GRAPHOIDS: A Graph-based Logic for
Reasoning About Relevance Relations *
or
When would x tell you more about
y if you already know z

by
Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@LOCUS.UCLA.EDU)

and

Azaria Paz
Computer Science Department
Technion, Israel Institute of Technology
Haifa, Israel
(paz%tech-sel.csnet@CSNET-RELAY.ARPA)

1

GRAPHOIDS: A Graph-based Logic for
Reasoning About Relevance Relations *

by
Judea Pearl
and
Azaria Paz

ABSTRACT

We consider 3-place relations $I(x, z, y)$ where, x, y, and z are three non-intersecting sets of elements (e.g., propositions), and $I(x, z, y)$ stands for the statement: "Knowing z renders x irrelevant to y." We give sufficient conditions on I for the existence of a (minimal) graph G such that $I(x, z, y)$ can be validated by testing whether z separates x from y in G. These conditions define a GRAPHOID.

The theory of graphoids uncovers the axiomatic basis of probabilistic dependencies and ties it to vertex-separation conditions in graphs. The defining axioms can also be viewed as inference rules for deducing which propositions are relevant to each other, given a certain state of knowledge.

1.   INTRODUCTION

Any system that reasons about knowledge and beliefs must make use of information about relevancies. If we have acquired a body of knowledge z and now wish to assess the truth of proposition x, it is important to know whether it would be worthwhile to consult another proposition y, which is not in z. In other words, before we consult y we need to know if its truth value can potentially generate new information relative to x, information not available from z. For example, in trying to predict whether I am going to be late for a meeting, it is normally a good idea to ask somebody on the street for the time. However, once I establish the precise time by listening to the radio, asking people for the time becomes superfluous and their responses would be irrelevant. Similarly, knowing the color of X's car normally tells me nothing about the color of Y's. However, if X were to tell me that he almost mistook Y's car for his own, the two pieces of information become relevant to each other. What logic would facilitate this type of reasoning?

In probability theory, the notion of relevance is given precise quantitative underpinning using the device of conditional independence. A variable x is said to be independent of y given the information z if

$$P(x, y \mid z) = P(x \mid z) P(y \mid z)$$

However, it is rather unreasonable to expect people or machines to resort to numerical verification of equalities in order to extract relevance information. The ease and conviction with which people detect relevance relationships strongly suggest that such information is readily available from the organizational structure of human memory, not from numerical values assigned to its components. Accordingly, it would be interesting to explore how assertions about

3

relevance can be tested in various models of memory and, in particular, whether such assertions can be derived by simple manipulations on graphs.

Graphs offer useful representations for a variety of phenomena. They give vivid visual display for the essential relations in the phenomenon and provide a convenient medium for people to communicate and reason about it. Graph-related concepts are so entrenched in our language that one wonders whether people can in fact reason any other way, except by tracing links and arrows and paths in some mental representation of concepts and relations. Therefore, if we aspire to use non-numeric logic to mimic human reasoning about knowledge and beliefs, we should make sure that most derivational steps in that logic correspond to simple operations on some graphs.

When we deal with a phenomenon where the notion of neighborhood or connectedness is explicit (e.g., family relations, electronic circuits, communication networks, etc.) we have no problem configuring a graph which represents the main features of the phenomenon. However, in modelling conceptual relations such as causation, association, and relevance, it is often hard to distinguish direct neighbors from indirect neighbors and, so, the task of constructing a graph representation becomes more delicate. Moreover, once we construct such a graph it is not always clear which of its topological properties carries meaningful information about the relations under study.

This paper studies the feasibility of devising graphical representations for relational structures in which the notion of neighborhood is not specified in advance. Rather, what is given explicitly is the relation of "in betweenness." In other words, we are given the means to

test whether any given subset S of elements *intervenes in a relation between* elements x and y, but it remains up to us to decide how to connect the elements together in a graph that accounts for these interventions.

The notion of conditional independence in probability theory is a perfect example of such a relational structure. For a given probability distribution P and any three variables x, y, z, it is fairly easy to verify whether knowing z renders x independent of y. However, P does not dictate which variables should be regarded as direct neighbors. That decision is left to the conceptualizer who must decide which dependencies to encode in the graph and what decoding techniques to use to recover them.

In the case of probabilistic dependencies, we are fortunate to have the theory of Markov-Fields. It tells us how to construct an edge-minimum graph G such that each time we observe a vertex x separated from y by a subset S of vertices, we can be guaranteed that variables x and y are independent given the values of the variables in S. Moreover, the set of neighbors assigned by G to each x coincides exactly with the boundary of x, i.e., the smallest set of variables needed to shield x from the influence of all other variables in the system.

The theory of graphoids extends this construction to cases where the notion of independence is not given probabilistically or numerically. We now ask what *logical* conditions should constrain the relationship:

$$I(x, z, y) = \text{``knowing } z \text{ renders } x \text{ irrelevant to } y\text{''}$$

so that we can validate it by testing whether z separates x from y in some graph G.

