

Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers

Ronald Breiger, Kathleen Carley, and Philippa Pattison,
Editors, Committee on Human Factors, National Research
Council

ISBN: 0-309-51916-0, 392 pages, 8.5 x 11, (2003)

**This PDF is available from the National Academies Press at:
<http://www.nap.edu/catalog/10735.html>**

Visit the [National Academies Press](http://www.nap.edu) online, the authoritative source for all books from the [National Academy of Sciences](http://www.nap.edu), the [National Academy of Engineering](http://www.nap.edu), the [Institute of Medicine](http://www.nap.edu), and the [National Research Council](http://www.nap.edu):

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the “[Research Dashboard](#)” now!
- [Sign up](#) to be notified when new books are published
- Purchase printed books and selected PDF files

Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, [visit us online](#), or send an email to feedback@nap.edu.

This book plus thousands more are available at <http://www.nap.edu>.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. [Request reprint permission for this book](#).

SESSION I

Social Network Theory Perspectives



Photograph by Ben Shahn, Natchez, MS, October, 1935

Library of Congress, Prints & Photographs Division, FSA-OWI Collection, [reproduction number, LC-USF33-006093-M4]

Finding Social Groups: A Meta-Analysis of the Southern Women Data¹

Linton C. Freeman
University of California, Irvine

1. Introduction

For more than 100 years, sociologists have been concerned with relatively small, cohesive social groups (Tönnies, [1887] 1940; Durkheim [1893] 1933; Spencer 1895-97; Cooley, 1909). The groups that concern sociologists are not simply categories—like redheads or people more than six feet tall. Instead they are social collectivities characterized by interaction and interpersonal ties. Concern with groups of this sort has been—and remains—at the very core of the field.

These early writers made no attempt to specify exactly what they meant when they referred to groups. But in the 1930s, investigators like Roethlisberger and Dickson (1939) and Davis, Gardner and Gardner (1941) began to collect systematic data on

¹ The author owes a considerable debt to Morris H. Sunshine who read an earlier draft and made extensive suggestions all of which improved this manuscript.

interaction and interpersonal ties. Their aim was to use the data both to assign individuals to groups and to determine the position of each individual—as a core or peripheral group member. But, to assign individuals to groups and positions, they needed to specify the sociological notions of group and position in exact terms.

Over the years a great many attempts have been made to specify these notions. In the present paper I will review the results of 21 of these attempts. All 21 tried to specify the group structure in a single data set. And 11 of the 21 also went on and attempted to specify core and peripheral positions.

My approach to this review is a kind of meta-analysis. Schmid, Koch, and LaVange (1991) define meta-analysis as “. . . a statistical analysis of the data from some collection of studies in order to synthesize the results.” And that is precisely my aim here. A typical meta-analysis draws on several data sets from a number of independent studies and brings them together in order to generalize their collective implications. Here I am also trying to discover the collective implications of a number of studies. But instead of looking at the results produced by several data sets, I will be looking at the results produced by several different analytic methods.

In this meta-analysis I will compare the groups and the positions that have been specified by investigators who examined data collected by Davis, Gardner and Gardner (1941) [DGG] in their study of southern women. My comparison draws on a number of techniques, including consensus analysis (Batchelder and Romney, 1986, 1988, 1989), canonical analysis of asymmetry (Gower, 1977) and dynamic paired-comparison scaling (Batchelder and Bershad, 1979; Batchelder, Bershad and Simpson, 1992). I will address two questions: (1) do the several specifications produce results that converge in a way that reveals anything about the structural form of the data? And, (2) can we learn anything about the strengths and weaknesses of the various methods for specifying groups and positions?

2. Southern Women Data Set

In the 1930s, five ethnographers, Allison Davis, Elizabeth Stubbs Davis, Burleigh B. Gardner, Mary R. Gardner and J. G. St. Clair Drake, collected data on stratification in Natchez, Mississippi (Warner, 1988, p. 93). They produced the book cited above [DGG] that reported a comparative study of social class in black and in white society. One element of this work involved examining the correspondence between people's social class levels and their patterns of informal interaction. DGG was concerned with the issue of how much the informal contacts made by individuals were established solely (or primarily) with others at approximately their own class levels. To address this question the authors collected data on social events and examined people's patterns of informal contacts.

In particular, they collected systematic data on the social activities of 18 women whom they observed over a nine-month period. During that period, various subsets of

these women had met in a series of 14 informal social events. The participation of women in events was uncovered using “interviews, the records of participant observers, guest lists, and the newspapers” (DGG, p. 149). Homans (1950, p. 82), who presumably had been in touch with the research team, reported that the data reflect joint activities like, “a day’s work behind the counter of a store, a meeting of a women’s club, a church supper, a card party, a supper party, a meeting of the Parent-Teacher Association, etc.”

This data set has several interesting properties. It is small and manageable. It embodies a relatively simple structural pattern, one in which, according to DGG, the women seemed to organize themselves into two more or less distinct groups. Moreover, they reported that the positions—core and peripheral—of the members of these groups could also be determined in terms of the ways in which different women had been involved in group activities.

At the same time, the DGG data set is complicated enough that some of the details of its patterning are less than obvious. As Homans (1950, p. 84) put it, “The pattern is frayed at the edges.” And, finally, this data set comes to us in a two-mode—woman by event—form. Thus, it provides an opportunity to explore methods designed for direct application to two-mode data. But at the same time, it can easily be transformed into two one-mode matrices (woman by woman or event by event) that can be examined using tools for one-mode analysis.

Because of these properties, this DGG data set has become something of a touchstone for comparing analytic methods in social network analysis. Davis, Gardner and Gardner presented an intuitive interpretation of the data, based in part on their ethnographic experience in the community. Then the DGG data set was picked up by Homans (1950) who provided an alternative intuitive interpretation. In 1972, Phillips and Conviser used an analytic tool, based on information theory, that provided a systematic way to reexamine the DGG data. Since then, this data set has been analyzed again and again. It reappears whenever any network analyst wants to explore the utility of some new tool for analyzing data.

3. The Data Source

Figure 1, showing which women attended each event, is reproduced from DGG (p. 148). DGG examined the participation patterns of these women along with additional information generated by interviews. As I discussed in Section 1 above, they had two distinct goals in their analysis: (1) they wanted to divide women up into groups within on the basis of their co-attendance at events, and (2) they wanted to determine a position—in the core or periphery—for each woman. As they put it (p. 150):

Where it is evident that a group of people participate together in these informal activities consistently, it is obvious

that a clique² had been isolated. Interviewing can then be used to clarify the relationship. Those individuals who participate together most often and at the most intimate affairs are called *core members*; those who participate with core members upon some occasions but never as a group by themselves alone are called *primary members*; while individuals on the fringes, who participate only infrequently, constitute the *secondary members* of a clique.

NAMES OF PARTICIPANTS OF GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN <i>Old City Herald</i>													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson.....	X	X	X	X	X	X		X	X					
2. Miss Laura Mandeville.....	X	X	X		X	X	X	X	X					
3. Miss Theresa Anderson.....		X	X	X	X	X	X	X	X					
4. Miss Brenda Rogers.....	X		X	X	X	X	X	X						
5. Miss Charlotte McDowd.....			X	X	X		X							
6. Miss Frances Anderson.....			X		X	X		X						
7. Miss Eleanor Nye.....					X	X	X	X						
8. Miss Pearl Oglethorpe.....						X		X	X					
9. Miss Ruth DeSand.....					X		X	X	X					
10. Miss Verne Sanderson.....							X	X	X			X		
11. Miss Myra Liddell.....								X	X	X		X		
12. Miss Katherine Rogers.....								X	X	X		X	X	X
13. Mrs. Sylvia Avondale.....							X	X	X	X		X	X	X
14. Mrs. Nora Fayette.....						X	X	X	X	X	X	X	X	X
15. Mrs. Helen Lloyd.....							X	X	X	X	X	X		
16. Mrs. Dorothy Murchison.....								X	X					
17. Mrs. Olivia Carleton.....								X		X				
18. Mrs. Flora Price.....								X		X				

Figure 1. Participation of the Southern Women in Events

Unfortunately, the data as presented in Figure 1 are not definitive. Indeed, two pages after presenting their data, DGG (p. 150) presented them again in another format. Their second version of the data is reproduced here as Figure 2. I have added annotations showing important comparative features in red.

The existence of that second presentation raises a problem; the data shown in Figure 1 do not agree with those shown in Figure 2. Specifically, in Figure 2 woman 15 was reported to have been a participant in events 13 and 14, but she was not so reported in Figure 1. In addition, in Figure 2 woman 16 was reported as participating in events 8, 9, 10 and 12. But in Figure 1, she was reported to have been a participant only in events 8 and 9. The extra events reported in Figure 2 but missing in Figure 1 are outlined in red.

² Note that DGG wrote before “clique” was defined as a technical term by Luce and Perry (1949). DGG used the word “clique” to mean what I am calling a “group.”

We are faced with a dilemma then. Are we to believe the data presented in Figure 1 or those presented in Figure 2? Homans (1950) was the first to use these data in a secondary analysis. He reported additional details that suggest that he was probably in touch with the original research team. And in his report he used the data as they are displayed in Figure 1.

Moreover, there is additional ancillary evidence for the correctness of the data as presented in Figure 1. It turns out that there is a contradiction in the presentation of Figure 2 that makes it difficult to accept the data presented there as correct. Compare, for example, the participation patterns displayed by the two women designated with red arrows in Figure 2: woman 11 and woman 16. According to Figure 2, these two women displayed identical patterns of participation. Yet, in that figure, woman 11 was classified as a primary member of her “clique” while woman 16 was called secondary. This contradiction implies that the correct data are those shown in Figure 1.

TYPE OF MEMBERSHIP	MEMBERS	EVENTS AND PARTICIPATIONS																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14			
<i>Clique I:</i>	1	C	C	C	C	C	C	-	C	C								
	Core.....	2	C	C	C	-	C	C	C	C	-							
		3	-	C	C	C	C	C	C	C	C							
	Primary...	4	C	-	C	C	C	C	C	C	-							
		5			P	P	P	-	P	-	-							
		6			P	-	P	P	-	P	-							
	Secondary.	7					P	P	P	P	-							
		8					-	S	-	S	S							
<i>Clique II:</i>	Secondary.	9	S						-	S	S	S						
		10							S	S	S	-	-	S				
	Primary...	11							-	P	P	P	-	P	←			
		12							-	P	P	P	-	P	P	P	←	
	Core.....	13							C	C	C	C	-	C	C	C		
		14							C	C	-	C	C	C	C	C		
		15							C	C	-	C	C	C	C	C		
	Secondary.	16							S	S	S	-	S	←				
		17							S	-	S							
		18							S	-	S							

Figure 2. Participation of the Southern Women in Events

Most analysts have apparently reached this conclusion. Most have used the data as shown in Figure 1. A few, however, have analyzed the data as presented in Figure 2, and this produces a problem for any attempt to compare the results of one analysis with those of another. When the two analyses are based on the use of different data sets comparisons are, of course, not possible.

