



Loglinear and multidimensional scaling models of digital library navigation

BARBARA P. BUTTENFIELD

Department of Geography, CB-260, University of Colorado, Boulder, CO 80309-0260, USA. email: babs@colorado.edu

RENE F. REITSMA

Department of Accounting, Finance and Information Management, College of Business, Oregon State University, 200 Bexell Hall, Corvallis, OR 97331-2603, USA. email: rreitsma@juliet.stfx.ca

(Received 4 January 1999 and accepted in revised form 25 January 2000)

Computer transaction log analysis comprises the modeling of patterns of machine–user transactions. Due to the web-like structure of the Internet, interactions can be represented in a three-dimensional, origin–destination–time flow/transaction matrix. Loglinear models are used to model these transactions in terms of their time, origin and destination components and to detect the signature and stability of patterns of navigation through an Internet-based digital library. Through the use of metric multidimensional scaling, transaction patterns are then visualized and interpreted. The methods are applied to a 12-month time series of transaction logs taken from the Alexandria Digital Library, a web-based digital library of maps and other cartographic materials. Among other things, results show that transaction patterns do not change significantly despite changes in the user interface and that user training has a significant impact on people’s navigation of the library.

© 2002 Published by Elsevier Science Ltd.

KEYWORDS: loglinear models; multidimensional scaling; transaction patterns.

1. Introduction

Evaluation of computer interface design has evolved into a multi-phased process, establishing user performance and satisfaction as a software testbed is designed, implemented and deployed (Buttenfield, 1998). Resolving discrepancies between what users think they do, and what they actually do (Landauer, 1995), forms a guiding principle for most evaluation efforts. The use of entry and exit surveys, focus groups and paper-based interface walkthroughs lends insight to user expectations. The use of transaction logs documents actual behavior. However, two concerns constrain wider adoption of transaction logging as an effective evaluation tool. The logs generate a huge volume of data, which must be parsed and substantially reduced to uncover use

patterns. Additionally, it is difficult to establish whether the use patterns which are uncovered are idiosyncratic and which patterns are stable over time and context of use.

The first problem has been essentially eliminated with the advent of automated parsing algorithms and similar automated technologies. The second constraint has been addressed by pattern matching tools, identifying frequently occurring “templates” which match logged behaviors. This form of deductive analysis has illuminated numerous studies (cited in the section below). It has the disadvantage that creation of a template may predispose the identification of particular use patterns. This paper presents an alternative, inductive approach to recovering stable patterns of use from large volume transaction log data. The advantages of the inductive methods include establishment of statistical significance for identified patterns, interpretation of specific dimensions of use, and robust identification of emergent (and previously unidentified) patterns of use.

In this paper, loglinear models are used to detect the signature and stability of patterns of activity for the Alexandria Digital Library, an Internet-based testbed cataloging distributed archives of maps, satellite images and geographical data sets. Through multidimensional scaling these patterns are then visualized and interpreted. Four dimensions of use are identified, and it is demonstrated that the use patterns along these dimensions vary for particular subgroups of the user population.

2. Deductive analysis of transaction logs

Transaction logs are comprised of data collected by electronically logging system interactions between an application and a user, as in an information retrieval package, for example, where users browse on-line archives for information (Figure 1). The common application relates to use of electronic library catalogs (Peters, 1993; Peters, Kurth, Flaherty, Sandore & Kaske, 1993). A second, more recent branch of applications can be found in the domain of software engineering and software usability testing (Landauer, 1995). With the merging of Internet hypertext transfer protocols with hypertext databases, transaction log analysis (Gay & Mazur, 1993; Wildemuth, 1993; Buttenfield & Kumler, 1996; Buttenfield, 1997) and on-line monitoring

```
204.133.134.17 - beta [29/Jan/1997:10:09:29 -0800] "GET /ADL/Catalog/3?USER_ID=2
713.426b&cat_search_technique=contain&on_page=cat_search_regions&cat_search_regi
ons=current_map&cat_Beginning_date_of_data_rel=greater_equal&cat_Beginning_date_
of_data_month=1&cat_Beginning_date_of_data_day=1&cat_Beginning_date_of_data_year
=&cat_Ending_date_of_data_rel=less_equal&cat_Ending_date_of_data_month=1&cat_Ending_
date_of_data_day=1&cat_Ending_date_of_data_year=&cat_Geographic_datatype_for
m_genr_rel=equal&cat_Geographic_datatype_form_genr=remote+sensing&on_page=cat_File_t
ype_%28e.g.%2C_GIF%29_of_brow&cat_File_type_%28e.g.%2C_GIF%29_of_brow=%25&M
ap.x=57&Map.y=15 HTTP/1.0" 302 360

128.111.96.72 - demo [29/Jan/1997:09:56:14 -0800] "GET /ADL/CatalogSearchResults
/1?USER_ID=2026.335a&on_page=cat_tagged&Gazetteer.x=97&Gazetteer.y=12 HTTP/1.0"
302 372
```

FIGURE 1. Examples of transaction log entries for 29 January 1997. Two separate entries, showing port of entry (demo and beta), numeric user IDs, IP addresses and timestamps. The first entry initiates a catalog search, and the second requests information on selected items returned (tagged) from a catalog search.

(Penniman & Dominick, 1980; Rice & Borgman, 1983; Birbeck, 1986; Borgman, 1995) have gained substantial visibility.

Transaction patterns can identify specific typical and atypical behaviors (Nielsen, 1993; Rubin, 1994); for example, cycles of search actions, instances of confusion (cycling back and forth between two options), and recurring misuse of system commands.

Wildemuth, de Blik, He and Friedman (1992) and Wildemuth, de Blik, He and Friedman (1993) utilize three methods to analyse use patterns for on-line medical databases. State-transition matrices highlight first-order search links, similar to the data collected for this paper. Identification of frequently repeating patterns helps to delineate model searching strategies. Grouping similar sequences of links, and comparing the circumstances in which they occur, provides a descriptive context for search behavior patterns.

Gay and Mazur (1993) interlace transaction logs of keystrokes and mouseclicks with videotape and voice transcripts. Logs comprise one of five types of tracking tools in their taxonomy and provide information about navigation strategies, comparative effectiveness of specific menu selections, and evidence of knowledge construction and relational webs of information. Using spreadsheet analysis, they have documented which databases in a collaborative computing project were utilized, and at what points in the project database information was accessed or ignored (Gay, Mazur & Lentini, 1994).

Patterns in the logs can uncover specific reasons for failed queries and other sources of user frustration. "For example, queries in the transaction logs which produced no results were analysed. With the addition of better word stemming and a spell checker, over half of those queries would have become meaningful and would have returned results, and so these features were added to the new Informedia Digital Video Library (IDVL)." (Christel & Pendyala, 1996).