We show that two main conditions (together with symmetry and subset closure) are sufficient:

(1)   weak closure for intersection

$$I(x, z \cup w, y) \ \& \ I(x, z \cup y, w) \Rightarrow I(x, z, y \cup w) \quad \text{and}$$

(2)   weak closure for union

$$I(x, z, y \cup w) \Rightarrow I(x, z \cup w, y)$$

Loosely speaking, (1) states that if y does not affect x when w is held constant and if, simultaneously, w does not affect x when y is held constant, then neither w nor y can affect x. (2) states that learning an irrelevant fact (w) cannot help another irrelevant fact (y) become relevant. Condition (1) is sufficient to guarantee a unique construction of an edge-minimum graph G that validates $I(x, z, y)$ by vertex separation. Condition (2) guarantees that the neighborhoods defined by the edges of G coincide with the relevance boundaries defined by I. These two conditions are chosen as the defining axioms of graphoids and are shown to account for the graphical properties of Markov-Fields.

This paper is organized as follows: In Section 2 we summarize the properties of probabilistic independencies and their graphical representations. This is done in order to exemplify and motivate the generalizations embodied in the theory of graphoids and can be skipped by readers preferring a purely axiomatic approach. Section 3 introduces an axiomatic definition of graphoids and proves their graph-representation properties. Section 4 discusses a few extensions and outlines open problems awaiting further theoretical development.

2. AN EXAMPLE: PROBABILISTIC DEPENDENCIES AND THEIR GRAPHICAL REPRESENTATION

Let $U = \{\alpha, \beta, \cdots\}$ be a finite set of discrete-valued random variables characterized by a joint probability function $P(\cdot)$, and let x, y, and z stand for any three subsets of variables in U. We say that x and y are conditionally independent given z if

$$P(x, y \mid z) = P(x \mid z) P(y \mid z) \qquad \text{when} \quad P(z) > 0 \tag{1}$$

Eq.(1) is a terse notation for the assertion that for any instantiation z_k of the variables in z and for any instantiation x_i and y_j of x and y, we have

$$P(x=x_i \text{ and } y=y_j \mid z=z_k) = P(x=x_i \mid z=z_k) P(y=y_j \mid z=z_k) \tag{2}$$

The requirement $P(z) > 0$ guarantees that all the conditional probabilities are well defined, and we shall henceforth assume that $P > 0$ for any instantiation of the variables in U. This rules out logical and functional dependencies among the variables; a case which would require special treatment.

We shall use the notation $(x \perp z \perp y)_P$ or simply $(x \perp z \perp y)$ to denote the independence of x and y given z. Thus,

$$(x \perp z \perp y)_P \quad \text{if} \quad P(x, y \mid z) = P(x \mid z) P(y \mid z) \tag{3}$$

Note that $(x \perp z \perp y)$ implies the conditional independence of all pairs of variables $\alpha \in x$ and $\beta \in y$, but the converse is not necessarily true.

The conditional independence relation $(x \perp z \perp y)$ satisfies the following properties:

$$(x \perp z \perp y) \Longleftrightarrow P(x \mid y, z) = P(x \mid z) \tag{4.a}$$

$$(x \perp z \perp y) \Longleftrightarrow P(x, z \mid y) = P(x \mid z) P(z \mid y) \tag{4.b}$$

$$(x \perp z \perp y) \Longleftrightarrow \exists f, g : P(x, y, z) = f(x,z)g(y, z) \tag{4.c}$$

$$(x \perp z \perp y) \Longleftrightarrow P(x, y, z) = P(x \mid z) P(y, z) \tag{4.d}$$

$$(x \perp z \perp y) \Longrightarrow (x \perp z, f(y) \perp y) \tag{5.a}$$

$$(x \perp z \perp y) \Longrightarrow (f(x, z) \perp z \perp y) \tag{5.b}$$

$$(x \perp y \perp z) \quad \& \quad (x, y \perp z \perp w) \Longrightarrow (x \perp y \perp w) \quad \text{(chaining)} \tag{5.c}$$

Symmetry:

$$(x \perp z \perp y) \Longleftrightarrow (y \perp z \perp x) \tag{6.a}$$

Closure for subsets:

$$(x \perp z \perp y, w) \Longrightarrow (x \perp z \perp y) \quad \& \quad (x \perp z \perp w) \tag{6.b}$$

Weak closure for intersection

$$(x \perp y, z \perp w) \quad \& \quad (x \perp y, w \perp z) \Longrightarrow (x \perp y \perp z, w) \tag{6.c}$$

Weak closure for union:

$$(x \perp y \perp z, w) \Longrightarrow (x \perp y, z \perp w) \tag{6.d}$$

Contraction:

$$(x \perp y, z \perp w) \quad \& \quad (x \perp y \perp z) \Longrightarrow (x \perp y \perp w) \tag{6.e}$$

