7

Using Correspondence Analysis for Joint Displays of Affiliation Networks

Katherine Faust

University of California, Irvine

This chapter describes and illustrates methods for studying affiliation networks, with special attention to methods for spatial representations that jointly display the actors and events in the network. Although affiliation networks have been the focus of methodological research for decades (Levine 1972; Breiger 1974; Seidman 1981; McPherson 1982; Wilson 1982), more recent analyses of affiliation networks have raised a number of issues concerning appropriate methods for their study. At the same time, research has pointed to the empirical and theoretical generality of this perspective (Freeman and White 1993; Wasserman and Faust 1994; Borgatti and Everett 1997; Faust 1997; Skvoretz and Faust 1999; Breiger 2000; Mische and Pattison 2000; Roberts 2000; Brazill and Groffman 2002; Faust et al. 2002; Pattison and Breiger 2002).

7.1 Background

Representing the two modes in the affiliation network in a "joint space" in which both actors and events are depicted simultaneously is of particular interest in both earlier and more recent work on affiliation networks. Such graphic displays commonly use scaling (e.g., correspondence analysis) or algebraic approaches (e.g., lattices). An important, but often neglected, aspect of some applications is clear specification of the formal relationships embodied in the configuration and explicit description of how the result corresponds to the original data. These omissions produce rather casual depictions and consequent ambiguity in interpretation. They also contribute to misunderstanding and fuel debate about the usefulness of the approach. The following passages are typical of such descriptions for affiliation networks or similar two-mode data arrays.

In describing correspondence analysis for the joint display of actors and events in an affiliation network of Chinese political actors' involvement, Schweizer (1991) interpreted the result in terms of a "preference" model: "In this application of the model to an actor-by-event matrix, actors are placed as points into their region of maximal involvement ('preference') for certain events" (p. 33). However, he neglected to reveal what scores were used for the display.

Similarly, in their reanalysis of the classic Davis, Gardner, and Gardner (1941) observations of southern women's attendance at social events, Borgatti and Everett

(1997) described the joint display as follows:

> Applied to the Davis, Gardner and Gardner data, a correspondence analysis results in a map in which (a) points representing the women are placed close together if the women attended mostly the same events, (b) points representing the social events are placed near each other if they were attended by mostly the same women, and (c) women-points are placed near event-points if those women attended those events. (p. 246)

However, as with the previous example, the authors failed to reveal which scores were used for the display, so we are left with no precise understanding of what "near" means in the plots.

Admittedly, these passages are intended as simplified descriptions to aid substantive interpretation of the results, but their informality and the absence of information about exactly which sets of scores are used in the figures prevent precise interpretation of the results. In addition, absence of formal specification contributes to debate about potential problems with correspondence analysis for studying affiliation networks. Criticisms have included application of the approach to dichotomous data, possible inadequacy of representations in two dimensions, and proper interpretations of distances in the displays (see the exchange between Borgatti and Everett 1997, and Roberts 2000).

This chapter takes the modest step of laying out some aspects of the formal basis for joint representations of actors and events in an affiliation network using correspondence analysis. The goal is to provide the precise formal specification of the model and of the relationship between the model and the input data in such a way that users can select among some possible alternatives and interpret the results appropriately. This chapter describes and illustrates some of the methodological issues using both a small hypothetical affiliation network and an affiliation network of Western Hemisphere countries and their memberships in regional trade and treaty organizations.

7.2 Affiliation Networks

Many social situations bring together actors in sets of two, three, or more in collectivities of arbitrary size. Corporate boards of directors, scientists attending sessions at a professional meeting, members of voluntary organizations in a community, activists gathered in protest demonstrations, fans watching sporting events, countries forging alliances through membership in trade and treaty organizations, and members of committees in a university are all examples of this sort of social situation. These situations are varied in nature. Some are quite informal social gatherings, whereas others are well-defined assemblages. In some situations people can be expected to interact quite intensely with one another, whereas in other situations direct interaction among all members is unlikely. Some situations are fleeting one-time events, whereas others are recurrent. Nevertheless, all the examples mentioned previously share a number of common features. In each, joint social participation brings together sets of actors rather than simply linking pairs or dyads. Thus, joint participation constitutes a social relation among *collections* of actors. Moreover, when actors participate in multiple interaction occasions, the social occasions themselves are linked to one another through their common participants. Finally, all the situations involve two different kinds of social entities: the individuals (referred to as *actors*) and the social occasions (referred to as *events*).

Affiliation networks have been used to study a wide range of social situations, and a partial list highlights their generality. A classic example is Davis et al.'s (1941) study of southern women's participation in informal social gatherings (see also Homans 1950; Breiger 1974; Doreian 1979; Freeman and White 1993; Freeman 2002). Many studies of corporate interlocks have used affiliation networks (Levine 1972; Mariolis 1975; Mariolis and Jones 1982; Mizruchi 1982), as have studies of corporate CEOs and their memberships in civic, cultural, and corporate boards (Galaskiewicz 1985). Participation in community ritual celebrations has been studied by Foster and Seidman (1984) and Schweizer, Klemm, and Schweizer (1993). Affiliation networks have also been used to study social movements (Rosenthal et al. 1985; Mische and Pattison 2000; Osa 2003) and other political situations, including roll call votes (Stokman 1977), opinions by U.S. Supreme Court justices (Breiger 2000; Brazill and Groffman 2002), winners and losers in Chinese political struggles (Schweizer 1991), and the participation of Soviet political elites in official and social occasions (Faust et al. 2002). Academic associations have been the focus of a number of affiliation network studies, including sociologists' memberships in disciplinary specialty sections (Cappell and Guterbock 1992; Ennis 1992).

The situations mentioned previously can be viewed as instances of affiliation networks (also called *membership networks*, *dual networks*, or *hypernetworks*). Affiliation networks are a general class of networks with several important properties. In particular, three characteristics distinguish affiliation networks from the more standard social network in which relations are measured on pairs of actors from a single set. Linkages in an affiliation network occur between two different kinds of social entities, referred to as "actors" and "events." As with all social networks, the actors may be any meaningful social unit, including individual or collective entities. The events in an affiliation network are collections of actors. The events may either be well-defined collectivities with official membership lists, or they may be less formal gatherings. Each of these kinds of entities constitutes a "mode" of the network. Thus, affiliation networks are *two-mode* networks. A second important characteristic is that the affiliation relation links collections of entities – actors belong to multiple events, and events may include multiple actors. Thus, affiliation networks are *nondyadic*. These two characteristics permit a third important property – the *duality* of perspectives in the relation between actors and events (Breiger 1974). Viewed from the perspective of actors, participation in events links actors to one another. Viewed from the dual perspective of events, the actors multiple memberships link events together. Putting these together gives a *joint* perspective of the simultaneous linking of actors through events and events through actors.

The distinction between social ties based on *membership relations* (seen in affiliation networks) and *social relations* (the typical one-mode network linking pairs of actors) is nicely discussed in Breiger's foundational work on membership networks (Breiger 1974). Both of these kinds of social ties give rise to networks, but networks with rather different properties. Social relations are dyadic – they link pairs of actors in a single mode directly to one another. However, membership relations link individuals to collectivities, and then indirectly to each other through these shared memberships.

These considerations provide the foundation for affiliation network methodology and highlight why methodology for affiliation networks deserves special attention, beyond that for standard one-mode networks.

Of particular importance in analyzing affiliation networks is producing an interpretable simultaneous or joint model of actors, events, and the relationships between them. This chapter describes how to do this appropriately using correspondence analysis.

7.3 Example

This chapter uses as an example the memberships of twenty-two Western Hemisphere countries in fifteen regional international organizations. The actors in this example are sovereign nations in North, Central, and South America. The events are regional international trade and treaty organizations. These organizations primarily promote economic interests or political, social, or cultural cooperation among member nations. The list of organizations and their members was compiled from *Keesing's Record of World Events*, and consists of the regional organizations listed for the Americas, excluding the Caribbean (East 1996). Membership in these organizations includes all countries that are full members, but excludes observers, nations outside the hemisphere, and territories. Organizational memberships were verified using information from the *CIA Yearbook* (CIA 2000) and publications of the individual organizations, when available. Brief descriptions of the organizations are presented in the Appendix. This substantive example illustrates the political and economic alliances among countries in one part of the world, and reveals the more local basis for some of these alliances.

7.4 Notation

An affiliation network is presented in a two-mode sociomatrix. The rows of the matrix index actors and the columns index events. The set of actors is denoted by N, with g being the number of actors, and the set of events is denoted by M, with h being the number of events. We use the notation $\mathbf{A}$ for the matrix, with entries a_{ij}, where $a_{ij} = 1$ if actor i is in event j and 0 otherwise. The sociomatrix for the twenty-two countries and fifteeen organizations is presented in Table 7.4.1. In this table, both countries and organizations are listed in alphabetical order.

In this form, it is difficult to see any patterns that might be present in the network. Representational and graphic methods can help reveal and communicate patterns in the data.

7.5 Bipartite Graph

Visualization is an integral part of social network analysis (McGrath, Blythe, and Krackhardt 1997; Freeman 2000). Well-drawn graphs or diagrams bring attention to

Table 7.4.1. *Sociomatrix of Western Hemisphere Countries and Memberships in Regional Trade and Treaty Organizations*

	ACS	ALADI	Amazon Pact	Andean Pact	CARICOM	GENPLACEA	Group of Rio	G-3	IDB	MERCOSUR	NAFTA	OAS	Parlacén	San José Group	SELA
Argentina	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1
Belize	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1
Bolivia	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1
Brazil	0	1	1	0	0	1	1	0	1	1	0	1	0	0	1
Canada	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0
Chile	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1
Colombia	1	1	1	1	0	1	1	1	1	0	0	1	0	0	1
Costa Rica	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1
Ecuador	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1
El Salvador	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
Guatemala	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
Guyana	1	0	1	0	1	1	0	0	1	0	0	1	0	0	1
Honduras	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
Mexico	1	1	0	0	0	1	1	1	1	0	1	1	0	0	1
Nicaragua	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1
Panama	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1
Paraguay	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1
Peru	0	1	1	1	0	1	1	0	1	0	0	1	0	0	1
Suriname	1	0	1	0	0	0	0	0	1	0	0	1	0	0	1
United States	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0
Uruguay	0	1	0	0	0	1	1	0	1	1	0	1	0	0	1
Venezuela	1	1	1	1	0	1	1	1	1	0	0	1	0	0	1

important features of the network, such as the presence of subgroups, the relative importance or centrality of actors (McGrath et al. 1997), and often convey descriptive information in a form that is more easily appreciated than are numeric summaries or matrices.

