2-1991

Predicting Cause-Effect Relationships from Incomplete Discrete Observations

E Boros  
*Rutgers University*

P. L. Hammer  
*Rutgers University*

John N. Hooker  
*Carnegie Mellon University, john@hooker.tepper.cmu.edu*

Follow this and additional works at: [http://repository.cmu.edu/tepper](http://repository.cmu.edu/tepper)

Part of the [Economic Policy Commons](http://repository.cmu.edu/tepper) and the [Industrial Organization Commons](http://repository.cmu.edu/tepper)

Published In  

This Article is brought to you for free and open access by Research Showcase @ CMU. It has been accepted for inclusion in Tepper School of Business by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.
Predicting Cause-Effect Relationships from Incomplete Discrete Observations *

E. Boros  
DIMACS and RUTCOR, Rutgers University  
P.O. Box 5062, New Brunswick, NJ 08903

P. L. Hammer  
RUTCOR, Rutgers University  
P.O. Box 5062, New Brunswick, NJ 08903

J. N. Hooker  
GSIA, Carnegie Mellon University  
Pittsburgh, PA 15213

February 1991

Abstract
We address a prediction problem that frequently occurs in practice. We wish to predict the value of a function on the basis of discrete observational data that are incomplete in two senses. Only certain arguments of the function are observed, and the function value is observed only for certain combinations of values of these arguments. We solve the problem under a monotonicity condition that is natural in many applications, and we discuss applications to tax auditing, medicine, and real estate valuation. In particular, we display a special class of problems for which the best monotone prediction can be found in polynomial time.

1 Introduction

The problem of establishing cause-effect relationship based on incomplete observations was studied in [4]. In this paper we address the problem of finding a good approximation of an unknown discrete function on the basis of a set of observations, which is incomplete in two senses. We observe the values of only

*GSIA Working Paper 1991-12. The first and second authors gratefully acknowledge the partial support of NSF (Grant DMS 89-06870) and AFOSR (Grants 89-0512 and 90-0008), and the third author that of AFOSR (Grant 87-0292).
some of the arguments of the function, and we observe the function value only for certain combinations of values of these arguments. Our goal is to predict the value of the function for any combination of values for these arguments. Such problems occur frequently, and we begin with some examples.

Suppose that a tax bureau must decide which tax forms to audit, with the goal of auditing only those for which the increased return justifies the auditing expense. The observation set consists of data records corresponding to forms audited in the past. Each record indicates whether the audit was justified and lists some attributes of the taxpayer. The problem is to determine, on the basis of these attributes, when future tax forms should be audited.

A taxpayer’s attributes form a vector $x$ of boolean (0-1) variables, where each $x_j$ indicates whether a certain threshold is exceeded. For example, one $x_j$ may have the value 1 when the taxpayer claims too many dependents for a person of his age, and another may have the value 1 when he claims too many charitable contributions relative to his income.

The observation set partially defines a function $g$ whose value is 1 when auditing was justified and 0 otherwise. Whether an audit was justified depends not only on the recorded attributes used to make the auditing decision, but on a number of hidden factors as well. The arguments of $g$ therefore consist of a vector $x$ of the recorded attributes and a vector $y$ of hidden attributes. So, each data record can be written in the form $(x, g(x, y))$. In practice the function $g$ is only partially defined because its value $g(x, y)$ is not given for every possible vector $(x, y)$ of attributes. We cannot even assume there is an observation for every possible $x$.

Our problem is to predict the value of $g(x, y)$ on the basis of $x$ alone. To do this we derive an approximation $f$ of $g$ that is a function of $x$ only. Deriving $f$ requires in part that we interpolate a value when $x$ has not been observed. But even when $x$ has been observed, we may need to reconcile different observed values of $g(x, y)$ for different $y$'s.

We will focus on problems in which $f$ is monotone, in the sense that $f(x) \geq f(x')$ whenever $x \preceq x'$, where $\preceq$ is a partial order on the outcomes and $\preceq$ a partial order on the $x$'s. In the tax auditing example, the two outcomes are linearly ordered (0 < 1), and $\preceq$ is defined by $x \preceq x'$ whenever $x_j \leq x'_j$ for all $j$. Monotonicity is a reasonable assumption for this problem, because every 1 among the attributes is another reason for suspecting that the form should be audited. If one form with certain suspicious traits should be audited, then another form with these and still other suspicious traits should certainly be audited.

