The tree structure of graphs for various graphical models

Xiaofei Wang and Jianhua Guo*

After proper decompositions or separations, there is a common characteristic of the secondary structures for various graphical models. In this paper, we show that the junction tree captures this common characteristic. To generalize all potential occurrences in different graphical models, we define junction trees on general set classes and show several equivalent properties of junction trees. For mixed graphical models and hierarchical models, we investigate in detail the M-decomposition of marked graphs and the H-decomposition of interaction graphs, and point out the junction tree structures of marked graphs and interaction graphs. Moreover, properties of separation trees and d-separation trees are discussed for undirected and directed graphs, respectively. Both separation and d-separation trees are closely associated with junction trees. Finally, we propose two algorithms for constructing junction tree structures for mixed graphical models and hierarchical models.

1. INTRODUCTION

In statistical studies there are many kinds of data structures, and most of which can be represented as graphs. Examples include directed acyclic graphs in Bayesian networks [21, 12, 4], interaction graphs in hierarchical models [1] and marked graphs in conditional Gaussian models [13]. Those models, where conditional independence relations among variables are presented as graphs, are called graphical models [11]. The graphical model, which has its origin in several scientific areas such as statistical physics [7] and genetics [28], is a major application of graph theories to data representation in statistics.

In our work, we investigate the structures of graphs for various graphical models of high dimensions and sparse relations among variables. These experiences [16, 26] lead us to a conclusion that, after proper decompositions or separations, there should be a common characteristic for the secondary structures of those graphs. The junction tree captures this common characteristic, and holds natural tree

*Corresponding author.

properties beyond general graph structures. By using the junction tree structure, there exist efficient algorithms for probabilistic propagation computations [12, 18]. Structure learnings [30, 31] for Bayesian networks and estimates [32] on undirected graphical models of high dimensional variables can also exploit tree structures.

The junction tree is closely related to the notion of clique trees in chordal graphs, which model the sparsity structure of the Cholesky factor of a sparse positive definite matrix [22]. The clique trees [6, 24], which consist of all the cliques that are basic blocks of vertices in chordal graphs, are also junction trees. In the present paper, our results on junction trees are not limited to chordal graphs; they apply to various graphs used in graphical models such as general undirected graphs in Markov random field, interaction graphs in hierachical models and marked graphs in conditional Gaussian models.

The junction tree structure demonstrates the conditional independence relations among blocks of variables within the framework of graphical models. For different models and graphs, the descriptions of basic blocks are different, but the secondary junction tree structure can be always found. To generalize all potential occurrences in different graphical models, we present and study junction trees in an abstract setting. We consider four equivalent properties of junction trees under families of subsets, which are the junction property, the induced-subtree property, the running-intersection property, and the maximum-weight spanning tree property. The corresponding results for clique trees in chordal graphs also exist [3], but we keep our framework as general as possible.

For mixed graphical models, after considering M-decompositions with statistical meanings, we study in detail basic blocks of vertices in marked graphs. Furthermore, we point out the existence of the junction tree structures for marked graphs. Similar results for interaction graphs are also given, and corresponding corollaries for general undirected graphs and chordal graphs are shown. For more general considering on tree structures, not limited to basic blocks, we study separation trees for undirected graphs and d-separation trees for directed acyclic graphs.

To give a specific method for the tree construction of mixed graphical models, we discuss the relationship between the M-decomposition of marked graphs and the decomposition of star graphs, and show that basic blocks of marked
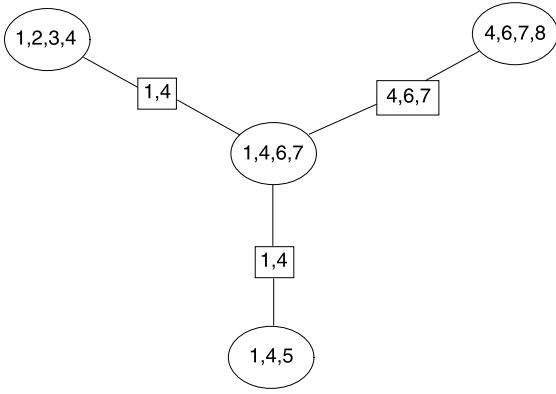
Figure 1. A secondary tree of
$\{\{1,2,3,4\},\{1,4,6,7\},\{1,4,5\},\{4,6,7,8\}\}.$

graphs can be obtained from star graphs. By using this transformation, we propose an algorithm for constructing junction trees for marked graphs. Moreover, for hierarchical models, we apply a revised edition of Leimer's decomposition algorithm [15] to interaction graphs for constructing junction trees for interaction graphs.

2. PRELIMINARIES

A family $\mathcal{U}_V$ is a collection of some subsets of a finite vertex set V. Every such subset of V is called an element in $\mathcal{U}_V$. Assume that V is the union of all the elements, i.e., $V = \bigcup_{U \in \mathcal{U}_V} U$. If one element does not contain another, we say that $\mathcal{U}_V$ is reduced. Let $T = (\mathcal{U}_V, \mathcal{E}_T)$ be a tree, of which every node is an element in $\mathcal{U}_V$, and every edge $(U_i, U_j) \in \mathcal{E}_T$ associated with an intersection set $S = U_i \bigcap U_j$, which we call an obstructor, connects nodes U_i and U_j in T. We call T a secondary tree of $\mathcal{U}_V$. Removing an obstructor S from the tree T splits T into two subtrees T_1 and T_2 with node sets $\mathcal{U}_1$ and $\mathcal{U}_2$, respectively. Let $V_i = \bigcup_{U \in \mathcal{U}_i} U$ be the union of the nodes in the subtree T_i for $i = 1$ and 2. The term "node" is used for a secondary tree to distinguish the term "vertex" for a graph. The node and obstructor of T are displayed as an ellipse and a rectangle, respectively. For example, consider a finite vertex set $V = \{1,2,3,4,5,6,7,8\}$, and assume that the family $\mathcal{U}_V$ consists of four elements $\{1,2,3,4\}$, $\{1,4,6,7\}$, $\{1,4,5\}$ and $\{4,6,7,8\}$. A secondary tree $T = (\mathcal{U}_V, \mathcal{E}_T)$ is shown in Figure 1. Its three edges are $(\{1,2,3,4\},\{1,4,6,7\})$, $(\{1,4,6,7\},\{1,4,5\})$, and $(\{1,4,6,7\},\{4,6,7,8\})$ and the three corresponding obstructors are $\{1,4\}$, $\{1,4\}$ and $\{4,6,7\}$.

Here, we choose the symbol $T = (\mathcal{U}_V, \mathcal{E}_T)$ to denote the tree structure. And it is well known that the pair $(V, \mathcal{U}_V)$ forms a hypergraph and every element $U \in \mathcal{U}_V$ is called as a hyperedge. Since U represents a node of secondary tree in this paper and is completely irrelative to the meaning of edges, we avoid using the terminology of hypergraph community even though most of the results below can be rewritten in the language of hypergraphs.

Without loss of generality, we assume that the family $\mathcal{U}_V$ is reduced and any obstructor for a secondary tree T is a non-empty set.

Definition 2.1. A secondary tree T of $\mathcal{U}_V$ has the junction property if the set $U \cap U'$ is contained in every node on the path connecting U and U' in the tree T for every pair of distinct elements $U, U' \in \mathcal{U}_V$. And T is called a junction tree of $\mathcal{U}_V$.

For any given family $\mathcal{U}_V$, $\mathcal{T}_{\mathcal{U}_V}^{jp}$ denotes the set of all the secondary trees $T = (\mathcal{U}_V, \mathcal{E}_T)$ that have the junction property. Then every $T \in \mathcal{T}_{\mathcal{U}_V}^{jp}$ is a junction tree of $\mathcal{U}_V$. The collection $\mathcal{M}_T$ of all the obstructors in a junction tree T is called a multiset. A multiset may contain two obstructors, formed by different edges in a junction tree, but with the same vertices. For instance, the secondary tree T of $\mathcal{U}_V$ in Figure 1 is a junction tree, and the edge $(\{1,2,3,4\},\{1,4,6,7\})$ forms the obstructor $\{1,4\}$, which is the same as one formed by the edge $(\{1,4,5\},\{1,4,6,7\})$.

We use $\mathcal{U}_V(v) \subseteq \mathcal{U}_V$ for the set of elements in $\mathcal{U}_V$ containing the vertex v.

Definition 2.2. A secondary tree T of $\mathcal{U}_V$ has the induced-subtree property if the set $\mathcal{U}_V(v)$ induces a subtree of T for any vertex $v \in V$.

For example, the secondary tree T of $\mathcal{U}_V$ in Figure 1 has the induced-subtree property. Let $\mathcal{T}_{\mathcal{U}_V}^{ist}$ denote the set of all the secondary trees $T = (\mathcal{U}_V, \mathcal{E}_T)$ that have the induced-subtree property.

Definition 2.3. A subset sequence $\{U_1, \ldots, U_m\}$ of V has the running intersection property (RIP) if there exists some $1 \leq i \leq j-1$ such that $U_j \bigcap (\bigcup_{k=1}^{j-1} U_k) \subseteq U_i$ for all $2 \leq j \leq m$. And the set $\{S_j | S_j = U_j \bigcap (\bigcup_{k=1}^{j-1} U_k), 2 \leq j \leq m\}$ is called a S-system of the RIP sequence $\{U_1, \ldots, U_m\}$.

For any fixed RIP sequence of a family $\mathcal{U}_V$, we can construct a secondary tree T_{rip} of $\mathcal{U}_V$ by making each element U_j adjacent to a "parent" element U_i, which is identified by "$U_j \bigcap (\bigcup_{k=1}^{j-1} U_k) \subseteq U_i$". Thus the multiset $\mathcal{M}_{T_{rip}}$ of T_{rip} is consistent with the S-system of the RIP sequence of $\mathcal{U}_V$. For instance, the secondary tree T of $\mathcal{U}_V$ in Figure 1 can be constructed from the RIP sequence $\{U_1 = \{1,2,3,4\}, U_2 = \{1,4,6,7\}, U_3 = \{1,4,5\}, U_4 = \{4,6,7,8\}\}$. Let $\mathcal{T}_{\mathcal{U}_V}^{rip}$ be the set of all the secondary trees $T = (\mathcal{U}_V, \mathcal{E}_T)$, which can be constructed from an RIP sequence of $\mathcal{U}_V$.

The following lemma given by Leimer [15] characterizes the properties of RIP sequences.