Because the affiliation relation always links actors to events and vice versa, all ties in an affiliation network are between entities from different sets – the two modes of the network. This means that an affiliation network can be represented as a bipartite graph. In a bipartite graph, the nodes can be partitioned into two mutually exclusive sets and all edges link nodes from different sets.

Figure 7.5.1 presents a graph of the network of countries and their memberships in regional organizations. In this figure, the points are located to highlight the fact that the graph is bipartite. Countries are roughly arrayed from south to north in terms of their geographic position, calling attention to the regional basis for many of the organizations. Consistent with the fact that a number of organizations have regional economic or political interests as their express intent, it can be seen in the graph that

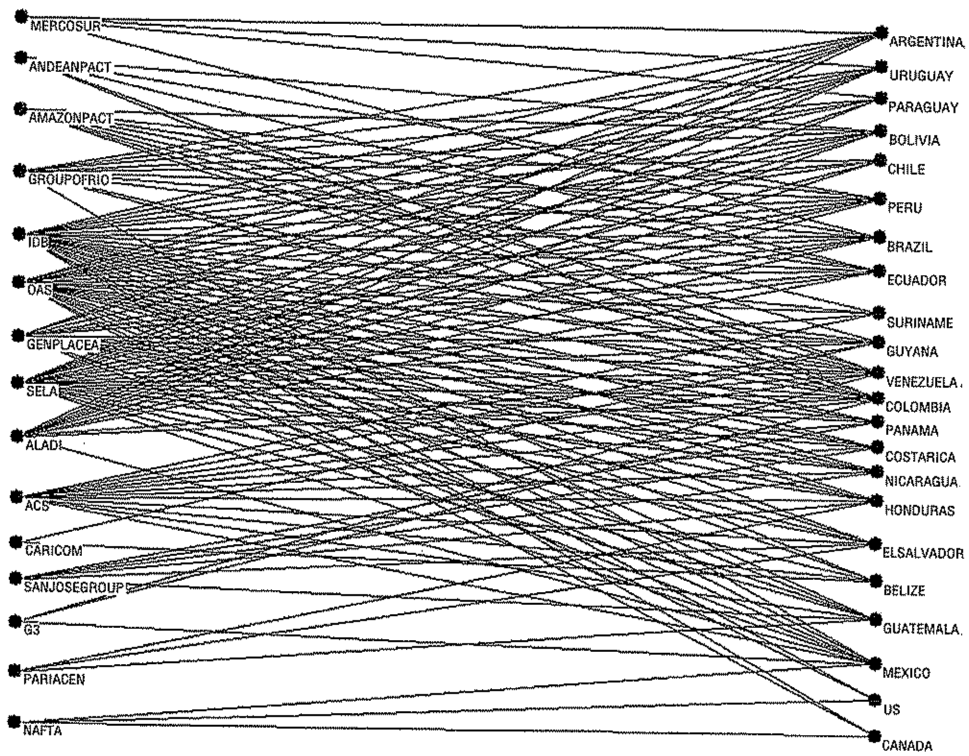

Figure 7.5.1. Bipartite graph of countries and organizations.

some organizations only have as members countries from within one region. Notably, MERCOSUR, Andean Pact, and Amazon Pact are all composed entirely of South American countries; San José Group and Parlacén include only Central American countries; and NAFTA consists entirely of North American countries. In contrast, other organizations clearly span the entire hemisphere (for example, IDB, OAS, and SELA); in fact, all countries in the set belong to IDB and OAS. (See Appendix for descriptions of these organizations.) We can also see differences among the countries in their level of participation. Some countries (for example, the United States and Canada) belong to relatively few of these organizations, whereas others belong to more than one-half of them (Ecuador, Peru, Bolivia, Colombia, Venezuela, and Mexico, for example).

In Figure 7.5.1, locations of the points are arbitrary in the sense that the formal information represented in the figure consists only of the nodes and the edges between nodes. Locations of points and the proximity of pairs or sets of points are not related in any specifiable way to the input data, nor do distances in the figure relate in an explicit way to associations between the countries and the organizations.

Alternatively, one can construct a graphic display in which location of points and distances between them convey precise information about properties of the network. The next section discusses how to accomplish this using correspondence analysis for joint graphic displays.

7.6 Joint Representation of Actors and Events

The goal is to represent the affiliation network graphically so both actors and events are presented in the same display. The problem is to find locations for points representing both actors and events so the resulting configuration provides a low-dimensional approximation to the input data and the locations of the points in the configuration correspond in an explicit way to specified aspects of the data. In the configuration, the proximity of points can show relationships among actors, among events, or between actors and events. More generally, a representation for two-mode data that include entities from both modes is called a *joint space* because it jointly displays entities from two different sets (Jacoby 1991; Coombs 1964). An important feature of the models described here is that they explicitly specify the relationship between the input data and the locations of points in the resulting configuration.

Affiliation networks are often analyzed using correspondence analysis and the resulting coordinates used in graphic displays. There is great appeal in this approach because it does provide a joint representation of the two modes in the network. Nevertheless, this approach has been the topic of considerable recent discussion and debate (Borgatti and Everett 1997; Breiger 2000; Roberts 2000; Faust et al.).

This section presents the formal basis for correspondence analysis, with special attention to how this model can be used for a joint representation of an affiliation network when interpretable distances between points from different modes are desired. Three alternatives are illustrated, first using a small hypothetical example and then using data on Western Hemisphere countries' memberships in trade and treaty organizations.

Correspondence analysis (Weller and Romney 1990; Greenacre and Blasius 1994; Blasius and Greenacre 1998) is a scaling approach usually used for studying relationships between variables in two-way arrays. It is one of a number of closely related approaches, including dual scaling (Nishisato 1994), homogeneity analysis (Gifi 1990), and optimal scaling. There are numerous articles and monographs describing the general approach. A few useful references include Weller and Romney (1990), Greenacre (1984), Greenacre and Blasius (1994), Blasius and Greenacre (1998), Nishisato (1994), Claussen (1998), and Gifi (1990).

Correspondence analysis is often used to study relationships between categorical variables in contingency tables or incidence matrices, but it has also been used to model social networks, particularly affiliation networks (Noma and Smith 1985; Wasserman and Faust 1989, 1994; Wasserman, Faust, and Galaskiewicz 1990; Schweizer 1991, 1993; Faust and Wasserman 1993; Nakao and Romney 1993; Kumbasar, Romney, and Batchelder 1994; Breiger 2000; Roberts 2000; Faust et al. 2002).

To appreciate and interpret correspondence analysis as a way of providing a joint display of an affiliation network, it is useful to clearly describe several aspects of the approach. In particular, the following points are important:

1. Decomposition of a matrix into its basic structure using singular value decomposition
2. Geometric features of correspondence analysis, especially as they pertain to the relationship between the resulting configuration and the input data

3. Alternative "scalings" that may be used for a joint representation and implications of the alternatives
4. Dimensionality of the result

Each of these issues is described and illustrated in turn.

7.7 Matrix Decomposition

Correspondence analysis is accomplished through the decomposition of a matrix into its basic structure (Digby and Kempton 1987; Weller and Romney 1990; Clausen 1998). In general, singular value decomposition is defined as the decomposition of a matrix, **A**, of size g by h, as:

$$\mathbf{A} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{Y}', \tag{7.1}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of singular values, $\{\lambda_k\}$, **X** is the matrix of left singular vectors, and **Y** is the matrix of right singular vectors. If **A** has g rows and h columns (with h less than or equal to g), then **X** is of size $g \times h$ and **Y** is size $h \times h$. Both **X** and **Y** are orthonormal. In other words, rows of **X** and similarly columns of **Y** are orthogonal and of unit length. Formally,

$$\mathbf{X}\mathbf{X}' = \mathbf{I} \tag{7.2}$$

$$\mathbf{Y}'\mathbf{Y} = \mathbf{I}, \tag{7.3}$$

where **I** is an identity matrix.

The same relationships can be expressed in terms of the elements of **X** and **Y**:

$$\sum_{k}^{W} x_{ik}^2 = \sum_{k}^{W} y_{jk}^2 = 1 \tag{7.4}$$

$$\sum_{k}^{W} x_{ik}x_{i'k} = \sum_{k}^{W} y_{jk}y_{j'k} = 0. \tag{7.5}$$

The number of singular values and singular vectors, and hence the dimensionality of the matrix, is equal to rank of the matrix **A**. Generally, this is the number of nonnegative singular values, which is no greater than the minimum of the number of rows (g) or columns (h). We denote the rank of **A** as W. When the full set of W dimensions are used, then $\boldsymbol{\Lambda}$, **X**, and **Y** perfectly reproduce the entries in **A**. When fewer than the full set of W dimensions are used, the result approximates the entries in **A**.

Correspondence analysis is a singular value decomposition not of **A**, but of a "normalized" version of **A**. Entries in the original matrix are divided by the square root of the product of the row and column marginal totals prior to singular value decomposition. Let **A** be a rectangular matrix of positive entries with g rows and h columns (where $g \geq h$). Two diagonal matrices $\mathbf{R}^{-\frac{1}{2}}$ and $\mathbf{C}^{-\frac{1}{2}}$ have entries equal to reciprocals of the

Table 7.7.2. *Hypothetical Affiliation Network*

	Event 1	Event 2	Event 3
Affiliation matrix, **A**			
Actor 1	1	0	1
Actor 2	0	0	1
Actor 3	0	1	1
Actor 4	1	1	0
"Normalized" matrix, $\mathbf{R}^{-1/2}\,\mathbf{A}\mathbf{C}^{-1/2}$			
Actor 1	0.500	0.000	0.408
Actor 2	0.000	0.000	0.577
Actor 3	0.000	0.500	0.408
Actor 4	0.500	0.500	0.000

row and column totals of **A**, respectively:

$$\mathbf{C}^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{a_{+j}}}\right) \tag{7.6}$$

$$\mathbf{R}^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{a_{i+}}}\right). \tag{7.7}$$

Correspondence analysis consists of a singular value decomposition of the matrix $\mathbf{R}^{-\frac{1}{2}}\mathbf{A}\mathbf{C}^{-\frac{1}{2}}$ defined as:

$$\mathbf{R}^{-\frac{1}{2}}\mathbf{A}\mathbf{C}^{-\frac{1}{2}} = \mathbf{X}\Lambda\mathbf{Y}', \tag{7.8}$$

where Λ is a diagonal matrix of singular values, $\{\lambda_k\}$, and **X** and **Y** are the left and right singular vectors (Digby and Kempton 1987; Weller and Romney 1990; Clausen 1998).