The proof of these assertions can be derived by elementary means. The properties in (4) and (5)

are taken from Lauritzen (1982) and those in (6) were added for completeness. The properties

in (4) characterize the numeric representation of P, while the rest are purely logical, void of any

association with numerical forms. (6.d) is a derivative of (5.a) using $z \subseteq y$ for $f(y)$, and (5.c)

is a consequence of (6.d) and (6.e). A graphical interpretation for properties (5.c) to (6.e) can be

obtained by envisioning the chain $x - y - z - w$ and associating the triplet $(x \perp z \perp y)$ with the

statement "z separates x from y" or "z intervenes between x and y." The five properties in (6) are logically independent and can form an axiomatic, possibly complete, basis for a logic of independencies. However, rather than pursuing this prospect, we focus on the correspondence between $\perp$ and vertex separation in graphs.

Ideally, we would like to display independence between variables by the lack of connectivity between their corresponding nodes in some graph G. Likewise, we would like to require that if the removal of some subset S of nodes from the graph renders nodes x and y disconnected, written $<x \mid S \mid y>_G$, then this separation should correspond to conditional independence between x and y given S, namely

$$<x \mid S \mid y>_G \Longrightarrow (x \perp S \perp y)_P$$

and conversely,

$$(x \perp S \perp y)_P \Longrightarrow <x \mid S \mid y>_G$$

This would provide a clear graphical representation for the notion that x does not affect y directly, that its influence is mediated by the variables in S. Unfortunately, we shall next see that these two requirements might be incompatible; there might exist no way to display *all* the independencies embodied in P by vertex separation in a graph.

Definition

An undirected graph G is a *dependency map* (D-map) of P if there is a one-to-one correspondence between the variables in P and the nodes of G, such that for all non-intersecting subsets, x, y, S of variables we have:

9

$$(x \perp S \perp y)_P \implies <x \mid S \mid y>_G \tag{7}$$

Similarly, G is an *Independency map* (I-map) of P if:

$$(x \perp S \perp y)_P \impliedby <x \mid S \mid y>_G \tag{8}$$

A D-map guarantees that vertices found to be connected are indeed dependent; however, it may occasionally display dependent variables as separated vertices. An I-map works the opposite way: it guarantees that vertices found to be separated always correspond to genuinely independent variables but does not guarantee that all those shown to be connected are in fact dependent. Empty graphs are trivial D-maps, while complete graphs are trivial I-maps.

Lemma

There are probability distributions for which no graph can be both a D-map and an I-map.

Proof

Graph separation always satisfies $<x \mid S_1 \mid y>_G \implies <x \mid S_1 \cup S_2 \mid y>_G$ for any two subsets S_1 and S_2 of vertices. Some P's, however, may induce both $(x \perp S_1 \perp y)_P$ and NOT $(x \mid S_1 \cup S_2 \perp y)_P$. Such P's cannot have a graph representation which is both an I-map and a D-map because D-mapness forces G to display S_1 as a cutset separating x and y, while I-mapness prevents $S_1 \cup S_2$ from separating x and y. No graph can satisfy these two requirements simultaneously. Q.E.D.

A simple example illustrating the conditions of the proof involves an experiment with two coins and a bell that rings whenever the outcomes of the two coins are the same. If we ignore the bell, the coin outcomes are mutually independent, i.e., $S_1 = \emptyset$. However, if we notice the bell (S_2), then learning the outcome of one coin should change our opinion about the other coin.

Being unable to provide a graphical description for *all* independencies, we settle for the following compromise: we will consider only I-maps but will insist that the graphs in those maps capture as many of P's independencies as possible, i.e., they should contain no superfluous edges.

Definition

A graph G is a *minimal* I-map of P if no edge of G can be deleted without destroying its I-mapness.

Theorem

Every P has a (unique) minimal I-map G_0 produced by connecting *only* pairs (α, β) for which:

$$(\alpha \perp U - \alpha - \beta \perp \beta)_P \quad \text{is FALSE} \tag{9}$$

[i.e., deleting from the complete graph *all* edges (α, β) for which $(\alpha \perp U - \alpha - \beta \perp \beta)_P$].

Proof

(1) G_0 is an I-map -- The proof found in the literature is rather lengthy and involves the properties of Markov Fields (see Lauritzen, 1982). In section 3 we show that it follows directly from (6.c).

(2) G_0 is minimal and unique -- deleting any of its (α, β) edges would render α separable from β by the complementary set $U - \alpha - \beta$ and would lead us to predict (by I-mapness) $(\alpha \perp U - \alpha - \beta \perp \beta)_P$. However, this prediction would be false; otherwise (by the construction method of G_0), (α, β) would not be connected in the first place. Q.E.D.

We call $G_0(P)$ -- The *MARKOV-NET* of P.

Definition

A *Markov boundary* $B_P(\alpha)$ of variable α is a minimal subset S that renders α independent of all other variables, i.e.,

$$(\alpha \perp S \perp U - S - \alpha)_P, \ \alpha \notin S, \tag{10}$$

and simultaneously, no proper subset S' of S satisfies $(\alpha \perp S' \mid U - S' - \alpha)_P$. If no S satisfies (10), define $B_P(\alpha) = U - \alpha$.