Lemma 2.1. *Let $\{U_1, \ldots, U_m\}$ be an RIP sequence, then:*

(1) *let t, $1 \leq t \leq m$, be fixed. If there is an $s \neq t$, s minimal, such that $U_t \subseteq U_s$, then:*

    (i) *$\{U_1, \ldots, U_{t-1}, U_{t+1}, \ldots, U_m\}$ is an RIP sequence if $s < t$;*

    (ii) *$\{U_1, \ldots, U_{t-1}, U_s, U_{t+1}, \ldots, U_{s-1}, U_{s+1}, \ldots, U_m\}$ is an RIP sequence if $s > t$.*

(2) *for a permutation* $\sigma : \{1, \ldots, m\} \to \{1, \ldots, m\}$, *if* $\{U_{\sigma(1)}, \ldots, U_{\sigma(m)}\}$ *is also an RIP sequence, then* $\{U_1, \ldots, U_m\}$ *and* $\{U_{\sigma(1)}, \ldots, U_{\sigma(m)}\}$ *have the same S-system.*

(3) *for any t, there exists a permutation* $\sigma : \{1, \ldots, m\} \to \{1, \ldots, m\}$ *such that* $\sigma(1) = t$ *and* $\{U_{\sigma(1)}, \ldots, U_{\sigma(m)}\}$ *is also an RIP sequence.*

For any family $\mathcal{U}_V$, there is a weighted secondary intersection graph $W_{\mathcal{U}_V}$, which is defined as follows. The vertex set of $W_{\mathcal{U}_V}$ is the family $\mathcal{U}_V$. The edge set of $W_{\mathcal{U}_V}$ is the set of all the pairwise distinct elements in $\mathcal{U}_V$, whose intersection is nonempty. Furthermore, every such edge (U, U') is assigned a weight given by $|U \bigcap U'|$. Let $\mathcal{T}_{\mathcal{U}_V}^{mst}$ denote the set of all the maximum-weight spanning secondary trees $T = (\mathcal{U}_V, \mathcal{E}_T)$ of $W_{\mathcal{U}_V}$. A secondary tree T of $\mathcal{U}_V$ has the maximum-weight spanning tree property if $T \in \mathcal{T}_{\mathcal{U}_V}^{mst}$. It is well known that $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a maximum-weight spanning tree if and only if every pair of elements $U, U' \in \mathcal{U}_V$ for which $(U, U') \notin \mathcal{E}_T$, the weight of every edge on the path joining U and U' in T is no smaller than $|U \bigcap U'|$. For example, the secondary tree T of $\mathcal{U}_V$ in Figure 1 is a maximum-weight spanning tree of $W_{\mathcal{U}_V}$.

The notion and properties of junction trees are studied in different scientific areas such as a relational database and graph theory. In the database community, acyclic database schemes [2] have many properties equivalent to the junction property. In the graph community, clique trees [3] of chordal graphs have the four properties mentioned above. However, in the following section, we will point out that the equivalence of those properties always hold and are not limited to chordal graphs.

3. CHARACTERIZATIONS OF JUNCTION TREES

In this section, we point out the four equivalent properties of junction trees, and show an invariance property of multisets.

Theorem 3.1. *For any given* $\mathcal{U}_V$, *we have* $\mathcal{T}_{\mathcal{U}_V}^{jp} = \mathcal{T}_{\mathcal{U}_V}^{ist} = \mathcal{T}_{\mathcal{U}_V}^{rip}$. *Furthermore, if* $\mathcal{T}_{\mathcal{U}_V}^{jp} \neq \emptyset$, *then* $\mathcal{T}_{\mathcal{U}_V}^{jp} = \mathcal{T}_{\mathcal{U}_V}^{ist} = \mathcal{T}_{\mathcal{U}_V}^{rip} = \mathcal{T}_{\mathcal{U}_V}^{mst}$.

Proof. If we see the elements in $\mathcal{U}_V$ as prime blocks of general undirected graphs, the proof here is similar as that of [25]. □

The condition that $\mathcal{T}_{\mathcal{U}_V}^{jp} \neq \emptyset$ for the theorem above is necessary. For example, let us consider the weighted secondary intersection graph $W_{\mathcal{U}_V}$ in Figure 2. It is easy to verify that any maximal-weight spanning tree of $\mathcal{U}_V$ may not be a junction tree, though the weighted secondary intersecting graph $W_{\mathcal{U}_V}$ and its maximum-weight spanning trees always exist.

For acyclic hypergraphs, the relationship between the junction property and the running intersection property is discussed in [2] and the maximum-weight spanning tree
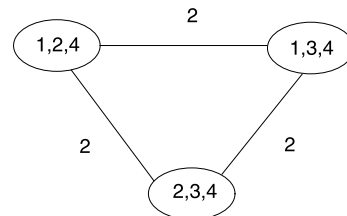

Figure 2. An example of weighted secondary intersection graphs $W_{\mathcal{U}_V}$, *in which* $\mathcal{U}_V = \{\{1, 2, 4\}, \{2, 3, 4\}, \{1, 3, 4\}\}$ *and every edge is assigned the weight* 2.

property is studied in [9] and [20]. Acyclic hypergraphs are closely related to hypertrees, and many fundamental facts about relations of acyclic hypergarphs to hypertrees are discussed in detail in [23].

The following theorem, which is also mentioned in [29], demonstrates an invariant property of the multiset.

Theorem 3.2. *For any given* $\mathcal{U}_V$ *and* $T_1, T_2 \in \mathcal{T}_{\mathcal{U}_V}^{jp}$, *we have* $\mathcal{M}_{T_1} = \mathcal{M}_{T_2}$.

Proof. For any $T \in \mathcal{T}_{\mathcal{U}_V}^{jp}$, $T \in \mathcal{T}_{\mathcal{U}_V}^{rip}$ and $\mathcal{M}_T$ is a S-system of some RIP sequence of $\mathcal{U}_V$ from Theorem 3.1. Thus we have $\mathcal{M}_{T_1} = \mathcal{M}_{T_2}$ by Lemma 2.1 (2). □

Under multivalued dependencies in relational databases, the result of Theorem 3.2 can be implied by Corollary 8.6 in [2]. Figure 3 shows that two different junction trees of $\mathcal{U}_V$ have the same multiset.

4. THE TREE STRUCTURE OF GRAPHS FOR VARIOUS GRAPHICAL MODELS

This section discusses the secondary structure of graphs for various graphical models from the viewpoint of junction trees. Without loss of generality, we assume that graphs are connected. Let $G = (V, E)$ be a undirected graph, where V is the set of vertices and E the set of undirected edges. A set of distinct vertices $[x_0, x_1, \ldots, x_k]$ is called as a path L in G between x_0 and x_k if $(x_{i-1}, x_i) \in E$ for all $i = 1, \ldots, k$. For $G = (V, E)$, a vertex subset S is called a separator for two disjoint vertex subsets A, B if every path in G between some $x \in A$ and $y \in B$ contains a vertex in S. We also say that S separates A and B in G. The separations in graphs under statistical models always mean conditional independence relations among variables.

Given two vertices $u, v \in V$, S is a uv-separator if S separates u and v in G and S is a minimal uv-separator if no proper subset of S separates u and v in G. Furthermore, S is a relative minimal separator of G if there exist two vertices u and v in G such that S is a minimal uv-separator. A clique is a maximal complete vertex set. A chord of a path is an edge joining two nonconsecutive vertices on the path. A graph is chordal (or triangulated, or decomposable) if every cycle with length greater than three has a chord.
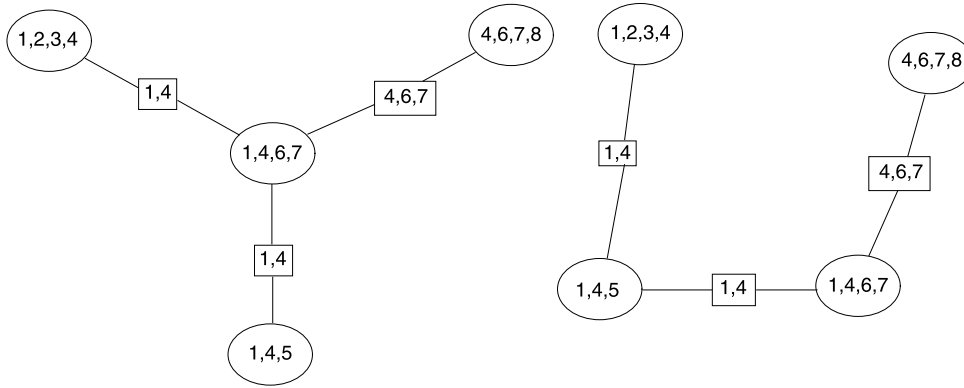
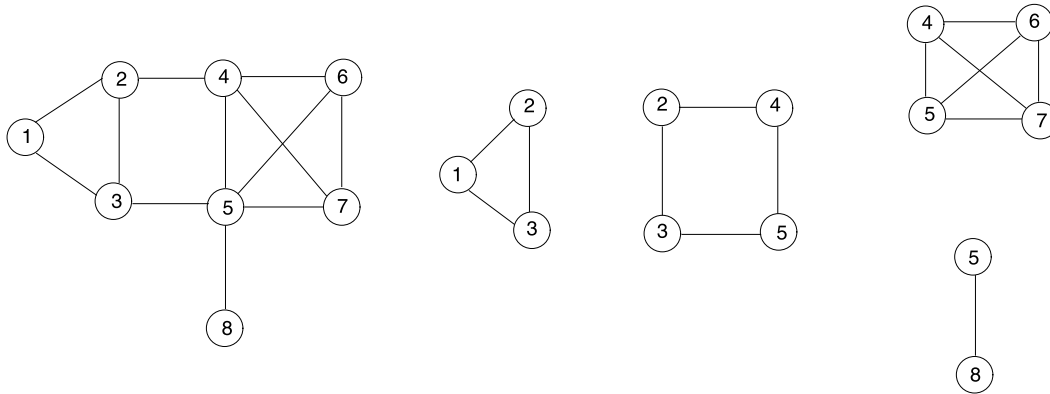*Figure 3. Two junction trees with the same multiset.*

Figure 4. A graph with four prime blocks.

If $V = A \bigcup B \bigcup S$ and $A \bigcap B = \emptyset$, $A \bigcap S = \emptyset$, $B \bigcap S = \emptyset$, we call (A, B, S) as a partition of V. For a graph $G = (V, E)$, a partition (A, B, S) of V is a decomposition of G if A, B are separated by S in G, and S is complete. The decomposition of graphs for graphical models shows that a complex system is built by combining simpler parts; estimates and tests on the whole models can be decomposed into those on submodels.