I have assumed that the “correct” data are those shown in Figure 1. For the relatively small number of results that have been produced by analyses of the data of Figure 2, I have asked the original analysts to redo their analyses using the Figure 1 data or I have redone them myself.

4. Finding Groups and Positions in the DGG Data

4.1 Davis, Gardner and Gardner’s Intuition-Based Groups (DGG41)

In their own analysis Davis Gardner and Gardner did not use any systematic analytic procedures. They relied entirely on their general ethnographic knowledge of the community and their intuitive grasp of the patterning in Table 1 to make sense of the data. As Davis and Warner (1939) described it, they drew on “. . . records of overt behavior and verbalizations, which cover more than five thousand pages, statistical data on both rural and urban societies, as well as newspaper records of social gatherings . . .”

DGG drew on all this material and used it both to assign the women to groups and to determine individuals’ positions within groups. They indicated that the eighteen women were divided into two overlapping groups. They assigned women 1 through 9 to one group and women 9 through 18 to another. They assigned three levels in terms of core/periphery participation in these groups. They defined women 1 through 4 and 13 through 15 as core members of their respective groups. Women 5 through 7 and 11 and 12 they called primary. Women 8 and 9 on one hand and 9, 10, 16, 17 and 18 on the other were secondary. Note that woman 9 was specified as a secondary member of both groups because, they said, “in interviews” she was “claimed by both” (**DGG**, p. 151).

4.2 Homans’ Intuition-Based Analysis (HOM50)

Like Davis, Gardner and Gardner before him, Homans (1950) interpreted these data from an intuitive perspective. Unlike those earlier investigators, however, Homans did not have years of ethnographic experience in Natchez to draw upon. His intuitions, therefore, had to be generated solely by inspecting the **DGG** data and, presumably, by conversations with the ethnographers.

Homans implied that he had re-analyzed the data using a procedure introduced by Forsyth and Katz (1946) whom he cited. Forsyth and Katz had suggested permuting the rows and columns of a data matrix so as to display its group structure as clusters around the principal diagonal of the matrix (upper left to lower right). Their procedure required that both the rows and columns be rearranged until—as far as possible—more or less solid blocks of non-blank cells are gathered together. Such blocks of cells, they suggested, represent “well-knit” groups.

It is doubtful that Homans actually used the Forsyth and Katz procedure. **DGG** had already arranged the matrix in such a way that it displayed group structure. Seemingly they had anticipated the Forsyth and Katz approach by six years. Homans,

then, did not rearrange the data matrix at all; he simply copied the arrangement of Figure 1, exactly as it was reported by **DGG**.

In any case, after inspecting the arrangement shown in Figure 1, Homans grouped 16 of the women and distinguished two levels of core and peripheral positions. In his report Homans (p. 84) report wrote:

. . . we generalize these observations by saying that the 18 women were divided into two groups. The pattern is frayed at the edges, but there is a pattern. The first seven women, Evelyn through Eleanor, were clearly members of one group; numbers 11 through 15, Myra through Helen, were just as clearly members of another. Some women participated about equally with both groups but not very much with either; Pearl [woman 8] is an example. And some participated, though not very often, only with the second group. Pearl, Olivia, Flora [women 8, 17 and 18] and their like are marginal group members.

This statement is somewhat ambiguous. It does assign women 1 through 8 to one group and 11 through 15, along with 8, 17 and 18 to the other. Because woman 8 (Pearl) is assigned to both, the two groups overlap. In addition Homans characterized women 8, 17 and 18 to “marginal positions” but it is difficult to know what he intended by the phrase “and their like.” His statement, moreover, makes no mention at all of woman 9 (Ruth) or woman 16 (Dorothy). They were simply not assigned to either group or to any position.

4.3 Phillips and Conviser’s Analysis Based on Information Theory (P&C72)

Phillips and Conviser (1972) were the first to use a systematic procedure in the attempt to uncover the group structure in the **DGG** data.³ They reasoned that a collection of individuals is a group to the extent that all of the members of the collection attend the same social events. So, to examine the **DGG** data, they needed an index of the variability of attendance. They chose the standard information theoretic measure of entropy, **H** (Shannon, 1964). **H** provides an index of the variability of a binary (yes/no) variable. In this case, it was applied to all the women (and all the events) in the **DGG** data. Thus, **H** was used to provide an index of the degree to which different collections of women attended different sets of events (and different sets of events attracted different collections of women).

Phillips and Conviser set about to find groups by comparing various ways of partitioning the women into subsets. They argued that any given partitioning produced social groups if the entropy **H** summed for all of the subsets was less than the entropy for

³ It should be noted that Phillips and Conviser attributed the southern women data to Homans. Nowhere in their paper did they acknowledge **DGG**.

the total set of women. In such an event, the women assigned to each subset would be relatively homogeneous with respect to which events they attended.

They evaluated the utility of any proposed partitioning by calculating the information theoretic measure α . α is an index of the degree to which the overall entropy of the total collectivity H is reduced by calculating the value H_i within each of the i designated subsets and summing the results (Garner and McGill, 1956). α is large when all the women who are classed into each subset are similar with respect to their attendance patterns. It is maximal only when the within-subset patterns are all identical.

To employ this approach, then, it is necessary to partition the women in all possible ways, calculate α for each partitioning, and see which partitioning produces the largest value.

There is, however, a major difficulty with this approach. The number of possible partitionings grows at an exponential rate with an increase in the number of individuals examined. The number grows so rapidly that the partitions cannot all be examined, even with as few as 18 women to be considered.

So Phillips and Conviser worked out a way to simplify the problem. Like Homans, they cited again the procedure suggested by Forsyth and Katz (1946). That procedure rearranges the rows and columns in the data matrix in such a way that women who attended the same events and events that were attended by the same women are grouped together. When this is done, only those women who are close together in the matrix are eligible to be in the same group. That being the case, only those relatively few partitionings that include or exclude individuals in successive positions in the data matrix need to be considered.

As I indicated above, the **DGG** data had already been arranged in the desired order by the original authors. So, like Homans, Phillips and Conviser did not actually have to rearrange them. They could proceed directly to partitioning. They began by partitioning the women into two classes (1 versus 2 through 18, 1 and 2 versus 3 through 18, 1 through 3 versus 4 through 18, etc.). They reported that, of all these two-group partitions, the split of 1 through 11 versus 12 through 18 yielded the largest value of α .

In checking their results, however, I discovered that their result was based on an error in calculation. When I recalculated I discovered that the maximum value of α is actually achieved with the 1 through 9 versus 10 through 18 split. This approach simply partitions; it cannot distinguish core or peripheral positions, nor can it permit overlapping.

4.4 Breiger's Matrix Algebraic Analysis (BGR74)

Breiger (1974) used matrix algebra to show that the original two-mode, woman by event, **DGG** data matrix could be used to generate a pair of matrices that are,

mathematically, dual. First, multiplying the original matrix by its transpose produces a woman by woman matrix in which each cell indicates the number of events co-attended by both the row and the column women. Second, multiplying the transpose by the original matrix yields an event-by-event matrix where each cell is the number of women who attended both the row event and the column event. The woman-by-woman matrix is shown in Figure 3.⁴ And its dual, the event-by-event matrix, is shown in Figure 4.

		1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1
		E	L	T	B	C	F	E	P	R	V	M	K	S	N	H	D	O	F
1	EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
2	LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
3	THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
4	BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
5	CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
6	FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	0	0	0
7	ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
8	PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
9	RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
10	VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
11	MYRA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
12	KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
13	SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
14	NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
15	HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
16	DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
17	OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
18	FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

Figure 3. The One-Mode, Woman by Woman, Matrix Produced by Matrix Multiplication

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
1	E1	3	2	3	2	3	3	2	3	1	0	0	0	0	0
2	E2	2	3	3	2	3	3	2	3	2	0	0	0	0	0
3	E3	3	3	6	4	6	5	4	5	2	0	0	0	0	0
4	E4	2	2	4	4	4	3	3	3	2	0	0	0	0	0
5	E5	3	3	6	4	8	6	6	7	3	0	0	0	0	0
6	E6	3	3	5	3	6	8	5	7	4	1	1	1	1	1
7	E7	2	2	4	3	6	5	10	8	5	3	2	4	2	2
8	E8	3	3	5	3	7	7	8	14	9	4	1	5	2	2
9	E9	1	2	2	2	3	4	5	9	12	4	3	5	3	3
10	E10	0	0	0	0	0	1	3	4	4	5	2	5	3	3
11	E11	0	0	0	0	0	1	2	1	3	2	4	2	1	1
12	E12	0	0	0	0	0	1	4	5	5	5	2	6	3	3
13	E13	0	0	0	0	0	1	2	2	3	3	1	3	3	3
14	E14	0	0	0	0	0	1	2	2	3	3	1	3	3	3

Figure 4. The One-Mode, Event by Event, Matrix Produced by Matrix Multiplication

⁴ Breiger renamed DGG’s “Myra.” He listed her as “Myrna” in his tables and diagrams. Breiger’s designation has been picked up in a number of later works—including the data set released as part of the UCINET program (Borgatti, Everett and Freeman, 1992).

After examining the woman by woman matrix Breiger's conclusion was that ". . . everyone was connected to virtually everyone else." It was difficult, therefore, to separate the women into subgroups. So, in order to do that he turned to the dual, event by event, matrix. He reasoned that ". . . only those events that have *zero* overlap with at least one other event are likely to separate the women into socially meaningful subgroups." He found that events 6, 7, 8 and 9 were all linked to all of the other events; they contained no zero entries. So he eliminated their columns from the original two-mode data matrix. When those four columns were eliminated, women 8 and 16 were not participants in any of the remaining events so they were dropped from the analysis.

The next step was to recalculate a new woman-by-woman matrix from the reduced woman by event matrix from which the four linking events and the two uninvolved women had been removed. He then dichotomized the reduced data matrix and determined its clique⁵ structure. The result was a clear separation of the women into three cliques, two of which overlapped. Women 1 through 7 plus woman 9 formed one clique, women 10 through 15 formed another. In addition, women 14, 15, 17 and 18 formed a third. Thus, the latter two groups overlapped; women 14 and 15 were members of both the second and the third cliques. Breiger's procedure, then, permits the generation of more than two groups and it allows groups to overlap. But like the Phillips and Conviser approach, it cannot assign core or peripheral positions.

4.5 Breiger, Boorman and Arabie's Computational Analysis (BBA75)

Breiger, Boorman and Arabie (1975) reported on a new computational technique, **CONCOR**, designed for clustering binary relational data. **CONCOR** partitions the points in a graph into exactly two similarity classes, or blocks. Such blocks result when a data matrix can be permuted in such a way that some rectangular areas of the permuted matrix contain mostly ones, and other rectangular areas are predominately filled with zeros.

CONCOR begins with a data matrix, for example, the **DGG** women by event data. Then, using only the rows (women) or only the columns (events) **CONCOR** calculates either a row-by-row or column-by-column matrix of ordinary Pearsonian correlations. Then the rows (or columns) in this correlation matrix are again correlated and this process is repeated, again and again, until all the correlations are uniformly either plus or minus one. The convergence to correlations to plus or minus one seems always to occur. And the result is partitioning of the original data into two relatively homogeneous groups.

