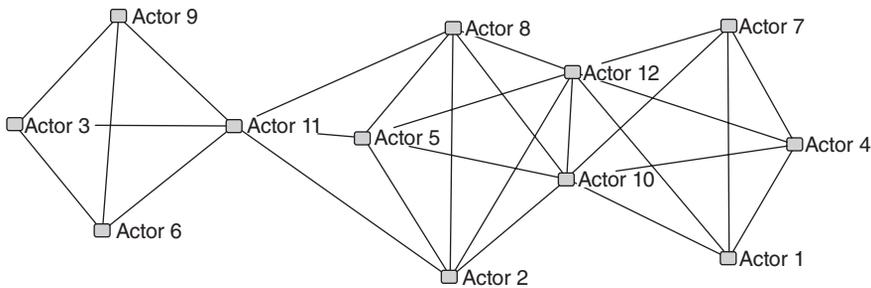
Two-mode networks can be captured in matrices, like single-mode networks, with one mode (e.g. people) represented along the rows and the other (e.g. events) represented down the columns. These matrices are referred to as incidence matrices. Similarly, two-mode networks can be represented as graphs, as in Figure 1.4, where events are represented by grey squares and participants by black circles.



**Figure 1.4** A two-mode network

Many of the whole network measures introduced above have two-mode equivalents, making it possible for a two-mode network to be analysed in its basic form. It is very common, however, to ‘affiliate’ two-mode networks, deriving two single-mode networks from them which can then each be analysed in the normal way. A people and events network, for example, might be affiliated to give us a network of people (linked where they attend the same events) and a separate network of events (linked where they are attended by the same people). Figure 1.5 visualises the single-mode participant-to-participant network which can be derived from Figure 1.4. In this case participants are linked to one another where they have attended at least one of the same events.

The data that we get when we affiliate two-mode data are weighted because participants might attend more than one of the same events. We might analyse the affiliated network in this weighted form. Alternatively, however, we might dichotomise it (simplifying it by deeming ties simply absent or present) on the basis of a threshold value. For example, we might decide to call two events connected when they share three or more of the same participants.



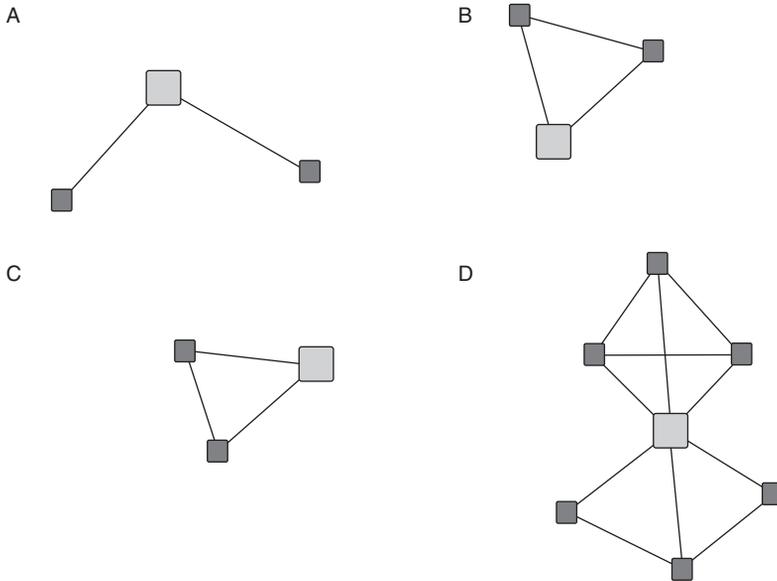
**Figure 1.5** A single-mode network of participants derived from Figure 1.4

In historical research, using archival sources, it is often impossible to get whole, single-mode data and we may have to use an affiliated two-mode data source as a proxy. Indeed, there may be many reasons why we resort to two-mode data gathering as a means of deriving a single-mode network. If the resulting network is meaningful and fits with the theories driving the research this is often fine. Researchers should be aware, however, both that this method of data gathering tends to shape the resulting network in a number of ways and that the move from two to one mode (affiliation) involves a loss of information (Everett and Borgatti 2013). This latter problem may be avoided, in relation to some analytic routines, however, if both modes are analysed and the results of these two analyses recombined (*ibid.*)

## Ego-Nets

Having briefly introduced whole and two-mode networks we can now turn to the main focus of this book: ego-nets. As explained at the beginning of this chapter, an ego-net is the network of contacts (alters) that form around a particular node (ego). Ego herself is sometimes removed for analytic purposes. That varies. Similarly, whilst it is often preferable to have data regarding (relevant) ties and the absence of such ties between alters, and whilst much of what we discuss in the book assumes access to such information, ego-net analysis may, in some cases, focus simply upon ego's ties, bracketing the question of ties between alters. For present purposes, however, we will assume that an ego-net involves ego, her alters and all relevant ties between alters.