A subgraph is prime if there is no decomposition of it. An induced subgraph $G(U)$ is called a prime block of G if $G(U)$ is prime and $G(X)$ is not prime for all X with $U \subsetneq X \subseteq V$. For convenience, if $G(U)$ is a prime block of G, then we also call U a prime block of G. Assume that (A, B, S) is a decomposition of G. Furthermore, if the prime blocks of $G' = G(A \bigcup S)$ and $G'' = G(B \bigcup S)$ are pairwise different and they are all the prime blocks of G, then (A, B, S) is called a P-decomposition of G where the letter "P" denotes the word "prime" (see [15]). An example is shown in Figure 4. This graph has four prime blocks $\{\{1, 2, 3\}, \{2, 3, 4, 5\}, \{4, 5, 6, 7\}, \{5, 8\}\}$. And it can be verified that $(\{1, 2, 3\}, \{6, 7, 8\}, \{4, 5\})$ is a P-decomposition of this graph, while $(\{1, 2, 3\}, \{7, 8\}, \{4, 5, 6\})$ is a decomposition but not a P-decomposition.

4.1 Marked graphs for mixed graphical models in statistics

Mixed graphical models under conditional Gaussian distributions introduced by Lauritzen and Wermuth [13] are important in statistics for mixtures of qualitative and quantitative data. Those models are represented by marked graphs with marked vertices Δ and unmarked vertices Γ corresponding to discrete and continuous variables, respectively. In mixed graphical models, discrete variables Δ follows multinomial distributions and continuous variables Γ given Δ follows Gaussian distributions.

Assume that $G = (\Delta \bigcup \Gamma, E)$ is a marked graph with marked vertices Δ and unmarked vertices Γ, which are displayed as round rectangles and circles, respectively. In a special case, decomposable marked graphs are considered by Leimer [14], which can be applied to decomposable graphical models as discussed in Lauritzen [11]. And in this section, we mainly discuss general marked graphs for mixed graphical models.

If a partition (A, B, S) of $V = \Delta \bigcup \Gamma$ is a decomposition of $G = (V, E)$ and any of the three conditions $A \subseteq \Gamma$, $B \subseteq \Gamma$ or $S \subseteq \Delta$ holds, we call (A, B, S) an M-decomposition
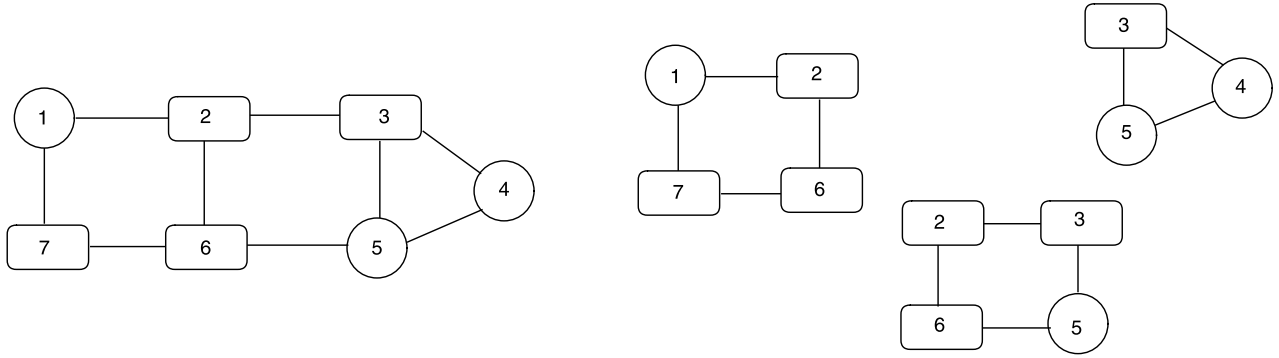
Figure 5. A marked graph with three MP-blocks.

of G where the letter "M" denotes the word "marked". A subgraph is M-prime if there is no M-decomposition of it. An induced subgraph $G(U)$ is an MP-block of G if $G(U)$ is M-prime and $G(X)$ is not M-prime for all X with $U \subsetneq X \subseteq V$. For convenience, if $G(U)$ is an MP-block of G, then we also call U an MP-block of G. Assume that (A, B, S) is an M-decomposition of G. Furthermore, if the MP-blocks of $G' = G(A \bigcup S)$ and $G'' = G(B \bigcup S)$ are pairwise different and they are all the MP-blocks of G, then (A, B, S) is called an MP-decomposition of G. For example, there are two MP-decompositions $(\{1, 7\}, \{3, 4, 5\}, \{2, 6\})$ and $(\{1, 2, 6, 7\}, \{4\}, \{3, 5\})$ of the graph G in Figure 5. $\{1, 2, 6, 7\}$, $\{2, 3, 5, 6\}$ and $\{3, 4, 5\}$ are all the three MP-blocks of G.

A subset sequence $\{U_1, \ldots, U_m\}$ of V has a marked running intersection property (MRIP) if $\{U_1, \ldots, U_m\}$ is an RIP sequence, and any of the two conditions $(U_j \setminus S_j) \subseteq \Gamma$ or $S_j \subseteq \Delta$ holds for any $2 \leq j \leq m$. The sequence with the marked running intersection property is called a SD-ordered sequence in [14].

For a marked graph G, a mixed graphical model $\mathcal{P}_G$ on it consists of conditional Gaussian distributions which are Markovian with regards to G. A graphical model $\mathcal{P}_G$ is said to be collapsible onto $B \subseteq V$ if $P(x_B) \in \mathcal{P}_{G(B)}$ for any $P(x) \in \mathcal{P}_G$. For mixed graphical models, the model collapsibility is equivalent to the estimation collapsibility that the marginality $\hat{P}(x_B)$ of the maximum-likelihood estimate on the whole model for G is just the maximum-likelihood estimate $\hat{P}_B(x_B)$ on the marginal model for $G(B)$. Frydenberg [5] also shows that a mixed model of marked graph $G = (\Gamma \bigcup \Delta, E)$ is collapsible onto A if and only if for any connected component C in A^c, the boundary of C is complete, and either $C \subseteq \Gamma$ or Δ contains the boundary of C. In fact, this equivalent condition for collapsibility expressed by graph language is consistent with the definition of M-decompositions of marked graphs, which provides statisticians a convenient way of estimating and testing for mixed models [26]. And an example for applying M-decomposition to collapsibility is described at the end of this subsection.

The following Lemma 4.1, Theorem 4.1 and Corollary 4.1, which characterize the properties of the M-decomposition and MP-blocks, are inspired by Leimer [15].

Lemma 4.1. *Let (A, B, S) be an M-decomposition of a marked graph G, then:*

(i) *If U is an MP-block of G, then $U \subseteq A \bigcup S$ or $U \subseteq B \bigcup S$ and U is an MP-block of $G(A \bigcup S)$ or $G(B \bigcup S)$, respectively.*

(ii) *Every MP-block U of $G(A \bigcup S)$ or $G(B \bigcup S)$ with $U \neq S$ is an MP-block of G.*

(iii) *If U_1 and U_2 are different MP-blocks of G, then $U_1 \bigcap U_2$ is complete.*

Proof. (i) Assume that $A \bigcap U \neq \emptyset$ and $B \bigcap U \neq \emptyset$. $(A \bigcap U, B \bigcap U, S \bigcap U)$ is an M-decomposition of $G(U)$, which is a contradiction. The second part of (i) is obvious.

(ii) Let U be an MP-block of $G(A \bigcup S)$ with $U \neq S$ (and similarly for $G(B \bigcup S)$). We can not have $U \subset S$ since $G(S)$ is M-prime, hence $U \bigcap A \neq \emptyset$. There is an MP-block X of G with $U \subseteq X$, i.e., $X \bigcap A \neq \emptyset$. It follows from part (i) of the lemma that X is an MP-block of $G(A \bigcup S)$, hence $U = X$.

(iii) $U := U_1 \bigcup U_2$. $G(U)$ is not M-prime since U_1 and U_2 are different MP-blocks of G. Let (A', B', S') be an M-decomposition of $G(U)$. Then U_1 and U_2 are also MP-blocks of $G(U)$ and part (i) of this lemma implies $U_1 \subseteq A' \bigcup S'$ and $U_2 \subseteq B' \bigcup S'$ or vice versa. Hence $U_1 \bigcap U_2 \subseteq S'$. Thus we have that $U_1 \bigcap U_2$ is complete. $\square$

By this lemma, we know that an M-decomposition (A, B, S) of G is an MP-decomposition if and only if S is neither an MP-block of $G(A \bigcup S)$ nor $G(B \bigcup S)$. If (A, B, S) is an M-decomposition of G, and S is an MP-block of both $G(A \bigcup S)$ and $G(B \bigcup S)$, then S is an MP-block of G.

Theorem 4.1. *For a marked graph $G = (\Delta \bigcup \Gamma, E)$ and a complete set $C \subseteq \Delta$, there is an MRIP sequence $\{U_1, \ldots, U_m\}$ of all the MP-blocks in G such that $C \subseteq U_1$.*

Proof. We proceed by induction on $n = |V|$. The case $n = 1$ is trivial. Consider $G = (V, E)$ with $n \geq 2$ and assume that the theorem is true for all graphs with less than n vertices.
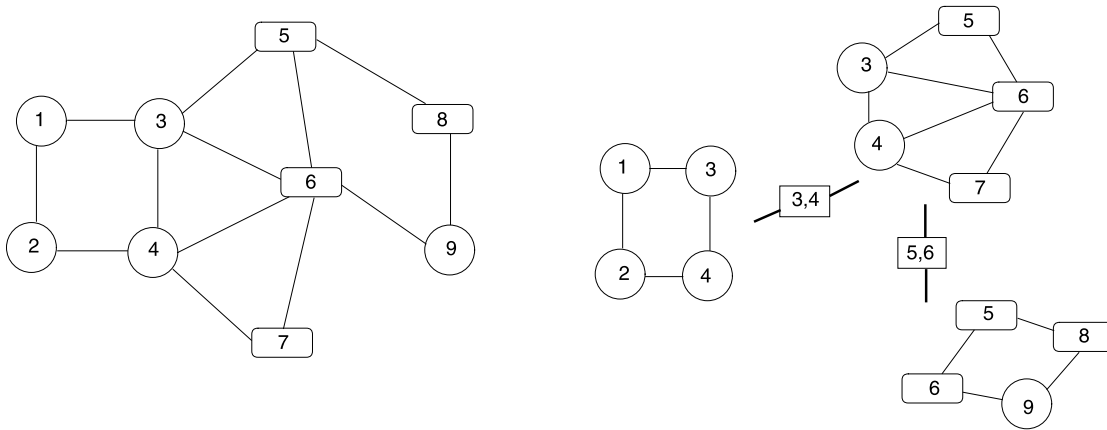
Figure 6. The marked graph $G = (V, E)$ with marked vertices $\Delta = \{5, 6, 7, 8\}$ and unmarked vertices $\Gamma = \{1, 2, 3, 4, 9\}$ which has three MP-blocks $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6, 7\}$ and $\{5, 6, 8, 9\}$.

