# On Information-Theoretic Characterizations of Markov Random Fields and Subfields

Raymond W. Yeung[*]    Ali Al-Bashabsheh[†]    Chao Chen[‡]    Qi Chen[*]

Pierre Moulin[§]

January 18, 2018

## Abstract

Let $X_i, i \in V$ form a Markov random field (MRF) represented by an undirected graph $G = (V, E)$, and $V'$ be a subset of $V$. We determine the smallest graph that can always represent the subfield $X_i, i \in V'$ as an MRF. Based on this result, we obtain a necessary and sufficient condition for a subfield of a Markov tree to be also a Markov tree. When $G$ is a path so that $X_i, i \in V$ form a Markov chain, it is known that the $I$-Measure is always nonnegative and the information diagram assumes a very special structure [11]. We prove that Markov chain is essentially the only MRF such that the $I$-Measure is always nonnegative. By applying our characterization of the smallest graph representation of a subfield of an MRF, we develop a recursive approach for constructing information diagrams for MRFs. Our work is built on the set-theoretic characterization of an MRF in [15].

[*]Raymond Yeung and Qi Chen are with Institute of Network Coding and Department of Information Engineering, The Chinese University of Hong Kong, N.T., Hong Kong. Email: {whyeung, qichen}@ie.cuhk.edu.hk

[†]Ali Al-Bashabsheh is with the Big Data and Brain Computing Center, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing, Postcode 100191. Email: entropyali@gmail.com

[‡]Chao Chen is with School of Electronic Engineering, Xidian University, 710071, China. Email: cchen@xidian.edu.cn

[§]Pierre Moulin is with ECE Department and Coordinated Science Laboratory, University of Illinois, Urbana IL 61801. Email: moulin@ifp.uiuc.edu

# 1 Introduction

A Markov random field (MRF) is often regarded as a generalization of a one-dimensional discrete-time Markov chain in the sense that the time index for the latter is replaced by a space index for the former. Historically, the study of MRFs stems from statistical physics. The classical Ising model, which is defined on a rectangular lattice, was used to explain certain empirically observed facts about ferromagnetic materials. In statistics, the dependencies between variables in a contingency table may also be modeled as an MRF [7]. In image processing and computer vision, the dependencies between pixels or image features are also commonly modeled by MRFs [21]. MRFs have also been used in wireless and ad hoc networking [17, 20, 19]. In recent years, MRFs have been used as a model for studying social networks [22, 23] and big data [24].

The foundation of the theory of MRFs may be found in [5] or [3] (also see [6] and [13]). It was described in [5] that the theory can be generalized to the context of an arbitrary graph. In this paper, we discuss such MRFs whose random variables are discrete. Before we present their formulation, we first introduce some notations that are used throughout the paper.

In this paper, all random variables are discrete. Let $X$ be a random variable taking values in an alphabet $\mathcal{X}$. The probability distribution for $X$ is denoted as $\{p_X(x), x \in \mathcal{X}\}$, with $p_X(x) = \Pr\{X = x\}$. When there is no ambiguity, $p_X$ is abbreviated as $p$. The *support* of $X$, denoted by $\mathcal{S}_X$, is the set of all $x \in \mathcal{X}$ such that $p(x) > 0$. If $\mathcal{S}_X = \mathcal{X}$, we say that $p$ is *strictly positive*, denoted by $p > 0$. Otherwise, $p$ contains zero probability masses, and we say that $p$ is not strictly positive. Note that probability distributions with zero probability masses are in general very delicate, and they need to be handled very carefully (see Example 1 below). All the above notations naturally extend to two or more random variables.

**Proposition 1.** *For random variables $X$, $Y$, and $Z$, $X \perp\!\!\!\perp Z \mid Y$ if and only if*

$$p(x, y, z) = a(x, y)b(y, z) \tag{1}$$

*for all $x$, $y$, and $z$ such that $p(y) > 0$, where $a$ is some function of $x$ and $y$ and $b$ is some function of $y$ and $z$.*

The example below illustrates the subtlety of conditional independence when the probability distribution contains zero probability masses.

