Multivariate Methods for Inter-Structural Analysis

Carter T. Butts*
Department of Social and Decision Sciences
Center for the Computational Analysis of
Social and Organizational Systems
Carnegie Mellon University

Kathleen M. Carley
Department of Social and Decision Sciences
H.J. Heinz III School of Public Policy and Management
Center for the Computational Analysis of
Social and Organizational Systems
Carnegie Mellon University

March 31, 2001

Abstract

Inter-structural analysis – the analysis of data sets consisting of multiple social structures – is an important extension of classical social network analysis. Building on past research regarding metric inference and graph covariance analysis, we here present a general approach to the analysis of social structure sets with arbitrary labeling assumptions. We define a family of structural distance and analogous structural covariance measures, and show that these measures satisfy the conditions necessary for their use in conjunction with traditional multivariate analysis procedures. Demonstrations of the use of various exploratory data analytic procedures on inter-structural data sets are provided, and suggestions are made regarding the further development of this approach.

Keywords: multivariate methods, metric inference, unlabeled graphs, structural comparison, graph covariance

*This material is based upon work supported under a National Science Foundation graduate fellowship, and was supported in part by the Center for the Computational Analysis of Organizational Systems and the Institute for Complex Engineered Systems at Carnegie Mellon University.

1 Introduction

A central problem for the structural analyst is the comparison of disparate social structures in order to identify underlying commonalities among them. This is especially true where large populations of structures are concerned (e.g., group studies (Hare et al., 1965), organizations (McKelvey, 1982; Hannan and Freeman, 1989)), mental models (Carley and Palmquist, 1992), or the like); in such cases, it is often reasonable to ask whether there may be groups of structures which are especially similar to others within the population. Likewise, in measuring structural similarity, it is often important to contrast the similarities and differences which are due to assignment of particular individuals (or types of individuals) to similar positions, versus similarities and differences which arise from underlying structure. Such questions, however, raise fundamental issues regarding the meaning of "underlying structure," and of "similarity" between structures. These issues, ultimately, are central to both the theory and practice of inter-structural analysis.

For a variety of problems of interest, there is no theoretical justification for treating some or all structural elements as *á priori* distinct from one another. This lack of distinction is equivalent to a notion of *exchangeability*: certain elements may be exchanged within the structure without doing violence to the theoretical basis for comparison[1]. This situation, common as it is, poses serious difficulties for comparative work. In particular, the methods which exist for directly assessing the differences between structures are based on oriented representations such as labeled graphs; unlabeled or unoriented structures (the sort implied above) cannot be treated in this way.

While previous research has given us a valuable set of tools for the dyadic comparison of social structures and for one type of point estimate of central tendency within graph sets (the central graph), we have had relatively few techniques for addressing the broader question of general tendencies towards similarity within large graph sets. For instance, given some large collection of social structures, we would like to be able to determine the extent to which the larger set may be expressed in terms of a smaller set of highly similar structures. In traditional data analysis, such questions are commonly answered via methods of *cluster analysis* (Johnson, 1967; Ward, 1963), which identify groups of observations satisfying various properties of within-group similarity and between-group difference[2]. Another, related, question is that of the identification of archetypal structures in terms of which the structures of a given data set may be expressed; this problem is roughly analogous to those addressed classically by principal component analysis (Jackson, 1991; Joliffe, 1986). In general, then, we find that a number of problems arise in social network analysis which have clear classical analogues, but for which classical methods have not previously been applied due to their differing assumptions (e.g., exchangeability of observations, distributional

[1] Exchangeability of *observations* is a common assumption in classical statistical analysis; the motivation for this assumption is often similar, though exchangeability of observations and exchangeability of structural elements have very different implications.

[2] The various methods of cluster analysis typically vary primarily by the notion of similarity/difference employed, by the manner in which similarities and differences are weighted, and by whether the procedures employed are hierarchical. See Romesburg (1984) for a fairly pragmatic overview.

requirements (especially normality), non-permutability of labels across variables[3]).

Here, we develop a number of techniques whose basis lies in traditional multivariate analysis for the purposes of analyzing structural data. Unlike typical multivariate analyses, of course, our "variables" will often represent entire social structures (rather than vectors of attribute values), a fact which will introduce some special concerns. Nevertheless, we here demonstrate that multivariate analysis of both labeled and underlying (i.e., unlabeled or unoriented) social structures is both possible and informative, illustrating the use of several such methods on a number of classic network data sets.

2 The Problem of Structural Comparison

Just as the analysis of positions within individual social structures poses particular theoretical and methodological challenges for the social researcher, so too does the problem of structural comparison pose difficulties of its own. Here, we briefly consider four basic perspectives on the structural comparison problem, along with some of the merits and liabilities of each. After reviewing these approaches, we will then move to some data considerations specifically relating to the analysis of multiple distinct structures. The discussion of these issues – particularly that of theoretical exchangeability of vertices – will provide important foundational material for the sections which follow.

2.1 Four Approaches

Although, as noted, the paucity of large multi-structure data sets has heretofore limited the need for methods of structural comparison, four primary approaches currently exist in the network literature. These approaches are comparison of graphs using graph-level indices (GLIs), comparison via algebraic decomposition, comparison using parameters of fitted edge models, and direct comparison of edge sets. While certain other methods (e.g., comparison of node-level index (NLI) distributions, visual inspection and comparison of blockmodels induced by nodal equivalences) have also been employed for purposes of structural comparison, these applications have thusfar been sufficiently ad hoc to preclude discussion here. Though all of the methods presented here belong to the fourth category (the direct comparison approach), it is useful to understand these techniques in the larger context of inter-structural analytic methods. Each approach to the comparison of social structures has its own advantages and limitations, and the structural analyst should employ the methodologies which are most appropriate for his or her theoretical problem.

[3]This generally tacit assumption is routinely violated by exchangeability of structural elements, e.g. in the case of graph correlations between unlabeled graphs.

2.1.1 Comparison Using Graph-Level Indices

Much of what we shall here call "classical" social network analysis[4] involves the construction and computation of structural indices – deterministic functions of graphs or positions which express structural properties of theoretical interest – and it is therefore unsurprising that one standard means of comparing structures is by computing one or more such indices and subjecting the resulting data to statistical analysis. Where motivated by substantive theory, GLIs such as density, reciprocity, various forms of centralization and hierarchy, number of components, diameter, etc. can simplify data analysis by permitting direct consideration of specific structural features (for an experimental example, see Freeman et al. (1980)). Similarly, specific families of structures may be characterized by a particular set of GLIs, as with Krackhardt's set of outtree "dimensions" or organizational structure (Krackhardt, 1994). For problems in which the salient features of network structure are well-known (and readily quantified), then, analysis of structural populations using indices is an entirely reasonable approach.

In contrast with the above, however, we cannot always reduce our properties of interest to a small number of unidimensional quantities. Likewise, the behavior of GLIs is not always conducive to such analyses. Path-based indices, for instance (i.e., betweenness and closeness distributions (and hence centralization scores), diameter), are sufficiently computationally expensive to render calculations on large numbers of high-order graphs laborious, and indeed the calculation of some indices (e.g., clique number) is NP complete. Furthermore, the distributional properties of many GLIs are poor, changing shape dramatically and losing information as graph size increases (Anderson et al., 1999). These factors limit the use of structural indices in many real-world contexts, where they may be difficult to calculate[5] and/or largely uninformative. As many structural indices interact with each other in complex and decidedly nonlinear ways, analyses of such indices with conventional statistical techniques must be performed with caution (Anderson et al., 1999). While some specialized, model-based approaches can aid in disentangling the effects of multiple structural indices (see below) these cannot completely obviate the problem[6], and it is still necessary for the analyst to specify *ex ante* a set of GLIs which adequately capture the dimensions on which

[4] "Classical" social network analysis as we use the term includes visualization of sociograms, utilization of node and graph-level indices, component/clique/group identification, nodal equivalence analysis, algebraic decomposition of relational structures, and walk/path distance analysis (see Wasserman and Faust (1994) for a canonical treatment). Classical SNA has given rise to offshoots such as ego network analysis (which focuses on local network properties), network statistics (which applies stochastic modeling to problems of structural inference), network process modeling (which examines processes affecting or affected by social structures), cognitive social structure analysis (which studies individual perceptions of social structure), and macrostructural analysis (which considers structural properties of relations on large, spatially embedded populations). Most later developments in network analysis build explicitly upon classical techniques, but differ in their adoption of novel approaches designed for particular problems of structural analysis.

[5] Or impossible, in some cases: as conventionally deployed, measures such as closeness are undefined on disconnected structures, and some others (such as eigenvector centrality) lose their normal interpretations in such circumstances (Wasserman and Faust, 1994).

[6] The problem being largely that it is meaningless to speak of "independent" effects of structural indices which are *inherently* related to one another; see Anderson et al. (1999) for a discussion of this issue.

the graphs in question are to be compared.

2.1.2 Comparison via Algebraic Methods

Another approach to the problem of inter-structural comparison is algebraic; algebraic methods (as we use the term here) attempt to express certain members of a graph set in terms of a set of operations applied to other set members, and/or attempt to find common algebraic constraints within structures which are shared by many members of the graph set. These approaches have seen their most complete development in research on kinship networks (White, 1963; Boyd, 1969, 1990), where the technique has been used to develop algebraic characterizations of cultural norms regarding family structure (Boorman and White, 1976), but have been applied to a number of other areas including organizational structure (White, 1963) and interactions among elites (Breiger and Pattison, 1978). More limited variants include the familiar work on structural equivalence (Lorrain and White, 1971), transitivity (Holland and Leinhardt, 1971), and structural balance (Harary, 1953). As these examples suggest, the great strength of the algebraic paradigm is its ability to precisely define structural constraints which are at once general (being applicable to a wide range of structures at multiple scales) and precise. Where algebraic reductions exist, one can very parsimoniously express the relationships among multiple structures, and in some cases it may even be possible to identify a set of basic structural elements from which (by application of the algebraic operators) an entire set of structures may be formed. In certain instances, shared internal algebraic constraints (e.g., obeyance of transitivity) may provide clues as to the processes from which networks arise, and/or cultural or other constraints (e.g., marital taboos) faced by actors within the social system. When applicable, then, it is quite clear that the algebraic approach to inter-structural comparison can be quite powerful.

Historically, the greatest difficulty with algebraic analysis of graphs has been its inflexibility; even minor deviations can be sufficient to "break" a hypothesized pattern. This situation may be changing: recent work by Pattison et al. (2000) has provided a basis for statistical inference of *approximate* algebraic structure within graphs, which will doubtless increase the empirical viability of this approach[7]. Nevertheless, there are many contexts of theoretical interest in which algebraic decomposition seems unlikely to prove helpful. Sets of independently collected group structures, for instance (e.g., organizational forms, friendship patterns in small groups, etc.) are unlikely to be algebraically related, and yet analysis of such structural populations is clearly motivated by problems in neo-institutional theory, human factors and organizational design, group behavior, and the like. Questions of central tendency, graph distribution, and underlying dimensions of structural similarity cannot readily be posed in an algebraic framework, and there do not currently exist feasible exploratory methods for detecting nontrivial algebraic structure within graph sets. In the absence of such techniques, algebraic relations must generally be postulated and evaluated by the researcher on a case-by-case basis. This, and the above considerations, limits the utility of the algebraic approach to inter-structural comparison in most current research

[7] See also the additive models implied by structural regression, described below.

settings.

2.1.3 Model-Based Comparison

Yet another approach to the problem of inter-structural analysis is that of model-based comparison. Methods of this family involve the specification of a specific stochastic model by which the structures in question are presumed to be well-predicted; this model is then fit to the data, and variations in model parameters across graphs are used to assess differences in structure (Feinberg et al., 1985; Wasserman et al., 1990). The most well-known model family which can be deployed for this purpose is the log-linear formulation known generically as p^*, along with its logistic variants (Wasserman and Pattison, 1996; Pattison and Wasserman, 1999; Robins et al., 1999). Under the p^* family, edge set probabilities are modeled as exponential functions of a linear predictor, the elements of which can include both structural indices (GLIs and NLIs) and supplemental covariates. (In the logistic variant, the log-odds for each edge are modeled by the *change* in the linear predictor induced by said edge's presence or absence (holding all other edges constant). Given certain regularity conditions, this model converges asymptotically to the full log-linear formulation. (C. Anderson et al., 1999).) Another very large family of stochastic models is based on the Bernoulli graph formulation, in which edges are taken to be independent Bernoulli trials conditional on a (possibly dependent) matrix of edge probabilities. Many of the biased-net formulations of Rapoport (1957; 1979), Fararo (1981; 1983; with Skvoretz, 1984), and Skvoretz (1985; 1990) are of this type, as are the spatial network models of Butts and Carley (2000). In all of these various guises, model-based comparison techniques can serve to connect theories regarding structural influences and/or constraints with structural data sets. Given an appropriate model, the researcher may infer the effects of particular predictors on network structure (e.g., homophily, distance, or popularity). Likewise, where competing explanations are present, model-based comparison allows for alternatives to be explicitly formulated and compared side-by-side. In this respect, then, model-based comparison in the context of inter-structural analysis is no different from the use of explicit stochastic modeling in other scientific contexts.

While model-based comparison is a powerful and useful approach to inter-structural analysis, it has certain limitations as well. Most important among these is the specification of the model on which the analysis is based. Log-linear models, for instance, rest on the assumption that a linear predictor can be related (via the appropriate functional form) to edge probabilities; where the relationship between covariates and edge probabilities cannot be specified in this fashion, such models will produce misleading results. Similarly, failure to specify the appropriate set of covariates can result in biased estimation of coefficients, and high degrees of multicollinearity between certain types of covariates (e.g., multiple centrality measures[8]) can lead to unstable coefficient estimates for insufficiently large structures (as

[8]It is well-known that average correlations between centrality measures are quite high for most graphs, an effect which can *worsen* with increasing network size; recent work by Everett et al. has, in fact, suggested that this effect is reliable enough to allow degree to be used as an effective proxy for betweenness in certain large structures.

well as biased estimates in the presence of measurement error). For this reason, model-based comparison is most appropriate where structural processes are well-understood and hypothesized effects are clearly specified. Other approaches may be more appropriate for the initial, exploratory stages of research.

2.1.4 Direct Comparison of Edge Sets

A fourth approach to graph comparison attempts to treat the problem by establishing a mapping between the edge sets of graph pairs and then proceeding to evaluate the similarities or distances between the sets; as the edge sets contain the structural information of the graph, such an approach provides a direct means of assessing similarities and differences between structures. Currently, the best-known methods in this family are the graph correlation and network regression[9] techniques introduced by Krackhardt (1997, 1998) for the analysis of labeled graphs, themselves closely related to the methods of cophenetic correlation employed in numerical taxonomy for comparing classification structures (Farris, 1969). More recently, work by Banks and Carley (1994), Sanil et al. (1995), Butts and Carley (1998), and Butts (1998) has focused on the use of metric distances (particularly the Hamming distance) as a useful tool for comparing both labeled and unlabeled graphs. What both of these lines of research share is their focus on the copresence of edges, and their treatment of structures as whole entities rather than as collections of indices and/or model parameters.

The edge-set comparison approach is powerful, but (as currently implemented, at least) it has limitations of its own. Because these methods seek to compare graphs *in toto*, they are not attuned to *particular* structural properties; thus, an analyst who knows *ex ante* which properties are of theoretical importance for his or her purposes might be better served by an approach (such as direct index comparison or model fitting) which examines *only* these dimensions. The methods presented here, in particular, are primarily data-analytic, and are best suited for exploratory analyses. After general distributional patterns have been assessed (e.g., by the identification of clusters of typical structures or of canonical sub-structures), more focused tools can then be employed to tease out specific interstructural relationships.

2.2 Data Considerations for Inter-structural Analysis

The grist for the proverbial mill of inter-structural analysis consists of one or more sets of social structures, representing various relations among various objects of sociological interest. Throughout this paper, we shall refer to the set of structures to be examined as the *structure set*, S, whose elements are taken to be graphs (G). (Except as noted otherwise, we will take the term "graph" here somewhat generically to refer to simple graphs, directed graphs, and valued graphs, with or without loops; in particular we will

[9] It might be noted that the inferential (as opposed to data-analytic) use of network regression is also a model-based technique (and is somewhat related to p^*, in particular); thus, the categories we have imposed here should be thought of as at least somewhat permeable.

assume directedness and allow loops unless otherwise indicated[10].) In order to compare these individual structures, we must consider certain properties of the graphs in question which do not commonly arise in classical network analysis: exchangeability of vertices (non-unique labeling); exchangeability of edge variables across structures; edge value distributions; and heterogeneity in network size. Before advancing to the proposed methods, then, we first examine some preliminary considerations regarding data to be employed for inter-structural analysis.

2.2.1 Exchangeability of Vertices (Labeling)

As we have suggested, a potential complication of inter-structural analysis lies in the fact that the elements represented by vertices – be they human actors, organizations, articles, groups, or any other objects of theoretical interest – are not always distinct with respect to substantive theory. A structural theory of gendered interaction, for instance, will certainly treat males and females as distinct, but may not discriminate between *particular* males or females. Similarly, a Marxist analysis may treat individuals as interchangeable within social classes, and a general theory of group structure may consider *all* actors to be interchangeable! The same, of course, goes for organizations: one may have an *a priori* reason to view non-profit corporations, government agencies, and for-profit firms as qualitatively distinct classes, but one's theory may not make differential predictions regarding structural elements within these classes. In the most extreme cases, we may in fact be interested only in underlying structures corresponding to the unlabeled graphs associated with observed structures. In the unlabeled case, obviously, no vertices are distinct.

Such circumstances suggest a form of *exchangeability* among vertices: on the basis of our substantive knowledge, we cannot always identify a uniquely permissible labeling of vertices within the networks under study. While this might not seem problematic at first blush, it is an unfortunate fact that many theoretically reasonable means of comparing structures (including Hamming distances and graph correlations, as we shall see) are extremely sensitive to vertex labeling. To appreciate why this is so, one need only recall that a relabeling of vertices within a graph is equivalent to a permutation of the rows and columns of said graph's adjacency matrix[11]. Any measure, then, which can be affected by such a permutation will be sensitive to choice of labeling, and an inability to identify a unique vertex labeling will thus imply a lack of unique values on these measures as well.