Theorem

Each variable α has a unique Markov boundary $B_P(\alpha)$ that coincides with the set of vertices $B_{G_0}(\alpha)$ adjacent to α in the Markov net G_0.

Proof

(1) $B_P(\alpha)$ is unique because, from (6.c), the family of sets satisfying (10) is closed under intersection.

(2) $B_{G_0}(\alpha)$ satisfies the condition in (10) because G_0 is an I-map.

(3) $B_{G_0}(\alpha)$ is contained in every set S satisfying (10), because otherwise (following (6.d)) at least one of its members should not have been connected to α. Q.E.D.

The usefulness of the preceding theorem lies in the fact that in many cases it is the Markov boundaries $B_P(\alpha)$ that define the organizational structure of human memory (e.g., storing the immediate consequences and/or justifications of each action or event (Doyle, 1979)). The fact that $B_P(\alpha)$ coincides with $B_{G_0}(\alpha)$ guarantees that many independency relationships can be validated by tests for graph separation at the knowledge level itself (Pearl, 1985).

3.   GRAPHOIDS

Definition

A *graphoid* is a set I of triples (x, z, y) where x, z, y are three non-intersecting subsets of elements drawn from a finite collection $U = \{\alpha, \beta, \cdots\}$, having the following four properties. (We shall write $I(x, y, z)$ to state that the triple (x, y, z) belongs to graphoid I.)

Symmetry -- $I(x, z, y) \Longleftrightarrow I(y, z, x)$ (11.a)

Subset Closure -- $I(x, z, y \cup w) \Longrightarrow I(x, z, y) \ \& \ (x, z, w)$ (11.b)

Intersection -- $I(x, z \cup w, y) \ \& \ (x, z \cup y, w) \Longrightarrow I(x, z, y \cup w)$ (11.c)

Union -- $I(x, z, y \cup w) \Longrightarrow I(x, z \cup w, y)$ (11.d)

For technical convenience we shall adopt the convention that I contains all triples in which either x or y are empty, i.e., $I(x, z, \varnothing)$.

If U stands for the set of vertices in some graph G, and if we equate $I(x, z, y)$ with the statement: ''z separates between x and y,'' written $<x \mid z \mid y>_G$, then the conditions in (11) are clearly satisfied. However, not all properties of graph separation are required for graphoids. For example, in graphs we always have [$<\alpha \mid z \mid \beta>_G \ \& \ <\alpha \mid z \mid \gamma>_G$] iff $<\alpha \mid z \mid \beta \cup \gamma>_G$ while property (11.b) requires only the ''if'' part. Similarly, graph separation dictates:

$$<x \mid z \mid y>_G \Longrightarrow <x \mid z \cup w \mid y>_G \quad \forall w$$

while (11.d) severely restricts the conditions under which a separating set z can be enlarged by w.

Definition

A graph G is said to be an *I-map* of I if there is a one-to-one correspondence between the elements in U and the vertices of G, such that, for all non-intersecting subsets x, y, S we have:

$$<x \mid S \mid y>_G \implies I(x, S, y) \tag{12}$$

Theorem-1

Every graphoid I has a unique edge-minimum I-map G_0. $G_0 = (U, E_0)$ is constructed by connecting *only* pairs (α, β) for which the triple $(\alpha, U-\alpha-\beta, \beta)$ is not in I, i.e.,

$$(\alpha, \beta) \notin E_0 \quad \text{iff} \quad I(\alpha, U-\alpha-\beta, \beta) \tag{13}$$

Proof

(1) We first prove that G_0 is an I-map using descending induction:

  (i) Let $n = |U|$. For $|S| = n-2$ the I-mapness of G_0 is guaranteed by its method of construction (13).

  (ii) Assume the theorem holds for every S' with size $|S'| = k \leq n-2$, and let S be any set s.t. $|S| = k-1$ and $<x \mid S \mid y>_{G_0}$. We distinguish two subcases: $x \cup S \cup y = U$ and $X \cup S \cup y \neq U$.

  (iii) If $X \cup S \cup y = U$ then either $|x| \geq 2$ or $|y| \geq 2$. Assume, without loss of generality, that $|y| \geq 2$, i.e. $y = y' \cup \gamma$. From $<x \mid S \mid y>_{G_0}$ and obvious properties of vertex separation in graphs, we conclude $<x \mid S \cup \gamma \mid y>_{G_0}$ and

14

$\ll x \mid S \cup y \mid \gamma\gg_{G_0}$. The two separating sets, $S \cup \gamma$ and $S \cup y$, are at least

$|S| + 1 = k$ in size; therefore, by induction hypothesis

$$I(x, S \cup \gamma, y) \quad \& \quad I(x, S \cup y, \gamma)$$

Applying the intersection property (11.c) yields the desired result: $I(x, S, y)$.