The application of **CONCOR** to the women (rows) of the **DGG** data set produced a partition. Women 1 through 7 and 9 were assigned to one group and 10 through 18 along with 8 were assigned to the other. Like the information theoretic approach used by

⁵ Here the term "clique" is used in the technical sense. It is a maximal complete subset (Luce and Perry, 1949).

Phillips and Conviser, **CONCOR** partitions the data. That means that **CONCOR**'s results cannot assign core or peripheral positions nor can they display groups that overlap.

4.6 Bonacich's Boolean Algebraic Analysis (BCH78)

Bonacich (1978) focused on the same duality that Breiger had noted, but he used a different algebraic tool. Instead of using Breiger's matrix algebra, Bonacich drew on Boolean algebra to specify both homogeneous groups of women and homogenous groups of events.

Bonacich began with the two-mode data set reported by **DGG**. His aim was to find a procedure for simplifying the data in such a way that—as much as possible—all the unions (**a** or **b**), intersections (**a** and **b**) and complements (not **a**) contained in the original data are preserved. Like Breiger, he drew on the duality between women and events. He strategically selected a subset of events (3, 8 and 12) that permitted him to divide the women into two groups in terms of their attendance at those events. The first group contains women 1, 2, 3, 4, 5 and 6 who were present at Event 3. The second group contains women 10, 11, 12, 13 and 15 who attended Event 12. Most of the women from both groups (all but woman 5 from the first group and woman 14 from the second group) also attended Event 8 along with four others. So, because they avoided attending event 8—that “bridged” between the two groups—Bonacich reasoned that they were “purer” representatives of their groups. He therefore defined women 5 and 14 as occupants of the central, or core, positions in their respective groups.

The approach used by Bonacich, then, divided the women into two groups and it also determined core and peripheral group members. But, because it ignored all but three events, it eliminated one-third of the women (7, 8, 9, 16, 17 and 18) from the analysis.

4.7 Doreian's Analysis Based on Algebraic Topology (DOR79)

Doreian (1979) drew on Atkin's (1974) algebraic topology in order to specify subgroups and positions in the **DGG** data. Atkin's model defines each event as the collection of women who attended it. Dually, each woman is defined as the collection of events she attended.

From the perspective of Atkin's model, two women are assumed to be connected to the degree that they are linked through chains of co-attendance. Women **A**, **B** and **C** form a chain at level 5 if a woman **A** co-attends 5 events with woman **B** and **B** attended 5 events with **C**. Then, even if **A** and **C** never attended any event together, all three are put together and their connection is assigned a level of 5. Doreian used this procedure to uncover the patterning of connections among the **DGG** women.

Women who form chains linked by co-attendance at 4 or more events, are divided into two groups. One contains women 1, 2, 3, 4, 5, 6, 7 and 9. The other includes women 10, 11, 12, 13, 14 and 15. Moreover, by considering subsets of women who were

connected at higher levels, Doreian was able to specify degrees of co-attendance ranging from the core to the periphery according of each group. His results place women 1 and 3 in the core of the first group, followed by 2 and 4 at the next level and 5, 6, 7 and 9 at the third. In the second group there were only two levels. Women 12, 13 and 14 were core, and 10, 11 and 15 were placed at the second level.

4.8 Bonacich's Use of Singular Value Decomposition (BCH91)

Bonacich (1991) used a form of singular value decomposition (SVD) to uncover groups in the DGG data. SVD is a tool from linear algebra. It is closely related to principal components analysis. Richards and Seary (1997) described the procedure:

. . . choose a set of axes in the multidimensional space occupied by the data and *rotate* them so that the first axis points in the direction of the greatest variability in the data; the second one, perpendicular to the first, points in the direction of greatest remaining variability, and so on. This set of axes is a coordinate system that can be used to describe the relative positions of the set of points in the data. Most of the variability in the locations of points will be accounted for by the first few dimensions of this coordinate system.

SVD can be used to analyze either one-mode or two-mode data. In this case, Bonacich applied it directly to the two-mode, women by events, data set. And he proposed that the first of the new transformed axes reveals both overall group structure and individual positioning. Individuals who displayed similar patterns of attendance at events are assigned similar scores on that axis. Moreover, that first axis is bipolar—it assigns both positive and negative scores to individuals. Bonacich indicated that women with positive values belong in one group and those with negative values in the other. The magnitudes of individual scores—positive or negative—can be taken as indices of core versus peripheral group membership. A large score associated with a woman indicates that she tended to present herself only at those events that were attended by other members of her own group and that she avoided events that were attended by members of the other group.

One of Bonacich's groups included women 1 through 9, and the other women 10 through 18. The order for the first group placed woman 5 in the core, followed by 4, 2, 1 and 6 together, and 3, 7, 9, and 8 in that order. In the second group, women 17 and 18 were together at the core, followed by 12, 13 and 14 together, then 11, 15, 10, and 16 in that order.

4.9 Freeman's Analysis Based on G-Transitivity (FRE92)

Freeman (1992) was the first to analyze the **DGG** data in a strictly one-mode form. Like Breiger (1974) he used matrix multiplication to produce the woman-by-woman matrix shown above in Figure 3. The cells in that figure show the number of events co-attended by the row-woman and the column-woman. The cells in the principal diagonal indicate the total number of events attended by each woman.

Freeman's analysis was based on an earlier suggestion by Granovetter (1973), hence the name, G-Transitivity. Granovetter focused on the strengths of the social ties linking individuals. He argued that, given three individuals **A**, **B** and **C** where both **A** and **B** and **B** and **C** are connected by strong social ties, then **A** and **C** should be at least weakly tied. Freeman assumed that the frequency of co-attendance provided an index of tie strength. He developed a computational model that works from the larger frequencies of co-attendance to the smaller. It determines a critical level of co-attendance below which the data violate Granovetter's condition. At the level just below the critical one, there is at least one triple where an **A** and a **B** are tied at that level, **B** and some **C** are tied at that level, but **A** and **C** have no tie at all. But at the critical level and all higher levels, the condition is met; so all ties that involve co-attendance at or above that critical level are, in Granovetter's sense, strong.

Applied to the **DGG** data this procedure divided fifteen of the eighteen women into two groups. The first contained women 1 through 7 and 9. The second contained women 10 through 16. Within each of those groups each woman was connected to every other woman on a path involving only strong ties. But there were no strong ties linking women across the two groups. Thus, this method could uncover groups in the **DGG** data set, but it could not distinguish between cores and peripheries. Moreover, this procedure was unable to assign group membership to women 8, 17 and 18.

4.10 Everett and Borgatti's Analysis Based on Regular Coloring (E&B93)

In an earlier paper Everett and Borgatti (1991) had defined regular equivalence in terms of graph coloring. Let the vertices of a graph represent social actors and the edges represent a symmetric relation linking pairs of actors. The neighborhood of a vertex is the set of other vertices that are directly connected to that vertex. Each vertex is assigned a color. Then any subset of vertices can be characterized by its spectrum, the set of colors assigned to its members.

Assigning colors to vertices partitions the actors into equivalence sets. And those equivalence sets are regular when all the vertices are colored in such a way that they are all embedded in neighborhoods with the same spectrum.

Everett and Borgatti (1993) used the concept of hypergraphs to generalize their regular coloring to two-mode data sets. And Freeman and Duquenne (1993) restated that generalization in simpler terms, referring only to ordinary bipartite graphs.

This approach draws on the duality of two-mode data. Women are connected only to events and *vice versa*. So the women are assigned colors from one spectrum and the events are assigned colors from another. Women are regularly equivalent when they participate in events that are regularly equivalent. Events are regularly equivalent when they have participants who are regularly equivalent. To simplify their analysis, Everett and Borgatti dropped four women (8, 16, 17 and 18). The remaining 14 women were partitioned into two regular equivalence classes, and, at the same time, the events were partitioned into three classes: Any women who attended any of events 1 through 5 and also any of events 6 through 9 were assigned to one class. Those women who attended any of events 6 through 9 and also any of 10 through 14 were assigned to the other class. The resulting partitioning of the fourteen women assigned women 1 through 7 along with 9 to one group and women 10 through 15 to the other. No assignments to core or peripheral positions were made.

4.11 Two Groupings Resulting from Freeman's Use of a Genetic Algorithm (FR193 and FR293)

Freeman (1993) produced another one-mode analysis of the DGG data. He began with the definition of group proposed by Homans (1950) and Sailer and Gaulin (1984). They defined a group as a collection of individuals all of whom interact more with other group members than with outsiders.

To explore this idea, Freeman used the same one-mode woman-by-woman matrix of frequencies that he had used in his earlier analysis of G-transitivity. He assumed that the women who co-attended the same social events would almost certainly have interacted. So he took this matrix as an index of interaction.

To find groups, then, one might examine all the possible partitionings of these 18 women and find any that meet the Homans-Sailer-Gaulin condition. But, in the discussion of the approach used by Phillips and Conviser above, the impossibility of searching through all the partitionings was established. Phillips and Conviser avoided the problem by limiting their search to only a small subset of the possible partitionings. Freeman (1993) took another approach. He drew on a search optimization algorithm to enhance the probability of finding partitionings that yield groups.

Freeman search was based on Holland's (1962) genetic algorithm. This approach emulates an actual evolutionary process in which pseudo-organisms "adapt" to the demands of an environmental niche. In this case, each pseudo-organism was associated with a particular partitioning of the 18 women. The niche was defined in terms of the Homans-Sailer-Gaulin definition of group. And the "fitness" of each grouping was evaluated in terms of the extent to which it approached that condition.

The search for an optimum partitioning was enhanced by allowing those partitionings with the highest fitness to crossbreed. In that way they produced another generation of "offspring" partitionings, each with some of the traits of their two parents.

And, to avoid getting the whole process locked into some less than optimal pattern, each partitioning in the new generation was subjected to a small chance of a mutation that would vary its structure.

The **DGG** data were entered into the genetic program and 500 runs were made. Two solutions that met the Homans-Sailer-Gaulin criterion were uncovered. The first, that occurred 327 times, found the same optimum revealed in the corrected Phillips and Conviser analysis. It grouped women 1 through 9 together and women 10 through 18 together. The second pattern occurred less frequently. It turned up 173 times and assigned women 1 through 7 to one group and women 8 through 18 to the other. These were the only partitions that displayed more interaction within groups than between groups.

4.12 Two Solutions Provided by Freeman and White's Galois Lattice Analysis (FW193 and FW293)

Freeman and White (1993) drew on another algebraic tool. They used Galois lattices to uncover groups and positions in the **DGG** data. Mathematically, a Galois lattice (Birkhoff, 1940) is a dual structure. It displays the patterning of women in terms of the events that each attended. And at the same time it shows the patterning of events in terms of which women attended each. Moreover it shows the containment structure of both the women and the events. A woman **A** "contains" another **B** if never attends an event where **A** is not present. And an event **X** contains event **Y** if no woman is present at **Y** who is not also present at **X**.