Ego-nets can be abstracted from whole networks. Each node in a whole network is or has an ego-net. Each has, potentially, a number of alters, and those alters are either connected to one another or not. Thus, in Figure 1.6 we have abstracted four ego-nets from the whole net represented in Figure 1.3 (the egos are the slightly larger, grey vertices).



**Figure 1.6** Four ego-nets (extracted from Figure 1.3)

Note that each ego-net varies in size. A, B and C each have 2 alters but D has 6. If our egos ever needed extra muscle for a job they were contemplating then D would be better placed to get it, all other things being equal, than the others. On the other hand, she probably has many more people

asking her for favours and making demands upon her time, which may not always be a good thing.

The structures of the nets also vary. Although A, B and C are each of the same size, B and C are each cliques, according to SNA's technical definition, whereas A is not. A clique, it will be remembered, is a subset of nodes, each of whom enjoys a tie with each of the others. This is true of B and C, but A's alters are not tied to one another. This will reduce the likelihood of solidarity and consensus in A and may make coordination more difficult. It may also sometimes work to A's advantage, however, as she controls the flow of information and resources between her two alters (at least as far as we can tell (see below)) and she may benefit from that. She may take credit for the good ideas of one alter when passing them on to the other, for example, and/or may be rewarded in kind by the recipient when passing on resources from one alter to another.

Furthermore, the independence of her alters from one another means that they are more likely to provide access to different flows of information, which is an advantage. In a very famous paper on the information flows involved in securing a new job, Mark Granovetter (1973, 1982) observed that transitive ties (which were discussed above) are often 'redundant' in informational terms because ego's alters will tend to give her the same information. Whilst transitive ties are more conducive to the development of trust, cooperation, consensus and solidarity (Coleman 1990), which can be an advantage (see Chapter Two), the closure of contacts precludes access to external nodes and thus external sources of ideas and information, resulting in constant regurgitation of the same ideas and information and, potentially thereby, stagnancy (Burt 2005). For this reason they can be less useful to those involved in them. New information is much more likely to come from intransitive ties; ties to alters who are not connected to the ego's other alters.

Note that ego-net D, which is bigger than either A, B or C, combines elements of each of their respective structures. Ego D has access to two independent 'pools' of information, like A. D occupies a similar 'brokerage' position, albeit mediating between two groups of alters rather than just two alters. One of these two groups (located above her on the plot) is a clique, like ego-nets B and C. The other is one tie short of a clique. Perhaps ego D will enjoy both the benefits of transitivity and brokerage? She enjoys solidarity and trust with each of her two clusters of alters and also both the opportunity to broker between them (with the benefits that brings) and access to two distinct pools of information and other resources. However, her position may create its own constraints. Her two clusters may compete for her loyalty, for example, pressuring her to take sides, and they may make competing demands upon her time, energy and other resources (see, for example, Crossley 2008b).

## The Pros and Cons of Ego-Analysis

It is always possible to extract ego-nets from a whole network. Where this is done the ego-nets can, of course, be recombined into a whole network, allowing the researcher to move between the level of the whole and the level of individual ego-nets. However, it is not always possible and not always desirable to gather whole network data. Researchers often elicit ego-net data in ways which do not allow us to put the whole back together (although it is always there tacitly). There are three good reasons to do this.

Firstly, ego-net analysis affords a means of analysing big networks. If we are interested in a relatively small population of actors, such as participants in a local music scene, protest group or pupils in a school, then it is feasible for us to conduct a census survey of our node set and we can therefore do whole network analysis. If we are interested in processes affecting bigger populations, such as a whole town or 'the general public', however, then a census survey will not be possible in most cases, ruling out whole network analysis. Generally we are constrained to sample large populations, denying ourselves access to the information required for a whole network analysis. Ego-net analysis is entirely compatible with a sample survey, however. Indeed it will ordinarily involve a sample survey, and as such it is possible to use it in relation to much bigger populations.

It is important to add here both that randomisation and other strategies employed in sample surveys are entirely compatible with an ego-net approach and that questions which elicit ego-net data, of the type discussed in Chapter Three, can be added to any standard questionnaire. Indeed, it is becoming increasingly common for a small number of ego-net questions to be added to the various regular large national surveys conducted in many countries. Ego-net questions inevitably add bulk to a questionnaire and for this reason their inclusion has to be given careful consideration, but a small ego-net module will add no more bulk than any other module and may prove very enlightening.

