VISUALIZING SOCIAL GROUPS

Linton C. Freeman, University of California, Irvine
2143 SSPA, School of Social Sciences, Irvine, CA 92697-5100

Key Words: Social Networks, Visualization

Jacob Moreno [18, 19] introduced the idea of using visual images to reveal important features of social patterning. As he ([20], pp. 95-96) put it:

> A process of charting has been devised by the sociometrists, the *sociogram*, which is more that merely a method of presentation. It is first of all a method of exploration. It makes possible the exploration of sociometric facts . . . It is at present the only available scheme which makes *structural* analysis of a community possible.

Although they might not agree with the precise terms of Moreno's statement, most contemporary structural analysts would still subscribe to the view that visual images continue to play a key role in both exploring and presenting the patterns displayed by structural data.

Among the many kinds of structural patterns that have been explored and presented visually, one stands out. Structural analysts have devoted an enormous amount of attention to the problem of uncovering *social groups*. This attention to the group concept makes sense, since groups have been a concern of sociologists from the beginning (see Durkheim [4], Tönnies [27], Simmel, [25] and Cooley [3]).

All of these early sociologists saw groups in structural terms—as a patterning of interaction. Freeman and Webster [10] described this view:

> . . . whenever human association is examined, we see what can be described as thick spots—relatively unchanging clusters or collections of individuals who are linked by frequent interaction and often by sentimental ties. These are surrounded by thin areas—where interaction does occur, but tends to be less frequent and to involve little if any sentiment.

In addition, group members are believed to vary in terms of the degree to which they are involved in group activities. Some are core members; some are peripheral.

Visualization and the group concept, then, form a natural pair. Visualization is a tool for structural analysis, and group—in the sociologists' sense—is a structural concept. Investigators have been using visual tools to uncover and to communicate about groups for almost 70 years. And they continue to do so. In this paper I will review some of the earlier work and present an up to date description of contemporary approaches to visualization.

Much of the earlier work was based on the intuitive view that social connections are binary, or on/off. From this perspective, a pair of individuals is either linked or not linked. Some of the older and much of the newer work, however, has been based on a quantitative conception of social links. Pairs of individuals are seen as relatively more or relatively less strongly linked. These two approaches tend to yield different kinds of group images. They will be treated separately in the next two sections.

IMAGES OF BINARY STRUCTURES

Moreno viewed social linkages as binary. To reflect that binary view, he constructed his images out of points and lines. Individuals were pictured as points, and social ties linking pairs of individuals were shown as lines connecting pairs of points. One of his early images—one designed to show group structure—is shown in Figure 1.

Asking children to choose two others that should be assigned seats next to them produced the data. Directed lines indicate choices. Moreno drew all the boys (triangles) on the left and the girls (circles) on the right. This arrangement emphasizes the fact that there are many links within each gender, but only one link bridges the gender gap. Boys and girls, in effect, formed almost completely distinct groups.

Other investigators adopted Moreno's general approach, but many were uncomfortable with his *ad hoc* method for locating points. Since it did not involve any well-defined procedures, Moreno's method meant that no subsequent investigators could reproduce the results of earlier investigations. For that reason, Bock and Husain [1], Proctor [22], Laumann and Guttman [15] and Levine [16] all introduced computational procedures that were designed to locate points in such a way that members of the same group would automatically be placed in close proximity to each other. These procedures were all based on one of two approaches: 1) *multidimensional scaling* (Kruskal and Wish [14]) or 2) *singular value decomposition* (Weller and Romney [30]).

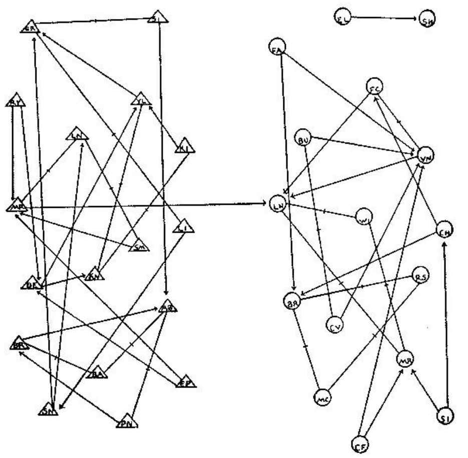

Figure 1. An Attraction Network in a Fourth Grade Class (from Moreno [19], p. 38).