**Example 1.** *Let $p$ denote the joint distribution of three random variables $X_1, X_2$, and $X_3$. In this example, we show that*

$$\left.\begin{array}{c} X_1 \perp\!\!\!\perp X_2 \mid X_3 \\ X_1 \perp\!\!\!\perp X_3 \mid X_2 \end{array}\right\} \Rightarrow X_1 \perp\!\!\!\perp (X_2, X_3) \tag{2}$$

*holds if $p > 0$, but does not hold in general.*

*Assume that $p > 0$. Then for all $x_1, x_2$, and $x_3$, by $X_1 \perp\!\!\!\perp X_2 \mid X_3$, we have*

$$p(x_1, x_2, x_3) = \frac{p(x_1, x_3)p(x_2, x_3)}{p(x_3)}, \tag{3}$$

*and by $X_1 \perp\!\!\!\perp X_3 \mid X_2$, we have*

$$p(x_1, x_2, x_3) = \frac{p(x_1, x_2)p(x_2, x_3)}{p(x_2)}. \tag{4}$$

*Equating (3) and (4), we have*

$$p(x_1, x_3) = \frac{p(x_1, x_2)p(x_3)}{p(x_2)}.$$

*Then*

$$\begin{aligned} p(x_1) &= \sum_{x_3} \frac{p(x_1, x_2)p(x_3)}{p(x_2)} \\ &= \frac{p(x_1, x_2)}{p(x_2)}, \end{aligned}$$

*or*

$$p(x_1, x_2) = p(x_1)p(x_2).$$

*Substituting this into (4), we have*

$$p(x_1, x_2, x_3) = p(x_1)p(x_2, x_3),$$

*i.e., $X_1 \perp\!\!\!\perp (X_2, X_3)$.*

*However, (2) does not hold in general, because if $X_1 = X_2 = X_3$, we see that $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_1 \perp\!\!\!\perp X_3 \mid X_2$ but $X_1 \not\perp\!\!\!\perp (X_2, X_3)$. Note that $p$ is not strictly positive if $X_1 = X_2 = X_3$.*

We now present the formulation of an MRF defined on an arbitrary graph. Let $G = (V, E)$ be an undirected graph, where $V = \{1, 2, \cdots, n\}$ is the set of vertices and $E \subset V \times V$ is the set of edges. We assume that there is no edge in $G$ which joins a vertex to itself. For any (possibly empty) subset $U$ of $V$, denote by $G \backslash U$ the graph obtained from $G$ by removing all the vertices in $U$ and all the edges joining a vertex in $U$. Let $s(U)$ be the number of components[1] in $G \backslash U$. Denote the sets of vertices of these components by $V_1(U), V_2(U), \cdots, V_{s(U)}(U)$. If $s(U) > 1$, we say that $U$ is a cutset in $G$. Throughout this paper, whenever we remove a subset of vertices $U$ from $G$, we always assume that we also remove all the edges joining a vertex in $U$.

Consider a collection of random variables $X_i, i \in V$ whose joint distribution is specified by a probability measure $p$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$, where random variable $X_i$ is associated with vertex $i$ in graph $G$. We now define a few Markov properties for random variables $X_1, \cdots, X_n$ pertaining to a graph $G = (V, E)$:

**Definition 1** (Pairwise Markov Property). *For all distinct $i, j \in V$ such that $\{i, j\} \notin E$, $X_i$ and $X_j$ are independent conditioning on $X_{V-\{i,j\}}$.*

**Definition 2** (Local Markov Property). *For all $i \in V$, $X_i$ and $X_{V-\overline{N}(i)}$ are independent conditioning on $X_{N(i)}$, where $N(i) = \{j \in V : \{i, j\} \in E\}$ is the set of neighbors of vertex $i$ and $\overline{N}(i) = N(i) \cup \{i\}$.*

**Definition 3** (Global Markov Property). *Let $\{U, V_1, V_2\}$ be a partition of $V$ such that the sets of vertices $V_1$ and $V_2$ are disconnected in $G \backslash U$. Then the sets of random variables $X_{V_1}$ and $X_{V_2}$ are independent conditioning on $X_U$.*

**Proposition 2.** *Random variables $X_i, i \in V$ satisfy the global Markov property if and only if for all cutsets $U$ in $G$, the sets of random variables $X_{V_1(U)}, \cdots, X_{V_{s(U)}(U)}$ are mutually independent conditioning on $X_U$.*

See [15] for a proof of Proposition 2. When $U = \emptyset$, this proposition states that if the graph $G$ has more than one component, i.e., $s(\emptyset) > 1$, then the sets of random variables $X_{V_1}(\emptyset), \cdots,$

---

[1] A component of an undirected graph is a subgraph in which any two vertices are connected, and which is not connected to any additional vertices in the supergraph.