Transaction logs have been studied to identify points where users leave the system. Neumann and Bishop (1998) report that transaction logs for the DeLlver digital library indicated a large number of users were apparently dissuaded by an overly long user-registration system. Once the registration process was streamlined, subsequent logs showed a significant increase in the number of users completing the registration and going on to browse system holdings. Battenfield and Kumler (1996) document the redesign of specific buttons (e.g. the query tool) in the UNIX prototype for the Alexandria Digital Library following examination of transaction logs indicating that the icon was confusing users.

Borgman (1995) summarizes results from two decades of on-line monitoring research. "While we learned early on that results vary widely, even when experienced searchers are given the same search task, we knew little about how the process that led to each of the results varied. We now can identify cycles of search actions, isolate typical and atypical search behavior and predict some user actions based on prior actions." Collecting transaction logs allows the system evaluator to monitor system use without intrusion. The method "...provides large volumes of data at relatively low cost, and can assist in qualitative interpretations of quantitative models" (Borgman, 1995). In counterpoint, transaction logs require parsing and pre-processing before interpretation may begin. Particularly on the Internet, it is difficult to isolate individual users and individual

sessions without adding special user-ids and session counters. Moreover, network delays on the Internet can obscure timestamp patterns of use (Buttenfield, 1995).

3. Inductive analysis of transaction logs

The analysis of transaction logs to date is deductive and relies on ex-post testing of selected patterns. Patterns are resolved from the raw logs by parsing and are then tested for statistical significance. An alternative used often by user interface designers (Landauer, 1995) is to create a normative model of expected use patterns and compare the content of transaction logs to this standard, noting where discrepancies occur and determining how users get “back on track” for a specified task (Buttenfield & Kumler, 1996). This approach is well-suited for classical deductive methodologies of theory formulation and subsequent empirical testing; i.e. pre-defined patterns are the operationalized hypotheses derived from an espoused theory. In case such hypotheses are lacking, however, such pattern definitions are not available and the deductive approach can provide neither reliable insights nor inference.

Lack of such hypothesis and theory is not uncommon, however. Designers of user interfaces and web pages employ a well-defined set of expected transactions as part of their evaluation toolkits. Hypermedia systems allow users to undertake sequences of actions that are not anticipated in the original design (Buttenfield & Weber, 1994). In a sense, the point of hypermedia is to facilitate rather than preclude these actions (Buttenfield, 1996). Consequently, actual transaction patterns can be highly complex. Internet-based system structures allow higher degrees of transaction pattern complexity than do traditional, hierarchically oriented computer systems. Hence, the availability of an adequate theory for predicting transaction patterns in complex systems is likely to be the exception rather than the rule.

In the absence of adequate theory, inductive methodologies can be applied. The objective of such a methodology is to sort, categorize and analyse the available data to uncover a viable theory. Instead of defining transaction patterns *a priori*, they are formulated by inference. The case study used here concerns the Alexandria Digital Library, or ADL (Andresen *et al.*, 1995; Buttenfield & Goodchild, 1996), a digital archive of maps, satellite imagery and air photos made available via the Internet (<http://alexandria.sdc.ucsb.edu>). Although the Alexandria Digital Library designers had certain navigational paths in mind when creating the interface, an infinite number of paths is possible, allowing, in principle, an infinite number of patterns in the transaction logs.

Although no definite theory on the nature of these (possible) patterns was available, certain trends and patterns were expected. It was hypothesized, for instance, that changes in the transaction patterns would follow modifications of certain elements of the user interface. Also, the transaction patterns exhibited by trained Alexandria Digital Library users will likely differ from those of untrained users.

4. Research questions

Despite the lack of a definite theory on ADL navigational behavior, the following general research questions can be presented.

- (1) What are the main patterns of transaction in ADL? What are their essential components?
- (2) Do transaction patterns change over time; for instance, as a consequence of modifications to the user-interface?
- (3) Do different user groups exhibit different transaction patterns? Can these patterns be understood in the context of the user group?

In this paper, we introduce a set of models for exploring these questions. Instead of defining patterns *a priori* and then testing for statistical significance, the models are used to first resolve the statistically stable patterns in the transaction data. Following this step, we generate explanatory theories. The models are based on the similarity between a transactional data matrix and a contingency or cross-classification table. Loglinear models developed for the analysis of such tables (Knoke & Burke, 1980; Raftery, 1986, 1995; Everett & Pecotich, 1991; De Maris, 1992) are used to analyse the transaction matrix. Next, the statistically stable patterns resolved through the use of these models are further explored through multidimensional scaling, a technique for reducing complex sets of data to their essential components or dimensions (Kruskal & Wish, 1978; Coury, 1991; Everett & Pecotich, 1991; Rhodes & Stern, 1993; Young & Lewycky, 1996). Finally, through interpretation of the resultant patterns, an initial theoretical framework for the explanation of the observed patterns is put forward.

5. Loglinear models of transactional data

Transactional data for web-page access can be represented as a three-dimensional origin–destination–time interaction matrix where $O_iD_jT_k$ represents the amount of interaction between origin O_i and destination D_j during the time interval T_k . A transaction O_iD_j is defined as the execution of a hyperlink from web page O_i to web page D_j . Hence, the cells of the matrix contain the frequencies of $O_iD_jT_k$ transactions (Figure 2). Note that this conceptualization implies the assumption that individual hyperlinks are independent; i.e. successive hyperlinks are unrelated and different users navigate the web pages independently of each other. Although this assumption implies the loss of information contained in the actual paths; i.e. the actual sequence of hyperlinks executed by ADL patrons, modeling the links as independent events is useful as a first step in revealing important stable patterns of use.

The transaction matrix can be conceptualized as a contingency table with nominal variables origin (O), destination (D) and time (T) having I, J and K levels, respectively. The loglinear model which fully describes the frequencies of such a contingency table is written as

$$O_iD_jT_k = \theta + \lambda_i^O + \lambda_j^D + \lambda_k^T + \lambda_{ij}^{OD} + \lambda_{ik}^{OT} + \lambda_{jk}^{DT} + \lambda_{ijk}^{ODT}, \tag{1}$$

where θ =effect of total transaction volume, λ_i^O =effect of origin i , λ_j^D =effect of destination j , λ_k^T =effect of time interval k , λ_{ij}^{OD} =first-order interaction between origin and destination, λ_{ik}^{OT} =first-order interaction between origin and time, λ_{jk}^{DT} =first-order

	Time _L	Dest. ₁	Dest. ₂	Dest. _K
	Time ₂	Dest.	Dest.	Dest.
Time ₁		Dest. ₁	Dest. ₂	Dest. _K
Orig. ₁	T _{1O₁D₁}	T _{1O₁D₂}	T _{1O₁D_K}	D _K
Orig. ₂	T _{1O₂D₁}	T _{1O₂D₂}	T _{1O₂D_K}	D _K
Orig. _J	T _{1O_JD₁}	T _{1O_JD₂}	T _{1O_JD_K}	D _K

