Brief Communications

A Method for Studying Intercorrelated Circulation Patterns in Library Systems

Charles H. Smith
Department of Geography, University of Illinois, Urbana, IL 61801

A means for studying intercorrelated circulation patterns among the branches of a library system is suggested. The focus is on the improvement of availability of library resources to the user. Under conditions in which the analyst is interested in generating a picture of system-wide utilization of library resources by an "average user" representative of the sum community, multidimensional scaling procedures might be applied to obtain relevant information. Following a brief description of multidimensional scaling, several kinds of specific analyses types concerning the allocation of library resources are discussed.

As a student of a highly interdisciplinary field (animal geography), it is especially important to me that a library system be organized in a fashion minimizing the time I spend locating and reaching materials shelved in an array of branch collections. This is certainly a problem that most users experience to one degree or another, and one that information systems analysts have been closely attuned to for some time. Two basic elements of the problem seem evident. The first concerns how to determine "what is located where," a difficulty whose remedy can be found in the employ of computerized shelf search systems and the like. The second, however, is a less tractable matter, concerning how to organize materials themselves within the sum library collection.

This second element involves system level organizational difficulties that are more resistant to minimization, constrained as solutions are by a variety of budgetary, operational, and user-accessibility considerations. Here I should like to suggest a route that might be followed to gather and analyze several types of information relevant to the last of these three considerations.

Our concerns are with the multicollection library system whose purpose it is to meet the needs of a number of special interest populations. This will most commonly mean a university community, one within which user needs are fixed by both instructional and research considerations. In all cases, however, the amount of time spent in retrieval of information will be minimized by a library system physically organized in a fashion compatible with the needs of the community as a whole.

The problem is how to reduce users' time expenditures through the efficient organization of materials in remote and/or central collections. This translates in a system sense into minimizing the number and length of trips among branch libraries for a hypothetical "average user." Of course, the identification of this notion as a goal is nothing new, but a straightforward approach toward its solution—even in user-services terms only—is something else again. Even when data on common utilization patterns are available, the matter of equitable and interpretable analysis of these data presents itself. The main difficulty is that we wish to identify not only which populations within the university are frequenting which branches, but correlations of such utilization among populations. Moreover, since we are trying to determine how all such correlations relate to one another to form a picture of how the system as a whole is organized, it becomes necessary to identify more than just isolated instances of correlated utilization of particular pairs of branches by particular subpopulations: *All* degrees of correlated use must be explicitly related to one another within the context of a single solution. Because we are considering discrete elements (branch collections) within a deterministically set out system (the sum collection), most multivariate statistical procedures are inappropriate: The object is not to predict varying circulation rates from user characteristics, but instead to identify degree of correlation of utilization by those users. Otherwise stated, we wish to investigate how the habits of our subjects (users) specify *relationships* among objects (collections), and not simply how one or the other specifies the second.

Received December 1, 1982; revised January 17, 1983; accepted March 7, 1983

One approach that seems as if it could be feasible involves the application of what are known as "multidimensional scaling" (MDS) techniques. This family of methods, closely related mathematically to other multivariate approaches, has found most of its use in psychology and sociology. Fundamentally, what MDS procedures accomplish is to reduce a symmetrical matrix of distances between pairs of objects to a single geometric representation of that set of distances. This is actualized by transforming the distances among the objects under consideration into a "map" of their relative locations within an n-dimensional coordinate field. In short, MDS provides a solution to the situation in which the distances among the objects under consideration are known *a priori*, and we wish to determine their relative locations in space. This is the inverse of the more common problem of starting with all locational coordinates given and all interlocation distances unknown.

A simple example will serve to clarify this operation. If a table of airline distances among 30 cities in the United States were fed into a typical MDS package such as KYST, the solution could consist of a list of 30 locations expressed as two-dimensional coordinates. If requested to do so, the KYST package would then produce a plot of these results in the form of a scattergram (in which each axis is literally a dimension). This two-dimensional "map" of coordinate locations would very closely correspond to the relative positions in the real world of all the cities represented. Thus, a multidimensional scaling operation has transformed the original distance matrix into a spatial representation of this information.

