Spatial Bayesian Learning Algorithms for Geographic Information Retrieval

Arron R. Walker
Queensland University of Technology
2 George St
Brisbane Q 4001 Australia
+61 7 3864 7044
ar.walker@qut.edu.au

Binh Pham
Queensland University of Technology
2 George St
Brisbane Q 4001 Australia
+61 7 3864 9324
b.pham@qut.edu.au

Miles Moody
Queensland University of Technology
2 George St
Brisbane Q 4001 Australia
+61 7 3864 7044
m.moody@qut.edu.au

ABSTRACT
An increasing amount of freely available Geographic Information System (GIS) data on the Internet has stimulated recent research into Geographic Information Retrieval (GIR). Typically, GIR looks at the problem of retrieving GIS datasets on a theme by theme basis. However in practice, themes are generally not analysed in isolation. More often than not multiple themes are required to create a map for a particular analysis task. To do this using the current GIR techniques, each theme is retrieved one by one using traditional retrieval methods and manually added to the map. To automate map creation the traditional GIR paradigm of matching a query to a single theme type must be extended to include discovering relationships between different theme types.

Bayesian Inference networks can and have recently been adapted to provide a theme to theme relevance ranking scheme which can be used to automate map creation [2]. The use of Bayesian inference for GIR relies on a manually created Bayesian network. The Bayesian network contains causal probability relationships between spatial themes. The next step in using Bayesian Inference for GIR is to develop algorithms to automatically create a Bayesian network from historical data. This paper discusses a process to utilize conventional Bayesian learning algorithms in GIR. In addition, it proposes three spatial learning Bayesian network algorithms that incorporate spatial relationships between themes into the learning process. The resulting Bayesian networks were loaded into an inference engine that was used to retrieve all relevant themes given a test set of user queries. The performance of the spatial Bayesian learning algorithms were evaluated and compared to performance of conventional non-spatial Bayesian learning algorithms.

This contribution will increase the performance and efficiency of knowledge extraction from GIS by allowing users to focus on interpreting data, instead of focusing on finding which data is relevant to their analysis.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Retrieval Models.

General Terms
Algorithms, Experimentation.

Keywords
Spatial Bayesian learning, learning Bayesian networks, geographic information system, geographic information retrieval, information retrieval.

1. INTRODUCTION
The amount of GIS data on the internet has been growing explosively in recent years. Some of the reasons for this growth are that spatial data itself is getting cheaper, in fact some governments and organisations give data away free of charge [3]. For example, Geoscience Australia [4] and ESRI's Geography Network [5] provide access to some free GIS data. Another contributing factor was the released of GIS Web standards like Web Map Service (WMS) by Open GIS Consortium (OGC) [6]. Since its introduction, the major commercial GIS companies have supported the WMS specification. Numerous free web based WMS viewers like Intergraph's "OGC WMS Viewer" [7] and ESRI's ArcExplorer have been developed. ArcExplorer not only supports WMS, but also provides the ability to download maps of a variety of standards, thus, allowing different datasets from different servers as well as local data to easily be combined into single map visualisation.

GIR techniques have not kept pace with this explosion of GIS web based data. Typically, GIR looks at the problem of retrieving GIS datasets on a theme by theme basis. However in practice, themes are generally not analysed in isolation. More often than not multiple themes are required to create a map for a particular analysis task. To do this using the current GIR techniques, each theme is retrieved one by one using traditional retrieval methods and manually added to the map. To automate map creation the traditional GIR paradigm of matching a query to a single theme type must be extended to include discovering relationships between different theme types. GIR should be able to create multi theme maps from a simple user query.

Bayesian Inference networks offer one such technique to retrieve multiple spatial themes given a simple user query. Recently a

GIR system was developed by Walker [2] that used a Bayesian network to assign causal relationships between spatial theme. The system used Bayesian inference theory to rank all available themes given that one theme has been found relevant to the query. In effect, the initial theme is retrieved via a typical GIR method, but once selected it becomes the evidence in a Bayesian Inference network to allow the related themes to it to be ranked and subsequently retrieved. Bayesian Inference will be explained in more detail later in this paper.

At the heart of a Bayesian Inference retrieval system is the Bayesian network which contains the intelligence about which themes are related to each other and how strong the relationships between them are. Walker's system required an initial Bayesian network to be calculated manually by domain experts. This was a slow and difficult job.

For Bayesian Inference networks to become feasible solutions for GIR, the creation of this Bayesian network must be performed automatically using readily available stored expert knowledge. The process of identifying the best Bayesian network and its associated conditional probabilities is known to experts in this area as "Learning Bayesian Networks".

This paper describes three new spatial Bayesian learning algorithms that automatically create Bayesian networks. The Bayesian learning approaches presented here considers spatial data aspects as part of their learning algorithms, therefore, addressing the deficiency in the "non-spatial" Bayesian learning approaches. The significance is the removal of the last manual process in Walker's Bayesian GIS retrieval system.

The remainder of this paper is structured as follows. Sections 2, proves background to how GIS data are managed in spatial themes. Section 3 reviews the current use of Bayesian networks for information retrieval. Section 4 investigates work in the area of learning Bayesian networks. Section 5 outlines a process to extract spatial relationship between GIS datasets. Section 6 proposes two spatial Bayesian learning algorithms and one parameter learning algorithm that incorporate spatial relationships. Section 7 reviews the suitability of using current IR evaluation measures for GIR. Section 8 outlines the experiment conducted to evaluate the proposed algorithms. Finally, section 9 presents the conclusions and comments on future work.

2. GIS DATA MANAGEMENT

This section gives a quick overview of how maps are created and then stored for later use in GIS. It will be shown later in this paper how this information can be used as to way to access existing stored knowledge of relationships between themes.

2.1 SPATIAL THEMES

The main principle of data organisation of a GIS is to group the spatial data into themes or spatial data layers. Each theme has an associated dataset and meta data. These themes are generally layered one on top of the other in the visualisations interface of the GIS, as shown in figure 1.

Categorising data into themes increases the efficiency of data querying. It allows easy addition of new data by simply overlaying a new theme layer to create a map which is useful for some analysis task.

