6.99A

SOCIAL NETWORK ANALYSIS

{author}
{affiliation}

Keywords

Social networks, structuralism, cohesion, brokerage, stratification, network analysis, methods, graph theory, statistical models

Contents

Summary

Social network analysis (SNA) focuses on the structure of ties within a set of social actors, e.g., persons, groups, organizations, and nations, or the products of human activity or cognition such as web sites, semantic concepts, and so on. It is linked to structuralism in sociology stressing the significance of relations among social actors to their behaviour, opinions, and attitudes. Social network analysis is felt to be appropriate for analyzing social cohesion, brokerage and exchange, as well as social ranking within or among social groups.
Two perspectives dominate SNA: the socio-centred and ego-centred perspective. The socio-centred perspective analyses overall network structure. It looks for patterns of ties that indicate cohesive social groups, central actors that may be paramount to the integration of the social network, and asymmetries that may reflect social prestige or social stratification. Recent advances are found primarily in the technique of blockmodelling. The ego-centred perspective focuses on the composition of local network structure. Do actors influence one another through their network ties (social influence model) and/or do actors adjust their ties to the characteristics of their peers and to their ties with them (social selection model)? Recent advances in this area include new types of statistical models.
The development and interest in SNA has increased sharply over the last few decades due to the application of mathematics – notably graph theory and statistical models – and the wide availability of software for network analysis both commercial and freely available through the internet. In addition to the formal, quantitative approach to social network analysis, a qualitative approach to social networks is developing.

1 Introduction

Social network analysis (SNA) focuses on the structure of ties within a set of social actors, e.g., persons, groups, organizations, and nations, or the products of human activity or cognition such as web sites, semantic concepts, and so on. Any social process or system that can be conceptualized as a set of units and a set of lines connecting pairs of units can be studied as a social network. Examples of social structures that have been studied as networks are friendship among children in a school, family relations among members of a social elite, shared board members of corporations, trade relations between countries, and hyperlinks between websites.

This is not to say that it is always useful or necessary to apply social network analysis to data that can be conceptualized as networks. For instance, if a researcher is just interested in knowing the number of people that a person can turn to for help, the number of ties instead of the structure of ties is relevant and network analysis is not needed. For network analysis to be applicable, theory from sociology or other social and behavioural sciences should give reasons to believe that the structure of ties is linked to behaviour, opinions, or social position of the members of the network. Three types of sociological concepts appear repeatedly in most applications of social network analysis: cohesion or solidarity, brokerage or influence, and ranking, prestige or status. They are discussed in Sections 4, 5, and 6 respectively.

2 Definition of a network

There are several ways of formally defining a network, depending on the branch of mathematics used. The most usual and flexible definition is derived from graph theory, which conceptualizes a social network is conceptualized as a graph, that is, a set of vertices (or nodes, units, points) representing social actors or objects and a set of lines representing one or more social relations among them.

A network, however, is more than a graph because it contains additional information on the vertices and lines. Characteristics of the social actors, for instance a person's sex, age, or income, are represented by discrete or continuous attributes of the vertices in the network, and the intensity, frequency, valence, or type of social relation are represented by line weights, line values, line signs, or line type.

Formally, a network **N** can be defined as $\mathbf{N} = (U, L, F_U, F_L)$ containing a graph $\mathbf{G} = (U, L)$, which is an ordered pair of a unit or vertex set U and a line set L, extended with a function F_U specifying a vector of properties of the units ($f : U \rightarrow X$) and a function F_L specifying a vector of properties of the lines ($f : L \rightarrow Y$). The set of lines L may be regarded as the union of a set of undirected edges E and a set of directed arcs A ($L = E \cup A$). Each element e of E (each edge) is an unordered pair of units u and v (vertices) from U, that is, $e(u{:}\ v)$, and each element a of A (each arc) is an ordered pair of units u and v (vertices) from U, that is, $a(u{:}\ v)$.

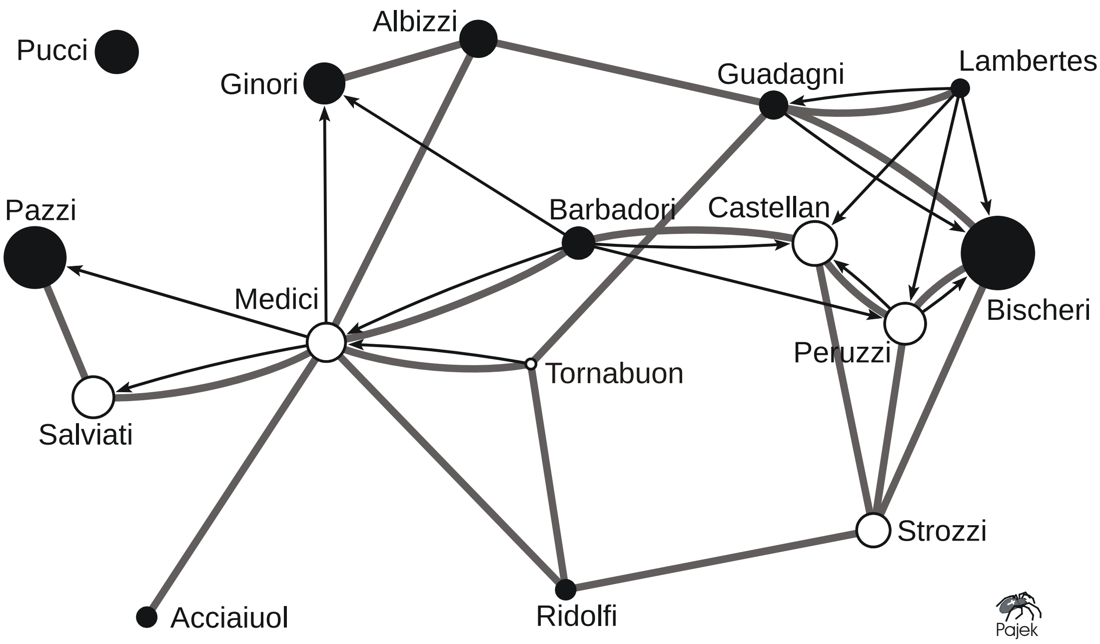

Figure 1 - Sociogram of marriage (grey edges) and business (black arcs) relations between 16 Florentine families (circa 1400 AD).

Figure 1 illustrates the graph theoretical definition of a network. It shows a sociogram of marriage and business relations among 16 families in Florence (Italy) around 1400 AD. The data are part of a larger dataset collected and analyzed by John F. Padgett. In a sociogram, vertices are represented by circles, so each family is represented by a circle. Lines, connecting two circles in the sociogram, can be of two types: directed or undirected. Here, directed lines represent business relations among families. They are drawn as black arcs in Figure 1, pointing towards the more prosperous family. Marriage relations between families are undirected lines or edges in the network. They are represented by grey lines in the sociogram.

As stated before, a network is more than a graph, it contains additional information on the vertices or the lines. In the sociogram, the family names link the vertices to historical people. This information goes beyond the pure structure of the network, which is represented by the graph. In addition, the sociogram shows two properties of the families: (1) whether the family had seats in the local civic council in a previous period (AD 1282-1344) and (2) the net wealth of the family in 1427 (measured in thousands of lira). Vertex colour shows the first property: black vertices represent families that used to be members of the civic counsel. Vertex size presents the wealth of the family: the bigger the circle, the more affluent the family in 1427. Finally, if information had been available on the number of marriages or business transactions between families, the network would also have included additional information on the lines. Then the width or colour of the line would be used to visualize properties of lines.

The network of Florentine families contains two different relations. As a consequence, two families are sometimes connected by two lines: both a marriage edge and a business arc. Loops do not occur, although it is not impossible that members of the same family trade among themselves. In graph theory, a simple graph is a graph without multiple lines and, in the case of an undirected graph, without loops. A simple graph can easily be stored in a matrix, which is called an adjacency matrix or sociomatrix. So if we split the two relations in this network, we would obtain two simple graphs that can be represented by sociomatrices (Figure 2).

