We start our chapter with a discussion of social systems composed of positions and roles. This is followed by a set of methods for identifying positions and roles and for delineating social network structures. We finish by listing some important open problems that require solutions, so that we can better understand the structure and operation of role systems.

INTRODUCTION

The paired concepts of positions and roles are staples among social science terms. Intuitively, the idea of a position is a location in some social structure and a role has a set of expected behaviors corresponding to the location (e.g., Faust and Wasserman, 1992). Given a family as a social system, a parent is a location and the parental role includes appropriate behaviors for rearing children. The child is another position that carries age-graded expectations of appropriate conduct by children toward their parents. Expectations of parents and children are coupled into a system of roles. Similarly, in an organization there are locations in some structure - stereotypically, a hierarchy - and roles are coupled to these locations. Expectations include rules for how superiors and subordinates behave in relation to each other inside their organization. Roles are defined for all levels and positions in the hierarchy and form a coupled system of expectations. Of course, this simple description identifies an idealized form for which there can be many variations. Key empirical issues involve delineating positions in social systems, identifying roles that correspond to these positions, the nature of, and extent to which, these roles exist, and examining how both role systems and social structures change over time.

Social network analysis provides a set of tools that includes ways of mapping social structures in ways that help identify positions and roles. When used for studying social structures over time, these tools help analysts understand how social structures and role systems change over time. If there are only simple descriptions of networks over time, this activity can be called studying network dynamics. However, if we are able to identify process rules that generate the observed changes over time, then we are examining the evolution of social structures over time. While both network dynamics and network evolution have their place in studying positions and roles, characterizing evolution is both more demanding and more important for constructing cumulative knowledge and understanding of roles and positions.

Social networks

A simple social network consists of a set of units, called social actors, with a single relation defined over them. For example, the units could be children and the relation defined as "plays with." For a family, the units are parents and children.
and one relationship is “controls” for parents acting toward their children. A more complex network has multiple relations. For the children on a playground, the relations studied could be “plays with” and “likes.” An even more complicated network has multiple relations and multiple levels. For a family, the levels are for parents and for children and the relationships studied could include “controls,” “loves (and/or hates),” “respects,” and “confides.” Additional levels could include multiple generations. In formal organizations with hierarchies, units can be individuals occupying locations at multiple levels with the relations “reports,” “seeks work-related help,” “provides work-related help,” and “socializes at breaks.”

Units in a social network can also include groups, organizations, and nations as well as the individuals in these larger and more extensive units. If the groups are gangs, the relations between them include alliance ties and enemy ties. For organizations, relations can include sending goods or people between organizations, sharing information, or forming alliances. For nations, the relations can include exports, imports, providing aid, belonging to military alliances, and waging war. Networks can also be made up of objects that have no obvious action identity as actors in the sense of individuals, groups, organizations, or nations. An example is the set of scientific documents for one or more scientific fields. These units include books, articles, and research reports. One relation defined over these objects is citation. Each scientific document contains references to earlier relevant work and the relational ties are citations of earlier documents by later documents. Patents form a similar set of units where the citations are governed by legal requirements to acknowledge prior inventions and their patents. For scientific articles, depending on the field, there are differing volumes of solo authored publications and joint productions by two or more scientists. For the latter, co-authorship is a relation defined over authors of documents and this can be coupled to citation networks to create a network database with networks involving quite different units. Ties in networks often have values capturing dimension such as intensity, frequency, or volume depending on the relation considered. In general, data sets can be created with both very large networks and multiple types of units. Regardless of the complexity of a network, the notions of positions and roles can be considered for them. Obviously, the way this is done depends on the size and complexity of the networks considered. But it can be done, and the approach known as generalized blockmodeling provides the tools for establishing positions and roles and thereby delineating social structures in a very general way. For the ease of exposition, most of our discussion is focused on simple networks but the tools can be used for any degree of network complexity.

**BLOCKMODELING**

One publication changed dramatically the way network analysts viewed the delineation and examination of social structure (as network structure). Lorrain and White (1971) introduced the concept of structural equivalence (defined below) as a way of operationalizing both position and role. In doing so, they set the foundations for studying rigorously empirical social structure and examining role systems. This led to the creation of blockmodeling. Based on their insights, Breiger et al. (1975) presented a practical algorithm for establishing positions in a network. It was based on a particular way of operationalizing structural equivalence. Burt (1976) provided an alternative operationalization and with it a different algorithm. Bailer (1978) provided another way of thinking about blockmodeling. This was later formalized by White and Reitz (1983) with the introduction of regular equivalence as a formal generalization of structural equivalence. In 1992, the flagship journal of the field, Social Networks, devoted a special issue to blockmodeling featuring a variety of approaches that had been created since the early statements that helped define the field. This helped create the conditions for the emergence of generalized blockmodeling as a systematic statement of the approach and secured the foundations of blockmodeling (Doreian et al., 2005). In the following, we do not discuss the details of the various algorithms used for establishing blockmodels. Instead, we focus attention on the core ideas. Our discussion of blockmodeling distinguishes classic blockmodeling and generalized blockmodeling. In doing so, we put the formal/mathematical foundations to one side. The cited documents informing our discussion provide the technical and formal details behind the nonmathematical statement provided here.

