Pergamon

Computers Math. Applic. Vol. 34, No. 11, pp. 89–97, 1997

0898-1221/97 $17.00 + 0.00

PII: S0898-1221(97)00222-8

Chordal Graph Models of Contingency Tables

H. J. KHAMIS
Statistical Consulting Center
Wright State University, Dayton, OH 45435, U.S.A.

T. A. MCKEE*
Department of Mathematics & Statistics
Wright State University, Dayton, OH 45435, U.S.A.

Abstract—Chordal graph theory has recently found application by statisticians in the analysis of contingency tables. Specifically, what are called "decomposable loglinear models" correspond exactly to chordal graphs. We survey these results, translating the statistical application into conventional graph theory.

Keywords—Chordal graphs, Interaction graphs, Probability models, Graphical models, Decomposable models.

1. INTRODUCTION AND PRELIMINARIES

Beginning with the 1980 paper [1] by Darroch, Lauritzen and Speed, chordal graphs have emerged as an important type of model for the statistical analysis of contingency tables. We provide a "map of the terrain" for the graph theorist who is interested in exploring the burgeoning literature on these chordal graph models (a literature that seems impenetrable for nonstatisticians). We present a limited survey of this literature, concentrating on reformulating the statistical applications into traditional graph theoretical terminology.

A recent paper by Lauritzen and Spiegelhalter [2] promises to be equally stimulating in terms of using related ideas to study the propagation of probabilistic evidence in expert systems; see also, Pearl's text [3, Chapter 3]. In a different direction, recent work by the present authors [4] shows how multigraphs offer a separate approach to these same types of applications—an approach that has certain graph-theoretical advantages.

1.1. Contingency Tables

Suppose $\mathcal{X} = \{V_1, \dots, V_d\}$ is a set of variables that represent classification criteria (perhaps political affiliation, race, occupation, gender, etc.) that take values from sets $I_1, \dots, I_d$ of discrete values. We use lower case letters $v_1, \dots, v_d$, respectively, for specific values from $I_1, \dots, I_d$ of the upper case variables $V_1, \dots, V_d$.

A *d-dimensional contingency table* is obtained from a set $\mathcal{X}$ and a random sample of objects from some population. The table is then formed from the Cartesian product $I_1 \times \dots \times I_d$ of sets of values, with each particular choice $v_1, \dots, v_d$ of values called a *cell* of the table. To each cell is assigned a *cell frequency*, recording the number of objects from the particular random sample represented by the table for which each variable V_i takes the corresponding value v_i. Each cell

*Research partially supported by Office of Naval Research Grant N00014-91-J-1210.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX also corresponds to a *cell probability*, the probability that an object chosen randomly from the entire population falls into that cell. If $S \subset \mathcal{X}$, then the Cartesian product of the sets I_i for which $V_i \in S$ is called a *marginal table*. A cell frequency in the marginal table is the sum of the corresponding cell frequencies from the full table over all values of the sets I_j with $V_j \notin S$.

As an example, Table 1 shows a three-dimensional contingency table based on data reported by Bartlett [5] in his pioneering 1935 article. These data are from an experiment giving the Response (alive or dead) of 240 plants for each combination of two variables: Time of Planting (early or late) and Length of Cutting (high or low).

Table 1.

Time of Planting		Early		Late	
Length of cutting		high	low	high	low
Response:	alive	156	107	84	31
	dead	84	133	156	209

Table 2.

Length of Cutting		High	Low
Response:	alive	240	138
	dead	240	342

The marginal table shown as Table 2 expresses what is called the "contingency" between Length of Cutting and the Response. It is obtained by summing (or *"collapsing"*) over the Time of Planting variable.

The primary objective in the analysis of contingency tables is to determine the structure of associations among the variables in $\mathcal{X}$. The most popular technique for doing this is called "loglinear model analysis", as defined by Birch [6]. For our purposes, we need only note that loglinear models involve a family $\mathcal{P}$ of subsets of $\mathcal{X}$, with the members of $\mathcal{P}$ corresponding to what are called the *permissible interactions* of the model. In particular, singleton subsets in $\mathcal{P}$ are said to correspond to *main effects*, doubleton subsets in $\mathcal{P}$ to *first-order interactions*, etc.