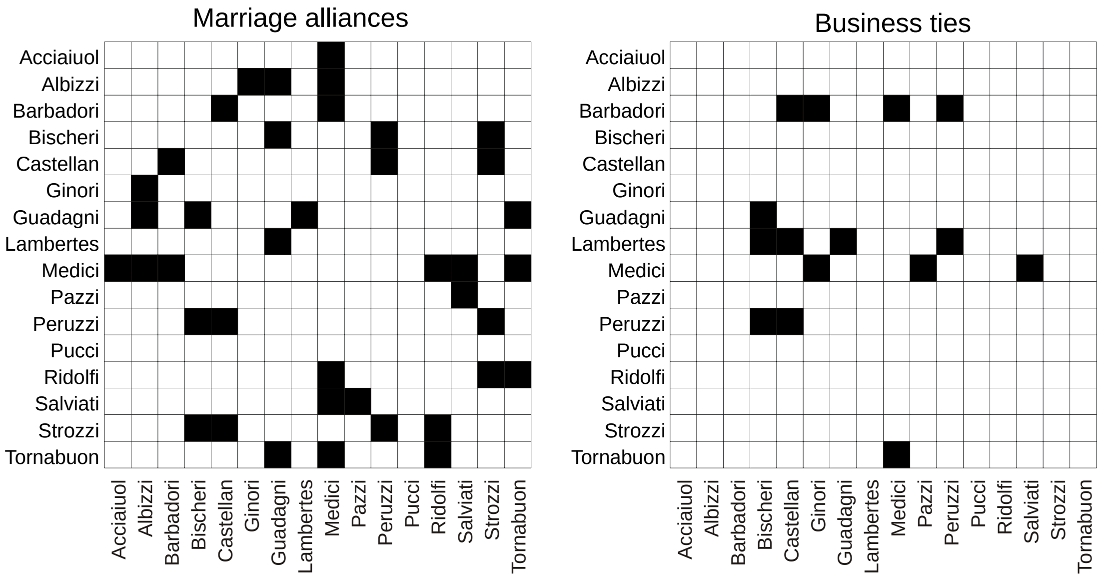

Figure 2 - Sociomatrices: marriage ties and business ties.

In the sociomatrices of the Florentine families network, the families define both the rows and columns. By convention, the entity in the row is the sender or tail of an arc and the entity in the column is the receiver or head of the arc. In these matrices, a black cell indicates the presence of a line. Black cells indicate the presence of lines, white cells the absence. Presence and absence re usually coded as 0 and 1 respectively, so the matrix may also contain numbers. We can see that the Acciaiuol family (1^{st} row) is married to the Medici family (9^{th} column) because the corresponding cell is black. Because the marriage relation is undirected, the latter family is also married to the former family, so the cell in the 9^{th} row and 1^{st} column is also black. The sociomatrix of an undirected relation (edges) is symmetric with respect to the diagonal running from the top left to the bottom right of the matrix. This is not the case with an asymmetric relation (arcs) such as the business ties.

In the example network, every family can be related to every other family by marriage or business ties. This is called a 1-mode network. There are many social relations, however, for which this is not the case. For instance, if persons are the vertices in the network, a marriage relation can only connect a man and a woman in many cultures. Then, we can split the vertices into two groups (men and women) such that ties can only exist between the two groups, not within one group. These groups are called modes and a network with two modes is called a bipartite or 2-mode network. A 2-mode network without multiple lines can be represented by a rectangular matrix in which the vertices of the first mode are found in the rows and the columns contain the vertices of the second mode. A very popular example of a 2-mode network is a network of affiliations between people and organizations, e.g., people (first mode) who are board members of large corporations (second mode).

3 Two perspectives: overall and local network structure

The conceptualization of social systems as graphs and networks offers the opportunity for systematic investigation and theorizing on the structure of ties among social actors beyond the pair. Whereas classical sociology tended to make a quantum leap from the individual and the pair to the triple, group, or society, e.g., Georg Simmel, graph theory offers the tools to formally describe and visualize social structure consisting of three and more actors. This has led to a new awareness of social structure as a system of ties that is both the product and the context and condition for human action.

Scientists that regard social structure as the product of ties and interaction between persons or other social objects are primarily interested in examining the overall network structure. What is the structure of ties within a social group, among social groups, within or among organizations, etcetera? Network analysis offers tools for describing overall network structure, disclosing, for instance, cohesive subgroups and ranked layers, or bottlenecks in exchange networks. This approach to social networks is known as the socio-centred approach. The substantive interest, then, is to find out whether cohesive subgroups identified in the network actually represent communities and whether ranked layers identify social strata. In a constructivist perspective, it may even be hypothesized that social identities and classifications are derived from network structure, e.g., groups are labelled as different because they occupy clearly different positions within a network of ties.

The other approach to social networks focuses on the individual actor and its immediate network neighbourhood. This is known as the ego-centred approach, which is currently being developed as the actor-oriented approach (see Section 7.2). It is assumed that social behaviour is orchestrated: actors adjust their behaviour and attitudes, opinions, and beliefs to the behaviour (etc.) of other members of the social system in which they participate. This is the social influence model of networks. As a network of ties, the system defines to whom an actor is exposed. The immediate contacts – the neighbours in graph theory – of an actor are usually most salient to its behaviour, but indirect contacts such as their neighbours' neighbours may be taken into account as well. In other words, an actor's local context or ego-network is likely to affect its behaviour.

At the same time, however, actors decide on which ties to establish, maintain, or end. This has been labelled the social selection model. Properties of alters usually play a role here, e.g., a preference for interaction with people that are similar to you, known as the homophily principle. Local network structure may also affect the creation or dissolution of ties, e.g., the often encountered phenomenon that people tend to become friends of their friends' friends. By ending ties or creating new ones, the individual changes both local network structure and overall network structure, that is, the system. Overall network structure, then, is regarded as the outcome of individual action. To the actors, the change of network structure is not necessarily predictable, so the interplay between individual action and network structure may offer surprising results.

The subsequent sections present the three main theoretical approaches to social networks: cohesion, brokerage, and ranking. Whenever applicable, both the socio-centred and ego-centred approaches are discussed.

4 Cohesion

One of the first intuitions in SNA concerns the tendency of human beings to form cohesive subgroups. This is a classical topic in the social sciences, see, for instance, George C. Homans' book *The Human Group*, and it was central to the sociometry tradition started by Jacob L. Moreno. Intuitively, the concept of social cohesion translates into relatively densely connected sections within the network: group members relate more extensively, frequently, or more positively among themselves than to members of other subgroups. In the history of SNA, the concept of relatively densely connected subnetworks has yielded a large number of graph theoretical ways for identifying cohesive subgroups at the level of overall network structure. Limiting the discussion to one-mode networks, that is networks in which there can be a tie between any pair of vertices, four approaches can be distinguished.

In the first and strictest approach, a cohesive subgroup is defined as a set of vertices in which all vertices are adjacent, that is, directly linked, to one another. In other words, cohesive subgroups are maximal complete subgraphs, which are called cliques. The term *maximal* indicates that no vertex can be added to the subgraph without violating the criterion. In the Florentine families network (Figure 1), three cliques of size three, that is, containing three families each, are found: the Medici – Ridolfi – Tornabuon clique, the Castellan – Peruzzi – Strozzi clique, and the Bischeri – Peruzzi – Strozzi clique. Note that the last two cliques overlap, which is typical for cliques. This poses an interpretive challenge to the researcher: should they be considered as separate cohesive subgroups or not?

The second approach is based on the notion of reachability and closeness of members within a subgroup. Members of a subgroup must be reachable in the sense that there are paths between them, i.e., a sequence of lines such that the end vertex of one line is the starting vertex of the next line such that no vertex in between the first and last vertex occurs more than once. If the direction of the lines is ignored, the sequence of lines is called a semipath. In addition, the shorter the geodesics (shortest paths) between them, the closer the vertices are in a graph theoretical sense, so the more cohesive the network structure of the subgroup is.