**Classic blockmodeling**

Some terms are used to provide a way of describing networks precisely. Actors are represented by vertices and social ties between them are represented by lines. A shorthand way of labeling a network is $N = (V, R)$, where $V$ represents the set of vertices and $R$ is a label for the relation. Networks with one relation can be viewed
as simple networks. For representing networks with multiple relations, this notation extends naturally to $N = (V, R_1, R_2, \ldots, R_r)$ for a set of $r$ relations. The ideas discussed below apply to all networks and, for ease of exposition, we use simple networks.

Some relations are inherently symmetric, for example, co-authoring a scientific paper. For such relations, a line representing a symmetric tie is called an edge. Other social relations are inherently asymmetric. For example, "parent of" is a relation represented by a line that goes from the parent to a child. A child can never be a parent of their parent(s).

When two actors are structurally equivalent they are connected in exactly the same way to other actors in the network. A formal definition can be found in Doreian et al. (2005: 172). In essence, they are structurally identical. A set of structurally equivalent actors is called a position. If the network has only sets of structurally equivalent actors then it is fully consistent with structural equivalence. This means the set of vertices $V$, can be partitioned into a set of $k$ clusters, $\{C_1, C_2, \ldots, C_k\}$ so that the vertices in a cluster $C_i$ are structurally identical. In this sense, given this definition of equivalence, there are $k$ positions in the network (representing a social structure). This provides a precise definition and operationalization of the term "position" as a cluster. Given two positions, $C_i$ and $C_j$, the set of ties from all actors in $C_i$ to all actors in $C_j$ forms a block. Given $k$ positions there are $k^2$ blocks and the whole structure is represented by these blocks. There are $k$ blocks where the ties are within positions and these are called diagonal blocks. There are $k(k-1)$ off-diagonal blocks containing ties between positions. If there are $n$ actors in the network and if $n$ is much larger than $k$, then the large network is represented by the blockmodel image where there are only positions and blocks. Put differently, the original network is modeled by the sets of positions and blocks – hence, the term blockmodel.

Most empirical networks are not described perfectly by the blockmodel just described. Pairs of actors are more likely to be "almost structurally equivalent" in the sense that their ties are with almost the same other actors. These differences, if small in number, are assumed to not matter that much empirically, which permits the retention of the idea of representing an observed network as a blockmodel. This raises the issue of empirically determining the positions and blocks. While there are many ways of doing this we confine our attention to those used most frequently. Two of these methods hinge on ways of representing the extent to which pairs of actors are structurally equivalent. Breiger et al. (1975) provided an algorithm based on having correlations represent the notion of "almost structurally equivalent." The location of an actor in a network is the vector of ties (both present and absent) involving an actor. For an undirected graph, the row (or column) is the location. For a directed graph, both the row and the column of ties for an actor represent the actor's location. If two actors are structurally equivalent, the correlation of their locations will be 1. Two actors are "almost structurally equivalent" if the correlation of their positions is "close enough to 1." The algorithm proposed by Breiger et al. (1975) iteratively uses correlations of locations to identify positions and, therefore, blocks.

Burt (1976) proposed an alternative way of operationalizing "almost structurally equivalent" by using the Euclidean distances between locations. If two actors are structurally equivalent, then the Euclidean distance between their locations is 0. "Almost structurally equivalent" became "the Euclidean distance is close enough to 0." The matrix of locations is turned into a matrix of distances, which is then subjected to a standard clustering method. Doreian et al. (2005) describe algorithms of the sort suggested by Burt and Breiger et al. as "indirect methods" because the network data are converted into (dis)similarities that are then clustered. There are three broad problems with these approaches: (i) there are many ways of constructing dis(similarities) and not all of them are compatible with structural equivalence; (ii) there are thousands of clustering algorithms and choosing one of them seems arbitrary, and (iii) the methods can be used only in an inductive fashion. The third is due to the fact that clustering diagrams or dendrograms are examined in order to discern the clusters that are then labeled positions. There is no upfront conceptualization beyond the idea of structural equivalence being applicable and analysts tend to accept what is returned by the joining of a clustering algorithm with a measure of (dis)similarity. These problems led Batagelj et al. (1992a) to...
pursue an approach they called a “direct approach” to blockmodeling. This was formalized further into “generalized blockmodeling” (Doreian et al., 1994, 2005) as a general method for partitioning networks into positions and blocks. Their analyses suggest that the direct approach produces better fitting partitions based on equivalence concepts than the indirect approach. Even so, the indirect approach remains useful and can be used more effectively for larger networks than the direct approach (described below) can handle.

**Generalized blockmodeling**

Generalized blockmodeling (Batagelj, 1997; Batagelj et al., 1992a, 1992b; Doreian et al., 1994, 2005), as a direct approach to network data, rests on some simple ideas. First, rather than think about structural equivalence as being approximated by a measure of (dis)similarity, it is more useful to think about what are the kinds of blocks that are consistent with structural equivalence. They are few in number: (i) diagonal blocks can have only two forms and (ii) off-diagonal blocks also have just two forms consistent with structural equivalence. The off-diagonal ideal or permitted blocks have only 0s in them or only 1s in them. They are, respectively, called *complete blocks* and *null blocks*. This captures exactly the idea of pairs of actors in different positions being connected in the same way to other actors. The same logic applies to diagonal blocks but the diagonal permitted blocks look slightly different. One permitted diagonal block has 1s everywhere except on the diagonal, which has only 0s. The other permitted form has 0s everywhere and 1s on the diagonal. The second block type of these two is rare empirically. We use the terms “complete blocks” and “null blocks” to describe both diagonal and off-diagonal blocks.