$X_{V_2}(\emptyset), \cdots, X_{V_{s(\emptyset)}}(\emptyset)$ are mutually independent. Here we regard unconditional mutual independence as a special case of conditional mutual independence.

Denote the Pairwise Markov Property, the Local Markov Property, and the Global Markov Property by (P), (L), and (G), respectively. It can readily be seen from their definitions that (G) $\Rightarrow$ (L) $\Rightarrow$ (P).

**Definition 4** (Markov Random Field). *The probability measure $p$, or equivalently, the random variables $X_i, i \in V$, are said to form an MRF represented by a graph $G = (V, E)$ if and only if the Global Markov Property is satisfied by $X_i, i \in V$.*

If $X_i, i \in V$ form an MRF represented by a graph $G$, we also say that $X_i, i \in V$ form a Markov graph $G$, $X_i, i \in V$ are represented by $G$, or $G$ is a (graph) representation for $X_i, i \in V$. When $G$ is a path,[2] we say that $X_i, i \in V$ form a Markov chain. When $G$ is a tree, we say that $X_i, i \in V$ form a Markov tree.[3] When $G$ is a cycle graph,[4] we say that $X_i, i \in V$ form a Markov ring.

In general, $X_i, i \in V$ can be represented by more than one graph. In particular, $X_i, i \in V$ are always represented by $K_n$, the complete graph with $n$ vertices. The graph $K_n$ specifies a degenerate MRF, because for every $U \subsetneq V$, $U$ is not a cutset in $K_n$. In other words, no Markov constraints are imposed on $X_i, i \in V$ by $K_n$.

Suppose the random variables $X_i, i \in V$ are represented by both $G = (V, E)$ and $G' = (V, E')$, where $E' \subsetneq E$, i.e., $G'$ is a proper subgraph of $G$. Then $G'$ imposes a larger set of Markov constraints on $X_i, i \in V$ than $G$, because a cutset in $G$ is also a cutset in $G'$ (but not vice versa). Thus we are naturally interested in the "smallest" graph (to be discussed in Section 2.4) that represents $X_i, i \in V$.

**Definition 5** (Subfield). *A subset of the random variables forming an MRF is called a subfield of the MRF.*

**Definition 6.** *A $n$-tuple $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ is called a configuration. A probability measure $p$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ is strictly positive, denoted by $p > 0$, if $p(\mathbf{x}) > 0$ for*

---

[2]A path is a graph whose vertices can be linearly ordered so that every pair of consecutive vertices forms an edge.

[3]The term "Markov tree" is also used in the number theory literature in the context of the Markov number, but it is not to be confused with the Markov tree in this paper.

[4]A cycle graph is a graph that consists of a single cycle.

*all configurations* **x**.

If $p > 0$, it can be shown that (G) = (L) = (P) (see for example [13]). In general, however, a probability measure $P$ may contain zero probability masses, i.e., $p(\mathbf{x}) = 0$ for some configuration **x**. For example, if some random variables in $X_1, X_2, \cdots, X_n$ are functions of other random variables, then $p$ is not strictly positive.

In this paper, we study the structure of MRFs by means of an information-theoretic approach. Specifically, structural properties of MRFs are obtained through the investigation of the set-theoretic structure of Shannon's information measures under the constraints imposed by the MRF. With this approach, we do not have to manipulate the underlying probability measure directly.

An identity involving only Shannon's information measures (i.e., entropy, mutual information, and their conditional versions) is referred to as an information identity. The set-theoretic structure of Shannon's information measures was first studied in [1], where it was proved that for every information identity, there is a corresponding set identity. This was further developed into the theory of $I$-Measure in [9]. Under this framework, every Shannon's information measure can formally be regarded as the value of a unique signed measure called the $I$-Measure, denoted by $\mu^*$, on a set corresponding to that Shannon's information measure. This establishes a complete set-theoretic interpretation of Shannon's information measures.

Subsequent to [9], the structure of the $I$-Measure for a Markov chain and more generally an MRF was investigated in [11] and [15], respectively. In particular, it was proved in [11] that the $I$-Measure for a Markov chain is always nonnegative, and an information diagram that displays the special structure of the $I$-Measure for a Markov chain was obtained.

