5

# TECHNIQUE FOR ANALYZING OVERLAPPING MEMBERSHIPS

*Phillip Bonacich*
UNIVERSITY OF CALIFORNIA, LOS ANGELES

Sociologists study the structure or pattern of relationships among individuals and among groups. The sharing of common members is an important relationship, as is the pattern of overlapping members. I will first discuss few instances of sociological concern with the pattern of overlapping memberships in order to clarify the problems that this chapter will solve.

The pattern of interlocking directorates among business organizations can give information about the power centralization in a society. Lieberson (1971) suggests that an analysis of the pattern of interlocking directorates among the largest business organizations sheds light on whether the power-elite view or the pluralist view of American society is the more accurate. Lieberson confines himself to a few selected facts (for instance, that the boards of the seven largest New York City banks in 1965 included officials from 51 of the largest 500 industrial companies). Although suggesting the value of a thorough analysis of interlocking directorates, Lieberson does not attempt to provide it.

Perrucci and Pilisuk (1970) examine the pattern of overlapping leadership among business and volunteer organizations in a community. They view power as residing particularly in individuals in key positions within the structure of organizations. Perrucci and Pilisuk (1970, p. 1044) state that "It is not the potency of the individual but the shape of the web (in which he is a node) which depicts the structure of enduring community power." The first part of their article explores the correlates of being a leader in many or a few community organizations, but, in the latter part of their work, they begin to explore the *pattern* of overlapping leadership in the community.

A failure to correct for the various sizes of the organizations is one defect in their analysis. The size of an organization affects the number of members it has in common with other organizations. The maximum overlap between two groups is limited by the size of the smaller group. Thus, a measure of overlap that is independent of the sizes of the organizations should be developed. At one point, Perrucci and Pilisuk rate groups according to the number of their members belonging to other groups (1970, p. 1047). This procedure is clearly inadequate.

A second suggestion considers the way in which overlaps among groups combine to form a centrality index for organizations. At one point, respondents were asked to rate the power of "34 community organizations having the greatest number of overlapping members" (1970, p. 1047). Presumably, just as individuals who belong to many organizations are more powerful than individuals who belong to few, organizations whose members belong to many other organizations should be more powerful than those who do not. However, not all groups are equal. Overlap with central groups (whose members in turn belong to many groups) contributes more to the centrality of a group than overlap with isolated groups. Overlap with a central group gives access to a greater number of powerful individuals, whereas overlap with relatively isolated groups gives access to relatively few. The number of overlaps with other groups alone does not involve such a weighting.

What, then, is needed to analyze the pattern of overlapping memberships? It is desirable to have a measure of overlap independent of group size. Then, an index of the centrality of groups in the structure of overlapping groups could be developed. The centrality of each group should be a function of the centrality of the groups with which it overlaps.

## *MEASURE OF OVERLAP*

We recognize four basic properties of a measure of overlap between groups that are implied by the discussion of centrality in the preceding section. First, the measure of overlap, like the proportion of common members in two groups, is zero if there is no overlap (in which case no communication or influence can be transmitted). Second, it is convenient if the maximum value is 1.00. Third, regardless of the size of the groups, the measure of overlap should have the same intermediate base line value whenever membership in the two groups is statistically independent. Independence of membership presumably indicates no special tendency for members of one group to avoid or pursue membership in the other group.

Fourth, and most important, the measure should be logically independent of group size. Larger groups necessarily overlap more, and we wish our measure of structural positional centrality to be uncontaminated by size. This independence of the measure from group size is given the following specific meaning. Let the matrix $N$ represent the overlap between two groups of size $n_{1.0}$ and $n_{0.1}$ in a system with $n$ members. (See Table 1.)

We insist that increasing or decreasing the size of a group in the simplest way, by multiplying all the elements in a row or a column by a positive number, does not affect the overlap relation between them. The matrix $N$ is "equivalent" to any matrix generated from it by multiplying a row or a column by a positive number. The overlap coefficient for the matrix $N$ is equal to the overlap coefficient for any matrix equivalent to it in this sense.