The criterion of reachability does not necessarily yield very dense subgroups. In sparse networks, any maximal connected subgraph (strong component) may represent a cohesive subgroup: there is a path between each pair of vertices within a component. In the Florentine families example, all families are directly or indirectly linked by marriage ties, except the isolated Pucci family. Increasing the required number of independent paths between any pair of vertices within a cohesive subgroup yields slightly denser subgroups, e.g., requiring at least two independent paths produces bi-components, which may be generalized to *k*-components for higher minimum numbers of independent (vertex-disjoint) paths between all vertices within a subgroup. In our example, there are at least two independent marriage paths among ten families. There are single paths only to the Pazzi, Salviati, Acciaiuol, Lambertes, and Ginori familes and as noted before, there are no marriages with the Pucci family.

Focusing on graph-theoretical distance between vertices usually yields denser subgroups. One may, for instance, set a maximum *n* to the distance between any two vertices within a subgroup, which is the concept of an *n*-clique. In our example, the families Bicheri, Castellan, Peruzzi, Ridolfi, and Strozzi are a *2*-clique because they are connected by paths of maximum length two in the marriage network and there is no other family connected to all of them by paths of this maximum length. Adding the restriction that the diameter of an *n*-clique is *n* or less, one obtains *n*-clans. Alternatively, one may define a cohesive subgroup as a maximal subgraph of diameter *n*, which is called an *n*-club.

The third approach focuses on the minimum number, strength, or multiplicity of ties among subgroup members. Subgraphs that are maximal with respect to the minimum number of neighbours within the subgraph are called k-cores and a maximal subgraph with respect to the maximum number of vertices in the subgraph that are not adjacent are known as k-plex. In a similar vein, restrictions can be imposed on the minimum strength or multiplicity of ties among members of a cohesive subgroup, generalizing Seidman's concept of a k-core to a valued core, which is called an m-core or m-slice: maximal connected subgraphs considering only lines with minimum value (or multiplicity) m.

In the fourth approach, cohesive subgroups are based on the relative frequency of ties among subgroup members in comparison to non-members: cohesive subgroups are relatively dense sections within the network, that is, relative to the sections outside (and between) subgroups. An LS set is a maximal subgraph such that any of its subsets has more ties to its complement within the LS set than to vertices outside the LS set. This idea was generalized to the concept of the lambda set, which requires the number of edge-disjoint paths between any pair of vertices within the lambda set to be larger than between any vertex within and any vertex outside the lambda set. In a signed network, e.g., a network of affective relations in which positive affections are represented by positive lines and negative lines show negative affections, cohesive subgroups can be delineated as sets of vertices which have relatively many positive lines among themselves and relatively many negative lines between members of different subgroups. Finally, clustering techniques and some types of blockmodels also detect clusters of vertices that have relatively many ties within clusters and few among clusters. These models offer an alternative way for finding cohesive subgroups, see Section 7.1.

Thus far, the socio-centred approach has been presented for detecting cohesive subgroups in the overall network. Now let us turn to the ego-centred approach: where do cohesive subgroups come from and what do they do?

The first and most general behavioural hypothesis merely states that similar people tend to interact more easily and people who interact tend to become or perceive themselves as more similar provided that the interaction is characterized as positive, friendly, and so on. In SNA, this tendency is mainly known as homophily, a concept coined by Paul F. Lazarsfeld and Robert K. Merton, but it is known under other names in several scientific disciplines, e.g., the phenomenon of attribution and affect control in social-psychology, assortative or selective mixing in epidemiology and ecology, and assortative mating in genetics.

If we concentrate on the graph theoretical aspects of this behavioural hypothesis, that is, the structure of ties, and take the (dis)similarities among the actors for granted, we find several characteristics of local structures that measure cohesive subgroup formation. At the level of a pair of actors, reciprocity of ties in directed networks (arcs) signals subgroup formation: both actors are hypothesized to choose each other when they are similar. At the level of the triple, transitivity results from tendencies toward cohesion. If actor u establishes a tie with actor v because they are similar, and actor v establishes a tie to actor w for the same reason, actors u and w are also similar, so actor u is expected to establish a tie with w as well, creating a so-called transitive triad (Figure 3). Stated differently, the path from u via v to w is closed by an arc from u to w. In general, the closure of paths or semipaths both in directed and undirected networks signals cohesive subgroup formation at the local level. Closure within an ego-network may be calculated as the percentage of all possible ties among a vertex' neighbours that are present, which is one of the definitions of the clustering coefficient. The concept of closure can be extended beyond a vertex' immediate neighbours, e.g., the number of semicycles (closed semipaths) of length 4 or larger in which a vertex is involved, e.g., balanced semicycles in signed networks.

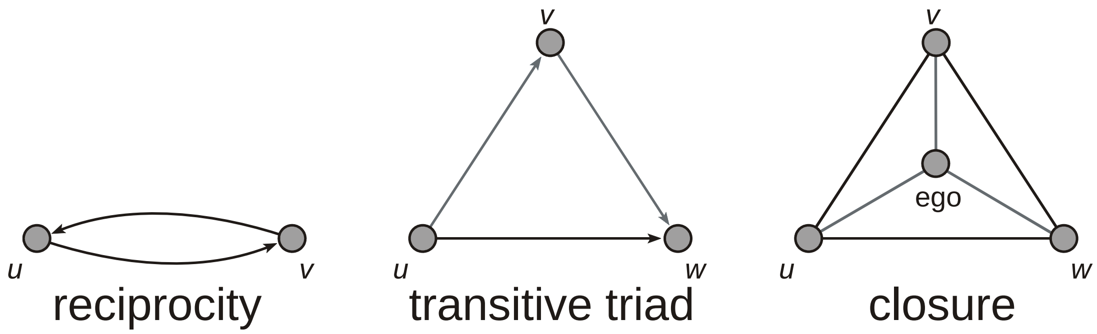

Figure 3 - Reciprocity, transitivity, and closure.

If we include measured attributes of the vertices in our indicators of cohesive subgroup formation, we can calculate homophily quite simply as the probability or ratio of ties between vertices that share a particular characteristic to ties between vertices that do not. Extending this idea to the ego-network, the homogeneity of actors involved in an ego-network may be taken as a measure of tendencies toward homophily. For qualitative attributes of the actors, Blau's index of variability or heterogeneity can be used ($1-\sum p_i^2$ where p is the proportion of group members in a particular category and i is the number of different categories), which is conceptually related to the Herfindahl-Hirschman Index in economics measuring the extent of monopoly within an industry. It is interesting to note that Blau's theory hypothesizes that heterogeneity rather than homogeneity of actors within a group enhances the operation and efficiency of the group. If improving group efficiency is the aim of actors, we would have to use a behavioural hypothesis that is quite the opposite of the homophily hypothesis.

In addition to homophily, there is a second behavioural hypothesis related to cohesion in SNA. This hypothesis is based on the idea that social action is embedded in networks. Named after the sociologist Georg Simmel, Simmelian ties are ties that are embedded in other ties, e.g., business ties are embedded in family ties, or in complete triads and cliques. Embedded ties are hypothesized to enforce group norms and enhance trust, hence pressure people into the same behaviour because there are parallel ties or because the two actors involved in a tie share common neighbours who supervise their behaviour. In the Florentine families example, we see that eight out of fifteen business ties are backed up my marriage alliances. Three out of seven cases in which there is a business tie but no marriage, involve the Lambertes family.

Just as with the homophily hypothesis, the embeddedness hypothesis predicts that tightly connected actors will be more similar in their behaviour and attitudes. In addition, it predicts that embedded ties are more stable and new ties are more likely to be established when they are embedded in existing cliques or existing ties. Closure again is an important indicator of tendencies to establish and maintain embedded ties but so is the multiplicity of relations: the extent to which a tie on one social relation duplicates a tie on another social relation.

If data on vertex attributes are available, especially if they concern public behaviour, that is, behaviour that is easily noticed by third parties such as publicly expressed opinions and statements, Simmelian ties are hypothesized to produce a special effect. Involvement in different groups (cliques) then exposes actors to possibly conflicting sets of norms and loyalties, which may urge them to cut their ties with some or all of these groups. In this case, actors are hypothesized to withdraw from stressful relations, so they discontinue ties that incorporate them into cliques (with a preference for cutting a minimum number of ties) or they discontinue ties such that they are no longer connected to actors voicing different opinions or norms.