The current work, consisting of the following three main results, is built on [1, 9, 11, 15]:

1. Let $X_i, i \in V$ be any set of random variables that form an MRF represented by a graph $G$, and let $X_i, i \in V'$, where $V' \subset V$, be any subfield of the MRF. We determine the smallest graph $G^*(V')$ that can always represent $X_i, i \in V'$.

2. The $I$-Measure of an MRF is always nonnegative if and only if the MRF is represented by either a path or a forest of paths.[5]

---

[5] A forest of paths is a graph with at least two components such that each component is a path.

3. We develop a recursive approach for constructing an information diagram that displays the special structure of the $I$-Measure for an MRF.

The rest of the paper is organized as follows. Section 2 contains an overview of the concepts and tools to be used in this paper. In Section 3, we define the graph $G^*(V')$ and establish that $G^*(V')$ is the smallest graph that can always represent the subfield $X_i, i \in V'$. Applying this result to Markov trees, we obtain in Section 4 a necessary and sufficient condition for a subfield of a Markov tree to form a Markov subtree. In Section 5, we establish that Markov chains are essentially the only MRFs for which the $I$-Measure is always nonnegative. In Section 6, we develop a recursive approach for constructing an information diagram that displays the special structure of the $I$-Measure for an MRF. The paper is concluded in Section 7.

## 2 Preliminaries

In this section, we introduce the notations and present the preliminaries for the rest of the paper. For a detailed discussion, we refer the readers to [18, Chapters 3 and 12] and the references therein.

### 2.1 $I$-Measure

We first give an overview of the basics of the $I$-Measure. Let $X_i, i \in V = \{1, 2, \cdots, n\}$ be jointly distributed discrete random variables, and $\tilde{X}$ be a set variable corresponding to a random variable $X$. We note that the $I$-Measure does not have to be defined in the context of an MRF, but here we use $V$ (the vertex set of a graph) as the index set of the random variables for the sake of convenience. Here we assume that $H(X_i) < \infty$ for $1 \le i \le n$, so that the $I$-Measure [9] for $p$ is well-defined.

Define the universal set $\Omega_V$ to be $\bigcup_{i \in V} \tilde{X}_i$ and let $\mathcal{F}_V$ be the $\sigma$-field generated by $\{\tilde{X}_i, i \in V\}$. The atoms of $\mathcal{F}_V$ have the form $\bigcap_{i \in V} Y_i$, where $Y_i$ is either $\tilde{X}_i$ or $\tilde{X}_i^c$. Let $\mathcal{A}_V \subset \mathcal{F}_V$ be the set of all the atoms of $\mathcal{F}_V$ except for $\bigcap_{i \in V} \tilde{X}_i^c$, which is equal to the empty set because

$$\bigcap_{i \in V} \tilde{X}_i^c = \left( \bigcup_{i \in V} \tilde{X}_i \right)^c = (\Omega_V)^c = \emptyset.$$

Note that $|\mathcal{A}_V| = 2^n - 1$. In the rest of the paper, when we refer to an atom of $\mathcal{F}_V$, we always mean an atom in $\mathcal{A}_V$ unless otherwise specified.

6

To simplify notation, we will use $X_U$ to denote $(X_i, i \in U)$ and $\tilde{X}_U$ to denote $\bigcup_{i \in U} \tilde{X}_i$ for any $U \subset V$. We will not distinguish between $i$ and the singleton containing $i$. It was shown in [9] that there exists a unique *signed* measure $\mu^*$ on $\mathcal{F}_V$ which is consistent with all Shannon's information measures via the following substitution of symbols:

$$
\begin{aligned}
H/I &\rightarrow \mu^* \\
, &\rightarrow \cup \\
; &\rightarrow \cap \\
| &\rightarrow -
\end{aligned}
$$

where "$-$" denotes the set difference. For example,

$$\mu^*((\tilde{X}_1 \cup \tilde{X}_2) \cap \tilde{X}_3 - \tilde{X}_4) = I(X_1, X_2; X_3 | X_4).$$

For all $A \in \mathcal{A}_V$, $\mu^*(A)$ is a linear combination of $H(X_B)$ for nonempty subsets $B$ of $V$.

Note that $\mu^*$ in general is not nonnegative. However, if $X_i, i \in V$ form a Markov chain, then $\mu^*$ is always nonnegative [11]. See Section 5 for further discussions.