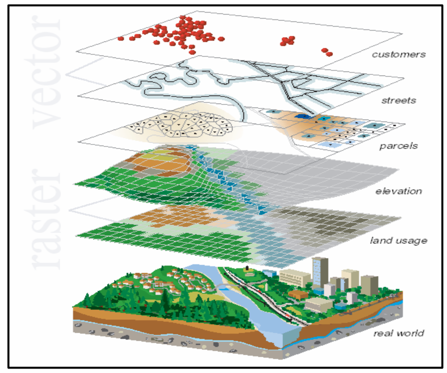


Figure 1. Spatial Themes [1]

2.2 WORKSPACES

Themes in isolation don't provide the user with a spatial context when they are analysing GIS data. Therefore, it is common practice to load multiple themes into a GIS before analysis begins. Most GIS remember this multiple theme configuration through a manual process called "workspace creation"[1]. Here users must manually select the themes of interest and explicitly save them to a workspace. Consequently, a workspace for each user analysis task is required. These workspaces are a static record of the themes loaded into the GIS, and thus do not dynamically update themselves as new themes become available. Large organizations that use GIS will generally have experts initially setup and maintain these workspace files.

There is an opportunity to use these historical workspace files to create a Bayesian network that assimilates the expert knowledge stored in them. Such a Bayesian network should illustrate good performance in retrieving GIS datasets. This paper presents such a method for learning a Bayesian network from workspace files.

3. BAYESIAN INFERENCE FOR IR

Traditional information retrieval (IR) systems were developed to retrieve text documents that are relevant to a given query. The use of Bayesian Inference for IR is widely accepted [8-10]. These systems use key terms in the query as evidence in the Bayesian network. From this evidence, the documents are ranked, in order of relevance, to the query. Over the years, Bayesian inference has been used in other information retrieval areas outside document retrieval. For example, Heckerman and Horvitz [11] developed a Bayesian help program, which retrieves relevant help topics given a user query. Their Bayesian network establishes a casual relationship between help topics and the query terms. Their system has become the basis for the Microsoft Office Help program. It demonstrates the usefulness of Bayesian inference in all areas of information retrieval. Some extensions to the traditional inference model combine Bayesian inference with heuristics with the aim of understanding queries [12]. These systems have shown retrieval performance improvements over the traditional method in certain domain areas.

Bayesian inference, as mentioned in the introduction, has been used by Walker [2] for GIR with good retrieval performance. His

[1] Workspace is a MapInfo term and the same process is known as Projects in ArcGIS.

system, establishes a casual relationship between spatial themes. The next section gives some background into Bayesian networks.

3.1 Bayesian Networks

Bayesian networks are graphical models for defining probabilistic relationships between variables. These relationships can involve uncertainty, unpredictability or imprecision. The relationships may be discovered automatically from data files, or created by experts, or developed by a combination of the two. An advantage of Bayesian networks is that they capture knowledge in a form people can understand intuitively, and which allows a clear visualisation of the relationships involved.

Bayesian probability of an event x is a person's degree of belief in that event. This is somewhat different from a classical probability of an event x which is a physical property of that event in the world (e.g. probability that a coin will land heads).

Bayesian networks used a directed acyclic graph (DAG) to represent assertions of conditional independence (See figure 2). The nodes in the graph represent the variables and the directed arcs define the conditional relationships. The advantages of directed graphic models over undirected models are the notion of causality. Causality indicates that if an arc is directed from A to B in the network, then A causes B. Bayes' theorem is used to calculate causal inference about the variables. Bayes' theorem states:

$$p(B_i \mid A) = \frac{p(A \mid B_i)p(B_i)}{p(A)} \qquad where \quad (i = 1, 2, ..., \quad r)$$

Bayes' theorem allows the updating of the probabilities regarding uncertain events when fresh information is received [13]. That is, once you know certain events have occurred then one can recalculate the probability of other events occurring. The graphical and probabilistic structure of a Bayesian network represents a single joint probability distribution. This distribution is obtained using the Product (Chain) Rule for Bayesian networks:

$$p(X_1 .... X_n) = \prod_{i=1}^{n} p(X_i \mid pa(X_i))$$

Applying Bayes' decision rule performs classification [14]. For example, assume that there are two hypotheses in the classification domain, Bayes' decision rule states that A should be assigned to the hypothesis for which the posterior probability is a maximum. That is, choose;

$$\begin{aligned} B_0 &: \quad if \quad P(B_0|A) > P(B_1|A) \\ B_1 &: \quad if \quad P(B_1|A) > P(B_0|A) \end{aligned}$$

Where $P(B_0|A)$ and $P(B_1|A)$ can be calculated using Bayes' rule. If the above example was extended to include more than just two hypotheses, then the problem can be viewed as searching through the set of all possible hypotheses with the goal of finding the best hypothesis. The best hypothesis can be defined as the most probable hypothesis given the "evidence" of the data D in the hypothesis space H. Such a hypothesis is referred to as the maximum a posterior (MAP) hypothesis [15].

$$h_{MAP} \equiv \max_{h \in H} P(h \mid D)$$

From Bayes' rule,

$$h_{MAP} \equiv \max_{h \in H} \frac{P(D \mid h)P(h)}{P(D)}$$

Because P(D) is independent of h, it can be dropped, resulting in

$$h_{MAP} \equiv \max_{h \in H} P(D \mid h)P(h)$$

This process of ranking hypotheses given evidence is known as Bayesian Inference. Bayesian networks have been used in many different domains [12, 16-19] to provide decision support. For example, medical diagnostic systems based on Bayesian networks compute the best diagnoses given the existence of certain patient symptoms (or evidence) [17]. Of more interest to this research, is the use of Bayesian networks in IR, this will be discussed in the next section.

3.2 Bayesian Networks in GIR

To rank spatial theme in order of relevance to each a Bayesian network must be constructed that represents the causal relationships between themes. An example of a Bayesian network for spatial themes used by the Gold Coast City Council is shown in figure 2. If we consider the query as the selection of one theme, then this theme can be used as the evidence for Bayesian inference. Consequently, the MAP hypothesis for each theme in the network can be calculated and ranked. The resulting ranking represents the themes most related to the query theme. More details on how Bayesian inference has been adapted to GIR can be found in [2].