We will also generalize our approach to problems in which more than two outcomes are possible, where these outcomes are partially ordered. The tax auditor, for example, might classify forms as needing no audit, needing a second look to determine the desirability of an audit, and needing an audit right away. These outcomes happen to be linearly ordered in a natural way (in the order listed), and it is reasonable to assume that $f$ is monotone.
The attributes as well as the outcome can have more than two values. Suppose for example we have a battery of 45 biochemical tests for carcinogenicity, each of which can have outcomes negative, indefinite, and positive\(^1\). We do not apply every test to a given chemical; when a test is not applied, we say that the test “result” is no data. The 45 test results are viewed as attributes of the chemical tested. Since the tests are not foolproof, we first apply them to a number of chemicals that have been previously tested clinically for carcinogenicity, with possible outcomes harmless, undetermined (indicating inconclusive clinical experience), potentially dangerous (meaning that the chemical causes cancer when ingested in very large doses), and dangerous. The problem is to find a function \(f\) that predicts the outcome of clinical trials for a new chemical.

To check for monotonicity we must impose a partial ordering on the attribute vectors and on the outcomes. The four possible values of each attribute \(x_j\) submit to a partial order: negative is less than indefinite and no data, which are less than positive; but indefinite and no data are incomparable. We can therefore say that \(x_j \leq x'_j\) when \(x_j \leq x'_j\) for all \(j\). The four outcomes have a similar partial ordering: harmless is less than potentially dangerous and undetermined, both of which are less than dangerous; but we may be unable to order potentially dangerous and undetermined with respect to each other. It is reasonable to assume that \(f\) is monotone with respect to these orderings.

Whenever \(f\) incorrectly predicts the outcome for a given observation, a penalty is incurred, where the size of the penalty depends on the correct and predicted outcomes. For us, the best approximation \(f\) is one that minimizes the total penalty. In [2] we investigate the issue of defining “best approximation” from a statistical point of view and distinguish the analysis described here from logit and categorical data analysis.

If there is no restriction on \(f\), finding the best approximation is relatively easy, since we can treat each \(x\) separately. That is, for any fixed \(x\), \(f(x)\) should have the value that minimizes the sum of the penalties over observations of the form \((x, y)\). If \(x\) occurs in no observations, \(f(x)\) can be set to an arbitrary value. But if \(f\) is required to be monotone (or to have some other restrictive property), the problem cannot in general be decomposed this way, and the observed values of \(f(x)\) for unobserved \(x\) are in general restricted. In [2] we consider some other possible restrictions on \(f\) (when it is a boolean function). More examples for similar problems can be found in [3, 14, 15].

In this paper we show that a network flow model can be used to determine the best approximation \(f\) when the partial ordering of the outcomes is an interval order. An interval order is one in which every element can be associated with an interval of real numbers, such that \(a \leq \beta\) if and only if the upper end of the interval associated with \(a\) is less than or equal to the lower end of that associated with \(\beta\). A special case of an interval order is one that is layered, in

\(^1\) We thank Dr. Paul B. Kantor of the School of Communication, Information, and Library Studies, Rutgers University, for kindly providing us this example.
the sense that the elements are partitioned into a sequence of sets $S_1, \ldots, S_m$, such that all the elements in each $S_i$ are incomparable, but everything in $S_i$ is less than or equal to everything in $S_{i+1}$. The outcomes of the clinical trials for carcinogenicity have this sort of ordering. A linear order, as in the tax auditing problem, is of course an interval order.

Another instance that might call for an interval ordering is estimating the value of a piece of property. Suppose we suspect that certain combinations of property attributes justify assessing the value to be within certain ranges. The ranges may overlap, but the outcomes nonetheless have an interval order. If higher ranked attribute values are more desirable, it may be reasonable to assume monotonicity.