Time ₂	Dest. ₁	Dest. ₂	Dest. _K
Orig. ₁	T _{2O₁D₁}	T _{2O₁D₂}	T _{2O₁D_K}
Orig. ₂	T _{2O₂D₁}	T _{2O₂D₂}	T _{2O₂D_K}
Orig. _J	T _{2O_JD₁}	T _{2O_JD₂}	T _{2O_JD_K}

Time _p	Dest. ₁	Dest. ₂	Dest. _K
Orig. ₁	T _{pO₁D₁}	T _{pO₁D₂}	T _{pO₁D_K}
Orig. ₂	T _{pO₂D₁}	T _{pO₂D₂}	T _{pO₂D_K}
Orig. _J	T _{pO_JD₁}	T _{pO_JD₂}	T _{pO_JD_K}

FIGURE 2. Web-based transactional data as a time–origin–destination matrix.

interaction between destination and time, λ_{ijk}^{ODT} = second-order interaction between origin, destination, and time.

The parameters of equation (1) are constrained by the following equations:

$$\sum_{i=1}^I \lambda_i^O = \sum_{j=1}^J \lambda_j^D = \sum_{k=1}^K \lambda_k^T = 0, \tag{2}$$

$$\sum_{i=1}^I \sum_{j=1}^J \lambda_{ij}^{OD} = \sum_{i=1}^I \sum_{k=1}^K \lambda_{ik}^{OT} = \sum_{j=1}^J \sum_{k=1}^K \lambda_{jk}^{DT} = 0, \tag{3}$$

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \lambda_{ijk}^{ODT} = 0. \tag{4}$$

Interpretation of the parameters is straight-forward in that λ values smaller than, larger than or equal to 0 indicate a disproportionately small, large or no effect, respectively. For instance, a value of $\lambda_{ik}^{OT} > 0$ indicates that during time interval k , averaged over all D destinations, origin i generated a disproportionate number of interactions. Similarly, a value of $\lambda_{ij}^{OD} < 0$ indicates that the interaction between origin i and destination j , averaged over time, is disproportionately small. Finally, a value of $\lambda_{ijk}^{ODT} > 0$ indicates that the interaction between origin i and destination j during time interval k is disproportionately large when compared to the same interaction in other

time intervals. This latter parameter is of special importance to this research as it establishes a statistical measure for change in transaction patterns. With this, we mean to say that by testing for the statistical significance of the λ_{ijk}^{ODT} second-order interaction, we can assess if an existing transaction pattern λ_{ij}^{OD} changes over time; for instance, as a consequence of changes to a system's user-interface or changes in the user-community.

The model can, of course, be extended with additional variables. For instance, in case transactions can be attributed to different user groups, the model can be written as

$$\begin{aligned}
 O_i D_j T_k U_l = & \theta + \lambda_i^O + \lambda_j^D + \lambda_k^T + \lambda_l^U \\
 & + \lambda_{ij}^{OD} + \lambda_{ik}^{OT} + \lambda_{il}^{OU} + \lambda_{jk}^{DT} + \lambda_{jl}^{DU} + \lambda_{kl}^{TU} \\
 & + \lambda_{ijk}^{ODT} + \lambda_{ijl}^{ODU} + \lambda_{ikl}^{OTU} + \lambda_{jkl}^{DTU} \\
 & + \lambda_{ijkl}^{ODTU}, \tag{5}
 \end{aligned}$$

where λ_i^U = effect of user group i ; λ_{il}^{OU} = first-order interaction between origin and user group; λ_{jl}^{DU} = first-order interaction between destination and user group; λ_{kl}^{TU} = first-order interaction between time and user group; λ_{ijl}^{ODU} = second-order interaction between origin, destination and user group; λ_{ikl}^{OTU} = second-order interaction between origin, time and user group; λ_{jkl}^{DTU} = second-order interaction between destination, time and user group; λ_{ijkl}^{ODTU} = third-order interaction between origin, destination, time and user group.

As equations (1) and (5) constitute saturated models; i.e. all possible effects are included, the models fit the observed transaction matrix perfectly. To test for the statistical significance of individual parameters, the predictions of simpler, nonsaturated models are compared with the observed frequencies. If the fit of the simpler model is acceptable, the omitted parameters are considered nonsignificant (Knoke & Burke, 1980). The likelihood ratio (L^2) goodness-of-fit statistic is chi-squared (χ^2) distributed with degrees of freedom df = number of parameters excluded from the model.

Loglinear modeling of interaction matrices has, on occasion, been used in Geography for the analysis of spatial flow phenomena such as migration or commuting (Scholten, 1982; Willekens & Baydar, 1983a, b; Aufhauser & Fisher, 1985). In Sociology, this type of modeling has been used to analyse patterns of mobility between, for instance, social class structures (DiPrete, 1990; Miller & Hayes, 1990). Similarly, Everett and Pecotich (1991) use loglinear models to analyse structure of inter-journal citations. We know, however, of no applications which model time as a third, nominal variable to be included in the model as in equation (1) or (5).

6. Loglinear models of Alexandria Digital Library transactions

To explore the three research questions, 12 months of Alexandria Digital Library transaction data (April 1996–May 1997) were collected and processed. Data for July 1996 were not available due to a server hard disk failure. Internet gateway problems surfaced intermittently in June 1997, damaging the reliability of monthly transaction logs. During this period, system designers turned their attention to a JAVA version of the library, and the transaction logs for the html version developed large temporal gaps,

forcing termination of the logging experiments. However, the hypertext transfer protocol version of ADL created a statistically reasonable environment for inference, with reliable transaction logs collected for 12 months during that 13-month period, and more than 175 000 transactions logged overall.

The html version of ADL included 155 web pages, and hyperlinks were embedded to navigate between some—but not all—pages. For example, the Alexandria Digital Library *catalog* consisted of eight different pages, including a page to formulate a query, a page to search a subject-heading thesaurus, pages for search results, pages for abbreviated and full metadata listings, a page for displaying thumbnail versions of images, and so forth. As another example, the user *tutorial* incorporated nearly 35 web pages. In the course of a query, a user would venture into the *catalog* or into the *tutorial* section, and within these “rooms” of the library, the only way to move to another “room” (the map *browser*, or the place name *gazetteer*, for example) was to return to the first page of the *catalog* or *tutorial* “room” and hyperlink from there to another part of the library. Access to the Alexandria Digital Library *homepage* and to the on-line *help* functions was available on every page. To simplify modeling computations and to mirror the logistics of the library navigation, the 155 pages were partitioned into twelve “rooms” (Table 1).