If all groups were of equal size, the proportion of members of one group who were in the other group satisfies the first three criteria and the fourth is irrelevant. However, when groups differ in size, the proportion of common members is affected by these differences in a number of ways. For example, the expected proportion of overlap is larger for pairs of large groups than for pairs of small groups. Our strategy is to alter matrix $N$ by multiplying the two rows and two

TABLE 1
Matrix $N$ Showing the Overlap Between Two Groups

| | | Group B: Member | Group B: Nonmember | |
|---|---|---|---|---|
| Group A | Member | $n_{11}$ | $n_{12}$ | $n_{1.0}$ |
| | Nonmember | $n_{21}$ | $n_{22}$ | $n_{2.0}$ |
| | | $n_{0.1}$ | $n_{0.2}$ | $n$ |

columns by four positive numbers so that both groups are made equal in size. This procedure is legitimate because multiplication of a row or a column does not change the matrix "essentially," by assumption four. The measure of overlap between two groups is the proportion of common members in this modified, yet equivalent, matrix.

Multiplying rows and columns by numbers is identical to pre- and postmultiplication of $N$ by diagonal matrices $P$ and $Q$. If $U = PNQ$, then by the fourth assumption, $U$ is "equivalent" to $N$.[1] Given $N$ and the desired row and column sums of $U$, the diagonal matrices $P$ and $Q$ can be determined. If we set $u_{1.0} = u_{0.1} = 1$, then $r = u_{11}$, which is the measure of overlap, is the proportion of members in each group also in the other group if the groups are made equal in size by using the equivalence relation defined by the fourth assumption.

The only ambiguity is the value of $u_{2.0} = u_{0.2}$. This value is the number of system members not in each group in the standardized overlap matrix. The simplest assumption appears to be that $u_{2.0} = u_{0.2} = u_{1.0} = u_{0.1} = 1$. In this case, $r$ has the following formula: (See Appendix.)

$$r = 0.5 \quad \text{if} \quad n_{11}n_{22} = n_{12}n_{21}$$

Otherwise,

$$r = (n_{11}n_{22} - \sqrt{n_{11}n_{22}n_{12}n_{21}})/(n_{11}n_{22} - n_{12}n_{21})$$

As an illustration, Table 2 shows the overlap relation in a certain high school, between membership in the school orchestra and membership in the academic honor society.

The matrix $U$ of standardized overlaps is uniquely determined by the values in Table 2. However, an infinity of sets of row and column multipliers will produce it. The reader can verify that matrix $U$ is produced (with slight rounding error) by multiplying the first row of Table 2 by 1.00, the second row by 0.0271, the first column by 0.298, and the second column by 0.0126.

TABLE 2
Overlap Between Orchestra and Honor Society

| | | Honor Society | |
|---|---|---|---|
| | | Member | Nonmember |
| Orchestra | Member | 2 | 32 |
| | Nonmember | 50 | 1735 |

[1] It is easy to show that this relation is also an equivalence relation in the usual mathematical sense. It is reflexive, symmetric, and transitive.

TABLE 3
Standardized Overlap Between Orchestra and Honor Society

| | | Honor Society | |
|---|---|---|---|
| | | Member | Nonmember |
| Orchestra | Member | 0.596 | 0.404 |
| | Nonmember | 0.404 | 0.596 |

Each row and each column of Table 3 sums to 1.00. If the process of increasing or decreasing the size of a group by multiplying a row or a column of the original matrix does not alter the relation between the groups, then the degree of overlap between the two unequal groups in Table 2 is equivalent to the overlap between the two equal groups in Table 3, where almost 60 per cent of each group overlaps with the other.