5 Centrality and Brokerage

The notion of centrality in social networks has a long history in SNA. It is attributed to Alex Bavelas in 1948. In discussions of centrality, network ties are usually regarded as channels for the exchange of information, goods, services, and so on. Being central in this exchange system has always been hypothesized to be related to influence and satisfaction. Centralization, as a characteristic of a network, has been linked to the efficiency of a network as an exchange system. More centralized groups, for example, have often been shown to be more efficient.

In 1979, Linton C. Freeman argued that the approaches to centrality are based on three ideas about what being central means: (1) being active within the network, that is, maintaining many ties, (2) being efficient or independent of go-betweens by having short distances to other vertices in the network, and (3) being an important go-between, that is, being part of many paths between other vertices in the network. Although alternative classifications and approaches exist, Freeman’s classification is used here, adding concepts of brokerage that have been developed elsewhere in SNA.

5.1 Activity

Being active or prominent in the network means that an actor has many ties, hence access to many sources of information (etc.). The degree of a vertex (the number of lines incident with a vertex), then, is the relevant graph theoretical measure of local structure, which is also known as degree centrality. In the Florentine families example, the Medici family has highest degree centrality both in the marriages network (it has marriage alliances with six families) and in the business ties network if we disregard the direction of business ties (it has business ties with 5 families). Perhaps this explains why this family was most powerful even though it was by far not the richest family and it did not use to have a seat in the city council (compare vertex sizes and vertex colours in Figure 1).

Centralization is the corresponding structural property of the overall network and it is defined as the variation in the centrality scores of the vertices in the network because this variation shows the extent to which there is a centre (very central vertices) and a periphery (vertices with very low centrality scores). The star and ring networks are defined as respectively the most and least centralized networks (centralization scores of 1 and 0). They are known to exhibit the highest and lowest variation in centrality scores in simple networks. See Figure 4 for an illustration, showing a star and a ring network labelling the vertices with their degree centrality scores. The degree centralization of the Florentine marriage network is .27 and for the business ties it is .12. The centralization is far from maximal (1.00), as is to be expected in most social networks but the difference between the marriage and business networks is remarkable.

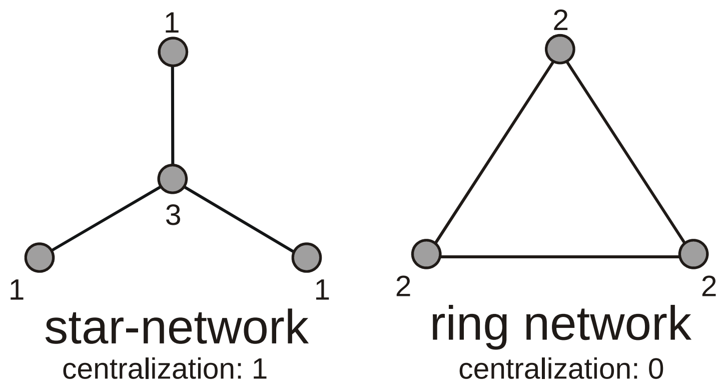

Figure 4 - Centrality and centralization in a star network and a ring network.

Where does degree centrality and centralization come from? From research on the power law in networks, it is known that preferential attachment to degree, that is, a tendency to connect to the best connected actors in the network, creates networks with a peculiar degree distribution, including many vertices with low or modest degree and few vertices with high degree. According to the definition of centralization in network analysis, this implies large variation in degree centrality scores, hence high degree centralization. In other words, the behavioural hypothesis of preference for high degree actors produces centralized overall network structure.

5.2 Efficiency and weak ties

The second approach to centrality focuses on graph theoretical distances between vertices. The central idea here is that actors have better access to information or other resources if fewer go-betweens are needed to reach or be reached by all other actors in the network. A minimum number of go-betweens yields maximum independence and maximum efficiency in the exchange network to vertices and the overall network.
Graph theoretical measures of local structure focus on graph theoretical distance, that is, the minimum length of paths between vertices because path length equals the number of go-betweens in the network plus one. Closeness centrality is a straightforward implementation of this idea because it merely normalizes the average graph theoretical distance between a vertex and all other reachable vertices in the network. In addition, paths can be weighted by the centrality of the vertices on them, which is done by Phillip Bonacich's eigenvector centrality. Finally, the difference between incoming paths and outgoing paths may be added.
The closeness centrality of the Medici family in the marriage network is .56, which is the number of reachable vertices (14 because the Pucci family is isolated) divided by the sum of the distances to (or from) these families (25 here). Note that the result may be slightly different if the unreachable Pucci family is taken into account as well. Again, the normalized variation of closeness centrality scores of the vertices in the network yields the appropriate measure of centralization of overall network structure.
The strength of weak ties argument proposed by Mark Granovetter may be regarded as a special application of the notion of efficiency. In his research on finding a job, Granovetter noticed that relatively superficial ties, ties with infrequent contact, give access to new information because they are more likely to link you to someone with whom you are not linked directly or at a short distance. Strong or intense ties tend to be situated within cohesive subgroups, so they are more likely to offer redundant

information already received through other ties. Granovetter is only interested in the effects of having weak ties, but if we turn his idea into a behavioural hypothesis, it suggests a preference to relate to distant vertices that are neither connected to yourself nor to your neighbours (or your neighbours' neighbours, and so on).

5.3 Control and structural holes

The third approach to centrality focuses on control over flows within the network: the more you are in between other vertices in the network, the more they depend on you to pass on information, the more you are able to control exchange within the network and profit from your control. Using this type of control is called brokerage.
The notion of being in between other vertices has a straightforward translation to graph theory as being part of a path between two other vertices. Limiting paths to the shortest paths between vertices both betweenness centrality and rush of a vertex are based on the proportion of all geodesics between other vertices that include this vertex. This measure has been extended to handle directed ties. In addition, information centrality takes into account all paths between vertices, not just the geodesics and flow betweenness or entropy also consider the values of lines. Betweenness centralization is the normalized variation of betweenness centrality of the vertices in the network.
In the network of business ties among Florentine families, just two families are included in paths between other families if we take the direction of lines into account: the Medici and the Peruzzi families. The former is in between, on the one hand, the Barbadori and Tornabuon families and on the other hand the Pazzi and Salvati families. Direction, however, is not very important here because it only tells us the wealthier partner in a business tie. If we ignore the direction of business lines, four more families are in between other families (Figure 5, vertex size and numbers between brackets expresses betweenness centrality), Now the Barbadori family scores even higher on betweenness centrality than the Medici family because the former family connects the business cluster around the latter with the rest of the business network.

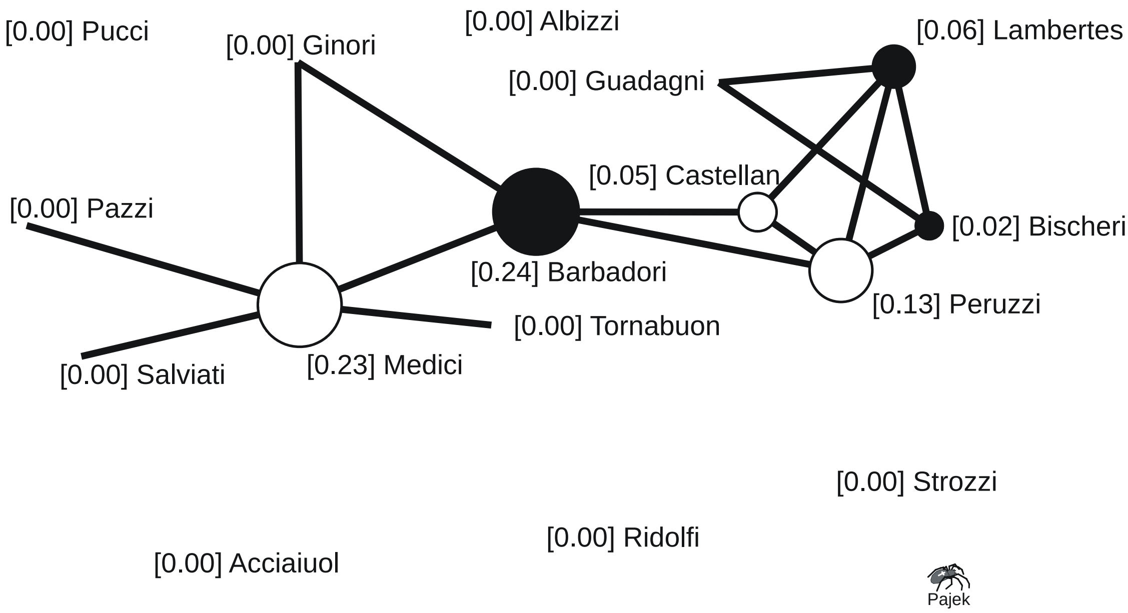

Figure 5 - Sociogram of undirected business relations between 16 Florentine families (circa 1400 AD).