To illustrate, consider a hypothetical example of an affiliation network of four actors and three events. The matrix, **A**, is presented in Table 7.7.2, panel A. The "normalized" version of this matrix, $\mathbf{R}^{-\frac{1}{2}}\mathbf{A}\mathbf{C}^{-\frac{1}{2}}$, is in panel B of Table 7.7.2.

The singular values, Λ, and the left and right singular vectors, **X** and **Y**, are presented in Table 7.7.3.

7.8 Relationship Between Scores and Input Data

One important feature of the decomposition is that the resulting scores (the right and left singular vectors) are explicitly related to the input data. As seen in (7.8), the decomposition $\mathbf{X}\Lambda\mathbf{Y}'$ reproduces the matrix, $\mathbf{R}^{-\frac{1}{2}}\mathbf{A}\mathbf{C}^{-\frac{1}{2}}$. Rearranging terms shows that it also reproduces the original matrix, **A**:

$$\mathbf{A} = \mathbf{R}^{\frac{1}{2}}\mathbf{X}\Lambda\mathbf{Y}'\mathbf{C}^{\frac{1}{2}}. \tag{7.9}$$

Table 7.7.3. *Singular Value Decomposition of "Normalized" Affiliation Matrix*

	Dimension		
Left singular vectors, **X**	1	2	3
Actor 1	−0.535	0.120	0.707
Actor 2	−0.378	0.676	0.000
Actor 3	−0.535	0.120	−0.707
Actor 4	−0.535	−0.717	0.000
Right singular vectors, **Y**			
Event 1	−0.535	−0.463	0.707
Event 2	−0.535	−0.463	−0.707
Event 3	−0.655	0.756	0.000
Singular values, **Λ**	1.000	0.645	0.500

Or, in terms of the elements of **A**:

$$a_{ij} = \sqrt{a_{i+}a_{+j}} \sum_{k=1}^{W} x_{ik}\lambda_k y_{jk}. \tag{7.10}$$

When the number of dimensions and, hence, sets of row scores, column scores, and singular values are equal to the rank of the matrix, W, then the data are perfectly reproduced. When fewer than W dimensions are used, the data are approximated in the lower dimensional solution. Using only the first singular value and first singular vectors reproduces the expected frequencies under the model of statistical independence in the matrix **A**. Formally,

$$\frac{a_{i+}a_{+j}}{a_{++}} = \sqrt{a_{i+}a_{+j}}x_{i1}\lambda_1 y_{j1}. \tag{7.11}$$

Consequently the first "trivial" dimension is usually ignored because it is simply a function of the marginal totals and does not represent the pattern of relationship between rows and columns.

7.9 Scores for Correspondence Analysis

Correspondence analysis uses one of a number of possible rescalings of the right and left singular vectors, **X** and **Y**. The first alternative, referred to as *optimal scores*, *standard scores*, or *standard coordinates*, multiplies values in each left singular vector **X** by the square root of the reciprocal of its row proportion and multiplies each right singular vector **Y** by the square root of the reciprocal of its column proportion. These new scores, which we denote $\tilde{u}_{ik}$ and $\tilde{v}_{jk}$, are:

$$\tilde{u}_{ik} = x_{ik}\sqrt{\frac{a_{++}}{a_{i+}}} \tag{7.12}$$

for row scores, and

$$\tilde{v}_{jk} = y_{jk}\sqrt{\frac{a_{++}}{a_{+j}}} \tag{7.13}$$

for column scores. On each dimension, these scores have weighted means equal to 0.0 and weighted variances equal to 1.0:

$$\sum_{i=1}^{g} \tilde{u}_{ik}\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} \tilde{v}_{jk}\frac{a_{+j}}{a_{++}} = 0 \tag{7.14}$$

$$\sum_{i=1}^{g} \tilde{u}_{ik}^{2}\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} \tilde{v}_{jk}^{2}\frac{a_{+j}}{a_{++}} = 1. \tag{7.15}$$

Because the variance is equal to 1.0 on each dimension, standard scores do not express the relative importance of each dimension in accounting for the data (Weller and Romney 1990). An alternative, referred to as *principal scores* or *principal coordinates*, and which we denote u_{ik} and v_{jk}, are given by:

$$u_{ik} = \lambda_k x_{ik}\sqrt{\frac{a_{++}}{a_{i+}}} \tag{7.16}$$

for row scores, and

$$v_{jk} = \lambda_k y_{jk}\sqrt{\frac{a_{++}}{a_{+j}}} \tag{7.17}$$

for column scores. On each dimension, these scores have weighted means equal to 0.0 and weighted variances equal to the singular value squared:

$$\sum_{i=1}^{g} u_{ik}\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} v_{jk}\frac{a_{+j}}{a_{++}} = 0 \tag{7.18}$$

$$\sum_{i=1}^{g} u_{ik}^{2}\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} v_{jk}^{2}\frac{a_{+j}}{a_{++}} = \lambda_k^2. \tag{7.19}$$

On each dimension, the principal coordinates, u_{ik} and v_{jk}, express the importance of the dimension in terms of the singular value squared, λ_k^2. As follows, we show how the singular values are related to the inertia or total variation in the data.

In summary, correspondence analysis results in three sets of information: a set of g scores for rows of the matrix, $\mathbf{U} = \{u_{ik}\}$, for $i = 1, 2, \ldots g$ and $k = 1, 2, \ldots W$; a set of h scores for columns of the matrix, $\mathbf{V} = \{v_{jk}\}$, for $j = 1, 2, \ldots h$ and $k = 1, 2, \ldots W$; and the singular values $\Lambda = \{\lambda_k\}$ for $k = 1, 2, \ldots W$, expressing the importance of each dimension.

7.10 Asymmetric Duality

An important feature of correspondence analysis is the inherent duality in the relationship between scores for rows $\{u_{ik}\}$ and scores for columns $\{v_{jk}\}$. This relationship is critical for the "asymmetric" interpretation of some correspondence analysis displays that use $\{u_{ik}\}$ and $\{v_{jk}\}$, or some rescaling of them, as coordinates for joint display of row and column entities. The duality can be seen in that, on each dimension, the score for an object in one set is the weighted average of the scores for all objects in the other set, where the weightings are the marginal row or column proportions. This duality is expressed in the following set of equations:

$$\lambda_k u_{ik} = \sum_{j=1}^{h} \frac{a_{ij}}{a_{i+}} v_{jk} \tag{7.20}$$

and

$$\lambda_k v_{jk} = \sum_{i=1}^{g} \frac{a_{ij}}{a_{+j}} u_{ik}. \tag{7.21}$$

For an affiliation network, the score for an actor is the weighted average of the scores for the events with which it is affiliated and the score for an event is the weighted average of the scores of its constituent actors.

7.11 Chi-Square Distance Interpretations

Scores for row and column objects may be used as coordinates in graphic displays, but appropriate interpretation of the display depends on the formal relationship between distances between points representing row and/or column entities and chi-square distances calculated on the input data. This relationship is at the heart of the interpretative debate about which of a number of alternative correspondence analysis scores should be used for graphic displays. The chi-square distances also express the overall variability (inertia) in the data and are used to measure how well the configuration fits the input data.

Distances in correspondence analysis displays represent the chi-square distances between row or column profiles. The profile for a row is defined as the entry in each cell divided by its corresponding row total, $\{\frac{a_{ij}}{a_{i+}}\}$, for $j = 1, 2, \ldots h$. A column profile similarly is defined as the entry in each cell divided by the column total, $\{\frac{a_{ij}}{a_{+j}}\}$, for $i = 1, 2, \ldots g$. The row and column profiles for the hypothetical example are presented in Table 7.11.4. Profiles of different rows (or different columns) can be compared with each other to measure the distances between rows (or between columns). The chi-square distance between profiles for rows i and i', denoted, $d(i, i')$ is given by:

$$d(i, i') = \sqrt{\sum_{j=1}^{h} \frac{\left(\frac{a_{ij}}{a_{i+}} - \frac{a_{i'j}}{a_{i'+}}\right)^2}{\frac{a_{+j}}{a_{++}}}}. \tag{7.22}$$

Table 7.11.4. *Row and Column Profiles for Hypothetical Affiliation Network*

Row profiles

	Event 1	Event 2	Event 3	Sum
Actor 1	0.500	0.000	0.500	1.000
Actor 2	0.000	0.000	1.000	1.000
Actor 3	0.000	0.500	0.500	1.000
Actor 4	0.500	0.500	0.000	1.000
Average row profile	0.286	0.286	0.429	

Column profiles

	Event 1	Event 2	Event 3	Average Column Profile
Actor 1	0.500	0.000	0.333	0.286
Actor 2	0.000	0.000	0.333	0.143
Actor 3	0.000	0.500	0.333	0.286
Actor 4	0.500	0.500	0.000	0.286
Sum	1.000	1.000	1.000	

A parallel definition for the chi-square distance between two column profiles, j and j', is:

$$d(j, j') = \sqrt{\sum_{i=1}^{g} \frac{\left(\frac{a_{ij}}{a_{+j}} - \frac{a_{ij'}}{a_{+j'}}\right)^2}{\frac{a_{i+}}{a_{++}}}}. \tag{7.23}$$

These are the interpoint distances depicted in graphic displays of correspondence analysis. The formal relationship between the chi-square distances and the row and column scores is presented in detail as follows (7.29 and 7.30).

Chi-square distances are also used to compare row or column profiles to the "average," or marginal, row or column profile to assess the total variation in the data. The average row profile is defined as the set of marginal column proportions, $i^+ = \{\frac{a_{+j}}{a_{++}}\}$ for $j = 1, 2, \ldots h$, and similarly the average column profile is defined as the set of marginal row proportions, $j^+ = \{\frac{a_{i+}}{a_{++}}\} i = 1, 2, \ldots g,$. The chi-square distance between an individual row (or column) and the average row (or column) profile is defined as:

$$d(i, i^+) = \sqrt{\sum_{j=1}^{h} \frac{\left(\frac{a_{ij}}{a_{i+}} - \frac{a_{+j}}{a_{++}}\right)^2}{\frac{a_{+j}}{a_{++}}}} \tag{7.24}$$

and

$$d(j, j^+) = \sqrt{\sum_{i=1}^{g} \frac{\left(\frac{a_{ij}}{a_{+j}} - \frac{a_{i+}}{a_{++}}\right)^2}{\frac{a_{i+}}{a_{++}}}}. \tag{7.25}$$

Table 7.11.5. *Chi-Square Distances Between Row Profiles and Between Column Profiles, and Distances Between Profiles and Mean Profile on Diagonal*

Chi-square distances between row profiles

	Actor 1	Actor 2	Actor 3	Actor 4
Actor 1	0.677			
Actor 2	1.208	1.155		
Actor 3	1.323	1.208	0.677	
Actor 4	1.208	2.021	1.208	0.866

Chi-square distances between column profiles

	Event 1	Event 2	Event 3
Event 1	0.866		
Event 2	1.323	0.866	
Event 3	1.462	1.462	0.745

Table 7.11.5 presents the chi-square distances between row profiles and between column profiles for the hypothetical example. These are the distances that are represented in correspondence analysis displays. On the diagonals of the arrays, Table 7.11.5 presents the distances between rows (and columns) and the average row (and column) profile.