## *INDEX OF CENTRALITY*

The matrix $R$ of overlap coefficients $r$ is much like a correlation matrix. It is symmetric. None of its values exceeds one. The closeness to a correlation matrix suggests factor analysis. For simplicity of exposition it is assumed that all groups are indirectly or directly connected. The eigenvector of the largest positive eigenvalue (the first factor) contains all positive (or all negative) values, and it is the only vector with this property. (See Bonacich, 1971, for proofs and a general discussion of the technique applied to any symmetric structure.) Make all values positive, and standardize the vector so that its length is its eigenvalue. This standardized eigenvector $S$ of the largest eigenvalue has the following properties:

(1) The outer product of the column vector $S$ with its transpose $SS'$ is the least-squared-error approximation to the matrix of standardized overlaps $R$, just as the first principal components factor approximates the correlation matrix. Specifically, let $D$ be the sum of the squared errors in approximating $R$ by the product of a column vector $S$ and its transpose.

$$D = \sum_i \sum_j (SS' - R)_{ij}^2 = \sum_i \sum_j (S_i S_j - r_{ij})^2$$

By setting $\partial D/\partial S = 0$, it can be shown that the best $S$ is an eigenvector of the largest eigenvalue, standardized so that its length is its eigenvalue. This condition is true whenever $R$ is symmetric. It does not matter whether $R$ is a correlation matrix or a matrix of standardized overlaps. The $i$th element of $S$ is the joining potential of members of group $i$ in the sense that $S_i S_j$ tends to be close to $r_{ij}$ for every other group $j$. Since $S_i S_j$ should be close to $r_{ij}$, one interpretation of $S_i^2$ is that it is the expected proportion of overlap between two groups with centrality indices of $S_i$.

(2) This "first factor" is also the vector of centrality scores we are looking for. Let the centrality of each group be a weighted combination of its overlaps with other groups, each overlap being weighted by the centrality of that group.

$$S_i = r_{i1}S_1 + \cdots + r_{in}S_n$$

This set of equations is likely to have no nonzero solution. However, a small modification, a multiplication of values on the left by a constant, ensures a solution.

$$\lambda S_i = r_{i1}S_1 + \cdots + r_{in}S_n$$

The solution to these equations is mathematically identical to the factor analytic solution. The set of weights $S_i$ is an eigenvector of the eigenvalue $\lambda$. Thus, $S$ is the vector of centrality scores we were looking for. For each group $i$, $S_i$ (actually $\lambda S_i$) is the sum of the overlaps of group $i$, where each overlap with another group is weighted by the centrality of that group. The ratio $S_i/S_j$ is the ratio of the two groups' contributions to the centrality of the other groups.

(3) The differences between this "factor" approach and Hubbell's (1965) "input/output" approach are discussed at length in another paper (Bonacich, 1971). In brief, whereas in the factor approach $S$ is a solution to the homogeneous equations $(R - \lambda I)S = 0$, Hubbell's status scores are a solution to the nonhomogeneous equations $(R - I)S = E$, where $E$ is a vector of ones. Each element of Hubbell's $S$ is the sum of all the paths emanating from a given vertex in the structure described by $R$; $S = \sum^{\infty} R^k$. However, Hubbell's $S$ vector is not a factor of the $R$ matrix nor is it a centrality measure in the sense used here.

## *ILLUSTRATION*

To illustrate the computation of centrality indices, I selected ten of the largest athletic activities from a convenient high school yearbook. Table 4 gives the amount of overlap among them. There were 929 male students who were potential members. The size of each team is given by the entries in the main diagonal. Table 5 gives the standardized overlaps for all pairs of teams.

TABLE 4
Overlaps Among Ten Athletic Activities in a High School

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 Varsity football | 49 | 3 | 1 | 1 | 4 | 12 | 4 | 1 | 1 | 1 |
| 2 Varsity basketball | 3 | 16 | 0 | 3 | 0 | 3 | 4 | 1 | 1 | 0 |
| 3 "B" basketball | 1 | 0 | 10 | 1 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 Varsity water polo | 1 | 3 | 1 | 16 | 0 | 1 | 0 | 2 | 4 | 1 |
| 5 Wrestling | 4 | 0 | 0 | 0 | 20 | 2 | 1 | 0 | 0 | 0 |
| 6 Varsity track | 12 | 3 | 0 | 1 | 2 | 28 | 1 | 0 | 0 | 0 |
| 7 Varsity baseball | 4 | 4 | 2 | 0 | 1 | 1 | 22 | 0 | 0 | 0 |
| 8 Golf | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 8 | 0 | 0 |
| 9 Varsity swimming | 1 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 12 | 0 |
| 10 "B" swimming | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 6 |

If this study were complete, centrality indices for these ten activities would be computed from a matrix of overlaps between all 65 clubs and activities existing in the high school. In this illustration, centrality indices are computed solely on the basis of the overlaps in Table 5 so that the reader can get a sense of the relationship between the raw data and the results.