(iv)   If $x \cup S \cup y \neq U$, then there exists at least one element δ which is not in

$x \cup S \cup y$, and for any such δ two obvious properties of graph separation hold:

$$\ll x \mid S \cup \delta \mid y \gg_{G_0} \tag{14}$$

and:

either $\ll x \mid S \cup y \mid \delta \gg_{G_0}$,   or $\ll \delta \mid S \cup x \mid y \gg$,   or both.   (15)

The separating sets above are at least $|S| + 1 = k$ in size; therefore, by induction

hypothesis:

$$I(x, S \cup \delta, y) \quad \& \quad I(x, S \cup y, \delta), \tag{16}$$

or:

$$I(x, S \cup \delta, y) \quad \& \quad I(\delta, S \cup x, y) \tag{17}$$

Applying the intersection property (11.c) of graphoids to either (16) or (17)

yields $I(x, S, y)$, which establishes the I-mapness of G_0.

(2)   Next we show that G_0 is edge-minimum and unique, i.e., that no edge can be deleted

from G_0 without destroying its I-mapness. Indeed deleting an edge $(\alpha, \beta) \in E_0$ leaves α

separated from β by the complementary set $U - \alpha - \beta$, and if the resulting graph is still an

I-map, we can conclude $I(\alpha, U - \alpha - \beta, \beta)$. However, from the method of constructing G_0

and from $(\alpha, \beta) \in E_0$ we know that $(\alpha, U - \alpha - \beta, \beta)$ is not in I. Thus, no edge can be re-

moved from G_0, and its minimality and uniqueness are established. Q.E.D.

15

Note that the union property (11.d) is not needed for the proof.

Definition

A *relevance sphere* $R_I(\alpha)$ of an element $\alpha \in U$ is any subset S of elements for which

$$I(\alpha, S, U-S-\alpha) \quad \text{and} \quad \alpha \notin S \tag{18}$$

Let $R_I^*(\alpha)$ stand for the set of all relevance spheres of α. A set is called a *relevance boundary* of α, denoted $B_I(\alpha)$, if it is in $R_I^*(\alpha)$ and if, in addition, none of its proper subsets is in $R_I^*(\alpha)$.

$B_I(\alpha)$ is to be interpreted as the smallest set that "shields" α from the influence of all other elements. Note that $R_I^*(\alpha)$ is non-empty because $I(x, z, \varnothing)$ guarantees that the set $S = U - \alpha$ satisfies (18).

Theorem 2

Every element $\alpha \in U$ in a graphoid I has a unique *relevance boundary $B_I(\alpha)$*. $B_I(\alpha)$

coincides with the set of vertices $B_{G_0}(\alpha)$ adjacent to α in the minimal graph G_0.

Proof

(i) $B_I(\alpha)$ is unique because the intersection property of graphoids (11.c) renders $R_I^*(\alpha)$

closed under intersection. Moreover, $B_I(\alpha)$ equals the intersection of all members of

$R_I^*(\alpha)$.

(ii)   Conversely, every relevance sphere $R \in R_I^*(\alpha)$ remains in $R_I^*(\alpha)$ after we add to it an

arbitrary set of elements S', not containing α. This follows from the union property of

graphoids (11.d). In particular, if there is an element β outside $\beta_I(\alpha) \cup \alpha$ then $U - \alpha - \beta$

is in $R_I^*(\alpha)$.

(iii)   From (ii) we conclude that for every element $\beta \neq \alpha$ outside $B_I(\alpha)$, we have

$I(\alpha, U - \alpha - \beta, \beta)$, meaning β could not be connected to α in G_0. Thus,

$$B_{G_0}(\alpha) \subseteq B_I(\alpha) \tag{19}$$

(iv)   To prove that $B_{G_0}(\alpha)$ actually coincides with $B_I(\alpha)$ it is sufficient to show that $B_{G_0}(\alpha)$ is

in $R_I^*(\alpha)$, but this follows from the fact that G_0, as an I-map, must satisfy (18).

Q.E.D.

Corollary 1

The set of relevance boundaries $B_I(\alpha)$ forms a *neighbor system*, i.e., a collection

$B_I^* = \{B_I(\alpha) : \alpha \in U\}$ of subsets of U such that

(i) $\alpha \notin B_I(\alpha)$, and

(ii) $\alpha \in B_I(\beta)$ iff $\beta \in B_I(\alpha)$, $\alpha, \beta \in U$

Corollary 2

The edge-minimum I-map G_0 can be constructed by connecting each α to all members of its relevance boundary $B_I(\alpha)$.

Thus we see that the major graphical properties found in probabilistic independencies are consequences of the intersection and union properties, (11.c) and (11.d), and will therefore be shared by all graphoids.