7.12 Inertia

In correspondence analysis, *inertia* quantifies the total amount of variation in the data (Greenacre 1984; Greenacre and Blasius 1994; Clausen 1998). Inertia is calculated as the weighted sum of the squared chi-square distances between the row profiles and the average row profile, where the weights are marginal row proportions, or similarly, the weighted sum of the squared chi-square distances between the column profiles and the average column profile, where the weights are the marginal column proportions (Greenacre 1984; Greenacre and Hastie 1987). The total inertia is given by the following equations:

$$\sum_{i=1}^{g} \frac{a_{i+}}{a_{++}} d(i, i^+)^2 \tag{7.26}$$

or

$$\sum_{j=1}^{h} \frac{a_{+j}}{a_{++}} d(j, j^+)^2. \tag{7.27}$$

These expressions show that total inertia can be decomposed into the contributions from each of the entities (rows or columns) in the data. The total inertia is also equal to the sum of the squared singular values:

$$\sum_{k=1}^{W} \lambda_k^2. \tag{7.28}$$

This allows another decomposition of the total inertia into the amount of variation that is accounted for by each dimension in the model by considering each λ_k^2, for $k = 1, 2, \ldots W$.

7.13 Within-Set Distance Comparisons

Distances between points in graphic display using correspondence analysis scores are interpretable with respect to specific patterns in the input data, but proper interpretation requires both correct selection of scores and recognition that the distances are chi-square distances. First, consider the distance between two rows, i and i'. This distance is the chi-square distance between the profiles for rows i and i' (7.22). The reproduced or fitted distance, $\hat{d}(i, i')$, is calculated from the correspondence analysis row scores, $\{u_{ik}\}$ and $\{u_{i'k}\}$ as:

$$\hat{d}(i, i') = \sqrt{\sum_{k=1}^{W} (u_{ik} - u_{i'k})^2}. \tag{7.29}$$

Similarly, the distance between two columns, j and j' (7.23), is calculated from the column scores $\{v_{jk}\}$ and $\{v_{j'k}\}$ as:

$$\hat{d}(j, j') = \sqrt{\sum_{k=1}^{W} (v_{jk} - v_{j'k})^2}. \tag{7.30}$$

When the full set of W dimensions is used, the chi-square distances between row profiles and between column profiles are perfectly reproduced, so $\hat{d}(i, i') = d(i, i')$ and $\hat{d}(j, j') = d(j, j')$ for all pairs of rows and all pairs of columns. When fewer than W dimensions are used, the fitted distance approximates the chi-square distance between row or column profiles in the original data. These distance interpretations hinge on correct selection of scores. The relationship between chi-square distances between row (or column) profiles and fitted distances (7.29 and 7.30) holds when scores are scaled as principal coordinates (7.16 and 7.17).

7.14 Between-Set Comparisons

Equations (7.29) and (7.30) express distances between entities within the same set. The problem of interpreting distances arises when comparing locations of points from different sets. The general problem has been widely discussed in the correspondence analysis literature (Carroll, Green, and Schaffer 1986, 1987, 1989; Greenacre 1989). In fact, some authors argue that interset distance comparisons are not even justified

(Nishisato 1994: p. 113). Unfortunately, these are the distances between actors and events in an affiliation network, and are exactly the comparisons of most interest if we take seriously the duality argument and the goal of constructing an interpretable joint space. Under what conditions are these interset relationships interpretable?

Returning to (7.16) and (7.17), suppose we want to express the score for an actor, u_{ik}, as function of the scores for the events to which it belongs, v_{jk}, for $j = 1, 2, \dots h$. This is accomplished by using actor scores expressed in principal coordinates paired with events expressed as standard coordinates. The relationship between row scores, u_{ik}, and column scores, $\tilde{v}_{jk}$, is:

$$u_{ik} = \sum_{j=1}^{h} \frac{a_{ij}}{a_{i+}} \tilde{v}_{jk}. \qquad (7.31)$$

With the row scores in principal coordinates, u_{ik}, and the column scores in standard coordinates, $\tilde{v}_{jk}$, the score for a row point is the weighted average of the scores for the column points. However, the reverse is not true. If one wants to express the score for a column point, v_{jk}, as the average of the scores for the rows, u_{jk}, for $i = 1, 2, \dots k$, then columns in principal coordinates, v_{jk}, are paired with standard coordinates for rows. The relationship between column scores, v_{jk}, and row scores, $\tilde{u}_{ik}$, is:

$$v_{jk} = \sum_{i=1}^{g} \frac{a_{ij}}{a_{+j}} \tilde{u}_{ik}. \qquad (7.32)$$

With column scores in principal coordinates, v_{jk}, and row scores in standard coordinates, $\tilde{u}_{jk}$, the score for a column point is interpreted as the weighted average of the scores for the row points.

In summary, for an asymmetric joint display there are two possible sets of scores that might be used: row scores, u_{ik}, paired with column scores, $\tilde{v}_{jk}$, or column scores, v_{jk}, paired with row scores, $\tilde{u}_{jk}$. Either of these gives an "asymmetric" depiction of the relationship between entities from the two modes. The choice of which to use depends on which relationships in the data one wants to highlight. Unlike distances within sets, interset comparison is only legitimate when the location of a single point from one set is compared with the locations of *all* points in the other set, but not when two individual points from different sets are compared.

7.15 Symmetric Representation

An alternative approach, which permits a "symmetric" view of the distances between row and column points, and in which interset distance comparisons are interpretable, has been proposed by Carroll et al. (1986). Carroll et al. viewed the problem as one of *multiple correspondence analysis*. Using this approach, observations in a two-way contingency table are reexpressed as a cases-by-variables data array called a *pseudocontingency table*. In the pseudocontingency table, a two-way contingency table with r rows, c columns, and N observations is transformed into an array in which each row in the new array is an observation in the original table, and the row and column entities in

Table 7.15.6. *Pseudocontingency Table for Hypothetical Affiliation Network*

Actor 1	Actor 2	Actor 3	Actor 4	Event 1	Event 2	Event 3
1	0	0	0	1	0	0
1	0	0	0	0	0	1
0	1	0	0	0	0	1
0	0	1	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	1	0	0
0	0	0	1	0	1	0

the new array are coded as indicator variables for the variables in the original table. The pseudocontingency table has N rows (one for each observation) and $r + c$ columns (the total number of categories of the row variable plus the total number of categories of the column variable). This arrangement is standard in multiple correspondence analysis where more than two variables can be studied by including additional sets of indicator variables. For an affiliation network, the pseudocontingency table, which we denote by **F**, has $g + h$ columns, one for each actor and one for each event, and as many rows as there are entries of 1 in the affiliation network, a_{++}.

Table 7.15.6 shows the pseudocontingency table for the hypothetical example of four actors and three events. This table has seven columns (four actors plus three events) and seven rows (the number of 1s in the affiliation matrix). Notice that the first row of the table codes actor 1 attending event 1, the second row codes actor 1 attending event 3, and so on. Each row in the table has exactly two entries equal to 1 and the remaining entries equal to 0. The column totals are the marginal row and column totals from the original affiliation network – for actors the totals are a_{i+}, and for events the totals are a_{+j}.

Correspondence analysis of this array results in two sets of scores: one for each column of the table and one for each row of the table. Usually, the scores for the rows will not be of interest. As in correspondence analysis of a two-way table, scores for categories of the column variables represent the chi-square distances between the column profiles. Because all row totals in the pseudocontingency table are equal to the number of variables (two in our case) and there are as many rows as there are observations (a_{++}), the total number of 1s in the table is equal to $2 \times a_{++}$ and all marginal row proportions are equal:

$$\frac{f_{i+}}{f_{++}} = \frac{1}{a_{++}}. \tag{7.33}$$

As a consequence, the chi-square distance between two column profiles simplifies and is equal to:

$$d(j, j') = \sqrt{a_{++} \sum_{i=1}^{N} \left(\frac{f_{ij}}{f_{+j}} - \frac{f_{ij'}}{f_{+j'}} \right)} \tag{7.34}$$

(also see Carroll et al. 1986: p. 275).

The approach proposed by Carroll, Green, and Schaffer uses correspondence analysis of the matrix, **F**, through a singular value decomposition of the matrix $\mathbf{R}^{-\frac{1}{2}}\mathbf{F}\mathbf{C}^{-\frac{1}{2}} = \mathbf{X}\Delta\mathbf{Y}'$. The scores for rows and columns using Carroll, Green, and Schaffer's approach, which we denote $\ddot{u}_{ik}$ and $\ddot{v}_{jk}$, are then scaled as principal coordinates. Because the columns of **F** index both actors and events, using these scores, distances between points from different sets are interpretable as chi-square distances between columns of the pseudocontingency table.