If G is M-prime, there is nothing to prove. Otherwise, let (A, B, S) be an M-decomposition of G. Based on the induction hypothesis there are two MRIP sequences $\{A_1, \ldots, A_a\}$ and $\{B_1, \ldots, B_b\}$ of all the MP-blocks in $G(A \bigcup S)$ and $G(B \bigcup S)$, respectively.

If $S \subseteq \Delta$ and $C \subseteq A \bigcup S$, then we can assume that $C \subseteq A_1$ and $S \subseteq B_1$ according to the induction hypothesis since S and C are both complete. Thus $\{A_1, \ldots, A_a, B_1, \ldots, B_b\}$ is an MRIP sequence. The case that $S \subseteq \Delta$ and $C \subseteq B \bigcup S$ is similar.

If $B \subseteq \Gamma$ and $C \subseteq A \bigcup S$, then we can assume that $C \subseteq A_1$ according to induction hypothesis since C is complete. Thus $\{A_1, \ldots, A_a, B_1, \ldots, B_b\}$ is also an MRIP sequence. If $B \subseteq \Gamma$ and $C \subseteq B \bigcup S$, then we have $C \subseteq S$. We can assume that $C \subseteq A_1$ according to induction hypothesis since C is complete. Thus $\{A_1, \ldots, A_a, B_1, \ldots, B_b\}$ is still an MRIP sequence. The case that $A \subseteq \Gamma$ is similar.

Without loss of generality, we assume that $\{A_1, \ldots, A_a, B_1, \ldots, B_b\}$ is an MRIP sequence and $C \subseteq A_1$. If (A, B, S) is an MP-decomposition, then the above joint sequence is the desired MRIP sequence. Otherwise, we see that S is an MP-block of $G(A \bigcup S)$ or/and $G(B \bigcup S)$. In both case, we get the desired sequence after omitting S once in the joint sequence by Lemma 2.13 of Lauritzen [11]. □

Corollary 4.1. *For a marked graph G, if $\{U_1, \ldots, U_m\}$ is an MRIP sequence of all the MP-blocks of G, then the partition $(A, B, S) := ((\bigcup_{k=1}^{m-1} U_k) \setminus U_m, U_m \setminus (\bigcup_{k=1}^{m-1} U_k), (\bigcup_{k=1}^{m-1} U_k) \bigcap U_m)$ is an MP-decomposition of G into $G' = G(A \bigcup S) = G(\bigcup_{k=1}^{m-1} U_k))$ and $G'' = G(U_m)$. $\{U_1, \ldots, U_{m-1}\}$ is an MRIP sequence of all the MP-blocks of G'.*

Proof. S is complete by Lemma 4.1(iii). Using the definition of the MRIP sequence, it remains to show that (A, B, S) is a decomposition of G and that S is not an MP-block of G'.

If $(b, v) \in E$ for some $b \in B, v \in V \setminus U_n$, then there is an MP-block U of G with $\{b, v\} \subseteq U$, and we have $b \in S$ by the definition of S. It is a contradiction. This shows that (A, B, S) is a decomposition of G. Furthermore, $S \subseteq U_p$ for some $p < n$, hence S is not an MP-block in G'. Thus (A, B, S) is an MP-decomposition of G. The second part of this corollary is obvious. □

We denote the set of all the MP-blocks in G as $\mathcal{U}_G$. Then $\mathcal{U}_G$ can form junction trees and we also have the following corollary.

Corollary 4.2. *For a marked graph G, if $\mathcal{U}_G$ is the set of all the MP-blocks in G, then $\mathcal{T}_{\mathcal{U}_G}^{jp} = \mathcal{T}_{\mathcal{U}_G}^{ist} = \mathcal{T}_{\mathcal{U}_G}^{rip} = \mathcal{T}_{\mathcal{U}_G}^{mst}$.*

An example for the tree structure of marked graphs is shown in Figure 6. $\{\{1, 2, 3, 4\}, \{3, 4, 5, 6, 7\}, \{5, 6, 8, 9\}\}$ is the set $\mathcal{U}_G$ of all the MP-blocks of G, and a junction tree for $\mathcal{U}_G$ is also shown. Estimates and tests on the whole mixed graphical models can be decomposed into those on sub-models of three MP-blocks. If we are interested in variables $\{3, 7\}$, variables $\{1, 2, 8, 9\}$ can be collapsed over.

Figure 7 shows a graphical representation of an emission problem considered in Lauritzen [10]. The variables, filter state (F), waste type (W) and burning regimen (B), are conceived as discrete variables. The remaining variables are measured on a continuous quantitative scale: metals in waste (M_{in}), metals emission (M_{out}), filter efficiency (E), dust emission (D), CO_2 concentration in emission (C), and light penetrability (L).

From the junction tree structure showed in Figure 8, we have an M-decomposition $(\{B, C, E, F, L\}, \{M_{in}, M_{out}\}, \{D, W\})$. Thus $\{M_{in}, M_{out}\}$ can be collapsed over by the graph condition of collapsibility for mixed graphical models. If we are interested in statistical inferences on some variables which are contained in $\{B, C, E, F, L, D, W\}$, we can directly do the estimate on the local graphical model induced by $\{B, C, E, F, L, D, W\}$, rather than first estimate for the
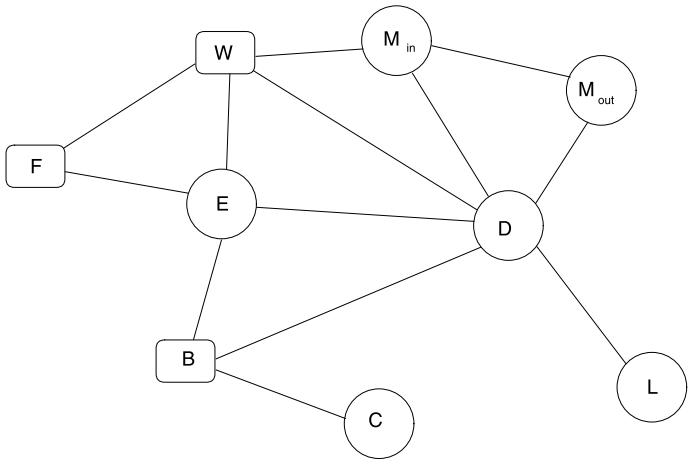
Figure 7. A marked graph from emission problems.

whole model and then do the marginality. This shows that M-decompositions and tree structures provide us a method to enhance the efficiency and accuracy of estimate. Moreover, even if $\{M_{in}, M_{out}\}$ are unobserved variables or variables with missing data, they can be negligible when we are only interested in variables in $\{B, C, E, F, L, D, W\}$.

4.2 Interaction graphs of generating classes in hierarchical models

A hierarchical model in statistics is characterized by a reduced class of variable sets, which is called a generating class. Let $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ be a generating class and

$V = \bigcup_{k=1}^m C_k$. For a subset $B \subseteq V$, we define $\mathcal{C}_B = \{C \bigcap B : C \in \mathcal{C}\}$ and assume that $\mathcal{C}_B$ is reduced for convenience.

A hierarchical model $\mathcal{P}_\mathcal{C}$ on $\mathcal{C}$ consists of multinomial distributions $P(i_V)$ satisfying that $P(i_V) = \prod_{C \in \mathcal{C}} \phi_C(i_C)$, where $\phi_C(i_C)$ is a function on C. For $B \subseteq V$, denote $\mathcal{P}_{\mathcal{C}_B}$ as the marginal model on $\mathcal{C}_B$, and $\hat{P}_{[B]}(i_B)$ as the maximum-likelihood estimate on $\mathcal{P}_{\mathcal{C}_B}$. $\mathcal{P}_\mathcal{C}$ is collapsible onto B if $P(i_B) \in \mathcal{P}_{\mathcal{C}_B}$ for any $P(i_V) \in \mathcal{P}_\mathcal{C}$. As shown in [1], $\mathcal{P}_\mathcal{C}$ is collapsible onto B if and only if the marginality $\hat{P}(i_B)$ of the maximum-likelihood estimate on the whole model is just the maximum-likelihood estimate $\hat{P}_{[B]}(i_B)$ on the marginal model $\mathcal{P}_{\mathcal{C}_B}$.

The graph $G_\mathcal{C} := (V, E_\mathcal{C})$ is the interaction graph of $\mathcal{C}$, where an edge (e, f) belongs to $E_\mathcal{C}$ if and only if $e \neq f$ and $e, f \in C_t$ for some t. If a partition (A, B, S) of V is a decomposition of $G_\mathcal{C}$ and $S \subseteq C$ for some $C \in \mathcal{C}$, we call (A, B, S) an H-decomposition of $G_\mathcal{C}$ where the letter "H" denotes the word "hierarchical". An induced subgraph of $G_\mathcal{C}$ is H-prime if there is no H-decomposition of it. An induced subgraph $G_\mathcal{C}(U)$ is called an HP-block of $G_\mathcal{C}$ if $G_\mathcal{C}(U)$ is H-prime and $G_\mathcal{C}(X)$ is not H-prime for all X with $U \subsetneq X \subseteq V$. For convenience, if $G_\mathcal{C}(U)$ is an HP-block of $G_\mathcal{C}$, then we also call U an HP-block of $G_\mathcal{C}$. Assume that (A, B, S) is an H-decomposition of $G_\mathcal{C}$. Furthermore, if the HP-blocks of $G'_\mathcal{C} = G_\mathcal{C}(A \bigcup S)$ and $G''_\mathcal{C} = G_\mathcal{C}(B \bigcup S)$ are pairwise different and they are all the HP-blocks of $G_\mathcal{C}$, then (A, B, S) is called an HP-decomposition of $G_\mathcal{C}$.