## 2.2 Full Conditional Mutual Independency

**Definition 7.** *Let $\{T, Q_1, Q_2, \cdots, Q_k\}$ be a partition of $V'$, where $k \geq 2$ and $V' \subset V$. The tuple $K = (T; Q_i, 1 \leq i \leq k)$ defines the following conditional mutual independency (CMI) on $X_i, i \in V$:*

$$X_{Q_1}, X_{Q_2}, \cdots, X_{Q_k} \text{ are mutually independent conditioning on } X_T.$$

*If $V' = V$, $K$ is called a full conditional mutual independency (FCMI).*

**Example 2.** *For $n = 6$, $K = (\{4\}; \{1,3\}, \{2,5\}, \{6\})$ defines the FCMI*

$$(X_1, X_3), (X_2, X_5), X_6 \text{ are mutually independent conditioning on } X_4.$$

*However, for $n = 7$, $K$ is not an FCMI because $\{\{4\}, \{1,3\}, \{2,5\}, \{6\}\}$ is not a partition of $\{1, 2, \cdots, 7\}$.*

**Definition 8.** *Let $K = (T; Q_i, 1 \leq i \leq k)$ be an FCMI on $X_i, i \in V$. The image of $K$, denoted by $Im(K)$, is the set of atoms of $\mathcal{F}_V$ of the form*

$$\left( \bigcap_{i=1}^{k} \bigcap_{j \in W_i} \tilde{X}_j \right) - \tilde{X}_{T \cup (\bigcup_{i=1}^{k} (Q_i - W_i))} \tag{5}$$

*where $W_i \subset Q_i$, $1 \leq i \leq k$, and there exist at least two $i$ such that $W_i \neq \emptyset$.*

The following proposition gives a more explicit expression for $Im(K)$. The proof is elementary and so is omitted.

**Proposition 3.** *Let $K = (T; Q_i, 1 \leq i \leq k)$ be an FCMI on $X_i, i \in V$. Then*

$$Im(K) = \left\{ A \in \mathcal{A}_n : A \subset \bigcup_{1 \leq i < j \leq k} (\tilde{X}_{Q_i} \cap \tilde{X}_{Q_j} - \tilde{X}_T) \right\}.$$

In the rest of paper, we denote the atom $\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3^c$ of $\mathcal{F}_{\{1,2,3\}}$ by $12\bar{3}$, etc.

**Example 3.** *Let $n = 3$ and consider the FCMI $K = (\emptyset; \{1\}, \{2\}, \{3\})$. Then $Im(K)$ is the set containing all the atoms in $(\tilde{X}_1 \cap \tilde{X}_2) \cup (\tilde{X}_1 \cap \tilde{X}_3) \cup (\tilde{X}_2 \cap \tilde{X}_3)$, as given by Proposition 3. This is illustrated in Fig. 1. Equivalently, the atoms in $Im(K)$ are $12\bar{3}$, $1\bar{2}3$, $\bar{1}23$, and $123$, as given by Definition 8. For example, for the atom $12\bar{3}$, we have $W_1 = \{1\}$, $W_2 = \{2\}$, and $W_3 = \emptyset$, so there are at least two $i$ such that $W_i \neq \emptyset$.*

The following theorem from [18] will be useful for proving some of the results in this work.

**Theorem 1.** *Let $K$ be an FCMI on $X_i, i \in V$. Then $K$ holds if and only if $\mu^*(A) = 0$ for all $A \in Im(K)$.*

Thus the effect of an FCMI $K$ on the joint probability distribution of $X_i, i \in V$ is completely characterized by $Im(K)$. We remark that if $\{T, Q_1, Q_2, \cdots, Q_k\}$ is a partition of $V'$ where $V' \subsetneq V$, then $K = (T; Q_1, Q_2, \cdots, Q_k)$ holds if and only if $\mu^*$ vanishes on all the sets prescribed in (5), although these sets are no longer atoms of $\mathcal{F}_V$.

**Example 4.** *Following Example 3, the random variables $X_1, X_2$, and $X_3$ are mutually independent if and only if $\mu^*$ vanishes on the atoms $12\bar{3}$, $1\bar{2}3$, $\bar{1}23$, and $123$.*

Let $A = \bigcap_{i \in V} \tilde{Y}_i$ be a nonempty atom of $\mathcal{F}_V$. Define the set

$$U_A = \{i \in V : \tilde{Y}_i = \tilde{X}_i^c\}. \tag{6}$$

Note that $A$ is uniquely specified by $U_A$ because

$$A = \left(\bigcap_{i \in V - U_A} \tilde{X}_i\right) \cap \left(\bigcap_{i \in U_A} \tilde{X}_i^c\right) = \left(\bigcap_{i \in V - U_A} \tilde{X}_i\right) \cap \left(\bigcup_{i \in U_A} \tilde{X}_i\right)^c = \left(\bigcap_{i \in V - U_A} \tilde{X}_i\right) - \tilde{X}_{U_A}.$$

Also note that in the definition of $U_A$, its dependence on $V$ is implicit, and what the set $V$ is should be clear from the context.

