

## An Application of Graphical Modeling to the Analysis of Intranet Benefits and Applications

Raffaella Settimi, Linda V. Knight,  
Theresa A. Steinbach and James D. White  
*DePaul University*

*Abstract:* Applications of multivariate statistical techniques, including graphical models, are seldom found in e-commerce studies. However, as this paper demonstrates, we find that probabilistic graphical models are useful in this area, both because of their ability to handle large numbers of potentially interrelated variables, and because of their ability to communicate statistical relationships clearly to both the researcher and the ultimate business audience. We show an application of this methodology to intranets, internal corporate information systems employing Internet technology. In particular, we study both the interrelationships among intranet benefits and the interrelationships among intranet applications. This approach confirms some hypothesized relationships, and uncovers heretofore-unanticipated relationships among intranet variables, providing guidance for business professionals seeking to develop effective intranet systems. The techniques described here also have potential applicability in other e-commerce arenas, including business-to-consumer and business-to-business applications.

*Key words:* Contingency tables, graphic models, log-linear models.

### 1. Introduction

This paper presents an application of graphical modeling to the analysis of data on features and benefits of corporate intranets. Graphical modeling provides an efficient statistical tool for the analysis of our data, both because of the ability of representing complicated relationships among variables in an intuitive and easy way using graphs, and because of the availability of efficient computational algorithms for parameter estimation and model selection.

Consider now a brief introduction to the intranet technology. An intranet is an internally focused application of Internet technology used to disseminate information within a company (Riggins and Rhee, 1998). Driven by promises of wide ranging benefits, from cost savings through fundamental business process improvement, companies are aggressively embracing intranet technology. Following its introduction in 1994, the number of companies implementing corporate

intranets has grown dramatically. More than three-quarters of all Web servers being installed are for intranet purposes, and the market for intranet applications, platforms and related technology is substantially outstripping that for the public Internet and all other IT areas (Gerstner, 2002).

A comprehensive list of intranet benefit variables has never been definitively identified. The information technology practitioner press is filled with anecdotal reports of a variety of benefits realized by companies that have implemented intranets, including monetary savings, improved productivity, enhanced customer service, and improved customer relationships. Much of the relevant research, however, addressed intranet benefits tangentially as the byproduct of studies that were focused on other issues. In six studies conducted between 1998 and 2001, categories of intranet benefits included increased productivity, reduced costs, improved customer service, improved data access, generation of new business, and improved internal communications (Bhattacharjee, 1998; Ptak, 1998; Bidgoli, 1999; Baker, 2000; Leung, 2001; Lai, 2001). Lai (2001), working from survey data collected in 1996 when intranets were first emerging, was the first to analyze the relationship between benefit variables and concluded that data access and cost savings were the most important benefits.

Despite the number of organizations that have installed intranets, there is little empirical evidence suggesting that expected benefits are being realized. Furthermore, relationships between benefit variables have not been explored in-depth. This study addresses these deficits.

In this paper we report the results of the statistical analysis of data collected in a survey on the main features and the characteristics of the benefits associated with corporate intranets. Section 2 presents the data collection methods, briefly describing the survey and the variables that are analyzed in this paper. The survey data are analyzed using graphical models, a class of multivariate statistical methods that are effectively applied to the analysis of problems with a large number of variables. Graphical models are estimated using computationally efficient algorithms. They provide a graphical representation of the associations between variables, allowing an easier interpretation of the model assumptions. In Section 2.3 we present the basic definitions and results for undirected graphical models, and in Section 2.4 we discuss the parameter estimation techniques and the model selection methods that are used to analyze the survey data. The results of the statistical analysis are discussed in Section 3.

## **2. Methods**

### **2.1 Data collection**

Between late 2000 and mid-2001, a survey of 110 working professionals was

conducted to examine the key aspects and characteristics of the benefits associated with implementation of corporate intranet systems. The survey consisted of questions regarding three main areas: (1) technical characteristics and features of the corporate intranets, (2) the benefits and the advantages of having an intranet system, and (3) general information about the company such as number of employees, type of industry, and the year the intranet was initiated. The respondents to the survey were employed by companies representing a broad range of industries, including, among others, IT consulting, financial services, education, manufacturing, telecommunications, and health care. About one-third of the respondents held positions in Information Technology departments, while the remainder was equally distributed in the sales and marketing sector, and in the financial and accounting sector.