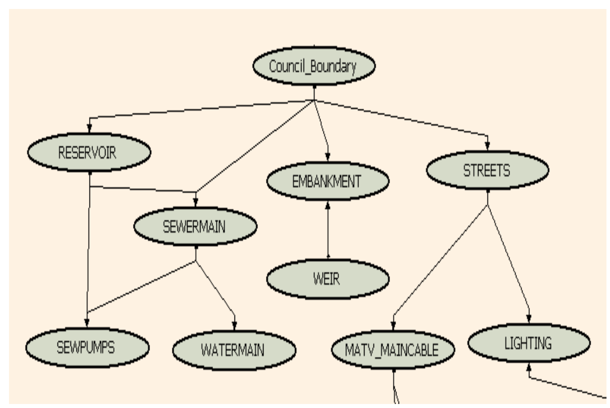


Figure 2. Bayesian Network of GIS Themes

4. LEARNING BAYESIAN NETWORKS

The process of learning a Bayesian network from data has two parts: 1) Structure Learning and 2) Parameter Learning. In Bayesian networks, the DAG is known as the structure and the conditional probability distributions are known as the parameters [20]. In the past, these properties had to be learnt manually with input from domain experts. For large networks, this task becomes impossible; as a result, researchers have developed methods to learn both the structure and the parameters of Bayesian networks automatically. In the following section, the main methods for Bayesian Learning will be outlined.

4.1 Structure Learning

Structured learning is the process of discovering the DAG that best describes the causal relationships in the data. The number of possible DAGs grows exponentially with the number of nodes. Robinson [21] equation below gives the number of DAGs:

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i) \qquad n > 2$$

For example n=10, gives approximately 4.2×10^{18} and n=106 approximately 1×10^{306}, therefore, exhaustively considering all DAG patterns is computationally infeasible [20]. Consequently, approximate algorithms that find the most probable structure have been developed. The two most popular methods are DAG search algorithm and K2 algorithm.

Both of these algorithms approach the structure learning problem by assigning equal prior probabilities to all DAG patterns and thus search the for the pattern that maximises the probability of the data, d, given the DAG, G, (i.e. P(d|G)). This probability is known as the Bayesian score ($score_B$).

$$score_B(d, G) = \prod_{i=1}^{n} \prod_{j=1}^{q_i^{(G)}} \frac{\Gamma(N_{ij}^{(G)})}{\Gamma(N_{ij}^{(G)} + M_{ij}^{(G)})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk}^{(G)} + s_{ijk}^{(G)})}{\Gamma(a_{ijk}^{(G)})}$$

The main difference between the two algorithms is how they determine the search space of DAG patterns to score. DAG search was developed by Chickering [22, 23] and uses a straightforward greedy search method and a set of operations. The operations are Add, Delete and Reverse edges in the DAG. The algorithm proceeds as follows: The initial DAG has no edges. At each step of the search, links are added, deleted and reversed and the new DAG score calculated. The algorithm stops when no operation increases the score. In contrast, the K2 algorithm developed by Cooper and Herskovits [24] is a greedy search method with a single operation. This operation is the addition of a parent to a node. K2 relies on the assumption that the order of allowable parents is know. This prior node ordering is created manually from expert knowledge.

The Bayesian score can only be calculated from data when the probabilities are relative frequencies. In the GIS domain, historical records of previous "workspaces' provide a measure of relative frequency of the use of datasets.

This paper present two new structure learning algorithms that are based on DAG search and K2, but include spatial relationships between datasets to alter the search space set of DAG patterns. These algorithms are presented in section 6.

4.2 Parameter Learning

Parameter values can only be learnt from data when the probabilities are relative frequencies. As mentioned above, in the GIS domain, historical records of previous "workspaces' provide a measure of relative frequency of the use of datasets. In addition, it is assumed that a theme's presence in a workspace is binominal, that is, it has only two values (present and not present). Finally, parameter learning assumes that relative frequencies have a beta distribution. With these assumptions, a standard parameter learning equation from [20] was used to discover the probability distribution for the Bayesian network. To update the distributions from the workspace data, the relative frequency given the data are calculated by using:

$$p(f \mid d) = beta(f : a + s, b + t)$$

Where d is a binomial sample with parameter f, f is the relative frequency of variable, a and b are the initial beta function parameters (set to 1 for equally likely) s is the number of variables in d equal to 1 (present) and; t is the number of variables in d equal to 2 (not present).

Because in parameter learning the network is known, the updated distributions can now be used to calculate all conditional probabilities.

4.2.1 Maximum Likelihood Parameter Estimation

The parameter Learning algorithm used in this experiment was maximum likelihood parameter estimation (MLE). MLE is a well known algorithm and an implementation of MLE was available in Bayes Net [25]. This code was easily modified and also provided a control for comparison with the new spatial MLE algorithm.

The idea behind MLE is to determine the parameters that maximise the probability (likelihood) of the sample data. MLE algorithm wants to maximise the likelihood of the parameter set θ given by dataset D and solve for θ :

$$L(\theta | D) = \prod_{j=1}^{m} p(x_j | \theta)$$

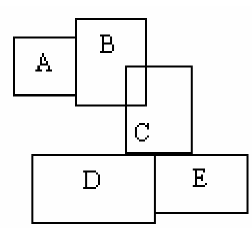
For binomial data, gives θ as:

$$\theta = \frac{N_1}{N_1 + N_0}$$

Where N_0 and N_1 are the number of times x equals 0 and 1 respectively. Therefore, MLE hypothesis asserts that the actual proportion of a parameter is equal to the observed proportion in the training set. This paper will utilise spatial relationships to introduce prior weights into MLE algorithm. This algorithm is presented in section 6.3.

5. SPATIAL RELATIONSHIPS

The new contribution of this work is the inclusion of spatial relationships into Bayesian learning algorithms. This section describes the method used to discover the spatial relationships between the GIS themes. Tobler's first law of Geography states: everything is related to everything else but nearby things are more related than distant things [26]. In spatial data analysis this Tobler's inter-dependence between spatial data can not be ignored [27]. One way to measure spatial relationships is to use the Moran's I measure[27]. Moran's I measure is dependent on the design of a contiguity matrix W which reflects the influence of neighbourhood. For example, a spatial neighbourhood contiguity matrix is shown in figure 3.



	A	B	C	D	E
A	0	1	0	0	0
B	1	0	2	0	0
C	0	2	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

(a) Map **(b) Contiguity Matrix W**

Figure 3. A spatial neighbourhood and its contiguity

The matrix can have multiple weights to record different degrees of relationship between objects. In figure 3, the spatial relationships of adjacent and overlaps are represented by the weights of 1 and 2 respectively.

