Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 1999 Proceedings

Americas Conference on Information Systems (AMCIS)

December 1999

Inductive Explorations of Information Space Geography

Rene Reitsma
University of Colorado

Barbara Buttenfield
University of Colorado

Inductive Explorations of Information Space Geography

René F. Reitsma, University of Colorado, reitsma@cadswes.colorado.edu and
Barbara P. Buttenfield, University of Colorado, babs@colorado.edu

Information Space Mapping

Traditional approaches to mapping information space derive the space's coordinate system from the postulates of a theory pertaining to the objective characteristics of the space. Examples of such approaches can be found in Dömel (1994), Girardin (1995), Hauck (1996) or Nielsen (1995). A good overview of these approaches is provided by Skupin (1998).

Notwithstanding the usefulness of these approaches, especially for purposes of mapping the contents of information spaces, one could question the validity of the chosen coordinate system. Alternatively, one might try to inductively infer the coordinate system from navigational data generated by those traversing the space while in search of information. Unlike the traditional and normative approaches, such an inductive approach would result in the mapping of the information space *as it is used* rather than *as it is defined.* For information spaces implemented on the world-wide web, the navigational data for such an inductive approach exist in the transaction logs kept by the server through which the information is served.

Methodology

A space's coordinate system can be computed from the distances between the objects through a process known as trilateration (Tobler 1995). Modern extensions of this technique are known as multidimensional scaling (Kruskal and Wish 1978; Tobler 1968; Young and Lewyckyj 1996). These techniques allow the scaling of any set of locations (i,j) and the distance between these locations (δ_{ij}) into a coordinate system of any dimensionality (D) and any variant of Minkowski or Lp-norm metric (1).

$$d_{ij} = \sqrt[r]{\left\{ \sum_{d=1}^{D} \left| {}^{i}d - {}^{j}d \right|^{r} \right\}} \tag{1}$$

The fit of the distances in the data (δ_{ij}) to those in the resultant space (d_{ij}) is expressed using the F-stress goodness-of-fit function (2).

$$F = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \left(f\left(\delta_{ij}\right) - d_{ij} \right)^{2}}{k}} \tag{2}$$

The question then becomes as to how to derive the input distances (δ_{ij}) from the users' navigational record. For this purpose, we adopted a method sometimes used for the numerical decomposition of interaction matrices, spatial or otherwise (Aufhauser and Fisher 1985; DiPrete 1990; Everett and Pecotich 1991; Miller and Hayes1990; Scholten1982; Willekens 1983). The method uses loglinear models (DeMaris 1992; Knoke and Burke 1980, Raftery 1995) to decompose an interaction-matrix (frequencies of, for example, trips made, migrations, marriages between social classes, interjournal citations, etc.) into its origin (O), destination (D) and interaction (OD) components.

For an N x M interaction matrix, the (saturated) loglinear model is as in (3).

$$Ln(F_{ij}) = \theta + \lambda_{i}^{O} + \lambda_{j}^{D} + \lambda_{i\,j}^{O\,D} \tag{3}$$

This model can be easily extended to include time (T). In (4), for instance, each of the parameters is interacting with the time variable indicating that the origin (O), destination (D) and origin-destination (OD) interaction patterns change over time.

$$Ln(F_{ijk}) = \theta + \lambda_{i}^{O} + \lambda_{j}^{D} + \lambda_{k}^{T} + \lambda_{i\,j}^{O\,D} + \lambda_{i\,k}^{O\,T} + \lambda_{j\,k}^{D\,T} + \lambda_{i\,j\,k}^{O\,D\,T} \tag{4}$$

Through testing for the statistical significance of each of the parameters, one arrives at the most parsimonious model which satisfactorily explains the observed interaction phenomena.

We propose using the $\lambda_{i\,j}^{O\,D}$ parameters as the inverse measure of distance; i.e., the higher the value, the "closer" we project the points in information space. Hence, the $\lambda_{i\,j}^{O\,D}$ values can directly serve as the input values (similarities) for a multidimen-
sional scaling for reconstructing the information space.