The group of subjects involved in the survey formed a convenience sample that was selected among IT professionals attending a development program on Web Commerce. This program was offered through a major university in the Midwestern USA. Students attended several lectures that varied from e-commerce technology and environment to Web technology infrastructures and corporate intranet systems. The survey was completed after attending a lecture on corporate intranets. Therefore, the respondents completing the survey had a common understanding of the meaning of the terms used in the questions, regardless of their different background and experience. Notice that, although the sample of respondents was not randomly selected, the survey is still a source of valuable information, providing data for exploratory studies on intranet implementation. Results of these studies allow researchers to gain significant insights on the intranets benefits.

Data were collected from 68 completed surveys. Because earlier research had been limited and prior studies did not build upon each other, the resulting list of potential intranet benefits and intranet applications was somewhat burdensome and redundant. Consequently, a concise number of variables of interest were identified for the survey and subsequent statistical analysis. Further details on the survey can be found in Knight *et al.* (2003). We decided to analyze two main sets of variables separately with the groups determined according to the researchers' interests.

## 2.2 Variables

The first selected set of variables, named "intranet benefits", describes the benefits accruing from the implementation of corporate intranets. The binary variables (Yes or No) record the answers to the following question: "What have been the significant benefits of your corporate intranet?". The list of variables is: (SA) increased savings, (AI) faster access to information, (QI) better quality

of information, (WE) more cooperative work environment, (PA) reduced paper work, (CS) improved customer service/relationships, and (PR) improved productivity.

The second set of variables, named “intranet applications”, describes the primary applications implemented on corporate intranets. The binary variables (Yes or No) record the answers to the question “What applications are on your corporate intranet?”. The list of variables is: (HR) human resources, (CS) customer service, (OE) order entry/tracking, (IN) inventory, (PC) production control, (SA) sales, (MA) marketing, and (AC) accounting.

For each set of variables, the table of counts is rather sparse, thus some of the table cells contain zero observations. This complicates the data analysis, since some combination of variable levels is systematically missing, and therefore never observed. We were unable to conclude if the empty cells could be considered “structural zeroes”, meaning that the corresponding classifications had a zero-probability of occurrence. We did not wish to set such strong constraints on our model, so we regarded the empty cells as systematically missing data. For this reason, the statistical inference using graphical models is very efficient. In fact, graphical models have a modular structure that enable us to decompose the data analysis into the analysis of smaller marginal tables. This is very helpful when the tables of data are sparse, as in this case. Typically the marginal tables that are defined in the decomposition, involve a smaller subgroup of variables, and therefore have cells with larger counts. Thus, the marginal tables are less likely to have empty cells.

### 2.3 Undirected graphical models

In this paper we present an application of graphical modeling to the analysis of survey data on the association structure between benefits accrued from the implementation of corporate intranets. Graphical models are multivariate statistical models that are often used to describe complicated problems involving a large number of variables. They assume a joint probability distribution among the observed variables that satisfy some independence relations that can be represented through a graph.

Although this statistical approach is relatively new, its origins may be found in path analysis for the analysis of genetics problems (Wright, 1934) and later applied in economics and social sciences (Wold, 1954). Only in the last decade, however, has the development and the properties of statistical graphical models been formally studied (Whittaker, 1990; Lauritzen, 1996; Cox and Wermuth, 1996). The investigation of this methodology remains an area of active research and the application of graphical models to the analysis of complex multivariate problems has begun to appear mainly in the social and medical sciences (see for

example Klein, Keiding and Kreiner (1995), Freidl, W (1997) and Gopnik A., Sobel D. M. and Schulz L. E. (2001)).

In addition to being computationally very efficient, graphical models have the advantage of providing a compact and natural representation of the relationships among large sets of variables and visually representing the probabilistic properties of the fitted model through a graph. This facilitates the interpretation of the model assumptions as well as communication between the statistician and the researcher, and the researcher and the target audience. Thus, in several cases, graphical modeling is preferable to more traditional data analysis methods. This is true especially when the data analysis involves a large number of variables and traditional models contain many parameters, making the interpretation of the model assumptions almost intractable.