5.1 Spatial Relationships for GIR

The spatial relationships used to construct the contiguity matrix for GIS data were: same point, contains, overlaps, adjacent, object extent overlaps, extent overlaps, separated and same object.

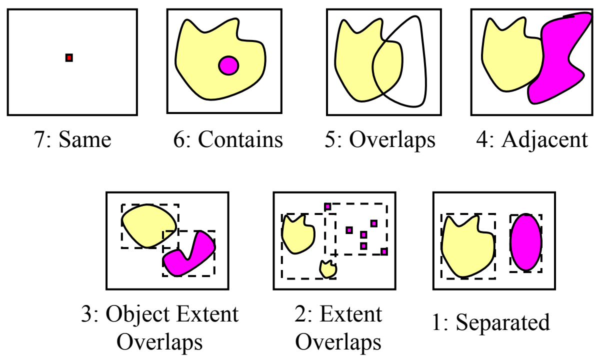


Figure 4: Spatial Relationships

The weights given to each relationship are listed in table 1 and the spatial relationships are illustrated in figure 4.

Table 1. Contiguity Matrix Weights

Spatial Relationship	Weight
One element in A at **Same point** *as one element in B*	7
At least one element in A **Contains** *at least one element in B*	6
At least one element in A **Overlaps** *at least one element in B*	5
At least one element in A is **Adjacent** *to at least one element in B*	4
At least one element in A **Object Extent Overlaps** *at least one element in B*	3
The extent of A **Extent Overlaps** *the extent of B*	2
All elements in A **Separated** *from all elements in B*	1
(Same object)	0

5.2 Spatial Causal Relationships

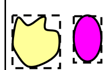
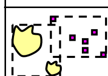
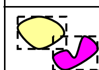
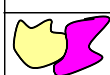
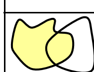
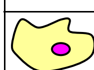
Because the spatial relationships are to be used in a Bayesian learning algorithm the relationships must be translated into a language similar to that used in Bayesian probability theory. Thus the notion of spatial causal relationships is presented here. Spatial relationships between objects can be considered as spatial causal relationships in the context of information retrieval. Consider three GIS datasets of *Australian Cities*, *American Cities* and the *Australian Continent*. From Tobler's first law of Geography, we can see that *Australian Continent* and *Australian Cities* will be more related than *Australian Continent* and *American Cities*. This is because *Australian Continent* is spatially closer to *Australian Cities* than to *American Cities*.

The Bayesian information retrieval system used in this paper is based on cause and effect relationships. Accordingly, the assumption is made that nearby objects have causal effects on each other. That is if a person queried the GIS for *Australian Continent* it is more probable that they are interested in *Australian Cities* than *American Cities* because of this assumed spatial causal relationship.

5.3 Translation of Spatial Relationships to Spatial Causal Relationships

The spatial relationships discovered in the contiguity matrix were translated into their equivalent spatial causal relationship. The resulting translation is listed table 2:

Table 2: Causal Spatial Relationships

Spatial Relationships		Spatial Causal Relationships		
	Same object	A	No link	0
	Extent separated	A B	No link	0
	Extent overlaps	A ←→ B	Link, direction unknown	weak
	Object extent overlaps	A ←→ B	Link, direction unknown	average
	Adjacent	A ←→ B	Link, direction unknown	good
	Overlaps	A ←→ B	Link, direction unknown	strong
	Contains	A → B	Directed link	strong
	At same point	A ←→ B	Link, direction unknown	Very strong

6. SPATIAL BAYESIAN LEARNING

Three new spatial Bayesian learning algorithms are discussed in this section. The first two are structure learning algorithms. They are called Spatial K2 and Spatial DAG. These algorithms incorporate spatial relationships into the two popular structure learning algorithms. Finally in this section a new parameter learning algorithm is discussed. This algorithm incorporates spatial relationships into MLE which is an existing Bayesian parameter learning algorithm.

6.1 Spatial K2

This algorithm automatically calculates the node order for use in the K2 algorithm. It uses the Spatial relationships between datasets (from contiguity matrix) to calculate the spatial causal ordering. Apart from this automated order calculation, the original K2 algorithm is unchanged

Spatial K2 Algorithm:

1) Calculate contiguity Matrix W;
2) Derive spatial causal order from Matrix W (See Deriving Spatial Ordering below);
3) Run K2 using this order.

Deriving Spatial Causal Ordering:

1) Sum the columns in W;
2) Place in descending order; (Note value that has maximum number has most influence on more neighbours)
3) Loop through this list;
   Swap the order of same value nodes if node has a greater score when only considering the current parents in list;
   If score the same just keep arbitrary order;

For example, consider the spatial relationships in figure 3, the resulting spatial causal order would be C, B, D, E, A

6.2 Spatial DAG

Spatial DAG modifies Chickering's DAG search algorithm to include spatial causal relationships. It only allows the operations of add, delete and reverse to be used within the search algorithm if they agree with the discovered spatial relationships. This should limit the number of DAG patterns that needs to be searched in order to find a maximum, thus it should take less time to discover the Bayesian network. In addition, the resulting Bayesian network should match the causal spatial relationships in the data.

Spatial DAG Algorithm:

Calculate contiguity Matrix W;
Do
  Score only DAGs that are in neighbourhood (i.e. add, delete, reverse) AND in contiguity Matrix W;
    If (any increase the score)
      Modify result DAG to the one that increases the score the most;
While (some operation increases the score);

6.3 Spatial MLE

Prior knowledge of the Bayesian relationships can be used to influence the MLE process. The method proposed here takes spatial relationships between data and converts them into a prior metric for use in an MLE algorithm.

The process for introducing prior knowledge about data into the MLE algorithm is detailed in Bayes Net [25]. It states; if we let N_{ijk} = the number of times X_i=k and Pa_i =j occurs in the training set, where Pa_i are the parents of X_i, then the maximum likelihood estimate is: $T_{ijk} = N_{ijk}/ N_{ij}$ (where N_{ij} = sum k'$N_{ijk'}$), which will be 0 if N_{ijk}=0. To prevent us from declaring that (X_i = k, Pa_i = j) is impossible just because this event was not seen in the training set, we can pretend we saw value k of X_i , for each value j of Pa_i some number (α_{ijk}) of times in the past. The MLE is then:

$$\theta_{\mathrm{ijk}} = \frac{\left(\mathrm{N}_{\mathrm{ijk}} + \alpha_{\mathrm{ijk}}\right)}{\left(\mathrm{N}_{\mathrm{ij}} + \alpha_{\mathrm{ij}}\right)}$$

This paper proposes to modify the above, so $\alpha_{ijk} = \alpha_{ij}$ = spatial prior weight. This will have the affect of biasing the probability toward the parent that has a spatial relationship with the child in the Bayesian network.

