

The papers should be printed in the following order:

Jan T.A. Koster

Gibbs and Markov Properties of Graphs

Steen A. Andersson, David Madigan, Michael D. Perlman, and Christopher M. Triggs

A Graphical Characterization of Lattice Conditional Independence Models

Steffen L. Lauritzen and Finn V. Jensen

Local Computation with Valuations from a Commutative Semigroup

Milan Studený

Semigraphoids and Structures of Probabilistic Conditional Independence

František Matúš

Conditional Independence Structures Examined via Minors

## **Advances in the Understanding and Use of Conditional Independence**

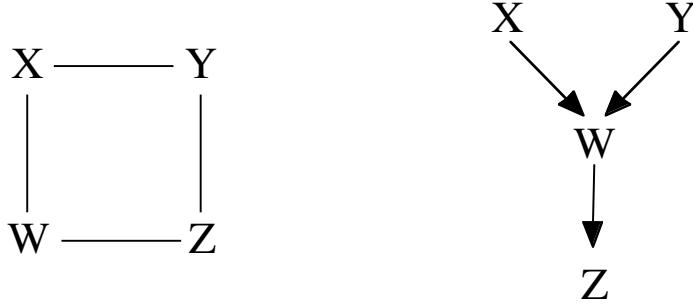
Glenn Shafer  
Rutgers University  
11 December 1996

For over a decade, conditional probabilistic independence has been a flourishing and fruitful field of mathematical research, drawing together theorists and practitioners working in artificial intelligence, statistics, and philosophy. This collection of papers, which grows out of a session at the International Workshop on Mathematics and Artificial Intelligence at Fort Lauderdale, Florida, in January, 1996, provides a good introduction to the preoccupations and achievements of this decade of research.

The current interest in conditional independence has been inspired, to a very considerable degree, by the work of Judea Pearl, summarized in his 1988 book, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. In 1988, Pearl was primarily concerned with the organization of probabilistic reasoning in expert systems, which he sought to base on assumptions of conditional independence that can be represented graphically. The graphical representation of conditional independence is an old idea in a variety of fields, including applied statistics (Wright 1934), econometrics (Simon 1954), and philosophy (Reichenbach 1956), but Pearl brought new energy and clarity to the idea and posed a wide variety of appealing new

mathematical questions. The papers in this collection show that reasonably thorough answers have now been provided to some of these questions.

Conditional independence is a relation among variables (one variable or set of variables may be independent of another given a third), and so the graphs that represent it have variables as their nodes. There are several different ways, however, to represent conditional independence relations among variables using arrows or undirected edges between nodes, and hence there are several types of “graphical models.” Two types stand out: undirected graphs (UGs) and directed graphs without directed cycles (usually called directed acyclic graphs, or DAGs, or alternatively and perhaps more justly, acyclic directed graphs, or ADGs). We interpret an UG by saying that two families of variables A and B are independent given a third family C whenever all paths between A and B go through C. We interpret a DAG by saying that each variable X is independent, given its parents (variables with arrows to X), of its nondescendants (variables to which there is no directed path from X).



An UG that encodes two independence relations:  
X and Z are independent given W and Y, and W and Y are independent given X and Z.

A DAG that encodes several independence relations, most prominently: X and Y are independent, and Z is independent of X and Y given W.

Figure 1.

Graphical models are attractive for several reasons. Most importantly, perhaps, they facilitate the construction of probability models. The conditional independence assumptions represented by the graph are equivalent to the assumption that the joint probability density for the variables in the graph can be decomposed into factors involving only neighboring variables (cliques in an UG, each variable and its parents in a DAG), and thus the graph represents the first step in the construction of a density; we first adopt the conditional independence assumptions and then we choose the individual factors. Since the conditional independence assumptions and the individual factors often have relatively clear substantive or causal interpretations, this

manner of construction facilitates explanation, whether by a statistician to her client or by an expert system to its user.

Graphical representation also serves to facilitate computation, especially updating following observation of some of the variables. In the case of expert systems, both research and application have emphasized the case where the model contains a larger number (perhaps thousands) of variables, each relatively simple (having only a few possible values), so that the graphical structure can be seen as a way of avoiding computational intractability arising from the number of the variables. This works when the graph is relatively sparse (each variable is linked to only a few other variables), so that the overall computation can be broken into many exponentially smaller “local” computations. (See, for example, Jensen 1996 or Shafer 1996a.) In the statistics literature, attention has more often focused on smaller models with continuous variables, where updating can lead to intractable numerical integrations even if there are only a handful of variables. In these problems, the graphical structure facilitates updating not because it permits exact local computation but rather because it facilitates Monte Carlo Markov Chain methods; this works when the individual factors used to construct the joint density are sufficiently well behaved to allow effective sampling from conditional distributions defined by multiplying together neighboring factors. The same conditions usually also permit computation of maximum likelihood estimators and related statistics. (See, for example, Lauritzen 1996.)