We begin in the next section with a precise statement of the problem and a small example. In Section 3 we show that finding the best monotone approximation is equivalent to a generalization of the “maximal closure” problem on a directed graph [11]. When there are two ordered outcomes, the generalization coincides with the maximal closure problem. In Section 4 we show that the generalized maximal closure problem can be reduced to the maximal closure problem when the outcomes have an interval order. Since the maximal closure problem can be solved with a minimum cut computation, we can solve the best approximation problem likewise. The complexity of our algorithm is $O(n^3m^3)$, where $n$ is the number of possible vectors $x$ and $m$ the number of possible outcomes.

2 Problem Statement

Let $(X, \prec)$ be a finite partially ordered set of attribute vectors $x$ and $(V, \prec)$ a finite partially ordered set of outcomes. The mapping $\mu : X \times V \mapsto \mathbb{Z}_+$ indicates the number of times an attribute vector $x$ resulted in a given outcome; that is, $\mu(x, \alpha)$ is the number of observations $(x, y)$ for which $g(x, y) = \alpha$. The penalty function is $c : V \times V \mapsto \mathbb{Z}_+$, where $c_\beta, \alpha$ is the penalty for predicting outcome $\beta$ when the observed outcome is $\alpha$. The problem is to define $f(x)$ on the observed values of $x$ so as to minimize the total penalty,

$$\epsilon[f, \mu] \overset{\text{def}}{=} \sum_{x \in X} \sum_{\alpha \in V} c_{f(x), \alpha} \mu(x, \alpha).$$

For an unobserved value $x'$, we can let $f(x')$ take any value for which $f(x') \not\geq f(x)$ for all observed values $x$ satisfying $x' \preceq x$, and for which $f(x) \not\geq f(x')$ for all observed values $x$ satisfying $x \preceq x'$.

We will use the following as a running example through the rest of the paper. Suppose we wish to determine the carcinogenicity of chemicals on the basis of only two tests. We have clinical data for five chemicals (Table 1). $N, D, I$, and $P$ respectively denote the possible test results (negative, no data, indefinite, and positive), and $h, u, p$, and $d$ denote the four possible outcomes (harmless,
undetermined, potentially dangerous, and dangerous. The table displays the number \( \mu(x, \alpha) \) of times chemicals with attribute vector \( x \) led to clinical result \( \alpha \).

To define the penalty for error, let us rank the outcomes: \( h \) has rank 1, \( u \) and \( p \) have rank 2, and \( d \) has rank 3. Then we might say that the penalty for an error is one more than the distance between the ranks of the predicted and observed outcome when they are different, and zero otherwise. For instance, predicting a potentially dangerous (\( p \)) chemical to be harmless (\( h \)) brings penalty 2, and predicting it to be undetermined (\( u \)) brings penalty 1. The total penalty \( \sum_{\beta} c_{\alpha \beta} \mu(x, \beta) \) for predicting that chemicals described by \( x \) have clinical outcome \( \alpha \) appears in the middle of the table. For instance, the penalty for predicting that chemicals with test results (\( N \)) are dangerous is 8, because they were observed to be harmless on 2 occasions and to have undetermined effect on 1 occasion.

If there are no restrictions on the approximating function \( f \), the best approximation is clearly the function \( f_0 \) shown in the table. We set \( f_0(N, N) = h \), for instance, since predicting \( h \) results in the least penalty (2). In general, \( f_0 \) is defined by the equation \( \mu(x, f_0(x)) = \max_{\alpha \in V} c_{f_0(x), \alpha} \mu(x, \alpha) \) for \( x \in X \). But in this example \( f_0 \) is not monotone, since \( f_0(I, N) > f_0(I, U) \) even though \( (I, N) \not\leq (I, U) \). Our task is to find the best monotone approximation.

### 3 Maximal V-Chains

Let \( G = (N, A, w) \) be a directed graph, where \( N \) stands for the set of nodes, \( A \) denotes the set of arcs, and \( w_x, x \in N \) are real weights associated with the nodes. A subset \( C \subseteq N \) of the nodes is called a closure of \( G \), if \( (x, y) \in A, x \in C \) implies \( y \in C \). A maximal closure of the graph \( G \) is a closure \( C \) for which \( w(C) \) is maximal.