These scores are related to the principal coordinates and standard coordinates for a correspondence analysis of **A** by the following equations:

$$\ddot{u}_{ik} = \tilde{u}_{ik}(1+\lambda_k)^{\frac{1}{2}} = \frac{u_{ik}}{\lambda_k}(1+\lambda_k)^{\frac{1}{2}} \tag{7.35}$$

$$\ddot{v}_{jk} = \tilde{v}_{jk}(1+\lambda_k)^{\frac{1}{2}} = \frac{v_{jk}}{\lambda_k}(1+\lambda_k)^{\frac{1}{2}}. \tag{7.36}$$

The singular values, $\ddot{\lambda}_k$, from the Carroll et al. (1986) approach are related to the singular values for the original two-way contingency table as:

$$\ddot{\lambda}_k = \frac{\lambda_k + 1}{2}. \tag{7.37}$$

The Carroll, Green, and Schaffer coordinates have weighted means of zero on each dimension:

$$\sum_{i=1}^{g} \ddot{u}_{ik}\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} \ddot{v}_{jk}\frac{a_{+j}}{a_{++}} = 0. \tag{7.38}$$

On each dimension, the weighted variances are equal to the squares of the singular values of the normalized pseudocontingency table. Carroll et al. (1986) demonstrated that for an array with two variables this is equal to $\frac{\lambda_k+1}{2}$, where the λ_k are the singular values of the original contingency table. Thus, for the Carroll, Green, and Schaffer coordinates, the weighed variances on each dimension are:

$$\sum_{i=1}^{g} \ddot{u}_{ik}^2\frac{a_{i+}}{a_{++}} = \sum_{i=1}^{h} \ddot{v}_{jk}^2\frac{a_{+j}}{a_{++}} = \frac{\lambda_k + 1}{2}. \tag{7.39}$$

There is now a *symmetric* interpretation of distances between objects from different sets. However, this symmetry comes at a cost. The chi-square distances between objects in the *same* set are completely determined by the row and column marginal totals in the original matrix, **A** (Carroll et al. 1986; Greenacre 1989). For example, the chi-square distance between rows i and i' of the original matrix is equal to:

$$d(i,i') = \sqrt{\frac{1}{\frac{a_{i+}}{a_{++}}} + \frac{1}{\frac{a_{i'+}}{a_{++}}}} \tag{7.40}$$

(Carroll et al. 1986: p. 275; Greenacre 1989: p. 359). A similar result holds for the distance between two columns of the original matrix. As Greenacre (1989) observed, because these within-set distances are fundamental to correspondence analysis of a two-way array, the fact that they are completely determined by the marginal totals in the pseudocontingency table limits the usefulness of this approach. Nevertheles,

Table 7.16.7. *Correspondence Analysis Scores for Hypothetical Network*

	Principal Coordinates		Standard Coordinates		Carroll, Green, Schaffer	
	1	2	1	2	1	2
Actor 1	−0.144	0.661	−0.224	1.323	−0.203	1.146
Actor 2	−1.155	0.000	−1.789	0.000	−1.623	0.000
Actor 3	−0.144	−0.661	−0.224	−1.323	−0.203	−1.146
Actor 4	0.866	0.000	1.342	0.000	1.217	0.000
Event 1	0.559	0.661	0.866	1.323	0.786	1.146
Event 2	0.559	−0.661	0.866	−1.323	0.786	−1.146
Event 3	−0.745	0.000	−1.155	0.000	−1.047	0.000

because of the interpretability of interset distances, it remains an option when a joint representation is desired.

7.16 Comparing Solutions

Let us now compare the results of these three possible scalings of correspondence analysis scores, beginning with the small hypothetical example. Table 7.16.7 provides sets of scores for the hypothetical example using each of the three approaches. Figures 7.16.2,

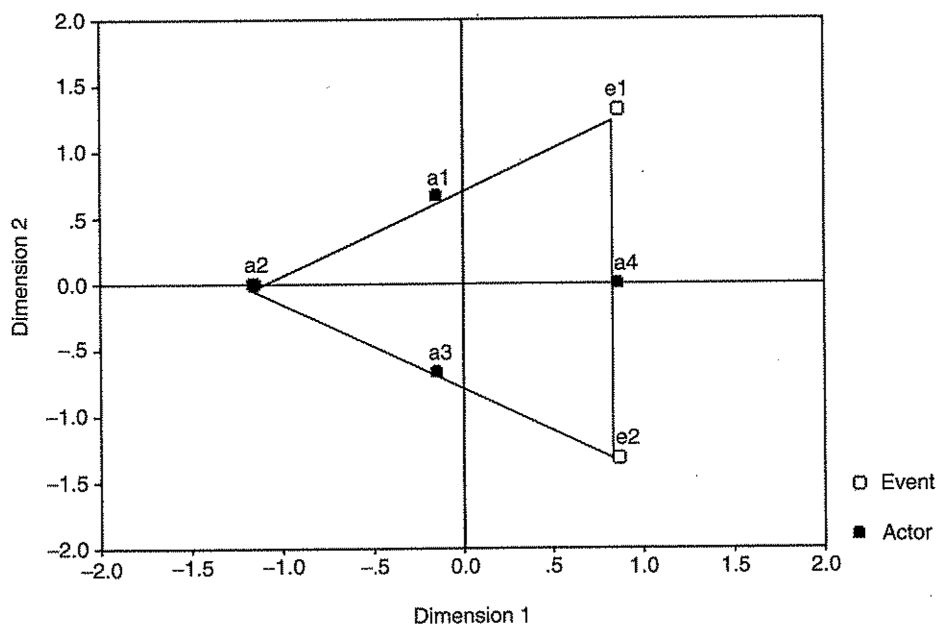

Figure 7.16.2. Correspondence analysis of hypothetical network: actors in principal coordinates and events in standard coordinates.

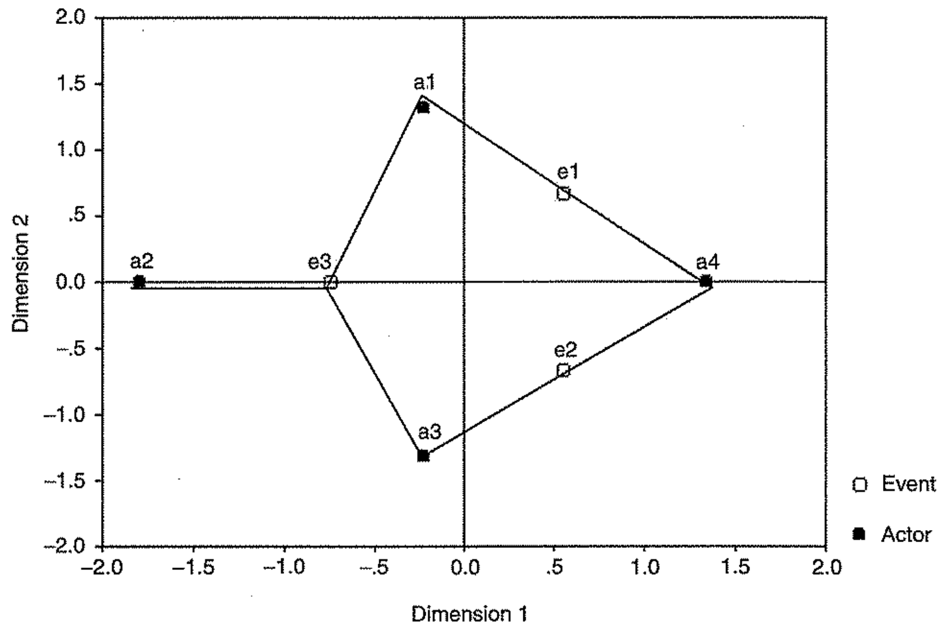

Figure 7.16.3. Correspondence analysis of hypothetical network: actors in standard coordinates and events in principal coordinates.

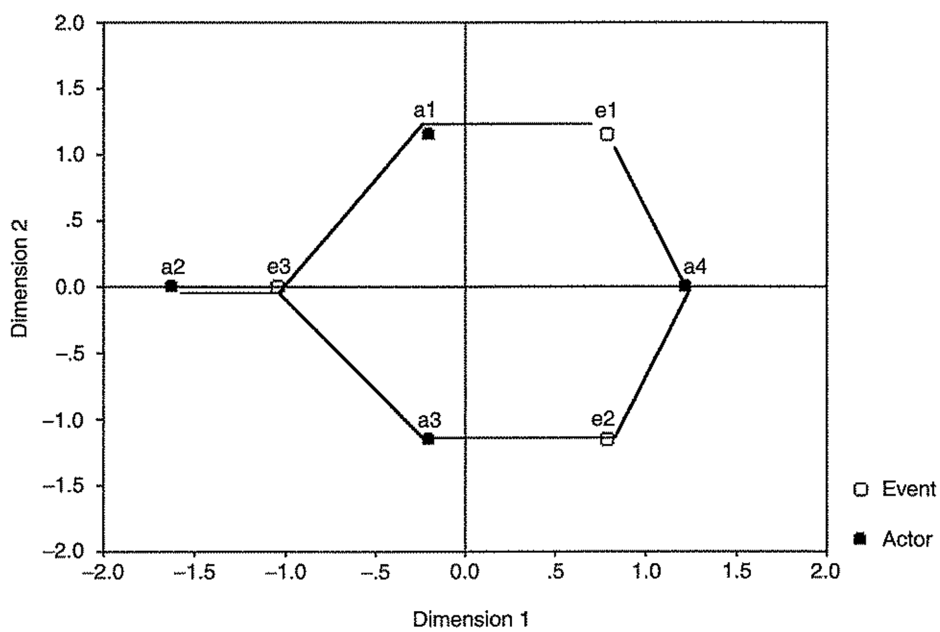

Figure 7.16.4. Correspondence analysis of hypothetical network: actors and events in Carroll, Green, and Schaffer coordinates.

7.16.3, and 7.16.4 present the first two nontrivial dimensions for each approach. Because there are four rows and three columns in the hypothetical network, the two-dimensional solution perfectly accounts for the data. The edges in the affiliation network are drawn on each figure to highlight the relationship between point locations and ties in the original network. Figure 7.16.2 shows the actors in principal coordinates and the events in standard coordinates. In this figure, the score for an actor (row) is the weighted average of the scores for the events (columns) to which it belongs. Because the two-dimensional solution perfectly fits the data, this interpretation is perfectly reflected in the two-dimensional graph, as can be seen from the fact that actors are literally "between" the events to which they belong. Notice that in this figure actor 2 and event 3 are in the same location because actor 2 only belongs to event 3. If a higher-dimensional solution were required to reproduce the data, the interpretation would only be approximated in a lower dimensions. Notice that in this figure the points for events (in standard coordinates) are around the perimeter of the figure, and points for actors (in principal coordinates) do not fall outside this polygon defined by the event points. This is generally the case because points in standard coordinates define the space onto which the other set of points are plotted (Greenacre and Hastie 1987). In contrast, Figure 7.16.3 reverses the roles of actors and events, presenting the events in principal coordinates and actors in standard coordinates. In this display, events are at the centroids of their constituent actors. The actors (in standard coordinates) are on the perimeter of the display, and events do not fall outside the polygon defined by the actor points. In both Figures 7.16.2 and 7.16.3, interpretation of the distances between actors and events is asymmetric and depends on which scaling is used. Distances relating points from different sets require situating a point from one set (in principal coordinates) in relation to *all* points from the other set (in standard coordinates). Figure 7.16.4 uses the Carroll, Green, and Schaffer scaling to produce display with a symmetric interpretation. In this figure, distance between a pair of points from different sets is interpreted as the chi-square distance between their respective column profiles in the pseudocontingency table.