While noting the advantages of the graphical representation of conditional independence, we should note also that the representation is imperfect. On the one hand, not all sets of conditional independence relations that might be satisfied by a probability distribution can be represented by a graphical model. On the other hand, two or more graphical models often represent the same conditional independence relations. When graphical models do represent the same independence relations, we say they are “Markov equivalent.” Figure 2 gives an example.

$$X \rightarrow Y \rightarrow Z \quad X \leftarrow Y \rightarrow Z$$

$$X \leftarrow Y \leftarrow Z \quad X — Y — Z$$

These four graphs encode the same conditional independence relation:  $X$  and  $Z$  are independent given  $Y$ .

Figure 2.

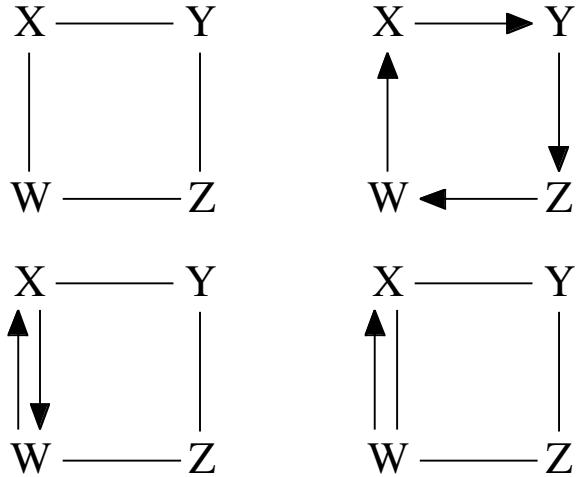
## Koster: A Yet More General Class of Graphs

One of the goals of the past decade's work has been a unified and generalized understanding of graphical models. An important step along this road was taken by Lauritzen and Wermuth (1989), who introduced chain graphs, in which a pair of variables can be linked either by an arrow or by an undirected edge. Both UGs and DAGs are special types of chain graphs. Further steps were recently taken by Spirtes (1995) and Koster (1996), who showed how DAGs (in the case of Spirtes) and more generally chain graphs (in the case of Koster) can be generalized so as to permit directed cycles.

Koster's paper in the current collection, a tour de force of quiet elegance, brings this work of generalization to an apparently definitive conclusion. This paper adds only one small insight to Koster's preceding paper on the topic (Koster 1996), but this additional insight allows Koster to provide a probabilistic interpretation for any graph that consists of a finite number of variables with arrows and undirected edges among them. (Between any two nodes X and Y in such a graph, we can have up to three links: an arrow from X to Y, an arrow from Y to X, and an undirected edge between X and Y. Any subset of these three links may be used, so that there are altogether eight ways X and Y can be related.) In the preceding literature on chain graphs, there were only four ways X and Y could be linked: (i) not at all, (ii) with an arrow from X to Y, (ii) with an arrow from Y to X, and (iv) with an undirected edge. Formally, one pretended to consider only arrows, and the undirected edge was said to be an abbreviation for a pair of arrows, one in both direction. In Koster's graphs, in contrast, we distinguish between an undirected edge and arrows in both directions.) Koster's general theory for such graphs generalizes and unifies all earlier results on the equivalence between factorability and various conditional independence conditions for UGs, DAGs, chain graphs, and several other classes of graphical models.

Koster's general theory is of more than formal interest. By bringing a wide variety of statistical models into a framework understood and used in artificial intelligence, it promotes the use of statistical tools in artificial intelligence and, more generally, increases the possibilities for communication between the two fields. Perhaps even more importantly, it should improve communication among statisticians. By bringing together graphical models that have received intensive attention from mathematical statisticians (e.g., hierarchical Bayesian models, log-linear models for contingency tables, and chain graph models) and models that have been widely used in the social sciences (e.g., the generalized path models associated with LISREL) but have sometimes been regarded with suspicion by mathematical statisticians, it opens new channels of understanding between mathematical and applied statisticians, which may lead to clearer standards for the construction and criticism of models.