We can learn something about how these two procedures work by applying them both to a single data set. In the 1930s Roethlisberger and Dixon [23] directed an extensive study of the organization of a part of the Western Electric Company. As a part of that study, they placed an observer for several months in a "wiring room" where telephone switches were constructed. The observer collected ethnographic data and described interaction patterns among nine wiremen, two solderers and an inspector in detail. The same observer also collected systematic data on various kinds of links among the workers. One data set recorded who played games with whom during off-hours. The data are shown in Table 1.

	I1	W1	W2	W3	W4	W5	W6	W7	W8	W9	S1	S4
I1	0	1	1	1	1	0	0	0	0	0	0	0
W1	1	0	1	1	1	1	0	0	0	0	1	0
W2	1	1	0	1	1	0	0	0	0	0	1	0
W3	1	1	1	0	1	1	0	0	0	0	1	0
W4	1	1	1	1	0	1	0	0	0	0	1	0
W5	0	1	0	1	1	0	0	1	0	0	1	0
W6	0	0	0	0	0	0	0	1	1	1	0	0
W7	0	0	0	0	0	1	1	0	1	1	0	1
W8	0	0	0	0	0	0	1	1	0	1	0	1
W9	0	0	0	0	0	0	1	1	1	0	0	1
S1	0	1	1	1	1	1	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	1	1	1	0	0

Table 1. Game Playing at Western Electric (from Roethlisberger and Dixon [23]).

Instead of drawing a point and line image with points placed in arbitrary locations, we can use one of our standard computational procedures to construct an image of the game-playing matrix. Let us begin with multidimensional scaling (MDS).

There are various versions of MDS, but they are all essentially search programs. They accept as input a matrix of proximities or a matrix of distances. Then they seek to arrange the points in such a way that the distances between pairs of points in the image correspond to the distances between individuals in the data matrix. With binary data, the aim is to place points close to each other if they are associated with individuals who are directly connected and to place points at a greater distance from each other if they are associated with individuals that are not directly connected. The resulting picture should display the structure contained in the data.

The investigator begins by specifying how many spatial dimensions the final image should have. For actual viewing, one, two or three dimensions are desirable. The program then produces a picture with the desired number of dimensions and reports an index of *stress* that tells how much distortion the picture produces.

A two-dimensional MDS image of the game-playing data is shown in Figure 2. The reported stress level is 0.0, so the image turns out to reproduce the patterning of who played games with whom quite well. And it is fairly obvious from looking at the figure that these individuals are divided into two groups. There are a good many ties linking the individuals within each group, and only one cross-group link.

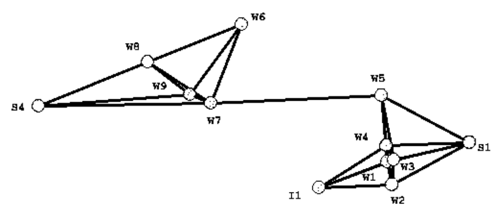

Figure 2. MDS of Game Playing at Western Electric (Raw Data).

This analysis was based on proximities. An alternative standard practice creates a distance-based image by calculating the graph theoretic distances—based on the shortest paths linking each pair of points. An MDS based upon these distances is shown in Figure 3.

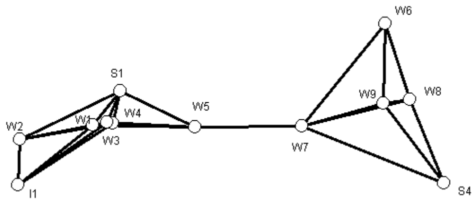

Figure 3. MDS of Game Playing at Western Electric (Graph Theoretic Distances).

The image in Figure 3 is very like the one shown in Figure 2. Going through the intermediate step of calculating graph theoretic distances seems to have had little impact here. This is probably the result of the fact that we are using non-metric MDS and it is producing only a monotone, not a linear, mapping of the input matrix to the visual image.

We can also use singular value decomposition (SVD) to assign locations to the points. Like MDS, there are several versions of SVD. But they are all alike in that they transform a data matrix into its *basic structure*. That transformation maps each observed variable (in this case one variable is associated with each of the 12 workers) into a new variable, or vector. These new vectors are produced in such a way that most of the variance in the original data is associated with the first new vector, next most with the second and so on.