Having set out a motivating intuition, we now proceed to describe the vertex exchangeability problem in more formal terms. We begin with the concept of exchangeability itself, as it applies to the inter-structural comparison problem. As described above, our notion of exchangeability is built on the idea of theoretical interchangeability; thus, we assume that we possess some *a priori* theory which makes predictions on some set of observables (here referred to as the theory's *scope*), conditional on the observed structure(s). Two vertices, then, will be exchangeable if and only if "swapping" them within the structure does not

[10] In symmetric and/or loopless cases, the relevant edges should be omitted from the analysis.

[11] And, in fact, we shall use the terms "labeling" and "permutation" interchangeably throughout this paper.

alter the predictions of the theory of interest. Formally, we define our notion of theoretical exchangeability as follows:

Definition 1 (Theoretical Exchangeability). Assume a graph $G = \{V, E\}$ and a theory $\mathcal{T}$, and let $L(G)$ represent a permutation (relabeling) of the vertices of G such that $L(v_i) \rightarrow v_j$ and $L(v_j) \rightarrow v_i$ for some $v_i, v_j \in V(G)$, and such that $L(v) \rightarrow v \ \forall \ v : v \in V(G), v \notin \{v_i, v_j\}$. The two vertices v_i and v_j, then, are said to be *theoretically exchangeable* iff $p(y|L(G), \mathcal{T}) = p(y|G, \mathcal{T}) \ \forall$ observables y within the scope of $\mathcal{T}$.

Note that the above can be seen as being closely related to the more standard notion of exchangeability, wherein parameters $\theta_1, \theta_2, \ldots \theta_n$ are said to be exchangeable in their joint distribution if and only if $p(\theta_1, \theta_2, \ldots \theta_n)$ is invariant to all permutations of the $1 \ldots n$ subscripts. Our application is less restrictive in that only the conditional distributions of the observables need be invariant, and in that we explicitly condition on our prior theory in all cases. Note too that our concept of "theory" can be thought of as resembling a Bayesian notion of prior knowledge combined with a set of scope restrictions on the predicted observables[12], although the prior "knowledge" in this case is taken to be hypothetical (possibly reflecting mere working assumptions) and is not assumed to reflect the actual beliefs of any particular observer. With these notions, we can proceed to develop a number of definitions which will aid in clarifying the analyses which follow.

Having defined what we mean by theoretical exchangeability of vertices, we now find it helpful to consider a logic-valued function which, when given some pair of vertices, will indicate the truth of the proposition that they are interchangeable. We call this function a *discrimination function*[13], given explicitly in Definition 2 below.

Definition 2 (Discrimination Function). Let the *discrimination function*, $\mathcal{F}_{\mathcal{D}}(x, y|\mathcal{T})$, for theory $\mathcal{T}$ and graph $G = \{V, E\}$, take $x, y \in V(G)$ into the set $\{\text{TRUE}, \text{FALSE}\}$ such that $\mathcal{F}_{\mathcal{D}}(x, y|\mathcal{T}) = \text{TRUE}$ iff x and y are exchangeable with respect to theory $\mathcal{T}$, $\forall \ x, y \in V(G)$.

The discrimination function has a number of properties which are important for our development of the exchangeability problem. Of particular import is the fact that, when taken as a binary relation on the vertex set, the discrimination function forms an equivalence relation. This result is shown in Lemma 1 below.

Lemma 1 ($\mathcal{F}_{\mathcal{D}}$ Forms an Equivalence Relation). *Given a graph. $G = \{V, E\}$, and a theory, $\mathcal{T}$, the discrimination function $\mathcal{F}_{\mathcal{D}}$ forms an equivalence relation on $V(G)$.*

Proof. As any relation which is reflexive, symmetric, and transitive is by definition an equivalence relation, we proceed to show that $\mathcal{F}_{\mathcal{D}}$ applied to $V(G)$ satisfies each of these properties.

[12] It is not our intention to argue here regarding the generic adequacy of this concept of "theory"; we have kept our specification of the term minimal for the purpose of this particular application.

[13] This notion clearly bears some resemblance to the concept of the *discriminant function*, and technically could be thought of as a special case of the latter. In common usage, however, the latter term is strongly associated with a particular set of classification methods which are not pertinent to this discussion (see, for instance, Dillon and Goldstein (1984).

(Reflexivity) $\mathcal{F}_{\mathcal{D}}(v, v \,|\mathcal{T}) = \textsc{True}$ iff $p(y \,|G, \mathcal{T}) = p(y \,|G, \mathcal{T})\ \forall\ y$ in the scope of $\mathcal{T}$. As the latter is trivially true for all $v \in V(G)$, it follows that $\mathcal{F}_{\mathcal{D}}$ is reflexive.

(Symmetry) Let $L(G)$ be a relabeling of the vertices of G such that $L(v_i) \to v_j$, $L(v_j) \to v_i$, and $L(v) \to v\ \forall\ v \notin \{v_i, v_j\}$. By definition, then, $\mathcal{F}_{\mathcal{D}}(v_i, v_j \,|\mathcal{T}) = \mathcal{F}_{\mathcal{D}}(v_j, v_i \,|\mathcal{T})$ iff $(p(y \,|G, \mathcal{T}) = p(y \,|L(G), \mathcal{T})) \Leftrightarrow (p(y \,|L(G), \mathcal{T}) = p(y \,|G, \mathcal{T}))$. This is true by the reflexive property of equality, and therefore $\mathcal{F}_{\mathcal{D}}$ symmetric.

(Transitivity) Let $L_{ab}(G)$ be a relabeling of the vertices of G such that $L_{ab}(v_a) \to v_b$, $L_{ab}(v_b) \to v_a$, and $L_{ab}(v) \to v\ \forall\ v \notin \{v_a, v_b\}$, for any $a, b \in V(G)$. Then assume that there exist three vertices $v_i, v_j, v_k \in V(G) : \mathcal{F}_{\mathcal{D}}(v_i, v_j \,|\mathcal{T}) = \mathcal{F}_{\mathcal{D}}(v_j, v_k \,|\mathcal{T}) = \textsc{True}$. By the definition of $\mathcal{F}_{\mathcal{D}}$, this implies that $p(y \,|G, \mathcal{T}) = p(y \,|L_{ij}(G), \mathcal{T}) = p(y \,|L_{jk}(G), \mathcal{T})$. As $L_{ik}(G) = L_{ij}(L_{jk}(L_{ij}(G)))$, it follows that

$$\begin{aligned} p(y \,|L_{ik}(G), \mathcal{T}) &= p(y \,|L_{ij}(L_{jk}(L_{ij}(G))), \mathcal{T}) \\ &= p(y \,|L_{jk}(L_{ij}(G)), \mathcal{T}) \\ &= p(y \,|L_{ij}(G), \mathcal{T}) \\ &= p(y \,|G, \mathcal{T}) \end{aligned}$$

and, therefore, $\mathcal{F}_{\mathcal{D}}(v_i, v_k \,|\mathcal{T}) = \textsc{True}$. $\mathcal{F}_{\mathcal{D}}$ is hence transitive on $V(G)$.

Because $\mathcal{F}_{\mathcal{D}}$ applied to $V(G)$ is reflexive, symmetric, and transitive, it follows that $\mathcal{F}_{\mathcal{D}}$ forms an equivalence relation on $V(G)$. □

With the discrimination function, we can proceed to determine the exchangeability of vertex pairs. Since such exchangeability has been shown (by Lemma 1) to form an equivalence relation on such pairs of vertices, it follows that we can use this relation to define equivalence classes on the vertex set. These classes will then correspond to groups of mutually exchangeable nodes. The set of these classes (a set of sets of vertices) is referred to here as the *exchangeability set*, and is a useful means of summarizing the information of $\mathcal{F}_{\mathcal{D}}$. We define this set as follows:

Definition 3 (Exchangeability Set). Given a graph $G = \{V, E\}$ and a discrimination function $\mathcal{F}_{\mathcal{D}}$ for theory $\mathcal{T}$, let the *exchangeability set* $\mathcal{E}_G$ be defined by $\mathcal{E}_G = \{e : e \subseteq V(G), \mathcal{F}_{\mathcal{D}}(v_i, v_j \,|\mathcal{T}) = \textsc{True}\ \forall v_i,$

Observe that, trivially, the exchangeability set will be defined for all graphs and will form a partition on the vertex set of G: since $\mathcal{F}_{\mathcal{D}}$ is reflexive, every vertex will be in a set containing at least itself; and because $\mathcal{F}_{\mathcal{D}}$ is also symmetric and transitive, no vertex can be in more than one set (as this would require there to be some pair of other vertices exchangeable with the initial vertex but not exchangeable with each other)[14]. With this partition in hand, we are now ready to move from considering the exchangeability of pairs of vertices to defining the associated set of relabelings – or, equivalently, permutations – implied by our theory of interest. We refer to these permutations as being *accessible*, in the sense that they can be thought of reflecting possible microstates of the structural system

[14]This is a general property of equivalence relations, but we describe it in fuller detail here for reasons of clarity.

which are consistent with the constraints imposed by the *a priori* theory[15]. This is specified formally by Definition 4:

Definition 4 (Accessible Permutations). Given a graph $G = \{V, E\}$ with permutation group (labeling set) $\mathbf{P}_G$ and exchangeability set $\mathcal{E}_G$, the set of *accessible permutations*, $\mathcal{P}_G$, is given by $\mathcal{P}_G = \{p \in \mathbf{P}_G : p(v_i) \rightarrow v_j \text{ iff } \forall\, v_i \in V(G)\ \exists\, e \in \mathcal{E}_G : v_i, v_j \in e\}$.

The set of accessible permutations, like the discrimination function, turns out to have a number of interesting and useful mathematical properties. In particular, it happens that the set of accessible permutations must always be a permutation group; this is trivially true when $\mathcal{P}_G = \mathbf{P}_G$, but is less obvious in the general case. We state this result formally in Theorem 1:

Theorem 1 (Accessible Permutation Group). *Given a graph $G = \{V, E\}$, the set of accessible permutations, $\mathcal{P}_G$, where combination of two permutations is interpreted as applying them successively, forms a permutation group.*

Proof. First, we note that proving that $\mathcal{P}_G$ together with the combination operation (denoted $P_1 \circ P_2$) forms a permutation group requires only proving that it satisfies the definition of a group (because the elements of $\mathcal{P}_G$ are permutations). We then proceed to show each required property in turn.

(Identity) As noted above, $\mathcal{E}_G$ covers $V(G)$. Therefore, the trivial automorphism $P_I = \begin{pmatrix} 1 & 2 & \ldots & |V(G)| \\ 1 & 2 & \ldots & |V(G)| \end{pmatrix}$ must be a member of $\mathcal{P}_G$. Because $P_I \circ P = P \circ P_I = P$ for all $P \in \mathbf{P}_G$, P_I acts as an identity element under combination, and thus the identity property is satisfied.

(Inverse) Assume that there exists some $P \in \mathcal{P}_G : \nexists P' \in \mathcal{P}_G : P \circ P' = P' \circ P = P_I$. Then it follows that there must exist some $v_i, v_j \in V(G) : \mathcal{F}_{\mathcal{D}}(v_i, v_j \,|\mathcal{T}) = \textsc{True}$, $\mathcal{F}_{\mathcal{D}}(v_j, v_i \,|\mathcal{T}) = \textsc{False}$. This violates the symmetric property of $\mathcal{F}_{\mathcal{D}}$ (shown in the proof of Lemma 1), however, and thus there exists no permutation in $\mathcal{P}_G$ without an inverse.

(Associativity) Since the elements of $\mathcal{P}_G$ are permutations, it is a standard result that their combination is associative.

(Closure) Assume that there exist two permutations $P_i, P_j \in \mathcal{P}_G : P_i \circ P_j \notin \mathcal{P}_G$. From this it follows that there must exist vertices $v_i, v_j, v_k \in V(G) : \mathcal{F}_{\mathcal{D}}(v_i, v_j \,|\mathcal{T}) = \mathcal{F}_{\mathcal{D}}(v_j, v_k \,|\mathcal{T}) = \textsc{True}$ but $\mathcal{F}_{\mathcal{D}}(v_i, v_k \,|\mathcal{T}) = \textsc{False}$. As we have seen in the proof of Lemma 1, however, $\mathcal{F}_{\mathcal{D}}$ is transitive, and this therefore implies a contradiction. Thus, $\mathcal{P}_G$ is closed under combination.

Finally, as $\mathcal{P}_G$ together with the operation of combination satisfy all of the above properties, we conclude that they comprise a permutation group. □

We now have all of the elements needed to treat formally the problem of vertex exchangeability. Within the group of possible permutations ($\mathbf{P}_G$) on G, only some are legal given our prior theory $\mathcal{T}$. These accessible permutations form a group, $\mathcal{P}_G$, whose members comprise the total set of vertex exchanges which are consistent with our theoretical assumptions.

[15] This usage is intentionally allusive to the concept of "accessible states" within statistical mechanics (Kittel and Kroemer, 1980), to which it is roughly analogous.

Maximal sets of vertices which are not theoretically distinct from one another comprise equivalence classes under the discrimination function $\mathcal{F}_{\mathcal{D}}$, and each such class is an element of the exchangeability set ($\mathcal{E}_G$). The members of $\mathcal{P}_G$ are thus those of $\mathbf{P}_G$ which exchange vertices only within the elements of $\mathcal{E}_G$. For an extremely restrictive $\mathcal{T}$ – such as a theory of interaction postulating individual differences[16] – the accessible permutation group may comprise only the identity permutation, $\begin{pmatrix} 1 & 2 & \ldots & |V(G)| \\ 1 & 2 & \ldots & |V(G)| \end{pmatrix}$. At the opposite extreme, a theory concerned only with underlying structure (e.g., the baseline structural theories of Mayhew (1984)) would implied an accessible permutation group equal to $\mathbf{P}_G$, the entire permutation group on the vertices of G. In between these extremes, a wide range of theories may posit different sets of exchangeable actors, each implying in its turn a particular accessible permutation group. By working with all graphs labeled under $\mathcal{P}_G$, then, we can accommodate a wide range of theoretical assumptions within a single unifying formalism.

2.2.2 Exchangeability of Structures

In addition to the theoretical exchangeability of vertices, we generally assume that edge variables are exchangeable across structures. Note that this assumption is not actually required for a strictly data-analytic interpretation of the analyses considered here, but is generally necessary if one wishes to draw inferences on structural populations from structural samples. Since the distance methods considered here are nearly always employed in an exploratory fashion, exchangeability should rarely be relevant to their use. The structural correlation and covariance methods discussed, by contrast, can be employed either data-analytically or inferentially; in the latter case, violations of the exchangeability assumption will almost certainly reduce the efficiency of estimation, and may bias the resulting inferences.

2.2.3 Edge Value Distributions

The standard formalism of classical network analysis is the graph, in which relations between vertex pairs are either present or absent. This lends itself, for obvious reasons, to a dichotomous data representation (with 0 commonly employed to represent "no relation" and 1 employed to represent "relation"), and indeed this is currently the representation most commonly employed. For purposes of this paper, we shall consider each potential edge between vertices to be associated with an indicator variable, δ, and all analyses will operate on the relevant δ values. Any special restrictions on the distribution of edge values (e.g., dichotomous, positive, etc.) will be specified along with the definition of δ employed for the method in question; note, however, that the methods presented here are generally applicable to any real-valued edges.

[16] For instance, one might compare friendship and advice ties within a particular set of actors. In this case, the individuals involved are theoretically distinct.

2.2.4 Heterogeneity in Network Size

In addition to those considerations already discussed, heterogeneity in network size (synonymously graph order, or vertex set cardinality) is another difficulty which must be dealt with in inter-structural analysis. At its core, this problem stems from foundational considerations: structures (as we are considering them) are defined in terms of relations (edge sets) among sets of elements (vertex sets). Thus, for inter-structural comparison we must account not only for the manner in which elements are connected within structures, but also for the manner in which elements within one structure are understood to map onto elements within the other structures being examined. One aspect of this problem – exchangeability of vertices – has been treated above. Here, we deal with the related issue of mapping the vertex sets of individual structures ($V(G)$ for G in structure set S) onto a common vertex set (denoted $V(S)$), so that the resulting edge sets may be compared directly.

Addition of Isolates Perhaps the most obvious means of solving the size heterogeneity problem is by defining $V_S = \bigcup_{G \in S} V(G)$, and then substituting $V(G) = V_S$ for all $G \in S$. Since the edge sets of each G contain edges only on $V(G)$, it thus follows that the nodes $V_S - V(G)$ will be isolates for any given G; we may thus think of this approach as "adding isolates" to certain graphs within the graph set in order to make vertex sets comparable. Theoretically, this can be understood in terms of envisioning a (perhaps hypothetical) combined vertex set, for which the observed relations graphs were subgraphs formed by the removal of isolates. This is sensible in a wide range of situations: e.g., if one's observed structures were to consist of intertemporal observations of intra-group relations within a group of changing membership, it would be intuitively sensible to treat non-members at any given point as isolates with respect to the observed relations (since they implicitly exist, but were not part of the group and hence not present in the intra-group tie structure).

With respect to the distance measures employed here, it is notable that neither the Hamming nor the structural distance are affected by the addition of isolates. (Indeed, this result extends to the generalized structural distances on all Minkowski metrics, prominently including the euclidean distance.) Thus, use of the union rule would seem highly appropriate when conducting inter-structural analyses using distance-based approaches. In the case of graph covariance, by contrast, the addition of isolates does exert an impact on results: since the graph variance is suppressed by the addition of isolates, it follows that graph covariances are likewise attenuated. Although the exact degree of attenuation depends on a number of considerations, adding m isolates to two graphs of size n will generally decrease the magnitude of their covariation by a factor of roughly $n^2/(n+m)^2$ (for $n \gg 1, m \ll n$). Graph correlations are more robust to such effects (since they rescale the covariance by the geometric mean of the variances), and adding isolates generally produces only minor variation in the case of graph correlation measures. The union rule, hence, should be applied with some caution to graph covariance analysis, but is reasonable when one is working with correlation matrices.