Interestingly, Koster's work does not enlarge our capacity for representing conditional independence relations. Any set of conditional independence relations that can be represented by one of Koster's graphs can already be represented by a chain graph. So if we take the view that graphical models have no meaning beyond the conditional independence relations they represent, then Koster has merely increased the redundancy available to us; in addition to redundancies like those in Figure 2, we now have redundancies like those in Figure 3. Many users, however, see more than conditional independence relations in the arrows in a graphical model; they see causal relations that have conditional independence relations as only one of their manifestations. The ultimate degree of importance of Koster's work seems to depend on the extent to which we can make sense of this attribution of causal meaning.



These four graphs encode the same two conditional independence relations: X and Z are independent given Y and W, and Y and W are independent given X and Z.

Figure 3.

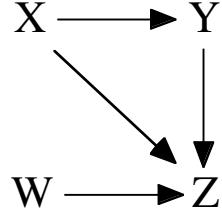
### **Andersson et al.: Lattice Conditional Independence Models and Graphical Models**

Another example of how graphical models have brought statistics and artificial intelligence together is provided by recent work on lattice conditional independence (LCI) models. The idea of an LCI model was discovered in the 1980s by Andersson and Perlman, who observed that the computational theory for many slightly nonstandard statistical problems (seemingly unrelated regressions, missing data problems in multivariate normal, multinomial, and analysis of variance models, etc.) depends principally on conditional independence relative to the lattice structure of these problems. The data structure generates a lattice of subsets of the variables in the problem, and

computation is straightforward only when any two of these subsets are independent given their intersection. As it turns out (Steffen Lauritzen was the first to see this), an LCI model imposes the same conditional independence relations as a graphical model in which the graph is a transitive DAG—a DAG in which there is an arrow from X to Y whenever there is a directed path from X to Y (see Figure 4), and Andersson, Perlman, Madigan, and Triggs have now devoted a series of articles to the elaboration of this equivalence.

$$X \rightarrow Y \rightarrow Z$$

A DAG that is not transitive: there is a directed path but not arrow from X to Z.



A transitive DAG.

Figure 4.

The article in this collection reviews earlier work, provides a completely general proof of the equivalence of LCI and transitive DAG models (the authors' previously published proof depended on the assumption that the variables involved have a positive joint density), and then considers the problem of understanding how general the class of transitive DAG models is. As the authors point out, the condition of transitivity is not quite as restrictive as one might think at first, because a DAG that is not transitive can be Markov equivalent to a DAG that is. They provide a feasible algorithm for deciding whether a given DAG is Markov equivalent to a transitive one.

Transitive DAGs are not very interesting as models for expert systems involving large numbers of variables, because transitivity is contrary to the sparseness required for efficient computation, and this problem is not alleviated by Markov equivalence. (As the authors note, two Markov equivalent DAGs have the same skeleton; only the directions of the arrows are allowed to change.) On the other hand, as Koster points out in his paper, the idea of a transitive DAG plays an important theoretical role for graphical models in general. Roughly speaking, we can form the transitive closure of any graphical model, and the factorization corresponding to this transitive closure (the outer factorization, as Koster calls it) can be thought of as the first step in the factorization associated with the graphical model. The outer factorization has a clear probabilistic interpretation; the factors are conditional probabilities for successive clusters of variables. The finer structure of the graph takes us beyond the outer factorization by giving additional interpretations to the conditional probabilities (this is the role of missing arrows, arrows that are in the transitive closure but not in the graph itself) and by providing further nonprobabilistic factorizations of the conditional

probability factors (this is the role of the undirected and cycle structure within the clusters).

### Lauritzen and Jensen: Local Computation

Just as the papers by Koster and by Andersson et al. can serve as an introduction to the literature on the construction of graphical models, the paper by Lauritzen and Jensen can serve as an introduction to the literature on computation with these models, which has evolved into a literature on the general problem of computation in join trees.

The basic question involved in computation with graphical models is directly encountered by every student of probability theory: How do we take advantage of a factorization when performing summations, integrations, or optimizations over nuisance variables? If we are summing a product involving many variables, and not all the variables appear in every factor, then we seek to sum the variables out in an order that allows us to remove as many factors as possible from each sum. Different orders will result in the clustering together in the same sum of different families of variables, and efficiency demands an order to minimizes the sizes of the largest clusters. The clusters themselves, together with the flow of the computation, defines a join tree, a tree of clusters satisfying the condition that any variable found in two clusters is also found in all the clusters on the path between the two. The computation itself can be thought of as a series of local computations—computations within the clusters—connected by messages passed along the links in the tree.