We will now briefly define the central concepts and the main results in graphical modeling, that will be used in this paper. The notation is the one adopted in the monograph by Lauritzen (1996). The readers are referred to Lauritzen (1996) for further details on graphical modeling.

Let  $V = \{Y_1, \dots, Y_k\}$  be a set of discrete random variables with a finite number of states, that are denoted for convenience by  $1, \dots, r_j$ ,  $j = 1, \dots, k$ . The data to be analyzed consist of  $N$  observations on the  $k$  variables in  $V$ . The data are represented as a  $k$ -dimensional contingency table of counts, formed by cross-classifying the  $k$  variables. Each cell in the table is specified as a  $k$ -tuple  $i = (i_1, \dots, i_k)$ . Such cell will represent the observation such that  $Y_1 = i_1, \dots, Y_k = i_k$ . The number of observations in the cell  $i$  is written as  $n(i)$  and the probability of observing cell  $i$  is  $p(i) = p(Y_1 = i_1, \dots, Y_k = i_k)$ . The contingency table of counts is defined as  $\mathcal{N} = \{n(i)\}_{i \in \mathcal{I}}$ , where  $\mathcal{I}$  is the set of all cells in the table.

There are several classes of graphical models that can be used to identify the relationships among the variables in  $V$ . In this paper we consider undirected graphical models because of the simplicity of their structure. An *undirected graphical model*  $U = \{\mathcal{G}, \mathcal{P}_U\}$  consists of two components. The first is an undirected graph  $\mathcal{G} = \mathcal{G}(V, E)$  with set of nodes  $V$  and set of undirected ('—') edges  $E$ , that connect pairs of distinct nodes in  $V$ . Thus, each vertex in the graph corresponds to a random variable. The second component is a family of probability distributions  $\mathcal{P}_U$  on the nodes of the graph that satisfy the set of conditional independence properties represented by the graph  $\mathcal{G}$ . The probability distribution is specified through the cell probabilities  $\{p_U(i)\}_{i \in \mathcal{I}}$ . The probability distribution  $\mathcal{P}_U$  defined by the graphical model  $U$  is such that each cell probability  $p_U(i)$  can be decomposed into a product of non-negative functions  $\psi_a(i_a)$  that depend on complete sets  $a$  of variables in  $V$ . A set  $a$  of variables in  $V$  is said to be *complete* if each pair of nodes in  $a$  is connected by an edge in the graph  $\mathcal{G}$ . The cells  $i_a$  are the cells of the marginal tables obtained by classifying the observations only by

the variables in  $a$ . The probability decomposition is expressed as follows:

$$p_U(i) = \prod_{a:a \subset V} \psi_a(i_a) \quad (2.1)$$

for all complete sets  $a$  in  $V$ . A discrete probability distribution  $\mathcal{P}_U$  that satisfies the property in (2.1) is said to be Markov with respect to the undirected graph  $\mathcal{G}$ .

The factorization of the probability distribution in (2.1) is a key property of graphical modelling. This permits us to perform locally the statistical computations using a modular approach. Large contingency tables of counts are decomposed into smaller marginal tables, and each marginal table can be analyzed separately. Such a decomposition is defined considering all the marginal tables relative to the complete subsets  $a$  of variables in  $V$ .

The probability distribution  $\mathcal{P}_U$  on  $V$  satisfies the set of conditional independence assumptions that are encoded in the undirected graph  $\mathcal{G}$ . These conditional independence relationships can be easily read off the graph by using the separation Markov property that is defined as follows. Let  $A, B$  and  $C$  be three disjoint subsets of  $V$  such that if  $C$  separates  $A$  from  $B$  in  $\mathcal{G}$ , then each variable in  $A$  is conditionally independent of the variables in  $B$ , given the variables in  $C$ . A subset  $C$  is said *to separate*  $A$  from  $B$ , if, for any node  $v$  in  $A$  and  $w$  in  $B$ , any sequence of adjacent vertices in  $V$  that connects  $v$  to  $w$  intersects  $C$ .