Removal of Subgraphs If one obvious means of dealing with heterogeneity is by adding substructures to as to establish a common vertex set, another is to *remove* subgraphs in some reasonable way, until all structures to be examined are once again defined on a shared set of vertices. In situations for which certain structural elements are shared across multiple structures, one reasonable removal rule is that of intersection; specifically, we define $V_S = \bigcap_{G \in S} V(G)$, and then substitute $V(G) = V_S$ for all $G \in S$ (removing all edges with one or more endpoints not in V_S in the process). Intersection is a harsh rule, and one which may not be practical when structural elements are not uniquely labeled. An alternative, therefore, is to remove theoretically uninteresting subgraphs (such as isolates) from large members of S until a shared vertex set is reached. Still another variant on this theme is the decomposition of large graphs into strongly connected components or other meaningful subunits, and the treatment of each subunit as a separate structure.

The intersection rule, as indicated, is difficult to apply when one is working with graph sets for which no meaningful intersection can be computed; still, it may be quite applicable when working with intertemporal, organizational, or other data in which nodes can be uniquely identified and the intersection of vertices is of theoretical import. The additional removal rules described here are necessarily somewhat ad hoc, and as such it is difficult to argue for them in general terms; their use should be theoretically motivated, and it is there one should look for a compelling rationale[17]. One partial exception to this is the removal of isolates. As with addition of isolates, removal of isolates will not affect distance measures, and can thus be sensibly employed in such analyses. Covariance measures will generally be *inflated* by the removal of isolates in a reversal of the attenuation described above; correlation will not be greatly affected, in most cases. Other removals or decompositions may have unpredictable effects, and caution (along with robustness testing[18]) is advised when employing such strategies.

Theoretical Reduction One alternative to either addition or removal of substructures is the *reduction* of some or all structures to a common vertex set, using a vertex equivalence rule. One family of such reductions – blockmodeling by means of structural (Lorrain and White, 1971), automorphic (Everett, 1985), or regular (White and Reitz, 1983) equivalence classes – is already familiar to structural analysts, and needs little elaboration here. Suffice it to say that to apply this method one uses a theoretically appropriate equivalence relation and uses it to identify a set of classes with the required cardinality. The reduced form blockmodel induced by these classes on the original structure is then used in the subsequent analysis. Another approach is to use theoretical exchangeability (see Definition 1) to define the classes from which to build the block model[19]. Though less common, this procedure is

[17] E.g., if one is interested in multiple relations on a work team, one may be justified in discarding actors outside of the team who may have been included in the original data set for one reason or another.

[18] For instance, computing jackknife confidence intervals by randomly removing the same number of vertices from the graphs in question and re-executing the analyses multiple times.

[19] For instance, one might create a set of reduced form blockmodels by lumping together all governmental, all non-profit, and all for-profit organizations within each observed structure. The resulting data set would represent *aggregate* relations between *categories*, and could be interpreted accordingly.

often more in keeping with prior theory than the structurally induced alternatives. (Indeed, structurally defined blockmodels are no less dependent on prior theory than are blockmodels based on nodal attributes, and their use must likewise be justified on the grounds of prior knowledge.)

Theoretical reduction (particularly when performed using structurally defined equivalences) is in some ways an attractive solution to the problem of size heterogeneity. Even when one's observed structures vary greatly in size, it may be possible to find a theoretically appropriate set of categories whose relations can be readily analyzed. In some cases, indeed, it is precisely these aggregate relations which are of theoretical interest in the first place, and a reduction is *required* by the motivating theory. On the other hand, the temptation to use structural or other equivalences to "force" a theoretically inappropriate reduction should be resisted. Without a strong motivation, interpretation of relations between classes can be difficult or impossible; interpretation of inter-structural analyses on such ill-defined relations is likely to be even more problematic! Similarly, it must be remembered by the analyst that any results obtained by inter-structural analyses on reduced graphs pertain only to the set of reduced graphs themselves, and not to the original data[20]. Theoretical reduction, then, is an important method of dealing with size heterogeneity, but must be employed with care.

3 Approaches Based on Metric Inference: Clustering and Scaling for Graph Sets

The first family of multivariate methods for inter-structural analysis considered here are those approaches based on metric distances between structures. Formally, for any collection of objects S, any function $f(x,y)$ is a *distance function* on S if and only if, for all $x,y \in S$: $f(x,y) \geq 0$ (nonnegativity); $f(x,y) = f(y,x)$ (symmetry); and $f(x,y) = 0$ if and only if $x = y$ (identification). Further, $f(x,y)$ is said to form a *metric distance* on S if and only if $f(x,y) + f(y,z) \geq f(x,z)$ for all $x,y,z \in S$ (the triangle inequality). Intuitively, the notion of distance can be thought of as quantifying a more general notion of *dissimilarity*: the more dissimilar the objects, the farther they are from one another; the dissimilarity of x with respect to y is the same as the dissimilarity of y with respect to x; no x and y can be less dissimilar than the dissimilarity between x and x (or y and y); and the total absence of dissimilarity between x and y implies that x and y are identical. To this, the notion of metric distance adds the intuition that two objects cannot be more dissimilar to each other than the sum of their dissimilarities to a third object[21]. By defining and measuring distance between social structures, then, we seek to quantify in a general fashion the degree to which they are different from one another. These differences can then be

[20] E.g. it may be the case that aggregate relations involving employee transfer and joint ventures among NPOs, firms, and governmental organizations are very similar, but these relations may vary radically for certain subsets of the above populations.

[21] An alternate intuition for the triangle inequality is that there are no "shortcuts": the distance between x and y cannot be longer than the distance from x to y through a third object.

used to identify subsets of structures which are particularly similar (as in clustering), or to map the resulting differences onto a lower-dimensional, intuitively accessible layout (as in multidimensional scaling).

Because the distance methods presented here make few assumptions regarding the distributional properties of their inputs, these methods are applicable to a wide range of problems. Similarly, the intuitive nature of the distance-based approach makes it ideal as a starting point for examining the multivariate analysis of inter-structural data. We begin, then, by defining what we mean by the distance between structures; after developing a set of distance measures, we proceed to illustrate their use in cluster analysis and multidimensional scaling of inter-structural data.

3.1 Hamming and Structural Distance Metrics

Following earlier work by Banks and Carley (1994) and Butts and Carley (1998), we here employ the Hamming distance (Hamming, 1950) as our basic measure of difference between structures. In particular, given two labeled digraphs, H_i and H_j with vertex sets $V(H_i) = V(H_j) = V_U$ and edge sets $E(H_i)$ and $E(H_j)$ respectively, we may define a metric distance between them as per Hamming (1950). First, we define an indicator function $\delta_H(x,y)$ such that

$$\delta_H(x,y) = \begin{cases} 1 & \text{if } (v_x, v_y) \in E(H), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The function δ permits us to count directed edges within a given labeled digraph. To derive the Hamming distance between our two labeled digraphs, then, we simply count the number of directed edges which exist in one graph and not the other. This gives us the following expression for the Hamming distance:

$$D(H_i, H_j) = \sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \left| \delta_{H_i}(x,y) - \delta_{H_j}(x,y) \right| \tag{2}$$

As noted above, previous work by Butts and Carley (1998) has shown that the observed Hamming distance between two labeled graphs may be decomposed into a minimal, structural distance which depends only on the underlying unlabeled graphs, and an additional labeling distance which is a function both of the underlying unlabeled graphs and their respective labelings. For $H_i = L_i(G_i)$ and $H_j = L_j(G_j)$ (where L represents a labeling of the vertices of G), this decomposition gives us

$$D(L_i(G_i), L_j(G_j)) = D_S(G_i, G_j) + D_L(L_i(G_i), L_j(G_j)) \tag{3}$$

where D_L represents the labeling distance, and the structural distance, D_S, is given by

$$D_S(G_i, G_j) = \min_{L_a \in \mathcal{P}_{G_i}, L_b \in \mathcal{P}_{G_j}} (D(L_a(G_i), L_b(G_j))) \tag{4}$$

Clearly, then, calculation of the structural distance can be viewed as equivalent to minimization of the Hamming distance over the set of accessible permutations, which is in turn equivalent to minimization of the labeling distance[22]. Simultaneous minimization of the labeling distance between all pairs of graphs within a set has been attempted via the use of canonical labeling algorithms (Butts and Carley, 1998), but current results are heuristic for cases in which $D_S > 0$[23]. When examining graph pairs dyadically, it is always possible to perform the appropriate minimization by exhaustively searching the space of accessible permutations, but when $|\mathcal{P}_G| \rightarrow |\mathbf{P}_G|$ this becomes a factorial-time algorithm. Though even this is not infeasible for sets of very small groups (e.g., $|V_U| < 10$) given current computing capabilities, and while many research contexts motivate use of accessible permutation groups with very low cardinality (the obvious case being that in which all nodes are uniquely labeled), such computational complexity renders the exhaustive search tactic useless for medium to large graphs in the unlabeled case. Happily, other approaches are available. As Butts and Carley (1998) suggest, heuristic search techniques such as Monte Carlo sampling and genetic algorithms provide viable alternatives for finding the structural distance between a given pair of graphs, and have the distinct advantage of being tunable to the specific properties of the problem under study. Because the search space of the labeling optimization problem has a well-defined local neighborhood (given by the set of all dyadic exchanges), simulated annealing is particularly well-suited to this task: the method has performed well in preliminary tests by the authors, and is employed for all labeling optimization problems presented here[24]. Though an optimal solution to the problem of calculating the structural distance between two graphs in the general case is not presently available, then, workable heuristics do exist. Refinement of these techniques is an important topic for further research.

Given values (or estimates) for the structural distances between graphs, we may construct a two-dimensional matrix of distances between all graphs within the set under analysis. This distance matrix, because it reflects a metric on the space of unlabeled graphs[25], can be unproblematically used as input data to standard cluster analysis routines, and, with some modification, can be used in conjunction with other methods as well. Since both Hamming-based distances can be interpreted directly as the minimum number of edges needed to take one graph (labeled, unlabeled, or partially labeled) into another, these distance metrics avoid the problems of interpretation which commonly plague multivariate methods based on distances. The above, then, serve as the basic notions of comparison for the methods

[22] As this implies, the structural distance is always conditional on the theory, $\mathcal{T}$, used to define the accessible permutation group. Although we will generally leave this unstated in our notation, it is always implicit.

[23] When $D_S = 0$, the minimization of D_L corresponds to the familiar graph isomorphism problem, for which a number of algorithms are available. Unfortunately, algorithmic performance in the isomorphism problem does not appear to relate strongly to performance in the more general case (Butts and Carley, 1998).

[24] In particular, all optimization results shown here were obtained using a simulated annealing algorithm with dyadic exchanges defining search space neighborhoods and an (exogenous) exponential cooling rule. Annealing parameters were adjusted in each case based on convergence diagnostics; multiple trials were employed to reduce the probability of convergence to a local optimum.

[25] This result is shown in Butts and Carley (1998).

which follow[26].

3.1.1 Extension to Generalized Structural Distance Measures

Before proceeding to demonstrate methods for the analysis of structural distances, it would seem worthwhile to note that the logic applied in deriving the structural distance from the Hamming distance can also be applied generically to any distance measure. In general, if $\mathcal{D}$ is some arbitrary distance measure on the space of labeled graphs, then we can define its structural counterpart, $\mathcal{D}_S$, by

$$\mathcal{D}_S\left(G_i, G_j\right) = \min_{L_a \in \mathcal{P}_{G_i}, L_b \in \mathcal{P}_{G_j}} \left(\mathcal{D}\left(L_a\left(G_i\right), L_b\left(G_j\right)\right)\right) \tag{5}$$

The form of the distance decomposition here is identical to that in the previous section (though the generalized structural distance itself may or may not have a reasonable interpretation, depending on the underlying distance measure). Although we will not here consider the full range of available alternatives to the Hamming metric, two reasonable options are the absolute distance and the euclidean distance. The former can be obtained from Equation 2 by replacing the definition of δ by

$$\delta_H\left(x, y\right) = \mathbf{A}_{xy} \tag{6}$$

where $\mathbf{A}$ is the valued adjacency matrix of H. To obtain the euclidean distance, we use the following:

$$\mathcal{D}^E\left(H_i, H_j\right) = \left(\sum_{x=1}^{|V_U|}\sum_{y=1}^{|V_U|}\left(\delta_{H_i}\left(x, y\right) - \delta_{H_j}\left(x, y\right)\right)^2\right)^{1/2} \tag{7}$$

with δ as given in Equation 6 above. For dichotomous data, it therefore follows that the relationship between the Hamming and euclidean distances is given by

$$\mathcal{D}^E\left(H_i, H_j\right) = \sqrt{D\left(H_i, H_j\right)} \tag{8}$$

a relationship which will be true of their structural counterparts as well. Thus, the euclidean distance will (in the dichotomous case) tend to place decreasing marginal weight on tie discrepancies as their number increases; whether this is considered theoretically desirable will depend on the phenomenon under observation.

3.2 Cluster Analysis of Graph Sets

As we have seen, there exist a variety of ways of identifying the structural distance, D_S between two graphs with arbitrary accessible permutation groups. Given a set of structures,

[26] With the exception of the graph correlation (Wasserman and Faust, 1994), which is employed in certain cases as well. See below.

then, it is possible to construct a *distance matrix* by accumulating the structural distances between each pair of structures. Such a distance matrix lends itself quite readily to the family of classification techniques known collectively as cluster analysis (Romesburg, 1984; Mardia et al., 1979), which are methods for the identification of homogeneous subgroups within a larger group. In this application, "homogeneous" means structurally similar (as measured by the structural distance), and the clusters which are identified as such may be studied for association with non-structural variables, used to construct a structural taxonomy, or possibly even used to infer "lineage" in the sense of population ecology[27]. Identified clusters can also be summarized using the central graph (Banks and Carley, 1994), whose properties can then be examined using classical network analysis methods. Such an approach could be used to scale down large data sets (e.g., organizational populations) by weeding out similar structures, resulting in a smaller and more manageable collection of archetypes.

One benefit of the deployment of cluster analysis as a multivariate network analysis tool is the fact that existing methods can be used directly on the structural distance matrix; standard approaches can be used, and no additional assumptions are necessary. (Indeed, cluster analysis of *positions* (as opposed to the inter-structural analysis pursued here) is already widely deployed in network analysis for the study of structural, automorphic, regular, and other equivalences.) To demonstrate its application, then, we now turn to an empirical example of the use of cluster analysis on a set of social relations.

3.2.1 An Example of Cluster Analysis of Graph Sets

The data set for our first example comes from Sampson (1969), whose famous study of initiates in a monestary during a period of internal political upheaval has become a classic of network analysis. During his stay in the monestary (as an observer), Sampson collected subjective interpersonal evaluations among the initiates on several dimensions: "liking" and "disliking"; "esteem" and "disesteem"; "positive" and "negative" influence; and "praise" and "blame". Furthermore, information on "liking" was collected at three points in time, the last of which coincided with the other measured relations. This data was collected as rank-order sociometric choices: up to three nominations were allowed for each person for each relation, coding in descending order of strength (from 3 down to 1 for the weakest choice, and 0 for those not chosen). Thus, we here use distances with δ defined as in Equation 6.

As the above would seem to indicate, the multiple relations collected on the same actors correspond to very different sociometric constructs; we would thus expect to see some distance between relations, depending on the similarity of the constructs in question. Similarly, we would expect that the positive and negative valance relations would be quite distinct from one another. [28]

[27] Such methods are widely deployed within evolutionary biology for purposes of phylogeny reconstruction; see Ridley (1994).

[28] Results obtained with average-link, single-link, and complete-link clustering rules were qualitatively identical to those presented here.

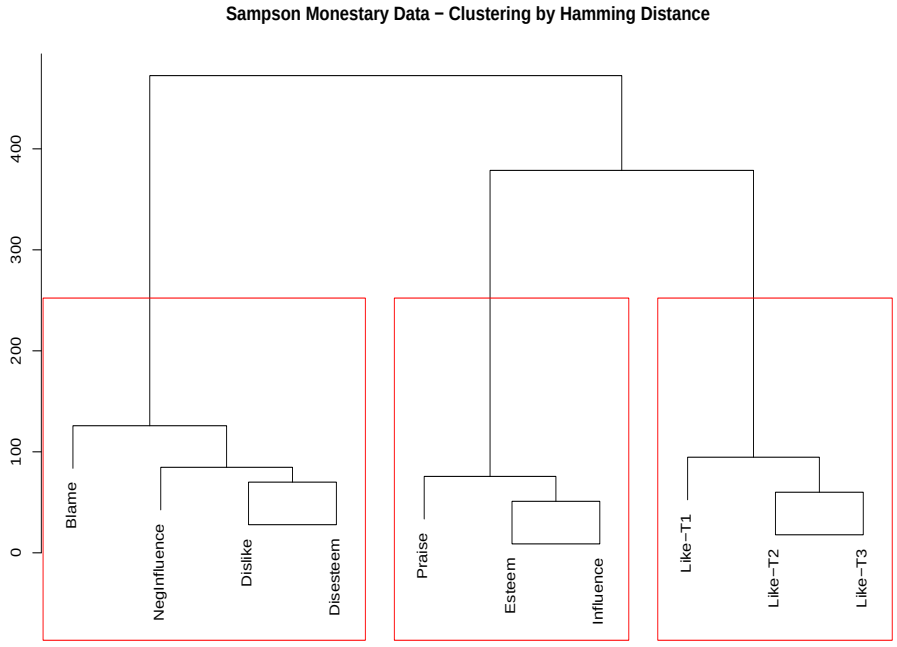

Figure 1: Ward's Clustering of Sampson Monestary Data, Uniquely Labeled Case

A binary tree representation of the hierarchical clustering of the Sampson data in the uniquely labeled case is presented in Figure 1; branchings of the tree represent partitioning decisions, with the height of the branching event indicating (via the y axis) the Ward distance at which the decision occurs. Note that the results of this analysis correspond excellently to our intuition regarding the relations in question. As we expected, the algorithm has immediately partitioned the positive and negative relations from one another, indicating that this is the initial division which results in the greatest reduction in variance. The next partitioning separates the "liking" relations from the other positive relations. Recall that these relations were observed at three points in time, with only the last being coterminous with the other data in the sample; interestingly, the distance between these networks appears related to time, with the T1-T3 distance being greater than the T1-T2 and T2-T3 distances), and with the mean distance to the other relations in the set similarly increasing with temporal distance. As we shall see, this suggestion of a steady evolution of structure over time is not unique to this data set.