1.2. Hierarchical Models

Our attention will be confined to the most important models in practical applications called "hierarchical models". *Hierarchical models* are loglinear models such that $S \subset S' \in \mathcal{P}$ implies $S \in \mathcal{P}$. Each hierarchical model can be characterized by a *generating class* $\mathcal{C}$ consisting of the family of inclusion-maximal sets in $\mathcal{P}$. Thus, $S \in \mathcal{P}$ if and only if there exists $S' \in \mathcal{C}$ such that $S \subseteq S'$. For simplicity, we assume, as in [1], that $\mathcal{X} = \bigcup\{C : C \in \mathcal{C}\}$; i.e., that all main effects are in $\mathcal{P}$.

In the example in Table 1, let V_1, V_2, and V_3 represent the Time of Planting, Length of Cutting, and Response variables, respectively. There are nine possible hierarchical models that contain all three main effects; they have generating classes $\{\{V_1\},\{V_2\},\{V_3\}\}$, three of the form $\{\{V_i,V_j\},\{V_k\}\}$, three of the form $\{\{V_i,V_j\},\{V_j,V_k\}\}$, and $\{\{V_1,V_2\},\{V_1,V_3\},\{V_2,V_3\}\}$ and $\{\{V_1,V_2,V_3\}\}$. Observe that the generating class $\mathcal{C} = \{\{V_1,V_3\},\{V_2\}\}$ represents the hierarchical model with $\mathcal{P} = \{\emptyset,\{V_1\},\{V_2\},\{V_3\},\{V_1,V_3\}\}$; i.e., there are three main effects, one first-order interaction, and no second-order interaction among the variables.

A result from loglinear model theory (see [1]) asserts that for any hierarchical model with generating class $\mathcal{C}$, if you knew the cell probabilities for the marginal tables with indices given by the members of $\mathcal{C}$ then you would know the cell probabilities for the full table; in short, the cell probabilities are *uniquely determined* by $\mathcal{C}$. However, there is a sense in which it is not possible in general to determine the cell probabilities from the marginal probabilities corresponding to $\mathcal{C}$ in a "nice" manner (see [7, Section 5.2.5; 8, Proposition 6 (iv)]). This is where chordal graphs will come in later; our Theorem 2 (taken from [1]) will show how this "nice" determination can be done (and will illustrate the technical meaning of "nice").

1.3. Interaction Graphs

Define the *interaction graph* of a hierarchical model to be the undirected graph having the vertex set $\mathcal{X} = \{V_1, \ldots, V_d\}$ with edges corresponding to the first-order interactions in the model. For instance, the interaction graph of the model with $\mathcal{X} = \{V_1, V_2, V_3\}$ and generating class $\mathcal{C} = \{\{V_3, V_4\}, \{V_1, V_3, V_5\}, \{V_2, V_4, V_5\}\}$ is shown in Figure 1. One natural use of interaction graphs is to determine the "conditional independencies" among the variables in the contingency table.

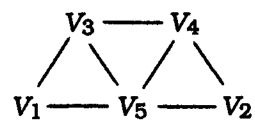

Figure 1.

With respect to choosing an object randomly from the entire population, let $\text{Prob}(v)$ abbreviate $\text{Prob}(V = v)$, $\text{Prob}(\{v_i : V_i \in S\})$ abbreviate $\text{Prob}(\{V_i = v_i : V_i \in S\})$, and $\text{Prob}(v \mid \{v_i : V_i \in S\})$ be the conditional probability that $V = v$, given that $V_i = v_i$ for each $V_i \in S$. The relationship between adjacency in the graph and conditional independence is developed in [1] using so-called "Markov random fields". The result is that, if G is an interaction graph of a hierarchical model (and so, remember, the vertices of G are precisely the variables of $\mathcal{X}$), then variables V and W are *non*adjacent in G if and only if they are *conditionally independent* relative to all the other variables; in symbols, if and only if

$$\text{Prob}\,(v \mid \{v_1, \ldots, v_d\} \setminus \{v\}) = \text{Prob}\,(v \mid \{v_1, \ldots, v_d\} \setminus \{v, w\}).$$

Informally, this means that knowing the value of W provides no additional information about the probability distribution of V when all the other values are known. Such a conditional independence is called a *zero partial association* (or *ZPA*) in [8], and when variables V and W *are* adjacent in G they are said to be *partially associated* relative to $\mathcal{X} \setminus \{V, W\}$. If a set S of variables separates V from W in the graph, then V and W are conditionally independent relative to S. Unconditionally independent variables lie in different connected components of G.