As an example, consider the graph \( G_0 \) in Fig 1, which has nodes corresponding to the five values of \( x \) in Table 1. (Ignore the large circles for the moment.) \( G_0 \) contains an arc \( (x, x') \) whenever \( x \preceq x' \). The closure of a set \( C \) of nodes (\( x \)'s) is the set of all \( x \)'s greater than or equal to the \( x \)'s in \( C \). For instance, the
Figure 1: Illustration of a V-chain.

closure of \{3\} is \{3, 4, 5\}.

J.-C. Picard [12] introduced the problem of finding a maximal closure in a given directed graph, and he showed that the maximal closure can be found as the source side of a minimum cut in an associated network. This problem has many interesting applications, e.g. [1, 7, 8, 9, 10, 12, 13, 16].

In this section we introduce a generalization of this problem, which is shown to be equivalent with the problem of V-monotone approximation, and which is reducible in certain cases to the problem of maximal closures.

Let us consider a partially ordered finite set \((V, \prec)\), and let \(\alpha \parallel \beta\) denote that \(\alpha\) and \(\beta\) are incomparable elements of \(V\).

A collection \(C = \{C_\alpha | \alpha \in V\}\) of subsets of \(G\) will be called a V-chain if

\[
\cup_{\alpha \in V} C_\alpha = N, \tag{2}
\]

\[
\alpha \geq \beta \implies C_\alpha \subseteq C_\beta, \tag{3}
\]

\[
\alpha \parallel \beta \implies C_\alpha \cap C_\beta \subseteq \bigcup_{\gamma \geq \alpha, \gamma \geq \beta} C_\gamma, \tag{4}
\]

and for every \(\alpha \in V\)

\[
(i, j) \in A, \ i \in \left(C_\alpha \cup \bigcup_{\beta \parallel \alpha} C_\beta\right) \implies j \in \left(C_\alpha \cup \bigcup_{\beta \parallel \alpha} C_\beta\right). \tag{5}
\]

For example, suppose we let \(V = \{h, u, p, d\}\) as in Table 1. One possible V-chain in the graph of Figure 1 is illustrated by circles; that is, \(C_h = \{1, 2, 3, 4, 5\}\), \(C_u = \{2, 3, 5\}\), \(C_p = \{4, 5\}\), and \(C_d = \{5\}\).

Note that the sets \(C_\alpha\) of this V-chain are determined by an approximating function, namely the function \(f_1\) displayed in Table 1. \(C_d\) is the set of nodes classified dangerous \((d)\) by \(f_1\), \(C_p\) the set of nodes classified dangerous or potentially dangerous \((d\ or\ p)\), \(C_u\) the set of nodes classified dangerous or undetermined \((d\ or\ u)\), and \(C_h\) the set of all nodes. We show that any V-chain is determined by some approximating function.

**Lemma 1** If \(C = \{C_\alpha | \alpha \in V\}\) is a V-chain of \(G\), then for every \(x \in N\) there is a unique element \(f_C(x) \in V\) such that

\[
x \in C_\alpha \text{ iff } \alpha \leq f_C(x). \tag{6}
\]

**Proof.** For any \(x \in N\) let \(V(x) = \{\alpha | x \in C_\alpha\}\). With this notation, the lemma states that \(V(x)\) has a unique maximal element. The condition (2) implies that \(V(x) \neq \emptyset\) for every \(x \in N\). Let us assume now that there is an \(x \in N\) for which
$V(x)$ has at least two different maximal elements, say $\alpha$ and $\beta$. Obviously, then $\alpha \parallel \beta$, and thus by (4), there must be an element $\gamma$ of $V$ with $\gamma \in V(x)$, $\gamma \geq \alpha$ and $\gamma \geq \beta$. This contradicts the maximality of $\alpha$ and $\beta$ in $V(x)$, and this contradiction proves the lemma.

\[ \square \]

If we introduce the notation $C^V_\alpha \overset{\text{def}}{=} C_\alpha \cup \bigcup_{\beta \parallel \alpha} C_\beta$, then condition (5) is equivalent to requiring that the sets $C^V_\alpha$ are closures of $G$.