The use of a join tree to represent the structure of probability computation signals a structural similarity between probability computation and other recursive computations, for join trees have been encountered in many problems, ranging from retrieval in relational databases (where the name “join tree” was coined), to the solution of sparse systems of linear equations, to the combination of evidence using belief functions. A number of authors, including myself, have sought to deepen our understanding of the commonalities among these problems. One result is the Shafer-Shenoy axioms for join-tree computation. Another is an analysis of the differences between competing architectures for message-passing and computation in join trees, published recently in Shafer (1996a). Both of these are reviewed by Lauritzen and Jensen.

From an abstract viewpoint, join-tree computation works for probability because of the distributivity of multiplication over addition (which allows us to remove from a sum factors that do not involve the variables being summed over) and the transitivity of marginalization (which allows variables to be summed out successively rather than all at once). The Shafer-Shenoy axioms formalize these properties in terms of two abstract objects: the semigroup defined by the multiplication and the lattice of domains formed by different

subsets of variables. These axioms are sufficient to allow join-tree computation, and they are satisfied in a broad range of problems. But there remains a bewildering array of closely similar problems that do not quite satisfy the Shafer-Shenoy axioms. Moreover—and this is the issue addressed in the paper by Lauritzen and Jensen—these axioms, while they suffice to assure that join-tree computation will work in the Shafer-Shenoy architecture, do not similarly suffice for the Lauritzen-Spiegelhalter and Aalborg architectures. (The Lauritzen-Spiegelhalter architecture is the architecture used by those authors in their widely read 1988 paper, while the Aalborg architecture is the one implemented in the widely used HUGIN software.) The latter architectures gain efficiency by using division, and while it is clear that the divisions needed are always available in the probability case, it has not been so clear how to characterize abstractly the extent to which division is needed and hence not so clear to what extent these architectures will work for other computations to which we would like to apply them.

The problem, then, is to formulate conditions on a semigroup and its associated domains that are widely satisfied and suffice for the Lauritzen-Spiegelhalter and Aalborg architectures. Lauritzen and Jensen address this problem beginning with the assumption that the semigroup is separative—i.e., that it is embeddable in a semigroup that is a union of groups. This turns out to be a very appropriate assumption, first because it is easy to check, and secondly because it is satisfied by virtually all computations that have been of interest for probabilistic expert systems, including combination of belief functions, constraint satisfaction, and fast retraction. Once separateness is assumed, the availability of division can be analyzed in terms of the relation between the groups into which the semigroup separates and the domains; under natural and widely satisfied assumptions on this relation, the divisions needed are available. The authors' analysis indicates, moreover, how the desired computations can actually be implemented in these architectures.

### **Studený and Matúš : The Structure of Conditional Independence**

If the most fundamental qualitative features of probability models are expressed by conditional independencies among variables, and if the graphical representation of these independencies is imperfect, then we should explore alternative representations. Pearl made some suggestions in this direction, raising a number of questions, and the papers by Studený and Matúš give a good view of the extent to which these questions have been answered.

Since there are never more than a finite number of variables in an expert system, it is always possible, in principle, to list the conditional independencies among these variables. But more economical representations are desirable. The most obvious economy is obtained omitting independencies implied others. What do we mean by saying that one independency, say  $R$ , is implied by others, say  $R_1, \dots, R_n$ ? We mean that any probability distribution over the variables satisfying  $R_1, \dots, R_n$  also satisfies  $R$ . But we need efficient

means of checking whether this is so, and thus Pearl was led to consider rules that permit the deduction of independencies from other independencies. He began with four rules considered earlier by A. P. Dawid (1979):

- If  $(C, B|A)$ , then  $(B, C|A)$ .
- If  $(D, B \cup C|A)$ , then  $(C, B|A)$ .
- If  $(D, B \cup C|A)$ , then  $(D, C|A \cup B)$ .
- If  $(D, C|A \cup B)$  and  $(D, B|A)$ , then  $(D, B \cup C|A)$ .

(Here  $(C, B|A)$  means that the family of variables  $C$  is independent of the family  $B$  given the family  $A$ . Aside from the first rule, symmetry, these rules are easily understood if  $(C, B|A)$  is taken to mean that  $B$  is of no further help in predicting  $C$  once  $A$  is taken into account; see also §F.10 of Shafer (1996b).) These rules are clearly sound. Are they, Pearl asked, also complete? Can all the implications of a set of conditional independencies be deduced using these rules?