Algorithm for calculating α_{ij} (spatial prior weight):

For each i,j

  If ij have a parent child relationship

    If ij have spatial relationship with weight greater than 3

      α_{ij} = 5.

End for

7. RETRIEVAL EVALUATION

Once the Bayesian networks were learnt, the respective Bayesian networks had to be evaluated against each other. In order to do this, a recall and precision measure that suited GIS dataset retrieval was established.

7.1 Recall and Precision

In traditional IR, recall and precision are generally described in terms of documents retrieved. The standard recall and precision measures [28] for document retrieval are:

1. Recall is the fraction of relevant documents retrieved to the total number of relevant documents.

$$recall = \frac{|R_a|}{|R|}$$

2. Precision is the fraction of the retrieved documents, which are relevant.

$$precision = \frac{|R_a|}{|A|}$$

The same recall and precision measures were used in this experiment by simply considering datasets retrieval as equivalent to document retrieval as far as performance is concerned. To evaluate the retrieval performance of the Bayesian networks over all test queries, the precision figures are averaged at each recall level and a graph of recall versus precision is construct [28].

8. EXPERIMENT AND RESULTS

This section describes the experiment conducted to evaluate the proposed spatial Bayesian learning algorithms.

8.1 Gold Coast City Council Database

The experiment used GIS datasets supplied by the Gold Coast City Council (GCCC). The datasets were typical spatial data of interest to council planners (i.e. property boundaries, water mains, etc). In addition, 20 workspace files currently used by GCCC for GIS dataset retrieval provided a measure of relative frequency of the use of datasets. The workspace data was organised into a form suitable for the Bayesian learning algorithms (i.e. 'workspace' x 'GIS datasets'). Each dataset was a discrete variable and was marked as *present* or *absent* from a particular workspace.

8.2 Calculation of Spatial Relationships

A C# program using MapObjects 2.2 was written to calculate the contiguity matrix for 70 GIS datasets. The GIS datasets were real life datasets supplied by Gold Coast City Council in Australia.

The algorithm used for calculating contiguity matrix W was a simple heuristic search algorithm: *for each GIS dataset pair, determine the spatial relationship between them and add to contiguity matrix.*

In order to test the correctness of the calculated spatial relationships, eleven spatial datasets consisting simple point, lines and polygons were evaluated. The simple test GIS themes and the resulting contiguity matrix is shown in figure 5. The time required to calculate spatial relationships for large GIS datasets was considerable.

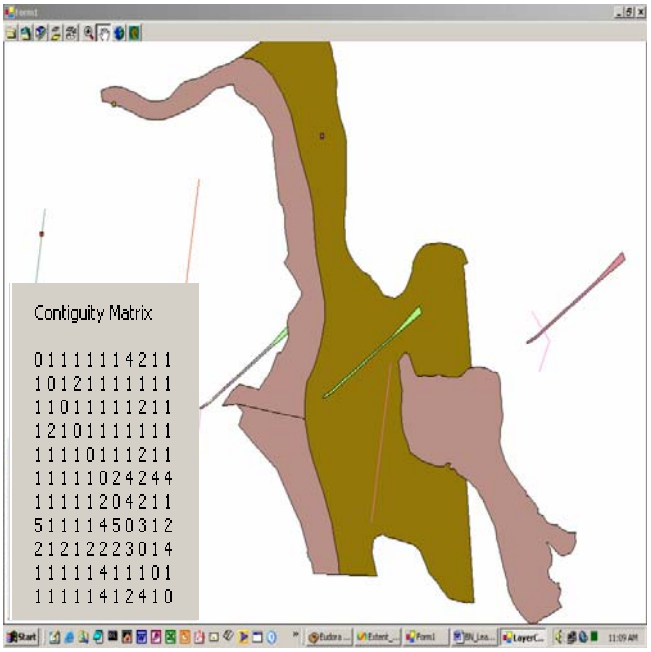


Figure 5. Testing Spatial Relationship Calculations

8.3 Implementing Spatial Bayesian Learning

All the Bayesian learning algorithms were implemented in Matlab using Bayes Net toolbox [25]. The main advantage was that the original K2, MLE and many Bayes related functions were already implemented in the toolbox. As a result, Spatial K2, DAG, Spatial DAG, Spatial MLE and Deriving Spatial Causal Ordering functions were written in Matlab.

Number of nodes (datasets)	Algorithm process time (sec)			
	K2 (non spatial arbitrary ordering)	Spatial K2 (spatial ordering)	DAG	Spatial DAG (with spatial rules)
3 (Cooper and Herskovits)	1.2603	(no spatial component)	4.5657	(no spatial component)
30 (GISData30)	248.0961	264.0287	1891.4584	883.2220
50 (GISData50)	609.7302	886.4125	4584.4303	2289.1130
70 (GISData70)	1344.3933	1337.6697	6897.5612	6994.8862

Table 3. Structure Learning Algorithm Process Time

The process time for each of the structure learning algorithms is shown in table 3. As expected, Spatial K2 and K2 are considerably faster than the DAG and Spatial DAG. The process time to calculate the contiguity matrix has not been included in the table. This would increase the processing time for both spatial methods equally.

Once the network structure was learnt, both the standard MLE and Spatial MLE parameter learning algorithms were used to calculate the a priori and conditional probability tables of the respective Bayesian networks. The original Bayes Net MLE algorithm allowed exact conditional probabilities of 1.0 and 0.0. This allowed for no uncertainty and affected the inference algorithm's ability to rank the posterior probabilities. As a result, the algorithm was modified to produce conditional probabilities of 0.9999 for 1.0 and 0.0001 for 0.0.

8.4 Display and Inference

The Bayes Net Toolbox does not have nice graphical output for Bayes networks, therefore, a function was written the save the resulting DAG to the Microsoft's MSBN file format. This allowed MSBN [29] to be used to display the BN and run inference calculations.