3.3 Multidimensional Scaling of Graph Sets

The family of classical techniques collectively known as multidimensional scaling (MDS) methods is based on the problem of configuring n points in a (generally low-dimensional) euclidean space in such a way as to preserve the observed distance relationships between these points within a higher-dimensional (and possibly non-euclidean) space (Dillon and Goldstein, 1984; Mardia et al., 1979). MDS methods are commonly used to facilitate visualization of multivariate data, particularly for the purpose of identifying objects or variables with similar characteristics; more specialized applications include seriation (Kendall, 1971) and low-dimensional parameterization of complex data sets. Like many forms of cluster analysis, MDS can generally be interpreted in a purely data-analytic fashion, with no stochastic modeling assumptions. This adds to the methods' versatility, and makes them particularly well-suited for idiosyncratic applications such as network analysis. Currently, multidimensional scaling is used within social network analysis for studying positions of structural elements within individual networks, and such applications are implemented in standard network analysis packages such as UCINET. Here, we extend this technique to the study of entire social networks within populations of structures, using the structural distance framework established above.

3.3.1 Classical MDS with Euclidean Representation

One simple example of the way in which MDS methods may be applied to inter-structural analysis serves to illustrate some aspects of the general approach. For this special case, we consider only labeled digraphs, and take the distance between structures to be the euclidean distance (defined in Equation 7), rather than the Hamming (or structural) distance. While the euclidean distance is less interpretable than the Hamming distance in the context of graph comparison, we employ it here due to its mathematical properties (discussed below). Using the above, we can now define a squared inverse distance matrix $\mathbf{M}$ on graph set S, as follows:

$$\mathbf{M} = (a_{ij}), \quad a_{ij} = -\frac{1}{2}\mathcal{D}^E\left(H_i, H_j\right)^2 \tag{9}$$

The centering matrix for S is then given by

$$\mathbf{C} = \mathbf{I} - |S|^{-1}\,\mathbf{1}\mathbf{1}' \tag{10}$$

and we are thus able to determine the centered inner product matrix of the scaling as

$$\mathbf{B} = \mathbf{CMC} \tag{11}$$

(The properties of $\mathbf{B}$ are guaranteed by the use of the Euclidean distance; see Mardia et al. (1979) for details.)

Given the above assumptions, the first k eigenvectors of $\mathbf{B}$ (interpreted as coordinates) provide the classical solution to the MDS problem in k dimensions for the set of labeled

digraphs S. The resulting coordinates may be examined for evidence of clustering, unidimensional ordering, and the like as per standard practice, provided that the labeling assumption is met and that the interpretation of euclidean distance is theoretically meaningful[29].

3.3.2 Generalized MDS with Structural Distances

The above provides an example of classical MDS on euclidean distances for labeled digraphs. In general, however, we are interested in graphs with arbitrary accessible labelings, and seek to perform our scaling on the matrix of structural distances (rather than their euclidean counterparts). Because these distance matrices are not in general Euclidean[30], the utility of the classical approach is here more limited. Nevertheless, well-established methods of multidimensional scaling exist for both metric and non-metric distance matrices, and these can be readily applied to the inter-structural analysis problem. In particular, the classical metric MDS procedure described above can be applied directly to structural distance matrices with interpretable results, so long as the first k eigenvalues are positive. As typical analyses consider only a small number of dimensions (e.g., 1-3), it is probable that this condition will be met in the vast majority of cases[31]. Non-metric methods such as those described by Kruskal (1964a; 1964b) will guarantee interpretable solutions, but sacrifice some of the information contained in the structural distance matrix. Generally speaking, it is recommended that researchers employ the classical metric MDS when using the metric distances described here, inspecting the solution to ensure that it provides a reasonable representation of the data; when this is insufficient, heuristic or non-metric methods should be employed.

3.3.3 An Example of MDS on Graph Sets

While multidimensional scaling and cluster analysis can be used for similar purposes, MDS compensates for its somewhat more restrictive nature by enabling the data analyst to extract dimensional structure from the data. Thus, if it is the case that dissimilarities between cases can be readily expressed in terms of a small set of dimensions, multidimensional scaling will often reveal this structure[32].

To illustrate the utility of this approach for inter-structural data analysis, we first turn to data collected in 1956 as part of the Michigan Group Study Project, under the supervision of Theodore Newcomb (Newcomb, 1961). As part of this project, 17 male transfer students at the University of Michigan were recruited to live in off-campus fraternity housing rented

[29] As we have seen the euclidean distance tends to suppress large distances relative to the Hamming distance, since the former grows on $O(d^{1/2})$ with respect to the latter for simple digraphs.

[30] Note that one may in principle arrive at a distance matrix which is Euclidean without utilizing Euclidean distance per se; the former property requires only that there exists a configuration in *some* Euclidean space such that the interpoint distances are equal to the observed distances.

[31] Classical metric MDS is known to be a fairly robust technique (Dillon and Goldstein, 1984), and where the classical solution fails, other metric alternatives do exist (i.e., heuristic optimization).

[32] Principal component analysis (PCA), which we shall consider presently, can also be used for this purpose; indeed, the latter can be seen as a special case of the former under classical assumptions (Mardia et al., 1979).

for the study, in exchange for agreeing to answer periodic questionnaires. One such questionnaire was a sociometric ranking task, in which each member of the group was asked to indicate (by ordinal ranking, no ties permitted) his preference for all other group members. This task was administered weekly for 15 weeks, thus providing a regular assessment of the affinitive relations within the group over an extended period.

In analyzing this data, one question which one might naturally ask is whether there is evidence of a systematic evolution of structure over time. Such a question can be asked in at least two senses: first, did the structure of preferences *among specific individuals* evolve systematically; and second, was there evolution in the *overall preference structure* over time[33]. The first question, being one in which we are concerned with the changes of attitudes among particular egos towards specific alters, clearly motivates treatment of vertices as uniquely identified. The second is concerned only with underlying structure, and treats all vertices as exchangeable. In both cases, however, we are interested in exploring the extent to which some underlying temporal dimension can potentially account for patterns of differences between structures, a task for which MDS is especially well-suited.

To proceed, we begin by forming two structural distance matrices on the 15 structures: in one matrix, the unique node labeling given by the individuals' identities is used, while in the other all vertices are taken to be exchangeable. As our measure of distance we employ absolute rank differences, taking δ_H as given in Equation 6. Applying the classical metric MDS algorithm to these matrices and plotting each observation on the first two dimensions yields the visualization of Figure 2. As the figure indicates, the MDS solution reveals a smooth curve in the first two dimensions for both the labeled and unlabeled cases, parameterized by time (indicated by the sequential numbering of observations). Such a curve is clear evidence of *seriation*, revealing a consistent pattern of change in the structure of interpersonal preference over time. Furthermore, the placement of points on the curve reveal much about the nature of this pattern: structural evolution was systematic (rather than erratic) and decelerated over time. The greatest changes in interpersonal preference rankings clearly occur in the first month (points 1-4), settling down substantially in subsequent weeks. By week 10 relatively little change occurs, and (more importantly) these changes are increasingly erratic; by the last two weeks, the system appears to have reached equilibrium.

Additional insight is provided by a comparison of the labeled and unlabeled cases. As Figure 2 indicates, the basic pattern of structural evolution was largely the same for both; this suggests that the changes which occurred in the Michigan Group Study did not simply involve a movement of actors between a fixed set of roles, but rather were fundamental to the underlying structure. Further support for this interpretation is provided by the Shepard diagram shown in the first panel of Figure 4. In the Shepard diagram, all observed distances are plotted against the distances resulting from the multidimensional scaling; in the metric case, these provide a useful heuristic for assessing the adequacy of the dimensional representation. As the figure shows[34], the predicted distance is a linear function of

[33] As an analogy, one may consider a flock of geese, which maintains a uniform flying formation over time despite the fact that individuals will periodically switch positions within the formation.

[34] Only the labeled case is depicted here; the Shepard plot for the unlabeled case is extremely similar.

the actual distance; some underprediction occurs at lower distances, but on the whole the representation appears to be a good one. Further analysis would examine additional dimensions, attempt to uncover the nature of the observed structural evolution, etc., but even this cursory examination demonstrates the power of this simple exploratory tool for revealing simple structure hidden in complex patterns of differences.

In the above example, we showed how an MDS could be used to reveal patterns of evolution in social structures. A second use of MDS is to uncover potentially meaningful latent dimensions in a structural data set. Though we will see other methods which also excel at this task, it is still worth considering a simple example which highlights the way in which the robust power of multi-dimensional scaling can uncover useful features in inter-structural data. The structures for this example are taken from Roethlisberger and Dickson's famous management study (Roethlisberger and Dickson, 1939); in particular they consist of various interactions among workers in the bank wiring room. A key question of the Roethlisberger and Dickson study was the interaction between work roles, working conditions, interaction, and productivity. As a result, Roethlisberger and Dickson considered a range of different kinds of interaction, ranging from horseplay and friendship to arguments and antagonism. Do these relations, when individuals are considered only as representatives of their work roles, be seen as exhibiting some underlying dimensional structure?

To examine this question, we take the structural distances between all structures (exchanging vertices within work roles), and apply the MDS algorithm; the results are shown in Figure 3. The Shepard plot for the wiring room MDS (Figure 4) appears to indicate an excellent fit on the first two dimensions, suggesting that the representation is sound, and there does indeed appear to be some intelligible meaning to the underlying dimensions. While somewhat subjective, one reasonable interpretation of the resulting distances is that the relations considered differ primarily on two dimensions, one of affect (positive versus negative) and one of formality (informal/nonprofessional to formal/professional). Likewise, the tendency for more strongly affective relations (e.g., antagonism and horseplay) to be less formal than less affective relations (e.g., job trades) suggests a potential explanation in terms of a trivariate, oblique structure. Such an observation could form the basis of subsequent investigation into the relationship between affectivity and formality, or could suggest the beginnings of a typology for interpersonal relations within organizations. As in the previous example, it would appear that the multidimensional scaling of structural distances can be a powerful exploratory tool for bringing order to a diverse data set.

4 Approaches Based on Graph Covariance: Structural Regression, Principal Components, and Canonical Correlation Analysis on Sets of Graphs

Distance-based approaches to inter-structural analysis are extremely general, but are likewise limited in their ability to deliver detailed information regarding the relationships within sets of structures. More theoretically refined methods for the analysis of multivariate data

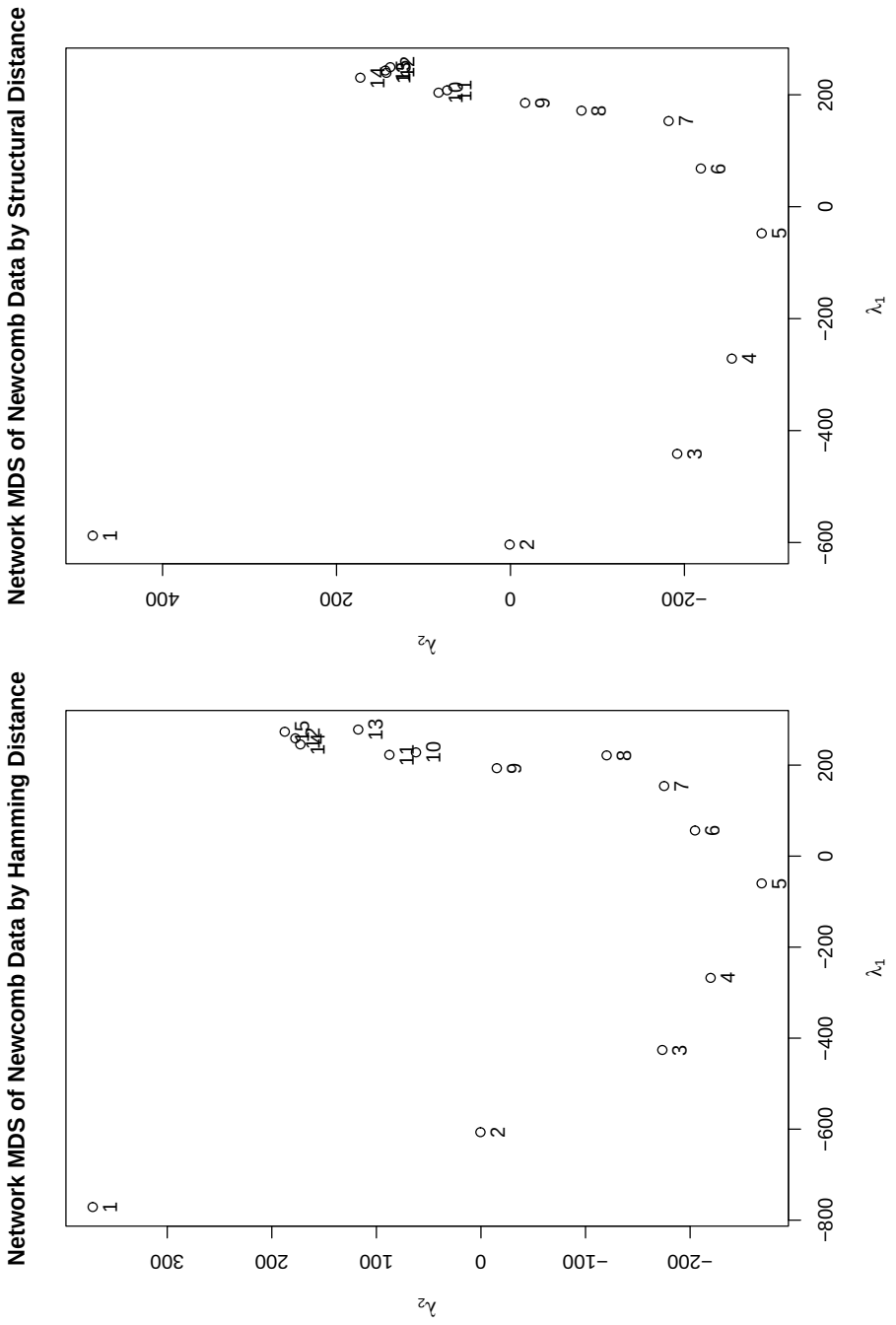

Figure 2: MDS of Newcomb Fraternity Data, Uniquely Labeled and Unlabeled Cases

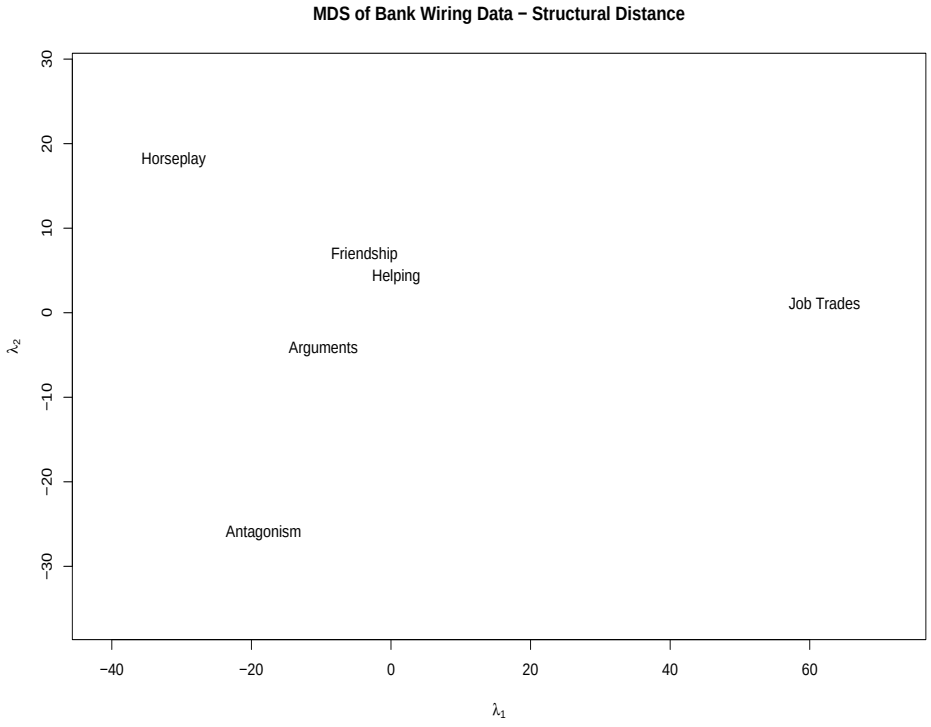

Figure 3: MDS of Bank Wiring Data, Actors Exchangeable within Roles

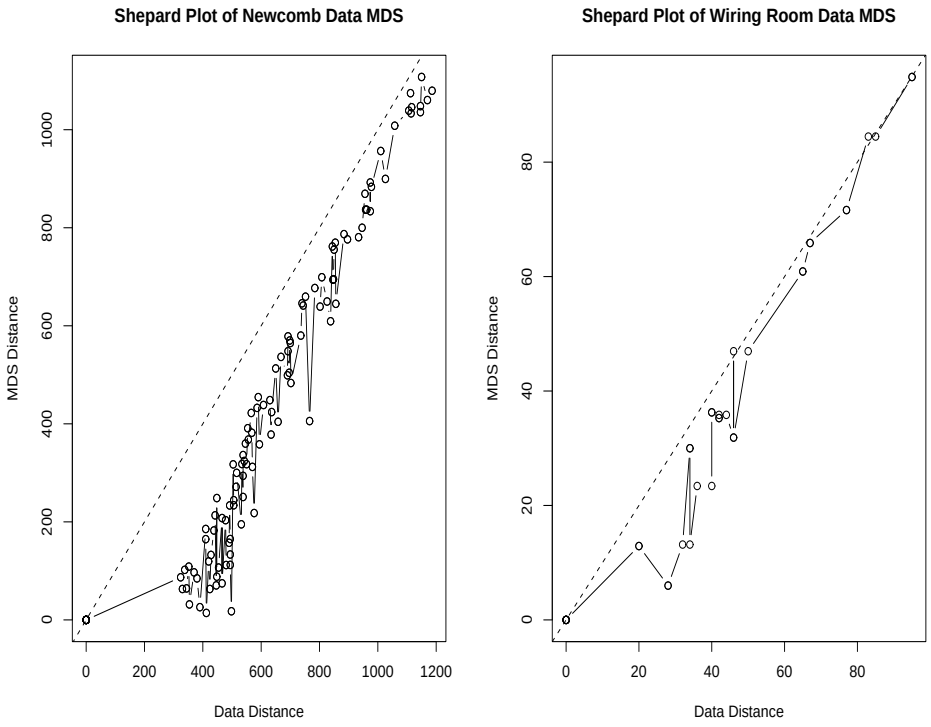

Figure 4: Shepard Diagnostic Plots, Newcomb and Bank Wiring Room MDS Fits

generally center on analysis of the covariance between variables; we, then, shall follow suit. While the methods we shall here employ have been well-developed within the statistical literature, their application to inter-structural data demands certain special considerations. In particular, many covariance-analytic methods rely on certain properties of the covariance matrix which are guaranteed under the conventional definition of the covariance, and we must here demonstrate that our measure also satisfies these properties before we can reliably utilize the methods in question.