If, furthermore, $w_{x,\alpha}$ are given real weights, $x \in N$ and $\alpha \in V$, then a maximal $V$-chain of $(G, w)$ is a $V$-chain $C$, which maximizes

\[ w(C) \overset{\text{def}}{=} \sum_{\alpha \in V} \sum_{x \in C_\alpha} w_{x,\alpha}. \]

We shall show below that the problem of monotone approximation is equivalent to a maximal $V$-chain problem in an associated graph.

Given an instance of the monotone approximation problem, let $G = (N, A)$ be an associated directed graph, where $N = X$, and $A$ is formed by the arcs $(x, y)$ for which $x \prec y$, and there is no other point $z \in X$ with $x \prec z \prec y$. ($G$ represents the Hasse diagram of the partially ordered set $(X, \prec)$.) We formed the graph in Figure 1 associated with the problem of Table 1 in this manner.

Note that the function $f_1$ that defines the $V$-chain depicted in Figure 1 is monotone. In fact we have the following result.

Theorem 2 Let $G$ be the directed graph associated with a given instance of the monotone approximation problem. Then there is a one-to-one correspondence between the monotone functions $f : X \mapsto V$ and the $V$-chains of $G$.

Proof. Let us assume first that $f : X \mapsto V$ is a $V$-monotone function. For $\alpha \in V$ let

\[ C_\alpha = \{x | f(x) \geq \alpha\}, \]  

(7)

and let $C_f = \{C_\alpha | \alpha \in V\}$. We shall show that $C_f$ is a $V$-chain. Condition (2) follows by the fact that $f(x)$ is defined for every $x \in X$. Conditions (3) and (4) are direct consequences of the definition of the sets $C_\alpha$ for $\alpha \in V$. Finally, condition (5) will be implied by the monotonicity of $f$, as follows. Let $(x, y) \in A$ be an arbitrary arc with $x \in C_\alpha$. We must show that either $y \in C_\alpha$ or $y \in C_\beta$ for some $\beta \parallel \alpha$. The monotonicity of $f$ implies that

\[ f(x) \nparallel f(y), \]  

(8)

and

\[ f(x) \geq \alpha \]  

(9)

follows by the definition of $C_\alpha$. Now, condition (8) implies that either $f(y) \geq f(x)$ or $f(y) \nparallel f(x)$. If $f(y) \geq f(x)$, then the transitivity of $\geq$ implies $f(y) \geq \alpha$,
and thus \( y \in C_\alpha \) in this case. If \( f(y) \mid f(x) \), then \( y \in C_{f(y)} \subseteq C_\alpha \), thus (5) holds in both cases.

For the converse relation, let \( C = \{ C_\alpha \mid \alpha \in V \} \) be a collection of subsets of \( G_M \), satisfying conditions (2)-(5). Then the function \( f_C \), defined by Lemma 1 is clearly an \( X \mapsto V \) mapping, and property (5) immediately implies the monotonicity of it.

We shall show that \( \{ x \mid f_C(x) \geq \alpha \} = C_\alpha \) for every \( \alpha \in V \). For this, let first \( x \in C_\alpha \). Then \( f_C(x) \geq \alpha \) follows by the definition of \( f_C \). On the other hand, if \( x \not\in C_\alpha \), then by condition (3) \( x \not\in C_\beta \) whenever \( \beta \geq \alpha \). Therefore \( f_C(x) \not\geq \alpha \).

We now know that finding a \( V \)-chain is equivalent to finding a monotone approximation \( f \), but to equate the maximal \( V \)-chain problem to the monotone approximation problem, we must relate the weight of a \( V \)-chain to the penalty associated with an approximation. This is done in the example of Table 1 as follows. The weight \( w_{x,h} \) contributed to a \( V \)-chain by each \( x \) lying in \( C_h \) but in no smaller \( C_\alpha \) is the negation of the penalty that results from predicting chemicals with attributes \( x \) to be harmless. The total weight \( w_{x,h} + w_{x,u} \) contributed by each \( x \) in \( C_u \setminus C_d \) is the negation of the penalty for predicting chemicals with attributes \( x \) to be undetermined, and similarly for \( w_{x,h} + w_{x,p} \). The total weight \( w_{x,h} + w_{x,u} + w_{x,p} + w_{x,d} \) contributed by each \( x \) in \( C_d \) is the negation of the penalty for predicting \( x \) to be dangerous. The resulting weights \( w_{x,\alpha} \) for each outcome \( \alpha \) appear on the right side of Table 1. Suppose, for instance, that the predictions are given by \( f_1 \). The penalty incurred for each \( x \) is shown in boldface in the middle of the table. The penalty for each \( x \) is the same as the negated sum of the boldface weights that appear at the end of that row. It is clear that finding a maximal weight \( V \)-chain over these weights is equivalent to finding predictions that minimize the penalty.