The result is a new matrix of the same size as the original, but arranged in such a way that each successive variable accounts for less variance. If we are to produce a good image of structure of the data, the first two or three new vectors account for most of the important patterning of the original data. When that occurs, SVD will produce a good image of the structure of the original data.

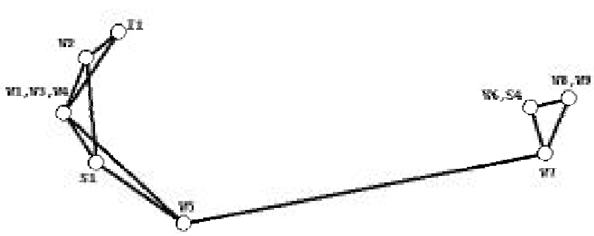

Figure 3. SVD of Game Playing at Western Electric.

In this case, the first two vectors of the new matrix captured almost 90% of the variance in the game-playing data. The image is shown in Figure 3. The analysis was generated by a version of SVD called *principal components* analysis. But any other version would have produced much the same image.

This image seems again to divide everyone up into two clusters. Here, however, some sets of individuals are collapsed; they are represented by a single point. This is the result of the fact that the collapsed points are *structurally equivalent*; they exhibit the same pattern of connections with the same others.

So, we have two images that seem to find groups, but it would be nice to have some baseline to compare them to. At this point all we know is that the algorithms both produce structures that seem to divide everyone up into two clusters, but we no nothing about whether these clusters represent groups or not. What we need is some sort of criterion that we can use in evaluating these displays.

Two such criteria are available: 1) groups as specified in ethnographic reports and 2) groups based on computations resulting from formal specification of group properties. I will use both of these criteria to evaluate the images produced by MDS and SVD.

First, let us examine the ethnographic report based on the observer's experience in the wiring room. The observer concluded that the workers were divided into exactly two groups. One contained four of the wiremen (designated ***W1***, ***W2***, ***W3*** and ***W4***), a solderer (***S1***) and an inspector (***I1***). The other also contained four wiremen (***W6***, ***W7***, ***W8*** and ***W9***) and the other solderer (***S4***). He concluded, moreover, that each of the groups had core and peripheral members. In the first group, ***W3*** was a "leader" and ***W2*** was "marginal." In the second, ***W6*** "was not entirely accepted by the group" and ***S4*** "was socially regarded as inferior." Finally, the remaining wireman, ***W5***, reportedly was not fully accepted by either group; he was designated an "outsider."

In broad terms, this report corresponds to both figures quite closely. In both cases the clusters shown match those produced by the ethnographic description, with the exception of ***W5*** who was assigned a peripheral position in the first group in both images. ***W3*** is certainly put in the center of the group in Figure 2, but is not singled out in Figure 3. ***W2***, ***W6*** and ***S4*** also appear to be more marginal in Figure 2. In this case, then, both MDS and SVD succeed in reflecting the overall structure reported by the ethnographer, but MDS captures more of the subtle detail.

Luce and Perry [17] introduced the first formal definition of social groups. They defined a *clique* as a maximal complete subgraph. That idea had, and continues to have great intuitive appeal. Yet, when cliques are calculated and compared with groups defined by ethnographers or by the participants themselves, they consistently show four weaknesses: 1) there are too many of them, 2) they are too small, 3) they overlap too much and 4) because they are complete they can display no internal structure.

There is, however, a procedure based on *Galois lattices* that has been used to define bounded groups in terms of the patterning of overlap among cliques (Freeman, [6]). The Galois lattice constructs a partial order among the individuals. That order for the Western Electric workers is shown in Figure 4. The five cliques, labeled 1 through 5, are shown at the top of the figure. The 12 individuals are shown below.

Clique overlaps are shown in terms of their memberships; Cliques 4 and 5, for example, share three members, ***W7, W8*** and ***W9***, and each has a single defining member (***S4*** for clique 5 and ***W6*** for clique 4).