Experiment

Twelve months of transaction logs from the web-based Alexandria Digital Library, or ADL, (Buttenfield, 1997) were used to compile a matrix of interactions between twelve groups of ADL web pages; i.e., each traversed hyperlink was recorded as an interaction. Loglinear modeling of these 175,606

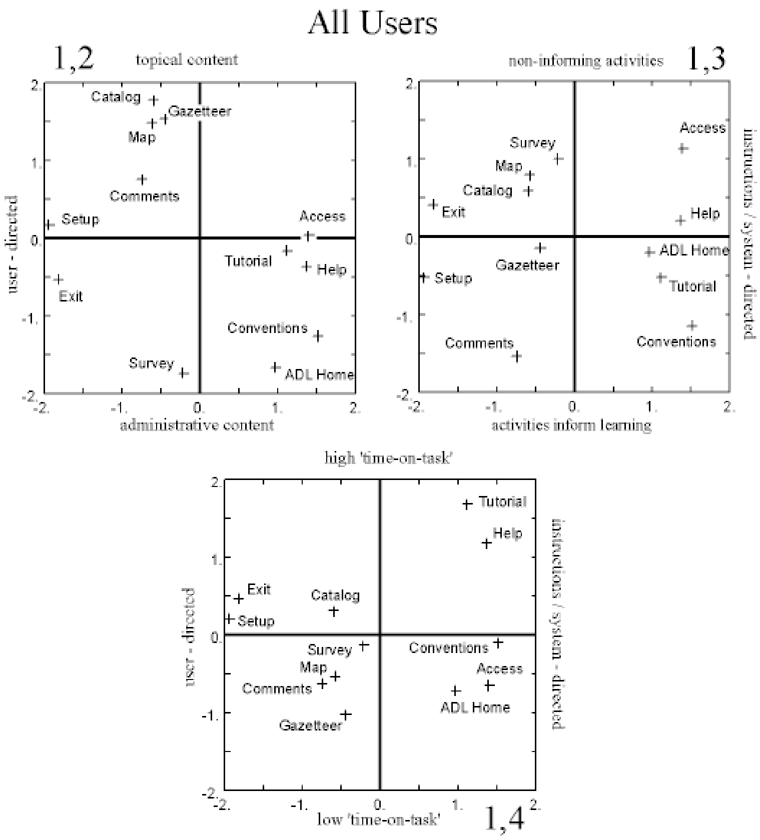

Figure 1. Four-dimensional scaling of ADL transactions

interactions resulted in a preferred model {OD}{OT}{DT}, indicating that navigational patterns {OD} are stable over time; the second order $\lambda_{i\,j\,k}^{O\,D\,T}$ effect was not statistically significant.

Figure 1 contains the results of a multidimen-sional scaling of the resultant $\lambda_{i\,j}^{O\,D}$ values. The solution converged well (low levels of F-stress) using Euclidian metric and four dimensions. In other words, the geometry *implied* by the 175,606 navigated links suggests a four-dimensional, orthogonal information space.

Figure 1 also contains the results of fitting the twelve ADL rooms into this four-dimensional space. By comparing the distributions of these rooms along the various axes, the dimensions of the ADL information space can be readily interpreted:

1. Dimension one illustrates that pages on the extreme left exhibit user-directed activity, such as searching the catalog, gazetteer, submitting comments or filling out the demographic survey. On the right hand are pages of instructions (the tutorial, conventions and access pages); that is, pages of activities that are system-directed.

2. The second dimension differentiates content of the Web pages. At the top are pages of library topical content, or for lack of a better term, the reference pages and the so-called stacks of ADL. At the bottom are pages of which the content is more administrative, advising users how to access the library, but not necessarily telling about the collection itself.

3. Dimension three differentiates uses that inform learning from those that do not. Pages at the bottom of the graph (the conventions page, tutorial and system setup pages) contain teaching materials to help users navigate the library more easily. At the top of the graph, the survey, access and exit pages exhibit no specific learning content. In the middle of the page, the gazetteer, setup and home pages do not contain direct content to facilitate navigation. The help pages contain definitions and explanations, but no explicit navigational support.

4. The fourth dimension differentiates time-on-task for specific activities on specific pages. Transaction logs indicate that when users navigate to the tutorial and help pages, they spend considerable time reading extensive

amounts of text and graphics. In contrast, the time spent in the gazetteer, for example, is relatively short. Users apparently go to this library "room" to find a place name, retrieve it and move to another part of the library fairly quickly.