**Theorem 3** Given an instance of the monotone approximation problem, the following equations have, for every \( x \in X \), a unique solution in variables \( w_{x,\alpha} \), \( \alpha \in V \).

\[
\sum_{\nu \leq \alpha} w_{x,\nu} = - \sum_{\nu \in V} c_{x,\nu,\alpha} \mu(x, \nu) \quad \alpha \in nV
\]  

(10)

Let furthermore \( f : X \mapsto V \) be an arbitrary function, and let \( C_f = \{ C_\alpha \mid \alpha \in V \} \) be defined by (7). Then

\[
d[f, \mu] = -w(C_f),
\]  

(11)

where \( (w_{x,\alpha}, \alpha \in V) \) is the unique solution of (10), and

\[
w(C_f) = \sum_{\alpha \in V} \sum_{x \in C_\alpha} w_{x,\alpha}.
\]

**Proof.** The existence and uniqueness of \( w \) follows easily from the fact that rearranging the variables in (10) by a linear extension of \( (V, <) \), the matrix of
the left hand side in (10) becomes a lower triangular \(0 - 1\) matrix, with 1’s in the main diagonal. From this, it follows that if \(\mu\) and \(c\) are integral, then \(w\) is also integral.

To prove (11) we can proceed as follows.

\[
  w(C_f) = \sum_{\alpha \in V} \sum_{x \in C_{\alpha}} w_{x, \alpha} = \\
  = \sum_{x \in X} \sum_{\alpha; \alpha \leq f(x)} w_{x, \alpha} = \\
  = \sum_{x \in X} \left(-\sum_{\nu \in V} c_{f(x), \nu} \mu(x, \nu)\right) = \\
  = -\epsilon[f, \mu].
\]

Here the second line comes from the first one by a simple rearrangement, the third line comes by (7), and to get the fourth equation we applied (10). For the last equation we simply applied (1).

\(\square\)

Corollary 4 For a given instance of monotone approximation, let \(w = (w_{x, \alpha}|x \in X, \alpha \in V)\) be determined by (10). Then finding the best \(V\)-monotone fit is equivalent to finding the maximal \(V\)-chain of the associated graph.

4 Maximal \(V\)-Chains via Maximal Closures

Let us consider in this section interval orders \((V, <)\). It is easy to see that such an order satisfies the following condition:

\[
  \text{There is an indexing } V = \{v_0, v_1, ..., v_{n}\} \text{ of } (V, <) \text{ such that for every } \alpha \in V, \{\nu \in V|\nu < \alpha\} \equiv \{\nu|i < i^-\} \text{ for some index } i^-.
\]

In this section we show that, in case of partially ordered sets \((V, <)\) satisfying condition (13), a maximal \(V\)-chain can be determined by computing a maximal closure in an associated graph.

Let \(G = (N, A)\) be a given directed graph, let \((V, <)\) be a finite partially ordered set, as in the previous section, and let \(w_{x, \alpha}\) be given reals for \(x \in N\) and \(\alpha \in V\).

It follows immediately from (13) that for any \(V\)-chain \(C = \{C_\alpha|\alpha \in V\}\), and for any index \(i,\)

\[
  C^V_{v_i} = \bigcup_{j = i^-}^{v} C_{v_j}.
\]

In the example we can let \(\{v_0, v_1, v_2, v_3\} = \{h, u, p, d\}\), so that for the \(V\)-chain illustrated in Figure 1 we have \(i^- = 2, 2, 4\) for \(i = 2, 3, 4\). Let \(k_C(x)\) be defined
for any $x \in N$ as the index of $f_C(x)$ in this linear order, where $f_C$ is again the function defined in Lemma 1. In other words, $f_C(x) = \nu_{k_C(x)}$ for every $x \in N$.