The more interesting applications of the technique come under less trivial conditions, of course. Three, four, five, or even more dimensions may be necessary to absorb the best part of the variation inherent in the initial distances matrix constructed. Moreover, the "distances" between the objects under consideration may be defined in a number of fashions representing, for example, the objects' degree of similarity or dissimilarity or how often one object is confused with the other or preferred over the other [1]. Such comparisons are all, in effect, relative distance measures. An example of the scaling of such subjectively assessed distances is provided in Figure 1. Another application arises under the "individual differences" method in which, for example, some number of subjects are asked to rate their relative degrees of preference for one object over another for each pairing of some number of objects. In this situation, the output configuration (= n-dimensional set of coordinate locations) will represent the "mean state" of ratings for that group of subjects; another way of putting this is to say that the set of ratings configurations associated with all the subjects has been collapsed into one such configuration representative of an "average subject."

Further discussion of the technique itself is not feasible here. The interested reader, however, will discover that literature on the subject is voluminous. (For example, see refs. 2–4.)

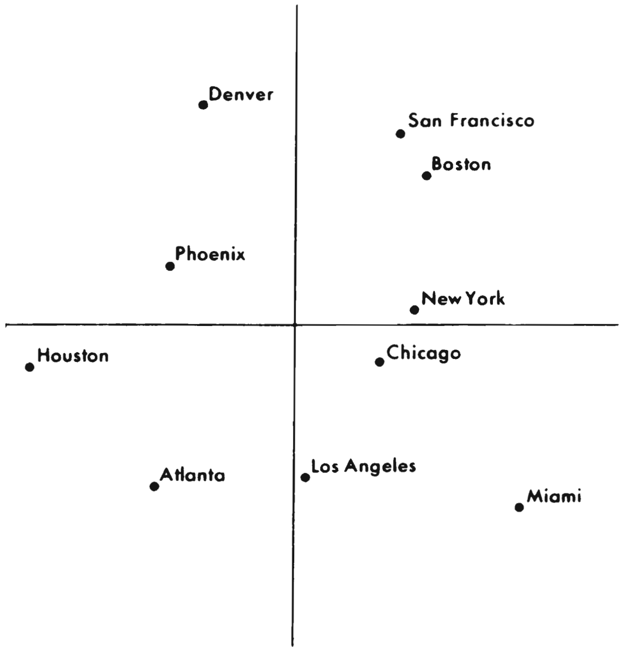

FIG. 1. Plot of a hypothetical two-dimensional scaling of some individual's perception of the relative similarities among ten cities. These results indicate this person views Boston and San Francisco and New York and Chicago as similar cities but, for example, Boston and Houston as very different cities. The reasons behind such views are not initially obvious here; perhaps, for example, this person associates comparable levels of cultural activity with Boston and San Francisco, comparable levels of economic activity with New York and Chicago, and with Boston and Houston no comparable levels of anything.

MDS could be employed in the present context of circulation patterns analysis in a number of fashions. For example, let us suppose that we have collected data for a very large number of users that allow us to determine all branch collections from which any given user has checked out books over some particular time span. A single input matrix of similarities ("distances") among branches could be compiled by simply tallying how many total individuals borrowed from each pairing of branch libraries over that span. Once the tally were completed, the resulting matrix could then be input to an MDS package such as KYST or TORSCA. A "best fit" solution to the system-wide set of interrelationships would then be generated via an iterative procedure. This solution would be output as a configuration of locations set in an n-dimensional space (the dimensionality of the solution is specified by the analyst). The visual representation of the results in two dimensions would be similar to the example given in Figure 1, with the difference that the located points would correspond to library branches instead of cities. This approach is convenient for its ability to accommodate a very large sample size.

Another way to approach the problem would start with the construction of a "profile matrix." Here, columns would represent branches, rows users, and values within the matrix the number of items checked out from each branch collection over some time period. If these data or something like them were unavailable, a survey study could provide similar information by obtaining from a sample of users an estimation of the number of trips out of 20 they would on the average make to each of the branch libraries listed. In either case, the initial profile matrix could then be transformed into a symmetrical correlation matrix and analyzed.

It must be understood here that in all the situations discussed above, the "distances" between branches portrayed in the output configurations represent *scaled* distances. These have been developed to delineate common utilization patterns of the "average user," and may well have little relation to relative real world geographic locations of the branches in question. How might such information be used? For one thing, it is quite apparent that from a user's perspective, branches that appear very close to one another in the output configuration might be consolidated. In this view, there seems little reason in maintaining two separate branch libraries that serve very nearly the same clientele (and that do *not* serve a large—and similar—proportion of the overall community as well). On the other hand, consolidation of branches appearing far apart from one another in the configuration would serve little purpose, since the results would involve bringing together collections serving almost entirely different clienteles.