The direct approach sets up comparisons of an ideal blockmodel based on structural equivalence (which can have complete and null blocks anywhere) and an empirical blockmodel with the same number of positions and blocks. In general, the empirical blockmodel approximates an ideal blockmodel. Differences between ideal and empirical blockmodels are easy to construct conceptually. Wherever a 1 appears in a null block there is one type of inconsistency, and wherever there is a 0 in a complete block there is another type of inconsistency. All we have to do is count the inconsistencies and seek an empirical partition that makes the number of inconsistencies as small as possible. To do this, a clustering problem is formulated with a criterion function that represents the sum of all of the inconsistencies from all of the blocks. The clustering problem is the following one: in the set of all possible partitions into k clusters we search for the best partition according to the selected criterion function. The two types of inconsistencies can be weighted differently if one type of inconsistency is deemed more important than the other. If null blocks are thought to be important parts of a network structure then the presence of 1s in them can be penalized heavily so that blocks identified as null blocks contain no 1s.

Solving the clustering problem is not easy. While it sounds useful to compare all possible ideal blocks with all possible empirical blocks this can be done only for very small networks. So some heuristic is needed. The one used in solving the generalized blockmodeling problem is a relocation algorithm that makes local comparisons of possible ideal and empirical partitions. Some number (k) of partitions is chosen and the network is partitioned randomly into k provisional positions. The *neighborhood* for such a partition is made up of other partitions that can be reached by either of the two types of changes. One is simply to move a vertex from one provisional position to another position. The second is to exchange a pair of vertices between two provisional positions. For each change, the criterion function can be calculated before and after the change. If the criterion function does not decline, then the new partition is discarded. But if it does decline then we move to the new partition and repeat the changes (moving a vertex or exchanging a pair of vertices between positions) from the new partition. If a change lowers the criterion function, we move again to the new partition and continue until no further reduction is possible. Because this is a local optimization method, this is repeated many times to maximize the likelihood of reaching a globally best (optimal) partition rather than only a locally best (local) partition.

White and Reitz (1983) generalized the idea of structural equivalence to regular equivalence. Two actors are *regularly equivalent* if they are connected in equivalent ways to equivalent others. (A formal definition can be found in Doreian et al., 2005: 173.) This somewhat mysterious definition is best illustrated by a formal hierarchy with multiple levels that reach down to the same extent in each division. A partition based on regular equivalence will identify each of the levels as a position and captures the idea of position roles nicely. Subordinates (in a position) are expected to behave in the same way with their superordinates (bosses) while bosses are expected to behave in the same way toward their subordinates. Structural equivalence does not lead to this kind of partition. White and Reitz (1983) proved that regular equivalence is a proper generalization of structural equivalence. This equivalence was included in
generalized blockmodeling through a theorem that regular equivalence permits only two block types: null blocks and blocks that have at least one 1 in every row and column of the block. Such blocks are called 1-covered blocks. (The proper generalization is shown with the complete block of structural equivalence being a very special case of a 1-covered block.) Again, a criterion function expressing inconsistencies between ideal and empirical blocks for regular equivalence can be constructed and the local optimization method described above can be used.

The underlying strategy of the generalized blockmodel is simple to describe and involves turning a definition of a type of equivalence into a set of permitted block types. Structural equivalence permits only null and complete blocks. Regular equivalence permits only null blocks and 1-covered blocks. The link between a type of equivalence and its permitted blocks is driven by theorems establishing the permitted blocks for an equivalence type. A natural avenue for expanding the types of equivalences for establishing new types of blockmodels is to expand the permitted block types. One set of expanded block types can be found in Doreian et al. (2005: Chapter 7). These new types included row-regular blocks (each row is 1-covered) and column-regular blocks (each column is 1-covered). These were used to partition baboon-grooming networks at two points in time where neither structural equivalence nor regular equivalence were useful. This general strategy of expanding block types, and hence types of blockmodels, permits an indefinite expansion of blockmodeling into different substantive domains where the new block types can be more general and more useful than relying only on structural or regular equivalence. Davis and Leinhardt (1972) explored the structure of small-group networks and suggested there was a tendency toward the formation of ranked-cluster systems. Based on these ideas, Doreian et al. (2000, 2005: Chapter 11) present a "ranked-cluster" blockmodel to study the structure of children's networks and the marriage system for nobility in Ragusa (now known as Dubrovnik) in the eighteenth and nineteenth centuries.

There are some important and useful results from using generalized blockmodeling. First, by setting up a clustering problem with a criterion function that is fully compatible with a type of blockmodel, the criterion function becomes an explicit measure of fit for the blockmodel. Of course, different criterion functions are defined for different types of blockmodels and each is tailored to the type of blockmodel used. This implies that the values of the criterion function cannot be used to compare the fits of different types of blockmodels. The logic of the approach is to formulate models and then fit them to the data by using an appropriate criterion function that is compatible with the blockmodel type. Each blockmodel stands or falls on its own to fit the data. Brusco and Steiner 2006, 2007) present results based on different optimization approaches that suggest the (relocation) heuristic used in generalized blockmodeling usually does return optimal partitions for small networks.