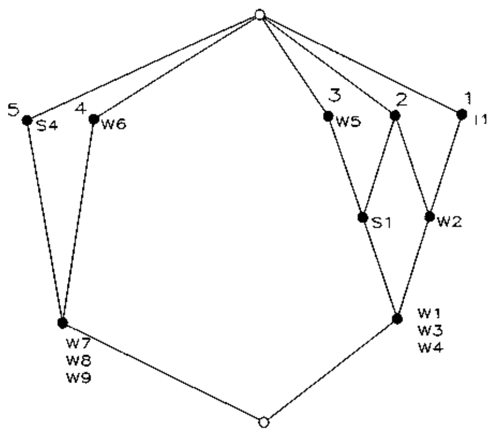

Figure 4. Galois Lattice of the Western Electric Cliques.

The cliques form two non-overlapping groups. Dependencies are bottom up. ***S1***, for example, never plays games unless ***W1, W3*** and ***W4*** are playing. And ***W5*** never plays unless those three and ***S1*** are involved.

So, one group contains ***W7, W8*** and ***W9*** as core members and ***W6*** and ***S4*** as peripheral. The other has a core of ***W1, W3*** and ***W4***, a semi-periphery of ***S1*** and ***W2***, and a periphery of ***W5*** and ***I1***.

This lattice-based structure corresponds closely both to the pattern described by the ethnographers and that revealed in the pictures produced by MDS and the SVD. All three methods apparently converge on the same specification of groups.

Recent research has increasingly embodied the idea that inter-actor links are not binary, but quantitative. Pairs of individuals are now viewed as linked by ties of varying in strength and each pair is assigned a number indicating the strength of their tie. This has led to the development of another approach. This is examined in the next section.

IMAGES BASED ON QUANTITATIVE ASSUMPTIONS

When network data come in the form of quantitative indexes of social proximities or distances, we would like to continue using MDS and SVD to uncover groups. As an example of quantitative network data, consider the observations of 13 male dolphins shown in Table 2. Records of the frequencies of pairs of dolphins observed swimming together were made over a period of time in a clear shallow bay.

	a	b	c	d	e	f	g	h	i	j	k	l	m
a	12	12	8	7	7	4	7	4	0	0	0	0	0
b	12	12	8	7	7	4	7	4	0	0	0	0	0
c	8	8	24	23	18	3	5	3	0	0	0	0	0
d	7	7	23	26	19	2	4	2	0	0	0	0	0
e	7	7	18	19	20	3	5	3	0	0	0	0	0
f	4	4	3	2	3	21	20	21	0	0	0	0	0
g	7	7	5	4	5	20	23	20	0	0	0	0	0
h	4	4	3	2	3	21	20	21	0	0	0	0	0
i	0	0	0	0	0	0	0	0	31	26	0	0	0
j	0	0	0	0	0	0	0	0	26	28	0	0	0
k	0	0	0	0	0	0	0	0	0	0	35	31	24
l	0	0	0	0	0	0	0	0	0	0	31	31	22
m	0	0	0	0	0	0	0	0	0	0	24	22	25

Table 2. Male Dolphins Observed Swimming Together (from Connor, Smolker and Richards, [2]).

When the data of Table 2 are entered into non-metric MDS, the two dimensional solution has a stress of .001. The resulting two-dimensional image is shown in Figure 5.

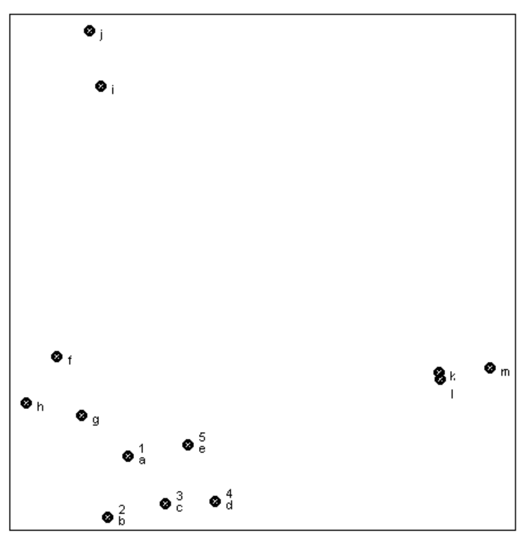

Figure 5. Two Dimensional MDS Representation of the Association among Dolphins.

When the principal components form of SVD was used with the data of Table 2, the first two axes were associated with 72% of the variance. The resulting image is shown in Figure 6.