Centralization of resources could also be studied by examining the overall output configuration. Branches appearing closer to the center of the configuration could be interpreted as those whose clientele cut across a wider spectrum of the community than those appearing nearer the periphery. The former might ideally be co-located in a central access main collection, whereas the latter would be more ideally left in more remote, perhaps departmental, locations. These points and related ones made in the last paragraph are illustrated in Figure 2.

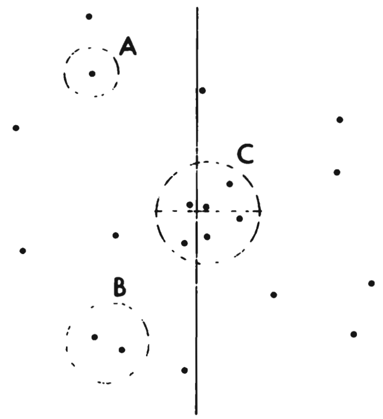

FIG. 2. A hypothetical two-dimensional configuration of scaled similarities among the twenty branch collections of a library system. (A) A typical remote collection serving a relatively specialized clientele. (B) Two collections that might logically be consolidated at a remote location. (C) Several collections whose degree of utilization by the community as a whole might merit their being located in a single central access facility.

Another application could involve minimizing disservice accompanying space utilization imbalances. On the whole, the least disservice would be entailed when books removed from overcrowded locations were relocated in roomier branches with whose output configuration location the original sites were most highly correlated (again, because closeness of location in the configuration indicates that these branches have much the same clientele).

An entirely different kind of user service problem could also be investigated through this general approach. Departmental relocation within the university system as a result of new construction or consolidation often generates library resource accessibility imbalances. The likelihood of creating such inconveniences might be lessened were user patterns within a potentially victimized group surveyed and predetermined in one of the manners already discussed. If the "average user" configuration generated suggested great utilization by that group of branches geographically remote from one another (and the location the group was being moved to), it might be argued that a new departmental remote collection should be created. Conversely, relocation might actually improve overall accessibility to other frequented libraries to the extent of justifying dissolving an already existing facility.

A final example of the possible use of MDS techniques in aiding the improvement of user service concerns adjustment of the system used to group subject areas in a collection. This could be accomplished in at least two ways: (1) by studying users' correlated borrowing of materials themselves (that is, subject categories such as "microbiology" and "biochemistry" would constitute the "objects" compared in the similarities matrix); and (2) by investigating the correlated trip behavior necessary to sustain users' correlated borrowing of materials. In the second case, the object would be to minimize users' necessary travel by physically relocating near to one another in a collection those subject categories determined to be most similar through prior analysis.

A number of other applications can also be envisioned, but the preceding exemplifies well enough the general range of possible types of information that could be gleaned from the application of MDS techniques. The data processing to obtain this information is not in itself difficult. Most of the main MDS packages are very easy to use [5–7]. It should be carefully noted, however, that proper application of the techniques is a complicated matter going far beyond the outline presented here and that study plans should be developed under professional guidance.

In sum, MDS would seem to offer an inroads to the study of the very complex problem of physical resource location optimization. Not only is it capable of sorting out patterned correlations among elements of a totally specified system, but it also maintains the individual identity of each element (here, branches of a library system, or subject collections) studied within the system in the results it produces. As both are central constraints on the analysis of circulation patterns in library systems, there may be a future in their application there.