1.4. Collapsibility Conditions

One of the important uses of an interaction graph is in characterizing "collapsibility conditions" for hierarchical loglinear models of multidimensional contingency tables. Some models have the property that the structural relationships among a set of classifying factors (variables), as determined by the hierarchical model, are unchanged upon *collapsing* (i.e., summing) over the remaining variables. Collapsibility is, therefore, useful for data reduction and for simplifying data analysis and model interpretation. (The content of this subsection will not be required in the later sections.)

Formally, let $\mathcal{C}$ be the generating class for a hierarchical model of a contingency table with $\mathcal{X} = \{V_1, \ldots, V_d\}$. Let $S \subset \mathcal{X}$ and consider the marginal table obtained by collapsing over $\mathcal{X} \setminus S$. The restriction $\mathcal{C}_S$ of $\mathcal{C}$ to the marginal table is defined by deleting all occurrences of variables in $\mathcal{X} \setminus S$ from $\mathcal{C}$ and then removing members that are contained in other members. For example, if $\mathcal{C} = \{\{V_1, V_2\}, \{V_2, V_3, V_4\}, \{V_1, V_4\}\}$ and $S = \{V_1, V_2, V_3\}$, then $\mathcal{C}_S = \{\{V_1, V_2\}, \{V_2, V_3\}\}$. We say that the hierarchical model having generating class $\mathcal{C}$ is *p-collapsible* onto S if the marginal probabilities $\text{Prob}(\{v_i : V_i \in S\})$ coincide with the probabilities determined by $\mathcal{C}_S$.

It is important to note that the p-collapsibility discussed here is defined in terms of probabilities and their invariance under collapsing. This is in contrast to a well-known concept of collapsibility defined by Bishop, Fienberg and Holland [9] and by Whittemore [10] in terms of invariance of loglinear model parameters (typically λ's); we would call this *λ-collapsibility*. The concept of p-collapsibility is particularly natural in the context of our present survey.

Assmussen and Edwards [11] provide necessary and sufficient conditions for p-collapsibility in terms of the interaction graph, including Theorem 1 below. Define the *boundary* of $S' \subset \mathcal{X}$ to be the set of variables (vertices) not in S' that are adjacent to some variable in S'.

THEOREM 1. *(See [11].) A hierarchical model having generating class $\mathcal{C}$ is p-collapsible onto S if and only if the boundary of each connected component of the subgraph induced by $\mathcal{X}\backslash S$ is contained in a member of $\mathcal{C}$ (i.e., in a generator of the model).*

The simplest example of non-p-collapsibility has $S = \{V_1, V_3\}$ and $\mathcal{C} = \{\{V_1, V_2\}, \{V_2, V_3\}\}$. The boundary of $\mathcal{X}\backslash S = \{V_2\}$ is $\{V_1, V_3\}$ and, since this is not contained in a generator, the hierarchical model with generating class $\mathcal{C}$ is not p-collapsible onto S. On the other hand, consider the five-dimensional model with $S = \{V_3, V_4, V_5\}$ and $\mathcal{C} = \{\{V_3, V_4\}, \{V_1, V_3, V_5\}, \{V_2, V_4, V_5\}\}$. The interaction graph is shown in Figure 1. Then $\mathcal{X}\backslash S = \{V_1, V_2\}$ and the connected components $\{V_1\}$ and $\{V_2\}$ of the induced subgraph $\langle V_1, V_2\rangle$ have boundaries $\{V_3, V_5\}$ and $\{V_4, V_5\}$, respectively. Since each of these is contained in a member of $\mathcal{C}$, the corresponding model is p-collapsible onto S.

1.5. Graphical Models

A hierarchical model is called a *graphical model* (a notion introduced by Darroch, Lauritzen and Speed [1]) whenever the inclusion-maximal cliques (which we call *maxcliques*) of its interaction graph correspond exactly to the members of the generating class of the hierarchical model. There are statistical techniques (see, for instance, [12, Section 5]) for selecting appropriate graphical models for a given observed contingency table.

The simplest example of a hierarchical model that is not graphical has $\mathcal{X} = \{V_1, V_2, V_3\}$ and generating class $\mathcal{C} = \{\{V_1, V_2\}, \{V_1, V_3\}, \{V_2, V_3\}\}$. The interaction graph would be a triangle, but it is not a graphical model since the triangle (maxclique) $\{V_1, V_2, V_3\}$ is not in $\mathcal{C}$. Likewise, the five-dimensional model with generating class $\{\{V_3, V_4\}, \{V_1, V_3, V_5\}, \{V_2, V_4, V_5\}\}$ with the interaction graph shown in Figure 1 is not graphical. But taking $\mathcal{C} = \{\{V_1, V_3, V_5\}, \{V_3, V_4, V_5\}, \{V_2, V_4, V_5\}\}$ gives a graphical model having that same graph.