After the data were converted into a $12 \times 12 \times 12$ origin–destination–time transaction matrix, a series of loglinear models were fit. The results are summarized in Table 2. Note that although results for the complete ADL interaction record are presented (All users), analyses were conducted for specific user groups as well. Ideally, this could be accomplished using the approach specified in equation (5). However, in this case such an approach was not feasible as the categorizations of users were not exclusive, i.e. the transaction data for certain users are part of more than one user group. Therefore, instead of figuring the user group as a separate variable in the model, models were independently estimated for each user group.

Statistical significance of the models listed in Table 2 was evaluated using two techniques: the traditional method compares increases in the likelihood ratio (ΔL^2) with increase in the model’s degrees of freedom (Δdf) using the chi-squared (χ^2)

TABLE 1
The twelve Alexandria Digital Library “Rooms”

Room	Groups
ADL Homepage (1) Access page (2) Conventions page (5) Exit page (6) Catalog (3) Placename gazetteer (8) Map browser (9) Comments (4) User demographic survey (7) Help files (11) Tutorial pages (12) System setup (10)	Getting into and out of the library Reference, indexes and the “stacks” Library patron feedback On-line documentation Setting the work environment

TABLE 2
Alexandria Digital Library transactions loglinear model fits

User group	Model	L ²	df	p	BIC)L ²)df	p)BIC
All users (n = 175,606)	{ODT}	0	0	–	0	–	–	–	–
	{OD}{OT}{DT} [†]	5,178.25	1,452	<0.01	–12,356.11	5,178.25	1,452	<0.01	–12,356.11
	{OT}{DT}	380,176.3	1,573	<0.01	361,180.80	374,997.75	121	<0.01	373,536.91
	{OD}{DT}	9,069.81	1,584	<0.01	–10,196.89	3,891.56	132	<0.1	2159.22
	{OD}{OT}	8,931.50	1,584	<0.01	–10,058.57	3,753.25	132	<0.1	2297.54
ADR panelists (n = 4,579)	{ODT}	0	0	–	0	–	–	–	–
	{OD}{OT}{DT} [‡]	839.75	1,452	>0.10	–11,399.50	839.75	1,452	>0.10	–3909.31
	{OT}{DT}	9,349.88	1,573	<0.10	–3,909.31	8,510.13	121	<0.1	7490.19
	{OD}{DT}	1,212.90	1,584	>0.10	–12,139.02	373.15	132	<0.1	–739.52
	{OD}{OT}	1,228.92	1,584	>0.10	–12,122.99	389.17	132	<0.1	–723.49
	{T}{OD}	2,672.30	1,716	<0.1	–11,792.27	1459.40	132	<0.1	346.75
Librarians (n = 15,297)	{ODT}	0	0	–	0	–	–	–	–
	{OD}{OT}{DT}	1,459.68	1,452	>0.10	–12,530.94	1,459.68	1,452	>0.10	–12,530.94
	{OT}{DT}	30,339.31	1,573	<0.01	15,182.81	2,8879.63	121	<0.01	27,713.75
	{OD}{DT}	2,922.098	1,584	<0.01	–12,340.39	1,462.42	132	<0.01	190.55
	{OD}{OT}	2,996.73	1,584	<0.01	–12,265.76	1,537.05	132	<0.01	265.18

[†] Preferred models are in *bold italics*.

[‡] Although the {OD}{DT} and {OD}{OT} models for the ADR panelists each generate a lower BIC than the {OD}{OT}{DT} model, dropping both parameters from the model ({T}{OD}) generates a bad fit.

distribution (p). An alternative measure, the Bayesian Information Criterion (BIC), was recently introduced by Raftery (1986, 1995). The BIC is a goodness-of-fit statistic which penalizes for the size of the sample and the number of model parameters. For loglinear models, it is defined as an adjustment to the following likelihood ratio:

$$BIC = L^2 - df \ln(n). \quad (6)$$

The BIC for a saturated loglinear model equals 0. Preferred models yield negative values.

Comparing the results of the two types of tests listed in Table 2 clearly illustrates the value of the BIC. The Alexandria Digital Library transaction data sets are large ($4579 < n < 175\,606$). At these sample sizes, model parameters are very likely to be considered statistically significant when using L^2 , the magnitude of which is linearly proportional to sample size. Therefore, when using BIC, many parameters which would be considered statistically significant using the L^2 test, will no longer be considered statistically significant when corrected for sample size. For example, when comparing the saturated model {ODT} with the model from which the second-order ODT interaction has been dropped {OD}{OT}{DT} for the “All users” group, a $\Delta L^2 = 5178.25$ with a $\Delta df = 1452$ yields a P -value of < 0.01 indicating that the λ_{ijk}^{ODT} parameter is statistically significant. However, comparison of the same models yields a $\Delta BIC = -12\,356.11$ indicating an improvement in model fit when the parameter is dropped from the model. Clearly, the P -value of < 0.01 must be attributed to sample size, and not to the significance of the parameter. We emphasize the importance and value for transaction log analysis of this new measure of statistical significance. As transaction logs are likely to contain very large samples, χ^2 -distributed and other statistics which are sensitive to sample size cannot be meaningfully applied. The BIC provides an elegant and easy-to-apply alternative for proper significance testing in those cases.

The results in Table 2 indicate that, according to the BIC, ADL transaction patterns as expressed by the λ_{ij}^{OD} parameters have not changed over time. Dropping the λ_{ijk}^{ODT} parameter from the models yields improvements in BIC model fit. In other words, even through the models are lacking a possible effect on the transaction matrix, their BIC fit increases. Therefore, one can safely assume that OD transaction patterns do not change significantly over time.

The results in Table 2 also indicate significant OD transaction patterns for all groups. In addition, user groups show significant first-order interactions between origin and time (OT) and time and destination (DT). These interactions indicate that the role of certain pages—as either origin or destination pages—changes over time. In the remainder of this paper, however, we concentrate on the more interesting OD transaction patterns.