A Galois lattice, therefore, permits the specification of classes of events. And it allows us to define subsets of actors in terms of those event classes. Overall, it allows an investigator to uncover all the structure that was displayed in the earlier algebraic work by Breiger (1974) and Bonacich (1978), without the necessity of choosing arbitrary subsets of events in order to classify the women.

Freeman and White first reported the structure that was revealed by examining the overall lattice. Those results assigned women 1 through 9 and 16 to one group and women 10 through 18 to the other. Woman 16, then, was a member of both groups. Core and peripheral positions were assigned according to the patterning of containment. At the core of the first group were women 1, 2, 3 and 4. Women 5, 6, 7, and 9 were in the middle. And woman 16 was peripheral. The core of the second group contained women 13, 14 and 15. Women 10, 11, 12, 17 and 18 were in the middle, and again woman 16 was peripheral.

Freeman and White's second analysis was based on examining the two sub-lattices of women that were generated by the partitioning of events in the overall lattice. Two sets of events, 1 through 5 and 10 through 14 shared no common actors. So Freeman and White examined the sub-lattices generated by considering only these events.

This is exactly the event set that Breiger used in his matrix algebraic analysis described above. And its overall results are similar to those produced by Breiger.

This second analysis excluded women 8 and 16 who had not attended any of these ten events. It produced two non-overlapping groups that contained the remaining sixteen women. One group included women 1 through 7 along with 9. And the other included women 10 through 15 along with 17 and 18. Woman 1 was at the core of her group, followed by 2, 3 and 4 at the next level, then 5, then 6, and finally 7 and 9 together at the periphery. Woman 14 was at the core of her group, followed by 12, 13 and 15 at the next level, 11, 17 and 18 next, and finally by woman 10 at the extreme periphery.

4.13 Borgatti and Everett's Three Analyses (BE197, BE297 and BE397)

As part of a broad examination of techniques for the analysis of two-mode data, Borgatti and Everett (1997) used three procedures for finding groups in the **DGG** data. They began by constructing a one-mode *bipartite* matrix of the **DGG** data. A bipartite matrix represents a graph in which the nodes can be partitioned in such a way that all the ties in the graph connect nodes that fall in one partition with nodes that fall in the other; there are no within partition ties. The bipartite matrix produced by Borgatti and Everett is shown in Figure 5. The partition is between women and events. All the ties go from women to events or from events to women. There are neither woman-woman ties nor are there any event-event ties.

		1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3					
		E	L	T	B	C	F	E	P	R	V	M	K	S	N	H	D	O	F	E	E	E	E	E	E	E	E	E	E	E	E						
1	EVELYN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	0	0	0	0					
2	LAURA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	0					
3	THERESA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0					
4	BRENDA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	0	0	0	0					
5	CHARLOTTE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0					
6	FRANCES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0					
7	ELEANOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0				
8	PEARL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0				
9	RUTH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0				
10	VERNE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0				
11	MYRA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0			
12	KATHERINE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1			
13	SYLVIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1			
14	NORA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1			
15	HELEN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	0			
16	DOROTHY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
17	OLIVIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0		
18	FLORA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0			
19	E1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
20	E2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
21	E3	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	E4	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	E5	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	E6	1	1	1	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	E7	0	1	1	1	1	0	1	0	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	E8	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	E9	1	0	1	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	E10	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	E11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	E12	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	E13	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	E14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 5. One-Mode Bipartite Representation of the DGG Data

Borgatti and Everett defined an object on this matrix called an *n-biclique*. Like an *n-clique*, an *n-biclique* is a maximal complete subgraph in which no pair of points is at a distance greater than *n*. In this case, they were interested only in such bicliques where *n* = 2.

Since women are connected to events, a 2-bicliques must include both women and events. It could include a one-step connection from at least one woman to at least one event and another one-step connection from those events to at least one other woman. Or it might include a one-step connection from at least one event to at least one woman and another one-step connection from those women to at least one other event. Originally, Luce and Perry (1949) had required that a clique contain at least three objects. Borgatti and Everett generalized that requirement and specified that their 2-bicliques must contain at least three women *and* at least three events.

Of the 68 2-bicliques, only 22 contained 3 or more women and 3 or more events. Women 8, 16, 17 and 18 were not included in any of these bicliques. Everett and Borgatti recorded the pattern of overlap among the 22 cliques and then used the average method of Johnson's (1967) hierarchical clustering to define groups, cores and peripheries on the matrix of overlaps. Their results reveal two groups. One contains women 1 through 7 and 9. The other contains women 10 through 15. Women 3 and 4 are the core of the first group. They are followed by 2, 1, 7, 6, 9 and 5 in that order. Women 12 and 13 are the core of the second group. They are followed in order by 11, 14, 10 and 15.

Borgatti and Everett used the same bipartite data in a second approach. There they used a search algorithm to find the two-group partitioning that maximized the fit between the observed data and an idealized pattern. In their idealized pattern all within-group ties are present and no between-group ties are present. An optimal partition is sought using Glover's (1989) *tabu search* algorithm, and the fit of any partition is measured by correlation between that partition and the ideal. The result was a simple partition of all the women and all the events. The positions of individuals were not calculated. The tabu search found that the best two-group partition for the women was 1 through 9 and 10 through 18.

In their third analysis Borgatti and Everett used the regular two-mode **DGG** data. Here again they sought an optimal partitioning. And again the criterion for an optimum was correlation between a particular partitioning and the ideal that contained solid blocks of zero ties between groups and one ties within groups. Like Freeman (1993), Borgatti and Everett used a genetic algorithm to search for an optimum. But, while Freeman's search was made on the one mode woman by woman data, Borgatti and Everett searched the two mode, woman by event data. The result of their application of the genetic algorithm assigned women 1 through 9 to one group and 10 through 18 to the other.

4.14 Skvoretz and Faust's p^* Model (S&F99)

Skvoretz and Faust (1999) explored the ability of p^* , a family of structural models proposed by Wasserman and Pattison (1996) to uncover the important structural properties of the **DGG** data. Skvoretz and Faust developed several models in which various conditioning factors were used in the attempt to reproduce the patterning of the **DGG** data from a small number of parameters. Their best fitting model embodied three key parameters, all concerned with bridging ties. One parameter was based on the number of triads in which two actors were linked by an event. The second was based on the number of triads in which two events were linked by an actor. And the third took into account the distances between pairs of events as measured by the number of actors on the shortest path linking them.

Together, these parameters did a good job of capturing the tendency of these women to attend those events that brought the same sets of individuals together, again and again. These three factors, then, were included in a model that predicted the likelihood that each woman attended each event. In effect, Skvoretz and Faust used the p^* model to produce an idealized version of the **DGG** data, one that removed minor perturbations and—hopefully—captured the essence of the overall pattern of attendance. They produced, in effect, an “improved” version of the **DGG** data.

In order to uncover the group structure in this idealized data set, I clustered the data produced by the model using Johnson's (1967) complete link hierarchical clustering algorithm. That algorithm produces an ordering among sets of women by linking first those who were assigned the closest ties according to the p^* model, then proceeding to the next lower level and so on. The results divide the women into two groups, and they assign core and peripheral positions in those groups. One group included women 1 through 9 and the other women 10 through 15 as well as 17 and 18. Woman 16 was not included in either group. The two positional orderings were: 1 and 3, 2, 4, 5, 6, 7, 9, 8 for the first group and 12 and 13, 14, 15, 11, 10, and finally 17 and 18 together at the extreme periphery for the second.

4.15 Roberts' Singular Value Decomposition of Dual Normalized Data (ROB00)

Roberts (2000) calculated a “marginal free” (Goodman, 1996) two-mode analysis of the **DGG** data. He pre-processed the data using the classical iterative proportional fitting algorithm (Deming and Stephan, 1940). This produced a “pure pattern” (Mosteller, 1968) in which all differences in row and column marginals were eliminated. The two-mode data transformed to constant marginals are shown in Figure 6.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
1	EVELYN	0.349	0.349	0.078	0.174	0.025	0.019	0.000	0.005	0.002	0.000	0.000	0.000	0.000	0.000
2	LAURA	0.414	0.415	0.093	0.000	0.030	0.023	0.021	0.006	0.000	0.000	0.000	0.000	0.000	0.000
3	THERESA	0.000	0.522	0.117	0.260	0.037	0.028	0.027	0.007	0.003	0.000	0.000	0.000	0.000	0.000
4	BRENDA	0.523	0.000	0.117	0.261	0.037	0.028	0.027	0.007	0.000	0.000	0.000	0.000	0.000	0.000
5	CHARLOTTE	0.000	0.000	0.265	0.591	0.085	0.000	0.060	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	FRANCES	0.000	0.000	0.616	0.000	0.197	0.150	0.000	0.038	0.000	0.000	0.000	0.000	0.000	0.000
7	ELEANOR	0.000	0.000	0.000	0.000	0.375	0.285	0.268	0.072	0.000	0.000	0.000	0.000	0.000	0.000
8	PEARL	0.000	0.000	0.000	0.000	0.000	0.729	0.000	0.184	0.087	0.000	0.000	0.000	0.000	0.000
9	RUTH	0.000	0.000	0.000	0.000	0.501	0.000	0.358	0.096	0.046	0.000	0.000	0.000	0.000	0.000
10	VERNE	0.000	0.000	0.000	0.000	0.000	0.000	0.320	0.086	0.041	0.000	0.000	0.554	0.000	0.000
11	MYRA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.052	0.025	0.588	0.000	0.335	0.000	0.000
12	KATHERINE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.003	0.072	0.000	0.041	0.439	0.439
13	SYLVIA	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.006	0.003	0.070	0.000	0.040	0.428	0.428
14	NORA	0.000	0.000	0.000	0.000	0.000	0.024	0.023	0.000	0.003	0.069	0.005	0.039	0.419	0.419
15	HELEN	0.000	0.000	0.000	0.000	0.000	0.000	0.160	0.043	0.000	0.486	0.034	0.277	0.000	0.000
16	DOROTHY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.678	0.322	0.000	0.000	0.000	0.000	0.000
17	OLIVIA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.376	0.000	0.624	0.000	0.000	0.000
18	FLORA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.376	0.000	0.624	0.000	0.000	0.000

Figure 6. DGG Data After Roberts’ Iterative Proportional Fitting

Roberts then subjected the transformed data to a singular value decomposition. The result is a procedure like correspondence analysis or principal components analysis, but one in which *all* of the marginal effects have been removed. The groups produced linked women 1 through 9 and 10 through 18. Women 1 and 12 are in the cores of the two groups. Woman 1 is followed by 2, 3 and 4 who are tied at the next level. They are followed by 5, 6, 7, 9, and 8 in that order. Woman 12 is followed by 13, 14, 11, 15, 10 and 16 in that order. And finally women 17 and 18 are grouped together at the extreme periphery of the second group.

4.16 Osbourn’s VERI Procedure for Partitioning (OSB00)

Osbourn (1999) is a physicist who conducted experimental research on visual perception. His experiment examined how individuals perceived collections of points presented on a two-dimensional plane and categorized them into clusters. His empirical results suggested that individuals rescale a dumbbell-shaped area around each pair of points in the plane. An individual clusters a pair of points together if and only if no other point falls in the dumbbell-shaped area between them. That result was used to determine the threshold for grouping as a function of variations in the separation of pairs of points. That threshold was called the visual empirical region of influence (**VERI**).