In this section, we will develop structural regression (which generalizes the network regression of Krackhardt (1988)), principal component analysis (which finds canonical structures in terms of which other structures can be expressed), and canonical correlation analysis (which expresses general linear relations among sets of graphs) for graph sets. As before, however, we begin by establishing the basic formalisms which will allow us to apply the methods of interest to inter-structural data. Here, these formalisms are the graph mean and variance, and the structural covariance and correlation measures. With these tools for graph comparison in hand, we then proceed to the multivariate analysis techniques themselves.

4.1 Graph Covariance and Correlation Measures

In order to analyze the structure of covariances (and/or correlations) among structures, it is clearly imperative for us to define what we mean by such terms. In general, the approach we follow is based on the treatment of edge structures as random variables, with the moments of these variables defined accordingly. From a data analytic perspective, we can be thought of as operating on (possibly permuted) vectorized representations of the graph adjacency matrix. While this approach seems intuitive enough, it is necessary to proceed carefully so as to ensure that measures defined in this fashion will in fact behave in a manner consistent with expectations. We start, then, by building our formalisms with the simplest, uniquely labeled case (for which applications already exist) and then moving to the more general problem of structural covariance on arbitrary graph sets.

4.1.1 Uniquely Labeled Case

We begin our discussion by reviewing the simplest case, in which $\mathcal{P}_G$ contains only one member and which therefore allows us the luxury of dealing solely with the uniquely labeled graph, H. As usual, we shall treat H as a digraph, and we will further allow loops. For purpose of this discussion, δ should be taken as defined in Equation 6, although the δ of Equation 1 can be employed in the dichotomous case. Missing data – including loops or off-triangle entries in the simple graphic case – should be omitted when calculating the values given below, and normalizing constants should be adjusted accordingly.

Given this, let us begin with the definition of the *graph mean*; this is an estimator of the expectation of a uniform draw from the graph's edge set. In the dichotomous case, the

graph mean corresponds to the graph's density, and it is here denoted by

$$\overline{\delta_H} = \frac{1}{|V_U|^2} \sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \delta_H(x,y) \tag{12}$$

With an expression for the graph mean, we are now prepared to define the covariance between two labeled graphs, H_i and H_j. The *graph covariance* is the first product-moment of two graphs' edge set variables about their graph means or formally

$$Cov(H_i, H_j) = \frac{1}{|V_U|^2} \sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \left(\left(\delta_i(x,y) - \overline{\delta_{H_i}} \right) \left(\delta_j(x,y) - \overline{\delta_{H_j}} \right) \right) \tag{13}$$

Note that the *graph variance*, a natural generalization of the standard variance, can be defined easily using the above as the covariance of a graph with itself. The interpretation, again, is one of the second moment of the distribution formed by uniform draws from the graph's edge set, and the definition is given by:

$$Var(H) = Cov(H,H) \tag{14}$$

$$= \frac{1}{|V_U|^2} \sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \left(\delta_H(x,y) - \overline{\delta_H} \right)^2 \tag{15}$$

We may now proceed to define the product-moment correlation between two labeled graphs, using the notions of graph variance and covariance defined above. This is presented in Equation 16:

$$\rho(H_i, H_j) = \frac{Cov(H_i, H_j)}{\sqrt{Var(H_i) Var(H_j)}} \tag{16}$$

Note that ρ has the usual interpretation as a product-moment correlation between the edge sets of H_i and H_j taken as random variables.

4.1.2 General Case

As with the structural distance, we have in the case of the graph covariance a measure of association between structures which we wish to generalize to sets with arbitrary accessible permutation groups.

First, we observe that the graph mean (or density), given in Equation 12 is labeling invariant; $\overline{\delta_H}$ is constant under all permutations. This in turn implies that the graph variance (Equation 14) is also labeling invariant, since $\sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \left(\delta_H(x,y) - \overline{\delta_H} \right)^2$ cannot be affected by permutation, either. This seems intuitively reasonable: the graph mean and graph variance are properties of the distribution of edge values, rather than their arrangement, and as such should not be affected by reordering vertices.

This last intuition is an important one in considering the more general case. Consider an unlabeled graph, G, with accessible permutation group $\mathcal{E}_G : |\mathcal{E}_G| > 1$. Now let H_i and H_j be the labeled graphs produced by $L_i(G)$ and $L_j(G)$, respectively, where $L_i, L_j \in \mathcal{E}_G, L_i \neq L_j$. Then, from Equation 13,

$$Cov(H_i, H_j) = \frac{1}{|V_U|^2} \sum_{x=1}^{|V_U|} \sum_{y=1}^{|V_U|} \left(\left(\delta_{H_i}(x,y) - \overline{\delta_{H_i}} \right) \left(\delta_{H_j}(x,y) - \overline{\delta_{H_j}} \right) \right) \tag{17}$$

which is clearly different from $Var(H_i) = Var(H_j) = Var(G)$ (where $Var(G)$ can be thought of as the variance of any labeling of G), since in general $\delta_{H_i}(x,y) \neq \delta_{H_j}(x,y)$[35]. For the generalized *structural covariance*, however, we naturally seek to define the measure such that the usual relationship between the variance and the covariance continues to hold. Specifically, we require that

$$Cov_S(L_i(G), L_j(G)) = Var(G) \quad \forall\, G,\ L_i, L_j \in \mathcal{P}_G \tag{18}$$

As it happens, the constraint of Equation 18 places a strong constraint on the structural covariance. For a single unlabeled graph G with specified accessible permutation group $\mathcal{P}_G$, the graph variance is necessarily the maximum covariance over all pairs of permutations of G within $\mathcal{P}_G$. This simple but important result is presented in Theorem 2:

Theorem 2 (Structural Covariance). *Given an unlabeled graph G with accessible permutation group $\mathcal{P}_G$, $Var(G) = \max_{L_a, L_b \in \mathcal{P}_G} Cov(L_a(G), L_b(G))$.*

Proof. First note that from the definition of the graph variance, $Var(G) = Var(L_a(G)) = Var(L_a(G)) \ \forall\ L_a, L_b \in \mathcal{P}_G$. For any two random variables X and $Y : Var(X) = Var(Y)$, however, $Cov(X,Y) \leq Var(X)$. It therefore follows that $Var(G) \geq Cov(L_a(G), L_b(G))$. Trivially, the identity permutation guarantees that $Var(G) = Cov(L_a(G), L_b(G))$ for at least one choice of $L_a, L_b \in \mathcal{P}_G$ (namely any in which $L_a = L_b$); thus $Var(G) = \max_{L_a, L_b \in \mathcal{P}_G} Cov(L_a(G), L_b(G))$. □

Observe that a direct consequence of the Structural Covariance Theorem is that, to satisfy the requirement of Equation 18, the structural covariance must be given by

$$Cov_S(G_i, G_j \,|\mathcal{P}_i, \mathcal{P}_j) = \max_{L_a \in \mathcal{P}_i, L_b \in \mathcal{P}_j} Cov(L_a(G_i), L_b(G_j)) \tag{19}$$

in the general case. Interestingly, this definition parallels the form of the structural distance (Equation 4), although here agreement is maximized whereas in the former we are lead to minimize disagreement. As this might suggest, the structural distance and structural covariance are connected in a number of respects; we shall not pursue this matter here, however.

[35] This follows from the fact that $\delta_{H_i}(x,y) = \delta_{H_j}\left(L_j\left(L_i^{-1}(x)\right), L_j\left(L_i^{-1}(y)\right)\right)$.

Given a concept of structural covariance, the generalization of the graph correlation follows quite naturally. Applying Equations 14 and 19 to the definition of the graph correlation, we define the structural correlation as

$$\rho_S\left(G_i, G_j \,|\mathcal{P}_i, \mathcal{P}_j\right) = \frac{Cov_S\left(G_i, G_j \,|\mathcal{P}_i, \mathcal{P}_j\right)}{\sqrt{Var\left(G_i\right) Var\left(G_j\right)}} \tag{20}$$

The most straightforward interpretation of ρ_S is that of the graph correlation between the two most strongly correlated labeled graphs in the sets of labeled graphs created by the action of two accessible permutation groups on their respective unlabeled graphs. If $\rho_S = \pm 1$, then there is a linear transformation and permutation which takes the edge set of one graph onto the edge set of the other; for dichotomous data, $\rho_S = 1$ implies that the graphs being compared are isomorphic, and $\rho_S = -1$ implies that they are complements of each other.

4.2 Structural Regression

One obvious application of the structural covariance is to regression analysis of social networks. The basic framework for network regression on labeled graphs was first articulated by Krackhardt (1988); here we generalize this approach for use on graph sets with arbitrary accessible permutation groups.

In order to set up the structural regression framework, it helps first to define some notation. In addition to the notions of the edge value function, the graph mean, variance, and covariance, and the structural counterparts to these measures, it is convenient to specify an interpretation of the application of arithmatic operators to labeled graphs. For purposes of this paper, then, we define such operations as follows:

Definition 5 (Graph Arithmatic). Let $\circ$ and $\triangle$ be binary and unary operators (respectively) taking $\mathbb{R} \rightarrow \mathbb{R}$. Given two labeled graphs H_i, H_j with vertex sets $V\left(H_i\right) = V\left(H_j\right) = V_U$ and adjacency matrices $\mathbf{A_i}$ and $\mathbf{A_j}$, define $H_i \circ H_j$ to be the graph $H_{H_i \circ H_j} = \{V_U, E\}$ with adjacency matrix $[\mathbf{A}_{xy} = \mathbf{A_i}_{xy} \circ \mathbf{A_j}_{xy}]$. Similarly, define $\triangle H_i$ to be the graph $H_{\triangle H_i} = \{V_U, E\}$ with adjacency matrix $[\mathbf{A}_{xy} = \triangle \mathbf{A_i}_{xy}]$.

Properly, then, arithmatic operations on one or more graphs are interpreted as elementwise operations on the graphs' respective adjacency matrices (or, equivalently, as elementwise operations on the graphs' edge set variables). With this shorthand, we can conveniently describe the conventional network regression of Krackhardt (1988) as fitting the following model to a single labeled response graph (H_y) given a set of labeled regressor graphs ($\{H_1, \cdots, H_k\}$):

$$E\left(H_y \,|\beta, H_1, \cdots, H_k\right) = \beta_0 + \beta_1 H_1 + \beta_2 H_2 + \ldots + \beta_k H_k \tag{21}$$

with the β values generally chosen to minimize the MSE. Issues involved in fitting the standard network regression model (as well as hypothesis testing) are detailed in Krackhardt (1987; 1988).

A major limitation of this model with respect to inter-structural analysis is the fact that it assumes that all graphs to be analyzed are uniquely labeled. In the more general case, of course, we cannot rely upon this assumption, and hence we require a generalization of this method which is applicable to graph sets with arbitrary permutation groups. In particular, we seek to identify a linear combination of edge variables together with a correspondence-maximizing choice of accessible permutations which will maximize the structural correlation between the graph formed by said linear combination and the response structure. We refer to this as the *structural regression model*, and define it as follows:

Definition 6 (Structural Regression). Given the graphs $H_y, H_1, H_2, \ldots, H_k$, the *least squares structural regression estimator* of H_y, $\hat{H}_y$, is given by $\hat{H}_y = \hat{\beta}_0 + \hat{\beta}_1 L_1(H_1) + \hat{\beta}_2 L_2(H_2) + \ldots + \hat{\beta}_k L_k(H_k)$ with $\hat{\beta} \in \mathbb{R}^{k+1}$ and $L_{H_i} \in \mathcal{P}_{H_i}\ \forall\ H_i$ chosen such that the multiple structural correlation, $\rho_S\left(H_y, \hat{H}_y\right)$, is maximized.

That the structural regression estimator is indeed a least squares estimator (conditional on the choice of permutations) follows trivially from the above definitions, and from the fact that the multiple correlation is maximized by the LSE. When $|\mathcal{P}_G| = 1$ for all G within the structural regression model, we are left with the familiar network regression model of Krackhardt (1988). This last leads us to one means of estimating the structural regression coefficients: choose the values of L uniformly from the acceptable permutation groups; fit a standard network regression conditional on the labelings selected, recording both the coefficients and the multiple structural correlation; repeat these steps multiple times, keeping the estimates associated with the maximum multiple structural correlation; and, finally, after a specified set of replications, keep the best estimates (as defined above). This Monte Carlo approach is simple to implement, albeit fairly inefficient, and improvements based on alternative heuristic optimization procedures (or exhaustive search, for very small graphs) are possible.

4.2.1 General Linear Structural Modeling

While we have focused here on extending network regression to the case of arbitrary labelings, it should be noted that this can be further generalized by considering alternatives to the model of Definition 6 above. Indeed, we can easily apply the same logic to the general linear model (McCullagh and Nelder, 1989)), which permits treatment of a much broader range of inter-structural relationships[36]. This model, which we call the general linear structural model (GLSM), may be defined as follows:

Definition 7 (General Linear Structural Model). Given a graph set $H_y, H_1, H_2, \ldots, H_k$, we define the *general linear structural model* as follows. Define $H_\eta = \beta_0 + \beta_1 L_1(H_1) + \cdots + \beta_k L_k(H_k)$ to be the *linear structural predictor* of H_y, where it is assumed that for some smooth, invertible *link function*, l, with inverse m, $\delta_{H_\eta} = l\left(E\left[\delta_{H_y}(i,j)\right]\right) =$

[36] As, for instance, logistic network regression, which is far more appropriate than conventional network regression for dichotomous data.

$m^{-1}\left(E\left[\delta_{H_y}(i,j)\right]\right)$. We then take the density of δ_{H_y} to be given by $f\left(\delta_{H_y}(i,j)\,|\theta_{ij},\psi\right) = exp\left(w_{ij}\left(\delta_{H_y}(i,j)-\gamma(\theta_{ij})\right)/\psi+\tau\left(\delta_{H_y}(i,j),\psi/w_{ij}\right)\right)$, where ψ is a scale parameter, θ_{ij} is an invertible function of $E\left[\delta_{H_y}(i,j)\right]$, and w_{ij} is a prior weight.

In general, we may attempt to fit the GLSM by maximizing the likelihood function across both the model parameters and the joint set of accessible permutations. Since the GLSM trivially becomes a standard GLM for fixed L, this maximization can be decomposed (as usual) into a standard model fitting problem and a problem of maximizing the best fit across the permutation set. Pursuing this problem is beyond the scope of this work, but it is expected that heuristic optimization techniques like those employed for other structural inference problems will be applicable here.

4.2.2 An Example of Structural Regression

For a brief demonstration of structural regression, we turn to a data set of Linda Wolfe concerning interaction among members of a troop of monkeys in Ocala, Florida. Information on interaction was collected over a three-month period, with interaction defined as co-presence at a particular site (summed for each dyad across all observation periods). In addition, Wolfe collected data on apparent rank within the troop, age and gender of troop members, and troop kinship ties (parentage). To what extent can the structure of interaction be expressed as a linear combination of primate attribute differences and kinship ties? To find out, we perform a structural regression of interaction on kinship, absolute differences in age (measured in years), and absolute differences in gender (dichotomously coded).

Because our dependent variable consists of event counts (discrete episodes of co-presence), we employ a GLSM using a poisson likelihood and logarithmic link function as our structural regression model. The results of the fit for the uniquely labeled case are shown in the first set of columns in Table 1. As can be seen, this model has surprisingly little explanatory power: although gender and rank differences have significant effects under a z test[37] (with gender and rank differences multiplying expected interaction by factors of approximately 1.4 and 0.97, respectively), the overall model fit is quite poor. In addition to the goodness-of-fit measures in Table 1[38], the top left panel of Figure 5 depicts the predicted versus actual interaction levels for all dyads; as the figure shows, kinship relations and attribute differences between particular individuals provide little or no indication of the extent to which those individuals will be observed to interact. Finally, as a somewhat better test of the null hypothesis that the structures in question are unrelated, we employ a conditional uniform graph (CUG) test of the AIC statistic. For this test we draw new sets of predictor graphs by bootstrapping (with replacement) from the edge distributions of the existing predictors, finding the AIC of the fitted model for each replication[39]. The proportion of replicate AIC

[37] Due to potential autocorrelation within the data set, these p-values should be taken as heuristic indicators of relationship strength only.

[38] The pseudo-R^2 measure used here is $1-\frac{\text{Resid Dev}}{\text{Null Dev}}$.

[39] Compare this to the QAP null hypothesis (Krackhardt, 1987), in which new predictors are drawn by randomly permuting rows and columns. More will be said about this presently.