Second, much can be done within the rubric of "formulating models and fitting them." Thus far, we have described an inductive approach to blockmodeling; all that matters is the statement of equivalence, usually structural or regular. Then some clustering algorithm is used to find positions and blocks. The notions of null, complete, and 1-covered blocks are all left implicit and can appear anywhere in the blockmodel returned by an algorithm. Yet, we often know more about a network than the applicability of a particular type of equivalence. The more we know, or think we know, about a network, the more we can "prespecify" a blockmodel. We may know not only the permitted block types but also where some of them are in a blockmodel. (Indeed, we might know where all of them appear in a blockmodel.) If we have some knowledge—which can be simply empirical or it could have a theoretical foundation—it is useful to prespecify a blockmodel that uses this knowledge. When this is done, generalized blockmodeling is used deductively.

A theoretical driven example of deductive blockmodeling is found with structural balance theory (Heider, 1946; as formalized by Cartwright and Harary, 1956). If a signed network is balanced, then the partition structure is known from a pair of structure theorems that state, depending exactly on how balance is defined, that the actors in the network can be partitioned into two positions (Cartwright and Harary, 1956) or into two or more positions (Davis, 1967) such that all of the positive ties are within positions and all of the negative ties are between positions. Doreian and Mrvar (1996) noted that this implies a particular blockmodel structure with two types of blocks. Positive blocks contain only positive or null ties and negative blocks contain only negative or null ties. The implied blockmodel has positive blocks on the diagonal and negative blocks off the diagonal. Doreian and Mrvar proposed a way of partitioning signed networks that are as close as possible to a partition expected for exact structural balance. The criterion function they proposed counts two types of inconsistencies: positive ties in negative blocks and negative ties in positive blocks. These inconsistencies can be weighted differently if desired. The relocation algorithm described earlier is used to solve this clustering
problem and is now an integral part of the generalized blockmodeling approach.

Another element of knowledge is that some actors belong together in a position or that some pairs of actors will belong to different positions. This kind of knowledge is expressed in the form of constraints on which positions actors can be placed within. The extreme level of knowledge is that we know the structure of the blockmodel image and the positions to which all actors belong and this would be expressed in a completely prespecified blockmodel. More likely, prespecification, if possible, will be partial. Examples of some prespecified blockmodels are found in Doreian et al. (2005: 233–46). Of course, without the knowledge for prespecifying a blockmodel, the inductive use of blockmodeling is the preferred option.

Nuskesser and Savitzi (2005) also present a formal overview of blockmodeling based on the conception of a position and link blockmodeling to a wider variety of network representations and methods.

**SOME RECENT EXTENSIONS TO GENERALIZED BLOCKMODELING**

We provide here some examples of ideas and analyses that can be viewed as extending generalized blockmodeling ideas in different ways that hold promise.

**Two-mode and three-mode network arrays**

Borgatti and Everett (1992), in a special double issue of *Social Networks* devoted to blockmodeling approaches, suggested applying blockmodeling ideas to multinetwork arrays and provided a way of doing so. This moved the approach beyond analyzing one-mode networks. This extension can also be formulated as a generalized blockmodeling problem where the network is defined by several sets of units and ties between them. Doreian et al. (2004, 2005) did this for blockmodeling two-mode networks. The examples they used included the classic Deep South data set and Supreme Court voting for a single term. More recently, Batagelj et al. (2006) applied these ideas to three-way network data. In addition to analyzing network data with three distinct types of social objects, they considered two special cases. One allowed that the two of the modes were the same in a three-way array and the second always has the same mode in the three-way array.

Krackhardt’s (1987) office data were used to illustrate the approach. Each of the 22 actors in the network provided their image of the one-mode relational data for the office. When these 22 perceptions are coupled, a full three-way array is created. Even this special case poses severe computational problems for the direct approach that have not been solved. Instead, Batagelj et al. (2007) proposed a dissimilarity measure for structural equivalence for all three cases and adopted an indirect approach. To do this they expressed structural equivalence as an interchangeability condition across modes for three-way networks. This allowed the construction of a compatible dissimilarity measure. Ward’s clustering method was used to obtain the three-dimensional partitioning via hierarchical clustering.

Roffilli and Lomi (2006) proposed a quite different approach to blockmodeling two-mode networks. It is based on a family of learning algorithms called Support Vector Machines (SVM). The analytical framework provided by SVM provides a flexible statistical environment for solving many classification tasks while redefining regression and density estimation problems. They also used the Deep South data and compared their results with other partitions of these data obtained by employing different methods. Their method acts as a data-independent preprocessing step and they were able to reduce the complexity of clustering problems. This reduction in complexity enabled the use of simpler clustering methods.

**Valued networks**

Doreian et al. (2005) confined their attention to binary networks where the ties are simply present or absent. Given the increased collection of valued network data, this is a clear limitation. This makes extending generalized blockmodeling to a valued network a necessary and important development. This task was picked up initially by Batagelj and Ferligoj (2000) and later by Žiberna (2007). Žiberna proposed three approaches to generalized blockmodeling for valued network data by assuming that the values of the ties are measured on at least an interval scale.