Osbourn generalized these results and they have been used to produce an all-purpose clustering and pattern recognition algorithm. The algorithm turns out to have important applications that deal with a wide range of complex phenomena. The model has successfully been applied in a number of areas. Among them it has been applied to interpretation of inputs of odors reported by mechanical “noses,” and it has been used to interpret pictures of brain tissue produced by magnetic resonance imagery (**MRI**).

The publications include general descriptions of the **VERI** procedure, but they do not specify the details. So when, in the year 2000, I met Osbourn I asked him if I might try **VERI** on the one-mode woman-by-woman version of the **DGG** data. Because my

data were not based on measured physical data, he concluded that it would be necessary to develop a special form of the algorithm. He did exactly that and sent me a program that was specifically designed to tolerate data that were not scaled.

I tried his adapted algorithm on the **DGG** data, but it was limited. It could only do one partitioning—splitting the women into two groups at a point that, according to the **VERI** criterion, provided the most dramatic separation. That partitioning assigned women 1 through 16 to one group and women 17 and 18 to the other.

4.17 Newman's Weighted Proximities (NEW01)

In analyzing data on co-authorship, Newman (2001) constructed a weighted index of proximity for two-mode data. He assigned each author of a publication a weight inversely proportional to the number of co-authors he or she had in that publication. This weighting was based on the assumption that a large collection of co-authors might be less well connected to each other than a small collection.

Newman had reported his general approach, but, like Osbourn, he had not spelled out the details of his weighting scheme. But I had already noticed that the bridging events in the **DGG** data (E6 through E9) were larger than the other events, therefore such weighting had intuitive appeal for those data. Therefore, I asked Newman to run his weighting algorithm on the **DGG** data, and he obliged. The results were a transformation of the **DGG** data that took his differential weighting into account.

As in the case of the Skvoretz and Faust result described above, I needed a way to convert the Newman data into groups and positions. Just as I did above, I used the complete link form of Johnson's (1967) hierarchical clustering to do that conversion. The clustering algorithm divided the women into a group containing women 1 through 7 and 9 and another with woman 8 along with women 10 through 18. It placed women 1 and 2 at the core of their group. They were followed, in order, by 3, 4, 6, 5, and 7 and 9 tied for the peripheral position. The core of the other group contained 13 and 14. They were followed by 12, 11, 15 and 10 in that order, then 17 and 18 were placed together and finally 8 and 16 were together at the extreme periphery.

That completes the review of 21 analyses of the **DGG** data. All of those analyses—either directly or indirectly—have specified groups among the women. In addition, 11 of them have indicated women's positions in the core or periphery. The question for the next section, then, centers on an examination of what we can learn by considering all of these results together and conducting a meta-analysis.

5. Meta-Analysis of the Results

In Section 4, twenty-one analytic procedures produced two kinds of substantive results when they were applied to the **DGG** data. In every case, groups were specified and individuals were assigned to those groups. And eleven procedures went on to specify

various levels of core and peripheral positions and to assign individuals to those positions. The next two sections will review and analyze these two kinds of results. Section 5.1 will deal with groups and Section 5.2 will examine positions.

In each of these sections, meta-analysis will be used to try to answer the two questions posed in Section 1 above: (1) By considering all of the results together, can we come up with an informed description of the group structure revealed in the **DGG** data? And (2), can we distinguish between those analytic tools that were relatively effective and those that were less effective in producing that description?

5.1 Finding Groups

The classifications of women into groups by each of the 21 procedures described in Section 4 yields a 21 by 153 matrix. The 21 rows represent the analytic procedures and the 153 columns represent the unordered pairs of women [(1,2), (1,3) . . . (17,18)]. But, since there is a good deal of agreement among the procedures and since no procedure generated more than two groups, we can simplify their presentation. Figure 7 shows the whole pattern of assignment of women to groups.

In Figure 7 the 18 columns represent the 18 women. And the 21 rows refer to the 21 analytic procedures. A woman in each cell is designated by a “W.” Groups are designated by colors. All the red “Ws” in a given row were assigned to the same group by the procedure designated in that row. All the blue ones were assigned to a second group. And in the fourth row there are green “W’s” that were assigned to a third group. Any woman who was assigned to two groups by the procedure in question, received a pair of color codes.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	DGG41	W	W	W	W	W	W	W	W	WW	W	W	W	W	W	W	W	W	W
2	HOM50	W	W	W	W	W	W	W	WW			W	W	W	W	W	W	W	W
3	P&C72	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
4	BGR74	W	W	W	W	W	W	W	W	W	W	W	W	W	WW	WW		W	W
5	BBA75	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
6	BCH78	W	W	W	W	W	W				W	W	W	W	W	W			
7	DOR79	W	W	W	W	W	W			W	W	W	W	W	W	W			
8	BCH91	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
9	FRE92	W	W	W	W	W	W			W	W	W	W	W	W	W	W		
10	E&B93	W	W	W	W	W	W			W	W	W	W	W	W	W	W		
11	FR193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
12	FR293	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
13	FW193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	WW	W	W
14	FW293	W	W	W	W	W	W			W	W	W	W	W	W	W		W	W
15	BE197	W	W	W	W	W	W			W	W	W	W	W	W	W			
16	BE297	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
17	BE397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
18	S&F99	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W		W	W
19	ROB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
20	OSB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
21	NEW01	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W

Figure 7. Group Assignments by 21 Procedures

What we need here is a way to evaluate the pattern displayed in Figure 7. Batchelder and Romney (1986, 1988, 1989) developed a method called *consensus analysis* that will do just that. Consensus analysis was originally designed to analyze data in which a collection of subjects answered a series of questions. It is based on three assumptions: (1) there is a “true” (but not necessarily known) answer to each question, (2) the answers provided are independent from question to question and (3) all the questions are equally difficult. Given these assumptions, consensus analysis uses the patterning of agreements among subjects to estimate both the “true” answer to each question and the “competence” of each subject. “True” answers are determined by the overall consensus of all the subjects. And the “competence” of a given subject is a function of the degree to which that subject provides answers that are close to the consensual answers.

For each pair of women, each analytic tool was, in effect, asked a kind of true/false question: does this pair of women belong together in the same group? And each analytic tool answered that question—yes or no—for each of the 153 pairs of women.

Thus, consensus analysis can be used to address both questions of interest in the current context. It can determine the “true” answers and it can determine the “competence” of each procedure. It uses an iterative maximum likelihood procedure to estimate the “true” answers. For these data, the answer sheet is shown in Figure 8. There, a woman in a row is classified as a member of the same group as a woman in a column if there is a 1 in their cell. The bold entries show complete agreement among the 21 procedures. All in all, there are 25 pairs of women where all 21 analytic procedures agreed that they should be placed together. But even in those cases where there was less than total agreement, the probability of misclassification approached zero very closely. The very worst case was the estimate that woman 8 and woman 9 belonged together; in that case the maximum likelihood probability of error was .0008. According to this figure, then, the consensus of the analytic procedures is to assign women 1 through 9 to one group and women 10 through 18 to the other.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2		1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
3			1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
4				1	1	1	1	1	0	0	0	0	0	0	0	0	0
5					1	1	1	1	0	0	0	0	0	0	0	0	0
6						1	1	1	0	0	0	0	0	0	0	0	0
7							1	1	0	0	0	0	0	0	0	0	0
8								1	0	0	0	0	0	0	0	0	0
9									0	0	0	0	0	0	0	0	0
10										1	1	1	1	1	1	1	1
11											1	1	1	1	1	1	1
12												1	1	1	1	1	1
13													1	1	1	1	1
14														1	1	1	1
15															1	1	1
16																1	1
17																	1

Bold font indicates that all 21 algorithms put this pair together

Figure 8. The Groupings Estimated by Consensus Analysis

Estimates of the “competence” of the 21 analytic procedures are obtained using factor analysis. First, the number of matches in the answers to the 153 questions is calculated for each pair of procedures. Then the covariance in those answers is calculated for each pair. And finally, loadings on the first principal axis are determined using singular value decomposition.

The first axis of each—matches and covariance—provides an index of competence. These two indices are useful in estimating the competence of procedures only if they are in substantial agreement. In this case the correlation between the axis based on matches and that based on covariances was .967. This suggests that the patterning of agreements is robust and that the results are not simply an artifact of the computational procedures.

The estimates of competence based on matches are shown in Figure 9. These are probably overestimates because of the large number of zeros in the data matrix. Nonetheless, because of their high correlation with covariances, they can be taken as monotonically linked to the success of each procedure in approximating the consensus classifications shown in Figure 9.

	Code	Analysis	Closeness to the Matching Criterion
1	DGG41	Davis, Gardner and Gardner, Ethnography	0.920
2	HOM50	Homans, Intuition	0.854
3	P&C72	Phillips and Conviser, Information Theory	0.968
4	BGR74	Breiger, Algebra	0.933
5	BBA75	Breiger, Boorman and Arabie, CONCOR	0.927
6	BCH78	Bonacich, Boolean Algebra	0.841
7	DOR79	Doreian, Algebraic Topology	0.923
8	BCH91	Bonacich, SVD	0.968
9	FRE92	Freeman, G-Transitivity	0.926
10	E&B93	Everett and Borgatti, Regular Coloring	0.916
11	FR193	Freeman, Genetic Algorithm 1	0.968
12	FR293	Freeman, Genetic Algorithm 2	0.842
13	FW193	Freeman and White, Galois Lattice	0.917
14	FW293	Freeman and White, Galois Sub-Lattice	0.954
15	BE197	Borgatti and Everett, Bi-Clique	0.916
16	BE297	Borgatti and Everett, Taboo Search	0.968
17	BE397	Borgatti and Everett, Genetic Algorithm	0.968
18	S&F99	Skvoretz and Faust, p* Model	0.957
19	ROB00	Roberts, SVD with Normalization	0.968
20	OSB00	Osbourn, VERI Algorithm	0.543
21	NEW01	Newman, Weighted Co-Attendance	0.932

Figure 9. The “Competence” Scores of the 21 Analytic Procedures

Figure 9 shows variation in the degree to which the various methods produce group assignments that meet this new criterion. Method 20, Osbourn’s **VERI** algorithm is dramatically poorer than any of the others. And at the opposite extreme, six of the methods, 3 (Phillips and Conviser’s information theory), 8 (Bonacich’s correspondence analysis), 15 (Borgatti and Everett’s taboo algorithm), 11 (Freeman’s first genetic algorithm solution), 16 (Borgatti and Everett’s tabu search), 16 (Borgatti and Everett’s genetic algorithm) and 19 (Roberts’ correspondence analysis) are tied for the best performance. They are followed very closely by two additional methods, 18 (Skvoretz and Faust’s **p*** analysis) and 14 (Freeman and White’s Galois sub-lattice). All in all, then, we have eight methods that perform very well. They all assigned individuals to groups in a way that is in substantial agreement with the assignments uncovered by consensus analysis.

