Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 51, No. 3 (2001) 563–571

Utilizing spatial information systems for non-spatial-data analysis

L. JOHN OLD

School of Library and Information Science, Indiana University, Bloomington, Indiana (USA)

Recent advances in the power and capabilities of personal computers have brought the algorithms and representational methods of Geographic Information Systems (GIS) to the desktop. Information that has relationships between elements may be represented spatially, especially if some distance metric can be brought to bear. This paper discusses the use of spatial methods for the display of non-Geographic data.

Introduction

Visualization of traditionally non-visual information (statistics, lists, tables and the like) is a growing art in such fields as data mining, geo-spatial information processing, chemistry and pharmaceuticals, medical diagnosis, and economics. Visualization usually relies heavily on the spatial representation of non-spatial data. The main problem for non-spatial-data visualization is how to identify or extract existing attributes in the data that can be used for spatial representation, or, alternatively, how to convert the data to a form that has spatial attributes. The use of distance metrics (dissimilarity, relevance, disutility, and so on) which are then converted to a spatial format using multidimensional scaling (MDS) techniques, is the method suggested here. Other possibly fruitful methods of conversion are factor analysis and Kohonen nets (*Small*, 1998), clustering and geometric triangulation (*Small*, 1999), and singular-value decomposition.

White and *McCain* (1998) applied MDS to mutual co-citation data of the most-cited authors in the field of information science, extracted from ISI's *Social Science Citation Index*. The resulting chart or map (Figure 1), (based on Figure 6, p. 350, *White* and *McCain*, 1998) allows the viewer to gain an intuition about the intellectual relationships between authors, as reflected by their proximity to each other in the MDS output. This paper aims to demonstrate the power of spatial information systems, in particular, Geographic Information Systems (GIS) to represent this type of data in new and

0138–9130/2001/US $ 15.00

interesting ways. This is not meant to be a re-analysis of *McCain* and *White*'s data but to show how spatial analysis can contribute to the researcher's set of analysis tools. Some skill is required, but any computer-literate researcher can apply the same techniques with a very short learning curve. Familiarity with the data is primary, as modeling requires an understanding of what objects are in the system and what relationships must hold among them.

Figure 1. INDSCAL map of 75 "canonical" information science authors.
Adapted from *White* and *McCain* (1998)

Method

Statistical packages such as SPSS can be used to create the X, Y coordinates via the MDS option. GIS systems such as ESRI's ArcView can be used to import and manipulate the data as maps. ArcView was used to create the graphics presented here, but Open Visualization Data Explorer (OpenDX) (an "open source" no-charge-for-use visualization system (*IBM*, 1999)) is one alternative.

Ignoring the potential pitfalls of various distance metrics and MDS options, such charts as in Figure 1 represent the position of objects in a Cartesian coordinate system. The X and Y values can be used to represent the same data in a GIS (analogous to cities and other GIS point data). Once the data are imported into the GIS, any of the powerful spatial analysis algorithms can be brought to bear.

Examples

Each object on a map may have many numeric attributes associated with it. In a GIS the displayed data are kept in a relational database that can be manipulated by the system, queried by the user, or used to represent the output of spatial queries applied to the map. For example, *White* and *McCain* provide discipline data about the authors in tabular form (Table 2., p. 333, *White* and *McCain*, 1998). This can be represented for each author by colored points or, as is done here in Figure 2, by symbols.