(We mention in passing that, just as graphical models are special kinds of interaction graphs, hierarchical models can be interpreted as "interaction hypergraphs". The (hyper)graph-theoretical foundations of [1] appear in [13], along with connections to game theory, a measure-theoretic problem, and Markov fields.)

Because graphical models are hierarchical, the maxcliques of the graph (since they correspond to elements of the generating class) identify marginal tables that contain all the necessary information needed to estimate the cell probabilities for the full table; see the discussion following Theorem 2. This can result in valuable data reduction, as illustrated in [12, p. 563]. In fact, the maxcliques are directly related to what are called "minimal sufficient statistics" in the "maximum likelihood estimation" of the cell probabilities as in [6].

An important advantage of graphical models is that, unlike most hierarchical models, the structural associations among their variables (including all second- and higher-order interactions) are immediate from the graph; in particular, graphical models can be interpreted exclusively in terms of conditional independence and this can be read directly from the graph as in Section 1.3. (In the context of Section 1.4, Theorem 1 implies that a graphical model is p-collapsible onto a vertex set S if and only if the boundary of each connected component of $\langle \mathcal{X}\backslash S\rangle$ is complete.)

Section 2 will discuss the central topic of this paper: a particularly useful sort of graphical model called a "decomposable model", and its equivalence to the interaction graph being chordal.

2. DECOMPOSABLE MODELS AND CHORDAL GRAPHS

Decomposable models were introduced as a special sort of hierarchical model by Goodman [14] under the name "models of Markov-type", and they were further developed by Haberman [15] and Andersen [16]. In [1], *decomposable models* are defined as graphical models that have *chordal graphs* (i.e., graphs in which every cycle of length four or more has a chord). That these notions of "decomposable" are equivalent is the content of [13, Theorem 2] (wherein *decomposable graphs* are defined as yet another name for chordal graphs, which are also frequently called *triangulated* or *rigid-circuit graphs*).

Much of the early literature of decomposable graphs involved rediscovering versions of the various characterizations of chordal graphs as are surveyed in [17, Chapter 4]. These new versions are always a bit different from the standard graph-theoretic formulations (e.g., they typically emphasize the maxcliques determined by simplicial vertices instead of the simplicial vertices themselves).

For example, Dirac's 1961 vertex separator characterization [17, Theorem 4.1], corresponds to the definition of decomposable on [1, p. 524]. (The word "decomposable" needs to be inserted in front of "generating" in the sixth line from the bottom of [1, p. 524].) Similarly, the "perfect vertex elimination scheme" characterization, see [17, Theorem 4.1], due to Fulkerson and Gross in 1965 and Rose in 1970, corresponds to what is called the "Markov interpretation" on [1, p. 528] and is essentially the approach used by Goodman and Haberman. The characterization of chordal graphs as intersection graphs of subtrees of a tree, see [17, Theorem 4.8], occurs in more recent work—see "junction trees" in [18] and "join trees" in [3]—but primarily as a data structure. One seemingly new characterization of chordal is also given, which we discuss as Theorem 3 below. If the dimension of the contingency table were to be very large, the existence of efficient algorithms (see [17, Algorithm 4.3]) for finding all the maxcliques of a chordal graph, and thereby the generating class of a decomposable model, would be important.

One major advantage of using decomposable models (i.e., chordal graphs) is that it becomes possible to "factor" a cell probability into probabilities from the marginal tables corresponding to the generating class C. We make this precise in Theorem 2 below, which we express in terms of graph neighborhoods and elimination schemes. The smallest *non*decomposable model is the graphical model having generating class $C = \{\{V_1, V_2\}, \{V_2, V_3\}, \{V_3, V_4\}, \{V_4, V_1\}\}$; its interaction graph is the cycle C_4, a nonchordal graph. Its cell probabilities cannot be expressed by an explicit multiplicative formula in terms of the two-dimensional marginal probabilities with indices corresponding to elements in C.