One area in which the analytic devices displayed no consensus at all has to do with overlap in-group memberships. Only four of the methods reviewed displayed overlapping groups. **DGG** and Homans both relied on intuition, so they could easily specify women who bridged between the groups they reported. **DGG** proposed that woman 9 was a member of their two groups: women 1 through 9 and 9 through 18. Homans saw woman 8 as a bridge between a group that included women 1 through 8 and one containing women 8, 11, 12, 13, 14, 15, 17 and 18. Breiger’s analysis was based in part on cliques. Thus, it could, and did allow overlaps. His method suggested that there were three groups, two of which overlapped. One group involved women 1 through 9, a

second involved women 10 through 15. And his third group included women 14 and 15 again, along with 17 and 18. Finally, the lattice analysis also allowed overlaps. Freeman and White found that, in lattice terms, woman 16 bridged between their two groups (women 1 through 9 plus 16 in the first, and women 10 through 18 in the second).

Clearly, there is nothing resembling a consensus in these reports of overlap. In fact, the closest thing there is to agreement can be found in the reports by **DGG** and Homans on one hand and the two solutions by Freeman using the genetic algorithm (**FR193** and **FR293**). The two Freeman solutions agreed that women 1 through 7 were a group, as were women 10 through 18. Women 8 and 9 were assigned to the first group by one solution and to the second group by the other. One possible interpretation of these results is that women 8 and 9 were both bridges between groups. This suggests that the lattice analysis supports both the **DGG** designation of woman 9 and the Homans designation of woman 8 as bridges. But beyond that, little can be generalized about these results. When it came to dealing with group overlap, these analyses certainly did not agree.

5.2 Specifying Positions: Core and Periphery

Next we turn to the question of the assignment of individuals to core and peripheral positions. In this case we are limited because only 11 of the analytic procedures produced such assignments. The designation of core and peripheral positions by those 11 procedures is shown in Figure 10. In each procedure, core and peripheral orders were assigned within each group. The first groups are shown on the left and the second groups on the right. Within each group, core/peripheral positions are shown left to right. The vertical lines show the divisions specified by the procedure. In the case of **DGG41**, for example, women 1, 2, 3 and 4 were at the core. They were followed by 5, 6 and 7. And finally, 8 and 9 were farthest from the core.

	First Group	Second Group
DGG41	1 2 3 4 5 6 7 8 9	13 14 15 11 12 9 10 16 17 18
HOM50	1 2 3 4 5 6 7 8	11 12 13 14 15 8 17 18
BCH78	5 1 2 3 4 6	14 10 11 12 13 15
DOR79	1 3 2 4 5 6 7 9	12 13 14 10 11 15
BCH91	5 4 2 1 6 3 7 9 8	17 18 12 13 14 1 15 10 16
FW193	1 2 3 4 5 6 7 8 9 16	13 14 15 10 11 12 17 18 16
FW293	1 2 3 4 5 6 7 9	14 12 13 15 11 17 18 10
BE197	3 4 2 1 7 6 9 5	12 13 11 14 10 15
S&F99	1 3 2 4 5 6 7 9 8	12 13 14 15 11 10 17 18
ROB00	1 2 4 3 5 6 7 9 8	12 13 14 11 15 10 16 17 18
NEW01	1 2 3 4 6 5 7 9	13 14 12 11 15 10 17 18 8 16

Figure 10. Core/Periphery Assignments by the 11 Analytic Procedures

In dealing with groups, consensus analysis was used to determine the “true” assignment of women to groups. That grouping drew upon the information provided by all the analytic devices considered simultaneously. In dealing with core and peripheral positions, it would be useful to be able to establish a similar criterion for the “true” positions in which to place individuals. But, unfortunately, consensus analysis could not be used here. The analytic devices displayed too much variability in assigning core and peripheral positions. Moreover, with only 11 analytic tools making assignments, there were fewer data points from which to generate reliable estimates.

As an alternative, I used two analytic tools. One is *canonical analysis of asymmetry* (Gower 1977; Freeman, 1997) and the other is *dynamic paired-comparison scaling* (Batchelder and Bershad, 1979; Batchelder, Bershad and Simpson, 1992; Jameson, Appleby and Freeman, 1999).

These two tools offer alternative ways of establishing the “true” or “best” ordering of individuals given something less than complete agreement among the eleven methods. The problem is simplified somewhat by partitioning the original 18 by 18 matrix. The “answer sheet” in Figure 8 above defined women 1 through 9 as one group and 10 through 18 as another. The assignments to core and peripheral positions reflect this division. All in all the eleven methods compared 537 ordered pairs of women in terms of which woman was nearer the core. Of these, 533 comparisons involved women from the same group; only four compared women from different groups. Therefore, since the overall pattern reflects the presence of the same groups that were specified above, I analyzed core/periphery positioning separately for each of those two groups. The core/periphery matrix for the first group is shown in Figure 11. That for the second is shown in Figure 12.

	1	2	3	4	5	6	7	8	9
1	0	4	4	5	8	8	9	7	9
2	2	0	3	3	8	9	9	7	9
3	1	3	0	3	7	7	8	6	8
4	2	2	2	0	8	8	9	7	9
5	2	2	2	2	0	5	5	6	6
6	0	0	1	0	2	0	5	6	7
7	0	0	0	0	1	1	0	6	5
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	4	0

Number of times the row woman was judged more central than the column woman

Figure 11. Matrix of Frequencies of Assignment to a “Closer to the Core” Position by 11 Analytic Procedures for the First Group of Nine

	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	1	4	3	3
2	7	0	0	0	1	4	5	5	5
3	8	7	0	2	4	6	5	6	6
4	9	9	3	0	3	6	5	7	7
5	10	9	5	2	0	8	5	7	7
6	7	4	2	0	0	0	5	7	7
7	0	0	0	0	0	0	0	1	1
8	2	1	1	1	1	1	3	0	0
9	2	1	1	1	1	1	3	0	0

Number of times the row woman was judged more central than the column woman

Figure 12. Matrix of Frequencies of Assignment to a “Closer to the Core” Position by 11 Analytic Procedures for the Second Group of Nine

These two data sets were first subjected to canonical analysis of asymmetry. As it is used here, the canonical analysis is based on an on/off—all or none—model. Each cell X_{ij} in the data matrix is compared with its counterpart X_{ji} . Whichever is the larger of the two is set equal to $\frac{1}{2}$ and the smaller is set equal to $-\frac{1}{2}$. If the two are equal, both are set to 0. Thus, either woman i is closer to the core than woman j , woman j is closer than i , or neither is closer.

The resulting matrix is called *skew-symmetric*. A skew-symmetric matrix displays an important property when it is analyzed using singular value decomposition. If a linear order is present, the first two axes of the output form a perfect arc around the origin. The arc arranges the points from the one at the extreme core to the one at the most extreme periphery.

The results for the data of Figures 11 and 12 both formed perfect arcs. They ordered the women in each of the two groups as shown in the third column of Figure 13. Those results show that the canonical all or none approach produces clear results. Women 5 and 6 fall at the same level, as do women 17 and 18. Other than these two positional ties, the canonical analysis displays a linear order for each of the two groups. It places woman 1 in the first group and woman 13 in the second in the extreme core positions. And it places women 8 and 16 in the extreme peripheries of their respective groups.

The other procedure, paired-comparison scaling, is not based on an all or none model. Instead, it provides real valued dominance scores that are sensitive to the proportion of judgments in which each woman is assigned a position nearer the core than each of the others. This approach is based on Thurstone’s (1927) method of paired comparisons. Again we begin by comparing each cell X_{ij} in the data matrix with its counterpart X_{ji} . Only this time we do not convert the numbers to dichotomous values. Instead, we convert them to probabilities:

$$P_{ij} = X_{ij} / (X_{ij} + X_{ji}).$$

These probabilities, then, are used to determine the core/periphery order.

Normal distribution assumptions are used to assign an initial level to each woman. Her initial level is determined by the number of others with whom she was compared, and the number of those comparisons in which she was determined to be nearer the core. These initial levels are then used in a recursive equation to assign each woman a final

Label	Woman	Rank on Canonical Analysis	Rank on Paired-Comparison
1	EVELYN	1	1
2	LAURA	2	2
3	THERESA	3	3
4	BRENDA	4	4
5	CHARLOTTE	5.5	5
6	FRANCES	5.5	6
7	ELEANOR	7	7
8	PEARL	9	9
9	RUTH	8	8
10	VERNE	6	6
11	MYRA	5	5
12	KATHERINE	3	3
13	SYLVIA	1	1
14	NORA	2	2
15	HELEN	4	4
16	DOROTHY	9	9
17	OLIVIA	7.5	7.5
18	FLORA	7.5	7.5

Figure 13. The Core-Periphery Rankings Assigned to Each Group of Women by the Two Scaling Procedures

level in the hierarchy. These final levels depend on the same two factors listed above. And, in addition, they depend on the number of comparisons in which a given woman was judged farther from the core and the average level of all those with whom she was compared

These calculations are repeated, again and again, each time adjusting for the changed values of the average levels of others. Finally, they converge and further computations are unnecessary. The final result is a scaling. Each woman is assigned a numerical score that represents her position in the continuum from core to periphery. The rank orders produced by this scaling are shown in the fourth column in Figure 13.

These two calculations involve approaches that are quite different. So it is heartening to discover that they order the women almost identically. The only difference is found in the first group. The canonical analysis shows a tie between woman 5 and

woman 6, but the paired-comparison analysis places woman 5 closer to the core than woman 6.

To evaluate the effectiveness of each of the 11 orders provided by the analytic procedures, I used **gamma**. **Gamma** provides an order-based measure of agreement. I compared each of the orders suggested by the procedures with the idealized orders provided by canonical analysis and paired-comparison analysis.

Because the two idealized orders are so similar, their **gammas** with the orders produced by the analytic procedures were, of course, nearly identical. The results for both the canonical and the paired-comparison standards are shown in Figure 14.

For both model-based standards, Homans' order produced the highest gamma. One must be careful, however, in looking at these values because different **gamma** calculations may be built on vastly different numbers of observations. In this case, the value of 1.0 associated with Homans' work was based on only 17 comparisons in the order of the women. In contrast the values associated with the two analyses by Newman were based on 58 and 59 comparisons respectively. Because Homans' report contained relatively less information about who was in the core and who was peripheral, it generated fewer predictions about positions. The predictions it did make happened to agree with the positional information produced by both criteria. But the Newman analyses both produced large numbers of predictions, and they were still mostly in agreement with those produced by the criteria.

Beyond Newman, the orders produced by Davis, Gardner and Gardner themselves, by Doreian, by Freeman and White in their first analysis, and by Skvoretz and Faust are consistently in agreement with the criteria. Their **gammas** are all above .9 and they are all based on at least 43 comparisons. At the opposite extreme, both Bonacich analyses and the Borgatti and Everett bi-clique analysis do not agree very well with the criteria.