An Illustration

To illustrate the role of these properties consider a simple graphoid defined on a set of four integers $U = \{ (1, 2, 3, 4 \}$. Let I be the set of twelve triples listed below:

$I = \{ (1, 2, 3), (1, 3, 4), (2, 3, 4), (\{1, 2\}, 3, 4), (1, \{2, 3\}, 4), (2, \{1, 3\}, 4), +$ symmetrical images $\}$

It is easy to see that I satisfies (11.a)-(11.d) and thus it has a unique minimal I-map G_0, shown in Figure 1. This graph can be constructed either by deleting the edges (1, 4) and (2, 4) from the complete graph or by computing from I the relevance boundary of each element, i.e.,

$B_I(1) = \{ 2, 3 \}, B_I(2) = \{ 1, 3 \}, B_I(3) = \{ 1, 2, 4 \}, B_I(4) = \{ 3 \}.$

Suppose that I contained only the last two triples (and their symmetrical images)

$$I' = \{(1, \{ 2, 3 \}, 4), (2, \{ 1, 3 \}, 4), +$ symmetrical images $\}$$

I' is clearly not a graphoid because the absence of the triples (1, 3, 4) and (2, 3, 4) violates the intersection axiom (11.c). Indeed, if we try to construct G_0 by the usual criterion of edge deletion, the graph in Figure 1 ensues, but it is no longer an I-map of I'; it shows 3 separating 1 from 4 while (1, 3, 4) is not in I'. In fact, the only I-maps of I' are the three graphs in Figure 2, and the edge-minimum graph is clearly not unique.

18

Now consider the list

$$I'' = \{ (1, 2, 3), (1, 3, 4), (2, 3, 4), (\{1, 2\}, 3, 4), + \text{ images} \}$$

I'' satisfies the first three axioms (11.a)-(11.c) but not the union axiom (11.d). Since no triple of

the form $(\alpha, U-\alpha-\beta, \beta)$ appears in I'', the only I-map for this list is the complete graph. How-

ever, the relevance boundaries of I'' do not form a neighbor set; e.g., $B_{I''}(4) = 3$,

$B_{I''}(2) = \{1, 3, 4\}$, so $2 \notin B_{I''}(4)$ while $4 \in B_{I''}(2)$.

Note that the graphoid I does not possess the contraction property (6.e) of probabilistic

dependencies. Therefore, there is no probabilistic model capable of inducing this set of

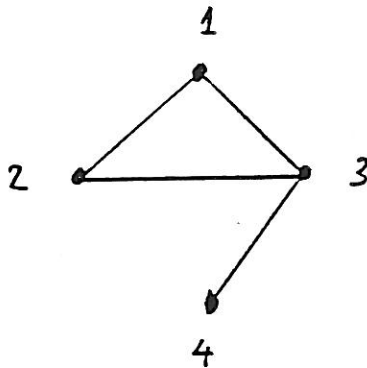relevance relationships unless we also add to I the triplet $(1, 2, 3)$.

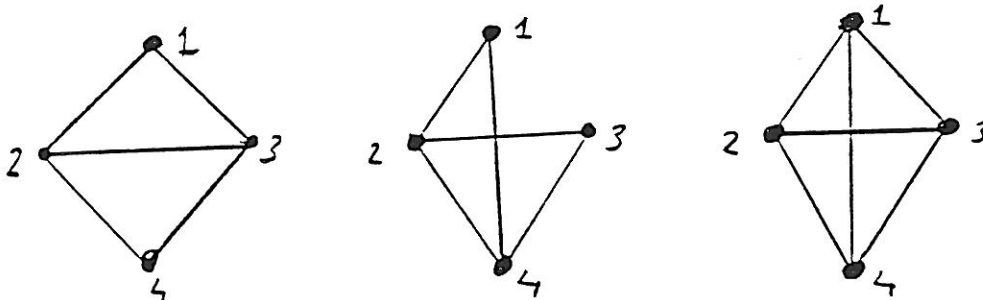FIGURE 1 — The Minimal I-map, G_0, of I.

FIGURE 2 — The Three I-maps of I'.

19

4. SPECIAL GRAPHOIDS AND OPEN PROBLEMS

In this section we study the special types of graphoid systems which have restricted domains of application. The most restricted type of graphoid is that which is isomorphic to some underlying graph, i.e., *all* triples (x, z, y) in I reflect vertex-separation conditions in an actual graph. We shall call this class of graphoids *graph-induced*.

4.1 Graph-induced Graphoids

Definition

A graphoid I is said to be *graph-induced* if there exists a graph G such that

$$I(x, z, y) <==> <x \mid z \mid y >_G \qquad (20)$$

The left-pointing implication is automatically guaranteed for all graphoids. It is the right-pointing implication that makes this class unique and restricted.

Definition

Let L be an arbitrary set of triples (x, z, y), where x, y, z are non-intersecting subsets of U. A graph G is said to be a *D-map* of L if there is a correspondence between U and the vertices of G such that

$$(x, z, y) \in L ==> <x \mid z \mid y >_G \ . \qquad (21)$$

With this terminology we can say that a graphoid I is graph-induced whenever there exists a graph which is both an I-map and a D-map of I.