Using a well known result stated in the Hammersley and Clifford theorem (Lauritzen 1996, page 36), the probability distribution  $\mathcal{P}_U$  defined by the factorization in equation (2.1), can be represented using a log-linear parameterization. So each cell probability  $p_U(i)$  can be written as below

$$\log p_U(i) = \sum_{a:a \subset V} \log \psi_a(i_a) =: \sum_{a:a \subset V} \phi_a(i_a) \quad (2.2)$$

where  $a$  is a subset of  $V$ ,  $\phi_a(i_a)$  is a function that depends on the variables in  $a$  only, and  $\phi_a(i_a) \equiv 0$  unless  $a$  is a complete subset of  $V$ .

When the number of variables is large, typically more than five variables, the log-linear parameterization may contain several parameters, making the interpretation of the model assumptions very complicated. Thus the use of a graph to depict the model conditional independence properties becomes extremely helpful.

## 2.4 The model selection and estimation

In graphical modelling there exist efficient algorithms for the estimation of the model parameters and for the model selection. Exploiting the factorization (2.1) of the probability distribution  $\mathcal{P}_U$ , the marginal probability model  $p_a$  on the

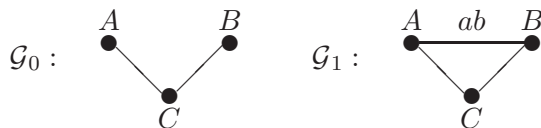
variables in a complete set  $a$  in  $V$  can be estimated separately using the maximum likelihood methods. The maximum likelihood estimates  $\hat{p}(i_a)$  are computed using the iterative proportional scaling algorithm described in Darroch and Ratcliff (1972). The normal equation is defined as

$$\hat{p}(i_a) = n(i_a)/N(I_a) \quad \text{for } i_a \in I_a$$

for each complete set  $a$  in  $V$ . The marginal counts  $n(i_a)$  are the cell counts in the contingency table for the variables in  $a$ ,  $I_a$  is the set of all counts in that table and  $N(I_a) = \sum_{i_a} n(i_a)$  is the total number of observations in the marginal table  $I_a$ .

In order to find the “best” predictive model, we use a backward selection procedure. We start with an initial model and edges are successively removed from it. Each edge is removed using an exact test criterion. The use of exact tests is motivated by the fact that the data set is relatively small. In fact, alternative asymptotic selection criteria such as  $\chi^2$ -tests on the model deviance rely on the assumption of large data sets (see for instance Lancaster, 1969). Such tests may not be reliable if applied to sparse tables of data with small or zero counts.

We will now describe in detail the model selection test used in this paper. Let  $A$  and  $B$  be two adjacent variables in the graph  $\mathcal{G}$  associated with the graphical model  $U = (\mathcal{G}, P_U)$ . Suppose we want to test if  $A$  and  $B$  are conditionally independent given the remaining variables  $C$ . In graphical terms, this is equivalent to a test that compares the null model  $M_0$  versus the full model  $M_1$ , with graphs  $\mathcal{G}_0$  and  $\mathcal{G}_1$  given below. Thus, the statistical test on the conditional independence assumption that  $A$  and  $B$  are independent given  $C$  is expressed as a test on the removal of the edge  $ab = e(A, B)$  that connects the variables  $A$  and  $B$  in  $V$ .



The likelihood ratio statistic for the test comparing the model  $M_0$  versus  $M_1$  is expressed as

$$G^2 = 2 \sum_{i_A=1}^{n_A} \sum_{i_B=1}^{n_B} \sum_{i_C=1}^{n_C} n(i_A, i_B, i_C) \log \frac{n(i_A, i_B, i_C)n(i_C)}{n(i_A, i_C)n(i_B, i_C)}$$

Such a test statistic is often called the deviance of  $M_0$  versus  $M_1$ . For small data sets, the probability distribution of the likelihood ratio statistic is typically computed using numerical algorithms that employ Monte Carlo sampling techniques. Such techniques are implemented in the software package MIM (Edwards,

2000). A backward selection procedure was used to select the “best” model to fit the data. Starting from an initial model, edges were removed one at a time. The edge whose test statistic  $G^2$  has the largest not significant  $p$ -value is removed. The procedure stops when there are no edges whose test statistic has a large  $p$ -value. Typically a significance level is chosen and edges are removed if the test  $p$ -values are larger than the selected significance level.

The next section will show the application of undirected graphical models to the analysis of the data sets collected in the survey on intranet features and benefits.