7.17 Distances Versus Dimensions

Figures 7.16.2, 7.16.3, and 7.16.4 appear to provide rather different depictions of the affiliation network. Certainly, their mathematical properties and the distance interpretations allowed by the three are different. Nevertheless, it is useful to explore more fully the formal relationships among the three approaches and also to display these relationships graphically. First, notice that the principal coordinates, u_{ik} and v_{jk}; standard coordinates, $\tilde{u}_{ik}$ and $\tilde{v}_{jk}$; and the Carroll, Green, and Schaffer coordinates, $\ddot{u}_{ik}$ and $\ddot{v}_{jk}$, are, on each dimension, linear functions of one another. These relationships are:

$$\tilde{u}_{ik} = \frac{u_{ik}}{\lambda_k} \tag{7.41}$$

$$\ddot{u}_{ik} = \frac{u_{ik}}{\lambda_k}(1 + \lambda_k)^{\frac{1}{2}} \tag{7.42}$$

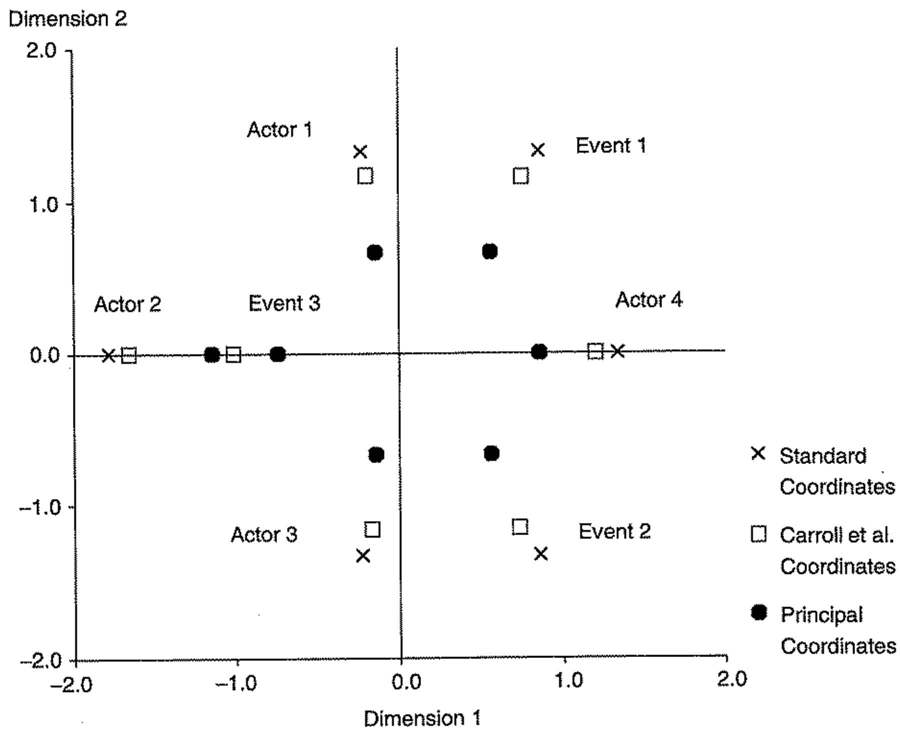

Figure 7.17.5. Three different scalings of correspondence analysis scores for hypothetical affiliation network.

$$\tilde{v}_{jk} = \frac{v_{jk}}{\lambda_k} \tag{7.43}$$

$$\ddot{v}_{jk} = \frac{v_{jk}}{\lambda_k}(1+\lambda_k)^{\frac{1}{2}}. \tag{7.44}$$

All three sets of scores have weighted means equal to 0, but they differ in their variances. Scores in principal coordinates have weighted variance equal to the square of the singular value on each dimension (7.19). Standard coordinates have weighted variance equal to 1 on each dimension (7.15). Finally, the Carroll, Green, and Schaffer coordinates have weighted variances equal to the squares of the singular values of the normalized pseudocontingency table (7.39). The implication of these relationships is that when the scores are used to define coordinate axes for graphic displays, the effect is a "stretching" or "shrinking" of each axis, as a function of the singular values on the various dimensions (Weller and Romney 1990). The overall impact of this stretching depends on how far the singular values depart from 1.0. For the hypothetical example, $\lambda_1 = 0.645$ and $\lambda_2 = 0.5$. The variances of principal coordinates, standard coordinates, and Carroll, Green, and Schaffer coordinates on the first dimension are 0.416, 1.0, and 0.822, and on the second dimension they are 0.25, 1.0, and 0.75, respectively.

The graphic impact of these alternatives is shown in Figure 7.17.5. In this figure, points for both actors and events in the hypothetical example are presented using all three alternatives in the same plot. By focusing on the positions of a single point across the different scalings, the stretching and shrinking effect is clear. Points in standard coordinates (variance of 1.0 on each dimension) are on the outside, points using the Carroll, Green, and Schaffer scores are next (variances of 0.822 and 0.75 on first and second dimensions), and the points in principal coordinates are on the inside (variances of 0.416 and 0.25).

7.18 Correspondence Analysis of Western Hemisphere Countries

Let us now return to the example of Western Hemisphere countries and their memberships in trade and treaty organizations. The affiliation network matrix for this example was presented in Table 7.4.1. Scores from the three different approaches are presented in Tables 7.18.8 and 7.18.9. Figures 7.18.6, 7.18.7, and 7.18.8 display the first two dimensions of each solution. Figure 7.18.6 uses principal coordinates for organizations and standard coordinates for countries. In this figure, each organization is the centroid of its member countries. In Figure 7.18.7, countries are presented in principal coordinates and organizations in standard coordinates. In this figure, each country is at the centroid of the organizations to which it belongs. Figure 7.18.8 uses the Carroll, Green, and Schaffer scores for coordinates. In this figure, the distance between a country and an organization is the chi-square distance between their respective columns in the pseudocontingency matrix (which is not presented).

Table 7.18.10 provides the squared singular values and percents of inertia for the first eleven dimensions. Using correspondence analysis of the affiliation matrix, **A**, the first five dimensions together account for 88.47% of the total inertia. In the Carroll, Green, and Schaffer approach the first five dimensions account for 94.66% of the variance. Table 7.18.10 also presents the contributions to the total inertia by each country and by each organization. Countries whose organizational memberships differ from the marginal profile (notably United States and Canada, but also Belize and Guyana) make larger contributions to the total inertia than do countries whose memberships are more similar to the marginal profile (notably, Chile, Bolivia, Ecuador, and Peru). Recall that this is measured as the chi-square distance between each profile and the marginal profile (7.22 and 7.23). Similarly, organizations whose memberships differ from the marginal distribution of members (e.g., CARICOM and NAFTA) contribute more to the total inertia than do organizations whose memberships are more similar to the marginal profile (e.g., SELA, OAS, and IDB).

Comparing Figures 7.18.6, 7.18.7 and 7.18.8, it can be seen that the major patterns are strikingly similar. Each has three clear branches reflecting the regional organization of this network. The first dimension contrasts South American countries and organizations on the one hand and Central American countries and organizations on the other hand. On the right are Parlacén and the San José Group, both organizations of Central American countries, along with El Salvador, Guatemala, and other Central American countries. On the lower left of each figure are Andean Pact, MERCOSUR, ALADI, Group of Rio, and Amazon Pact, organizations whose members are primarily in South America, along with Paraguay, Uruguay, Peru, Ecuador, and other South American countries. In all three figures, the second dimension clearly distinguishes Canada and the United States (both North American countries) along with NAFTA from other countries

Table 7.18.8. *Correspondence Analysis Scores for Organizations*

	Principal Coordinates			Standard Coordinates			Carroll, Green, and Schaffer Coordinates			Contribution to Inertia	
Organization	1	2	3	1	2	3	1	2	3	Total	Percent
ACS	0.73	−0.13	0.32	1.21	−0.24	0.71	1.08	−0.21	0.61	0.067	5.414
ALADI	−0.80	−0.15	−0.22	−1.34	−0.28	−0.49	−1.20	−0.25	−0.42	0.059	4.713
Amazon Pact	−0.52	−0.33	0.57	−0.87	−0.61	1.26	−0.77	−0.53	1.07	0.076	6.124
Andean Pact	−0.80	−0.37	0.33	−1.33	−0.69	0.73	−1.19	−0.60	0.62	0.081	6.518
CARICOM	0.74	−0.13	2.83	1.24	−0.25	6.26	1.11	−0.22	5.34	0.158	12.673
GENPLACEA	0.14	−0.22	−0.13	0.23	−0.42	−0.29	0.20	−0.36	−0.25	0.027	2.200
Group of Rio	−0.80	−0.15	−0.22	−1.34	−0.28	−0.49	−1.20	−0.25	−0.42	0.059	4.713
G-3	−0.55	0.22	0.33	−0.93	0.41	0.73	−0.83	0.36	0.62	0.083	6.696
IDB	0.08	0.23	0.01	0.14	0.43	0.01	0.12	0.38	0.01	0.015	1.209
MERCOSUR	−0.93	−0.26	−0.92	−1.56	−0.49	−2.03	−1.39	−0.43	−1.73	0.117	9.415
NAFTA	−0.02	3.41	−0.11	−0.04	6.37	−0.25	−0.04	5.58	−0.22	0.239	19.180
OAS	0.08	0.23	0.01	0.14	0.43	0.01	0.12	0.38	0.01	0.015	1.209
Parlacén	1.69	−0.38	−0.91	2.83	−0.72	−2.01	2.53	−0.63	−1.72	0.123	9.838
San José Group	1.44	−0.30	−0.61	2.40	−0.57	−1.36	2.15	−0.50	−1.16	0.114	9.167
SELA	0.08	−0.19	0.02	0.13	−0.36	0.05	0.12	−0.32	0.04	0.012	0.932
Total										1.246	100.00