Table 6 gives the centrality indices for the ten teams. The table suggests that participants in football, basketball, and water polo tended to be active in other (central) sports, whereas other sports tended to be more specialized. Participation in the latter sports was not strongly associated with participation in other sports, especially in the central sports. This aspect of a sport, namely, whether its participants are generally active in other sports, could be associated with its status in the school. Football in particular attracted athletes.

Table 7 is the product $SS'$, the outer product of the vector of centrality scores and its transpose. Table 7 shows the "expected" amount of overlap between the teams. Therefore, of special interest are

TABLE 5
Standardized Overlaps Among Ten Athletic Activities in a High School

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.68 | 0.59 | 0.52 | 0.69 | 0.81 | 0.67 | 0.62 | 0.56 | 0.66 |
| 2 | 0.68 | 0.00 | 0.00 | 0.80 | 0.00 | 0.74 | 0.80 | 0.75 | 0.70 | 0.00 |
| 3 | 0.59 | 0.00 | 0.00 | 0.72 | 0.00 | 0.00 | 0.77 | 0.00 | 0.00 | 0.00 |
| 4 | 0.52 | 0.80 | 0.72 | 0.00 | 0.00 | 0.60 | 0.00 | 0.82 | 0.86 | 0.78 |
| 5 | 0.69 | 0.00 | 0.00 | 0.00 | 0.00 | 0.66 | 0.59 | 0.00 | 0.00 | 0.00 |
| 6 | 0.81 | 0.74 | 0.00 | 0.60 | 0.66 | 0.00 | 0.55 | 0.00 | 0.00 | 0.00 |
| 7 | 0.67 | 0.80 | 0.77 | 0.00 | 0.59 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 |
| 8 | 0.62 | 0.75 | 0.00 | 0.82 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 9 | 0.56 | 0.70 | 0.00 | 0.86 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 10 | 0.66 | 0.00 | 0.00 | 0.78 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

TABLE 6
Centrality Scores for Ten Athletic Activities

| | |
|---|---|
| 1. Varsity football | 0.86 |
| 2. Varsity basketball | 0.79 |
| 3. "B" basketball | 0.42 |
| 4. Varsity water polo | 0.78 |
| 5. Wrestling | 0.38 |
| 6. Varsity track | 0.64 |
| 7. Varsity baseball | 0.59 |
| 8. Golf | 0.49 |
| 9. Varsity swimming | 0.47 |
| 10. "B" swimming | 0.32 |

pairs for which the actual overlap, in Table 5, is markedly different from the expected overlap. For example, the actual overlap between water polo and varsity swimming, 0.86, is much greater than the expected overlap, 0.37. These deviations indicate special affinities or dissimilarities between the sports. The difference $r_{ij} - 0.5$ shows whether there is more or less overlap between groups $i$ and $j$ than occurs if membership were statistically independent, but the general tendencies of members of these groups to belong to other groups is not controlled for. The difference $r_{ij} - S_iS_j$, however, gives the special attractions or repulsions between pairs of groups over and above their members' tendencies to be joiners or loners.