### 3. Results of the Analysis

#### 3.1 Intranet benefits

We first examined the responses relative to intranet benefits. The goal of the analysis is to understand the perceived benefits of corporate intranets in the work environment. The chart in Figure 1 shows the distribution of the responses given in the survey.

The data analysis started by using a full undirected graphical model on the variables representing the significant benefits of the intranets. A full undirected graphical model assumes that all the variables are related to each other. The corresponding undirected graph is complete, that is each pair of nodes is linked by an edge.

To improve the efficiency of the selection algorithms, we asked the researchers to identify a few edges that could be eliminated, based on their prior knowledge. The edges to be removed were elicited by identifying pairs of variables that could be assumed not to be directly associated. This preliminary step allowed us to simplify the model by assuming some conditional independence associations among the variables. This process makes the selection algorithm more efficient, since the model is not full and can be decomposed into smaller submodels. The selection algorithm can then be applied separately to the submodels.

The application of the backward selection algorithm gives rise to the graphical model with graph  $\mathcal{G}_B$  shown in Figure 2. The undirected graphical model is selected using the exact methods discussed in Section 2.4. Table 1 shows the exact  $p$ -values associated with each edge removal test.

The fitted graphical model with graph  $\mathcal{G}_B$  can be used to understand how users perceive the benefits associated with corporate intranets. The analysis of the conditional independence associations represented in the graph  $\mathcal{G}_B$  gives insights on the associations among the variables in the survey. Using the Markov property of the estimated probability model, conditional independence assumptions can easily be read off the graph by using the separation property defined in Section 2.3. Whenever two variables are separated by a set of variables in the



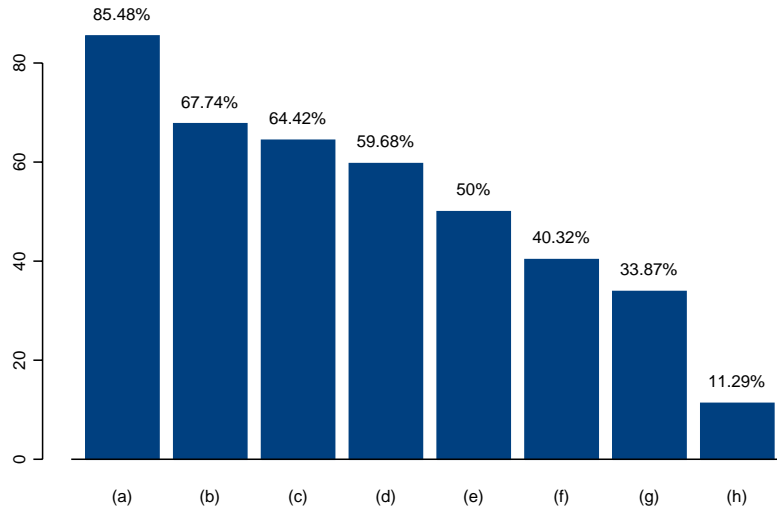


Figure 1: Chart of the response of intranet benefits. (a)= Faster access to information, (b)= savings on paper, telephone, etc., (c)= better quality of information, (d)= reduced paper work, (e)= improved productivity, (f)= improved customer service/relationships. (g)= more cooperative environment, (h)= other.

graph, then the two variables are conditionally independent given that separator.

The estimated conditional probabilities that are computed from the fitted graphical model are shown in Tables 2, 3, 4 and 5. The statistical analysis shows that higher savings are most often associated by the users of corporate intranets with improved customer service and reduced paperwork. This is supported by a common belief in Customer Relationship Management, a business theory where better managed customer relationships may lead to real savings and decreased sale costs (Karimi *et al.*, 2001).

Intuitively, the researchers hypothesized that improved productivity would be associated with better quality of information, faster access to that information, and reduced paperwork. However, the conditional independence assumptions in the fitted graphical model indicate that, given better quality of the information (QI), improved productivity (PR) is independent of faster access (AI) to that information and reduced paperwork (PA).

Moreover, the estimated values of the conditional probabilities in Table 3 show that corporate intranets are perceived to be beneficial for improved productivity

whenever the intranets are regarded as providing better quality information, improved customer service/relationship and a more cooperative work environment.

Another interesting result of the analysis shows that faster access to information is perceived as strongly associated with reduced paperwork and also with the quality and the relevance of the information available on the intranets. This confirms the researchers' intuitive hypothesis.