The first approach proposed by Žiberna (2007) is straightforward generalized blockmodeling of binary networks to valued blockmodeling. He uses a threshold parameter, and ties are assessed in relation to the value of the threshold (without binarizing the network). Patterns within blocks, as signatures of block types, are still examined to identify block types. One problem that emerges is that two blocks with the same pattern of ties but with different values for the ties that are present
cannot be distinguished. This implies that such differences in tie values cannot be used to locate optimal partitions. This problem led Ziberna (2008) to consider a second approach that he called homogeneity blockmodeling. In this approach the inconsistency of an empirical block, compared with a corresponding ideal block, is measured by variability of values within a block. Ideally, all of the values inside a block are the same. This is fully consistent with the founding idea of blocks being composed of identical ties. (Instead of seeing if ties are present with a value of 1, the homogeneity partitioning establishes blocks with minimum variation in the values of the ties within the blocks.) While this approach helps identify blocks with values that are as homogeneous as possible where ties need to be present, it runs into another problem. For binary blockmodeling there is a clear distinction between null blocks and other block types. Yet a null block is homogeneous (with a value of 0) and cannot be readily distinguished as a special distinctive block type under homogeneity blockmodeling. As a result, while homogeneity blockmodeling is well suited for distinguishing empirical blocks based on tie values and finding partitions based on such differences, it is less suited for distinguishing empirical blocks based on block types and finding partitions based on such distinctions. This led Ziberna to consider implicit blockmodeling of valued networks.

Implicit blockmodeling can distinguish empirical blocks based on tie values and block types. However, it is heavily influenced by the values of block maxima and often classifies blocks differently than it would be desired. As a result, the partitions that it finds are heavily influenced by the classification of blocks and can lead to unsatisfactory partitions. This is especially problematic if block maximum normalization is also used. The partitions that it produces can be improved, but this improvement comes at the price of one of the main advantages of implicit blockmodeling, which is its ability to distinguish between the null block type and other block types.

While the three approaches to blockmodeling valued relations proposed by Ziberna, all have problems that are best viewed as important first steps toward establishing better solutions to this class of network partitioning problems. They were proposed, in part, as a response to some results Ziberna obtained while examining an obvious strategy for blockmodeling a valued network. This is to select some threshold and use it to binarize the network: ties at, or above, the threshold are coded 1 while ties below the threshold are set to 0. The binarized networks were then treated with the approach advocated by Doreian et al. (2005).

Ziberna suggests binarization is a poor first step of a general strategy because the established partitions can be unstable and produce different blockmodels depending on the threshold selected. It follows that using one of the three types of generalized blockmodeling of valued networks proposed by Ziberna is preferable to using binary blockmodeling whenever we have networks measured on an interval scale. Generalized blockmodeling of valued networks produces better partitions and fewer equally well-fitting partitions because it measures the block inconsistencies more precisely. The problems encountered reflect the brute fact that moving from binary networks to valued networks moves us to a difficult set of partitioning problems.

We note that for structural equivalence, the indirect approach is available because similarities like correlations and dissimilarities like Euclidean distances can be computed for valued network data as long as they are compatible with structural equivalence. This is not the case for regular equivalence where we still do not have a widely accepted way of computing the extent to which vertices in networks are regularly equivalent.

Nordlund (2007) also tackled the problem of partitioning valued networks in terms of regular equivalence and his argument is consistent with Ziberna: applying techniques appropriate for binary networks directly to valued networks is problematic. He proposed a formal heuristic for viewing ties as regular based on their linkages given the role set of actors. He combined this idea with measures for block criteria fulfillment to establish reduced graphs where methods were more sensitive to patterns of ties rather than their strengths.

Weber and Denk (2007) proposed an approach for valued blockmodeling for input-output relations viewed as networks. Flows between industrial or economic sectors, as an aggregation of flows between businesses, are clearly valued. And the flows that go into national input-output tables can be analyzed also at a "lower" level with businesses as the units. Regardless of whether data for these flows express the volume of goods and services flowing or their monetary values, it is foolish to even think of these data as being binary. So if blockmodeling is contemplated, it has to deal with valued data whose values can vary greatly. Given that goods and services flow between units, it is natural to think of these data as two-mode data with the rows as transmitters (exporters) and the columns as receivers (importers). Doreian et al. (2005: 265–69) did exactly this for journal-to-journal citation networks. For business-to-business transaction patterns, businesses can remain a part of a trading network or leave while other businesses can join
the trading flows. Studying such economic networks includes evaluating businesses, characterizing the overall structure of these networks, and clustering both units and relations. It is possible to identify gaps, both in the form of missing actors and missing ties, in these networks. Additionally, the flows are not simply between pairs of units because indirect flows are important and merit attention. While indirect paths of varying length have been considered by network analysts thinking of what flows from one unit to another and then on to a third unit in social networks is much harder to conceptualize than for economic flow networks.