Studený gave a negative answer to this question at the end of the 1980s. He eventually showed, in fact, that there is no complete finite axiomatization of conditional independence. On the other hand, as other authors have shown, Dawid's rules are complete for certain restricted models: models in which all independencies have the same total set of variables, or in which all have the same conditioning variables. In the current paper, Studený establishes another respect in which the rules are complete: they account for all consequences of pairs of conditional independencies. What they do not do is allow us to deduce all consequences of three or more independencies.

Studený's analysis of conditional independence is an impressive tour de force, involving an exhaustive and sometimes exhausting examination of what he calls dominant triplets. An alternative approach, more closely allied to older work in graph theory and perhaps more powerful, is developed by Matúš in his paper in this collection. Matúš analyzes classes of models in terms of the minors they permit. When a model is a set of conditional independencies, a minor is a sort of submodel, obtained by conditioning on certain variables and omitting certain other variables. Given a model involving a set  $N$  of variables, we choose two disjoint subsets, say  $L$  and  $M$ , of  $N$ . We are only interested in variables in  $L \cup M$ , and so we omit from the model any  $(C, B|A)$  that does not satisfy  $A \cup B \cup C \subseteq L \cup M$ . Moreover, we are interested in the situation where we have observed and conditioned on values for the variables in  $M$ , and so we also omit any  $(C, B|A)$  that does not satisfy  $M \subseteq A$ , and then we simplify the  $(C, B|A)$  that remain by omitting explicit reference to  $M$ —i.e., we replace each  $(C, B|A)$  by  $(C, B|A \setminus M)$ . This definition of minor applies to any model consisting of triplets of sets; in addition to the case where the sets have variables as elements and  $(C, B|A)$  means that  $C$  and  $B$  are conditionally independent given  $A$ , Matúš also considers, for example,

the case where the sets have subspaces of a linear space as elements, and  $(C, B \setminus A)$  means that the intersection of the subspaces spanned by  $B$  and by  $C$  is contained in the subspace spanned by  $A$ .

As Matúš shows, many classes of models can be characterized most simply by their minors for a small number of variables. For example, a model can be represented by a UG if and only if all its minors for three variables have one of the two forms Matúš presents graphically in Figure 6 of his paper. Other models for three variables (there are only a finite number of these) are forbidden; they cannot appear as minors for a model representable by an undirected graph, and in fact a model can be represented by a UG if and only if it has none of these forbidden minors for three variables. Other classes of models, in contrast, have infinitely many forbidden minors. Matúš shows, for example, that this is the case for the class consisting of all conditional independence models. This provides a new proof for the theorem, established earlier and more laboriously by Studený, that there is no complete finite axiomatization for conditional independence.

## Other Current Research

While the papers in this collection provide a relatively thorough report on the most established threads of research on conditional independence, they do not represent so well areas now most in ferment and hence most important for the future. Three such areas should be mentioned in any review: generalization to decision models, learning, and causal interpretation.

Generalization to decision models has always been on the research agenda of students of conditional independence. Indeed, influence diagrams, which generalize DAGs by including decision variables as well as chance variables, have been an active topic of research in operations research and artificial intelligence since the early 1980s (see, for example, Oliver and Smith 1990). But the inclusion of decision variables creates enough additional complexity that the modelling and computational tools available for influence diagrams are still not as satisfactory as those available for purely probabilistic conditional independence graphs. It is reasonable to hope that they will improve substantially over the next few years.

The problem of learning graphical models from data can be broken into two parts—the problem of learning the graph—i.e., the conditional independence structure, and the problem of learning the probabilities once the graph is known. The problem of learning the probabilities has stimulated work of very high quality (see especially Lauritzen 1996) that is sufficiently close to classical statistical theory in aim and method that we can expect much similar work to follow. The problem of learning the graph corresponds to the statistical problem of model selection, but here classical statistical theory is too cautious and unsettled to provide the guidance needed by expert systems. We can hope that work currently under way will provide that guidance and

thereby bring new ideas into statistical theory. For references to current work on model selection, we may cite Lauritzen (1990, p. 230) and Spirtes, Glymour, and Scheines (1993).

