TECHNICAL REPORT
CSD 880052
R-101
March 1988

INFLUENCE DIAGRAMS AND D-SEPARATION

Thomas Verma
<verma@cs.ucla.edu>

Judea Pearl
<judea@cs.ucla.edu>

Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024-1600

ABSTRACT

Some properties of influence diagrams are examined in light of d-separation, a sound and complete criterion for identifying the independencies represented by influence diagrams. An axiomatic characterization of d-separation is presented and used to develop the necessary and sufficient conditions for the entailment and equivalence of influence diagrams. In addition, several previous results are re-examined in terms of d-separation.

* This work was supported in part by the National Science Foundation Grants #DCR 85-01234 and #IRI 86-10155

Influence Diagrams and d-Separation*

Thomas Verma <verma@cs.ucla.edu>

Judea Pearl <judea@cs.ucla.edu>

Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024-1600

ABSTRACT

Some properties of influence diagrams are examined in light of d-separation, a sound and complete criterion for identifying the independencies represented by influence diagrams. An axiomatic characterization of d-separation is presented and used to develop the necessary and sufficient conditions for the entailment and equivalence of influence diagrams. In addition, several previous results are re-examined in terms of d-separation.

INTRODUCTION

Influence diagrams are directed acyclic graphs which represent sets of conditional independence statements. They may be used in decision analysis, evidential reasoning and statistical modeling [Howard and Matheson, 1981; Pearl, 1986; Smith, 1987]. d-Separation is a graphical criterion for identifying independencies represented by influence diagrams. It is sound when used with influence diagrams based upon semi-graphoids [Pearl and Verma, 1987]. That is, if the independence statements used to design an influence diagram obey the axioms of semi-graphoids (see Table 1), then every independence that d-separation identifies is correct. Examples of semi-graphoids are the conditional independence relationship in probability theory or embedded multivalued dependencies (EMVD) in relational databases [Fagin, 1977]. Further if all that is known is that the independence statements obey the properties of semi-graphoids, or else, that they are based upon some probability distribution, then d-separation is complete with respect to the specification set [Geiger and Pearl, 1988]. That is, d-separation will correctly identify every possible independence statement which follows from the information used to design the influence diagram.

A dependency model is a set of conditional independence statements of the form $I(X, Z, Y)$. The statement $I(X, Z, Y)$ specifies that the values of the variables in the set X are independent of the values of the variables in the set Y once the values of the variables in the set Z are known. A semi-graphoid is any dependency model which obeys the axioms of Table 1. These axioms, common to most every conventional definition of conditional independence, are very similar to the axioms of *Generalized Conditional Independence* [Dawid, 1979; Smith, 1987] second axiom is omitted. In an alternate treatment of semi-graphoids the first two axioms are omitted, and only statements about mutually exclusive sets of variables are considered. Under this restriction, generalized conditional independence and semi-graphoids are equivalent. The differences between these two treatments are merely cosmetic, since no substantial use is made of the first two axioms.

* This work was supported in part by the National Science Foundation Grants #DCR 85-01234 and #IRI 86-10155

Reflexivity	$X \subseteq Z \Rightarrow I(X, Z, Y)$
Relative Disjunction	$I(X, Z, Y) \Rightarrow X \cap Y \subseteq Z$
Symmetry	$I(X, Z, Y) \Rightarrow I(Y, Z, X)$
Decomposition	$I(X, Z, YW) \Rightarrow I(X, Z, Y)$
Weak Union	$I(X, Z, YW) \Rightarrow I(X, ZY, W)$
Contraction	$I(X, Z, Y) \wedge I(X, ZY, W) \Rightarrow I(X, Z, YW)$

Table 1: semi-graphoid axioms

Since a perfect representation of an arbitrary semi-graphoid, probabilistic dependency model, or relational dependency model (EMVD) requires exponential space on average [Verma, 1987] it is necessary to settle for a partial representation. The influence diagram used in conjunction with d-separation is a good partial representation for these classes of dependency models.

D-SEPARATION AXIOMS

The definition of d-separation is algorithmic: two sets of nodes X and Y are d-separated given a third set Z if and only if there is no active bi-directed path from a node in X to a node in Y. A path is activated by a set Z if every node with converging arrows either is or has a descendant in Z and every other node along the path is not in Z. Further there is considered to be a vacuous path from a node to itself, and this path is only de-activated by the very same node. This algorithmic definition is good in that it insures computability, but it is theoretically awkward. An axiomatic definition would help shed light on the structure of d-separation.