Figure 9 shows an interaction graph $G_\mathcal{C}$ of the generating class $\mathcal{C} = \{\{1, 2\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{3, 6\}, \{4, 5\}, \{5, 6\}\}$. In $G_\mathcal{C}$, $(\{1, 2, 4\}, \{6\}, \{3, 5\})$ is
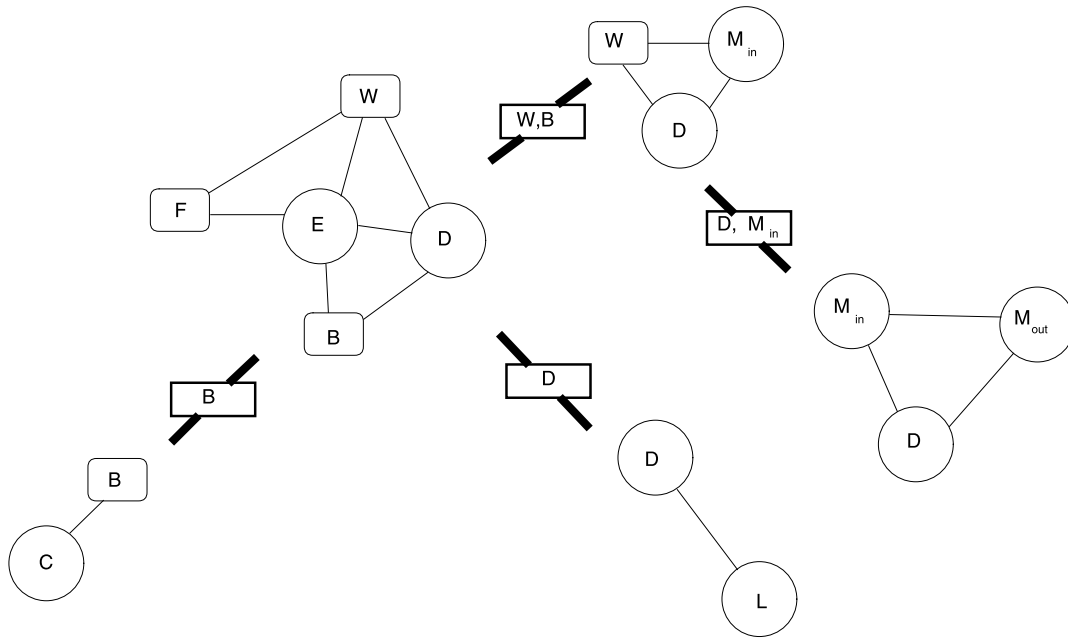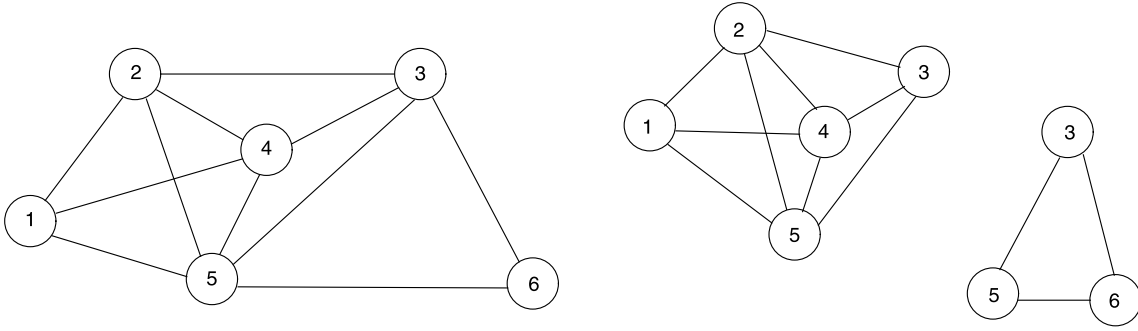

Figure 8. A junction tree consists of five MP-blocks.

Figure 9. The interaction graph $G_\mathcal{C}$ of the generating class $\mathcal{C} = \{\{1,2\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{5,6\}\}$ with two HP-blocks.

the only one HP-decomposition. $\{1,2,3,4,5\}$ and $\{3,5,6\}$ are the two HP-blocks of $G_\mathcal{C}$.

The definition of H-decompositions comes from the equivalence condition of the collapsibility of hierarchical models in statistics. Asmussen and Edwards [1] show that a hierarchical model is collapsible onto a if and only if the boundary of every connected component of a^c is contained in some generator, which is consistent with the definition of H-decompositions. For a generating class $\mathcal{C}$, if there is an H-decomposition (A, B, S) of $G_\mathcal{C}$, $\mathcal{P}_\mathcal{C}$ is collapsible onto $A \bigcup S$. Thus the maximum likelihood estimate on variables of $A \bigcup S$ directly from the sub-model on $\mathcal{C}_{A \bigcup S}$ is just the marginality of the estimate from the whole model on $\mathcal{C}$. If we are interested in some variables in $A \bigcup S$ and there are some unobserved variables in B, then the estimate on $A \bigcup S$ directly from the sub-model on $\mathcal{C}_{A \bigcup S}$ is valid and avoids unobserved variables.

The proofs of the following Lemma 4.2, Theorem 4.2 and Corollary 4.3 are similar as Lemma 4.1, Theorem 4.1 and Corollary 4.1. So we omit the proofs of Lemma 4.2, Theorem 4.2 and Corollary 4.3.

Lemma 4.2. *Let (A, B, S) be an H-decomposition of $G_\mathcal{C}$, then:*

(i) *If U is an HP-block of $G_\mathcal{C}$, then $U \subseteq A \bigcup S$ or $U \subseteq B \bigcup S$ and U is an HP-block of $G_\mathcal{C}(A \bigcup S)$ or $G_\mathcal{C}(B \bigcup S)$, respectively.*
(ii) *Every HP-block U of $G_\mathcal{C}(A \bigcup S)$ (or $G_\mathcal{C}(B \bigcup S)$) with $U \neq S$ is an HP-block of $G_\mathcal{C}$.*
(iii) *If U_1 and U_2 are different HP-blocks of $G_\mathcal{C}$, $U_1 \bigcap U_2$ is contained in some element of $\mathcal{C}$.*

From this lemma, an H-decomposition (A, B, S) is an HP-decomposition of $G_\mathcal{C}$ if and only if S is neither an HP-block of $G_\mathcal{C}(A \bigcup S)$ nor $G_\mathcal{C}(B \bigcup S)$. If (A, B, S) is an H-decomposition of $G_\mathcal{C}$, and S is an HP-block of both $G_\mathcal{C}(A \bigcup S)$ and $G_\mathcal{C}(B \bigcup S)$, then S is an HP-block of $G_\mathcal{C}$.

Theorem 4.2. *For a generating class $\mathcal{C}$, there is an ordering of all the HP-blocks $U_1, \ldots, U_m$ of $G_\mathcal{C}$ such that $\{U_1, \ldots, U_m\}$ is an RIP sequence.*

Corollary 4.3. *For a generating class $\mathcal{C}$, if $\{U_1, \ldots, U_m\}$ is an RIP sequence of all the HP-blocks of the interaction graph $G_\mathcal{C}$, the partition $(A, B, S) := ((\bigcup_{k=1}^{m-1} U_k) \setminus U_m, U_m \setminus (\bigcup_{k=1}^{m-1} U_k), (\bigcup_{k=1}^{m-1} U_k) \bigcap U_m)$ is an HP-decomposition of $G_\mathcal{C}$ into $G'_\mathcal{C} = G_\mathcal{C}(A \bigcup S)$ and $G''_\mathcal{C} = G_\mathcal{C}(U_m)$. $\{U_1, \ldots, U_{m-1}\}$ is an RIP sequence of all the HP-blocks of $G'_\mathcal{C}$.*

The decomposition algorithm proposed by Leimer [15] can be revised to find the set $\mathcal{U}_\mathcal{C}$ of all the HP-blocks in $G_\mathcal{C}$. Then $\mathcal{U}_\mathcal{C}$ can form junction trees and we also have the following corollary.

Corollary 4.4. *For a generated class $\mathcal{C}$, if $\mathcal{U}_\mathcal{C}$ is the set of all the HP-blocks in $G_\mathcal{C}$, then $\mathcal{T}_{\mathcal{U}_\mathcal{C}}^{jp} = \mathcal{T}_{\mathcal{U}_\mathcal{C}}^{ist} = \mathcal{T}_{\mathcal{U}_\mathcal{C}}^{rip} = \mathcal{T}_{\mathcal{U}_\mathcal{C}}^{mst}$.*

Let us consider a hierarchical model whose generating class is $\mathcal{C} = \{\{1,2,3\}, \{2,3,4\}, \{2,3,5\}, \{2,6\}, \{4,5,6,7\}\}$. This generating class forms an interaction graph $G_\mathcal{C}$, which has three HP-blocks, and a junction tree of all the three HP-blocks is shown in Figure 10. This tree structure from H-decomposition demonstrates the independence relations among subsets of variables within the framework of the globe. By the equivalent condition of collapsibility for hierarchical models, the estimates and tests on the whole models can be decomposed into those on sub-models of three HP-blocks. If we are only interested in variables $\{3,6\}$, we can directly do the estimate on the local model induced by $\{2,3,4,5,6\}$. This shows that even if $\{1,7\}$ are unobserved variables or variables with missing data, H-decompositions and tree structures can help us eliminate $\{1,7\}$.

Let G be an undirected graph and $\mathcal{C}$ be the set of all the cliques in G. G can be viewed as the interaction graph $G_\mathcal{C}$ of $\mathcal{C}$. Thus (A, B, S) is an HP-decomposition of $G_\mathcal{C}$ if and only if (A, B, S) is a P-decomposition of G, and U is an HP-block of $G_\mathcal{C}$ if and only if U is a prime block of G. The following corollary is also referred to by Wang and Guo [25].