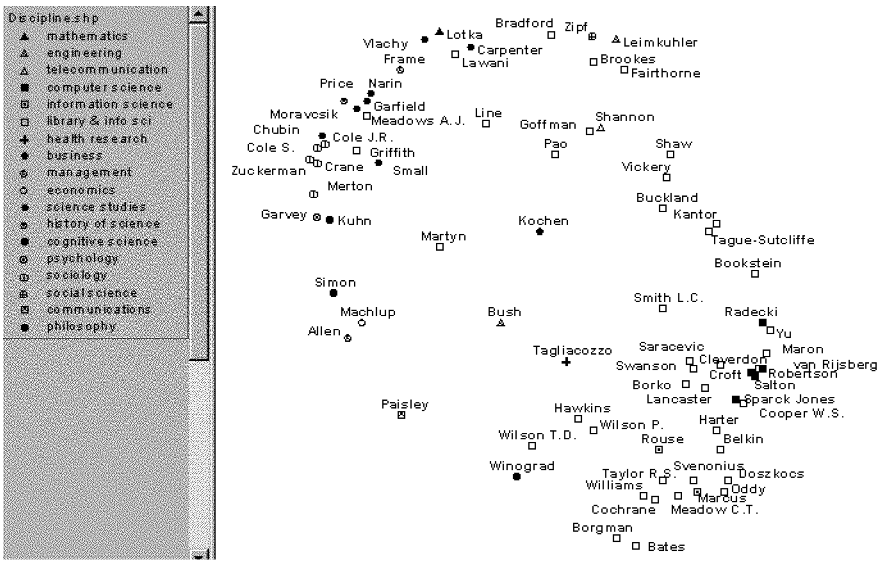

Figure 2. Authors by discipline. Data from from *White* and *McCain* (1998)

This uncovers a phenomenon previously hidden in the data, namely that what *White* and *McCain* identify as the authors interested in information retrieval (the bottom right hand cluster) are also the authors identified as belonging to the disciplines of Library and Information Science (L&IS), and Computer Science. Authors associated with Sociology and Science Studies are clustered at the upper left.

Mean co-citation count data from *White* and *McCain* (Table 4., p. 339, *White* and *McCain*, 1998) can be associated with each author, and contours of like counts can be calculated using the GIS (Figure 3). Numbers can also be associated with the contours, or the isolines can be color-coded and a key displayed with the graphic.

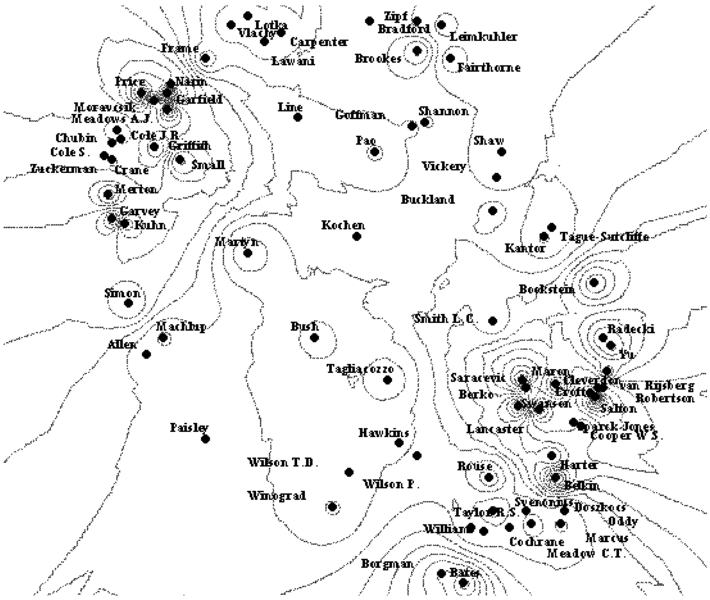

Figure 3. Contours associating authors of like co-citation count

GISs also allow the display of bar graphs (or pie charts) of data in place points or symbols. That would be an alternative method of displaying the co-citation counts for each author, while still retaining the inter-author relationship information.

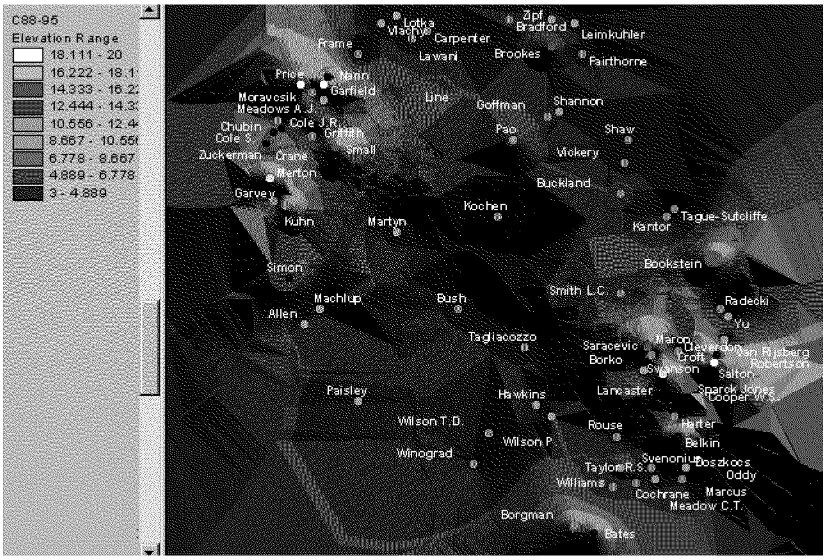

Figure 4. Surface of author co-citation counts (1988-1995) derived from the contours of Figure 3

The method demonstrated next builds on the contour information by extrapolating a surface between points on the contours. The process is analogous to connecting points plotted on a coordinate system to form a curve, but here it is drawn in two dimensions (Figure 3). Alternatively, this could be viewed as analogous to a two-dimensional bar chart, where each author's mean co-citation count determines the height of the bar at that point, with a dark sheet draped over the whole. The "hills" reflect high mean co-citation counts, while the "valleys" reflect low counts. (Keep in mind that "low" here is a relative term; these are the 75 *most* cited information scientists identified in *White* and *McCain*'s study.)