Recall that a vertex V is *simplicial* in a graph G if it is in a unique maxclique, or equivalently if its *open neighborhood* $N(V)$ (i.e., the set of all neighbors of V) induces a complete subgraph. An *elimination scheme* is an ordering $(V_1, \dots, V_d)$ of the vertices of G such that each V_i is simplicial in the induced subgraph $\langle V_i, \dots, V_d \rangle$. Let $N_i(V_i)$ denote the open neighborhood of V_i in $\langle V_i, \dots, V_d \rangle$, and $N_i[V_i]$ be the corresponding *closed neighborhood* $N_i(V_i) \cup \{V_i\}$. The defining property of an elimination scheme becomes that each $N_i(V_i)$ is complete in $\langle V_i, \dots, V_d \rangle$. The Perfect Vertex Elimination Scheme Theorem [17, Theorem 4.1] asserts that *a graph is chordal if and only if it has an elimination scheme.*

Theorem 2 below translates the top part of [1, p. 529; 3, Theorem 8 of Chapter 3] and simplifies the intricate method on [19, p. 342]. (In [1], b_t should be replaced by c_t in the displayed fraction.) The given formula determines the probability Prob $(\{v_1, \dots, v_d\})$ of an object being in any cell $(v_1, \dots, v_d)$ in terms of the probabilities for the marginal tables corresponding to $N_i(V_i)$ and $N_i[V_i]$.

THEOREM 2. *(See [1].) A graphical model G is decomposable (i.e., chordal), if and only if the following gives an explicit formula for the cell probabilities for the full table, where $(V_1, \dots, V_d)$*

is an arbitrary elimination scheme:

$$\text{Prob}\ (\{v_1,\dots,v_d\}) = \frac{\prod_{i=1}^{d} \text{Prob}\ (\{v : V \in N_i[V_i]\})}{\prod_{i=1}^{d-1} \text{Prob}\ (\{v : V \in N_i(V_i)\})}.$$

In fact, the parameter d in the quotient in Theorem 2 can be replaced by $\ell + 1$, where ℓ is the minimum number such that $N_{\ell+1}[V_{\ell+1}] = \langle V_{\ell+1},\dots,V_d\rangle$ (i.e., such that $\langle V_{\ell+1},\dots,V_d\rangle$ first becomes complete).

For instance, recall the five-dimensional contingency table having hierarchical model $\mathcal{C} = \{\{V_1,V_3,V_5\},\{V_3,V_4,V_5\},\{V_2,V_4,V_5\}\}$ and the interaction graph shown in Figure 1. (See [12, Section 6] for an actual five-dimensional contingency table with a loglinear model having this graph.) The graph is chordal and V_1,V_2,V_3,V_4,V_5 is an elimination scheme. By Theorem 2 (with $\ell = 2$ since $\langle V_3,V_4,V_5\rangle$ is a maxclique), we get

$$\text{Prob}\ (\{v_1,\dots,v_5\}) = \frac{\text{Prob}\ (\{v_1,v_3,v_5\})\cdot\text{Prob}\ (\{v_2,v_4,v_5\})\cdot\text{Prob}\ (\{v_3,v_4,v_5\})}{\text{Prob}\ (\{v_3,v_5\})\cdot\text{Prob}\ (\{v_4,v_5\})}.$$

(Taking factors past $\ell = 2$, would merely put identical factors into both numerator and denominator.) Denoting a cell probability in a marginal table by a subscripted $\wp$, this can be written

$$\wp_{fghij} = \frac{\wp_{f+h+j}\,\wp_{+g+ij}\,\wp_{++hij}}{\wp_{++h+j}\,\wp_{+++ij}},$$

where $f \in I_1, g \in I_2,\dots,j \in I_5$ and each "+" denotes a dimension collapsed over in the full table to form the marginal table.

A related advantage of using decomposable models (i.e., chordal graphs) is that we can obtain closed-form expressions for what are called "maximum likelihood estimators" of the cell probabilities. (In general, iterative techniques are needed to obtain such estimators.) Goodman [14,19] provides an involved method for obtaining the explicit maximum likelihood estimators for decomposable models; see also, [15]. Corresponding to the estimation formula in [1], the maximum likelihood estimator $\hat{\wp}_{fghij}$ of $\wp_{fghij}$ is

$$\hat{\wp}_{fghij} = \frac{\hat{\wp}_{f+h+j}\,\hat{\wp}_{+g+ij}\,\hat{\wp}_{++hij}}{\hat{\wp}_{++h+j}\,\hat{\wp}_{+++ij}} = \frac{n_{f+h+j}\,n_{+g+ij}\,n_{++hij}}{n_{++h+j}\,n_{+++ij}\,n_{+++++}},$$

where the subscripted n's represent marginal cell frequencies. This follows from each maximum likelihood estimator such as $\hat{\wp}_{f+h+j}$ (of a cell of a marginal table corresponding to a maxclique) being equal to the marginal cell frequency n_{f+g+j} divided by the total number n_{+++++} of objects in the table; see [6].