TABLE 7
Expected Overlap Among Ten Athletic Activities

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.68 | 0.36 | 0.68 | 0.33 | 0.56 | 0.51 | 0.42 | 0.41 | 0.28 |
| 2 | 0.68 | 0.00 | 0.33 | 0.61 | 0.30 | 0.51 | 0.46 | 0.38 | 0.37 | 0.26 |
| 3 | 0.36 | 0.33 | 0.00 | 0.33 | 0.16 | 0.27 | 0.25 | 0.20 | 0.20 | 0.14 |
| 4 | 0.68 | 0.61 | 0.33 | 0.00 | 0.30 | 0.50 | 0.46 | 0.38 | 0.37 | 0.25 |
| 5 | 0.33 | 0.30 | 0.16 | 0.30 | 0.00 | 0.24 | 0.22 | 0.18 | 0.18 | 0.12 |
| 6 | 0.56 | 0.51 | 0.27 | 0.50 | 0.24 | 0.00 | 0.38 | 0.31 | 0.30 | 0.21 |
| 7 | 0.51 | 0.46 | 0.25 | 0.46 | 0.22 | 0.38 | 0.00 | 0.29 | 0.28 | 0.19 |
| 8 | 0.42 | 0.38 | 0.20 | 0.38 | 0.18 | 0.31 | 0.29 | 0.00 | 0.22 | 0.15 |
| 9 | 0.41 | 0.37 | 0.20 | 0.37 | 0.18 | 0.30 | 0.28 | 0.22 | 0.00 | 0.10 |
| 10 | 0.28 | 0.26 | 0.14 | 0.25 | 0.12 | 0.21 | 0.19 | 0.15 | 0.10 | 0.00 |

## *CONCLUSIONS*

We have described a method for analyzing the pattern of overlapping memberships among groups. The key objective is to develop a measure of structural centrality in the pattern of overlapping memberships such that the centrality of a group is a function of the centrality of the groups with which it overlaps. First, we developed a measure of the overlap independent of group size. This measure is the proportion of members common to two groups after the sizes of the groups are equalized by using an equivalence relation described in the text. We showed that a "factoring" of the matrix of standardized overlaps yields the desired measure of centrality.

The structure or pattern of overlapping memberships is an essential sociological phenomenon. Possible applications of the technique, based on current interests of sociologists, include the study of interlocking directorates of businesses and overlapping memberships or leaderships among community organizations. In this chapter, the overlap among some activities in a high school is illustratively examined. The development of the technique suggests new applications. In studying the spread of rumor or other forms of information, centrality could be related to how fast information spreads from one group to other groups. The technique described in this chapter can be used whenever membership lists exist and it is hoped that this technique will enable researchers to use more effectively these often easily available data.

## *APPENDIX*

$N$ is the two-by-two matrix of the overlap in membership between two groups. By multiplying the rows and columns of $N$ by positive numbers, we wish to produce a matrix $U$ with the following properties:

$$U_{11} + U_{12} = 1$$
$$U_{21} + U_{22} = 1$$
$$U_{11} + U_{21} = 1$$

A little thought shows that multiplying rows and columns does not alter the following ratio:

$$U_{11}U_{22}/U_{12}U_{21} = n_{11}n_{22}/n_{21}n_{12}$$

Solving the first three equations in terms of $U_{11}$ and substituting into the fourth equation, we find that:

$$U_{11}^2(n_{11}n_{22} - n_{12}n_{21}) - 2U_{11}n_{11}n_{22} + n_{11}n_{22} = 0 \qquad (1)$$

If $n_{11}n_{22} = n_{21}n_{12}$, then (1) is linear and $U_{11} = 0.5$. Otherwise, $U_{11}$ has one root, less than or equal to 1.00

$$U_{11} = (n_{11}n_{22} - \sqrt{n_{11}n_{12}n_{21}n_{22}})/(n_{11}n_{22} - n_{21}n_{12})$$

A note of caution: For two exhaustive but not mutually exclusive groups, $r = U_{11} = 0$. This is the only instance I have discovered in which the coefficient misbehaves.

## REFERENCES

BONACICH, P.

1971 "Factoring and weighting approaches to status scores and clique identification." *Journal of Mathematical Sociology,* in press.

HUBBELL, C. H.

1965 "An input-output approach to clique identification." *Sociometry* 28 (December):377-399.

LIEBERSON, S.

1971 "An empirical study of military-industrial linkages." *American Journal of Sociology* 76 (January):562–584.

PERRUCCI, R. AND PILISUK, M.

1970 "Leaders and ruling elites: The interorganizational bases of community power." *American Sociological Review* 35 (December):1040–1057.