Stochastic blockmodeling

The generalized blockmodeling approach, as described above, is explicitly deterministic. A criterion function expressing the core of a clustering problem is minimized to determine the "best" partition(s) given the criterion function. An alternative strategy is to adopt a probabilistic approach and treat the underlying processes in a stochastic fashion. Mirroring structural equivalence as an underlying conception, two actors are stochastically equivalent if they have the same probability distribution of their ties to other units (Holland et al., 1983; Wasserman and Anderson, 1987; Anderson et al., 1992). Using probabilities for ties established from the data, positions are populated with stochastically equivalent units.

Nowicki and Snijders (2001) developed a Bayesian approach for stochastic blockmodeling where the units are dyads. The ties can have categorical values and the parameters of their model are estimated by a Markov chain Monte Carlo ("MCMC") procedure. Two features of their approach are noteworthy for this discussion: (1) missing data can be handled and (2) given the network data, some vertices may be unclassifiable into a position. Most discussions of blockmodeling quietly ignore the problems of missing data, which can have serious implications (see below). The idea of there being some units not belonging to positions was first noted by Burt (1976) in the form of a "residual cluster" and this, too, has been ignored in generalized blockmodeling.

Airola et al. (2007a, 2007b) introduced a family of stochastic blockmodels that combine features of mixed-membership models and blockmodels for relational data in a hierarchical Bayesian framework. They proposed a nested variance inference scheme for this class of models, which is necessary to successfully perform fast approximate posterior inference. Handcock et al. (2007) proposed a new model with latent positions under which the probability of a tie between two units depends on the distance between them in an unobserved Euclidean "social space," and the locations of units in the latent social space arise from a mixture of distributions, each corresponding to a cluster. Handcock et al. proposed two estimation methods: a two-stage maximum likelihood method and a fully Bayesian method that uses MCMC sampling. They also proposed a Bayesian way of determining the number of clusters that are present by using approximate conditional Bayes factors.

Generalized blockmodeling of signed networks

As noted above, structural balance theory implies a distinctive blockmodel structure for signed social networks. Yet structural balance as a process, or a set of processes, is not the only force creating signed social relations between human actors. It is quite possible that there are some actors that are universally liked (or liked by most members of a group), despite the presence of the kind of divisions predicted by structural balance theory. If this is present then, there will be positive blocks off the main diagonal of an image matrix. From the perspective of conflict resolution, when social groups are completely split into mutually hostile subgroups, with positive ties only within these groups, this is not a good situation because there are no actors in locations to mediate conflicts between these subgroups. Mediators would have positive ties to members of at least two of the mutually hostile subgroups. But were there to be one or more mediators between a pair of opposed subgroups this would also imply positive blocks off the image matrix's main diagonal. Finally, there are groups where there are some mutually hostile individuals and their presence would imply a negative block on the main diagonal of the image matrix. Structural balance, with its signature blockmodel structure, would bury all of these features into inconsistencies with structural balance. If these other processes left traces in the structure of a group, they could not be identified using structural balance.

To deal with these problems, Doreian and Mrvar (2009) relaxed the specification of structural balance by allowing positive and negative blocks to appear anywhere in a blockmodel. They retained the same criterion function as used for structural balance and called this relaxed structural balance. They proved that this is a proper generalization of structural balance and applied this revised version of balance to some of the classical signed social network data sets and obtained blockmodels with better fits to the data.
The original Heiderian balance theory had two types of relations. One type was social relations between people and the other took the form of "unit formation" relations between people and social objects (like values and beliefs). The generalization provided by Cartwright and Harary (1956) buried this distinction and used only signed relations. In effect, the unit formation relations were discarded. While the formalization changed dramatically the study of signed social relations in groups and permitted great progress, something was lost. Mrvar and Doreian (2009) formalized the idea of unit relations as signed two-mode networks and extended the relaxed balance partitioning algorithm to partition-signed two-mode data. The primary empirical example featured the voting patterns of Supreme Court justices for one term.

**Partitioning large or complex networks**

A complementary way to generalized blockmodeling for large networks was proposed by Hsieh and Magee (2008), who presented another algorithm for decomposing a social network into an optimal number of structurally equivalent classes. The k-means method is used to determine the best decomposition of the social network for various numbers of positions. This best number of positions is determined by minimizing the intra-position variance of similarity subject to the constraint that the improvement in going to more subgroups is better than partitioning a random network would achieve. They also describe a decomposability metric that assesses how closely the derived decomposition approaches an ideal network having only structurally equivalent classes.

Reichardt and White (2007) presented a framework for blockmodeling the functional classes of agents within a complex network. They derived a different measure for the fit compared to one used by Batagelj, Doreian, and Ferligoj of a network to any given blockmodel. Their method can handle both two-mode and one-mode data, and directed and undirected as well as valued networks, and allows for different types of links to be dealt with simultaneously. They applied their approach to a world trading network and were able to establish the roles played by countries as occupants of positions in world trading.

Wang and Lai (2008) picked up the problem of detecting positions and hence blockmodels in a complex network. Using the mixture models and the exploratory analysis presented by Newman and Leicht (2007), they developed an algorithm that is applicable to a network with any degree distribution for the vertices in the network. The language of "finding communities" in a network can be seen as a variant of identifying positions in networks but with some differences. In general, clusters of similar components are not necessarily identical with the communities in a community network; thus partitioning a network into clusters of similar components provides additional information of the network structure. Their proposed algorithm can be used for community detection when the clusters and the communities overlap.