Section 5 of [1] and Theorem 2 of [13] give what appears to be a new characterization of chordal graphs in terms of a function that associates to each complete subgraph C of G an "index" $\nu(C) \le 1$ as follows. (For simplicity, we suppose that G is connected.) If C is a maxclique, put $\nu(C) = 1$. If C is an *articulation clique* (i.e., C is a minimal complete subgraph such that $G - C$ becomes disconnected), put $\nu(C) = 1 - \delta$, where δ is the number of components H of $G - C$ such that C is *not* a maxclique of the subgraph induced by $H \cup C$. If C is any other complete subgraph, put $\nu(C) = 0$.

THEOREM 3. *(See [13].) A graph G is chordal if and only if $\sum \nu(C) = 1$, where the sum is taken over all complete subgraphs C of G.*

(In [13], it is also shown that this sum is at least 1 for all graphs.) Theorem 3 corresponds to the fact that, along an elimination scheme, each maxclique arises one time as a $N_i[V_i]$ and each articulation clique C arises $-\nu(C)$ times as a $N_i(V_i)$. The values $\nu(C)$ appear as exponents in the maximum likelihood estimator of Prob $(v_1,\dots,v_d)$ as on [1, p. 531]; the specific values of $\nu(C)$ put appropriate factors into the numerator and denominator (as in Theorem 2) the proper number of times.

As an example, consider the five-dimensional example discussed above. If C is one of the three maxcliques, then $\nu(C) = 1$; if C is one of the two articulation cliques $\{V_3, V_5\}$ and $\{V_4, V_5\}$, then $\nu(C) = 1 - \delta = 1 - 2 = -1$; if C is any other complete subgraph, then $\nu(C) = 0$. Thus, $\sum \nu(C) = 1$.

3. CAUSAL INTERPRETATIONS AND DIGRAPH MODELS

Section 2 shows how decomposable models admit an explicit factorization of the joint distribution and of the maximum likelihood estimators. Although this is of theoretical value, the practical value is eclipsed by the widespread use of iterative algorithms such as the Newton-Raphson method (see [20]). An important *practical* advantage of decomposable models is that they are the hierarchical models that are both graphical and "recursive" (defined below); roughly, the edges of the corresponding chordal graph can be oriented so as to correspond to a causal interpretation.

Recall that in a graphical model, no variable is considered as being "*posterior*" to another and the edges are undirected. In many applications, however, it is desirable to consider how certain variables are influenced by other variables, resulting in a directed graph model called a "recursive model", which was introduced in [8]. Recursive models are a special kind of so-called "path analysis model" introduced by Goodman [21], and each of these can be characterized by a nontrivial factorization of the joint distribution in terms of the "response variables". (Section 3.3 of [3] discusses general causal digraph models with a result analogous to Theorem 4 below, and recursive models are introduced in an exercise on pp. 137,138; see also, [11, p 576].)

While graphical models are based on conditional independence restrictions for variable pairs with respect to all other variables, recursive models are defined by conditional independence restrictions for variable pairs involving a variable with respect to those variables that might influence it. Specifically, a *recursive model* has vertex set $\{V_1, \dots, V_d\}$ and a specified value $k < d$, where the vertices are ordered such that each V_i having $1 \leq i \leq k$ is a *response variable* with respect to only variables among $V_{i+1}, \dots, V_d$. For $i < j \leq d$ and $i \leq k$, there is a one-headed arrow $V_i \leftarrow V_j$ whenever V_i and V_j are *not* conditionally independent relative to $\{V_{i+1}, \dots, V_d\} \backslash \{V_j\}$. (This conditional independence restriction is called a *zero partial dependency*, or *ZPD*, in [8].) The arrow $V_i \leftarrow V_j$ corresponds to V_i being *partially dependent* on V_j. Finally, there is a two-headed arrow $V_i \leftrightarrow V_j$ whenever $i, j \geq k + 1$.