Let us define now a directed graph, $\hat{G} = (\hat{N}, \hat{A})$ associated to $G$ and to $(V, <)$, as follows. The graph $\hat{G}$ will contain $v|N|$ nodes, each node $x \in N$ will have $v$ “relatives” in $\hat{N}$, denoted by $x_1, x_2, \ldots, x_v$. We include the arcs $(x_i, y_i)$ for $(x, y) \in A$ and for indices $i$ with $\nu_{i-1} < \nu_i$ in $(V, <)$. We also add the arcs $(x_i, x_{i-1})$ to $\hat{A}$ for every $x \in N$ and $i = 2, \ldots, v$. In other words, we place $v$ copies of the node set $N$ on the top of each other, and we include all the arcs of $G$ in the $i$th level, if $\nu_{i-1}$ and $\nu_i$ are comparable; otherwise the $i$th level is an empty graph. In addition to these, we include a directed arc from each vertex to its lower relative, too. The graph $\hat{G}$ associated with the graph of Figure 1 appears in Figure 2. We shall define real weights $\hat{w}_x$ associated to every vertex $x_i \in \hat{N}$ as the unique solution of the following system of linear equations.

$$\sum_{i=1}^{j} \hat{w}_{x_i} = -w_{x, v_0} + \sum_{\alpha \leq v_j} w_{x, \alpha} \quad j = 1, \ldots, v; \quad x \in N.$$  \hspace{1cm} (15)

This system of equations is clearly nonsingular and has determinant equal to 1, so that it not only has a unique solution $\hat{w}$, but $\hat{w}$ must also be integral whenever $w$ is integral. The weights $w_{x_i}$ for our example also appear in Figure 2.

**Theorem 5** There is a one-to-one correspondence between the $V$-chains of $G$ and the closures of $\hat{G}$. Furthermore, if $C = \{C_\alpha|\alpha \in V\}$ is a $V$-chain of $G$, and $\hat{C}$ is the corresponding closure of $\hat{G}$, then

$$w(C) = \sum_{x \in N} w_{x, v_0} + \hat{w}(\hat{C}),$$  \hspace{1cm} (16)

where $\hat{w}$ is the vector defined by (15).

Since $\sum_{x \in N} w_{x, v_0}$ is a constant here, it follows that finding a maximal $V$-chain in $G$ is equivalent with finding a maximal closure in $\hat{G}$.

**Proof.** Let us consider first a $V$-chain of $G$, $C = \{C_\alpha|\alpha \in V\}$, and define a subset of the nodes of $\hat{G}$ as follows

$$\hat{C}_C \defeq \bigcup_{i=1}^{v} \{x_i | x \in \bigcup_{j=m_i}^{v} C_{v_j}\}.$$  \hspace{1cm} (17)

We shall show that $\hat{C}_C$ is a closure in $\hat{G}$. For $x_i \in \hat{C}_C$ it follows that $x_{i-1} \in \hat{C}_C$ simply by the definition (17). Let us consider now an arc $(x_i, y_i) \in \hat{A}$ of $\hat{G}$,
where \( x_i \in \hat{C}_C \). By the definition of \( \hat{G} \), \( (x_i, y_i) \in \hat{A} \) only if \( \nu_{i-1} < \nu_i \). This
implies \( i^- = i \) by (13), and thus \( \cup_{j=i}^{\nu_i} C_{\nu_j} \) is a closure of \( G \) follows by (14) and
(5).

Now \( x_i \in \hat{C}_C \) implies, by definition that \( x \in C_{\nu_j} \) for some \( j \geq i \), i.e. that
\( x \in \cup_{j=i}^{\nu_i} C_{\nu_j} \).

From \( (x_i, y_i) \in \hat{A} \) it follows also that \((x, y) \in A\), therefore \( y \in \cup_{j=i}^{\nu_i} C_{\nu_j} \), and
thus \( y_i \in \hat{C}_C \).