	Uniquely Labeled Case			Unlabeled Case		
	$\hat{\beta}$	(StdErr)	p-value[a]	$\hat{\beta}$	(StdErr)	p-value[a]
(Intercept)	1.41^{***}	(0.0601)	<2e-16	1.67^{***}	(0.0540)	<2e-16
Kinship	0.148	(0.124)	0.233	0.354^{**}	(0.117)	0.00251
Gender Difference	0.353^{***}	(0.0620)	1.33e-08	-0.527^{***}	(0.0598)	<2e-16
Age Difference	-0.00798	(0.00879)	0.364	-0.185^{***}	(0.0110)	<2e-16
Rank Difference	-0.0256^{***}	(0.00675)	0.000153	0.0769^{***}	(0.00602)	<2e-16
Null deviance	734.46 (379 df)			734.46 (379 df)		
Residual deviance	696.17 (375 df)			286.89 (375 df)		
Pesudo-R^2	0.052			0.609		
	AIC	$p \leq$ AIC	$p \geq$ AIC	AIC	$p \leq$ AIC	$p \geq$ AIC
CUG AIC Test[b]	1825.6^{***}	> 0.999	< 0.001	1416.3^{***}	> 0.999	< 0.001

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, [a] p-values based on z-test; see text regarding interpretation

[b] CUG p-values reflect 1000 replications

Table 1: Structural Regression Model Fits, Wolfe Primate Data

values which are less than or equal to the original AIC value then provides a (one-tailed) p-value for the null hypothesis that the observed model likelihood is typical of what would be obtained by random structures with the same edge distributions as the original graphs. As Table 1 indicates, the AIC is well below that of the population of similar random structures, suggesting that there is at least *some* connection between the labeled structures. Taken as a whole, however, the analysis suggests that the relationship is extremely weak.

Although the interactions among particular individuals are only weakly accounted for by kinship and individual differences, what about the underlying structures? Could there still be a strong relationship between the general *patterns* of difference and kinship and those of intra-troop interaction? Some idea of this can be gleaned by repeating the above analysis under an assumption of full exchangability among individuals. Using the likelihood maximization approach discussed above, we estimate the GLSM model (using a simulated annealer) under conditions of full exchangeability. The results of this analysis are presented in the second set of columns of Table 1; as can be seen, there is a strong relationship between the *underlying* structure of interaction and the underlying structures of kinship and difference. In particular, the underlying structure of interaction closely resembles the a combination of the structure of kinship, rank differences, and the complements of age and gender differences (respectively). Although the p-values associated with the z-test should be interpreted only heuristically (as the null hypothesis they test is something of a straw man in this case), the properties of model fit are good (see also the top right panel of Figure 5) and a CUG test comparison with bootstrapped unlabeled graphs indicated that the fit obtained for these unlabeled graphs is far higher than what would be expected from

a more reasonable null model. Thus, we can reasonably conclude that while there is little connection between interaction rates and other structuaral properties for particular dyads in the Wolfe data, there are nontrivial relationships between the corresponding underlying structures.

4.3 Network Principal Component Analysis

Principal component analysis (Jackson, 1991; Joliffe, 1986) is a versatile technique which attempts express a variable set in terms of an ordered set of synthetic variables such that each accounts for the maximum possible variance remaining in the set, conditional on being orthogonal to the synthetic variables already defined, and on being a linear combination of the original variables. Mechanically, principal component analysis (PCA) is an analysis of the eigenstructure of the covariance (or sometimes correlation) matrix of a variable set; it is thus closely related both to classical MDS and to more modern techniques such as factor analysis and structural equation modeling. As a powerful technique with wide applicability, PCA would a seem to be a promising candidate for use in inter-structural analysis, and here we develop an extension of the method to the study of graph sets.

4.3.1 Description of the Approach

The basic approach taken here is to construct the matrix of structural covariances (or correlations) of networks within the set to be examined, and to subject this matrix to the PCA procedure. (This is without loss of generality; the discussion which follows will also apply to the special case of uniquely labeled graphs.) This procedure is equivalent to identifying the eigenvalues and eigenvectors of the matrix in question, and algorithms to perform this calculation are already implemented in a wide range of mathematical and statistical computing environments. Each extracted eigenvector of the covariance matrix will be real-valued, and corresponds to a given synthetic variable; the components of the eigenvector correspond to the elements of the linear combination taking the original variable set into the synthetic variable (the loadings). Under standard assumptions, the eigenvalues of the covariance matrix are further guaranteed to be real and nonnegative, corresponding (within a constant factor) to the variance accounted for by the associated synthetic variables. While few distributional assumptions are required for these conditions to be satisfied[40], it *is* necessary that the matrix to be analyzed be positive semi-definite (p.s.d). Before proceeding to the use of the method, then, we must first establish that this condition can be satisfied for graph and structural covariance matrices.

In the uniquely labeled case, the graph covariance is simply the standard product moment of the vectors corresponding to the vectorized adjacency matrices of a given pair of graphs about their means; thus, the covariance matrix is guaranteed to be positive semi-definite (Mardia et al., 1979) and its eigenstructure can be interpreted in the usual fashion. In the general case, by contrast, it is not at all obvious that the structural covariance matrix will be

[40] E.g., it is not required the data be drawn from any particular joint distribution.

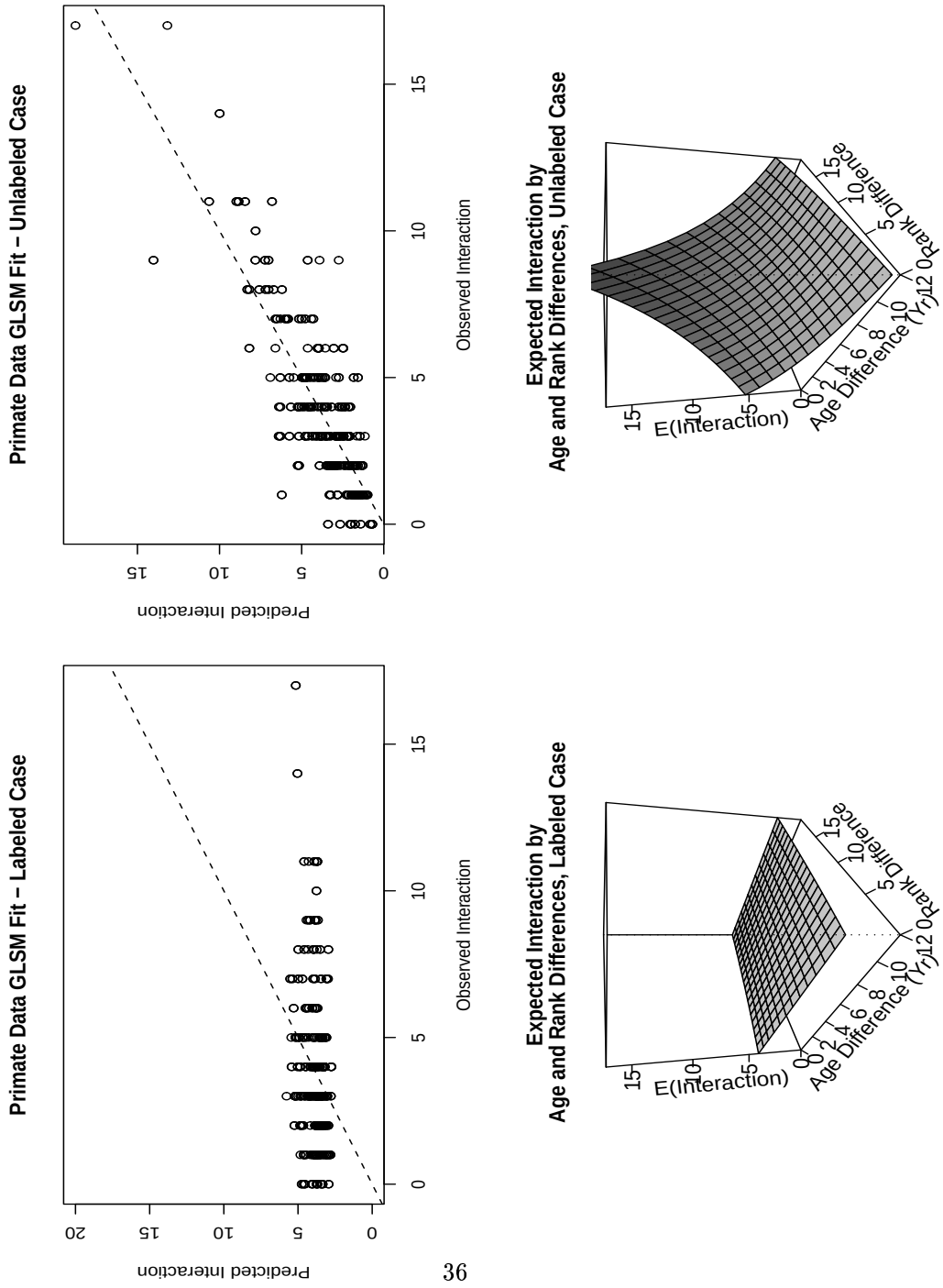

Figure 5: Structural Regression Model Fits and Expected Interaction by Differences in Age and Rank, Wolfe Primate Data

p.s.d., making interpretation of the eigenstructure potentially problematic. As it happens, however, we can show that the structural covariance matrix will be p.s.d. in all cases and p.d. so long as the edge sets of the all graphs within the set are linearly independent under all accessible permutations. This result is given in the following theorem:

Theorem 3 (Structural Covariance Matrices are P.S.D.). *Let Σ be the structural variance-covariance matrix formed by Cov_S on all pairs of graphs within the set $\{G_1, G_2, \ldots, G_k\}$ with accessible permutation groups $\{\mathcal{P}_{G_1}, \mathcal{P}_{G_2}, \ldots, \mathcal{P}_{G_k}\}$ (respectively). Then Σ is positive semi-definite. Further, iff $\nexists\, G \in \{G_1, \ldots, G_k\}, \Gamma \subset \{G_1, \ldots, G_k\}, G \notin \Gamma : L_G(G) = \sum_{i=1}^{|\Gamma|} a_i L_{\Gamma_i}(\Gamma_i)$ for any $L_a \in \mathcal{P}_a$, Σ is positive definite.*

Proof. We begin by noting that a necessary and sufficient condition for a real matrix $\mathbf{A}$ to be p.s.d. is that $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0\ \forall\ \mathbf{v} \in \mathbb{R}^n$. Our procedure is then to demonstrate that $\forall\ \mathbf{v} \in \mathbb{R}^n\ \exists\ G : Var(G) \leq \mathbf{v}^T\Sigma\mathbf{v}$; since the graph variance is strictly nonnegative, this implies that Σ is p.s.d.

The graph we shall use for purposes of this proof is $\sum_{i=1}^{k} v_i G_i$ (with $\mathbf{v}$ as given above), using the notion of graph addition developed earlier (see Definition 5). Note here that, as indicated previously, the expectation of a graph is interpreted in the same manner as the graph mean; that is, as the expectation of the edge set. The graph variance of $\sum_{i=1}^{k} v_i G_i$ is then given by

$$
\begin{aligned}
Var\left(\sum_{i=1}^{k} v_i G_i\right) &= E\left[\left(\sum_{i=1}^{k} v_i G_i - E\left[\sum_{i=1}^{k} v_i G_i\right]\right)^2\right] && \text{(by Eq. 14)} && (22)\\
&= E\left[\left(\sum_{i=1}^{k} v_i G_i - \sum_{i=1}^{k} v_i E[G_i]\right)^2\right] && \text{(by defn of } E[G]) && (23)\\
&= E\left[\left(\sum_{i=1}^{k} v_i \left(G_i - E[G_i]\right)\right)^2\right] && && (24)\\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} v_i v_j \left(G_i - E[G_i]\right)\left(G_j - E[G_j]\right) && && (25)\\
&\leq \sum_{i=1}^{k}\sum_{j=1}^{k} v_i v_j Cov_S\left(G_i, G_j \,\middle|\, \mathcal{P}_{G_i}, \mathcal{P}_{G_j}\right) && \text{(by Eq. 19)} && (26)\\
&= \mathbf{v}^T\Sigma\mathbf{v} && \text{(by defn } \Sigma) && (27)
\end{aligned}
$$

Since the graph variance is strictly nonnegative, it follows that Σ must be p.s.d.

To prove our last assertion, recall that a real p.s.d. matrix $\mathbf{A}$ is p.d. unless $\exists \mathbf{v} \in \mathbb{R}^n : \mathbf{v} \neq \mathbf{0}^n, \mathbf{v}^T\mathbf{A}\mathbf{v} = 0$. If $\exists G \in \{G_1, \ldots, G_k\}, \Gamma \subset \{G_1, \ldots, G_k\}, G \notin \Gamma : L_G(G) = \sum_{i=1}^{|\Gamma|} a_i L_{\Gamma_i}(\Gamma_i)$ for some accessible permutations L, then it follows that $\exists \mathbf{v} : \mathbf{v}^T\Sigma\mathbf{v} = 0$ (constructed by

setting $v_G = 1$, $v_{\Gamma_i} = a_i$, and all other elements equal to 0). The converse follows trivially from the properties of linear dependence. □

Thus we can, in fact, apply principal component analysis to structural covariances on graph sets with arbitrary accessible permutation groups. Given this preliminary, we may now consider how the output of the procedure should be interpreted when applied to graph sets.

For the labeled case, interpretation of the network PCA is as follows: each component corresponds to an abstract structure – with edge set composed of linear combinations of the observed structures – which accounts for a portion of the tie variance within the observed structure set. These structures, in a sense, form a set of "archetypes" or building-blocks from which the observed structures can be reconstructed. Indeed, taking the scores on each variable will construct the archetypal structure set; note, however, that these will necessarily be valued graphs, due to the nature of the procedure. By examining the archetypal structures accounting for the most variance, it should be possible to identify important features which exist across observed structures. Likewise, loading patterns should reveal clusters of structures with similar properties, as per their interpretation under standard PCA. Because of the dichotomous nature of standard network data, it is notable that there are limitations to the interpretations which can be made regarding the status of the identified components: they cannot, in particular, be assumed to have desirable sampling properties. Fortunately, however, the principal component analysis can be interpreted in a strictly data-analytic fashion (Dillon and Goldstein, 1984), and normality assumptions are not required for such a use.

In the more general case, the interpretation of components as archetypal structures continues to hold, but these structures are not (necessarily) uniquely labeled. Reconstruction of the observed structures from the archetypes constitutes both linear combination and relabeling operations, with the latter chosen so as to maximize structural correspondence. This does not affect the interpretation of loadings (eigenvectors) or eigenvalues, but does complicate the extraction of structural archetypes; for many purposes, however, it is not necessary to examine these synthetic structures directly (just as synthetic variables are often used only indirectly in conventional PCA), and hence this complication is not expected to affect use of the technique in the majority of situations.

4.3.2 An Example of Network PCA

To demonstrate the use of principal component analysis on multiple networks, we will examine a data set of Carley et al. (1993) consisting of cognitive association networks. As part of an ongoing study of team learning in an information systems (IS) context, 40 undergraduate students in an IS class were asked to explain the notion of "information system" in their own words. From these answers 130 distinct concepts were coded, and a concept network was formed for each subject wherein (for each pair of concepts, A and B) an edge went from concept A to concept B if and only if concept B was mentioned within a set number of words of a mention of concept A (i.e., the concepts were proximity coded). These concept

networks, then, can be taken to be structural encodings of each subject's communicated understanding of the term "information system." Given this, a number of questions come to mind. To what extent do the subjects' explanations of "information system" share one or more common substructures? Are subjects' explanations in broad agreement, or does a diversity of explanations exist? What is the nature of the strongest common theme within the subjects' explanations? While these are all qualitative questions, we may find some rather quantitative answers by applying the network PCA algorithm to the set of subject concept networks.

To perform a network principal component analysis, we first calculate the graph correlation matrix of all subject concept networks; in this case, concepts have been coded for the express purpose of cross-subject comparison, and hence we treat all graphs as uniquely labeled. From the graph correlation matrix we extract the eigenstructure, recalling that each eigenvector corresponds to a set of principal component loadings and that each associated eigenvalue indicates (to within a scalar multiple) the proportion of total variance associated with said component. A scree plot of these variance proportions (rescaled eigenvalues) is shown in Figure 6. Out of the 40 components extracted, the first accounts for approximately 13% of the total concept network variance (over twice that of the next largest component), with subsequent components accounting for under 6% each. In terms of our initial question, the pattern of eigenvalues suggests one archetypal structure (theme) which accounts for far more of the total concept structure variance than any other, with other archetypes being relatively similar in terms of their contributions to subjects' explanations. The presence of such a large number of subsidiary concept archetypes might suggest a relatively high degree of diversity in subject accounts, an impression which is also corroborated by an examination of the loadings (eigenvector elements) on the first component. These are presented in Figure 7, and plainly show a wide range in the extent to which even the most common structural archetype contributes to individual subject explanatory structures.