Current work on the causal interpretation of conditional independence is also very promising, even if its philosophical nature makes assessment difficult. The conundrums of causality often confuse more than they instruct, and some might argue that the progress of the past decade was possible only because these conundrums were put in the closet. But they cannot remain in the closet. Practitioners invariably appeal to causal interpretation when constructing graphical models, and a theory can hardly justify ignoring so vital an aspect of its application. Moreover, the theoretical progress just charted has accentuated the need for a better understanding of causality. In particular, the large and increasing scope of Markov equivalence calls out for an accounting: we need a way to give mathematical and philosophical content to the differences between Markov-equivalent graphs, differences that are none the less real to practitioners for the fact that they are not expressed by conditional independence. And the very complexity of conditional independence, as displayed to us by Studený and Matúš, suggests that this concept is only a partial representation of something more fundamental.

At the center of current work on the causal interpretation of conditional independence is the philosophical view, adopted especially by Spirtes, Glymour, and Scheines (1993) and Pearl (1995), that causation is a counterfactual relation. In this opinion of these authors, the most fundamental causal graphs are functional DAGs—graphs in which each variable is a function of its parents together with an independent random error. The functional form means that the graph tells us what value a variable  $X$  would have taken had an intervention fixed the values of its parents without otherwise interfering in the causal process described by the graph. This understanding, which resonates with views articulated in econometrics in the 1940s (see Morgan 1990) and in mathematical statistics more recently (see Holland 1986), has been used by Pearl to motivate a calculus of causal effects that can help statisticians with a number of problems of causal inference, including some that had previously resisted systematic understanding. But the generality and philosophical adequacy of this approach has been sharply disputed by a number of authors, as illustrated in the discussion of Pearl (1995).

In my own recent work (Shafer 1996b and 1996c), I have advanced an alternative understanding of probability and causality, based not on the idea of intervention but rather on the idea of prediction. According to this understanding, causal probabilities are those that would be used for prediction by an ideal scientist, who knows and observes everything that could conceivably be known or observed by any agent—human or artificial. Events, including interventions, have a causal effect when they change the

probabilities of this ideal scientist. This ideal scientist may have probabilities for all interesting events or (as in a decision problem) only for some. The mathematical theory that emerges from this understanding allows us to show how independence and (more often) uncorrelatedness can result from causal relations. And it gives mathematical content to the differences between causal graphs that yield the same conditional independence relations. In fact, it provides several alternative causal interpretations for any given causal graph, thus providing modellers with the means to express causal claims more precisely. More importantly, it provides causal concepts that are useful even when the causal regularity suggested by graphs is unavailable.

## Conclusion

These papers, and the references therein, provide a nearly comprehensive introduction to conditional independence and probabilistic expert systems. Readers who want a yet more comprehensive view may wish to consult the references in Lauritzen (1996) and Shafer (1996a). Electronic resources, including information on workshops and conferences on probabilistic expert systems, is available on the World Wide Web at <http://www.auai.org/>.

## References

- Almond, Russell (1995). *Graphical Belief Modelling*. London: Chapman and Hall.
- Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* **41** 1-31.
- Holland, Paul W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association* **81** 945-970.
- Jensen, Finn (1996). *An Introduction to Bayesian Networks*. London: University College Press.
- Koster, Jan T.A. (1996). Markov properties of non-recursive causal models. *Annals of Statistics*.
- Lauritzen, Steffen L. (1996). *Graphical Models*. Oxford: Oxford University Press.
- Lauritzen, Steffen L., and David J. Spiegelhalter (1988). Local computation with probabilities on graphical structures and their applications to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B* **50** 157-224.
- Lauritzen, Steffen L., and Nanny Wermuth (1989). Graphical models for associations between variables, some of which are qualitative and some of which are quantitative. *Annals of Statistics* **17** 31-57.
- Morgan, Mary S. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
- Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann.

- Pearl, Judea (1995). Causal diagrams for empirical research (with discussion). *Biometrika* **82** 669-710.
- Reichenbach, Hans (1956). *The Direction of Time*. Berkeley, California: University of California Press.
- Shafer, Glenn (1996a). *Probabilistic Expert Systems*. Philadelphia: SIAM.
- Shafer, Glenn (1996b). *The Art of Causal Conjecture*. Cambridge, Massachusetts: The MIT Press.
- Shafer, Glenn (1996c). The situation of causality. *Foundations of Science* **1**.
- Simon, Herbert A. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association* **49** 469-492.
- Spirites, Peter (1995). Directed cyclic graphical representation of feedback models. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (P. Besnard and S. Hanks, eds.). San Mateo, California: Morgan Kaufmann.
- Spirites, Peter, Clark Glymour, and Richard Scheines (1993). *Causation, Prediction, and Search*. New York: Springer-Verlag.
- Wright, Sewall (1934). The method of path coefficients. *Annals of Mathematical Statistics* **5** 161-215.