References

1. Shepard, R. N. "A Taxonomy of Some Principal Types of Data and of Multidimensional Methods for their Analysis." In: R. N. Shepard, A. K. Romney, and S. Nerlove, Eds. *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*. New York: Seminar Press; 1972: Vol. 1, 21-47.
2. Beals, R.; Krantz, D. H.; Tversky, A. "Foundations of multidimensional scaling." *Psychological Review*. 75:127-142; 1968.
3. Kruskal, J. B.; Wish, M. *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications; 1978.
4. Shepard, R. N. "Representation of structure in similarity data: problems and prospects." *Psychometrika*. 39:373-421; 1974.
5. Chang, J. J.; Carroll, J. D. "How to Use INDSCAL, A Computer Program for Canonical Decomposition of N-way Tables and Individual Differences in Multidimensional Scaling." Bell Telephone Laboratories; 1972 (unpublished).
6. Kruskal, J. B.; Young, F. W.; Seery, J. B. "How to Use KYST, A Very Flexible Program to do Multidimensional Scaling and Unfolding." Bell Telephone Laboratories; 1973 (unpublished).
7. Pruzhansky, S. "How to Use SINDSCAL, A Computer Program for Individual Differences in Multidimensional Scaling." Bell Telephone Laboratories; 1975 (unpublished).

Applications of Multidimensional Scaling: Comment on "A Method for Studying Intercorrelated Circulation Patterns in Library Systems"

William E. McGrath
School of Information and Library Studies, State University of New York at Buffalo, Buffalo, NY 14260

Applications of multidimensional scaling to library circulation, with implications for management decisions in the areas of collection building, housing of collections, allocation of funds, and on-line retrieval are discussed. As a data-descriptive, hypothesis-generating technique, implications for theory building, hypothesis testing, prediction, and explanation are also discussed. Cautions concerning the interpretation of clusters, dimensions, the form of input matrices, and care in the definition of objects submitted to scaling are mentioned.

Multidimensional scaling (MDS); generally considered to be data-descriptive and hypothesis-generating, is a fascinating approach to discovery of pattern. It is not intended as an alternative to hypothesis-testing or predictive techniques such as multiple regression—although some theoreticians have argued that it can be.

As a student of animal geography, Smith [1] is undoubtedly aware of the extensive use of MDS in the analysis of animal populations and ecological settings. The book by Joel E. Cohen is a good example [2]. If Smith is not a student of information science, however, he may not be aware of several applications of MDS to the problem he describes.

One application to circulation in a university library can be found in my two articles in which the correlation measure was the number of books checked out by students in pairs of subject categories, the objects being compared [3,4]. In both these articles, the subject categories cluster much as one might expect—in three dimensions, though two might suffice. In the first article, I suggest how the results might be used in library acquisitions. In the second, I suggest that the results might be used to defend the housing of similar subject areas together (basically the same argument as Smith's), assignment of subject area bibliographers, allocation of funds, and online retrieval.

Another application of MDS is my work at OCLC, Inc., in which I analyzed the similarity of library holdings in large networks [5–7]. In these articles, I suggest that discovered dimensions and clusters be used in forming network policy for interlibrary loan traffic, optimal storage of bibliographic data, and other applications.

The cocitation studies of Henry Small and Belver Griffith should not be overlooked in this discussion [8].* Their work, though employing algorithms other than those mentioned by Smith and the ones I have used, is highly innovative and has major implications for theoretical and practical application in information science.

To my knowledge MDS has not specifically been applied to branch library circulation. Co-use of branch library pairs as a measure of similarity, as Smith suggests, seems reasonable, but a sufficiently large sample to establish variability may be difficult to obtain. As the number of objects being compared increases, the number of pairs increases rapidly. Characteristic of these large matrices, is a large number of zero cells, in which there is no co-occurrence. Still, the MDS approach to branch library analysis would be most useful in a system with many branches, say 30 or more, and least useful in systems with a handful of branches, unless something other than branches were being compared, say subject areas within branches. Online circulation systems would certainly facilitate data collection in this situation.

Smith argues that the intent of library use studies "is not to predict varying circulation rates from user characteristics, but instead to identify degree of correlation of utilization by those users." I would argue that nothing is wrong with using predictive techniques if the research question is appropriately addressed by them. If these techniques can also help to explain a dependent variable, or a cluster, or a dimension, then the theoretical underpinnings of our discipline is thus strengthened. I would thus argue that MDS has great potential for contributing to this theory, but the results of MDS—specifically, the configuration of objects in n-dimensions, i.e., the clusters of objects and their proximities thus derived—do not, in themselves, have predictive or explanatory power. They are patterns perhaps never before seen or described—like the braids and spokes in the rings of Saturn. They explain nothing, but cry out for explanation. Discovered patterns

*Subsequent papers by Small, Griffith, and colleagues have made extensive use of multidimensional scaling in mapping scientific specialties.