Having answered our first two questions, we now turn towards the third: what is the nature of the archetypal structure associated with the first principle component? To answer this question, we must first find the structure in question; this is accomplished by finding the scores for all edge variables in the concept network on the first principal component. The resulting structure – a valued graph – can then be thought of as a basic "building block" from which individual explanations are partially constructed. Analysis of the archetypal structure can be performed in a number of ways, but here we limit ourselves to a visual examination of the sociogram formed by the strong edges (those with values greater than 0.5) of the graph associated with the first principal component. This is shown in Figure 8 (isolates omitted for clarity). Note that due to the nature of the proximity encoding, it is possible to infer approximate meaning from the directed relation: thus, for instance, we find that "information system" $\rightarrow$ "manages" $\rightarrow$ "information", "information system" $\rightarrow$ "aids" $\rightarrow$ "user", "computers" $\rightarrow$ "information system", etc. This "diamond"-shaped structure of associations (see Figure 8) implies that the strongest common theme in individual responses consists of a series of statements concerning the ways in which information systems act upon information. Further investigation might relate incorporation of this theme to individual performance or

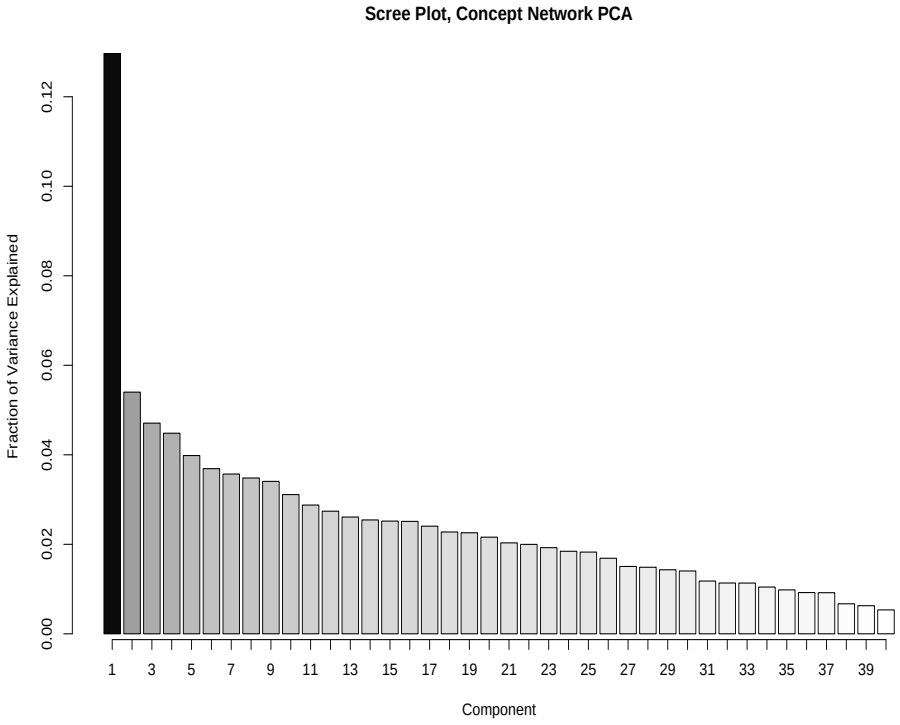

Figure 6: Scree Plot for PCA of Concept Network

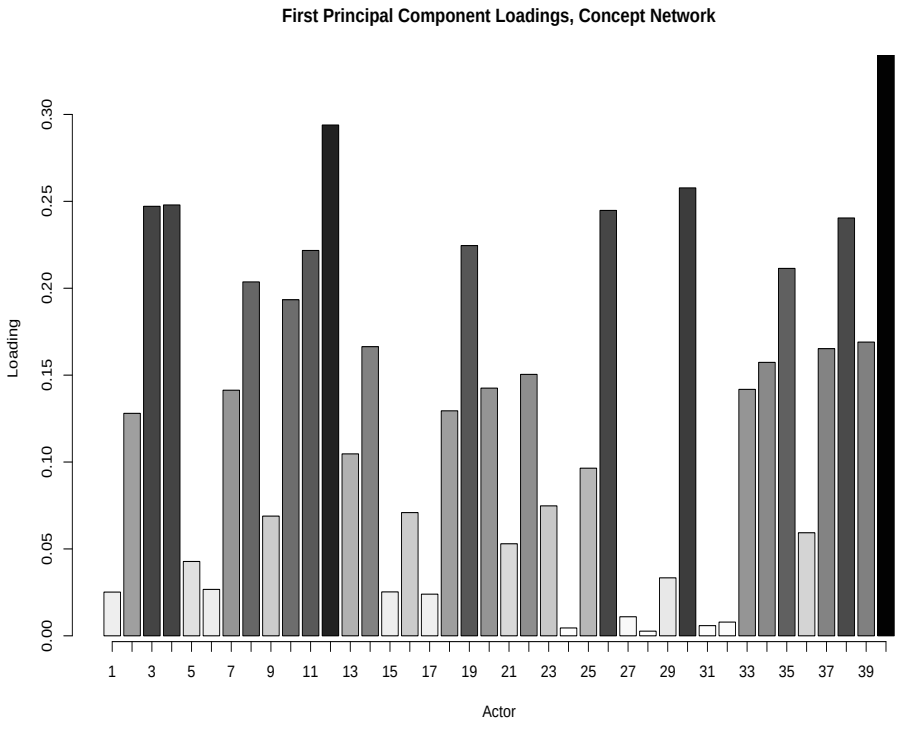

Figure 7: Loadings on the First Principal Component, Concept Network

competency (e.g., are higher-performing IS workers more likely to incorporate the common understanding of information system concepts into their explanations), or examine the structure of additional themes (e.g., do these less commonly-shared archetypes reflect specialized knowledge).

4.4 Canonical Correlation for Graph Sets

Canonical correlation analysis (Hotelling, 1936; Dillon and Goldstein, 1984) is a powerful – if often poorly-understood – multivariate analysis technique. It is a fairly "deep" method in the sense that it is closely related to (or, in some cases, can even be reframed as) multivariate regression, MANOVA, discriminant analysis, factor analysis, and PCA (among others). In its most common interpretation, canonical correlation analysis (CCA) can be thought of as identifying a set of synthetic variables (*canonical variates*) composed of linear combinations of the original variables in two variable sets, such that the correlation between synthetic variable pairs is maximal. For this reason, then, CCA is unaffected by linear transformations of either variable set independently, and is ideal when a quantity of theoretical importance is specified only up to a linear combination of secondary variables[41]. More generally, the canonical correlation analysis provides a versatile means of determining the extent to which two sets of variables are linearly related to one another, and further of identifying the nature of the specific relations which exist between the sets. Although (as is generally the case with such methods) multivariate normality is required for hypothesis testing, this can be satisfied approximately for large sample sizes (here, large graphs), and the strictly data analytic interpretation of the technique requires no distributional assumptions (Dillon and Goldstein, 1984). Given the above, then, it seems both reasonable and profitable to generalize canonical correlation analysis to the problem of inter-structural analysis.

4.4.1 Description of the Approach

Given the results we have already established, the application of CCA to inter-structural analysis is quite straightforward. To begin, we assume that we have identified two sets of structures, X and Y. Taking Σ to be the structural covariance matrix of the full structure set, we can express the Σ in terms of the following submatrix decomposition:

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{YY}} \end{bmatrix} \tag{28}$$

Thus, we can think of the combined structural covariance matrix in terms of the structural covariance matrices of each set, along with the between-set covariance matrices. The problem solved by the canonical correlation analysis, then, can be formulated as

$$\mathbf{B}^{-1}\mathbf{A}\hat{\mathbf{v}} = \lambda\hat{\mathbf{v}} \tag{29}$$

[41] E.g., as is often the case for problems in which multiple indicators are used to represent a hypothesized construct.

Figure 8: Strong Edges in the First Principal Component, Concept Network

where

$$\mathbf{A} = \begin{bmatrix} 0 & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Sigma_{\mathbf{XX}} & 0 \\ 0 & \Sigma_{\mathbf{YY}} \end{bmatrix}, \text{and } \hat{\mathbf{v}} = \begin{bmatrix} \overline{\delta_X}\hat{\mathbf{v}}_{\mathbf{X}} \\ \overline{\delta_Y}\hat{\mathbf{v}}_{\mathbf{Y}} \end{bmatrix}. \tag{30}$$

The elements of $\hat{\mathbf{v}}$ (the eigenvectors) are the canonical coefficients (analogous to beta weights in a standard regression context), and their corresponding eigenvalues are the associated canonical correlations. That we can perform this operation using the structural covariance matrix is guaranteed by the result of Theorem 3, assuming that there are no strict linear dependencies within either graph set. Likewise, interpretability of the resulting eigenstructure can be shown from the same theorem: from Theorem 3 and the above assumption, $\mathbf{B}$ is positive definite and $\mathbf{A}$ is at least positive semi-definite. It is then a standard result of linear algebra that all of the nonzero eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$ are positive, and the eigenstructure can be interpreted in the usual fashion.

It should be noted that the formulation of Equation 29 can be used to describe a variety of related procedures: letting $\mathbf{A} = \Sigma$ and $\mathbf{B} = \mathbf{I}$, for instance, results in a principal component analysis; replacing $\mathbf{B}$ with the identity matrix results in partial least squares; and setting the lower right element of $\mathbf{B}$ to the identity matrix provides the formulation for multivariate linear regression. With the exception of PCA (which has already been treated) we do not consider these alternative analyses at length. Nevertheless, similar procedures to those illustrated here should allow these techniques to be applied to inter-structural analysis.

4.4.2 An Example of Canonical Correlation on Graph Sets

For an illustrative example of canonical correlation on graph sets, we here turn to a well-known data set from Wasserman and Faust (1994). This set consists of five relations on a set of 24 countries chosen to exhibit a variety of developmental classes: trade in food and/or animals; trade in crude materials; trade in mineral fuels; trade in basic manufactured goods; and diplomatic exchange. For each nation (ego), incoming edges to ego were coded present if (for the year 1984) imports from alter amounted to at least 0.01% of ego's total imports of said commodity in US$. (Diplomatic exchange is an exception: there, ego sends an edge to alter if and only if ego had an embassy or high commission in alter for the year 1984.) In addition to these five network relations, we also consider the relations formed by pairwise absolute differences on four national attributes: annual population growth rate (1970-81); annual per capita GNP growth rate (1970-81); secondary school enrollment ratio (1980); and energy consumption per capita in kilograms of coal equivalent (1980). Taking these two sets of relations – network ties of trade and diplomacy, on the one hand, and absolute differences in development on the other – we may reasonably inquire as to what relationships exist between the sets, and to the extent that these relationships account for the variance in either. With this in mind, we now turn to a canonical correlation analysis of the trade data[42].

[42] As our purpose here is purely demonstrative, we do not examine alternative transformations of the variable set or the like; we note, however, that logarithmic transformation of various input combinations did not result in a superior model fit.

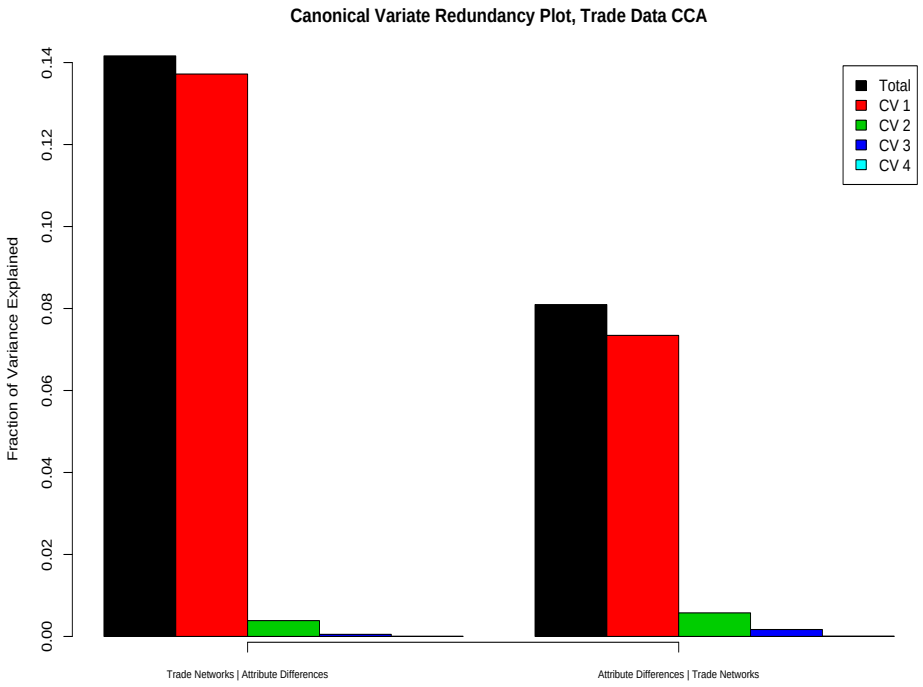

Figure 9: Canonical Variate Redundancies, Trade Data

Since our data consists of multiple relations on a single set of countries, and since our theoretical interest is in the relations among these countries *per se*, we take all graphs to be uniquely labeled by country. Applying the structural covariance to the graph sets and fitting the model of Equation 29, we obtain the four canonical variates. The canonical correlation coefficients for these variates were found to be 0.506, 0.167, 0.0728, and 0.0184 (respectively). The canonical variate redundancies – the proportion of variance explained in each set by the other given each canonical variate – provide an index (analogous to the R^2) of the extent to which each set of relations can be used to predict the other; these are shown in Figure 9. Note that nearly all of the predictive power between the two data sets lies in the first canonical variate, and that this variate can be used to predict approximately 14% of the total variance in trade relations from attribute differences. The trade relations, by contrast, can predict only about 8% of the total variance in attribute differences between nations, but given that the latter proceed the former in time this asymmetry is not particularly surprising.

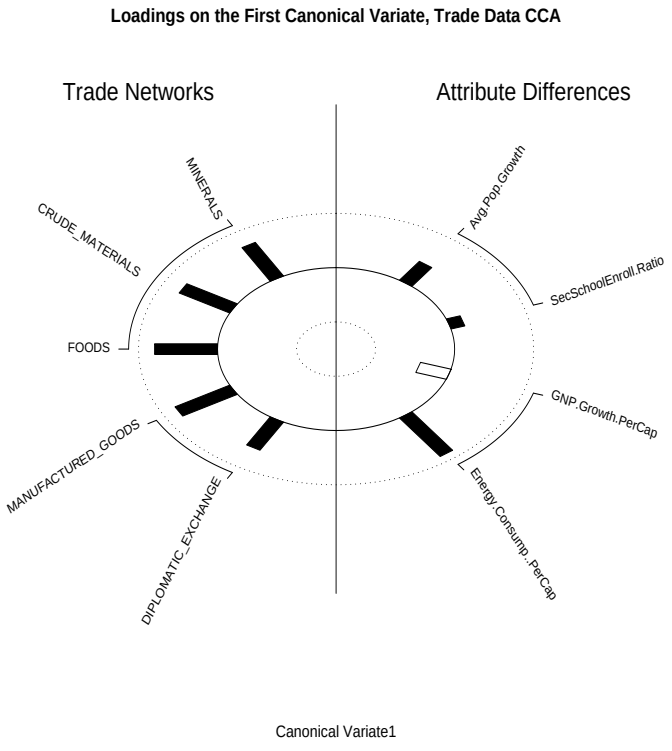

Figure 10: Loadings on the First Canonical Variate, Trade Data

Since the first canonical variate appears to account for nearly all of the shared variance between the two data sets (and for brevity), we shall focus exclusively on it. To interpret this variate, we first inspect the loadings[43] of our various trade and difference relations. These are shown in Figure 10. (In this Helio plot, positive loadings are represented with outward-pointing black bars and negative loadings are depicted with inward-pointing white bars; bar magnitude corresponds to correlation magnitude, with 1.0 and -1.0 at the outer and inner circles (respectively).) As can be seen, the first variate clearly represents a general tendency towards trade, with large positive loadings on all five of the trade/exchange relations. Simultaneously, this same variate also indicates a strong tendency to have had very different levels of per capita energy consumption, somewhat different rates of secondary education and population growth, and somewhat *similar* changes in per capita GNP growth.

Further insight into the nature of the relationships involved can be had by examining the

[43]These are also referred to as *structural correlations* in the literature, but we refrain from using this term here to avoid confusion.

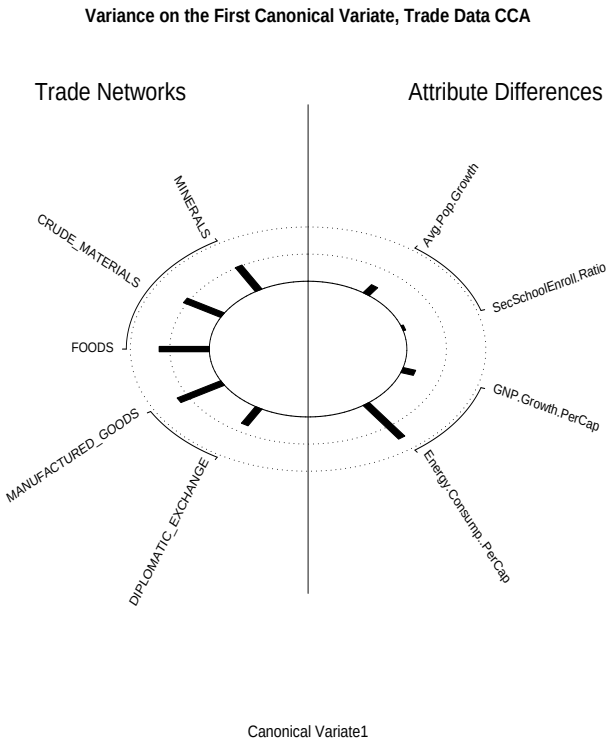

Figure 11: Variance Explained by the First Canonical Variate, Trade Data

proportion of variance in each original variable accounted for by the first canonical variate (the squared loadings)[44]. As the Helio plot of Figure 11 indicates, approximately half of the variance in each trade relation loads on the first canonical variate (with trade in foods and manufactured goods being notably higher, and diplomatic exchange being somewhat less). With respect to attribute differences, however, it is overwhelmingly the per capita energy consumption which loads heavily on the first canonical variate, the other variables having less than a quarter of their variances associated with it; overall, this variate accounts for approximately 29% of the variability of the attribute differences, versus approximately 54% of the variance in trade relations. Were it our intention to conduct a more in-depth analysis, we would examine other transformations of these variables, and would examine the full set of canonical variates; even this cursory analysis, however, serves to demonstrate the capacity of CCA to illuminate features of network data sets.

[44]Note that this is distinct from the fraction of each original variable which can be *predicted* from the other variable set on a given canonical variate.

5 Discussion

While we have tried to highlight the most critical aspects of the deployment of common multivariate methods for exploratory inter-structural analysis, many issues clearly remain. While limitations of space prevent us from considering all of these in detail, we do briefly discuss two – possible future expansions of the present methods and null hypothesis testing.