7. Alexandria Digital Library transaction logs as a multidimensional phenomenon

After selecting the preferred model of ADL transactions for each user group, the values of the 144 (12×12) λ_{ij}^{OD} interactions were estimated for each group. For purposes of illustration, Table 3 contains the λ_{ij}^{OD} values for the “All users” group. The lambda

TABLE 3
Interactions for the “All users” group {OD} {OT} {DT} model

ADL page groups	<i>Home</i>	<i>Acces</i>	<i>Catalog</i>	<i>Comments</i>	<i>Conventions</i>	<i>Exit</i>	<i>Survey</i>	<i>Gazetteer</i>	<i>Browser</i>	<i>Setup</i>	<i>Help</i>	<i>Tutorial</i>
<i>Home</i>	3.66[†]	0.28	-1.41	-1.23	3.87	-1.72	0.08	-1.16	-1.57	-1.1	0.98	-0.66
<i>Access</i>	-0.22	1.89	1.13	-1.65	-0.21	-2.23	0.79	1.19	1.91	-1.62	-0.14	-0.85
<i>Catalog</i>	-1.43	-0.89	4.11	0.66	-2.04	1.11	-1.47	0.18	0.66	0.76	-0.83	-0.82
<i>Comments</i>	-1.09	-1.26	0.78	3.31	-0.15	-0.22	-1.36	0.88	0.74	0.68	-1.7	-0.63
<i>Conventions</i>	1.0	2.71	-2.47	0.34	4.28	-1.43	-0.31	-1.87	-2.26	-1.59	1.37	0.23
<i>Exit</i>	-0.67	-0.32	-0.71	-0.41	-1.48	3.07	1.42	-0.8	-0.69	1.88	-0.23	-1.06
<i>Survey</i>	3.2	0.58	-1.51	-0.74	-0.88	0.55	3.66	-1.66	-1.41	-1.09	-0.07	-0.63
<i>Gazetteer</i>	-1.41	-1.21	0.72	0.87	-1.92	0.84	-1.16	3.96	0.94	0.35	-0.85	-1.12
<i>Browser</i>	-1.42	-0.91	1.12	0.8	-2.07	0.63	-0.75	1.21	2.94	0.38	-0.62	-1.31
<i>Setup</i>	-1.44	-1.58	0.08	0.71	-1.58	1.88	-0.51	0.01	0.48	4.19	-1.22	-1.01
<i>Help</i>	-0.16	0.45	-0.79	-1.58	1.94	-1.11	-0.22	-0.77	-0.4	-1.6	2.01	2.22
<i>Tutorial</i>	-0.03	0.28	-1.05	-1.08	0.25	-1.37	-0.17	-1.16	-1.35	-1.25	1.29	5.64

[†] $|\lambda_{ij}^{OD}| \geq 2.0$ are in ***bold italics***.

values indicate that the majority of the transactions occur within single groups of pages; i.e. almost all λ_{ij}^{OD} interactions for $i = j$ are strongly positive, indicating that the page groups represent centers of activity within the library.

Although it is of course possible to study the lambda matrices for patterns of high and low values, it was instead decided to apply a more aggregate approach in searching for the underlying structure of the transaction patterns. To this end, multidimensional scaling or MDS (Kruskal & Wish, 1978; Everett & Pecotich, 1991; Coury, 1991; Young & Lewyckyj, 1996) was applied to the lambda matrices. With this technique, a data set such as a λ_{ij}^{OD} matrix can be reduced to its essential contents. The elegance of the technique lies in the conceptualization of the information in a data set in so-called "dimensions." For instance, a $N \times N$ distance matrix representing all distances between N points on a map, can be modeled as a two-dimensional phenomenon because the distances are a representation of points in two space (Kruskal & Wish, 1978). Similarly, the λ_{ij}^{OD} matrix of the fitted loglinear models can be used as the measure of similarity between the various clusters of ADL Web pages, and a multidimensional scaling can be used to resolve the matrix' underlying structure or components. This approach is consistent with the method introduced by Everett and Pecotich (1991).

As the λ_{ij}^{OD} values are constrained by equation (3), each λ matrix contains negative values. However, as an MDS data set representing (dis)similarities may not contain negative values, all λ_{ij}^{OD} lambdas were shifted in positive direction with a value $|\lambda_{ij}^{OD}|_{\min}$ prior to submitting them to MDS.

Note that to avoid negative similarities one might suggest using the loglinear model's τ parameters (antilogs of the λ parameters). However, as the τ parameters are constrained by the following equation, this would result in nonlinearities in the similarities:

$$\prod_{i=1}^I \prod_{j=1}^J \tau_{ij}^{OD} = 1. \tag{7}$$

Table 4 shows MDS model fits for all user groups for solutions in two to six dimensions using both metric and nonmetric, Euclidean ALSCAL (Young & Lewyckyj,

TABLE 4
MDS model fits based on (transformed) λ_{ij}^{OD} interactions as similarity

User group and MDS type	Kruskal's F-stress [†] by dimensionality (percentage stress reduction)				
	2	3	4	5	6
All users-metric	0.278	0.207 (25.54)	0.158 (23.67)	0.145 (8.23)	0.141 (2.76)
All users-nonmetric	0.255	0.180 (29.41)	0.126 (30.00)	0.111 (11.90)	0.094 (15.31)
ADR panelists-metric	0.309	0.234 (24.27)	0.186 (20.51)	0.174 (6.45)	0.169 (2.87)
ADR panelists-nonmetric	0.275	0.181 (34.18)	0.137 (24.31)	0.107 (21.90)	0.090 (15.89)
Librarians-metric	0.284	0.202 (28.87)	0.174 (13.86)	0.163 (6.32)	0.158 (3.06)
Librarians-nonmetric	0.258	0.172 (33.33)	0.138 (19.77)	0.110 (20.29)	0.095 (13.64)

[†] $F = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J (f(\delta_{ij}) - d_{ij})^2}{k}}$.

1996). As can be expected, the nonmetric solutions generally yield lower stress values (only the ordinal information in the λ values is used). However, the metric solutions all indicate a stabilization of the stress at three- or four-dimensional configurations; something that facilitates interpretation of the dimensions and comparisons among user groups. As metric scaling uses more of the information in the data than nonmetric scaling, it was decided to base further analysis on the three- or four-dimensional metric solutions.

8. Alexandria Digital Library transaction theory formulation

Figure 3 contains the four-dimensional solutions for the aggregated group of all Alexandria Digital Library users. The solutions are represented in three two-dimensional plots: dimension one plotted against dimension two ($I,2$), three ($I,3$) and four ($I,4$), respectively.

The first research question solicits main patterns of transaction in ADL and essential components of use. Figure 3 shows the four dimensions of use for the aggregated user population (175 606 transactions). Direct examination of the plotted coordinates reveals ready interpretations of the four dimensions of use. The primary dimension (horizontal axis in the plot labeled $I,2$) 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 side are pages of instructions (the *tutorial*, *conventions*, and *access* pages); that is, pages of activities that are system-directed.

The second dimension of use (vertical axis in the plot labeled $I,2$) 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 the content of which is more administrative, telling users about how to access the library, but not necessarily telling about the collection itself.

The third dimension of use (vertical axis in the plot labeled $I,3$) differentiates uses that inform learning from those that do not. Users navigating to 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 (in contrast for example to the *tutorial*).

The fourth dimension differentiates time-on-task for specific activities on specific pages. Transaction logs indicate that when users navigate to *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.