Wermuth [22] proved that a graphical model is decomposable if and only if its vertices can be ordered as follows: if $h < k < j$ and if V_h is adjacent to both V_i and V_j, then V_i is adjacent to V_j. (Such an ordering is called a *reducible zero-pattern* in [8]. The definition of "reducible" on [8, p. 539] is the contrapositive of the above statement; see also parts (i) and (ii) of their Proposition 6.) This equivalence corresponds to yet another formulation of the perfect vertex elimination scheme theorem. Proposition 5 of [8] then interrelates the recursive digraph models (and ZPDs) with reducible zero-patterns in graphical models (and ZPAs). This is stated as Theorem 4 below, showing that decomposable models are the graphical models that can be oriented so as to have a causal interpretation. This is sometimes colloquially expressed as: *the decomposable models are precisely those that are both graphical and recursive.* (Recall the introduction of the number $\ell < d$ immediately following the statement of Theorem 2.)

THEOREM 4. *(See [8].) Given a decomposable model (i.e., a chordal graph G), each choice of elimination scheme for G and each choice of k such that $\ell \leq k < d$ determine a recursive model with underlying graph G.*

Namely, given an elimination scheme $(V_1, \dots, V_d)$ and a value k such that $\ell \leq k < d$, consider the first k variables as response variables. Direct an edge $V_i \leftarrow V_j$, whenever $i \leq k$ and $i < j$. Direct an edge $V_i \leftrightarrow V_j$, whenever $i, j \geq k+1$ (noting that $\langle V_{k+1}, \dots, V_d \rangle$ is complete since $\ell \leq k$). Interpret the one-headed arrows as partial dependencies. Note that the requirement that $\ell \leq k$ is consistent with the requirement in [8] that the ZPDs in a recursive model involve variable pairs (i, j) with $i < j$ and $i \leq k$ (i.e., that only one-headed arrows may be removed from the complete graph of a recursive system). This also shows that there must be at least ℓ response variables, providing a practical interpretation for ℓ.

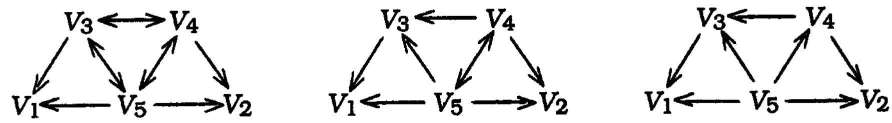

Figure 2.

For instance, the five-dimensional model discussed following Theorem 2 having the graph and elimination scheme shown in Figure 1 has $\ell = 2$ and could be given the three orientations shown in Figure 2 (taking $k = 2, 3, 4$, respectively). By associating incoming one-headed arrows with conditional probabilities and the clique of two-headed arrows with a joint probability, we can again derive an explicit formula for the cell probabilities for the full table (see [8, Formula 5.1]):

$$\operatorname{Prob}\left(\{v_1, \dots, v_d\}\right) = \left(\prod_{i=1}^{k} \operatorname{Prob}\left(v_i \mid \{v : V \in N_i(V_i)\}\right)\right) \cdot \operatorname{Prob}\left(\{v_{k+1}, \dots, v_d\}\right).$$

Whenever $\ell \leq k < d$, the formula in Theorem 2 follows from the above by the definition of conditional probability: $\operatorname{Prob}(v_i \mid \{v : V \in N_i(V_i)\})$ equals $\operatorname{Prob}(\{v : V \in N_i[V_i]\})$ divided by $\operatorname{Prob}(\{v : V \in N_i(V_i)\})$.

4. SUMMARY AND DIRECTIONS FOR FUTURE WORK

This survey attempts to consolidate recent applications of graph theory to the analysis, interpretation, and underlying theory of loglinear models of multidimensional contingency tables.

Any hierarchical loglinear model can be uniquely represented by a generating class and associated with an interaction graph that determines all first-order interactions (and so all conditional independencies). If the maxcliques of the interaction graph correspond exactly to the generators, then the model is graphical and the higher-order interactions are also determined by the interaction graph. Consequently, graphical models can be interpreted exclusively in terms of conditional independence. In addition to identifying conditional independencies, the interaction graph can be used to determine p-collapsibility conditions (Theorem 1).

For graphical models, being chordal is equivalent to decomposability, i.e., the ability to factor each cell probability with respect to certain marginal cell probabilities. This can be stated in terms of open and closed neighborhoods in a simplicial elimination ordering of the vertices (Theorem 2). Decomposable models are also important because they are the hierarchical models that are simultaneously graphical and recursive. Hence, such models allow for causal interpretations of the structural relationships between response and explanatory (i.e., nonresponsive) variables through the use of a directed graph. The parameter ℓ, a by-product of the simplicial elimination ordering, serves as a lower bound for the number of response variables permissible in the recursive scheme (Theorem 4).