It is straightforward to show that d-separation obeys all the axioms of semi-graphoids, but the complete set of axioms, listed in Table 2, are a bit more complex and dubious to discern.

Chordality $\quad I(x, zw, y) \wedge I(z, xy, w) \Rightarrow I(x, z, y) \vee I(x, w, y)$

Weak Transitivity $\quad I(X, Z, Y) \wedge I(X, Zw, Y) \Rightarrow I(X, Z, w) \vee I(w, Z, Y)$

d-Transitivity

$$I(X, Z, Y) \Leftrightarrow \forall_{x \in X,\, y \in Y \text{ or } x \in Y,\, y \in X} \quad \{x \in Z \vee y \in Z\} \ \textit{or} \ \left\{ \begin{array}{c} \{x \neq y\} \\ \textit{and} \\ \left\{ \forall_{a,b \in U} \exists_{c \in Z} I(x, Zy, a) \vee I(b, Zx, y) \vee a \text{—} b \vee a \not\text{—} b | c \right\} \\ \textit{and} \\ \left\{ \forall_{a \in U} I(a, Zx, y) \vee x \not\text{—} a \vee \left[\exists_{w \in U} x \not\text{—} w \wedge x \text{—} w | a \right] \right\} \end{array} \right\}$$

Table 2: d-separation axioms (lower case variables represent singletons)

The first two axioms are well known [Pearl, 1986], whereas the last most complex axiom is best posed in terms of adjacency and conditional adjacency. A direct consequence of the acyclicity of influence diagrams is that adjacency is equivalent to non-d-separability, as stated in the following lemma:

Lemma: Two nodes a and b of an influence diagram are adjacent, written $a — b$, if and only if there is no set S not containing a or b which d-separates a and b.

Proof: If a and b are adjacent, then the path consisting of the link between them will serve to connect them, and cannot be deactivated by any set of nodes except those sets containing either a or b. If a and b are not adjacent, then the set S of the parents of a and b will not contain either node and will serve to deactivate any paths between the nodes. Any path between the nodes which has an outward pointing arc on either end will be trivially deactivated by S. Any path between the nodes which has both ends pointing inward must contain at least one head to head node. Consider the head to head node on the path which is closest to a. There is a directed path from a to this node, and for the path to be active, either this node or one of its direct descendents must be in S. This node in S cannot be a parent of a (otherwise there would be a directed loop) so, either it is a parent of b or the path is not active. If it is a parent of b then there is a directed path from a to b. Analysis of the head to head node closest to b leads to the conclusion that either there is no active path between a and b or there is a directed path from b to a. Since there can be no directed loops, there must be no active path between a and b given S thus a and b are d-separable. □

This lemma can be used to define the term adjacency with respect to general dependency models. Thus, if a and b are variables of some dependency model M, then they are adjacent in M, written $a \underset{M}{—} b$, if and only if they are not independent in M given any set that does not contain a or b.

A natural extension of adjacency is conditional adjacency. For any three variables a, b and c of a dependency model M, a and b are conditionally adjacent given c, written $a \underset{M}{—} b|c$ if and only if once c is given, a and b can not become independent, i.e. a and b are not independent given any set which does not contain a or b but does contain c. Graphically, for some influence diagram D, $a \underset{D}{—} b|c$ whenever a and b are the common parents of c or some ancestor of c, or a and b unconditionally adjacent in D. This observation leads to the following lemma relating adjacency, conditional adjacency and head to head nodes.

Lemma: For any three nodes a, b and c of an influence diagram, $a \not— b$ and $a — b|c$ if and only if a and b are the non-adjacent common parents of some node d which is either equal to c or is an ancestor of c.