Secondly, because ego-net analysis is compatible with the range of sampling strategies routinely used in (quantitative) social science it is also compatible with most of the techniques of statistical analysis and modelling employed in such research. Whole network data contradict the assumptions of standard statistical approaches and, for this reason, can only be analysed, statistically, by means of a range of specially adapted methods (Borgatti et al. 2013, Lusher et al. 2013). Most obviously, for example, cases (nodes) in a whole network survey are not randomly sampled from a wider population (as is assumed in inferential statistics) and the connections between them contradict the assumption of case-wise independence. There are no such problems with ego-net data, however, at least where appropriate sampling techniques have been employed. We should add here, furthermore,

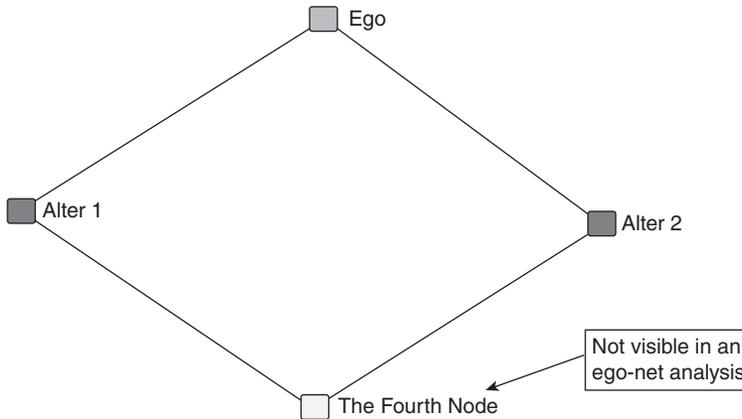
that ego-net analysis generates a range of measures, discussed in Chapter Four, which may be included, alongside other, more conventional measures, in such research.

This will be particularly salient where networks are one amongst a number of foci within a research project, each of which must be accommodated within a single research design. Whole network analysis makes very specific demands upon the researcher and allows little room for compromise. As such it cannot easily be added to a project which has a remit beyond networks. Ego-net analysis, by contrast, is often quite easy to slot into a more conventionally structured project.

A final advantage of ego-net research relates to what Simmel (1955) calls 'intersecting social circles' and what White (2008) calls network domains or 'net doms'. Both writers observe that in modern societies most people interact and form ties across a number of distinct 'social circles' or 'domains' whose membership, with the exception of ego herself, does not overlap. For example, the typical adult may have alters in their family, neighbourhood, workplace, gym and local pub. They are a point of intersection between these different circles but they are most likely the only point of intersection in many cases. Their gym buddies will probably know one another but won't know anybody else in ego's network and likewise for members of each of the other circles. These patterns of separation and intersection, which are essential to a proper understanding of the networked character of human social life, are much easier to get at by means of an ego-net survey. If we are interested in one domain, such as the gym, then a whole network approach may make sense because we have a relatively contained population (everybody who goes to the gym) (Crossley 2008b). If we are interested in many domains, however, some of which only overlap through a single node, then the size and complexity of the task at hand, coupled with the demands of bounding our object of study, will often rule out a whole network study. We need to ask individual egos about the different social circles in which they mix and the different sets of alters with whom they enjoy ties in each of those circles.

Note here that focusing upon a single domain, as we typically do in a whole network analysis, may result in a distorted picture of the social world because it separates that domain from others. Studies of social influence within the whole network of a school or a gym population, for example, may miss important actors from outside of that domain who influence those within it. Ego's activity in the gym may be affected by her relations to alters in a different domain, such as her family or workplace. Similarly, nodes who appear isolated may only be so in the one domain observed and only because they connect more strongly to alters in other domains not captured in the node set of a whole network study. By moving outwards from the individual and allowing us to tap into each of the various circles in which they mix, ego-net analysis helps to circumvent this potential problem.

Having said this, there is a comparable problem with ego-net analysis, which is avoided in whole network analysis. To give an example: when discussing the ego-nets in Figure 1.6 we referred to A's brokerage position and the advantages this creates for her. However, what if her two alters each have a tie to a common, fourth node who has no direct tie to A and who would not show up, therefore, in an ego-net study (see Figure 1.7)? Each of A's alters now has an alternative source of information and other resources, which reduces A's bargaining power. In Figure 1.7, A's two alters are each in a position to play a brokerage role too, mediating between A and this fourth node. Indeed, each node in this mini network is in exactly the same position vis-à-vis the others: each connects two otherwise unconnected nodes and is indirectly connected to one further node by each of their alters. Exactly how this will play out is not clear, but what matters for our purposes is that where our ego-net analysis (Figure 1.6) suggested that A was in an advantageous position, compared to her alters, a fuller analysis suggests that she is not. She is in the same position as each of her alters. And that is only how she appears on the basis of this snapshot. If we were to add further nodes then the picture may change again. Additional information, beyond that regarding A's ego-net, changes the picture that we derive from A's ego-net. This is only one example of the kinds of complications and qualifications (to an ego-net analysis) that might arise if we have access to what is, in this respect at least, the fuller structural information that we derive from a whole network study.