The second research question is directed towards changes in the patterns of transactions through the 13-month period, questioning whether these may be tied to specific interface design modifications. As described earlier, the OD transaction pattern

All Users

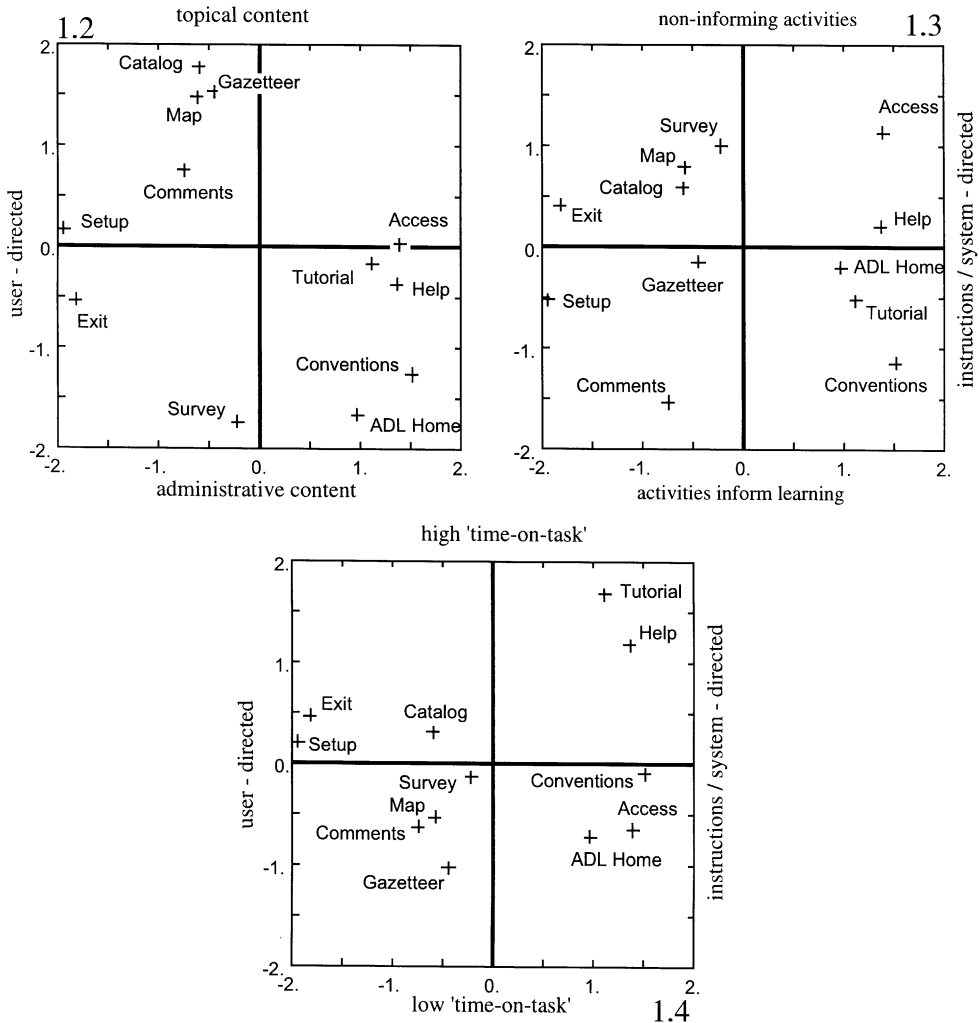


FIGURE 3. Coordinate output for aggregated users of the Alexandria Digital Library (April 1996–May 1997) based on four-dimensional MDS solution (stress=0.158). The dimensions are interpreted in the text.

does not change significantly over time. This result is somewhat problematic, as the interface design team spent considerable effort adjusting and refining various aspects of the look and feel of the interface, based on other methods of soliciting user feedback such as videotape, recorded interviews, focus groups, scenario-building and comments in the transaction logs. Text was modified to eliminate jargon and clarify directions. The wallpaper in various library “rooms” was color-coded to orient users as to *where* they were in the library. In addition, some pages were eliminated, others expanded. In spite of these changes, many of which were requested by users, the pattern of navigating between pages did not change during the 13 months. One reason for this apparent

paradox could be that the aggregated user population confound patterns of use for individual target user groups.

This leads to the third question, focusing on the difference between subgroups of the monitored user population. In the first year of the ADL project, three target user groups were initially defined: earth scientists, librarians and K-12 educators. It should be noted that in addition to a timestamp, each log entry includes an Internet Protocol (IP) address. Examination of the transaction IP domains and the surveys that could be tied to specific transaction logs indicates that there are actually six target user groups: earth scientists; educators (K-12, college and university levels); librarians; system designers; computer scientists and engineers; and a final group of individuals (mostly librarians) who underwent some training in ADL via periodic face-to-face meetings, intensive email and follow-up contact. The people in this group agreed to use the library more intensively than other (casual) users or beta testers and is referred to as the Alexandria Design Review, or ADR panelists. The panel met twice as a group during the 1993–1997 period, and ADL staff met individually with ADR panelists at irregular intervals throughout the project. As a result of this more intensive exposure, it can be said that the ADR panelists had more training, more focused attention of ADL staff, and a higher motivation to learn how to navigate the library.

Earlier, a working hypothesis was mentioned that levels of training should modify use patterns, and indeed, this seems to be reflected in the transaction logs. In comparing the dimensional plots for the librarians (15297 transactions) with plots for the ADR panelists (4579 transactions) in Figure 4, dimensions of use mirror those for the aggregated uses described above. However, the page locations along the third

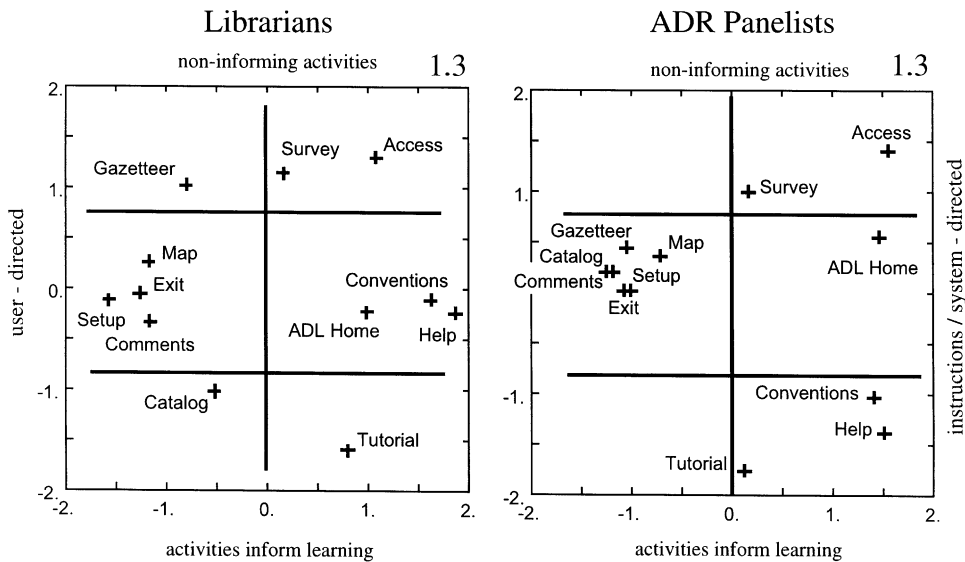


FIGURE 4. Coordinate output for dimensions one and three for two subgroups of users (librarians and ADR Panelists) based on a four-dimensional MDS solution (stress=0.174 and 0.186, respectively). Coordinate differences of pages in the two plots relate to training in system use and are interpreted in the text.