The resulting Bayesian networks created by the four Bayesian learning algorithms are shown in figures 6 to 9.

The main difference in the structures is the maximum number of parents. K2 and Spatial K2 was limited to 2 parents to allow the algorithm to process fast. The DAG and Spatial DAG network, which had no limit, averaged 6 parents per node and therefore, are more complex networks.

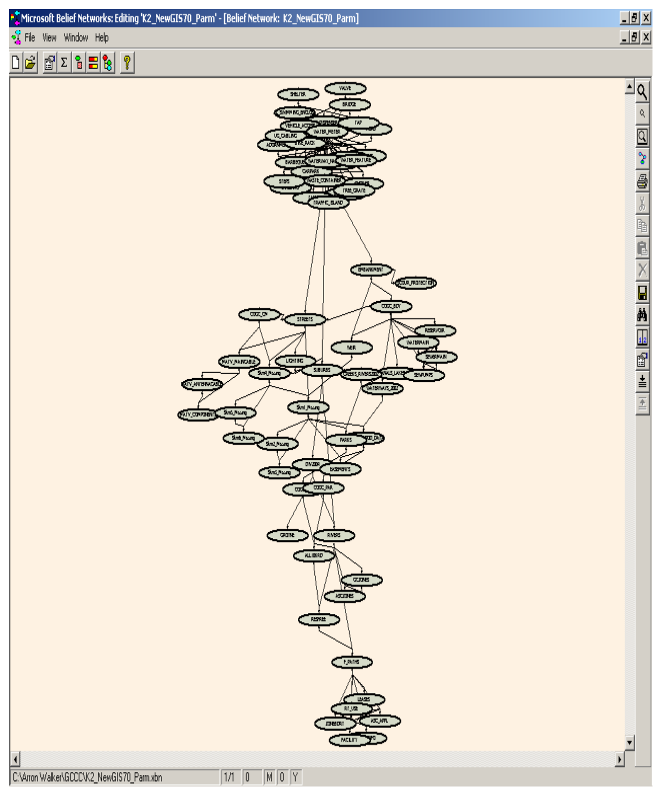


Figure 6. Network constructed by K2

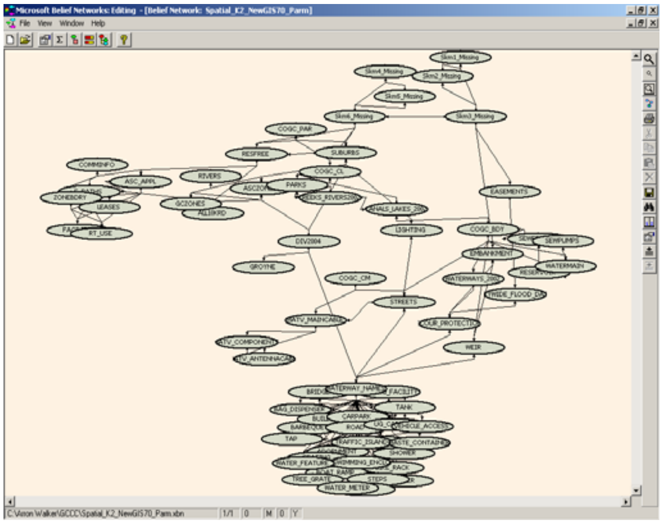


Figure 7. Network constructed by Spatial K2

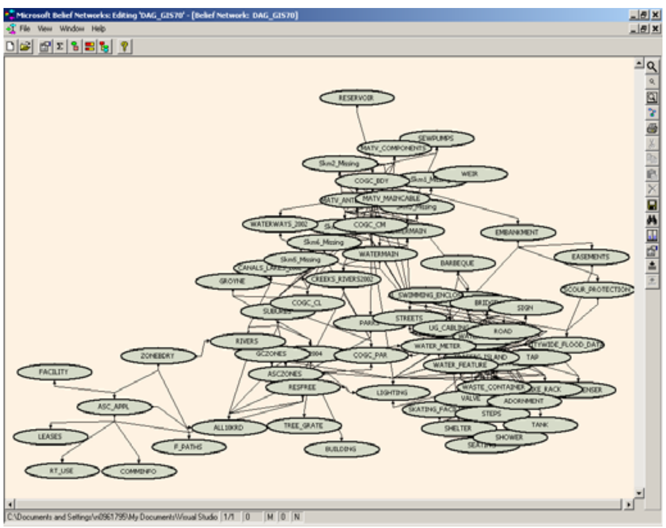


Figure 8. Network constructed by DAG

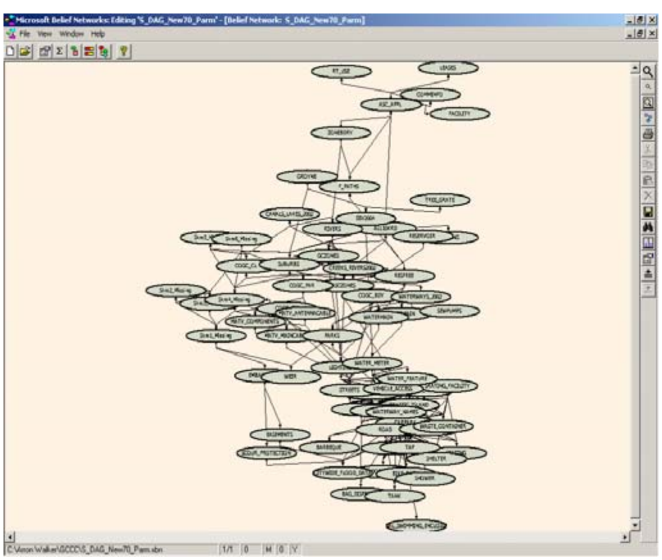


Figure 9. Network constructed by Spatial DAG

8.5 Evaluation

To obtain the recall and precision measures, a set of queries and answers was devised. Traditional GIR can match queries to datasets of similar themes. This evaluation tests the respective Bayesian Network's ability to match a theme to related, but not necessarily similar, themes. Therefore, to simplify our querying process we assume that a dataset has been matched to a general query using traditional methods. Following this assumption, we used a single dataset as query input, and evaluated its ability to retrieve all related datasets. The original workspace data provided a means of testing this type of retrieval. Each dataset in the workspace became a query and the answer to that query was all the remaining datasets in the workspace. The four distinct Bayesian networks were evaluated using 70 single word queries. A C# function was added to the Bayesian GIS dataset retrieval system developed in [2] to automatically query the learnt Bayesian networks.