In Figure 5 the accumulated mean co-citation counts for the three periods evaluated by *White* and *McCain* are displayed in 3D. The same methods have been applied to visualization of the relationships between senses of words taken from synonym sets in Roget's International Thesaurus (*Old*, 1999a).

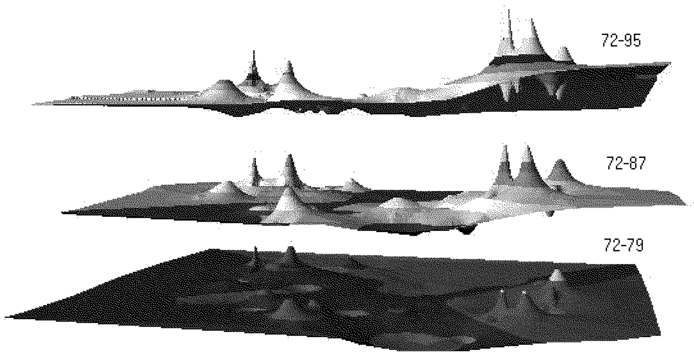

Figure 5. 3D view of accumulated author co-citation counts

Discussion

The data analysis method introduced here demonstrates just four of the many available options; the GIS system utilized allows for the manipulation of three-dimensional maps enabling rotation and inspection from different angles. Even the two-dimensional maps allow zooming, panning, and the hiding of data layers during inspection. The three-dimensional maps can also be exported as VRML files, viewable in virtual reality environments or web browsers.

Any of the data associated with the objects on a map can be identified and displayed concurrently using a GIS. Clicking on a point object displays a tabular output of all of the data associated with that point. For example, in Figure 3, the author's name, discipline, location, and mean co-citation counts for all periods, can be displayed with one click. Using a spatial selection (click and drag) the same data can be displayed for a range of selected authors. Alternatively an SQL-type query can be made against the tabular data, and the result set is highlighted both in the map and in the source table.

GIS can manipulate other data-types. For example polygons (which represent data as areas) and lines (which represent relationships between points, as in graphs) are not utilized here, but are powerful options for non-spatial data analysis.

Even automatic categorization methods (not as yet fully evaluated) may be applied. Figure 6 may be compared to the output of the Kohonen feature map method (demonstrated, for example, in Figure 2. of *White* et al. 1998), which uses an artificial neural network. Figure 6 was produced using a spatial method of classifying the information science authors. The method grows "spheres of influence" according to mean co-citation counts (years 1972-1995), in a simple clustering method. This method does not force authors such as Bush or Kochen into any particular camp. The authors identified in the imbedded query result table are those in the cluster at the bottom right hand corner of the map.