Corollary 4.5. *If $G = (V, E)$ is an undirected graph, and $\mathcal{U}_G$ is the set of all the prime blocks in G, then $\mathcal{T}_{\mathcal{U}_G}^{jp} = \mathcal{T}_{\mathcal{U}_G}^{ist} = \mathcal{T}_{\mathcal{U}_G}^{rip} = \mathcal{T}_{\mathcal{U}_G}^{mst}$.*
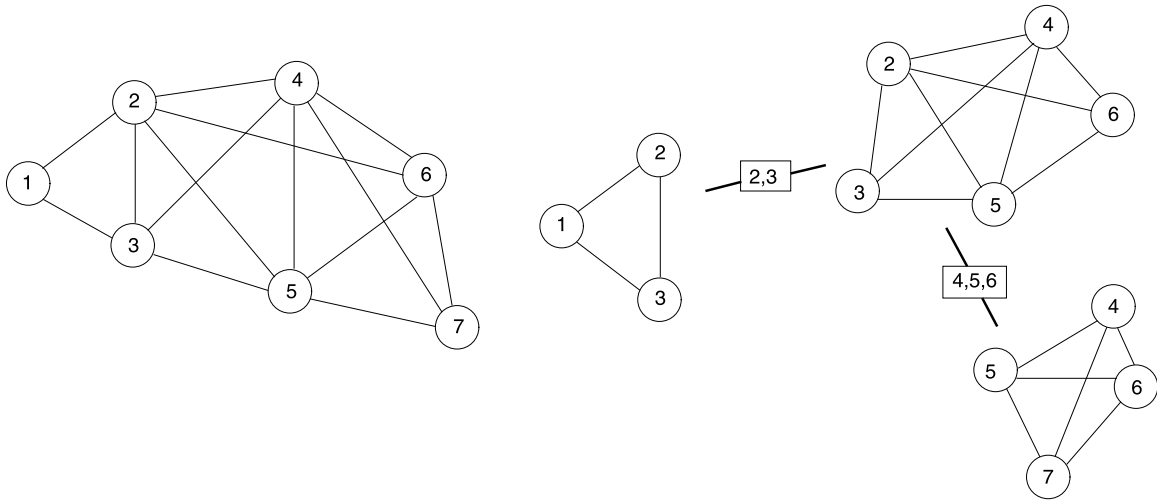
Figure 10. The interaction graph $G_\mathcal{C}$ of the generating class $\mathcal{C} = \{\{1,2,3\},\{2,3,4\},\{2,3,5\},\{2,6\},\{4,5,6,7\}\}$ which has three HP-blocks $\{1,2,3\}, \{2,3,4,5,6\}$ and $\{4,5,6,7\}$.
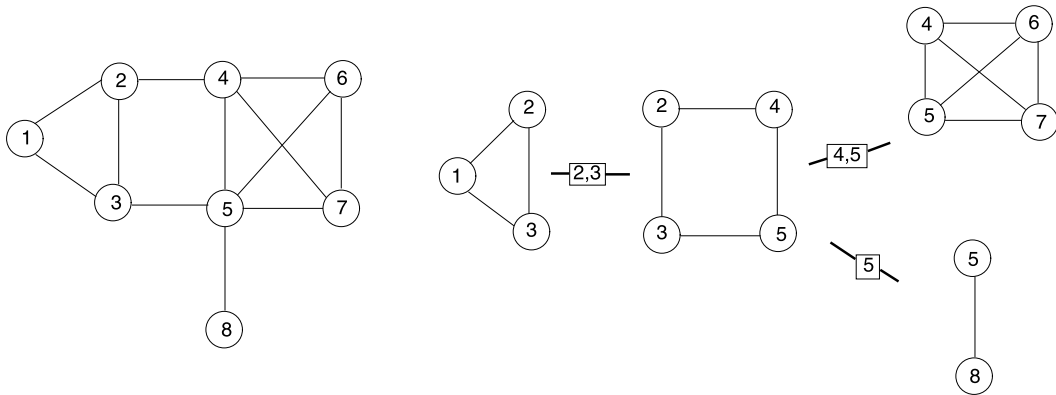

Figure 11. Graph G with four prime blocks and a junction tree T of the set of all the prime blocks in this graph.
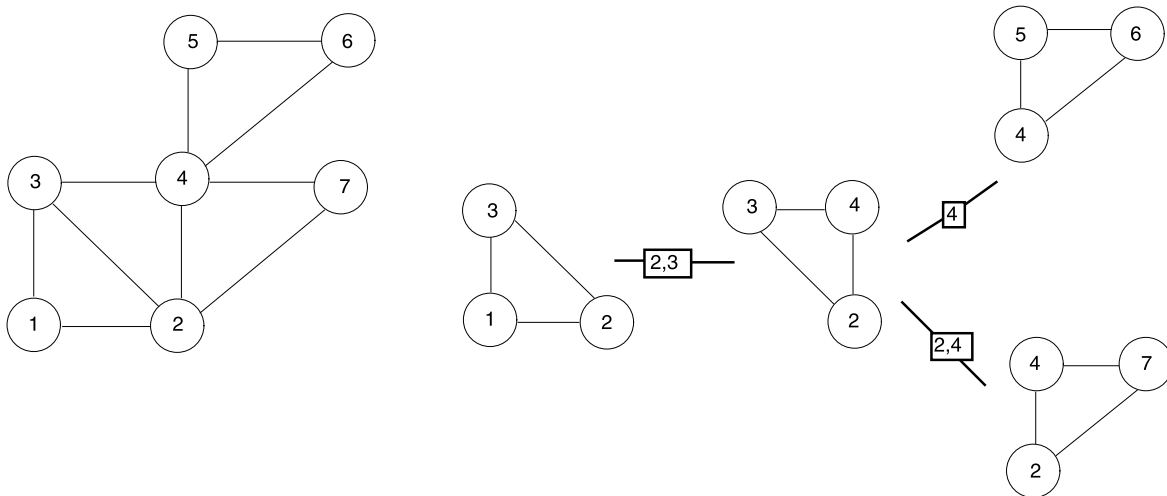

Figure 12. Chordal graph G with four cliques and a junction tree T of the set of all the cliques in this chordal graph.

Furthermore, if G is a chordal graph, the set of all the prime blocks of G is just the set of all the cliques of G. Then we have the following corollary which is also studied by Blair and Peyton [3].

Corollary 4.6. *If $G = (V, E)$ is a chordal graph, and $\mathcal{K}_G$ is the set of all the cliques in G, then $\mathcal{T}_{\mathcal{K}_G}^{jp} = \mathcal{T}_{\mathcal{K}_G}^{ist} = \mathcal{T}_{\mathcal{K}_G}^{rip} = \mathcal{T}_{\mathcal{K}_G}^{mst}$.*

4.3 Separation trees for general undirected graphs

Let $G = (V, E)$ be an undirected graph and $\mathcal{U}_V$ be a family of subsets of V. A secondary tree $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a separation tree for $G = (V, E)$ if for any $(U, U') \in \mathcal{E}_T$, $S = U \bigcap U'$ can separate the vertex sets $V_1 \setminus S$ and $V_2 \setminus S$ of two subtrees T_1 and T_2 obtained by removing the edge (U, U') from T (see [16]). The following Theorem 4.3 shows the relationship between separation trees and junction trees.

Theorem 4.3. *For an undirected graph $G = (V, E)$, $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a separation tree for G if and only if $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a junction tree and for any edge $(x, y) \in E$ there is a $U \in \mathcal{U}_V$ such that $x, y \in U$.*

Proof. To verify the necessity, we show that for every pair of distinct elements $U, U' \in \mathcal{U}_V$, the set $U \bigcap U'$ is contained in every element on the path connecting U and U' in the T. If U and U' are adjacent in T, it is trivial. If there is a path $[U_1 = U, U_2, \ldots, U_l, U_{l+1} = U']$ in T, then $U \bigcap U'$ is contained in $S_k = U_k \bigcap U_{k+1}$ for $k = 1, 2, \ldots, n$ because of the separation property of S_k. Thus T is a junction tree. For any $(x, y) \in E$, if there does not exist a $U \in \mathcal{U}_V$ such that $x, y \in U$, x and y are separated from the definition of separation trees. It is a contradiction to the fact that (x, y) is an edge.

To verify the sufficiency, we show that for any $(U, U') \in \mathcal{E}_T$, $S = U \bigcap U'$ can separate the vertex sets $V_1 \setminus S$ and $V_2 \setminus S$ of two subtrees T_1 and T_2 obtained by removing the edge (U, U') from T. If not, there is $x \in V_1 \setminus S$ and $y \in V_2 \setminus S$ such that $(x, y) \in E$. Then there is a $U \in \mathcal{U}_V$ such that $x, y \in U$. Without loss of generality, we assume that U in T_1. From the junction property, y is contained in S. So it is a contradiction to the fact that $y \in V_2 \setminus S$. $\square$

From Theorem 4.3, we know that the separation tree also has three other properties.

Corollary 4.7. *A separation tree $T = (\mathcal{U}_V, \mathcal{E}_T)$ for an undirected graph $G = (V, E)$ has the junction property, the induced-subtree property, the running-intersection property, and the maximum-weight spanning tree property.*

Proof. It follows from Theorem 4.3 and Theorem 3.1. $\square$

The following Theorem 4.4 shows a specific separation tree for an undirected connected graph G.

Theorem 4.4. *For an undirected graph $G = (V, E)$ and the set $\mathcal{U}_G$ of all the prime blocks in G, any junction tree $T = (\mathcal{U}_G, \mathcal{E}_T) \in \mathcal{T}_{\mathcal{U}_G}^{jp}$ is a separation tree for G.*

Proof. It follows from Theorem 4.3. $\square$

For any undirected graph $G = (V, E)$, we can construct a separation tree whose multiset consists of a maximal set of pairwise non-crossing relative minimal separators of G. Two relative minimal separators S and T are said to be crossing if S is a uv-separator for a pair of vertices $u, v \in T$, in which case T is an xy-separator for a pair of vertices $x, y \in S$. Let $\mathcal{S}_G$ denote a maximal set of pairwise non-crossing relative minimal separators of G. We can obtain a chordal graph G' by putting every separator in $\mathcal{S}_G$ into a complete vertex set. Then $\mathcal{S}_G$ is the set of all the relative minimal separators of G' by Property 5.3 of Heggernes [8]. The set $\mathcal{K}_{G'}$ of all the cliques in G', which is a reduced subset class of V, can form a junction tree and also a separation tree $T = (\mathcal{K}_{G'}, \mathcal{E}_T)$ for G' and G by Corollary 4.4 and Theorem 4.4. Thus the multiset $\mathcal{M}_T$ in the separation tree T consists of the set $\mathcal{S}_G$. An invariant property for separation trees is characterized in the following theorem.

Theorem 4.5. *For an undirected graph $G = (V, E)$ and a family $\mathcal{U}_V$ of V, if $T = (\mathcal{U}_V, E_T)$ is a separation tree for G and $T' = (\mathcal{U}_V, E_{T'})$ is a junction tree, then we have that T' is also a separation tree for G.*

Proof. It follows from Theorem 4.3. $\square$

4.4 D-separation trees for directed acyclic graphs

The D-separation tree is a useful notion in statistics, which is introduced by Xie, Geng and Zhao [30] to characterize the structure of conditional independencies among multiple variable sets. By using D-separation trees, the structural learning of a directed acyclic graph can be decomposed into problems related to small subgraphs. Thus both the efficiency of structural learning and the power of conditional independence tests can be improved. Recently, Liu, Guo and Jing [16] have studied a minimal d-separation tree under a partial ordering, by which the maximal efficiency of learning can be obtained. And in this subsection, we focus on the properties of D-separation trees.

Let $\overrightarrow{G} = (V, \overrightarrow{E})$ be a directed acyclic graph, where V is the vertex set and $\overrightarrow{E}$ is the set of directed edges. A directed edge from a vertex u to a vertex v is denoted by $u \rightarrow v$, and u is called a parent of v. The set of all parents of a vertex v is denoted by $pa(v)$. $u \rightarrow v \leftarrow w$ forms a v-structure if there is no edge from u to w or from w to u. A moral graph $\overrightarrow{G}^m$ for $\overrightarrow{G} = (V, \overrightarrow{E})$ is an undirected graph (V, E^m) where $E^m = \{(u, v) : u \rightarrow v \text{ or } v \rightarrow u \text{ in } \overrightarrow{E}\} \bigcup \{(u, v) : u \rightarrow w \leftarrow v$ forms a v-structure$\}$.