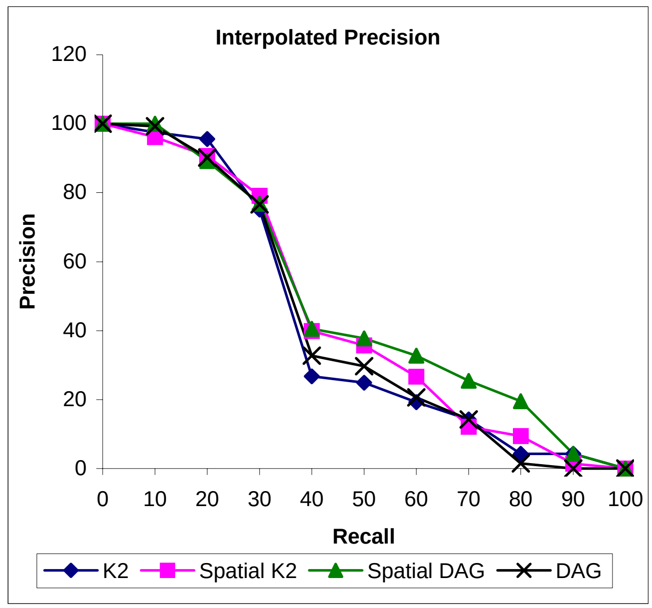


Figure 10: Bayesian Structure Learning Comparison

Figure 10 shows the recall and precision figures for the four Bayesian structure learning algorithms evaluated in this experiment. Each of the structure learning algorithms shown here used the MLE parameter learning algorithm. From figure 10, it can be seen that all algorithms performed similar between 0 to 30% recall, however, differences became obvious in the 40% to 90% recall range. Spatial DAG produced the best retrieval performance with Spatial K2 the next best performer. Both spatial algorithms performed better than the non-spatial algorithms of K2 and DAG.

Finally, the MLE and the Spatial MLE parameter learning algorithms were evaluated on the four Bayesian networks that have been constructed. From figures 11 and 12 we can see that including spatial parameter learning gives an advantage only when the algorithm is run on a Bayesian structure discovered using traditional non-spatial algorithms. That is, if the Bayesian structure was discovered using a spatial learning algorithm and then the parameters were discovered using a spatial algorithm we in fact get poorer retrieval performance than if we had a spatial structure with traditional parameters.

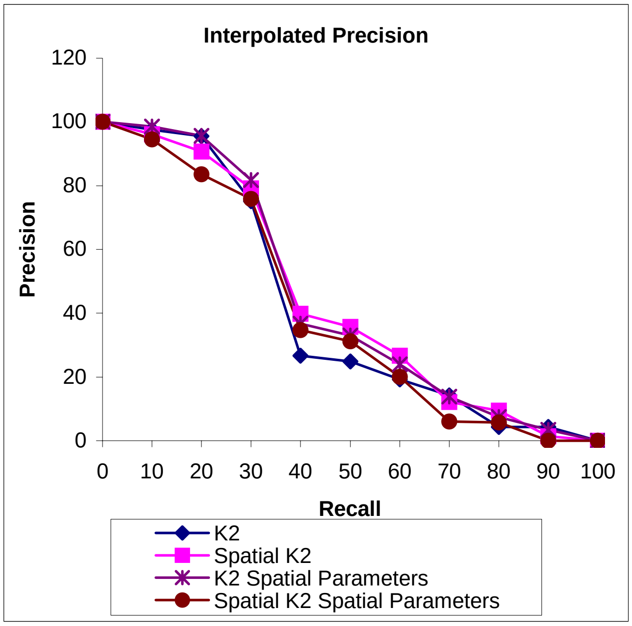


Figure 11: Parameter Learning Comparison (K2 & SK2)

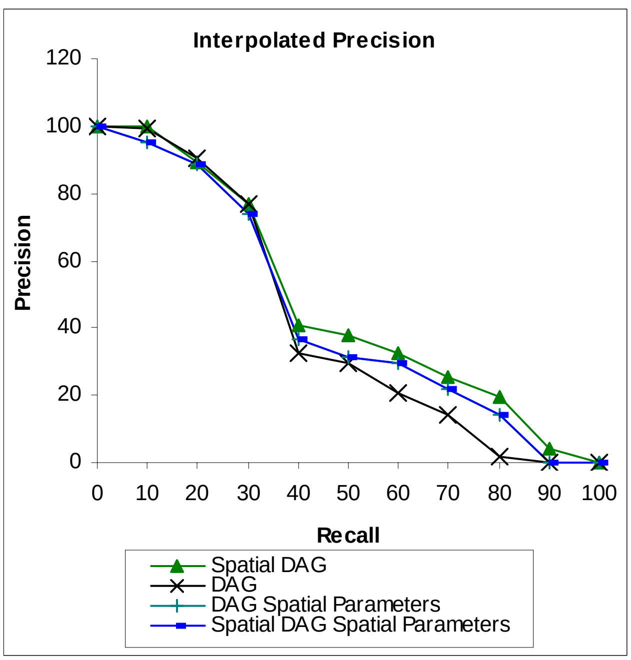


Figure 12: Parameter Learning Comparison (DAG & SDAG)

9. CONCLUSIONS AND FUTURE WORK

This paper shows that Bayesian learning algorithms can automatically create Bayesian networks suitable for use in a Bayesian inference GIR system. The algorithms utilised expert knowledge currently stored in GIS workspace data files. Consequently, no manual expert input would be required to set up a Bayesian inference GIR system.

The paper presented two spatial structure learning algorithms that demonstrate the advantages of incorporating spatial relationships when compared to traditional structure learning algorithms. The spatial structure learning algorithms yielded improved retrieval performance over non-spatial algorithms. This improvement was offset by the additional time required to calculate the network. The major overhead being the time required calculating the spatial relationships in large GIS datasets. The processing time is of less importance as the algorithms are only run once at the initialisation stage of a Bayesian GIS retrieval system.