Table 7.18.9. *Correspondence Analysis Scores for Countries*

	Principal Coordinates			Standard Coordinates			Carroll, Green, and Schaffer Coordinates			Contribution to Inertia	
Country	1	2	3	1	2	3	1	2	3	Total	Percent
Argentina	−0.52	−0.14	−0.46	−0.86	−0.26	−1.02	−0.77	−0.23	−0.87	0.043	3.487
Belize	0.57	0.00	1.41	0.96	0.00	3.12	0.86	0.00	2.66	0.111	8.914
Bolivia	−0.53	−0.22	0.10	−0.89	−0.41	0.22	−0.79	−0.36	0.19	0.035	2.787
Brazil	−0.56	−0.20	−0.25	−0.93	−0.37	−0.55	−0.84	−0.32	−0.46	0.041	3.288
Canada	0.08	2.41	−0.08	0.13	4.49	−0.17	0.12	3.94	−0.15	0.121	9.722
Chile	−0.45	−0.01	−0.18	−0.76	−0.02	−0.40	−0.68	−0.02	−0.34	0.031	2.469
Colombia	−0.40	−0.16	0.22	−0.66	−0.30	0.49	−0.59	−0.26	0.42	0.045	3.621
Costa Rica	0.71	−0.12	−0.14	1.18	−0.23	−0.32	1.06	−0.20	−0.27	0.035	2.811
Ecuador	−0.53	−0.22	0.10	−0.89	−0.41	0.22	−0.79	−0.36	0.19	0.035	2.787
El Salvador	1.01	−0.21	−0.41	1.69	−0.38	−0.91	1.51	−0.34	−0.77	0.065	5.224
Guatemala	1.01	−0.21	−0.41	1.69	−0.38	−0.91	1.51	−0.34	−0.77	0.065	5.224
Guyana	0.32	−0.14	1.15	0.53	−0.27	2.54	0.48	−0.24	2.16	0.083	6.657
Honduras	1.01	−0.21	−0.41	1.69	−0.38	−0.91	1.51	−0.34	−0.77	0.065	5.224
Mexico	−0.20	0.67	0.00	−0.33	1.25	0.00	−0.30	1.10	0.00	0.065	5.242
Nicaragua	0.71	−0.12	−0.14	1.18	−0.23	−0.32	1.06	−0.20	−0.27	0.035	2.811
Panama	0.71	−0.12	−0.14	1.18	−0.23	−0.32	1.06	−0.20	−0.27	0.035	2.811
Paraguay	−0.64	−0.09	−0.49	−1.07	−0.17	−1.08	−0.96	−0.15	−0.92	0.055	4.407
Peru	−0.53	−0.22	0.1	−0.89	−0.41	0.22	−0.79	−0.36	0.19	0.035	2.787
Suriname	0.15	−0.07	0.41	0.25	−0.13	0.90	0.23	−0.11	0.77	0.036	2.894
United States	0.08	2.41	−0.08	0.13	4.49	−0.17	0.12	3.94	−0.15	0.121	9.722
Uruguay	−0.52	−0.14	−0.46	−0.86	−0.26	−1.02	−0.77	−0.23	−0.87	0.043	3.487
Venezuela	−0.40	−0.16	0.22	−0.66	−0.30	0.49	−0.59	−0.26	0.42	0.045	3.621
Total										1.246	100.000

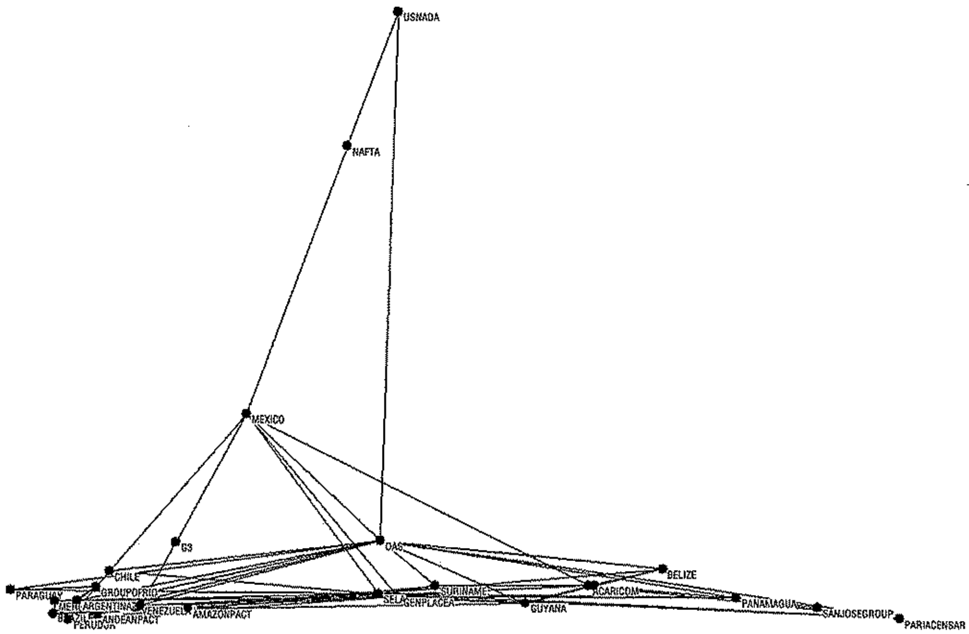

Figure 7.18.6. Correspondence analysis of Western Hemisphere countries and memberships in trade and treaty organizations. Countries in standard coordinates and organizations in principal coordinates.

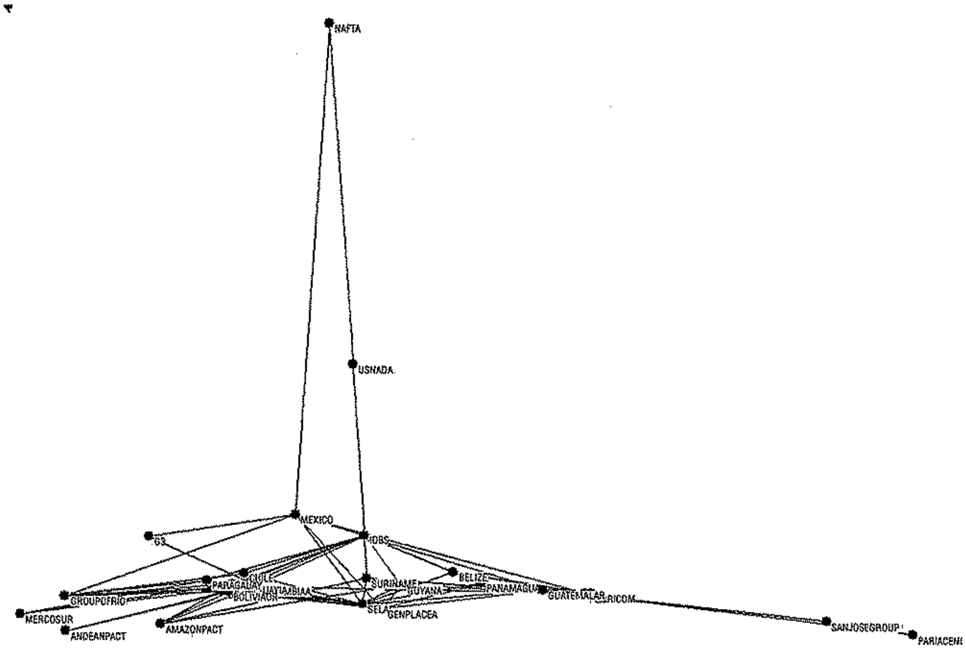

Figure 7.18.7. Correspondence analysis of Western hemisphere countries and memberships in trade and treaty organizations. Countries in principal coordinates and organizations in standard coordinates.

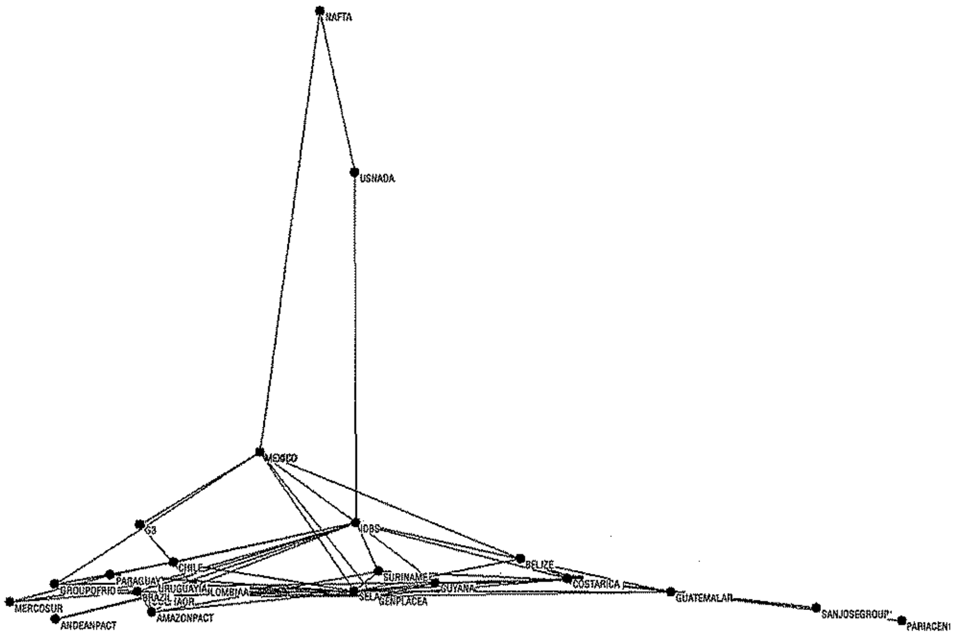

Figure 7.18.8. Correspondence analysis of Western Hemisphere countries and memberships in trade and treaty organizations, both countries and organizations in Carroll, Green, and Schaffer coordinates.

and organizations. Recall that Canada, the United States, and NAFTA all contributed substantially to the total inertia in the data. In all three figures, organizations whose members span the hemisphere (SELA, OAS, and IDB) are in the center at the bottom of the figure. As noted previously, SELA, OAS, and IDB all have membership profiles that are similar to the marginal profile, and their contributions to the total inertia are relatively small. The three figures do differ in minor details, which can be understood by recalling that the different approaches essentially stretch or shrink axes relative to one another.

7.19 Conclusion

In conclusion, let us return to the issue raised at the beginning of this chapter, in particular, the formal basis for an interpretable joint display of actors and events in an affiliation network. As the illustrations in this chapter demonstrate, joint graphic displays using correspondence analysis are possible, but require careful specification of which of a number of possible solutions is used for the display. The problem resides in appropriate interpretation of within-set and between-set distances between points. These interpretations require clear specification of which scores are used in order to avoid improper conclusions. That said, when viewed in concert, the various approaches are strikingly similar, at least for within-set comparisons.