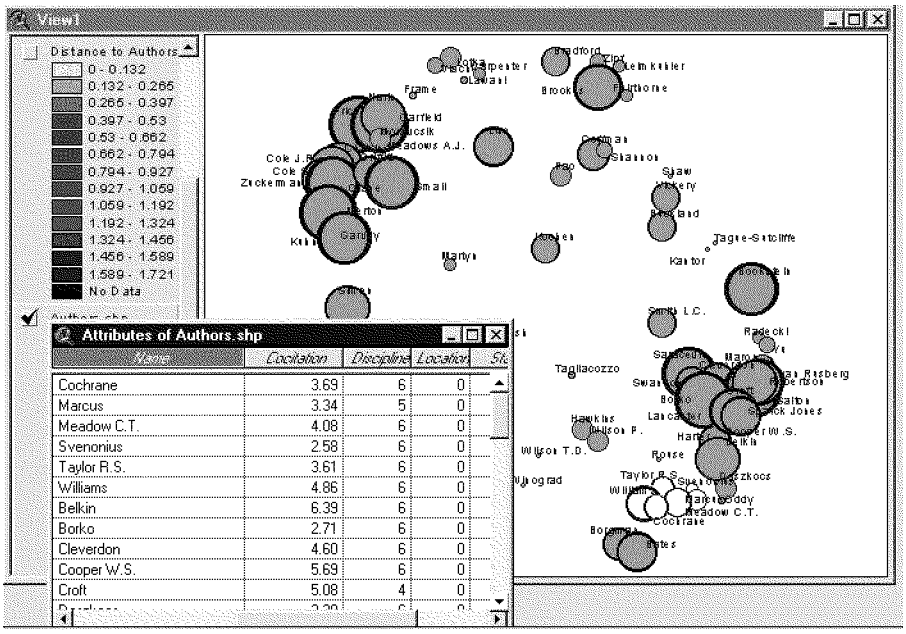

Figure 6. Clusters of authors based on connecting radii of influence (mean co-citation counts)

Conclusion

This paper has introduced a cross section of methods available for non-spatial-data analysis in the spatial information systems domain. The methods are applicable to a wide range of non-spatial data however some limitations still exist. Apart from the use of animation, time series data are difficult to deal with in a GIS. The development of a conceptual model and associated tools for the visualization of spatial-temporal process information is among the goals of the Commission on Visualization of the International Cartographic Association (*ICA*, 1997).

The data in spatial information systems are most easily interpreted in color graphics; interpretation of the gray-scale graphics given in Figure 4 is confounded by the 3D shadowing effects. Another disadvantage of these black and white pictures, as opposed to color pictures, is that less information can be displayed. As stated previously, the disciplines of the authors can be encoded by color, allowing for a more perceptually intuitive method of analyzing the data (where groupings in the data may be more easily identified). In GIS systems the elevation surface described by Figure 4 would be displayed by color ramps (gradients) where differences from author to author are also more easily discriminated (see *Old*, 1999b). Lines, analogous to roads or rivers in a map, may be used to define relationships between entities and allows for the use of lattice and graph methodologies (*Priss* and *Old*, 1998).

Further analysis and evaluation of these methods will be conducted in terms of usability, the evaluation dimensions for Visual Information Retrieval Interface (VIRI) (*Rorvig* and *Hemmje*, 1999), and *Tufte*'s (1983) principles of graphical excellence.

*

I wish to thank Dr Chuck *Davis* without whose support, encouragement, and enthusiasm this paper would never have been written.

References

ESRI (1999). *ArcView GIS*.
Available: http://www.esri.com

International Cartographic Association Commission on Visualization (August 1997), Overview.
Available: http://www.geovista.psu.edu/icavis/

IBM (1999). *Open Visualization Data Explorer*.
Available: http://www.research.ibm.com/dx/

OLD, L. J., (1999a). *Spatial Representation of Semantic Information.*MAICS99 presentation notes. Available: http://php.indiana.edu/~jold/maics/maics.htm

OLD, L. J., (1999b). *Spatial Representation and Analysis of Co-Citation Data on the "Canonical 75": Re-viewing White and McCain.* Available: http://php.indiana.edu/~jold/SLIS/L710.htm

PRISS, U., OLD, L. J. (1998). Information access through conceptual structures and GIS. In: *Information Access in the Global Information Economy. Proceedings of the 61st Annual Meeting of ASIS*, 1998, pp. 91–99.

Roget's International Thesaurus, 3rd Edition, Thomas Crowel Co., 1963.

RORVIG, M., HEMMJE, M. (1999). Conference Notes--1996: Foundations of Advanced Information Visualization for Visual Information (Retrieval) Systems, *Journal of the American Society for Information Science*. 50(9): 835–837.

SMALL, H., (1998). Personal communication.

SMALL, H. (1999). Visualizing science by citation mapping, *Journal of the American Society for Information Science*. 50(9): 799–813.

TUFTE, E. R. (1983). *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT, 1983.

WHITE, H., MCCAIN, K. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science,* 49(4): 327–355.

WHITE, H., XIA LIN, MCCAIN, K. (1998). Two modes of automated domain analysis: multidimensional scaling vs. Kohonen feature mapping of information authors, In: *Structures and Relations in Knowledge Organization, Proceedings 5th Int. ISKO-Conference,* Ergon Verlag, Würzburg, pp. 57–63.

Received March 6, 2001.

Address for correspondence:
L. JOHN OLD
School of Library and Information Science,
Indiana University,
P.O. Box 3065, Bloomington, Indiana 47402, USA
E-mail: jold(at)indiana.edu