While betweenness centrality situates an actor with respect to all other actors in the network, Ronald S. Burt proposed a local variant, focusing on control within the ego-network of an actor. His behavioural hypothesis rests on the *tertius gaudens* principle: the benefits that accrue to an actor that is in between two actors that are not directly linked because of the opportunity to broker information between them or, in a more malicious variant, to divide and conquer.
This hypothesis translates quite easily into graph theoretical structure. The absence of a tie between two neighbours of an actor is called a structural hole (Figure 6a). The behavioural hypothesis states that actors try to increase the structural holes that they can exploit. At the same time, however, they try to minimize the structural holes through which they can be exploited. This means, among other things, that an actor will not end a tie to one of its neighbours if the two neighbours are directly linked (see Figure 6b) because that would create an opportunity to broker at the expense of the actor. In this situation, the actor is constrained in its opportunities to change ties.

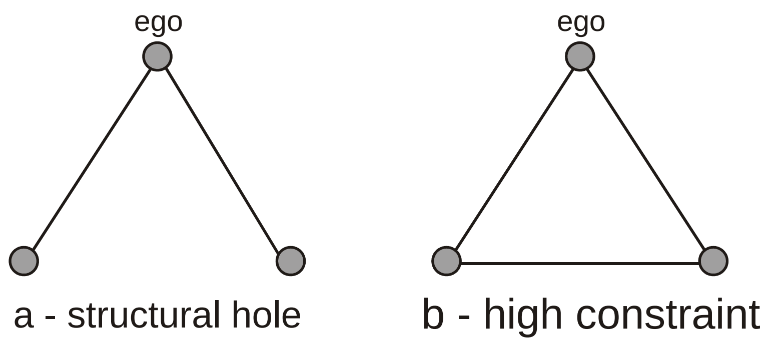

Figure 6 - A structural hole (a) and a triad with high constraint (b).

Structural holes and constraint are the flip sides of the same coin. A tie with low constraint indicates that the tie is involved in (many) incomplete triads (such as Figure 6a), so there are (many) structural holes offering the actor options for brokering. High constraint on a tie means that it is part of (many) complete triads (as in Figure 6b), so there are few or no possibilities for brokerage. In the undirected business ties network among Florentine families (Figure 5), the constraint is lowest for the Medici family because it is involved only in one complete triad (with the Barbadori and Ginori families) from which it cannot retreat without creating a structural hole around itself that can be exploited by one of the other families. However, the Medici family has structural holes between all other pairs of families in its ego-network, yielding many options for brokerage.
Because the presence or absence of ties among an actor's neighbours is key to the argument, network analysts have also used the density of the ego-network without the ego as a proxy of constraint: the higher the density, the higher the constraint on the ego. Alternatively, betweenness centrality for ego-networks can be used.

6 Prestige and Ranking

For prestige and ranking the direction of ties is crucial because asymmetry in networks is assumed to be linked to social prestige. The general idea here is that social inequalities are reflected and possibly created by asymmetric ties. For example, everyone invites the most popular boy or girl in class but s/he doesn't return each invitation and, at the same time, being invited may strengthen her popularity. Of course, the nature of the social relation determines the direction of choices; ties like "reports to" or "pays respect to" point toward higher levels in a hierarchy while "beating up" points in the opposite direction.

A central behavioural hypothesis concerns the popularity or attractiveness of actors. Actors tend to (want to) relate to actors with attractive structural properties or attributes, so attributes related to power or social status increase the probability that an actor will be chosen. From a constructivist point of view, however, being chosen often is also interpreted as a sign of importance and prestige, so receiving many choices (ties) increases the probability of receiving even more. In this way, networks may produce informal status hierarchies. The Matthew Effect, introduced in sociology by Robert K. Merton, comes to mind here: "For unto every one that hath shall be given, and he shall have abundance: but for him that hath not shall be taken away even that which he hath" (gospel of Matthew XXV, 29).

In graph theoretical terms, the structural attractiveness of an actor refers to the number of incoming arcs, which is simply the indegree of a vertex. This is called the popularity of a vertex and, of course, we must replace it by the vertex' outdegree if the relation is negative, e.g., submission, beating up, criticizing. If indirect choices must be taken into account as well, attractiveness is measured by proximity prestige, which is based on the average distance from all other vertices in the network – a directed variant of closeness centrality. Proximity prestige captures the idea that nominations or choices by actors who are themselves popular, contribute more to one's structural prestige. Bonacich's measure of power adds the idea that power may also be derived from being connected to powerless actors rather than to other powerful people.

Adding data on social attributes that make some actors more prestigious such as wealth, social class, beauty, and so on, we should expect preferential attachment to vertices that score high on these attributes. Note that vertex attributes play a slightly different role here than in the case of cohesion. Now we are concerned with attributes measured minimally at the ordinal level, expressing prestige that an actor possesses to a higher or lower degree. In the case of cohesion, we deal with nominal attributes, which merely express an identity. In contrast to homophily, the attributes of the actor who initiates the directed tie (the tail of the arc) does not matter here because the effect is solely related to structural characteristics or attributes of the actor at the receiving end of the tie (the arc's head).

There is a second, slightly different behavioural hypothesis relating to deference or submission rather than attractiveness. The idea is that actors mainly tend to create positive ties to other actors in their own status group or to actors in a higher status group – the people they are looking up to – to consolidate and improve their social position. Similarly, they tend to direct negative ties to actors in lower status groups. Note the difference with attractiveness: it is hypothesized that actors choose upwardly but they need not prefer the most attractive (top) actors in the network as they are supposed to do according to the attractiveness hypothesis.

The main difference between attractiveness and deference is that the former only takes into account structural properties or attributes of the tie's receiver, whereas the characteristics of both the sender and receiver of the tie matter to the latter. This distinction has important consequences to the structure of the overall network. Whereas the attractiveness hypothesis yields centralized networks, the deference hypothesis yields layered networks. The layers consist of sets of vertices that are symmetrically linked, e.g., by reciprocal ties, while the ties between layers are asymmetric, all pointing in the same direction. This behaviour may be a consequence of a formal hierarchy, e.g., positions within an organization with formalized relations such as reports to, or it may actually show or produce an informal social hierarchy, e.g., status differences between men and women in a particular social setting.

In simple directed networks, triads, that is, three vertices and the lines among them, are the key to measuring tendencies toward ranking at the local level. A network consisting only of symmetric and empty ties can be partitioned into subgroups such that all vertices within subgroups are linked by symmetric ties and no ties are present between subgroups. These networks can only contain three types of triads, which define the models of balance and clusterability (Figure 7). Networks containing only or unexpectedly many of these triads consist of two or more cohesive subgroups that are not ranked.

Assuming that asymmetric ties represent the ranking of clusters into a hierarchy, the model of ranked clusters allows five more types of triads (Figure 7). The ranked clusters model requires arcs from each vertex to all vertices on higher ranks. This requirement is usually too strict for empirical social networks and it is relaxed in the transitivity model by simply adding one type of triad to the set of allowed triads (Figure 7). The transitivity model stipulates that clusters of vertices at different ranks are either completely linked or not linked at all, yielding a partial order. Finally, the model of hierarchical clusters accounts for asymmetries within clusters. Note the nesting of the models for overall network structure, which is why the sets of permitted triads is extended for more general models.