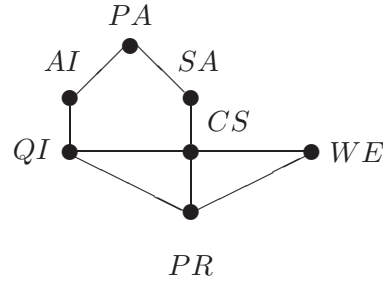


Figure 2: Graph  $\mathcal{G}_B$  representing the relationships among intranet benefits

Legend		
SA	=	Increased savings in paper, telephone, etc
AI	=	Faster access to information
QI	=	Better quality of information
WE	=	More cooperative work environment
PA	=	Reduced paperwork
PR	=	Improved productivity
CS	=	Improved customer service

Table 1: Exact  $p$ -values of the edge removal tests

Edge	$p$ -value	Edge	$p$ -value
PA $\bullet$ — $\bullet$ SA	< 0.001	CS $\bullet$ — $\bullet$ WE	0.003
PA $\bullet$ — $\bullet$ AI	0.023	QI $\bullet$ — $\bullet$ PR	0.006
AI $\bullet$ — $\bullet$ QI	0.008	SA $\bullet$ — $\bullet$ CS	< 0.001
PR $\bullet$ — $\bullet$ WE	< 0.001		

**Conditional probability tables estimated from the data on intranet benefits** (Tables 2 – 5)

Table 2: Conditional probabilities  $p(CS|WE, PR, QI)$ 

	WE	No				Yes			
		No		Yes		No		Yes	
		PR	QI	PR	QI	PR	QI	PR	QI
CS	No	0.894	0.706	1.000	0.400	0.000	0.000	1.000	0.372
	Yes	0.106	0.294	0.000	0.600	1.000	1.000	0.000	0.628

Table 3: Conditional probabilities  $p(SA|PA, CS)$ 

	SA	PA		CS	
		No		Yes	
		No	Yes	No	Yes
SA	No	0.737	0.218	0.239	0.030
	Yes	0.263	0.782	0.761	0.970

Table 4: Conditional probabilities  $p(PR|WE, QI)$ 

	PR	WE		QI	
		No		Yes	
		No	Yes	No	Yes
PR	No	0.870	0.493	0.333	0.157
	Yes	0.130	0.507	0.667	0.843

Table 5: Conditional probabilities  $p(AI|QI)$ 

	AI	QI	
		No	Yes
AI	No	0.318	0.050
	Yes	0.682	0.950

### 3.2 Intranet applications

The second data set contains observed responses concerning the type of applications available on the intranet.

The histogram in Figure 3 shows the percentage of the observed responses. More than 87% of the corporate intranets in the survey contained the human resources application. This may indicate that the need for a human resources

application is common across most of the companies surveyed, or that most companies seem to implement the human resources site first. Other applications appeared less frequently. The marketing application may be found on about 44% of the corporate intranets, while about one-third have accounting and customer service applications. Inventory and production control applications were implemented in less than 8% of corporate intranets. This is consistent with the mix of industries represented in the survey, which included IT consulting, financial services, education, telecommunications, energy, healthcare, retail, and manufacturing. It is conceivable that production control and inventory are significant systems in a minority of these industries.

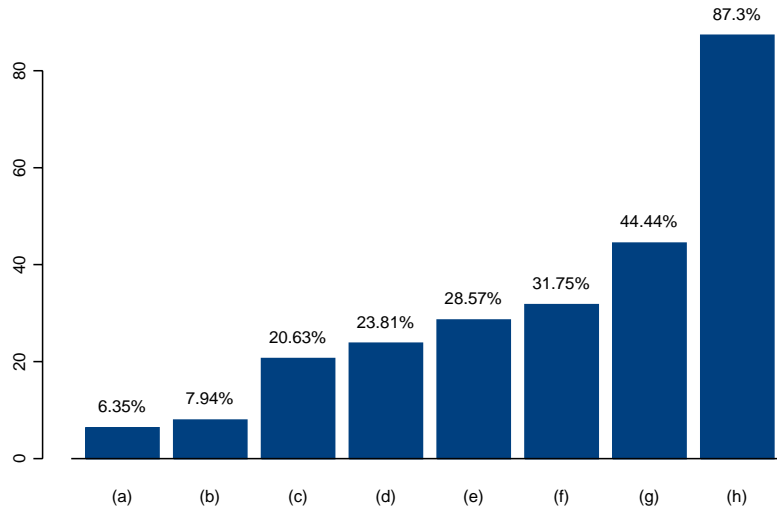


Figure 3: Chart of the survey data on intranet applications. (a)= Production control, (b)= inventory, (c)= sales, (d)= other entry/tracking, (e)= customer service, (f)= accounting, (g)= marketing, (h)= human resources.