Lemma-2

A sufficient condition for a graphoid to be graph-induced is that $G_0(I)$ is a D-map of I,

i.e.,

$$I(x, z, y) \Longrightarrow\ <x \mid z \mid y >_{G_0} \qquad\qquad (22)$$

where G_0 is constructed from I using the connecting criterion of (13).

Proof

The I-mapness of G_0, together with (22), implies (20).

Theorem-3

A necessary and sufficient condition for a graphoid I to be graph induced is that it satisfies the following five axioms:

$$I(x, z, y) \Longleftrightarrow I(y, z, x) \qquad\qquad \text{(symmetry)} \quad (23.a)$$

$$I(x, z, y \cup w) \Longrightarrow I(x, z, y) \ \& \ I(x, z, w) \qquad\qquad \text{(subset closure)} \quad (23.b)$$

$$I(x, z \cup w, y) \ \& \ I(x, z \cup y, w) \Longrightarrow I(x, z, y \cup w) \qquad\qquad \text{(intersection)} \quad (23.c)$$

$$I(x, z, y) \Longrightarrow I(x, z \cup w, z) \ \forall w \subset U \qquad\qquad \text{(strong union)} \quad (23.d)$$

$$I(x, z, y) \Longrightarrow I(x, z, \gamma) \ \text{ or } \ I(\gamma, z, y) \ \forall \gamma \notin x \cup z \cup y \qquad\qquad \text{(transitivity)} \quad (23.e)$$

Remarks

(23.c) and (23.d) imply the converse of (23.b), which makes I completely defined by the set of triples (x, z, y) in which x and y are individual elements of U. Equivalently, we can express the axioms in (23) in terms of such triples. Note also that the union axiom (23.d) is unconditional and therefore stronger than the one required for general graphoids (11.d). It allows us to construct G_0 by simply deleting from a complete graph every edge (α, β) for which a triple of the form (α, S, β) appears in I. These five axioms are independent and imply all the properties of probabilistic dependencies, especially chaining (5.c) and contraction (6.e).

21

Proof

1. The necessary part follows from the observation that all five properties are satisfied by vertex separation in graphs.

2. To prove sufficiency we need to show that for any set I of triples (x, z, y) satisfying (23.a)-(23.e) there exists a graph G such that (x, z, y) is in I iff z is a cutset in G that separates x from y. We show that G_0 is such a graph. In view of Lemma 2 and the remark above, it is sufficient to show that

$$I(\alpha, S, \beta) \Longrightarrow \; < \alpha \,|\, S \,|\, \beta >_{G_0} \quad \alpha, \beta \in U, S \subset U$$

This is proved by descending finite induction:

(i)   For $|S| = n-2$ the theorem holds automatically, from the way G_0 is constructed.

(ii) Assume the theorem holds for any S with size $|S| = k \le n-2$. Let S' be any set of size $|S'| = k-1$.

(iii) For $k \le n-2$, there exists an element γ outside $S' \cup \alpha \cup \beta$, and, using (23.d), we have: $I(\alpha, S', \beta) \Longrightarrow I(\alpha, S' \cup \gamma, \beta)$.

(iv) By (23.e) we have either $I(\alpha, S', \gamma)$ or $I(\gamma, S', \beta)$.

(v)   Choosing the first alternative in (iv) (the latter gives an identical result) and applying (23.d) gives $I(\alpha, S' \cup \beta, \gamma)$.

(vi) The middle arguments in (iii) and (v) are both of size k, so by induction hypothesis we have: $< \alpha \,|\, S' \cup \gamma \,|\, \beta >_{G_0}$ and $< \alpha \,|\, S' \cup \beta \,|\, \gamma >_{G_0}$.

22

(vii)   By the intersection property (23.c) for vertex-separation in graphs, these two

assertions imply $< \alpha | S' | \beta >_{G_0}$. Q.E.D.

4.2   Probabilistic Graphoids

DEFINITION

A graphoid is called *probabilistic* if there exists a probability distribution P on the variables in U such that $I(x, z, y)$ iff x is independent of y given z, i.e.,

$$I(x, z, y) <==> (x \perp z \perp y)_P \tag{24}$$

In other words, probabilistic graphoids capture the notion of conditional independence in Probability Theory. The properties of this family of graphoids were discussed in detail in Section 2, where it was shown that the graphical properties of Markov nets stem from the two graphoid axioms, Eqs. (6.c) and (6.d).

Since every probabilistic-independence relation satisfies (6.a)-(6.e) we have:

Lemma-3

A necessary condition for a graphoid to be probabilistic is that, in addition to (11), it also satisfies the contraction property (6.e), i.e.,

$$I(x, y \cup z, w) \quad \& \quad I(x, y, z) ==> I(x, y, w) \tag{25}$$

(25) can be interpreted to state that if we judge w to be irrelevant (to x) after learning some irrelevant facts z, then w must have been irrelevant before learning z. Together with the union property (11.d) it means that learning irrelevant facts should not alter the relevance status of other propositions in the system; whatever was relevant remains relevant and what was irrelevant

23

remains irrelevant.