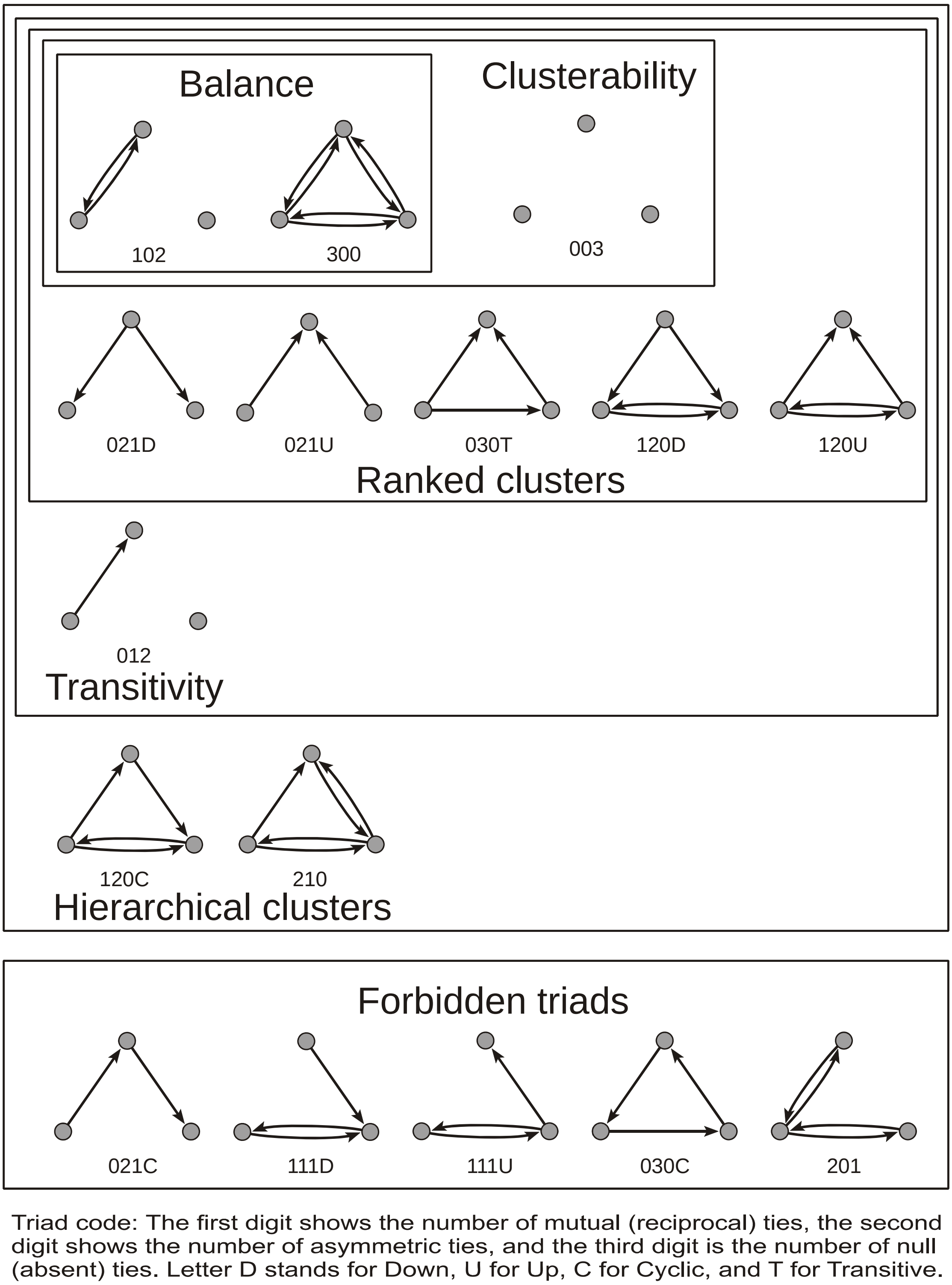

Figure 7 - Triad types and balance-theoretic models.

In the perfect case, there is a one-to-one relation between sets of occurring types of triads and the overall structure of the network. Therefore, the triad census, which is the frequency distribution of the sixteen types of triads in a directed network, offers an indication of overall network structure. In the imperfect case, the triad census of a network may be compared to the frequency distributions of triad types in random networks to test the tendency toward ranking. In the network of business ties among

Florentine families, for instance, triad types 003 (clusterability), 021D, 021U, 030T (ranked clusters), 012 (transitivity), and 021C (forbidden) appear. The forbidden triad indicates that the data do not perfectly fit any of the models. However, it appears less frequently than expected by chance given the number of vertices and arcs in the network. Actually, the 021D and 030T triads are the only ones that are more frequent than chance predicts, so a ranked clusters model is most appropriate for this network.

The triad census does not show the composition of the clusters and ranks; it does not identify the vertices belonging to particular clusters and ranks. This can be done in several ways. Realizing that ranks should be connected asymmetrically in directed networks, strong components cannot include more than one rank because vertices within strong components are by definition mutually reachable. Ties between strong components, then, are asymmetric, so it is easy to establish the ranking among strong components. Strong components, however, do not require a lot of symmetry in the ties; actually, not a single tie needs to be reciprocated. The latter requirement is made by the symmetric-acyclic decomposition, which looks for symmetric clusters defined as maximum subsets of vertices that are directly or indirectly linked by symmetric ties.

7 Future Directions

There are current developments both in the socio-centred approach and the ego-centred approach to social networks, which are expected to continue in the near future. Generalized blockmodelling is the prime example of a new, flexible technique for analyzing overall network structure. Exponential random graph models (ERGM) and Markov Chain Monte Carlo (MCMC) simulation models for network dynamics offer unprecedented possibilities for statistical testing of hypotheses on actor's behaviour and genesis of local network structure. In addition, SNA is increasingly becoming part of the toolkit of qualitative researchers.

7.1 Blockmodelling

In social theory, positions and roles are important and related theoretical concepts. A position, e.g., being an instructor at a university, is usually connected to a social role or a role set, e.g., tutoring students, conferring with colleagues, etc. It is hypothesized that this role or role set involves a particular pattern of ties and relations, e.g., towards students, colleagues, and superiors. Sociologists, social psychologists, and other social scientists investigate the nature of social roles and role sets by observing interactions and by interviewing people about their motives and their perceptions of their roles.

SNA concentrates on the patterns of ties. It tries to identify actors that have similar patterns of ties to other actors. Actors with similar patterns of ties are said to be relationally equivalent, to constitute an equivalence class, or to occupy equivalent positions in the network. If equivalent actors are detected, it is interesting to find out whether they are associated with a particular role or role set or, vice versa, if we have people with different social roles, it is possible to check whether this is reflected by different equivalence positions in the network.

In blockmodelling, the adjacency or sociomatrix is analyzed. Basically, this matrix is sorted (permuted) in a clever way such that a distinguishable pattern emerges. The pattern is summarized by partitioning the vertices into clusters such that the intersection of the rows belonging to a cluster and the columns belonging to a cluster are either empty or filled in a particular way. The intersections are called blocks, hence the name blockmodelling. The filling type of the blocks is specified in the image matrix in which each cluster defines a row and column. The cells in the blocks of the adjacency matrix are, so to speak, shrunk to one cell in the image matrix.

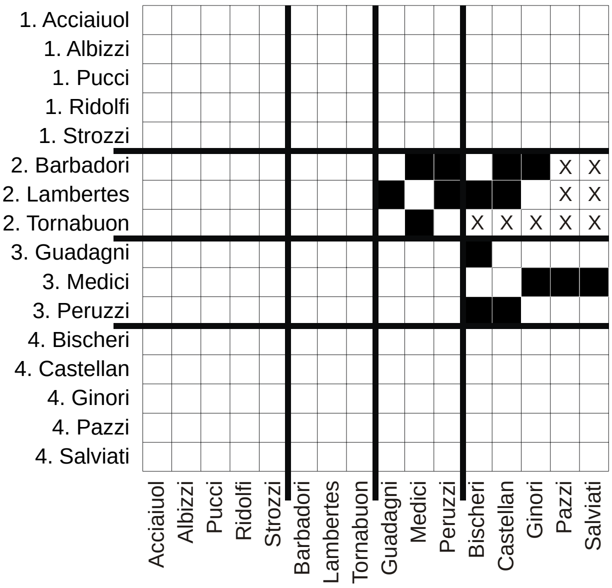

Image matrix

1	Empty	Empty	Empty	Empty
2	Empty	Empty	Regular	Regular
3	Empty	Empty	Empty	Regular
4	Empty	Empty	Empty	Empty

Figure 8 - Permuted adjacency matrix of business ties with 4 clusters.