The selected undirected graphical model for the data on intranet applications is displayed in Figure 4. Table 6 shows the exact  $p$ -values for the significant edges that were computed by the edge removal test in the model selection algorithm. Edges with  $p$ -values higher than 0.05 were removed.

The estimated conditional probabilities for the fitted graphical model are displayed in Tables 7, 8, 9, 10 and 11. The statistical analysis of the data shows that the order tracking application is often associated with the customer service application. Moreover, even though the production control application is found

in a small sample of intranets, it is always associated with customer service. The sales application is implemented more often in intranets that have accounting and marketing applications.

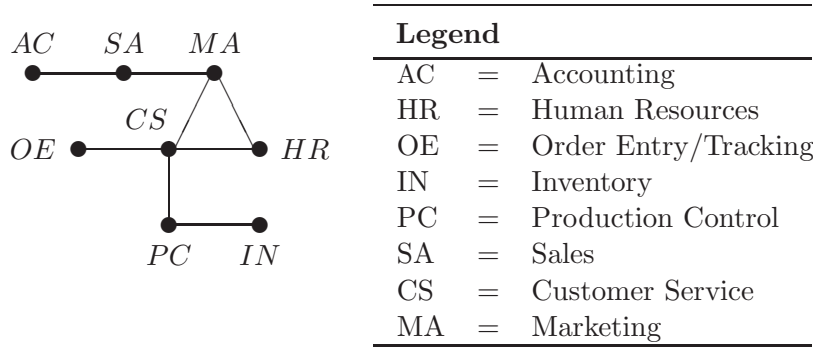


Figure 4: Graph  $G_A$  representing the relationships among intranet applications

Table 6: Exact  $p$ -values of the edge removal tests

Edge	$p$ -value	Edge	$p$ -value
AC ●—● SA	0.017	SA ●—● MA	< 0.001
MA ●—● CS	0.009	MA ●—● HR	0.003
CS ●—● PC	0.005	CS ●—● HR	0.024
PC ●—● IN	0.001	OE ●—● CS	< 0.001

**Conditional probability tables estimated from the data on intranet applications (Tables 7 – 11)**

Table 7: Conditional probabilities  $p(OE|CS)$

		CS	
		No	Yes
OE	No	0.889	0.444
	Yes	0.111	0.556

Table 8: Conditional probabilities  $p(AC|SA)$

		SA	
		No	Yes
AC	No	0.760	0.385
	Yes	0.240	0.615

Table 9: Conditional probabilities  $p(HR|CS, MA)$ 

MA		No		Yes	
CS		No	Yes	No	Yes
HR	No	0.036	0.143	0.353	0
	Yes	0.964	0.857	0.647	1

Table 10: Conditional probabilities  $p(PC|IN)$ 

CS		No	Yes
PC	No	0.983	0.400
	Yes	0.017	0.600

Table 11: Conditional probabilities  $p(PC|CS)$ 

CS		No	Yes
PC	No	1	0.778
	Yes	0	0.222

The results of the analysis in Table 9 show that the human resources and the marketing applications are not likely to be found together on the same corporate intranet. In fact, the human resources application is found in about 75% of the observed intranets that do not have a marketing application, and the marketing application is found in about 94% of the intranets that do not have a human resources application. So, it is more likely that the marketing application is not implemented if the human resources application is implemented. The researchers postulate that this is because some organizations initiate their intranet development process with an internal focus, beginning with the employee-oriented human resources application, whilst others initiate their intranet development process with an external focus, beginning with more customer-oriented marketing applications.

#### 4. Discussion

The contribution of this study is to uncover relationships among perceived intranet benefits and their associated applications. Specifically, the results of the statistical analysis upheld the researchers' hypotheses in some cases, whilst uncovering new relationships in other cases.