We define a set of distinct vertices $[x_0 = u, \ldots, x_{n-1}, x_n = v]$ as a path in G between u and v if $x_{i-1} \rightarrow x_i$ or $x_i \rightarrow x_{i-1}$ is contained in $\overrightarrow{E}$ for $i = 1, \ldots, n$. v is a descendant of u if there is a path

between u and v in $\overrightarrow{G}$ and all edges on this path point at the direction toward v. A path l is d-separated by a set of vertices S, if l contains three vertices u, r, v such that either $r \in S$ satisfying one of the three conditions $u \to r \to v$, $v \to r \to u$ and $u \leftarrow r \to v$, or $r \notin S$ satisfying $u \to r \leftarrow v$ and no descendant of r in S. Two distinct sets X and Y of vertices are d-separated by a set S if S d-separates every path between any vertex in X and any vertex in Y. Let $\overrightarrow{G} = (V, \overrightarrow{E})$ be a directed acyclic graph and $\mathcal{U}_V$ be a family of subsets of V. A secondary tree $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a d-separation tree for $\overrightarrow{G} = (V, \overrightarrow{E})$ if for any edge $(U, U') \in \mathcal{E}_T$, $S = U \bigcap U'$ d-separates, in $\overrightarrow{G}$, the vertex sets $V_1 \setminus S$ and $V_2 \setminus S$ of two subtrees T_1 and T_2 obtained by removing the (U, U') from T (see [30]). The following Theorem 4.6 shows the relationship between d-separation trees and junction trees.

Theorem 4.6. *For a directed acyclic graph $\overrightarrow{G} = (V, \overrightarrow{E})$, $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a d-separation tree for $\overrightarrow{G}$ if and only if $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a junction tree and for any $x \in V$ there is a $U \in \mathcal{U}_V$ such that $(\{x\} \bigcup pa(x)) \subseteq U$.*

Proof. It is well known that T is a d-separation tree for $\overrightarrow{G}$ if and only if T is a separation tree for $\overrightarrow{G}^m = (V, E^m)$ (see [16]).

If T is a d-separation tree for $\overrightarrow{G}$, then T is a separation tree for $\overrightarrow{G}^m$. From Theorem 4.3, T is a junction tree. Since $\{x\} \bigcup pa(x)$ is complete in $\overrightarrow{G}^m$ and T is a separation tree for $\overrightarrow{G}^m$, thus there is a $U \in \mathcal{U}_V$ such that $(\{x\} \bigcup pa(x)) \subseteq U$.

If for any $x \in V$ there is a $U \in \mathcal{U}_V$ such that $(\{x\} \bigcup pa(x)) \subseteq U$, then for any edge $(u, v) \in E^m$ there is a $U \in \mathcal{U}_V$ such that $u, v \in U$. Furthermore, if $T = (\mathcal{U}_V, \mathcal{E}_T)$ is a junction tree, then T is a separation tree for $\overrightarrow{G}^m$ from Theorem 4.3. Thus, T is a d-separation tree for $\overrightarrow{G}$. $\square$

From Theorem 4.6 we know that the d-separation tree also has three other properties.

Corollary 4.8. *For a directed acyclic graph $\overrightarrow{G} = (V, \overrightarrow{E})$ and a d-separation tree $T = (\mathcal{U}_V, \mathcal{E}_T)$ for $\overrightarrow{G}$, T has the junction property, the induced-subtree property, the running-intersection property, and the maximum-weight spanning tree property.*

If we substitute the d-separation tree for the separation tree in Theorem 4.5, the corresponding conclusion also holds.

Theorem 4.7. *For a directed acyclic graph $\overrightarrow{G} = (V, \overrightarrow{E})$, if $T = (\mathcal{U}_V, E_T)$ is a d-separation tree for $\overrightarrow{G}$ and $T' = (\mathcal{U}_V, E_{T'})$ is a junction tree, then we have that T' is also a d-separation tree for $\overrightarrow{G}$.*

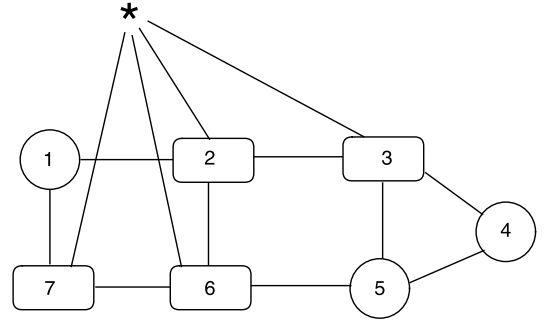*Proof.* It follows from Theorem 4.6. $\square$

Figure 13. Star graph constructed from Figure 5.

5. ALGORITHMS FOR CONSTRUCTING JUNCTION TREES FOR MIXED GRAPHICAL MODELS AND HIERARCHICAL MODELS

In this section, we discuss how to construct junction trees for mixed graphical models and hierarchical models. First, we consider the case for mixed graphical models. We need to introduce the notion of star graphs for mark graphs. For a marked graph $G = (V = \Gamma \bigcup \Delta, E)$, we can construct a star graph from G by adding $\star$ into its vertex set and connecting this $\star$ with every discrete variables, and denote this star graph as $G^\star = (V \bigcup \{\star\}, E^\star)$, where $E^\star = \{(\delta, \star) | \delta \in \Delta\} \bigcup E$. For any subset $B \subseteq V$, the star graph $(G(B))^\star$ of the induced graph $G(B)$ is just the induced subgraph $G^\star(B \bigcup \{\star\})$ of the star graph $G^\star$. Figure 13, in which $\star$ is connected with all the discrete variables, is the star graph constructed from Figure 5.

Lemma 5.1. *Let $G = (V, E)$ be a marked graph and (A, B, S) is a partition of V with $A \neq \emptyset$ and $B \neq \emptyset$, then:*

(i) *(A, B, S) is an M-decomposition of G and $A \subseteq \Gamma$ if and only if $(A, B \bigcup \{\star\}, S)$ is a decomposition of $G^\star$.*
(ii) *(A, B, S) is an M-decomposition of G and $B \subseteq \Gamma$ if and only if $(A \bigcup \{\star\}, B, S)$ is a decomposition of $G^\star$.*
(iii) *(A, B, S) is an M-decomposition of G and $S \subseteq \Delta$ if and only if $(A, B, S \bigcup \{\star\})$ is a decomposition of $G^\star$.*

Proof. It can be gotten by the relation between marked graphs and star graphs. $\square$

Theorem 5.1. *Let $G = (V, E)$ be a marked graph and $A \subseteq V$, then $G(A)$ is M-prime in G if and only if $G^\star(A \bigcup \{\star\})$ is prime in $G^\star$ or $G^\star(A)$ is prime in $G^\star$.*

Proof. $(\Rightarrow)$ If neither $G^\star(A \bigcup \{\star\})$ nor $G^\star(A)$ is prime in $G^\star$, then there is a decomposition (A', B', S') of $G^\star(A \bigcup \{\star\})$, and a decomposition (A'', B'', S'') of $G^\star(A)$. Then we consider three cases: First, let us consider that $\star \in B'$. If $B' = \{\star\}$, then $A'' \subseteq \Gamma$ or $B'' \subseteq \Gamma$ since $A \bigcap \Delta \subseteq S'$ and S' is complete. Thus (A'', B'', S'') is an M-decomposition of $G(A)$. This is a contradiction. If $\{\star\} \subsetneq B'$, then $(A', B' \setminus \{\star\}, S')$ is

an M-decomposition of $G(A)$ by Lemma 5.1. This is a contradiction. Second, let us consider that $\star \in A'$. It is similar to the above discussion. Third, let us consider that $\star \in S'$. We have that $(A', B', S' \setminus \{\star\})$ is an M-decomposition of $G(A)$ by Lemma 5.1. It is a contradiction.

($\Leftarrow$) If $G(A)$ is not M-prime in G, then there is an M-decomposition (A', B', S') of $G(A)$. Then we consider three cases: First, if $B' \subseteq \Gamma$, then $(A' \bigcup\{\star\}, B', S')$ is a decomposition of $G^\star(A\bigcup\{\star\})$ by Lemma 5.1 and (A', B', S') is a decomposition of $G^\star(A)$. This is a contradiction. Second, if $A \subseteq \Gamma$, then it is similar to the above discussion. Third, if $S' \subseteq \Delta$, $(A', B', S' \bigcup\{\star\})$ is a decomposition of $G^\star(A\bigcup\{\star\})$ by Lemma 5.1 and (A', B', S') is a decomposition of $G^\star(A)$. This is a contradiction. □

Theorem 5.2. *Let $G = (V, E)$ be a marked graph and $U \subseteq V$, then U is an MP-block of G if and only if $U \bigcup\{\star\}$ is the unique prime block containing U of $G^\star$ or U is a prime block of $G^\star$.*

Proof. ($\Leftarrow$) Since $G^\star(U\bigcup\{\star\})$ or $G^\star(U)$ is prime in $G^\star$, $G(U)$ is M-prime in G by Theorem 5.1. If U is not an MP-block, then there is a subset $B \subseteq V$ such that $U \subsetneqq B$ and $G(B)$ is M-prime in G. By Theorem 5.1, $G^\star(B\bigcup\{\star\})$ or $G^\star(B)$ is prime in $G^\star$. If $G^\star(B\bigcup\{\star\})$ is prime, then neither $U\bigcup\{\star\}$ nor U is a prime block of $G^\star$ since $U \subsetneqq U\bigcup\{\star\} \subsetneqq B\bigcup\{\star\}$. If $G^\star(B)$ is prime, U is not a prime block and $U\bigcup\{\star\}$ is not the unique prime block containing U of $G^\star$. It is also a contradiction.

($\Rightarrow$) Since U is an MP-bock of G, $G^\star(U\bigcup\{\star\})$ is prime or $G^\star(U)$ is prime.

If $G^\star(U\bigcup\{\star\})$ is prime, then $G^\star(U)$ is not a prime block of $G^\star$. Furthermore, if $U\bigcup\{\star\}$ is not a prime block of $G^\star$, then there is a subset $B \subseteq V$ such that $U \subsetneqq B$ and $G^\star(B\bigcup\{\star\})$ is prime. Thus $G(B)$ is M-prime in G by Theorem 5.1. It is in contradiction with the fact that U is an MP-block of G. Thus $U\bigcup\{\star\}$ is a prime block of $G^\star$. If there is a subset $B \subseteq V$ such that $U \subsetneqq B$ and $G^\star(B)$ is prime, then $G(B)$ is M-prime in G. It is in contradiction with the fact that U is an MP-block of G. Thus $U\bigcup\{\star\}$ is the unique prime block containing U of $G^\star$.