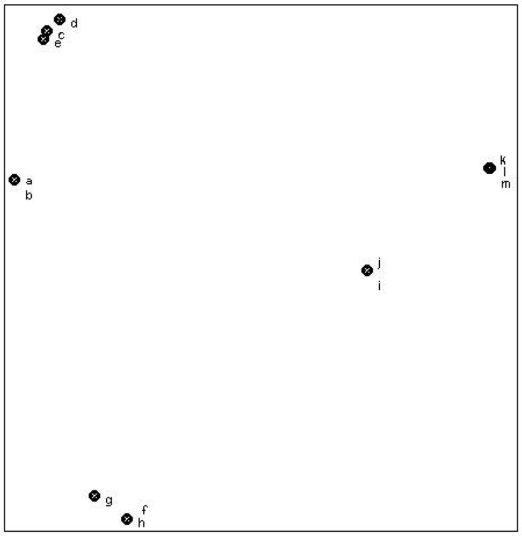

Figure 6. Two Dimensional SVD Representation of the Association among Dolphins.

Both figures appear to break the dolphins up into well-spaced subsets. The question is, do these subsets correspond to groups. To answer this question requires that we specify the structural properties of groups in terms of this kind of quantitative measure of social linkages.

Homans ([11], p. 84) suggested a definition of groups that was based on a quantitative conception of interpersonal links:

> If we say that individuals A, B, C, D, E, . . . form a group, this will mean that at least the following circumstances hold. Within a given period of time, A interacts more often with B, C, D, E, . . . than he does with M, N, L, O, P, . . . whom we choose to consider outsiders or members of other groups. B also interacts more with A, C, D, E, . . . than he does with outsiders, and so on for the other members of the group.

Sailer and Gaulin [24] made this idea a bit more precise when, in discussing groups of monkeys, they said:

> On the basis of co-occurrence, monkey ***m*** could be said to belong to group ***G*** if it spends more time with monkeys in group ***G*** than with other monkeys that are not.

This idea specifies clearly the conditions under which a collection of individuals can or cannot be considered to be a group. But it does not provide a procedure for uncovering such subsets. Given ***N*** individuals, it would be natural to examine all possible subsets and thus to reveal those that met the defining criterion. The problem with such an approach is that the number of subsets grows exponentially with ***N***. That means that whenever ***N*** is larger than some very small number, computation is not possible.

It turns out, however, that a simple genetic algorithm provides an effective way to search for the subsets that meet the group criterion (Freeman, [5]). In the case of the dolphin data, the algorithm finds three non-overlapping groups, one of which is divided further into three partially overlapping groups. The non-overlapping groups are {a,b,c,d,e,f,g,h},{i.j} and {k.l.m}. The first of these sets divides up into {a,b}, {c,d,e} and {f,g,h}which overlap to some degree.

The images displayed in both Figure 5 and Figure 6 arrange the points in such a way that they both reveal this group structure. In both cases, we can draw a convex polygon that contains all the members of each group. But the image produced by principal components analysis displayed in Figure 6 has an advantage. As Webster [29] showed, SVD images are oriented to rectangular X and Y coordinates; they place the points in such a way that the groups can all be bounded by rectangles. See Figure 7.

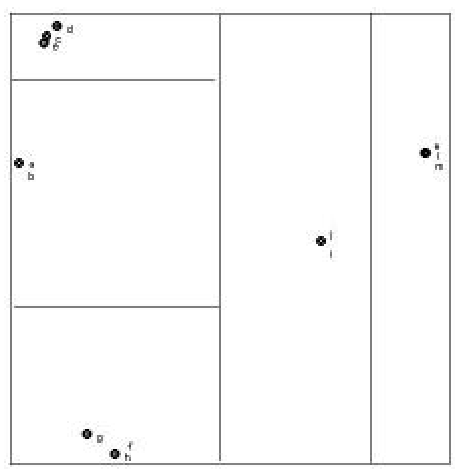

Figure 7. The Group Structure Contained in the Dolphin Data.

THREE DIMENSIONAL IMAGES

It should be remembered here that the first two axes of the SVD of the dolphin data accounted for 72% of the variance. But the first three axes account for more than 93% of the variance. It might be reasonable, then to display this pattern using three rather than two dimensions.

Three-dimensional figures are difficult to produce for the printed page. Klovdahl [13] used an early program, ORTEP, that had originally been developed for the display of molecules (Johnson, [12]). He adapted it in order to produce three-dimensional images of social networks.