By introducing a parameter that controls the involved effects of the heterogeneity, they investigated how the cluster structure can be coupled with the heterogeneity characteristics. They show how a group partition can evolve into a community partition in some situations when the involved heterogeneity effects are tuned. Their algorithm can be extended to valued networks.

Recently, two approaches were proposed for blending blockmodeling with graph theoretical constraints. At the Dagstuhl Seminar 08191 in May 2008, a group of participants (Batagelj et al., 2008) started developing a general framework for graph decompositions. Kemp and Tenenbaum (2008) also proposed an approach based on graph grammars.

**OPEN GENERALIZED BLOCKMODELING PROBLEMS**

There are a wide variety of open problems to merit attention. Some of them are general in the sense that they are relevant for all generalized blockmodeling problems. Doreian (2006) provided a partial characterization of these general problems. Other open problems pertain to specific applications. To some extent, this classification is arbitrary and we do not imply that the general problems are more important for the future of generalized blockmodeling. Both contain problems where their solutions will constitute advances for using generalized blockmodeling to identify positions and roles. We first discuss some specific open problems before moving on to consider some general open problems.

**Open specific generalized blockmodeling problems**

*Regular equivalence*

One of the vexing problems with regular equivalence is that while it has a conceptual superiority to structural equivalence as a definition
of role, it seems to be less successful empirically. As noted earlier, there is no compatible measure of the extent to which two locations are regularly equivalent, and this handicaps the indirect approach to blockmodeling networks in terms of regular equivalence. From the vantage point of generalized blockmodeling, there are two large issues. One is that so many blocks are consistent with the permitted 1-covered block type. They range from one block of structural equivalence to blocks with just one 1 in each row and column. This is a wide range of variation to include within a single definition of regular equivalence. One consequence is that many equally well-fitting blockmodels for a given number of positions can be identified in a given network under regular equivalence without nonarbitrary ways of selecting one of them. The second stems from the deep result of Borgatti and Everett (1989) that each network has a lattice of regularly equivalent partitions. Sometimes these lattices are trivial but in most cases they are not. This raises an obvious question for generalized blockmodeling. Given multiple exact partitions of a network that are consistent with regular equivalence, which is the most appropriate for an ideal block model when trying to establish an empirical blockmodel based on regular equivalence? A formal approach to role structures based on regular equivalence is provided by Lerner (2005).

Partitioning signed networks
A general criticism of the use of the relocation algorithm in the direct blockmodeling approach is that there is no guarantee that the use of the method provides the optimal solutions or yields all of the equally well-fitting partitions for a particular network. Brusco et al. (2010) provide convincing evidence that for “small problems” (where the number of vertices is less than 30 to 35) the relocation method does return all of the optimal partitions of a signed one-mode network. This was done through the comparison of the performance of the relocation method with a branch and bound algorithm that is guaranteed to return all of the optimal partitions. However, while the relocation algorithm can handle much larger signed networks, the branch and bound method cannot. So the guarantee cannot be extended to larger networks. Also, the two algorithms are not identical in terms of the difficulties they encounter with different network features such as size, density of ties, and the relative proportions of positive and negative ties. Brusco et al. (2010) advocate using both algorithms, rather than relying on only one, whenever this is possible. And the idea of a guarantee, at this time, has not been extended to signed two-mode networks. Potential applications of two-mode signed partitioning of two-mode networks include, in addition to Supreme Court voting patterns, U.S. congressional voting, voting in other deliberative bodies, and voting at the United Nations. For all of these potential examples, the networks are large and very dense. There is a clear need for establishing sound ways of partitioning these large networks.

The general method proposed for partitioning signed networks (Doreian and Mrvar, 1996, 2009; Mrvar and Doreian, 2009) allows for the differential weighting of positive inconsistencies in negative blocks and negative inconsistencies in positive blocks. However, the differential weighting has not been explored in any detail and the implications of using this differential weighting needs to be mapped for empirical networks. When one sign predominates, it may be necessary to weight the two types of inconsistencies differentially. The best ways of doing this are not known. It seems that potential null blocks in signed networks also have relevance and need to be specified. General ways of doing this in a principled fashion are not known and establishing such methods forms another open problem.

Open General Problems for Blockmodeling

Boundary problems
The blockmodeling approach is avowedly positional because the location of an actor is defined as the pattern of ties to and from all other locations in the network. This implies that identifying the boundary of a network correctly is very important. Laumann et al. (1983) pointed to “the boundary problem” as important for network analysis in general. However, its importance for positional approaches to the analysis of network structure is particularly acute. We simply do not know the implications of an incorrect identification of a network boundary for generalized blockmodeling beyond the intuition that the approach is highly vulnerable to identifying the boundary incorrectly. Establishing some bounds of the sensitivity to identified blockmodels (as positions and roles) is an important open problem.