When studying an affiliation network, the choice of which sets of scores should be used depends on theoretical and interpretative considerations. In an affiliation network,

Table 7.18.10. *Singular Values and Percent of Inertia Accounted for in Correspondence Analysis of Countries and Organizations*

	Principal Coordinates or Standard Coordinates				Carroll, Green, and Schaffer Scaling			
Dimension	Singular Values	Squared Singular Values	Percent of Inertia	Cumulative Percent of Inertia	Singular Values	Squared Singular Values	Percent of Inertia	Cumulative Percent of Inertia
1	0.598	0.358	28.69	28.69	0.894	0.799	29.71	29.71
2	0.536	0.287	23.05	51.73	0.876	0.767	28.53	58.24
3	0.452	0.204	16.39	68.12	0.852	0.726	26.99	85.23
4	0.416	0.173	13.88	82.00	0.416	0.173	6.43	91.66
5	0.284	0.081	6.47	88.47	0.284	0.081	3.00	94.66
6	0.247	0.061	4.89	93.37	0.247	0.061	2.27	96.93
7	0.196	0.038	3.08	96.45	0.196	0.038	1.43	98.35
8	0.169	0.029	2.29	98.74	0.169	0.029	1.06	99.42
9	0.110	0.012	0.97	99.71	0.110	0.012	0.45	99.87
10	0.049	0.002	0.19	99.90	0.049	0.002	0.09	99.95
11	0.035	0.001	0.10	100.00	0.035	0.001	0.05	1.000
Total		1.247				2.690		

one might view the social position of an actor as being defined by the social events in which it participates. With respect to the current example of Western Hemisphere countries, one could view a nation as described by the international organizations to which it belongs. Such a theoretical interpretation suggests that in a graphic display the location of an actor should be a function of the events with which it is affiliated. Alternatively, one could view the social location of an event as a function of its members. Again, for the current example, one would view an organization as defined by its constituent countries. A third possibility is that neither set of entities has precedence and that the relationships should be interpreted symmetrically. Correspondence analysis approaches exist for all three of these interpretations.

References

Blasius, Jorg, and Michael Greenacre. 1998. *Visualization of Categorical Data*. New York: Academic Press.

Borgatti, Stephen P., and Martin Everett. 1997. "Network analysis of 2-mode data." *Social Networks*, 19:243–269.

Brazill, Timothy J., and Bernard Groffman. 2002. "Factor analysis versus multidimensional scaling: Binary choice roll-call voting and the U.S. Supreme Court." *Social Networks*, 24:201–229.

Breiger, Ronald L. 1974. "The duality of persons and groups." *Social Forces*, 53:181–190.

Breiger, Ronald L. 2000. "A tool kit for practice theory." *Poetics*, 27:91–115.

Cappell, Charles L., and Thomas M. Guterbock. 1992. "Visible colleges: The social and conceptual structure of sociology specialties." *American Sociological Review*, 57:266–273.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1986. "Interpoint distance comparisons in correspondence analysis solutions." *Journal of Marketing Research*, 23:271–280.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1987. "Comparing interpoint distances in correspondence analysis solutions: A clarification." *Journal of Marketing Research*, 26:445–450.

Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer. 1989. "Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions." *Journal of Marketing Research*, 26:366–368.

CIA. 2000. *CIA World Fact Book*. Available at: http://www.odci.gov/cia/publications/factbook/.

Claussen, Sten-Erik. 1998. *Applied Correspondence Analysis*. Thousand Oaks, CA: Sage.

Coombs, Clyde. 1964. *A Theory of Data*. New York: Wiley.

Davis, A., B. Gardner, and M. R. Gardner. 1941. *Deep South*. Chicago: University of Chicago Press.

Digby, P. G. N., and R. A. Kempton. 1987. *Multivariate Analysis of Ecological Communities*. London: Chapman and Hall.

Doreian, Patrick. 1979. "On the evolution of group and network structure." *Social Networks*, 2:235–252.

East, Roger (ed.). 1996. *Keesing's Record of World Events*. London: Longman.

Ennis, James G. 1992. "The social organization of sociological knowledge: Modeling the intersection of specialties." *American Sociological Review*, 57:259–265.

Faust, Katherine. 1997. "Centrality in affiliation networks." *Social Networks*, 19:157–191.

Faust, Katherine, and Stanley Wasserman. 1993. "Correlation and association models for studying measurements on ordinal relations." In Marsden, Peter V. (ed.), *Sociological Methodology* Vol. 23. Cambridge, MA: Blackwell, pp. 177–215.

Faust, Katherine, Karin E. Willert, David D. Rowlee, and John Skvoretz. 2002. "Scaling and statistical models for affiliation networks: Patterns of participation among Soviet politicians during the Brezhnev era." *Social Networks*, 24:231–259.

Foster, Brian L., and Stephen B. Seidman. 1984. "Overlap structure of ceremonial events in two Thai villages." *Thai Journal of Development Administration*, 24:143–157.

Freeman, Linton. 2000. "Visualizing social networks." *Journal of Social Structures*, 1(1). Available at: http://www.heinz.cmu.edu/project/INSNA/joss/index1.html.

Freeman, Linton C. 2002. *Meta Analysis of Davis, Gardner, and Gardner Data*. Unpublished manuscript, University of California, Irvine.

Freeman, Linton C., and Douglas R. White 1993. "Using Galois lattices to represent network data." In Marsden, Peter V. (ed.), *Sociological Methodology*, pp. 127–146.

Galaskiewicz, Joseph. 1985. *Social Organization of an Urban Grants Economy*. New York: Academic Press.

Gifi, Albert. 1990. *Nonlinear Multivariate Analysis*. New York: John Wiley.

Greenacre, Michael. 1984. *Theory and Applications of Correspondence Analysis*. New York: Academic Press.

Greenacre, Michael. 1989. "The Carroll-Green-Schaffer scaling in correspondence: A theoretical and empirical appraisal." *Journal of Marketing Research*, 26:358–365.

Greenacre, Michael, and Jörg Blasius. (eds.). 1994. *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*. New York: Academic Press.

Greenacre, Michael, and Trevor Hastie. 1987. "The geometric interpretation of correspondence analysis." *Journal of the American Statistical Association*, 82:437–447.

Homans, George C. 1950. *The Human Group*. London: Routledge and Kegan Paul.

Jacoby, William G. 1991. *Data Theory and Dimensional Analysis*. Newbury Park, CA: Sage.

Kumbasar, Ece, A. Kimball Romney, and William H. Batchelder. 1994. "Systematic biases in social perception." *American Journal of Sociology*, 100:477–505.

Levine, Joel. 1972. "The sphere of influence." *American Sociological Review*, 37:14–27.

Mariolis, Peter. 1975. "Interlocking directorates and control of corporations: The theory of bank control." *Sociological Quarterly* 56:325–439.

Mariolis, Peter, and Maria H. Jones. 1982. "Centrality in corporate interlock networks: Reliability and stability." *Administrative Science Quarterly* 27:571–584.

McGrath, C., J. Blythe, and D. Krackhardt. 1997. "The effect of spatial arrangement on judgments and errors in interpreting graphs." *Social Networks*, 19:223–242.

McPherson, J. Miller. 1982. "Hypernetwork sampling: Duality and differentiation among voluntary organizations." *Social Networks*, 3:225–249.

Mische, Ann, and Philippa Pattison. 2000. "Composing in a civic arena: Publics, projects, and social settings." *Poetics*, 27:163–194.

Mizruchi, Mark S. 1982. *The American Corporate Network 1904–1974*. Beverly Hills, CA: Sage.

Nakao, Keiko, and A. Kimball Romney. 1993. "Longitudinal approach to subgroup formation: Reanalysis of Newcomb's fraternity data." *Social Networks*, 15:109–131.

Nishisato, Shizuhiko. 1994. *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Noma, E., and D. R. Smith. 1985. "Scaling sociomatrices by optimizing an explicit function: Correspondence analysis of binary single response sociomatrices." *Mulitvariate Behavioral Research*, 20:179–197.

Osa, Maryjane. 2003. *Solidarity and Contention: The Networks of Polish Opposition*. Minneapolis: University of Minnesota Press.

Pattison, Philippa, and Ronald Breiger. 2002. "Lattices and dimensional representations: Matrix decompositions and ordering structures." *Social Networks*, 24(4):423–444.

Roberts, John M., Jr. 2000. "Correspondence analysis of two-mode network data." *Social Networks*, 22:65–72.

Rosenthal, Naomi, Meryl Fingrutd, Michele Ethier, Roberta Karant, and David McDonald. 1985. "Social movements and network analysis: A case study of nineteenth-century women's reform in New York state." *American Journal of Sociology*, 90:1022–1054.

Schweizer, Thomas. 1991. "The power struggle in a Chinese community, 1950–1980: A social network analysis of the duality of actors and events." *Journal of Quantitative Anthropology*, 3:19–44.

Schweizer, Thomas. 1993. "The dual ordering of actors and posessions." *Current Anthropology*, 34:469–483.

Schweizer, Thomas, Elmar Klemm, and Margarete Schweizer. 1993. "Ritual as action in a Javanese community: a network perspective on ritual and social structure." *Social Networks*, 15:19–48.

Seidman, Stephen B. 1981. "Structures induced by collections of subsets: A hypergraph approach." *Mathematical Social Sciences*, 1:381–396.

Skvoretz, John, and Katherine Faust. 1999. "Logit models for affiliation networks." In Sobel, Michael E., and Mark P. Becker (eds.), *Sociological Methodology* Oxford: Basil Blackwell, pp. 253–280.

Stokman, Frans N. 1977. *Roll Calls and Sponsorship : A Methodological Analysis of Third World Group Formation in the United Nations*. Leyden, The Netherlands: A. W. Sijthoff.

Wasserman, Stanley, and Katherine Faust. 1989. "Canonical analysis of the composition and structure of social networks." In Clogg, Clifford C. (ed.), *Sociological Methodology* Vol. 19. Cambridge, MA: Basil Blackwell, pp. 1–42.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Wasserman, Stanley, Katherine Faust, and Joseph Galaskiewicz. 1990. "Correspondence and canonical analysis of relational data." *Journal of Mathematical Sociology*, 15:11–64.

Weller, Susan C., and A. Kimball Romney. 1990. *Metric Scaling: Correspondence Analysis*. Newbury Park, CA: Sage.

Wilson, Thomas P. 1982. "Relational networks: An extension of sociometric concepts." *Social Networks*, 4:105–116.

Appendix: List of Western Hemisphere Organizations

1. Association of Caribbean States (ACS): Trade group sponsored by the Caribbean Community and Common Market (CARICOM).
2. Latin American Integration Association (ALADI): Free trade organization.
3. Amazon Pact: Promotes development of Amazonian territories.
4. Andean Pact: Promotes development of members through economic and social integration.
5. Caribbean Community and Common Market (CARICOM): Caribbean trade organization; promotes economic development of members.
6. Group of Latin American and Caribbean Sugar Exporting Countries (GEPLACEA): Sugar-producing and exporting countries.
7. Group of Rio: Organization for joint political action.
8. Group of Three (G-3): Trade organization.
9. Inter-American Development Bank (IDB): Promotes development of member nations.
10. South American Common Market (MERCOSUR): Increases economic cooperation in the region.
11. North American Free Trade Agreement (NAFTA): Free trade organization.
12. Organization of American States (OAS): Promotes peace, security, economic, and social development in the Western Hemisphere.
13. Central American Parliament (PARLACÉN). Works for the political integration of Central America.
14. San José Group. Promotes regional economic integration.
15. Latin American Economic System (SELA): Promotes economic and social development of member nations.