In [4], the present authors consider the intersection multigraph of the generating class. Such "generator multigraphs" are natural for recognizing decomposable models and obtaining maximum likelihood estimators (as in the present paper), and also for finding conditional independencies. The graph theory in [4] focuses on maximum spanning trees and edge cutsets (rather than on chordal graphs and minimal vertex separators as is appropriate for interaction graphs).

Several possible directions for future research are suggested by the results surveyed above. We mention four possibilities.

(i) The concept of p-collapsibility differs substantively from the well-studied concept of λ-collapsibility. For instance, a given hierarchical loglinear model is always p-collapsible onto a subset of a generator, but the same is not true for λ-collapsibility. The differences and relationships between these two notions of collapsibility (and their manifestations in the interaction graph) should be investigated further.

(ii) Because of the fundamental importance of chordal graphs being the intersection graphs of subtrees of trees, the potential significance of permissible loglinear interactions having underlying, tree-like topologies should be investigated.

(iii) Other path analysis models besides the recursive models should be investigated, including various degrees and types of recursiveness.

(iv) Recent, stronger results from chordal graph theory should be investigated in this context. An attempt should be made to identify new graph-theoretic concepts or results suggested by these applications.

In conclusion, the application of graph theory to loglinear models of contingency tables provides a rich, unexpected connection between two rather diverse areas. Increased awareness of this connection can be expected to lead to mutual benefits for both areas.

REFERENCES

1. J.N. Darroch, S.L. Lauritzen and T.P. Speed, Markov fields and log-linear interaction models for contingency tables, *Annals of Statistics* **8**, 522–539, (1980).
2. S.L. Lauritzen and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their applications to expert systems, *J. R. Statist. Soc.* **B50**, 157–224, (1988).
3. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufman, San Mateo, CA, (1988).
4. T.A. McKee and H.J. Khamis, Multigraph representations of hierarchical loglinear models, *J. Statist. Plann Inference* **53**, 63–74, (1996).
5. M.S. Barlett, Contingency table interactions, *J. R. Statist. Soc.* **B2**, 248–252, (1935).
6. M.W. Birch, Maximum likelihood in three-way contingency tables, *J. R. Statist. Soc.* **B25**, 220–233, (1963).
7. A. Agresti, *Categorical Data Analysis*, Wiley, New York, (1990).
8. N. Wermuth and S.L. Lauritzen, Graphical and recursive models for contingency tables, *Biometrika* **70**, 537–552, (1983).
9. Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, (1980).
10. A.S. Whittemore, Collapsibility of multidimensional contingency tables, *J. R. Statist. Soc.* **B40**, 328–340, (1978).
11. S. Asmussen and D. Edwards, Collapsibility and response variables in contingency tables, *Biometrika* **70**, 567–578, (1983).
12. D. Edwards and S. Kreiner, The analysis of contingency tables by graphical models, *Biometrika* **70**, 553–565, (1983).
13. S.L. Lauritzen, T.P. Speed and K. Vijayan, Decomposable graphs and hypergraphs, *J. Austral. Math. Soc.* **A36**, 12–29, (1984).
14. L.A. Goodman, The multivariate analysis of qualitative data: Interaction among multiple classifications, *J. Amer. Statist. Assoc.* **65**, 226–256, (1970).
15. S.J. Haberman, *The Analysis of Frequency Data*, (IMS Monographs) University Chicago Press, Chicago, IL, (1974).
16. A.H. Andersen, Multidimensional contingency tables, *Scand. J. Statist.* **1**, 115–127, (1974).
17. M.C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, San Diego, CA, (1980).
18. F.V. Jensen, S.L. Lauritzen and K.G. Olesen, Bayesian updating in causal probabilistic networks by local computations, *Computational Statistics Quarterly* **4**, 269–282, (1990).
19. L.A. Goodman, Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables, *J. Amer. Statist. Assoc.* **66**, 339–344, (1971).
20. S. Kotz and N.L. Johnson, (Editors-in-Chief), *Encyclopedia of Statistical Sciences*, Volume 6, Wiley, New York, (1985).
21. L.A. Goodman, The analysis of contingency tables when some variables are posterior to others: A modified path analysis approach, *Biometrika* **60**, 179–192, (1973).
22. N. Wermuth, Linear recursive equations, covariance selection, and path analysis, *J. Amer. Statist. Assoc.* **75**, 963–972, (1980).