Conclusion

Using the statistically stable elements of navigational paths taken through a web-based information space, the geometry of that space, as well as the meaning of its dimensions can be reconstructed. Such spatial interpretation greatly simplifies the conceptualization and understanding of the paths travelled and people's perception of the information space. Similarly, when the interaction patterns change through time; i.e., the λ_{ijk}^{ODT} parameters are statistically significant, it can be concluded that the geometry of the information space changes through time, possibly as the effect of changes in the information system's user interface, changes in the user population or as a consequence of user training.

References

Aufhauser, E., and Fisher, M.M. "Log-linear Modelling of Spatial Analysis," Environment and Planning A (17), 1985, pp. 931-951.

Buttenfield, B.P. "Delivering Maps to the Information Society: A Digital Library for Cartographic Data," Proceedings, 17th Conference of the International Cartographic Association, Stockholm, Sweden, June 1997, pp. 1409-1416.

DeMaris, A. "Logit Modeling, Practical Applications," Series: Quantitative Applications in the Social Sciences, (86), Sage Publications, Newbury Park, CA, 1992.

DiPrete, T.A. "Adding Covariates to Loglinear Models for the Study of Social Mobility," American Sociological Review, (55), 1995, pp. 757-773.

Dömel, P. "WebMap - A Graphical Hypertext Navigation Tool," Second International WWW Conference; Chicago, IL, 1994, http://www.ncsa.uiuc.edu/SDG/IT($/Proceedings/Searching/doemel/www-fall94.html.

Everett, J.A., and Pecotich, A. A. "Combined Loglinear/MDS Model for Mapping Journals by Citation Index," Journal of the American Society of Information Science, (42), 1991, pp. 405-413.

Girardin, L. Cyberspace Geography Visualization, Springer-Verlag, Berlin, Germany, 1995.

Hauck, F. "Supporting Hierarchical Guided Tours in the World-Wide Web," Computer Networks and ISDN Systems, (28), 1996, pp. 1233-1242.

Knoke, D., and Burke, P.J. "Loglinear Models," Series: Quantitative Applications in the Social Sciences, (20), Sage Publications, Beverly Hills, CA, 1980.

Kruskal, J.B., and Wish, M. "Multidimensional Scaling," Series: Quantitative Applications in the Social Sciences, Sage Publications, Newbury Park, CA,1978.

Miller, R.L., and Hayes, B.C. "Female Marital Mobility Within Northern Ireland and the Republic of Ireland," Sociological Review, (38), 1990, pp. 111-133.

Nielsen, J. Multimedia and Hypertext; the Internet and Beyond; Academic Press Professional, San Diego, CA, 1995.

Raftery, A. "Bayesian Model Selection in Social Research," Sociological Methodology, (25), 1995, pp. 111-163.

Scholten, H.J. "Describing and Forecasting Migration Within a Housing Market Area: The Loglinear Model Used as a Spatial Interaction Model," Geografische en Planologische Notities, (22), Free University of Amsterdam, Amsterdam, 1982.

Skupin, A. "Organizing and Visualizing Hypermedia Information Space," Doctoral Dissertation, Dept. of Geography, State University of New York at Buffalo, 1998.

Tobler, W. A "Capadocian Speculation," Nature, (231), 1968, pp. 39-41.

Tobler, W. "A Graphical Introduction to Survey Adjustment," Cartographica, (32), 1995, pp. 33-42.

Willekens, F.J. "Specification and Calibration of Spatial Interaction Models: A Contingency-Table Perspective and an Application to Intra-Urban Migration in Rotterdam," Tijdschrift voor Economische en Sociale Geografie (TESG), (73), 1983, pp. 239-252.

Young, F. W., and Lewyckyj, R. ALSCAL User's Guide, L.L. Thurstone Psychometric Laboratory, University of North Carolina, 1996.

Acknowledgments

This paper forms a portion of the Alexandria Project, funded jointly by NSF (grant IRI-94-11330), ARPA, and NASA. Additional support from the University of Colorado is also gratefully acknowledged.