Figure 8 shows the results of a blockmodel applied to the business ties among the Florentine families. Four cluster are identified: (1) the isolated families that have no business ties, (2) the families that trade only with wealthier families, (3) the families that trade both with more and less wealthy families, and (4) the families that trade only with less wealthy families. When the adjacency matrix is sorted according to these clusters and the clusters are separated by fat lines (Figure 8, left) a clear pattern of ties emerges in comparison to the original matrix (Figure 2). Many blocks are empty, e.g., all blocks in the rows and columns of the first cluster, the isolated families. They neither send (rows) nor receive (columns) business ties. Furthermore, all blocks on and under the diagonal are empty, whereas several blocks above the diagonal are filled. This pattern is characteristic of a ranked structure (see Section 6) without symmetric ties.

Cohesion and brokerage yield different types of permuted adjacency matrices. A network consisting of cohesive subgroups has a permuted adjacency matrix with most or all ties in the blocks on the diagonal: ties concentrate within clusters. Highly centralized networks, e.g., core-periphery structures that are usually found in the world trade system are characterized by permuted adjacency matrices with ties clustered in particular rows and/or columns: central vertices send and/or receive ties from many vertices whereas the latter have relatively few ties among themselves (remember the star network, Figure 4).

The image matrix (Figure 8, right) shows that the non-empty blocks are constrained to be regular equivalent. In this type of equivalence, each block must be either empty or regular, which means that each row and each column must contain at least one line. Substantively, this means that each vertex in the sending cluster is linked to at least one vertex in the receiving cluster and vice versa. Note that regular equivalent vertices do not need to be tied to the same vertices; as a university instructor, I have the same

pattern of ties with colleagues and students as a colleague at another university but I do not relate to the same persons. This is an important characteristic of blockmodels.
From the adjacency matrix (Figure 8, left,) it is clear that one block is not entirely regular, namely the block containing business ties from the second cluster to the fourth cluster. One family in cluster 2, the Tornabuon family, does not have a business tie with a family from the fourth cluster. In addition, two families in the fourth cluster (Pazzi and Salviati) do not have business ties with a family in the second cluster. The X signs mark the cells that should not all be empty. Errors like the ones described here are usually found when blockmodelling social networks. The goal then is to find the best fitting blockmodel, which is usually accomplished with optimization techniques. Statistical models for testing blockmodels are rare.
Regular equivalence is just one type of equivalence. Another type is structural equivalence, which is stricter than regular equivalence because it only allows empty and complete blocks, implying that structural equivalent actors are related to the same alters. The notion of blockmodelling is being generalized by allowing several other types of blocks and by giving the researcher the possibility to decide exactly which block should be of what type(s) in the image matrix. This is known as generalized blockmodelling, which is a general framework including structural and regular equivalence as special cases.

7.2 Statistical actor-oriented models

The preceding sections presented behavioural hypotheses that have similar, different, or even opposite consequences for overall network structure. SNA increasingly focuses on local structure using overall network structure merely as a collection of (overlapping) local structures. Behavioural hypotheses translate much more directly to local structure, that is, to the ties of the actor and those of its neighbours and possibly their neighbours' neighbour. Local structure is the part of the network that an actor can easily survey and actually change.
The latest developments in techniques for modelling network structure and evolution apply the actor-oriented or local structure approach either in statistical models, notably Exponential Random Graph Models (ERGM) and Markov Chain Monte Carlo simulation models for repeated measurements of social networks esp. as implemented in the SIENA software package, or in simulation models, such as agent-based computational approaches. The techniques test behavioural tendencies by relating the existence, creation, maintenance, and ending of ties by individual actors to the local configuration of ties, to previous ties between the actor and the alter or to present ties on another social relation, and to characteristics of both the actor and the alter.
In principle, the actor-oriented approach is able to test all behavioural hypotheses presented in the preceding sections. If hypothesized local configurations appear more often than expected by chance, the underlying behavioural hypothesis is assumed to guide individual behaviour at least to some extent. If the behaviour in a set of actors or of an individual actor is in line with several behavioural hypotheses at the same time, the effect of each behavioural tendency can be separated. Thus, it is possible to link complex overall network structure to compound behaviour of the actors in the network.
The techniques for analyzing local structure are in development. Models for the co-evolution of relations and quantitative attributes of vertices over time have just been introduced. Non-standard types of networks such as signed relations, incomplete data, and external constraints on data collection or conceptual constraints on network

structure such as two-mode networks may limit the applicability of current models and will hopefully spur the development of new ones.

7.3 Qualitative network analysis

As applied graph theory, SNA is a quantitative method. However, its discrete nature, that is, its focus on patterns as the presence versus absence of lines, and the human capability to explore patterns by eye-balling network visualizations, which are easy to produce with current software, make it a tool for quantitative researchers as well. Already some time ago, handbooks for qualitative research such as Miles and Huberman's standard work advised drawing observed interaction within a social group as sociograms. Software for qualitative analysis produces hierarchical coding schemes, which are represented as networks. Complex coding schemes can be subjected to network analysis to explore their structure.

Alongside the application of network analysis as a tool in qualitative research, efforts are being made at defining a type of network analysis that is intrinsically qualitative. The focus is on what specific ties mean to actors as part of a thick description of the data rather than assembling data on general types of ties among group members. The identification of patterns of ties then does not seem to be relevant and it remains to be seen what kinds of analyses result.

Glossary

actor Actor refers to a person, organization, or nation that is involved in a social relation. Hence, an actor is a vertex in a social network.

acyclic network An acyclic network does not contain cycles.

adjacency matrix An adjacency matrix is a square matrix with one row and one column for each vertex in a network. The content of a cell in the matrix indicates the presence and possibly the sign or value of a tie from the vertex represented by the row to the vertex represented by the column.

adjacent Two vertices are adjacent if they are connected by a line.

arc An arc is a directed line. Formally, an arc is an ordered pair of vertices.

asymmetric dyad An asymmetric dyad is a pair of vertices connected by unilateral arc(s).

attribute An attribute is a characteristic of a vertex measured independently of the network.

balance model The balance model applies to an unsigned directed network if it consists of two cliques that are not interrelated.

betweenness centrality The betweenness centrality of a vertex is the proportion of all geodesics between pairs of other vertices which include this vertex.

betweenness centralization Betweenness centralization is the variation in the betweenness centrality of vertices divided by the maximum variation in betweenness centrality scores possible in a network of the same size.

bi-component A bi-component is a component of minimum size three that does not contain a cut-vertex.

bipartite network See: two-mode network.

block A block contains the cells of an adjacency matrix that belong to the cross-section of one or two classes.

blockmodel A blockmodel assigns the vertices of a network to classes and it specifies the permitted type(s) of relation within and between classes.

blockmodeling The technique to obtain a blockmodel is called blockmodeling.

cell A cell of a matrix is the intersection of a row and a column.

clique A clique is a maximal complete subnetwork containing three vertices or more.

closeness centrality The closeness centrality of a vertex is the number of other vertices divided by the sum of all distances between the vertex and all others.