dimension of use show interesting and dramatic differences. Notice that for the (untrained) librarians, the *gazetteer*, *map browser* and *catalog* are widely spaced. The nature of the MDS configuration is that pages spaced far apart are not so frequently the target of single links. That is, the librarians tended to jump from (for example) the *catalog* to some other page, and then perhaps to the *gazetteer*. In the plot of the ADR *panelists*, these three pages are clustered tightly, indicating that the frequency of direct links for this group is much more common. The ADL was designed such that users should jump directly from the *map browser* to the *gazetteer*, back to the *browser*, and then to the *catalog* and back once more to the *browser*. Three walk-through tutorials direct users to follow that precise pattern. Clearly, the ADR panelists, once trained by system designers and evaluators, tended to follow that pattern of usage.

What is clear is that the differences between these two groups are not disciplinary (both contain librarians, spatial data archivists, and those with disciplinary interests related to archiving and disseminating digital geographic information). The difference is not apparent in other dimensions of use (time-on-task, or user vs. system-directed activities). The difference appears in use of pages that inform learning, and those individuals who received training appear to have learned that the shortest path through this extensive web site (the shortest path to access the maps and imagery of interest) is to jump directly between these three “rooms”, iteratively.

9. Discussion

This paper documents the application of interaction models commonly applied in geographical and sociological analysis to the study of user navigation through the Alexandria Digital Library. User-navigation data are parsed from web transaction logs after which the navigation models are established using loglinear testing and visualized as coordinate output from multidimensional scaling. The loglinear models provide an inferential mechanism to sift through all the possible use patterns available in the transaction logs to resolve those patterns which are statistically stable, across target user groups and across a 13-month time period. We feel this is an important contribution to the study of transaction logging data in that it opens the door to induction and theory-building about user navigation through a web site.

Results of the models indicate that for the users as a whole, patterns of use demonstrate consistent structure (users are not moving at random through ADL). Furthermore, the patterns do not change over time, in spite of dramatic changes made to the interface design during the 13 months of study. One might erroneously conclude from this that major interface design revisions were unimportant. A different conclusion seems more constructive, namely that the library in its html incarnation was structurally complex. Modifications to appearance were unable to overcome this complexity in users' navigation paths (and by inference, in users' perceptions). One very major simplification in a new, JAVA-based interface has been to collapse the *catalog*, *gazetteer* and *browser* pages. What the user experiences is a single “search” window. Models of use patterns could establish whether this collapsing of “rooms” simplifies the navigational strategies of users in whole or in subgroups. It is doubtful that use of any large Internet site has been tested across multiple software testbeds, and this remains an avenue of important investigation.

Further analysis of the loglinear user parameters gives a set of multidimensional coordinates with reasonable stress values and four, readily interpretable dimensions of use. These dimensions are evident for the aggregate of all users as well as for selected subgroups, particularly librarians. The first dimension distinguishes user-directed from system-directed activities. The second distinguishes pages accessed to gain topical information from those accessed for administrative uses. The third dimension is a corollary to the second, distinguishing learning-based pages from other library pages. The fourth dimension distinguished web pages based on time-on-task. It should be remembered, however, that these pages are not clustered together because of the length of time users spent on them. Instead, the activities; i.e. the use patterns for the *tutorial*, *help* and similar pages are categorically different from use patterns for more quickly traversed library “rooms” such as the *home page*, *comments* page and *gazetteer*.

Although the above method has generated useful results, it is only the first step of a more comprehensive search for patterns and theory building. The next step would be to use the initial theory presented here to predict a set of specific navigational patterns and to test deductively for the presence of the patterns. Following this approach, the inductive and deductive methods can be used in a complimentary fashion in the search for more complete theories of library navigation.

Finally, as mentioned earlier, the analysis presented here ignores all information contained in the actual sequences or paths of hyperlinks followed by ADL patrons. Instead, the analysis involves patterns of single hyperlinks only; i.e. the under- or over representations of certain links relative to others. Although it was shown that on its own this information provides important insights into navigational behavior, a more complete analysis should take the context of each link—previous and next—into account. It should also be noted that transaction logging provides evaluative information about an interface design as it is deployed. A comprehensive evaluation plan should proceed through the initial planning, implementation, and deployment, and should incorporate other methodologies such as usability engineering in the early stages, to establish a complete portrait of system usability (Buttenfield, 1998). Here too, the results from the inductive analysis can drive the future development of more complex hypotheses that can then be used to predict specific navigational paths rather than single navigational moves.

This paper forms a portion of the Alexandria Digital Library Project, funded in 1994 jointly by NSF, ARPA, and NASA (NSF Grant IRI-94-11330). Matching support from the University of Colorado is gratefully acknowledged.

References

- ANDRESEN, D., CARVER, L., DOLIN, R., FISCHER, C., FREW, J., GOODCHILD, M., IBARRA, O., KOTHURI, R., LARSGAARD, M., MANJUNATH, B., NEBERT, D., SIMPSON, J., SMITH, T., YANG, T., & THING, Q. (1995). The WWW prototype of the Alexandria Digital Library. *Proceedings of the International Symposium on Digital Libraries*, Tsukuba, Japan, WWW at <http://alexandria.sdc.ucsb.edu/public-documents/papers/japan-paper/>.
- AUFHAUSER, E. & FISHER, M. M. (1985). Log-linear modelling of spatial analysis. *Environment and Planning A*, **17**, 931–951.
- BIRBECK, V. P. (1986). Unobtrusive testing of public library reference service. *Refer*, **4**, 5–9.