Measurement errors
We also have a limited understanding of the vulnerability of both generalized and classic blockmodeling to measurement errors. This problem takes two forms. One is that data can be
missing and that often, perhaps too often, a tie that is recorded as a null tie is really a case of missing data. The other is that, while some value for a tie is recorded, the actual value recorded may be of the wrong magnitude. The presence of the first type of measurement error can affect seriously the blockmodeling of both binary and valued networks and the presence of the second is most acute for blockmodeling valued networks. Obtaining a better understanding of the vulnerability of establishing blockmodels to errors of measurement is an important open problem. There are proposals for imputing values for missing data (e.g., Huisman, 2009) and we may need to assess the impact of imputation methods for missing data on delineated blockmodels or roles and positions. There are two broad ways of making these assessments. One is through controlled simulations where variations in (fictitious) recorded data are generated and the impacts on blockmodeling results are assessed. The other is to start with real data and introduce controlled amounts of measurement error and examine the results (Doreian, 2006).

Assessing fits of blockmodels
In making comparisons between partitions, based on either structural or regular equivalence, using direct and indirect blockmodeling approaches, Doreian et al. (1994) showed that the minimized values of the criterion functions for the indirect approach were never lower than those obtained via the direct approach. And, in many cases, the values of the criterion functions for partitions obtained when using the indirect approach were much higher than for the direct approach. However, their argument is partially circular as their methods are designed to minimize criterion functions that are defined for a particular equivalence. There may be alternative and better criterion functions than the ones they considered. The analyses of Reichardt and White (2007) reopen this issue, for they use a different criterion function for a particular definition of equivalence than the corresponding one used by Doreian et al. (1994, 2005). Establishing better criterion functions for many types of blockmodels remains an important open issue.

The value of the minimized criterion function for structural equivalence for empirical networks can be quite large. When the value of the minimized criterion function is zero, or close to zero, it is clear that the delineated blockmodel fits. While the criterion function has been minimized, a high value of the criterion function often stems from the stringency of it counting all discrepancies with null blocks or complete blocks in empirical networks. There is no obvious guarantee that, with a "high" value of the criterion function a delineated partition "really" is consistent with the idea of structural equivalence. Establishing better bounds is an important open problem and insights may well come from fitting blockmodels based on stochastic equivalence of the sort suggested by Nowicki and Snijders (2001). An impression about the nature and quality of obtained partitions can be gained also from where the value of their criterion functions lie in the distribution of values of criterion function obtained using the Monte Carlo method.

Blockmodeling large networks
For generalized blockmodeling, using the direct approach, the predominant methods are based on local optimization algorithms. However, the current versions of the algorithms can handle networks having some hundreds of units. Even for networks of this limited size we do not know if they will return all of the optimal partitions. And their use is impossible for larger networks. To make progress we need to formulate blockmodels that can be fitted for large networks and develop faster algorithms, or we need more effective heuristics. Indirect approaches appear to be much more useful for large networks, yet their applicability is for a very restricted range of equivalence types. In the main, structural equivalence - while very restricted - is the only viable option for partitioning large networks as a whole. However, various forms of preprocessing of large networks generates reduced networks that can be analyzed further with regard to positions and roles. Those based on different connectivity decompositions - for example, weak components, strong components, graph condensation, symmetric-acyclic decomposition, bi-connected components - can be determined very efficiently.

Numbers of positions
Determining the number of positions is a difficult problem for generalized blockmodeling even when the networks are not large. The methods proposed by Handcock et al. (2007) and by Reichardt and White (2007) both include a way of establishing the number of positions empirically from the data. This seems important as a general feature to be included in blockmodeling. Yet the most useful way of establishing the number of positions for a blockmodel may rest on a sound understanding of the substantive processes driving network tie formation and the generated network structures in given empirical contexts.
For example, friendship formation in a school system seems most constrained or driven by the classes and levels into which students are distributed. The best guess for the number of positions, most likely, is the number of grades in the school system.

**Dynamic blockmodels**

The origins of blockmodeling are located in attempts to analyze and understand the operation of role systems. Blockmodeling was established as an effective way of doing this because of its ability to identify positions and roles. As such, it became a useful empirical method for partitioning social networks. Over time, this meant that it could be used for partitioning any network, even networks where there appeared to be no obvious connection to role systems. Yet the notions of positions and roles acquired a generalized form that led to establishing well-defined positions, roles, and role structures in many networks. If there is an attempt to understand how networks, as role structures, form and change, there is a deep problem requiring a solution.

The presumption underlying most attempts to delineate network structures using blockmodeling ideas is that the “surface networks” we observe are the manifestation, or indicators, of an underlying “more fundamental structure” of the network. Blockmodeling, regardless of the specific forms used, allow us to identify this fundamental structure. And we can do this at various points over time. If the structure remains the same over the observation period this is useful knowledge. And if it changes over time then it is possible to create a sequence of blockmodels that capture the fundamental structure as it changes. This, too, will be useful knowledge but the sequence of blockmodels fitted at successive time points is only a sequence of descriptions. If a social structure as a network really is changing then it is the fundamental structure that is changing, with the observed changes being indicators of the underlying fundamental change. If these changes are not just random events, then we need to account for them. When a system is evolving over time, most likely, there will be coherent process rules driving these changes. Mere descriptions of the changes involved, even if couched in terms of blockmodels, seem insufficient. We need to understand the processes generating structural change and this implies understanding how blockmodels, as representations of positions and role systems, evolve. This is the biggest open problem for generalized blockmodeling and understanding the operation of role systems.

**REFERENCES**