ID	Code	Analysis	Gamma with Canonical	Number of Comps	Gamma with Paired-Comparison	Number of Comps
1	DGG41	Davis, Gardner and Gardner, Ethnog.	0.962	52	0.923	52
2	HOM50	Homans, Intuition	1.000	17	1.000	17
6	BCH78	Bonacich, Boolean Algebra	-0.111	9	0.000	10
7	DOR79	Doreian, Algebraic Topology	0.929	28	0.929	28
8	BCH91	Bonacich, Correspondence Analysis	0.313	67	0.324	68
12	FW193	Freeman and White, Full Lattice	0.953	43	0.953	43
13	FW293	Freeman and White, Sub-Lattice	0.867	45	0.870	46
14	BE197	Borgatti and Everett, Bi-Clique	0.385	39	0.350	40
17	S&F99	Skvoretz and Faust, p*	0.932	59	0.933	60
18	ROB00	Roberts, Correspondence Analysis	0.844	64	0.846	65
20	NEW01	Newman, Weighted Co-Attendance	0.966	58	0.932	59

Figure 14. Gammas Showing the Degree to which 11 Analyses Agreed with the Two Standards in Assigning Individuals to Core and Peripheral Positions

So again, we have been able to uncover something close to a consensus—this time with respect to core and peripheral positions. And we have again been able to find out something about the extent to which each of the analytic procedures approaches that consensus.

6. Summary and Discussion

6.1 Assignment to Groups

Each of the 21 analyses reported here assigned the **DGG** women to groups. Consensus analysis determined the agreement among the assignments. It turned out that there was a strong core of agreement among most of the analytic devices. The agreement was substantial enough to allow the model to be used specify a partition of the women into groups—one that captured the consensus of all the analyses. At the same time, the consensus analysis was also able to provide ratings of the “competence” of each of the analytic procedures.

The consensual assignment of women to groups and the “competence” ratings of the analytic methods were reported above. The “competence” scores were reflected in the first axis of an singular value decomposition of the matches generated by the methods in assigning pairs of women to the same or to different groups. In that earlier examination I reported only the first axis. But here, it is instructive to examine the second and third axes. They are shown in Figure 15.

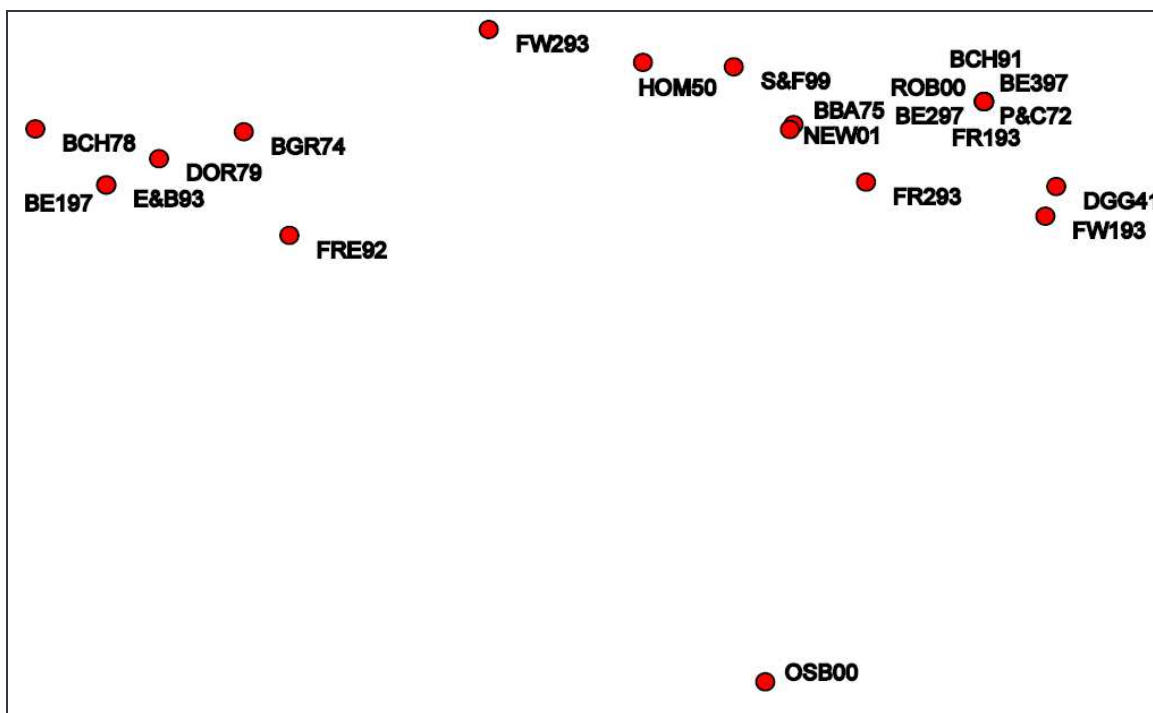


Figure 15. Axes 2 and 3 Produced by the Singular Value Decomposition of the Matches in the Assignments of Women

The arrangement of points representing the analytic methods in Figure 15 tells a good deal about both the partitioning of women to groups and the competence of the methods. The consensus put women 1 through 9 in the first group and women 10 through 18 in the second. In Figure 15, the basis for determining why this was the “best” partition becomes apparent. That partition was specified exactly by six of the analytic procedures: **P&C72** (The corrected version of Phillips and Conviser’s information theoretic algorithm), **BCH91** (Bonacich’s correspondence analysis), **FR193** (Freeman’s first genetic algorithm solution), **BE297** (Borgatti and Everett’s taboo search), **BE397** (Borgatti and Everett’s genetic algorithm) and **ROB00** (Roberts’ correspondence analysis of normalized data). These analyses are all placed at a single point at the upper right of the figure.

Other analyses that produced results that were quite close to that ideal pattern are clustered closely around that point. For example, **BBA75** and **NEW01** produced the same pattern with only one exception. They both assigned woman 8 to the second group. **FR292** assigned both women 8 and 9 to the second group. **DGG41** put woman 9 in both groups. And **FW193** put woman 16 in both. Finally, **S&F99** deviated only by failing to include woman 16 in either group. Thus, in addition to the six “perfect” partitionings, six additional procedures came very close to the ideal and are clustered in the region surrounding these “perfect” solutions. This clustering is the key. It shows a clear consensus around the 1-9, 10-18 division. This consensus is really remarkable in view of the immense differences among the analytic procedures used.

Figure 16 re-labels all the points such that their departure from the “perfect” partitioning is displayed. Note that the 1-9, 10-18 partition is labeled “**PERFECT.**” Note also that departures from that ideal are generally placed farther from the **PERFECT** point as the degree of their departure grows. They are, moreover, segregated in terms of the kinds of departure they embody. All the points that fall on the left of the vertical axis involve methods that failed to assign two or more of the women to groups. Overall, those points are arranged in such a way that those falling further to the left are those that are missing more women. Immediately to the right of the vertical, are the methods that located women in the “wrong” group. And their height indicates the number of women classified in “error.” All the way to the right are the methods that assigned women to both groups. And, to the degree that they assigned more women in that way, they are farther to the right. Finally, it should be noted that there are two analyses that assigned women to multiple groups on the left. But it is clear from their placement that this analysis was more responsive to their inability to place women in groups than it was to their dual assignments.

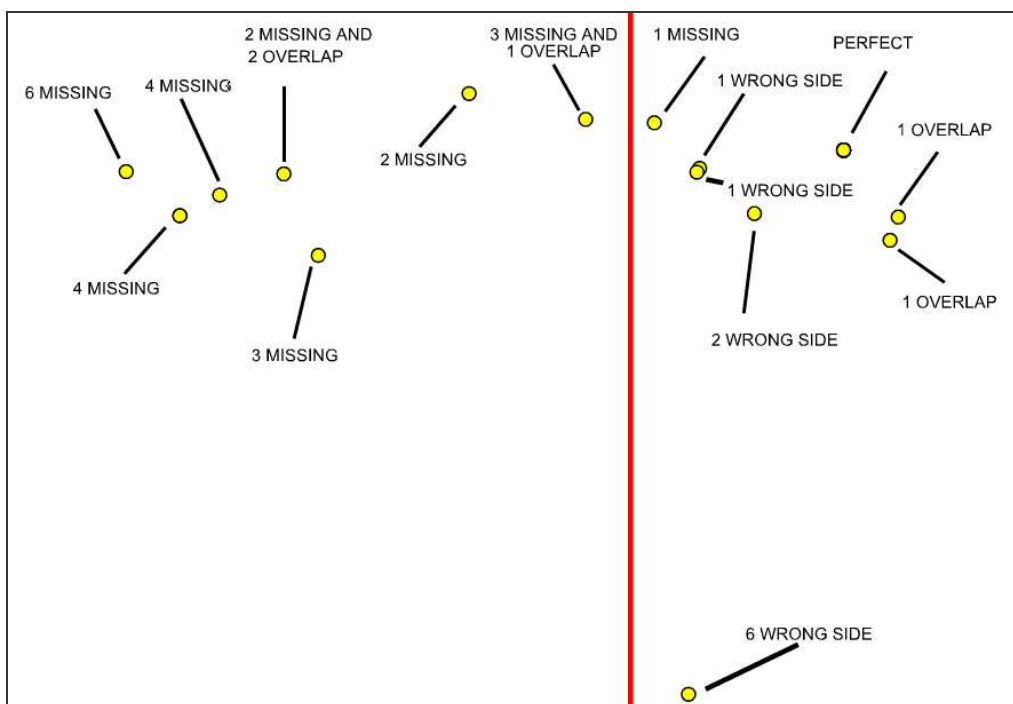


Figure 16. Axes 2 and 3 of the Matches Labeled by Structural Form

Overall, then, it is clear that there was a consensus about assigning women to groups. Six methods agreed, and most of the others departed relatively little from that agreed-upon pattern.

6.2 Positions in Groups

In assigning positions to individuals, I used two, quite different, scaling techniques. One was based on a dominance model. For each pair of women **A** and **B**, the model placed **A** closer to the core than **B**, if and only if more procedures placed **A** closer to the core than **B**. The other was probability-based. It placed **A** closer to the core than **B** with some probability based on the proportion of procedures that placed **A** closer to the core than **B**.

Despite their differences, the results of these two methods turned out to be almost identical. They were similar enough that either could be taken as providing something very close to an optimum assignment of individuals to positions. The effectiveness of each of the analytic procedures was evaluated by their monotone correlations with these optima. The results were very similar; the correlation between the gammas produced by the dominance model and those produced by the probability model was .983. So, even without consensus among the procedures, I was able to find the agreed-upon order—core to periphery—and to evaluate the ability of each method to uncover that order.

6.3 A Final Word

As a whole I believe that this meta-analysis has been productive. As far as assigning women to groups was concerned, the results were dramatic. When it came to assigning women to positions, the results were less dramatic, but still fairly convincing. We end, then, with four strong results: (1) We have a consensual partitioning of women into groups. (2) We have a consensual assignment of women into core and peripheral positions. (3) We have a rating of the methods in terms of their competence in assigning women to groups. And (4) we have a rating of the methods in terms of their ordinal correlations with the standard positional assignments.

I would like to wind up with two additional comparisons of the analytic procedures examined here. These final comparisons will be restricted to the published analyses of the **DGG** data; they will not include the unpublished analyses involving my use of Osbourn's **VERI** or the analysis Newman ran at my request.