If $G^\star(U\bigcup\{\star\})$ is not prime, then $G^\star(U)$ is prime. If U is not a prime block of $G^\star$, there is a subset $B \subseteq V\bigcup\{\star\}$ such that $U \subsetneqq B$ and $G^\star(B)$ is prime in $G^\star$. Since $G^\star(U\bigcup\{\star\})$ is not prime in $G^\star$, we have $U\bigcup\{\star\} \neq B$. If $\star \in B$, then $G(B\setminus\{\star\})$ is M-prime by Theorem 5.1. It is a contradiction. If $\star \notin B$, $G(B)$ is M-prime by Theorem 5.1. It is also a contradiction. □

Theorem 5.3. *If $\{U_1, \ldots, U_n\}$ is a RIP sequence of all the prime blocks of $G^\star$ and $\star \in U_1$, then $\{U_1 \setminus \{\star\}, \ldots, U_n \setminus \{\star\}\}$ is an MRIP sequence.*

Proof. It can be gotten by using induction. □

From Theorem 5.3 and Lemma 2.13 of Lauritzen [11], an MRIP sequence of all the MP-blocks of G can be obtained by a RIP sequence of all the prime blocks of $G^\star$, where it

only needs to judge the inclusion relationship among subsets. Thus we use Leimer's algorithm [15] for $G^\star$, which describes the P-decomposition for undirected graphs, to construct junction trees for marked graphs.

Algorithm Junction_ Trees_ For_ Mixed_ Graphical_ Models
Input: A marked graph G for a mixed graphical model.
Output: A junction tree for $\mathcal{U}_G$ of all the MP-blocks of G.
Begin
1. Construct a star graph G^* for G.
2. Call Leimer's algorithm [15] for finding a RIP sequence $\{U_1, U_2, \ldots, U_n\}$ of all the prime blocks of G^* such that $\star \in U_1$.
3. For $t = 1, \ldots, n$, drop the star in U_t if $\star \in U_t$. And we obtain an MRIP sequence $\{U_1', U_2', \ldots, U_n'\}$ where $U_t' = U_t\setminus\{*\}$ for $t = 1, \ldots, m$.
4. Repeatedly remove the element U in $\{U_1', U_2', \ldots, U_n'\}$ if U is contained in another element, and adjust the ordering by Lemma 2.13 of Lauritzen [11]. Then we obtain an MRIP sequence $\{U_1'', U_2'', \ldots, U_m''\}$ of all the MP-blocks of G.
6. Construct junction tree from this MRIP sequence.
End

Now we consider the case for hierarchical models. We can apply a slight revised edition of Leimer's algorithm to the interaction graph of the generating class for the hierarchical model.

In the HP-decomposition (A, B, S) for interaction graphs, S is required to be contained in some element of the generating class. So we only need to judge whether $C(f_t)$, which is defined in [15], is contained in some element in generating class. Thus we obtain an algorithm for finding junction trees for hierarchical models as follows:

Algorithm Junction_Trees_ For_ Hierarchical_ Models
Input: A generating class $\mathcal{C}$ of a hierarchical model.
Output: A junction tree for $\mathcal{U}_\mathcal{C}$ of all the HP-blocks of $G_\mathcal{C}$.
Begin
1. Construct $G_\mathcal{C}$ from $\mathcal{C}$.
2. Call Leimer's algorithm [15] for $G_\mathcal{C}$, and in addition, the decomposition step is successful only if $C(f_t)$ is contained in some element in $\mathcal{C}$. Then we obtain a RIP sequence $\{U_1, \ldots, U_m\}$ of all the HP-block of $G_\mathcal{C}$.
3. Construct junction from this RIP sequence.
End

6. DISCUSSION AND CONCLUSION

In this paper, we consider the structure of graphs in various graphical models from the viewpoint of junction trees, and these trees have four equivalent properties: junction properties, induced subtree properties, running intersection properties and maximum weight spanning tree properties. This common tree structure can efficiently demonstrate lots of conditional independence relations in statistical models,

which provides us convenience to divide and conquer estimates [32], tests [5] and computations [12].

To handle with all the potential occurrences in different graphical models, we present the definition of junction trees as general as possible, and investigate secondary structures for graphs used in different graphical models.

Our investigation shows that there exist junction tree structures for marked graphs in mixed graphical models. And a similar result is also for interaction graphs in hierarchical models. These secondary structures of graphical models can be exploited to find essential models, containing our interested variables, of lower dimensions and to collapse unobserved variables or variables of missing data over.

We study the relations among junction trees, separation trees and d-separation trees. Properties of d-separation trees can be applied to learnings of Bayesian networks, which efficiently improve the power of conditional independence tests [16, 30].

For mixed graphical models and hierarchical models, we propose two algorithms to construct junction trees for marked graphs and interaction graphs. And these junction tree structures obtained by our algorithms can be applied to collapsibility for mixed graphical models and hierarchical models, and enhance the efficiency of estimates and tests.

Statistical models and relational databases have a close relation, which is not further considered in this paper. Wong et al. [27] give a detailed discussion on the relation between Bayesian networks and relational databases for the implication problem. Some algorithms in database areas can also be applied to statistical problems. Graham's algorithm can be used for collapsibility in decomposable graphical models [17]. Fagin's decomposition algorithm [2] has reference values to the decomposition of statistical models.

In this paper, we have not discussed the association between marked graphs and mixed Bayesian networks. Actually, moral graphs for mixed Bayesian networks are marked graphs. By some proper triangulations of these marked graphs, we can obtain junction trees to provide efficient algorithms for probabilistic propagation computations [19]. A further work for applying junction trees to those computations in mixed Bayesian networks is a topic of our future research.

ACKNOWLEDGEMENTS

Received 12 January 2012

REFERENCES

[1] ASMUSSEN, S. and EDWARDS, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70** 567–578. MR0725370

[2] BEERI, C., FAGIN, R., MAIER, D. and YANNAKAKIS, M. (1983). On the desirability of acyclic database schemes. *JACM* **30** 479–513. MR0709830

[3] BLAIR, J. R. S. and PEYTON, B. W. (1993). *An introduction to chordal graphs and clique trees.* In: Sparse matrix computations: Graph theory issues and algorithms, IMA Volumes in Mathematics and its Applications, George J. A., Liu J. W. H., eds. Springer-Verlag, Berlin, **56**, pp. 1–30. MR1320296

[4] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGEL-HALTER, D. J. (1999). *Probabilistic Networks and Expert Systems.* Springer-Verlag, New York. MR1697175

[5] FRYDENBERG, M. (1990). Marginalization and collapsibility in graphical interaction models. *Ann. Statist.* **18** 790–805. MR1056337

[6] GAVRIL, F. (1974). The intersection graphs of subtrees in trees are exactly the chordal graphs. *J. Combin. Theory Ser. B* **16** 47–56. MR0332541

[7] GIBBS, W. (1902). *Elementary Principles of Statistical Mechanics.* Yale University Press, New Haven.

[8] HEGGERNES, P. (2006). Minimal triangulations of graphs: A survey. *Discrete Math.* **306** 297–317. MR2204109

[9] JENSEN, F. V. (1988). *Junction trees and decomposable hypergraphs.* Technical report, JUDEX, Aalborg, Denmark.

[10] LAURIZEN, S. L. (1992). Propagation of probabilities, means, and variances in mixed graphical association models. *JASA* **87** 1098–1108. MR1209568

[11] LAURITZEN, S. L. (1996). *Graphical Models.* Clarendon Press, Oxford. MR1419991

[12] LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 157–224. MR0964177

[13] LAURITZEN, S. L. and WERMUTH N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. MR0981437

[14] LEIMER, H. G. (1989). Triangulated graphs with marked vertices. *Ann. Discrete Math.* **41** 311–324. MR0976010

[15] LEIMER, H. G. (1993). Optimal decomposition by clique separators. *Discrete Math.* **113** 99–123. MR1212872

[16] LIU, B. H., GUO, J. H. and JING, B. Y. (2010). A note on minimal d-separation trees for structural learning. *Artif. Intell.* **174** 442–448. MR2655946

[17] MADIGAN, D. and MOSURSKI K. (1990). An extension of the results of Asmussen and Edwards on collapsibility in contingency tables. *Biometrika* **77** 315–319. MR1064803

[18] MADSEN, A. L. (2006). Variations over the message computation algorithm of lazy propagation. *IEEE T. Syst. Man. Cy. B* **36** 636–648.

[19] MADSEN, A. L. (2008). Belief update in CLG Bayesian networks with lazy propagation. *Int. J. Approx. Reason.* **49** 503–521.

[20] MAIER, D. (1983). *The Theory of Relational Databases.* Computer Science Press. MR0691493

[21] PEARL, J. (1988). *Probabilistic Reasoning in Intelligence Systems.* Morgan Kaufmann, San Mateo. MR0965765

[22] ROSE, D. J. (1972). *A graph-theoretic study of the numerical solution of sparse positive definite systems of liner equations in Graph Theory and Computing*, R. C. Read, eds., Academic Press, 18–217. MR0341833

[23] VOLOSHIN, V. I. (2009). *Introduction to Graph and Hypergraph Theory.* Nova Science Publishers, Inc, New York. MR2514872

[24] WALTER, J. (1972). *Representations of Rigid Cycle Graphs.* PhD thesis, Wayne State University. MR2622336

[25] WANG, X. F. and GUO, J. H. (2008). Junction trees of general graphs. *Front. Math. China* **3** 399–413. MR2425162

[26] WANG, X. F., GUO, J. H. and HE, X. (2011). Finding the minimal set for collapsible graphical models. *Proc. Amer. Math. Soc.* **139** 361–373. MR2729097

[27] WONG, S. K. M., BUTZ, C. J. and WU, D. (2000). On the implication problem for probabilistic conditional independency. *IEEE T. Syst. Man. Cy. A* **6** 785–805.

[28] WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5** 161–215.

[29] XIANG, Y. (2002). *Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach.* Cambridge University Press. MR1922807

[30] XIE, X., GENG, Z. and ZHAO Q. (2006). Decomposition of structural learning about directed acyclic graphs. *Artif. Intell.* **170** 422–439. MR2209208

[31] XIE, X. and GENG, Z. (2008). A recursive method for structural learning of directed acyclic graphs. *J. Mach. Learn. Res.* **9** 459–483. MR2417242

[32] XU, P. F., GUO, J. H. and HE, X. (2011). An improved iterative proportional scaling procedure for Gaussian graphical models. *JCGS* **20** 417–431. MR2847802

Xiaofei Wang
School of Mathematics and Statistics
Northeast Normal University
Changchun 130024, China
E-mail address: wangxf341@nenu.edu.cn

Jianhua Guo
School of Mathematics and Statistics
Northeast Normal University
Changchun 130024, China
E-mail address: jhguo@nenu.edu.cn