Conversely, let us consider now a closure \( \hat{C} \) of \( \hat{G} \). Let \( A_0 \overset{\text{def}}{=} N \), and let
\( A_i \overset{\text{def}}{=} \{ x|x_i \in \hat{C} \}, \) for \( i = 1, 2, ..., v \). Let furthermore,
\[
C_{\nu_i} \overset{\text{def}}{=} (A_i \setminus A_{i+1}) \cup \bigcup_{j: \nu_j < \nu_i} A_j,
\]
for \( i = 0, 1, ..., v-1 \), and let \( C_{\nu_v} = A_v \). We shall show that \( \mathcal{C} = \{ C_{\nu_i} | i = 0, 1, ..., v \} \)
is a V-chain of \( G \) such that \( \hat{C} \equiv \hat{C}_C \).

It is easy to observe that
\[
N = A_0 \supseteq A_1 \supseteq \cdots \supseteq A_v,
\]
simply by the definition of \( A_i \)'s. Using conditions (13) and (14) it follows also that
\[
C_{\nu_i}^{\nu_i} = A_i
\]
for every index \( i \), and, in particular, \( A_i \) is a closure whenever \( \nu_{i-1} < \nu_i \).

Properties (2)-(5) are now easy consequences of (18)-(20). It is also clear
that for any index \( 1 \leq i \leq v \), \( A_i = \cup_{j=i}^{\nu_i} C_{\nu_j} \), therefore \( x_i \in \hat{C} \) iff \( x \in A_i \),
implying \( \hat{C} \equiv \hat{C}_C \).

Finally, (16) follows easily by (15) and (17):
\[
\hat{u}(\hat{C}) = \sum_{x_i \in \hat{N}} \hat{w}_{x_i} = \\
= \sum_{x_i \in \hat{N}} \sum_{i=1}^{f(x)} \hat{w}_{x_i} = \\
= \sum_{x_i \in \hat{N}} \left( \sum_{a \leq f(x)} w_{x,a} - w_{x,\nu_i} \right) = \\
= w(\mathcal{C}) - \sum_{x \in \hat{N}} w_{x,\nu_i}.
\]

The maximal closure for Figure 2 in encircled in the figure. Note this solution
solution corresponds to the function \( f_1(x) \) defined in Table 1. Thus \( f_1 \) is a best
approximation to the data.

We conclude by indicating in Table 2 the possible values of \( f_1(x) \) for both
observed and unobserved values of \( x \). The observed values are in boldface. Note,
for instance, that \( f_1(N, U) \) cannot be \( d \), since \( (N, U) \preceq (I, U) \) and \( f_1(IU) = u \),
and \( d > u \). In this example the value of \( f_1(x) \) is completely determined for
only one unobserved \( x \), namely \( x = PP \).
Table 2: Function values consistent with an optimal approximation.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f_1(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N,N)$</td>
<td>$h$</td>
</tr>
<tr>
<td>$(U,N)$</td>
<td>$h,u,p$</td>
</tr>
<tr>
<td>$(N,U)$</td>
<td>$h,u,p$</td>
</tr>
<tr>
<td>$(I,N)$</td>
<td>$u$</td>
</tr>
<tr>
<td>$(N,I)$</td>
<td>$h,u,p,d$</td>
</tr>
<tr>
<td>$(U,U)$</td>
<td>$h,u,p$</td>
</tr>
<tr>
<td>$(I,U)$</td>
<td>$u$</td>
</tr>
<tr>
<td>$(P,N)$</td>
<td>$u,p,d$</td>
</tr>
<tr>
<td>$(U,I)$</td>
<td>$h,u,p,d$</td>
</tr>
<tr>
<td>$(I,I)$</td>
<td>$u,p,d$</td>
</tr>
<tr>
<td>$(N,P)$</td>
<td>$h,u,p,d$</td>
</tr>
<tr>
<td>$(P,U)$</td>
<td>$p$</td>
</tr>
<tr>
<td>$(P,I)$</td>
<td>$u,p,d$</td>
</tr>
<tr>
<td>$(U,P)$</td>
<td>$h,u,p,d$</td>
</tr>
<tr>
<td>$(I,P)$</td>
<td>$d$</td>
</tr>
<tr>
<td>$(P,P)$</td>
<td>$d$</td>
</tr>
</tbody>
</table>

References