Conjecture

The contraction property (25) is sufficient for a graphoid to be probabilistic.

Unlike the sufficiency proof for graph-induced graphoids, we found no way of construct-ing a distribution P that yields $I(x, z, y) \Longrightarrow (x \perp z \perp y)_P$ for every I that satisfies (25). We, therefore, leave this conjecture as an open question.

4.3   *Correlational Graphoids*

Let u consist of n random variables $u_1, u_2, \ldots, u_n$, and let z be a subset of u such that $|z| \leq n-2$. The *partial correlation coefficient* of u_i and u_j with respect to z, denoted $\rho_{ij \cdot z}$, measures the correlation between u_i and u_j after subtracting from them the best linear estimates using the variables in z (Cramér, 1946). In other words, $\rho_{ij \cdot z}$ measures the correlation that remains after removal of any part of the variation due to the influence of the variables in z.

Definition

Let x, y, z be three nonintersecting subsets of u. A relation $I_c(x, y, z)$ is said to be *correlation-based* if for every $u_i \in x$ and $u_j \in y$ we have:

$$I_c(x, y, z) \Longleftrightarrow \rho_{ij \cdot z} = 0 \tag{26}$$

In other words, x is considered irrelevant to y and relative to z if every variable in x is uncorre-lated with every variable in y, after removing the (linear) influence of the variables in z.

24

Theorem-4

Every correlation-based relation is a graphoid which, in addition to axioms (11), also satisfies the contraction property (25) and the converse of (11.d), i.e.,

$$I(x,y,z) \text{ and } I(x,z,w) ==> I(x,z,y \cup w) \tag{27}$$

Conjecture

Every graphoid satisfying (25) and (27) is isomorphic to some correlation-based relation.

5. CONCLUSIONS

We have shown that the essential qualities characterizing the probabilistic notion of conditional independence are captured by two logical axioms: weak closure for intersection (6.c), and weak closure for union (6.d). These two axioms enable us to construct an edge-minimum graph in which every cutset corresponds to a genuine independence condition, and these two axioms were chosen therefore as the logical basis for graphoid systems — a more general, nonprobabilistic formalism of relevance. Vertex separation in graphs, probabilistic independence and partial uncorrelatedness are special cases of graphoid systems where the two defining axioms are augmented with additional requirements.

The graphical properties associated with graphoid systems offer an effective inference mechanism for deducing, in any given state of knowledge, which propositional variables are relevant to each other. If we identify the relevance boundaries associated with each proposition in the system, and treat them as neighborhood relations defining a graph G_0, then we can correctly deduce irrelevance relationships by testing whether the set of currently known propositions constitutes a cutset in G_0.

REFERENCES

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J.

Doyle, J. (1979), "A Truth Maintenance System," *Artificial Intelligence*, vol. 12, no. 3.

Lauritzen, S.L. (1982), *Lectures on Contingency Tables*, 2nd Ed., University of Aalborg Press, Aalborg, Denmark.

Pearl, J. (1985), Fusion, Propagation and Structuring in Bayesian Networks, Technical Report CSD-850022, June 1985.
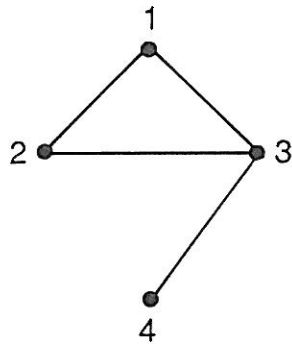
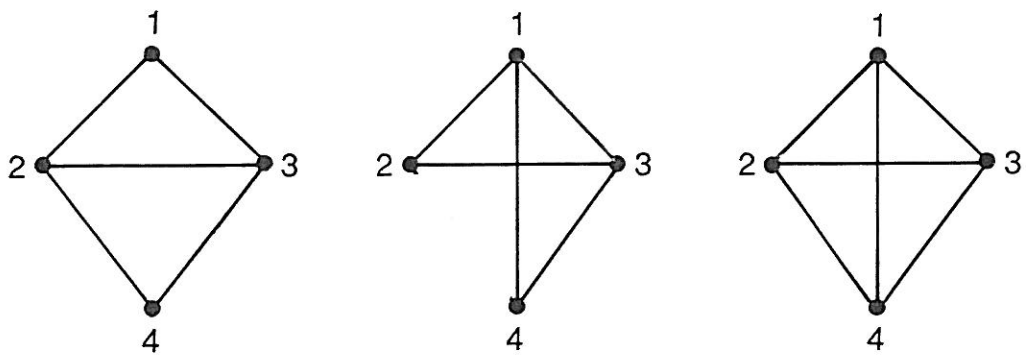Figure 1: The Minimal I-Map, G_o, of I

Figure 2: The Three I-Maps of I '