The incorporation of spatial relationships into the parameter learning MLE algorithm did improve retrieval performance, but not to the same degree as the spatial structured learning algorithms. Interestingly, the retrieval performance was the best for the spatial structured learning networks when they used a non-spatial parameter learning algorithm. If the spatial parameter learning algorithm was used with non-spatial structure learning then the retrieval performance was better than if just a non-spatial parameter learning algorithm was used. This was an unexpected result; it was thought that the best retrieval performance would have come from a combined spatial structure and spatial parameter learning algorithm. The only explanation for this is that the prior bias calculation for the spatial MLE algorithm is cancelling out the structure learnt during the spatial structure learning process. More investigation into this will be required in the future.

In future work, some additional parameter learning algorithms will be modified to include spatial relationship. It is also planned to improve the GIR system developed in [2]. The major benefit of this work is its ability to tap into stored expert knowledge (workspace files) to allow efficient retrieval of GIS themes. Not all users possess the expert knowledge to match spatial themes to analysis task. Furthermore, users usually do not have the time to study the meta data for all available datasets to make these decisions. Technology such as WMS has greatly increased the number of datasets available for analysis. With so many data sources available, the manual process of selecting datasets for particular analysis tasks is not trivial, hence the need for an automatic process. A static workspace requires users to constantly check for new datasets, but a dynamic GIR environment that automatically loads new datasets would ensure that users' decision making is based on the best available data.

10. ACKNOWLEDGMENTS

The Cooperative Research Centre for Satellite Systems (CRCSS) is established and supported under the Australian Government's Cooperative Research Centres Program. This research is support by the Built Environment Research Unit, Dept of Public Works of the Queensland Government. The Gold Coast City Council provided the GIS datasets for use in the experiment.

11. REFERENCES

[1] "ArcView," Environmental Systems Research Institute, 2003, http://www.esri.com/, accessed on: 14-Jan-2003.

[2] A. Walker, B. Pham, and A. Maeder, "A Bayesian framework for automated dataset retrieval in Geographic Information Systems," presented at The 10th International

Conference on Multi-Media Modeling, Brisbane, Australia, 2004.

[3] Queensland Government, "Using spatial information for a sustainable SEQ," presented at 2002 SEQ Spatial Information Expo, Brisbane, 2002.

[4] "Geoscience Australia," Australian Government, 2004, http://www.ga.gov.au, accessed on: 29-Nov-2004.

[5] "Geography Network," ESRI, 2005, http://www.geographynetwork.com/, accessed on: 31-May-2005.

[6] "Open GIS Consortium," 2003, http://www.opengis.org/, accessed on: 30-June-2003.

[7] "OGC WMS Viewer," 2003, http://www.wmsviewer.com/, accessed on: 4-July-2003.

[8] R. Fung and B. Del Favero, "Applying Bayesian networks to information retrieval," Communications of the ACM, ACM Press, vol. 38, pp. 42-57, 1995.

[9] H. Turtle and W. B. Croft, "Evaluation of an inference network-based retrieval model," ACM Transactions on Information Systems (TOIS), vol. 9, pp. 187 - 222, 1991.

[10] B. A. Ribeiro-Neto and R. Muntz, "A belief network model for IR," in Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. Zurich, Switzerland: ACM Press, 1996, pp. 253--260.

[11] D. Heckerman and E. Horvitz, "Inferring Informational Goals from Free-Text Queries: A Bayesian Approach," presented at Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, 1998.

[12] H. M. Meng, W. Lam, and K. F. Low, "A Bayesian approach for understanding information-seeking queries VO - 4," presented at IEEE International Conference on Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings., 1999.

[13] T. Leonard and J. S. J. Hsu, Bayesian Methods - An Analysis for Statisticians and interdisciplinary Researchers. Cambridge, United Kingdom: Cambridge University Press, 1999.

[14] A. A. Skabar, "Inductive Learning Techniques for Mineral Potential Mapping," in School of Electrical and Electronic Systems Engineering. Brisbane: Queensland University of Technology, 2000, pp. 226.

[15] T. M. Mitchell, Machine Learning. New York: McGraw-Hill, 1997.

[16] A. Jameson, B. Gro[ss]mann-Hutter, L. March, R. Rummer, T. Bohnenberger, and F. Wittig, "When actions have consequences: empirically based decision making for intelligent user interfaces," Knowledge-Based Systems, vol. 14, pp. 75-92, 2001.

[17] P. Haddawy, J. Jacobson, and C. E. Kahn Jr., "BANTER: a Bayesian network tutoring shell," Artificial Intelligence in Medicine, vol. 10, pp. 177-200, 1997.

[18] N. Liem and P. Haddawy, "Answering queries from context-sensitive probabilistic knowledge bases," Theoretical Computer Science, vol. 171, pp. 147-177, 1997.

[19] M.-L. Shyu and S.-C. Chen, "A Bayesian network-based expert query system for a distributed database system VO - 3," presented at IEEE International Conference on Systems, Man, and Cybernetics, 2000.

[20] R. E. Neopolitan, Learning Bayesian Networks. Upper Saddle River: Person Education, 2004.

[21] R. W. Robinson, "Counting unlabeled acyclic digraphs," Lecture notes in mathematics, 622: Combinatorial mathematics V, New York: Springer-Verlag, 1977.

[22] D. M. Chickering, "Learning Equivalence Classes of Bayesian Network Structures," in Proceeding of Twelfth Conference Uncertainity in Artificial Intelligence, E. Horvitz and F. Jensen, Eds., 1996, pp. 150-157.

[23] D. M. Chickering, "Learning Equivalence Classes of Bayesian-Network Structures," Journal of Machine Learning Research, vol. 2, pp. 445-498, 2002.

[24] G. F. Cooper and E. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data," Machine Learning, vol. 9, pp. 309-347, 1992.

[25] "Bayes Net Toolbox for Matlab," MIT, 2004, http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html, accessed on: May-2004.

[26] W. R. Tobler, Cellular Geography, Phillosophy in Geography. Dordrecht: Reidel, 1979.

[27] S. Chawla, S. Shekhar, W. Wu, and U. Ozesmi, "Modeling spatial dependencies for mining geospatial data: An introduction," in In Geographic data mining and Knowledge Discovery(GKD), H. Miller and J. Han, Eds. Taylor and Francis, 2001.

[28] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. New York: ACM Press, 1999.

[29] "MSBNx: Bayesian Network Editor and Toolkit," Microsoft, 2003, http://research.microsoft.com/adapt/msbnx/default.aspx, accessed on: 1-12-2003.