ORTEP images are drawn in such a way that differences in the sizes and shapes of points and convergences in the lines depicting links are used to create the impression of perspective. As Figure 8 shows, a third dimension is at least crudely displayed.

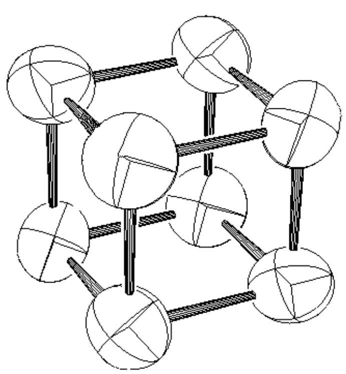

Figure 8. Apparent Perspective in an ORTEP Image.

More recently a large number of programs are available to facilitate producing apparent three-dimensional figures on computer monitors. When a monitor is used as a display device, the three-dimensional illusion can be much more convincing. There are, for example, a great number of screen-oriented programs that permit the creation of Virtual Reality Modeling Language (VRML) images (see Web 3D Consortium, [28]).

VRML images create the illusion of a third dimension by providing not only perspective, but also lighting, shading and color cues as well. In addition, they enhance the 3D illusion by permitting the viewer to "walk" around and through the image and thus to explore its details from a range of different perspectives.

I used VRML to illustrate these features in an earlier web publication [8]. There I took published data on the sexual links among the individuals in a gay group and constructed a VRML image (shown there as Figure 3). Most modern web browsers will permit their users to explore the details of that structure.

Other programs, like Xgobi from A.T.& T. Labs (see Swayne, Cook and Buja, [26]) and a great many statistics and data exploration programs can create the same kind of illusion of three dimensions simply by slowly rotating the image. In addition, many of these programs incorporate various "grand tour" options that permit viewers to explore the structural properties of images that contain more than three dimensions.

COLOR AND MOTION

Color is extremely important, not only in presenting images to others, but also in using them to gain new insights about structural properties of social network data. Unfortunately, in this black and white medium, I cannot illustrate the point. But I have often found it illuminating to produce an image (using MDS or SVD) and to color each point according to a series of properties of the individuals—age, sex, education and the like. If a color representing a property is spread out throughout the image, it is probably unrelated to the structure. But, to the degree that a property maps to a particular cluster in the image, it may provide a basis for, or be a consequence of the structural patterning displayed in the image.

In addition, animated images can provide powerful devices for uncovering structural processes. Like colors, animated images cannot be effectively displayed on the printed page. But also like colors, they are easy to display on a computer monitor. A particularly useful program for social network animation is called MOVIEMOL. It was produced by Ojamäe and Hermansson [21].

In another web publication [7] I have used MOVIEMOL to show some of the power of dynamic images to help to uncover important structural patterns in social network data. In that paper, Figures 21 and 22 illustrate how animation can help to reveal patterning that is otherwise difficult to find. And in another web publication [9], I have use a dynamic display to show how a weighted spring embedder can produce a meaningful arrangement of points representing interconnected individuals. That exercise can be viewed by anyone with a Java-equipped browser.

CONCLUSIONS

By way of summary, I would like to suggest that social network analysts have a number of available procedures for visualization that produce images that correspond quite closely to both ethnographic descriptions and formal definitions of social structure. These procedures, particularly MDS and SVD, are simple and quick and they arrange the points in such a way that their main structural properties are displayed.

In addition, visual displays of this sort often provide new structural insights. Three-dimensional images and those involving color and animation are particularly useful in this respect.

And finally, the aim of this essay was to present a "state of the art" review of visualization in social network analysis. When it was presented orally at the meeting, I used a computer to illustrate and expand many of the points I have tried to make here. Unfortunately, in print, I must often describe, rather than illustrate contemporary visual imagery. So, I encourage interested readers to use their computers and to explore more of the images—particularly those involving color, three dimensions and animation—described above. The relevant URLs are listed in the bibliography below.

ACKNOWLEDGEMENTS

This work was supported by MSH-Paris and the Maison Suger. Their generous support is gratefully acknowledged.

REFERENCES

[1] BOCK, R.D., AND HUSAIN, S.Z. (1952), "Factors of the tele: a preliminary report," *Sociometry*, 15, 206-219.