- BORGMAN, C. L. (1995). Online monitoring methods for user-based design and evaluation for digital libraries. *Proceedings, 37th Allerton Institute, Graduate School of Library and Information Science*, University of Illinois, Champaign-Urbana, 29–31 October.
- BUTTENFIELD, B. P. (1995). Evaluating user requirements for a digital library testbed. *Proceedings, AUTO-CARTO 12*, pp. 207–214, Charlotte, NC, 27 February–1 March.
- BUTTENFIELD, B. P. (1996). Scientific visualization for environmental modeling: interactive and proactive graphics. In M. F. GOODCHILD, L. T. STEYAERT, B. O. PARKS, C. JOHNSON, D. MAIDMENT, M. CRANE & S. GLEDENNING, Eds. *GIS and Environmental Modeling*, pp. 463–469. Fort Collins, CO: GIS World Books.
- BUTTENFIELD, B. P. (1997). Delivering maps to the information society: a digital library for cartographic data. *Proceedings, 17th conference of the International Cartographic Association*, pp. 1409–1416, Stockholm, Sweden, June.
- BUTTENFIELD, B. P. (1998). Usability evaluation of digital libraries. *Science and Technology Libraries*, **17**, 39–59.
- BUTTENFIELD, B. P. & GOODCHILD, M. F. (1996). The Alexandria Digital Library project: distributed library services for spatially referenced data. *Proceedings GIS/LIS'96*, pp. 76–84, Denver, CO, November.
- BUTTENFIELD, B. P. & KUMLER, M. P. (1996). Tools for browsing environmental data: the Alexandria Digital Library interface. *Proceedings, Third International Conference on Integrating Geographic Information Systems and Environmental Modeling*, Santa Fe, NM, 21–26 January, WWW http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD.ROM/sf_papers/buttenfield_babs/babs_paper.html.
- BUTTENFIELD, B. P. & WEBER, C. R. (1994). Proactive graphics for exploratory visualization of biogeographical data. *Cartographic Perspectives*, **19**, 8–18.
- CHRISTEL, M. G. & PENDYALA, K. (1996). Informedia goes to school: early findings from the digital video library project. *D-Lib Magazine*, September 1996. Web at <http://www.dlib.org/dlib/september96/informedia/09christel.html>.
- COURY, B. G. (1991). Probing the mental models of system state categories with multidimensional scaling. *International Journal of Man-Machine Studies*, **36**, 637–696.
- DEMARIS, A. (1992). *Logit Modeling, Practical Applications, Series: Quantitative Applications in the Social Sciences*, No. 86. Newbury Park, CA: Sage Publications.
- DIPRETE, T. A. (1990). Adding covariates to loglinear models for the study of social mobility. *American Sociological Review*, **55**, 757–773.
- EVERETT, J. A. & PECOTICH, A. (1991). A combined loglinear/MDS model for mapping journals by citation index. *Journal of the American Society of Information Science*, **42**, 405–413.
- GAY, G. & MAZUR, J. (1993). The utility of computer tracking tools for user-centered design. *Educational Technology*, **1**, 45–59.
- GAY, G., MAZUR, J. & LENTINI, M. (1994). The use of hypermedia data to enhance design. *Computer Graphics*, **28**, 34–37.
- KNOKE, D. & BURKE, P. J. (1980). *Loglinear Models, Series: Quantitative Applications in the Social Sciences*, No. 20. Beverly Hills, CA: Sage Publications.
- KRUSKAL, J. B. & WISH, M. (1978). *Multidimensional Scaling, Series: Quantitative Applications in the Social Sciences*, No. 1. Newbury Park, CA: Sage Publications.
- LANDAUER, T. K. (1995). *The Trouble with Computers. Usefulness, Usability, and Productivity*. Cambridge, MA: MIT Press.
- MILLER, R. L. & HAYES, B. C. (1990). Female marital mobility within Northern Ireland and the Republic of Ireland. *Sociological Review*, **38**, 111–133.
- NEUMANN, L. J. & BISHOP, A. P. (1998). From usability to use: measuring the success of testbeds in the real world. *Proceedings of the Annual GSLIS Clinic*, Urbana, IL, 12–14 April.
- NIELSEN, J. (1993). *Usability Engineering*. Boston, MA: Academic Press.
- NSF (1996). *National Science Foundation NSF/ARPA/NASA Digital Libraries Initiative*. Washington, D.C.: National Science Foundation.
- PENNIMAN, W. D. & DOMINICK, W. D. (1980). Monitoring and evaluation of online information system usage. *Information Processing and Management*, **16**, 17–35.

- PETERS, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, **11**, 41–66.
- PETERS, T., KURTH, M., FLAHERTY, P., SANDORE, B. & KASKE, N. K. (1993). An introduction to the special section on transaction log analysis. *Library Hi Tech*, **11**, 38–40.
- RAFTERY, A. (1986). Choosing models for cross-classifications. *American Sociological Review*, **51**, 145–146.
- RAFTERY, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, **25**, 111–163.
- RHODES, A. K. & STERN, S. E. (1993). Ranking harassment: a multidimensional scaling of sexual harassment scenarios. *Journal of Psychology*, **129**, 29–39.
- RICE, R. E. & BORGMAN, C. L. (1983). The use of computer-monitored data in information science and communication research. *Journal of the American Society for Information Science*, **34**, 247–256.
- RUBIN, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: John Wiley and Sons.
- SCHOLTEN, H. J. (1982). Describing and forecasting migration within a housing market area: the loglinear model used as a spatial interaction model. *Geografische en Planologische Notities*, No. 22, Amsterdam: Free University of Amsterdam.
- UPTON, G. J. G. & FINGLETON, B. (1979). Loglinear models in geography. *Transactions of the Institute of British Geographers*, **5**, 103–115.
- WILDEMUTH, B. M. (1993). Post-positivist research: two examples of methodological pluralism. *Library Quarterly*, **63**, 450–468.
- WILDEMUTH, B. M., DE BLIEK, R. & FRIEDMAN, C. (1993). Measures of searcher performance: a psychometric evaluation. *Information Processing and Management*, **29**, 533–550.
- WILDEMUTH, B. M., DE BLIEK, R., HE, S. & FRIEDMAN, C. (1992). Search moves made by novice end users. *Proceedings, 29th Conference of the American Society of Information Science*, pp. 154–161, Pittsburgh, PA, 26–29 October.
- WILLEKENS, F. J. (1983). 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**, 239–252.
- WILLEKENS, F. J. & BAYDAR, N. (1983a). *Hybrid-loglinear models*. Working Papers of the N.I.D.I., No. 41, Netherlands International Demographic Institute, Voorburg.
- WILLEKENS, F. J. & BAYDAR, N. (1983b). *Forecasting place to place migration with generalized linear models: an application to urbanization in the Netherlands*. Working Papers of the N.I.D.I., No. 42, Netherlands International Demographic Institute, Voorburg.
- YOUNG, F. W. & LEWYCKYJ, R. (1996). *ALSCAL User's Guide*. Chapel Hill, NC: L.L. Thurstone Psychometric Laboratory, University of North Carolina.