Received January 10, 1983; accepted March 7, 1983

can, of course, provide insights useful in decision making such as in optimizing location of library materials.

I generally have found users of MDS, myself among them, quite enthusiastic about its potential. That enthusiasm should be tempered with some caution. In my experience, the cautions expressed by Kruskal, Young, Shepard, and others are well worth heeding. For example, two dimensions are usually sufficient to accommodate similarities in matrices of up to 30 or 40 objects. Three dimensions may be found in larger matrices, whereas four or more dimensions may be simply noise. Though more dimensions can always be found in data, one should always employ criteria such as the best-fit measure STRESS (Kruskal calls it "badness of fit") and the familiar R-square in multiple correlation to determine the optimum number of dimensions. In my work, I have analyzed up to 100 objects and still cannot justify recognition of more than three dimensions, though the fourth dimension does increase R-square slightly.

Nonsymmetrical, rectangular matrices, in which the rows and columns are different objects (such as the branches and users proposed by Smith) are not easily handled by KYST, MD-SCAL, TORSCA, and other packages. Nonsymmetric matrices, particularly Coombs' unfolding, or conditional matrix model generate bizarre results with these packages. Even the superior package ALSCAL may generate uninterpretable results [9]. Smith correctly points out that nonsymmetric matrices should be symmetricized.

Shepard warns that MDS by itself is insufficient to exploit all the information in data, urging that cluster analysis and tree fitting also be used [10]. I suggest further that, after the MDS configurations are obtained, explanatory techniques such as multiple regression and discriminant analysis be used to explain why objects cluster as they do.

Lastly, I would urge the analyst to take great care in describing exactly what is being compared or correlated and in describing how the data were obtained. In the context of this discussion, for example, it is easy to confuse libraries, branches, subject areas, and persons. In the standard symmetric model, persons would be compared to persons, branches to branches, and so on. So little is known about cluster patterns in our field and so much can be done with this model (both classic metric and the newer nonmetric model based on ordinal data), that use of other models (nonsymmetric, conditional or unfolding, individual differences, 3-way, etc.) can wait respectfully for more complex problems.

Smith's ideas are welcome. I would hope that now he and other MDScalers will proceed with the analysis.

References

1. Smith, C. H. "A Method for Studying Intercorrelated Circulation Patterns in Library Systems." *Journal of the American Society for Information Science*. 34:(2):000–000; 1984.
2. Cohen, J. E. *Food Webs and Niche Spaces*. Princeton, NJ: Princeton University Press; 1978.
3. McGrath, W. E. "The Circulation Uncertainty Principle and the Cosmology of Collection Development." In: P. Spyers-Duran, and T. Mann, Jr., Eds. *Shaping Library Collections for the Eighties. Fourth International Conference on Approval Plans and Collection Development. Milwaukee, WI, 1980*. Phoenix, AZ: Oryx Press; 1980.
4. McGrath, W. E. "Multidimensional mapping of book circulation in a university library." *College and Research Libraries*. 44:103–115; March 1983.
5. McGrath, W. E. "Multidimensional Map of Library Similarities." *Proceedings of the 43rd Annual Meeting of the American Society for Information Science, Anaheim, CA, 1980*. Washington, DC: ASIS; 1980.
6. McGrath, W. E. "Implications for Cooperative Collection Development in a Random Group of Academic Libraries; or, Beyond Overlap." In: Association of College and Research Libraries, Eds.; *Options for the Eighties*. Greenwich, CT: JAI Press; 1982.
7. McGrath, W. E.; Hickey, T. B. *Research Report Prepared for OCLC on Multidimensional Mapping of Libraries Based on Shared Holdings in the OCLC Online Union Catalog*. Dublin, OH: OCLC Office of Research; 1983; OCLC/OPR/RR-83/5.
8. Small, H. G.; Griffith, B. C. "The structure of scientific literature I: Identifying and graphing specialties." *Science Studies*. 4:17–40; 1974.
9. Young, F. W.; Lewyckyj, R. *ALSCAL-4 User's Guide; a Guide for Users of ALSCAL-4: a Nonmetric Multidimensional Scaling and Unfolding Program with Several Individual Differences Options*. Carrboro, NC; Data Analysis and Theory Associates; 1979.
10. Shepard, R. N. "Multidimensional scaling, tree-fitting and clustering." *Science*. 210:290–298; 1980.