Proof: Assume $a — b|c$ then, by definition, any set along with c will activate a path between a and b, even the empty set. Thus there is a path between a and b which can be activated by c alone, but this path must not be active given only the empty set since $a \not— b$. Therefore there must be a head to head node d on this path, and it must either be equal to c or be an ancestor of c. Furthermore d must be a descendent of both a and b, otherwise there would be a set which separates d from either a or b, and this set along with c would then serve to separate a from b which would contradict the assumption that $a — b|c$. The converse follows trivially from the definitions of adjacency, conditional adjacency and d-separation. □

d-Transitivity is best understood in the contrapositive: if X and Y are not d-separated, then there must be a path between them. If there is a head to head node on the path, then the third part of the right side of the axiom holds, where a and b are the non-adjacent common parents for the head to head node. And if an arc on either end points outward, then the fourth part of the right hand side holds. The parenthesised portion of this part ensures that the last arc is not compelled to point inward. The first two parts correspond to the 'base-steps' of reflexivity and relative decomposition. At least one of the parts apply for any path. This exhaustiveness is the basic principle behind the proof of completeness of the axioms for d-separation.

Theorem 1: The axioms of Table 2 are a sound and complete characterization of d-separation.

Proof: The proof of soundness is straight forward from the definition of d-separation. The proof of completeness entails the construction of an influence diagram which perfectly represents any dependency model that satisfies the axioms. To construct such an influence diagram, simply place an arc between any two nodes whose corresponding variables are not separable in the dependency model. Then orient the arcs in the following manner, first for any three nodes which satisfy the three properties, $a — c$, $c — b$ and $a \not— b$ and $a — b \mid c$, orient the two such that $a \rightarrow c$ and $b \rightarrow c$. Finally orient any remaining arcs in any fashion such that no new non-adjacent parents are created. Weak transitivity insures that no arc will receive two directions in the first orientation phase, and chordality ensures that the second phase can be completed. It remains to show that any influence diagram generated by this algorithm is a perfect map of the dependency model.

(D-mapness) Consider an arbitrary non-separation $\neg I(X, Z, Y)$ in the influence diagram. It must be the case that there is a path from some node in X to some node in Y. Induction on the length of the path in conjunction with d-transitivity will yield that X and Y must be dependent given Z in the original model. Any path either contains a head to head node or an outward link on an end, thus can be broken into one or two shorter paths (possible of length zero corresponding to base steps) these along with the only if direction of the d-transitivity axiom complete the induction.

(I-mapness) Consider an arbitrary separation $I(X, Z, Y)$ in the influence diagram. This time upward induction and the if direction of the d-transitivity axiom are used to show that the corresponding statement must be in the original model. Suppose that $I(X, Z, Y)$ is not in the original model, then one of the two consequents of the d-transitivity axiom must be false. In either case there is one or two independence statements which must be false, and inductively mean that certain paths must exist. These paths along with the corresponding links will imply that there is a path between X and Y which is activated by Z in the influence diagram. This contradictory to the hypothesis, so the independence must hold in the original model. □

INFLUENCE DIAGRAM ENTAILMENT

The complexity of the d-separation axioms inhibits their direct practical use. In fact, they cannot be used as a set of (computable) inference rules, nor do they help to solve the membership problem. But the ideas behind the axioms do help to shed light on the structure of d-separation. One problem concerning influence diagrams that has received much attention is that of entailment: determination of the conditions under which one influence diagram entails another. The notions of adjacency and conditional adjacency permit a concise statement of the conditions for influenced diagram entailment.

Theorem 2: For any two influence diagrams, D and E over the same set of nodes N, $D \Rightarrow E$ if and only if all three of the following hold:

1. $$\forall_{a,b \in N} \; a \underset{D}{—} b \Rightarrow a \underset{E}{—} b$$

2. $$\forall_{a,b,c \in N} \; a \underset{D}{—} b \mid c \Rightarrow a \underset{E}{—} b \mid c$$

3. $$\forall_{a,b,c \in N} \; a \underset{E}{—} b \mid c \wedge a \underset{D}{—} c \wedge b \underset{D}{—} c \Rightarrow a \underset{D}{—} b \mid c \vee a \underset{E}{—} b$$

Proof: The only if portion follows directly from the definitions, suppose that $D \Rightarrow E$. The first part must hold, otherwise there would be a separation in E of a and b, which would not hold in D. The same argument holds for part two. The antecedent of part three implies that a and b are the head to head parents of c in E, and that they are linked to c in D. If they are not the head to head parents of c in D, and they are not linked in E, then they are separable in E and not in D, thus part three must hold.