The first comparison is based on time. Figure 17 shows the average competence ratings of procedures published at various points of time. The data in Figure 17 show an interesting secular trend. There has been a slow but consistent trend toward increasing competence through time. Thus, the overall tendency in published reports using the **DGG** data is clearly in the direction of greater competence.

In addition, it is possible to make some generalizations about the adequacy of the various kinds of analytic procedures that have been used to find groups and positions using the **DGG** data. Six procedures (**BGR74**, **BCH78**, **DOR79**, **E&B93**, **FW193** and **FW293**) all took essentially algebraic approaches. Five (**P&C72**, **FR193**, **FR293**, **BE197** and **BE297**) used various algorithms to search for an optimal partition. Three analyses (**BBA75**, **BCH91** and **ROB00**) employed various versions of singular value decomposition. Two (**DGG41** and **HOM50**) were based simply on the authors' intuitions. And three developed unique approaches. One (**FRE92**) looked at a kind of transitivity. A second (**BE197**) dealt with overlapping bicliques. And the third (**S&F99**) developed a statistical model.

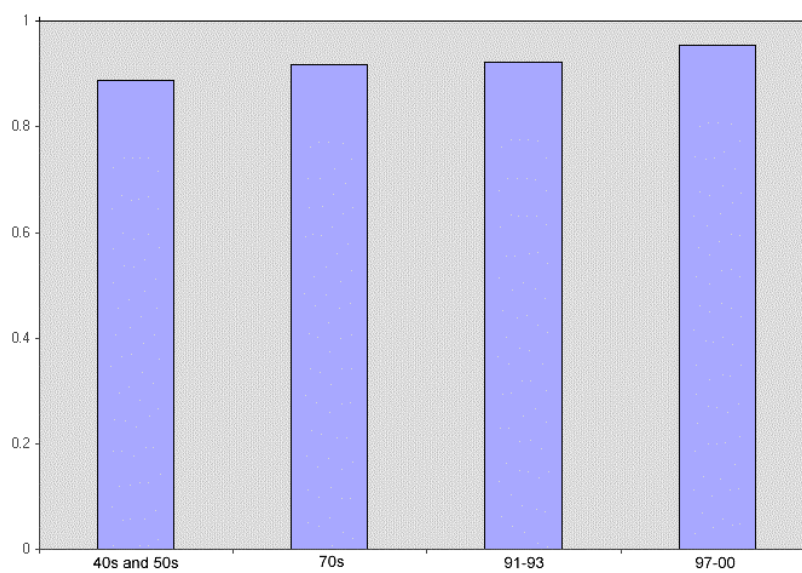


Figure 17. Competences of the Procedures Over Time

All in all, then, we have used seven distinct classes of procedures in analyzing the **DGG** data. Figure 18 shows the relative success of each class in terms of its average competence.

Procedure	N	Average Score
Statistical Model	1	0.957
Eigen Structure	3	0.954
Optimal Partition	5	0.941
Transitivity	1	0.926
Cliques	1	0.916
Algebraic Duality	6	0.914
Intuition	2	0.887

Figure 18. Average Competences of the Various Classes of Procedures

A number of features of Figure 17 are worth noting. First, the statistical model of the **DGG** data developed by Skvoretz and Faust was the winner. It won despite the fact that, unlike most of the other procedures, it was not explicitly designed to uncover groups. Group structure emerged as a sort of bi-product of a broader structural analysis.

The statistical model is not, however, the undisputed champion. It is followed so closely by the three singular value decomposition analyses, that it has to share the crown with them. And the five partitioning programs are right up there near the top.

There seems to be a step between all those procedures and the next three. Clearly transitivity, bicliques and the algebra-based approaches did not do as well. And, finally, the intuitive judgments fall at the bottom. In part that position is due to the vagaries of Homans' report, but **DGG** themselves did very little better. This result is particularly interesting given the fact that Davis, Gardner and Gardner's interpretation of their own data is often taken as privileged. The assumption has been that because they had a huge amount of ethnographic experience in the community, **DGG** had an edge—they somehow knew the “true” group structure. But, particularly in the light of the present results, there is no compelling reason to award **DGG** any special exalted status *vis-à-vis* their ability to assign individuals to groups. Indeed, their very intimacy with these 18 women might have led to various kinds of biased judgments.

References

- Batchelder, W. H. and N. J. Bershad
1979 The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology*, 19:39–60.
- Batchelder, W. H., N. J. Bershad and R. S. Simpson
1992 Dynamic paired-comparison scaling. *Journal of Mathematical Psychology*, 36:185–212.
- Batchelder, W. H. and A. K. Romney
1986 The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman and G. Owen (Eds.) *Information Pooling and Group Decision Making*. Greenwich, Connecticut: JAI Press, Inc.
- Batchelder, W. H. and A. K. Romney
1988 Test theory without an answer key. *Psychometrika* 53:71-92.
- Batchelder, W. H. and A. K. Romney
1989 New results in test theory without an answer key. In E. Roskam (Ed.) *Advances in Mathematical Psychology Vol. II*. Heidelberg New York: Springer Verlag.
- Birkhoff, G.
1940 *Lattice theory*. New York: American Mathematical Society.
- Bonacich, P.
1978 Using boolean algebra to analyze overlapping memberships.

Sociological Methodology 101-115.

Bonacich, P.

1990 Simultaneous group and individual centralities. *Social Networks* 13:155-168.

Borgatti, S. P. and M. G. Everett

1997 Network analysis of 2-mode data. *Social Networks* 19:243-269.

Borgatti, S. P., M. G. Everett and L. C. Freeman

1991 UCINET IV, Version 1.0. Reference Manual. Columbia, SC: Analytic Technologies.

Breiger, R. L.

1974 The duality of persons and groups. *Social Forces* 53:181-190.

Breiger, R. L., S. A. Boorman and P. Arabie

1975 An algorithm for clustering relational data, with applications to social network analysis and comparison to multidimensional scaling. *Journal of Mathematical Psychology* 12:328-383.

Cooley, C. H.

1909 *Social Organization*. New York: C. Scribner's sons.

Davis, A., B. B. Gardner and M. R. Gardner

1941 *Deep South*. Chicago: The University of Chicago Press.

Davis, A. and W. L. Warner

1939 A comparative study of American caste. In *Race Relations and the Race Problem*. (Ed.) E. T. Thompson. Durham: Duke University Press.

Deming, W. E. and F. F. Stephan

1940 On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* 11:427-444.

Doreian, P.

1979 On the delineation of small group structure. In H. C. Hudson, ed. *Classifying Social Data*, San Francisco: Jossey-Bass.

Durkheim, É.

[1893] 1933 *The division of labor in society*. Translated by G. Simpson. New York: Free Press.

- Everett, M. G. and S. P. Borgatti
1992 An Extension of regular colouring of graphs to digraphs, networks and hypergraphs. *Social Networks* 15:237-254.
- Forsyth, E. and L. Katz
1946 A matrix approach to the analysis of sociometric data. *Sociometry* 9: 340-347.
- Freeman, L. C.
1993 On the sociological concept of "group": a empirical test of two models. *American Journal of Sociology* 98:152-166.
- Freeman, L. C.
1994 Finding groups with a simple genetic algorithm. *Journal of Mathematical Sociology* 17:227-241.
- Freeman, L. C.
1997 Uncovering organizational hierarchies. *Computational and Mathematical Organization Theory* 3:5-18
- Freeman, L. C. and V. Duquenne
1993 A note on regular colorings of two mode data. *Social Networks* 15:437-442.
- Freeman, L. C. and D. R. White
1994 Using Galois lattices to represent network data. In *Sociological Methodology 1993*. P. Marsden, ed. Pp. 127-146. Cambridge, MA: Blackwell.
- Garner, W. R. and W. J. McGill
1956 The relations between information and variance analyses. *Psychometrika* 21: 219-228.
- Glover, F.
1989 Tabu search.-- Part 1. *ORSA Journal on Computing* 1: 190-206.
- Goodman, L. A.
1995 A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association* 91: 408-428.
- Gower, J.C.
1976 The Analysis of asymmetry and orthogonality. In J. Barra, F. Brodeau

and G. Romier (Eds.), *Recent Developments in Statistics*, North Holland, Amsterdam.

Granovetter, M.

1973 The strength of weak ties. *American Journal of Sociology* 81: 1287-1303.

Holland, J. H.

1962 Outline for a logical theory of adaptive systems. *Journal of the ACM*, 9:297-314.

Homans, G. C.

1950 *The Human Group*. New York: Harcourt, Brace and Company.

Jameson, K. A., M. C. Appleby and L. C. Freeman

1998 Finding an appropriate order for a hierarchy based on probabilistic dominance. *Animal Behaviour* 57:991-998.

Johnson, S. C.

1967 Hierarchical clustering schemes. *Psychometrika* 32:241-254.

Koch, G. G. and L. M. LaVange

1991 An overview of statistical issues and methods of meta-analysis
Journal of Biopharmaceutical Statistics 1:103-120

Luce, R. D., and A. Perry

1949 A method of matrix analysis of group structure," *Psychometrika*, 14:95-116.

Mosteller, F.

1968 Association and estimation in contingency tables. *Journal of the American Statistical Association* 63: 1-28.

Newman, M.

2001 The structure of scientific collaboration networks. *Proceedings of the National Academy of Science* 98:404-409.

Osborn, G. C.

1999 Clustering, pattern recognition and augmentation of data visualization using VERI algorithm. Invited talk, American Statistical Society Meeting.

Phillips, D. P. and R. H. Conviser

1972 Measuring the structure and boundary properties of groups: Some uses of information theory. *Sociometry* 35:235-254.

Richards, W. and A. Seary

- 1996 Introduction to eigen analysis of networks. Paper presented at the INSNA Sunbelt Meeting, San Diego.
<http://www.sfu.ca/~richards/Pages/wdr97.htm>

Roberts, J. M. Jr.

- 2000 Correspondence analysis of two-mode networks. *Social Networks* 22:65-72.

Roethlisberger, F. J. and W. J. Dickson

- 1939 *Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago.* Cambridge, Mass.: Harvard University Press.

Sailer, L. D., and S. J. C. Gaulin

- 1984 Proximity, sociality and observation: the definition of social groups. *American Anthropologist*, 86, 91-98.

Shanon, C. E.

- 1963 The mathematical theory of communication. In C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Skvoretz, J. and K. Faust

- 1999 Logit models for affiliation networks. *Sociological Methodology* 29:253-280.

Spencer, H.

- 1895-97 *The Principles of Sociology*. 3d ed., New York: D. Appleton.

Thurstone, L. L.

- 1927 A law of comparative judgment. *Psychological Review* 34:273-286.

Tönnies, F.

- [1855] 1936 *Fundamental Concepts of Sociology (Gemeinschaft und gesellschaft)*. Translated and supplemented by C. P. Loomis. New York: American Book Company.

Warner, M.

- 1987 *W. Lloyd Warner: Social Anthropologist*. New York: Publishing Center for Cultural Resources.

Wasserman, S. and P. Pattison

- 1997 Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika*, 60:401-426.