closeness centralization Closeness centralization is the variation in the closeness centrality of vertices divided by the maximum variation in closeness centrality scores possible in a network of the same size.
clusterability model The clusterability model applies to an unsigned directed network if it consists of two or more cliques that are not interrelated.
clusterable (semi)cycle A cycle or semicycle is clusterable if it does not contain exactly one negative arc.
complete network A complete network is a network with maximum density: all possible lines occur.
component A (weak) component is a maximal (weakly) connected subnetwork.
degree The degree of a vertex is the number of lines incident with it.
degree centrality The degree centrality of a vertex is its degree.
degree centralization Degree centralization of a network is the variation in the degrees of vertices divided by the maximum degree variation which is possible in a network of the same size.
density Density is the number of lines in a simple network, expressed as a proportion of the maximum possible number of lines.
digraph A digraph or directed graph is a graph containing one or more arcs.
distance The distance from vertex *u* to vertex *v* is the length of the geodesic from *u* to *v*.
dyad A dyad is a pair of vertices and the lines between them.
edge An edge is an undirected line. Formally, an edge is an unordered pair of vertices.
ego-centred approach An ego-centered approach to a network considers the structural characteristics of individual vertices.
ego-network The ego-network of a vertex contains this vertex, its neighbors and all lines among the selected vertices.
equivalent (class, position) Actors with similar patterns of ties are said to be relationally equivalent, to constitute an equivalence class, or to occupy equivalent positions in the network.
generalized blockmodeling In generalized blockmodeling, the permitted block types are specified for each individual block.
geodesic A geodesic is the shortest path between two vertices.
graph A graph is a set of vertices and a set of lines between pairs of vertices.
hierarchical clusters model The hierarchical clusters model applies to an unsigned directed network if it consists of connected clusters such that clusters within ranks are not related and clusters between ranks are related by null dyads or asymmetric dyads pointing towards the higher rank with the additional provision that a cluster does not contain cycles of asymmetric dyads.
homophily Homophily is the phenomenon that similar people interact a lot, at least more often than with dissimilar people.
image matrix An image matrix is a simplification of an adjacency matrix by shrinking each block to one new cell indicating the block type.
incident A line is defined by its two endpoints, which are the two vertices that are incident with the line.
indegree The indegree of a vertex is the number of arcs it receives.
k-connected component A *k*-connected component is a maximal subnetwork in which each pair of vertices is connected by at least *k* distinct (non-crossing) semipaths or paths.
k-core A ***k***-core is a maximal subnetwork in which each vertex has at least degree *k* within the subnetwork.
line A line is a tie between two vertices in a network: either an arc or an edge.
loop A loop is a line that connects a vertex to itself.
matrix A matrix is a two-way table containing rows and columns.
m-slice An ***m*-slice** is a maximal subnetwork containing the lines with a multiplicity equal to or greater than *m* and the vertices incident with these lines.
multiple lines If a particular arc or edge, that is, a particular ordered or unordered pair of vertices, occurs more than once, there are multiple lines.
multiplicity Line multiplicity is the number of times a specific line (ordered or unordered pair of vertices) occurs in a network.
neighbour A vertex that is adjacent to another vertex is its neighbor.
network A network consists of a graph and additional information on the vertices or the lines of the graph.

null dyad A null dyad is a pair of vertices that are not connected by lines.
one-mode network In a one-mode network, each vertex can be related to each other vertex.
optimization technique An analytic technique searching for the best solution according to a criterion function by repetition is called an optimization technique.
outdegree The outdegree of a vertex is the number of arcs it sends.
partial order In a partial order, some but not all pairs of units (e.g., vertices) are ordered.
partition A partition of a network is a classification or clustering of the vertices in the network such that each vertex is assigned to exactly one class or cluster.
path A path is a walk in which no vertex in between the first and last vertex of the walk occurs more than once.
permutation A permutation of a network is a renumbering of its vertices.
popularity The popularity or indegree of a vertex is the number of arcs it receives in a directed network.
position A position is a particular pattern of ties.
ranked clusters model The ranked clusters model applies to an unsigned directed network if it consists of cliques and ranks such that cliques within ranks are not related and cliques between ranks are related by asymmetric dyads pointing towards the higher rank.
reachable We say that a vertex is reachable from another vertex if there is a path from the latter to the former.
regular block A regular block is a block containing at least one arc in each row and in each column.
regular equivalence Vertices that are regular equivalent do not have to be connected to the same vertices, but they have to be connected to vertices in the same classes.
semicycle A semicycle is a closed semipath.
semipath A semipath is a semiwalk in which no vertex in between the first and last vertex of the semiwalk occurs more than once.
semiwalk A semiwalk from vertex *u* to vertex *v* is a sequence of lines such that the end vertex of one line is the starting vertex of the next line and the sequence starts at vertex *u* and ends at vertex *v*.
signed graph A signed graph is a graph in which each line carries either a positive or a negative sign.
simple graph A simple undirected graph contains neither multiple edges nor loops. A simple directed graph does not contain multiple arcs.
socio-centred approach A socio-centered approach to a network considers the structural characteristics of the entire network.
star-network A star-network is a network in which one vertex is connected to all other vertices but these vertices are not connected among themselves.
strong component A strong component is a maximal strongly connected subnetwork.
strongly connected A network is strongly connected if each pair of vertices is connected by a path.
structural equivalence Two vertices are structural equivalent if they have identical ties with themselves, each other, and all other vertices.
structural hole There is a structural hole in the ego-network of a vertex if two of its neighbors are not directly connected.
structural property A structural property is a characteristic (value) of a vertex that is a result of network analysis.
symmetric-acyclic model The symmetric-acyclic model applies to a directed network if it consists of clusters of vertices that are linked by symmetric ties directly or indirectly and if the ties among the clusters produce an acyclic network (when the clusters are shrunk).
symmetrize To symmetrize a directed network is to replace unilateral and bi-directional arcs by edges.
transitive triad In a transitive triad, each path of length two is ‘closed’ by an arc from the starting vertex to the end vertex of the path.
transitivity model The transitivity model applies to an unsigned directed network if it consists of cliques such that cliques within ranks are not related and cliques between ranks are related by null dyads or asymmetric dyads pointing towards the higher rank.
triad A triad is a (sub)network consisting of three vertices.
triad census The triad census of a directed network is the frequency distribution of the sixteen types of triads in this network.

two-mode network	In a two-mode network, vertices are divided into two sets and vertices can only be related to vertices in the other set.
undirected graph	An undirected graph does not contain arcs: all of its lines are edges.
valued network	A valued network is a network in which lines have (variable) values.
vertex (vertices)	A vertex (singular of vertices) is the smallest unit in a network
walk	A walk is a semiwalk with the additional condition that none of its lines is an arc of which the end vertex is the arc's tail. One might say that you always follow the direction of arcs in a walk.
weakly connected	A network is weakly connected if each pair of vertices is connected by a semipath.

Bibliography

Barabási A.-L. (2002). *Linked: The New Science of Network*. Cambridge, MA: Perseus Publishing. [The seminal book on scale free networks and the power law.]

Carrington P.J., Scott J., Wasserman S. (2005). *Models and Methods in Social Network Analysis*, 328 pp. New York, USA: Cambridge University Press. [Overview of recent developments in social network analysis and software overview.]

De Nooy W., Mrvar A., Batagelj V. (2005). *Exploratory Social Network Analysis with Pajek*, 334 pp. New York, USA: Cambridge University Press. [This handbook offers an overview of network ideas and methods and guidance for doing network analysis with the free software package Pajek.]

Degenne A., Forsé M. (1999). *Introducing Social Networks*, 248 pp. London: Sage Publications. [Overview of social network analysis in the context of sociological theory.]

Doreian P., Batagelj V., & Ferligoj A. (2005). *Generalized Blockmodeling*, 384 pp. New York: Cambridge University Press. [Presents the latest developments in blockmodelling.]

Freeman L.C. (2004). *The Development of Social Network Analysis*, 205 pp. Vancouver: Empirical Press. [A history of social network analysis as a scientific endeavour.]

Macy M.W., Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, **28**, 143-166. [Describes the application of agent-based models to social network analysis.]

Robins G., Snijders T., Wang P., Handcock M., Pattison P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks* **29**, 192-215. [Special issue on new developments in statistical models for network analysis.]

Scott J. (1991, 2000[2]). *Social Network Analysis: A Handbook*, 208 pp. London: Sage Publications. [Overview of social network analysis at an introductory level.]

Wasserman S., Faust K. (1994). *Social Network Analysis: Methods and Applications*, 825 pp. New York, USA: Cambridge University Press. [The standard handbook for social network analysts.]