**Figure 1.7** The fourth node

Beyond this we may be interested in the properties of the whole network and/or its broader sub-groupings. Some of these properties could be inferred from a sample of ego-nets. If we are confident that the number of alters ('degree') is normally distributed across a population, for example, and that our

sample of ego-nets is truly random, then we may infer the average degree, and by means of this, the density of a whole network from our sample. Furthermore, again on condition of assumptions regarding distribution and randomness, the density of individual ego-nets may be used to infer likely levels of clustering in the wider network. However, our assumptions may not hold. A body of literature has emerged in recent years pointing to the existence of a class of networks whose degree manifests a so-called 'power-law distribution'. In lay terms this means that a very small proportion of nodes in a network have a huge number of alters, whilst the vast majority have only a small number. Where this occurs the network has a very distinctive, centralised, structure but one that will probably elude a random sample survey because random sampling is unlikely to pick up the tiny minority of 'hub' nodes in the population. Again this is just one example, amongst many, of whole network level properties that are difficult and/or impossible to get at by means of a sample of ego-nets.

## Quantity and Quality

The discussion in this chapter hitherto has been largely framed in quantitative terms. We have spoken of random samples, measures and models. Network analysis, in all of its forms, is amenable to a mixing of methods, quantitative and qualitative, however, and we intend to reflect this in this book. Matrices are a good means of capturing who knows whom. They allow us to distinguish between types of ties, different strengths of connection and sometimes also between positive and negative ties. But we may wish to know more about the meaning of specific alters for ego, what they do together and the 'story', as White (2008) calls it, of their relationship. We may wish to embellish our data regarding ego's network with wider qualitative information about their life and outlook, gleaned through qualitative interviews, and/or to embed our ego-net data within an ethnographic understanding of its context. Indeed, much early and pioneering research on SNA emerged out of an ethnographic context, with graph theoretic methods being used to build upon, organise and systematise qualitative-observational data, and the authors of that work often included both quantitatively and qualitatively defined properties in their concept of a 'network' (e.g. Mitchell 1969).

This is significant for the present authors because several of these pioneers were members of the 'Manchester School' – a research cluster (to use the contemporary jargon) based in our own institution. Indeed, our own research centre, the Mitchell Centre, is named after one of these pioneers, and he, in turn, was the PhD supervisor of one of us (Martin Everett). Our vision of SNA, both in this book and more generally, reflects the pragmatic and mixed approach of the original Manchester School, combining methods

where it makes sense to do so and moving freely between qualitative and quantitative data.

## Conclusion and Chapter Plan

In this chapter we have introduced the idea of ego-net analysis, comparing it with analysis of whole networks and considering both where it is most appropriate and what its respective strengths and weaknesses are. We have also introduced the vocabulary of SNA and a number of concepts and measures which will be revisited in greater detail in later chapters. It only remains for this chapter to briefly map out the content of these chapters.

**Chapter Two – Social Capital and Small Worlds: A Primer:** Many recent developments in ego-net analysis have originated in the context of debates in two central substantive areas of research: social capital and small worlds. In order to facilitate proper understanding of these developments we use Chapter Two to briefly introduce these two areas of substantive research, explaining where and why they connect to innovations in ego-net research.

**Chapter Three – Getting Ego-Nets:** Here we consider a number of the most common ways of gathering ego-net data.

**Chapter Four – Analysing Ego-Net Data:** We discuss all of the main measures of ego-net properties typically used by network analysts, explaining where and why they should be used.

**Chapter Five – Narratives, Typologies and Case Studies:** This chapter introduces qualitative approaches to ego-net analysis and discusses the advantages of adopting a mixed-method approach.

**Chapter Six – Multilevel Models for Cross-Sectional Ego-Nets:** In this chapter and also Chapter Seven, we build on the discussion of ego-net measures in Chapter Four but also slightly shift our focus and gear, by considering a number of recent statistical developments in ego-net analysis. In Chapter Six specifically we focus in particular upon the way in which ego-net data might be used in the context of multilevel modelling.

**Chapter Seven – Statistical Analysis of Network Dynamics:** Sticking with a more advanced statistical approach, this chapter considers methods for modelling change within ego-nets across time.