5.1 Areas for Future Development

Although we have demonstrated a number of different methods for inter-structural analysis, it is clear that these are merely a few of the wide array of possible approaches to the general problem of structural comparison[45]. Though limitations of space prevent our treating them in detail, two families of general approaches to multivariate modeling would seem to hold particular promise for application to inter-structural analysis: structural equation modeling (SEM), which provides an extremely general framework for the analysis of covariance matrices; and hierarchical Bayesian modeling (HBM) which is a particular approach to model building under the Bayesian paradigm. Here, we provide an outline which gives some flavor of each, along with some ideas as to how each could be applied to the problem of inter-structural modeling and analysis.

5.1.1 SEM of Multiple Relations

In the family of multivariate methods, structural equation modeling (Loehlin, 1987; Dillon and Goldstein, 1984) is one of the more recent, and perhaps controversial, approaches to the study of multiple, related variables. While sometimes difficult to evaluate, SEM methods have proven useful for exploring complex relationships within large variable sets. Though founded on normality assumptions which make them unsuitable for most network data, it is possible to apply SEM to a specific family of network analysis problems. The approach to be taken in many ways follows that described above for network covariance methods, although the scope is more limited.

Because of the fairly restrictive assumptions underlying SEM methods, the application of this approach should be restricted to continuously valued sociometric data (e.g., ratings of affinity, trust, prestige, etc.) on labeled node sets[46]. (Extension to the unlabeled case is possible, but may require some changes in modeling assumptions.) Depending on the specific approach being employed, an inter-structural SEM analysis would proceed either by multi-stage structural regressions, or by analysis of the structural covariance matrix. In either case, it should be possible to answer questions regarding (possibly causal) relationships between different relations within the group. The benefits of such an approach are obvious:

[45] Indeed, as we noted in Section 4.4.1 above, the whole of linear subspace modeling can be applied to the analysis of structural covariance matrices, though we have treated only PCA and CCA in detail.

[46] This may be overly conservative, since many fewer assumptions are required for the actual *fitting* of structural equation models than for their *evaluation*. (Indeed, ordinary OLS regression is itself simply a special case of SEM.) As different approaches to SEM vary in their sensitivity to normality assumptions, we expect that some approaches will prove more amenable than others to inter-structural applications.

connections between interpersonal contact, affinity, and homophily, for instance, have been the subject of much theoretical activity, and network SEM methods would potentially allow a number of extant theories to be tested directly. Particularly in group research, then, the network SEM would be a valuable addition to the methodological arsenal; if it proves possible to extend the method to the unlabeled case, the benefits could be even greater.

5.1.2 Bayesian Modeling of Graph Sets

Like SEM above, hierarchical Bayesian modeling (Gelman et al., 1995) provides a potentially powerful means of assessing the effects of multiple relations on one another within individual node sets, given a sample of structures. Unlike SEM, however, the Bayesian approach has a number of distinct advantages in the context of network analysis. First, hierarchical Bayesian modeling (HBM) is not founded on asymptotic arguments and hence it allows for the analysis of small data sets which would be inappropriate for SEM techniques. Similarly, the persistent problems of goodness-of-fit assessment which plague the SEM approach (particularly with either especially large or especially small data sets) are obviated under HBM due to the replacement of goodness-of-fit with posterior uncertainty and expected Bayes loss (Robert, 1994). Another advantage of the HBM perspective is the lack of a strong framework of *ex ante* stochastic assumptions: HBM allows for the construction of models which fit the characteristics of the particular problem being studied, rather than requiring problems to be forced into a particular stochastic modeling framework (e.g. linear relations between normally distributed random variables). Finally, HBM is arguably preferable in that it grounds inference on a Bayesian foundation, which can be justified on an axiomatic basis (see Robert (1994)).

Despite these advantages, however, hierarchical Bayesian modeling has the disadvantage of being fairly demanding in terms of researcher expertise, formalization of prior information, and (often) computation. Initial construction of an HBM must generally be done on a problem by problem basis, such that while a class of particular cases can be treated with the same model, truly novel situations generally demand a new derivation. Construction of HBMs usually requires careful attention to details of the stochastic model in question, which requires that the researchers have a clear understanding of the distributional characteristics of the phenomenon under study; other information from past research must be formalized in the development of prior distributions, which can in some cases be a difficult task. Lastly, hierarchical Bayesian models can prove reasonably computationally expensive, particularly for high-dimensional problems of the sort faced by structural analysts.

For all of the above reasons, then, HBM has only recently emerged as a viable alternative to classical methods for addressing complex problems of inference (despite its many advantages). The problem of inter-structural analysis, however, offers an excellent opportunity to use this approach. Specifically, the network paradigm presents a case in which many of the difficult distributional problems have straightforward solutions, and the HBM framework provides a simple but effective means of dealing with hitherto difficult problems such as informant inaccuracy (see, for instance, recent work by Butts (2000)). Here, we suggest the use of hierarchical Bayesian modeling to address the question of the relationship between

multiple relations on in the case of dichotomous data. The resulting model should have many applications which are similar to the SEM approach discussed above, but would be applicable to the dichotomous data sets commonly employed in the network world; more importantly, by specifying a particular stochastic modeling framework, HBM methods would provide a stronger basis for generalization of findings across data sets. The HBM approach, then, should be seen as a complement to the more exploratory techniques presented here.

5.2 Hypothesis Testing

In our exposition thusfar, it is notable that we have generally avoided the topic of null hypothesis testing for inter-structural analysis methods. Although known to be fundamentally problematic[47] (Robert, 1994), null hypothesis tests are nevertheless useful heuristic devices for interpreting statistical findings. Unfortunately, the nature of the methods considered here renders most conventional hypothesis testing procedures inappropriate; what follows is a brief discussion of known results, with some guidelines for future research.

5.2.1 Classical Tests

Even in a conventional context, hypothesis testing is problematic for many of the methods discussed here. Cluster analysis and MDS, for instance, are primarily used as exploratory procedures, and it can be difficult to construct reasonable and informative null hypotheses for such uses of these techniques. Where results are available, these commonly focus on the multivariate normal case, which is rarely applicable to network data sets.

In the special case for which all structures to be compared are uniquely labeled, and for which all edges are distributed multivariate normal, classical results regarding the distribution of component variances and loadings for the PCA and CCA methods may *in principle* be applied their usual interpretations. The strong potential for within-network autocorrelation, however, suggests that this will rarely be appropriate for interpersonal network data. Similar difficulties exist with relate with respect to the use of such tests on dichotomous data; while estimation is fairly robust in such cases (Van der Geer, 1993), large samples may be required for estimator distributions to approach those assumed by the classical null hypotheses. As a rule, then, we strongly suggest that classical null hypothesis tests be used heuristically with network data, if at all.

[47] Among other weaknesses, the null hypothesis test considers the likelihood of the data under the null hypothesis without comparing this likelihood to any specified alternative. Since "rejecting the null hypothesis" must logically imply an acceptance of the proposition that some alternative hypothesis is *more* plausible, it follows from the above that such rejection can never be logically justified by the test alone. Common practice involving such tests (in which "rejection" of the null is interpreted as support of a particular alternative hypothesis) further runs afoul of the fallacy of false dichotomy, though this practice is justly eschewed by statisticians.

5.2.2 QAP and CUG Nonparametric Tests

Difficulties with the use of classical hypothesis tests in the context of simple network regression prompted Krackhardt (1987) to introduce the use of Hubert's QAP test for certain network-level hypotheses. A permutation test, QAP (in its network guise) tests an observed bivariate network statistic against the distribution of said statistic on the permutation groups of the graphs in question; a "significant" result, then, implies that the particular set of labelings in the original data induce a value of the statistic which is unusual given the underlying structures. While a useful hypothesis in certain circumstances, QAP is usually inapplicable for the types of analyses conducted here: since it *tests against* underlying (unlabeled) structure, it does not provide a means of testing underlying structures themselves[48].

In the context of exploratory inter-structural analysis, a more useful null hypothesis is often that the observed inter-structural relationships are typical of what would be observed were the graphs in question drawn uniformly from some structural population. Tests of this generalized hypothesis – conditional uniform graph (CUG) tests – allow the researcher to compare the results against a more useful measure of "no association." General procedures for performing CUG tests on graph-level indices have been presented by Anderson et al. (1999), and more recently work by Pattison et al. (2000) has suggested additional means for drawing from graph null hypotheses. In the case of the multivariate methods considered here, one basic procedure of fairly general applicability is to draw a population of structures by bootstrapping from the edge distributions of the original graphs, and then to re-run the analyses on the new population. Comparison of the observed results against the distribution of replicate results provides a test of the hypothesis that the observed results are typical for graph sets of equivalent size and edge distribution. More restrictive hypotheses, of course, can be applied by modifying the replication rule; in general, we support the position that the null hypothesis employed should be, inasmuch as possible, a theoretically credible baseline model (Mayhew, 1984) for the relationship to be tested.

6 Conclusion

Although theoretical traditions ranging from neo-institutionalism and small group theory to the theory of mental models pose hypotheses which are best explored via the direct comparison of social structures, relatively few tools have been available to facilitate such analyses. By building on previous research in the areas of metric inference and covariance analysis of graphs, we have here provided a general approach to graph comparison which is applicable under arbitrary assumptions of node exchangeability. One potential advantage of this approach is that it allows the researcher to deploy well-known and understood multivariate methods (e.g,. cluster analysis, MDS, PCA, and CCA) to data sets consisting of

[48]Indeed, similarity among underlying structures will generally lead to *less* significant results under the QAP test; in particular, for a bivariate comparison it is impossible to obtain a one-tailed p-value below $\max_{G \in S} \frac{|\mathbf{A}_G|}{|\mathbf{P}_G|}$, where $\mathbf{A}_G$ and $\mathbf{P}_G$ are the automorphism and permutation groups of G, respectively.

multiple partially labeled social structures. As we have demonstrated, these methods have the potential to serve as powerful exploratory tools for inter-structural analysis, and can serve as important adjuncts to more traditional network analytic techniques.

7 References

Anderson, B.S.; Butts, C.T.; and Carley, K.M. (1999). "The Interaction of Size and Density with Graph Level Indices." *Social Networks,* 21(3), 239-267.

Anderson, C.; Wasserman, S.; and Crouch, B. (1999). "A p* Primer: Logit Models for Social Networks. *Social Networks,* 21,37-66.

Banks, D., and Carley, K. (1994). "Metric Inference for Social Networks." *Journal of Classification*, 11(1), 121-49.

Boorman, S.A., and White, H.C. (1976). "Social Structure from Multiple Networks II. Role Structures." *American Journal of Sociology*, 81, 1384-1446.

Boyd, J.P. (1969). "The Algebra of Group Kinship." *Journal of Mathematical Psychology,* 6, 139-167.

Boyd, J.P. (1990). *Social Semigroups: A Unified Theory of Scaling and Blockmodeling as Applied to Social Networks.* Fairfax, VA: George Mason University Press.

Breiger, R.L., and Pattison, P.E. (1978). "The Joint Role Structure of Two Communities' Elites." *Sociological Methods and Research,* 7, 213-26.

Butts, C.T. (1998). "Cluster Analysis of Unlabeled Structures." ICES Research Report 88-04-98, Carnegie Mellon University.

Butts, C.T. (2000). "Network Inference, Error, and Informant (In)Accuracy: A Bayesian Approach." CASOS Working Paper, Carnegie Mellon University.

Butts, C.T. and Carley, K.M. (1998). "Canonical Labeling to Facilitate Graph Comparison." ICES Research Report 88-06-98, Carnegie Mellon University.

Butts, C.T., and Carley, K.M. (2000). "Spatial Models of Large-Scale Interpersonal Networks." CASOS Working Paper, Carnegie Mellon University.

Carley, K.; Kiesler, S.; and Wholey, D. (1993). "Learning Teamwork: Studies of Training in Software Development," *Proceedings of the 1993 Coordination Theory and Collaboration*

Technology Workshop. Symposium conducted for the National Science Foundation, Washington, D.C.

Carley, K.M., and Palmquist, M. (1992). "Extracting, Representing, and Analyzing Mental Models." *Social Forces*, 70, 601-36.

Dillon, W.R., and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications.* New York: John Wiley and Sons.

Everett, M.G. (1985). "Role Similarity and Complexity in Social Networks." *Social Networks*, 7, 353-59.

Fararo, T.J. (1981). "Biased Networks and Social Structure Theorems. Part I." *Social Networks*, 3, 137-59.

Fararo, T.J. (1983). "Biased Networks and the Strength of Weak Ties." *Social Networks*, 5, 1-11.

Fararo, T.J., and Skvoretz, J. (1984). "Biased Networks and Social Structure Theorems. Part II." *Social Networks*, 6, 223-258.

Farris, J.S. (1969). "On the Cophenetic Correlation Coefficient." *Systematic Zoology*, 18, 279-85.

Feinberg, S.E.; Meyer, M.M.; and Wasserman, S. (1985). "Statistical Analysis of Multiple Sociometric Relations." *Journal of the American Statistical Association,* 80, 51-67.

Freeman, L.C.; Roeder, D.; and Mulholland, R.R. (1980). "Centrality in Social Networks: II. Experimental Results." *Social Networks*, 2, 119-41.

Gelman, A.; Carlin, J.B.; Stern, H.S.; and Rubin, D.B. *Bayesian Data Analysis.* London: Chapman and Hall.

Hamming, H. (1950). "Error Detecting and Error Correcting Codes." *Bell System Technical Journal,* 29, 147-160.

Hannan, M.T., and Freeman, J. (1989). *Organizational Ecology.* Cambridge, MA: Ballinger.

Harary, F. (1953). "On the Notion of Balance of a Signed Graph." *Michigan Mathematical Journal*, 3, 37-41.

Hare, A.P.; Borgatta, E.F.; and Bales, Robert.F. (1965). *Small Groups: Studies in Social*

Interaction. New York: Knopf.

Holland, P.W., and Leinhardt, S. (1971). "Transitivity in Structural Models of Small Groups." *Comparative Group Studies*, 2, 107-24.

Hotelling, H. (1936). "Relations Between Two Sets of Variates." *Biometrika*, 28(3/4), 321-77

Jackson, J.E. (1991). *A User's Guide to Principal Components.* New York: John Wiley and Sons.

Johnson, S.C. (1967). "Hierarchical Clustering Schemes." *Psychometrika,* 32, 241-254.

Joliffe, I.T. (1986). *Principal Component Analysis.* New York: Springer-Verlag.

Kittel, C., and Kroemer, H. (1980). *Thermal Physics,* 2nd Edition. New York: W.H. Freeman and Company.

Krackhardt, D. (1994). "Graph Theoretical Dimensions of Informal Organizations." In *Computational Organization Theory*, edited by K.M. Carley and M.J. Prietula. Hillsdale, NJ: Lawrence Erlbaum Associates

Krackhardt, D. (1987). "QAP Partialling as a Test of Spuriousness." *Social Networks*, 9, 171-86

Krackhardt, D. (1988). "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data." *Social Networks,* 10, 359-381.

Kruskal, J.B. (1964a). "Multidimensional Scaling by Optimizing Goodness-of-fit to a Nonmetric Hypothesis." *Psychometrika*, 29, 1-27.

Kruskal, J.B. (1964b). "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika*, 29, 28-42.

Lorrain, F., and White, H.C. (1971). "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology*, 1, 49-80.

Mardia, K.V.; Kent, J.T.; and Bibby, J.M. (1979). *Multivariate Analysis.* London: Academic Press.

Mayhew, B.H. (1984). "Baseline Models of Sociological Phenomena." *Journal of Mathematical Sociology*, 9, 259-81.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall.

McKelvey, B. (1982). *Organizational Systematics: Taxonomy, Evolution, Classification.* Berkeley, CA: University of California Press.

Newcomb, T. (1961). *The Acquaintance Process.* New York: Holt, Reinhard, and Winston.

Pattison, P., and Wasserman, S. (1999). "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology,* 52, 169-193.

Pattison, P.; Wasserman, S.; Robins, G.; Kanfer, A.M. "Statistical Evaluation of Algebraic Constraints for Social Networks." *Journal of Mathematical Psychology,* 44(4), 536-568.

Rapoport, A. (1957). "A Contribution to the Theory of Random and Biased Nets." *Bulletin of Mathematical Biophysics,* 15, 523-33.

Rapoport, A. (1979). "A Probabilistic Approach to Networks." *Social Networks,* 2, 1-18.

Robert, C.P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation.* New York: Springer.

Robins, G.; Pattison, P.; and Wasserman, S. (1999). "Logit Models and Logistic Regressions for Social Networks, III. Valued Relations." *Psychometrika,* 64, 371-394.

Roethlisberger, F., and Dickson, W. (1939). *Management and the Worker.* Cambridge: Cambridge University Press.

Romesburg, H.C. (1984). *Cluster Analysis for Researchers.* Belmont, CA: Lifetime Learning Publications.

Sampson, S. (1969). *Crisis in a Cloister.* Unpublished doctoral thesis, Cornell University.

Sanil, A.; Banks, D.; Carley, K. (1995). "Models for Evolving Fixed Node Networks: Model Fitting and Model Testing." *Social Networks,* 17(1), 65-81.

Skvoretz, J. (1985). "Random and Biased Networks: Simulations and Approximations." *Social Networks,* 7, 256-61.

Skvoretz, J. (1990). "Biased Net Theory: Approximations, Simulations, and Observations."

Social Networks, 12, 217-38.

Van der Geer, J.P. (1993). *Multivariate Analysis of Categorical Data: Theory.* Newbury Park: Sage.

Ward, J. (1963). "Hierarchical Grouping to Maximize an Objective Function." *Journal of the American Statistical Association,* 58, 236-44.

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Wasserman, S.; Faust, K.; and Galaskiewicz, J. (1990). "Correspondence and Canonical Analysis of Relational Data." *Journal of Mathematical Sociology,* 15, 11-64.

Wasserman, S., and Pattison, P. (1996). "Logit Models and Logistic Regressions for Social Networks: I. An introduction to Markov Graphs and p*." *Psychometrika,* 60, 401-426.

White, D.R., and Reitz, K.P. (1983). "Graph and Semigroup Homomorphisms on Networks of Relations." *Social Networks,* 5, 193-235.

White, H.C. (1963). *An Anatomy of Kinship.* Englewood Cliffs, NJ: Prentice-Hall.