The if portion follows from induction on the a path in D. Suppose that one through three hold, it is enough to show that every dependence in D is also in E to complete the proof. Let $\neg I(X,Z,Y)$ be such a dependence, then there must be a bi-directed path from some node $x \in X$ to some $y \in Y$ which is activated by Z. But this same path must exist in E by part one. Further part two insures that any active head to head node on the path in D is also an active head to head node in E. The only way that this path would not be active in E is if it had a head to head node on it which was not on it in D, but part three curtails this possibility, thus the dependency is in E. □

Corollary: For any two influence diagrams, D and E over the set of nodes N, $D \equiv E$ if and only if both of the following hold:

1. $\forall_{a,b \in N} \; a \underset{D}{—} b \Leftrightarrow a \underset{E}{—} b$

2. $\forall_{a,b,c \in N} \; a \underset{D}{—} b \,|\, c \Leftrightarrow a \underset{E}{—} b \,|\, c$

APPLICATIONS

Theorem 2 clearly states the conditions under which sound transformations of influence diagrams can be preformed. The first condition states that no arcs can be removed, thus only reorientation and addition of arcs are premitted. The second and third conditions concern non-adjacent common parents. The second condition states that if a reorientation destroys a non-adjacent common parent triple then the parents must be joined by a link. The third condition states that if a reorientation creats a new pair of common parents then the pair must be joined by a link.

From this theorem, most results concerning manipulations of influence diagrams, e.g. arc reversal and node removal [Howard and Matheson, 1981] are immediate. For example consider arc reversal: if an arc is reversed, condition two implies that every set of nodes which were common parents but no longer are must be joined by a link. Condition three dictates which new parents must also be linked (only those which were created as a direct result of reversing the arc, and not those created by the addition of arcs). Theorem 2 implies that the arc reversal algorithm of Howrd and Matheson is minimal in the number of added arcs when only one arc is to be reversed, but that it is not minimal when several arcs are reversed. Another result which follows immediately from theorem 2 is the lisence to add any arcs, which follows directly since it does not violate any conditions of the theorem. The corollary to theorem 2 implies the decomposition theorem of [Smith, 1987] which gives a sufficient condition for equivalence.

One major limitation of this theorem is that it only applies to diagrams over the same sets of nodes. The [Howard and Matheson, 1981] result which gives a sufficient condition for removal of nodes follows directly from the fact that leaf nodes (nodes with no descendents) can be removed along with theorem 2. Leaf node removal follows from the definition of d-separation, thus to remove an arbitrary node, simply reorient any links direct out of it (by use of arc reversal), then remove the node and any links directed at it. If several arcs are reversed simultaneously, then an application of theorem 2 to maintain soundness will create a diagram with fewer links than the sequential application of the Howard and

Matheson method. This still may not be the most efficient diagram with the removed nodes as the determination of necessary conditions for general entailment remains an open problem.

REFERENCES

A.P. Dawid, ''Conditional Independence in Statistical Theory,'' *Journal Royal Statistical Society,* Vol B, #42, pp. 105-130. 1979.

R. Fagin, ''Multivalued Dependencies and a New Form for Relational Databases,'' *ACM Transactions on Database Systems, 2, 3;* pp. 262-278. 1977.

D. Geiger & J. Pearl, ''On The Logic of Influence Diagrams,'' *Technical Report R-112,* Cognitive Systems Laboratory, UCLA. 1988.

R.A. Howard & J.E. Matheson, ''Influence Diagrams,'' chapter 8, in *The Principles and Applications of Decision Analysis,* Vol. II, Strategic Decisions Group, Menlo Park, California. 1981.

J. Pearl, ''Markov and Bayes Networks: a comparison of two graphical representations of probabilistic knowledge,'' *Technical Report CSD 860024,* Cognitive Systems Laboratory, UCLA. 1986.

J. Pearl & TS Verma, ''The Logic of Representing Dependencies by Directed Graphs,'' *Proceedings,* AAAI-87, Seattle, WA, July 1987

J.Q. Smith, ''Models, Optimal Decisions and Influence Diagrams,'' *Technical Report 106,* Department of Statistics, Warwick England. 1987.

TS Verma, ''Some Mathematical Properties of Dependency Models,'' *Technical Report,* Cognitive Systems Laboratory, UCLA. 1987.