[2] CONNOR, R. C., SMOLKER, R. A., AND RICHARDS, A. F. (1992), "Dolphin alliances and coalitions," Coalitions and Alliances in Humans and Other Animals, eds: A.H. Harcourt, and F.B.M. deWaal, Oxford: Oxford University Press.

[3] COOLEY, C.H. (1909/1962), *Social Organization*, New York: Shocken Books.

[4] DURKHEIM, E. (1893/1933), *The Division of Labor in Society*, trans: G. Simpson, N.Y.: Free Press.

[5] FREEMAN, L.C. (1993), "Finding groups with a simple genetic algorithm," *Journal of Mathematical Sociology*, 17, 227-241.

[6] FREEMAN, L.C. (1996), "Cliques, Galois lattices, and the structure of human social groups," *Social Networks*, 18, 173-187.

[7] FREEMAN, L.C. Visualizing Social Networks (http://carnap.ss.uci.edu/vis.html).

[8] FREEMAN, L.C. Using Molecular Modeling Software in Social Network Analysis: A Practicum (http://eclectic.ss.uci.edu/~lin/chem.html).

[9] FREEMAN, L.C. Using Available Java Programs to Visualize Social Networks (http://carnap.ss.uci.edu/java/java.html).

[10] FREEMAN, L.C., AND WEBSTER, C.M. (1994), "Interpersonal proximity in social and cognitive space," *Social Cognition*, 12, 223-247.

[11] HOMANS, G.C. (1950), *The Human Group*, New York: Harcourt, Brace and Company.

[12] JOHNSON, C.K. (1965), *ORTEP Report ORNL-3794*, Oak Ridge, TN: Oak Ridge National Laboratory.

[13] KLOVDAHL, A.S. (1981), "A note on images of networks," *Social Networks*, 3, 197-214.

[14] KRUSKAL, J.B., AND WISH, M. (1978), *Multidimensional Scaling*, Beverly Hills, CA: Sage.

[15] LAUMANN, E.O., AND GUTTMAN, L. (1966), "The relative associational contiguity of occupations in an urban setting," *American Sociological Review*, 31, 169-178.

[16] LEVINE, J. (1979), "Joint-space analysis of "pick-any" data: Analysis of choices from an unconstrained set of alternatives," *Psychometrika*, 44, 85-92.

[17] LUCE, R.D., AND PERRY, A. (1949), "A method of matrix analysis of group structure," *Psychometrika*, 14, 95-116.

[18] MORENO, J.L. (1932), *Application of the Group Method to Classification*, New York: National Committee on Prisons and Prison Labor.

[19] MORENO, J.L. (1934), *Who Shall Survive?*, Washington, DC: Nervous and Mental Disease Publishing Company.

[20] MORENO, J.L. (1953), *Who Shall Survive?*, Beacon, N.Y.: Beacon House Inc.

[21] OJAMÄE, L., AND HERMANSSON, K., MOVIEMOL, (http://www.fos.su.se/moviemol.html).

[22] PROCTOR, C. (1953), "Informal social systems," *Turrialba*, eds: C.P. Loomis, J.O. Moralis, R.A. Clifford, and O.E. Leonard, Glencoe, IL: Free Press.

[23] ROETHLISBERGER, F.J., AND DICKSON, W.J. (1939), *Management and the Worker*, Cambridge, MA: Harvard University Press.

[24] SAILER, L.D., AND GAULIN, S.J.C. (1984), "Proximity, sociality and observation: the definition of social groups," *American Anthropologist*, 86, 91-98.

[25] SIMMEL, G. (1902), "The number of members as determining the sociological form of the group. I," American Journal of Sociology, 8, 1-46.

[26] SWAYNE, D.F., COOK, D., AND BUJA, A., Xgobi, (http://www.research.att.com/areas/stat/xgobi/).

[27] TÖNNIES, F. (1887/1955), *Gemeinschaft und Gesellschaft*, trans.:C. P. Loomis, London: Routledge and Paul .

[28] WEB 3D CONSORTIUM (1999), The VRML Repository (http://www.web3d.org/vrml/vrml.htm).

[29] WEBSTER, C.M. (1993), "Task-related and context-based constraints in observed and reported relational data," Ph.D. Thesis, University of California, Irvine.

[30] WELLER, S.C., AND ROMNEY, A.K. (1990), *Metric Scaling: Correspondence Analysis*, Beverly Hills, CA: Sage.