The role of graphical modeling in the statistical analysis has proven very beneficial, in particular for providing a very efficient tool of communication and

interpretation of the results of the analysis. This research has been useful in identifying relationships among variables that will form the focus of future studies. Specifically, the results about the intranet benefits can be used for future longitudinal studies that investigate in depth the strength of the relationships between savings, which is the variable that directly affects corporate profits, and the other variables.

A more extensive study is needed to investigate the relationship between the benefit variables and the application variables. The results presented here have established that organizations appear to begin intranet development efforts with one of two applications: human resources or marketing. It would be valuable for organizations to know if there is a difference in the benefits depending upon the initial application. As part of this broader survey, we may also want to seek more representative demographics across industries.

Beyond this case, the undirected graphical modeling techniques applied here also have potential application in other e-commerce related research areas involving large numbers of variables with relatively unknown relationships. For example, such techniques could be employed in studies of critical success factors in business-to-consumer (B2C) Web-based information systems, or in studies of business-to-business (B2B) supply chain management systems.

## References

- Baker, S. (2000). Getting the most from your intranet and extranet strategies. *The Journal of Business Strategy* **21**, 40-43.
- Bhattacharjee, A. (1998). Management of emerging technologies: Experiences and lessons learned at US West. *Information and Management* **33**, 263-272.
- Bidgoli, H. (1999). An integrated model for introducing intranets. *Information Systems Management* **16**, 78-87.
- Cox, D. R. and Wermuth, N. (1996). *Multivariate dependencies — Models, Analysis and Interpretation*. Chapman and Hall.
- Freidl, W (1997). The impact of anomia as a factor in a demand resource model of health. *Social Science and Medicine* **44**, 1357-65.
- Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics* **43**, 1470-1480.
- Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer-Verlag.
- Gerstner, J. (2002). Intranets mean business. *Communication World* **19**, 14-17.
- Gopnik, A., Sobel, D. M. and Schulz, L. E. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology* **37**, 620-629.

- Karimi, J., Somers, T. M. and Gupta, Y. P. (2001). Impact of information technology management practices on customer service. *Journal of Management Information Systems* **17**, 125-158.
- Klein, J. P., Keiding, N. and Kreiner, S. (1995). Graphical models for panel studies, illustrated on data from the Framingham heart study. *Statistics in Medicine* **14**, 1265-1290.
- Knight, L. V., Steinbach, T. A. and Settimi, R. (2003). An exploratory analysis of intranet benefits. Submitted to *Information Resources Management Journal*.
- Lai, V. (2001). Intraorganizational communication with intranets. *Communications of the ACM* **44**, 95-100.
- Lancaster, H. O. (1969). *The chi-squared distribution*, Wiley and Sons.
- Lauritzen, S. L. (1996). *Graphical Models*, Oxford University Press.
- Leung, H. K. N. (2001). Quality metrics for intranet applications. *Information and Management* **38**, 137-152.
- Ptak, R. (1988). Designing a business-justified intranet project. *Information Systems Management* **15**, 13-19.
- Riggins, F. J. and Rhee, H. S. (1998). Toward a unified view of electronic commerce. *Association for Computing Machinery, Communications of the ACM* **41**, 88-95.
- Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Wiley.
- Wold, H. O. A. (1954). Causality and Econometrics. *Econometrica* **28**, 443-463.
- Wright, S. (1934). The methods of path coefficients. *Annals of Mathematical Statistics* **5**, 161-215.

Received May 14, 2003; accepted October 6, 2003.

Raffaella Settimi  
School of Computer Science  
Telecommunications and Information Systems  
DePaul University  
243 S. Wabash Ave.  
Chicago, IL 60604, USA  
rsettimi@cs.depaul.edu

Linda V. Knight  
School of Computer Science  
Telecommunications and Information Systems  
DePaul University  
243 S. Wabash Ave.  
Chicago, IL 60604, USA



Theresa A. Steinbach  
School of Computer Science  
Telecommunications and Information Systems  
DePaul University  
243 S. Wabash Ave.  
Chicago, IL 60604, USA

James D. White  
School of Computer Science  
Telecommunications and Information Systems  
DePaul University  
243 S. Wabash Ave.  
Chicago, IL 60604, USA