Define $w(A) = n - |U_A|$ as the *weight* of the atom $A$, the number of $\tilde{X}_i$ in $A$ which are not complemented. We now show that an FCMI $K = (T; Q_i, 1 \le i \le k)$ is uniquely specified by $Im(K)$. First, by letting $W_i = Q_i$ for $1 \le i \le k$ in (5), we see that the atom

$$\left(\bigcap_{j \in \bigcup_{i=1}^k Q_i} \tilde{X}_j\right) - \tilde{X}_T$$

is in $Im(K)$, and it is the unique atom in $Im(K)$ with the largest weight. From this atom, $T$ can be determined. To determine $Q_i, 1 \le i \le k$, we define a relation $q$ on $T^c = V - T$ as follows. For $l, l' \in T^c$, $(l, l')$ is in $q$ if and only if one of the following is satisified:

   i) $l = l'$;

   ii) $l \ne l'$ and the atom

$$\tilde{X}_l \cap \tilde{X}_{l'} \cap \left(\bigcap_{j \in V - \{l, l'\}} \tilde{X}_j^c\right) \tag{7}$$

     is not in $Im(K)$.

The idea of ii) is that $(l, l')$ is in $q$ if and only if $l, l' \in Q_i$ for some $1 \le i \le k$, which can be seen as follows. If $l, l' \in Q_i$ for some $i$, then the atom in (7) is not in $Im(K)$ by Definition 8 because $\{l, l'\} \subset W_i$ and so $W_i \ne \emptyset$ but $W_j = \emptyset$ for all $j \ne i$ (an atom in $Im(K)$ has at least two $i$ such that $W_i \ne \emptyset$). On the other hand, if $l \in Q_i$ and $l' \in Q_{i'}$ where $i \ne i'$, then by letting $W_i = \{l\}$ and $W_{i'} = \{l'\}$, we see that the atom in (7) is in $Im(K)$.

Then $q$ is reflexive by i), and is symmetric because the definition of $q$ is symmetrical in $l$ and $l'$. Moreover, $q$ is transitive from the discussion above because if $l, l' \in Q_i$ for some $1 \leq i \leq k$ and $l', l'' \in Q_{i'}$ for some $1 \leq i' \leq k$, then $i = i'$ and $l, l'' \in Q_i$. In other words, $q$ is an *equivalence relation* that partitions $T^c$ into $\{Q_i, 1 \leq i \leq k\}$. Therefore, $K$ can be recovered from $Im(K)$, and so it is uniquely specified by $Im(K)$.

Let $\Pi = \{K_l, 1 \leq l \leq m\}$ be a collection of FCMIs on $X_i, i \in V$, and define

$$Im(\Pi) = \bigcup_{l=1}^{k} Im(K_l).$$

Since $\Pi$ holds if and only if $K_l$ holds for all $l$, it follows from Theorem 1 that $\Pi$ holds if and only if $\mu^*(A) = 0$ for all $A \in Im(\Pi)$. Thus the effect of a collection $\Pi$ of FCMIs on the joint probability distribution of $X_i, i \in V$ is completely characterized by $Im(\Pi)$.

Consequently, for two collections $\Pi_1$ and $\Pi_2$ of FCMIs, $\Pi_1 \Rightarrow \Pi_2$ if and only $Im(\Pi_1) \supset Im(\Pi_2)$, and $\Pi_1 = \Pi_2$ if and only $Im(\Pi_1) = Im(\Pi_2)$.

One can interpret $Im(K)$ as the "footprint" of an FCMI $K$. Then the footprint of a collection $\Pi$ of FCMIs is simply the union of the footprints of the individual FCMIs in $\Pi$. However, two different collections of FCMIs may have the same footprints, as shown in the next example. Thus unlike an FCMI, a collection of FCMIs is in general not uniquely specified by its image.

**Example 5.** *Let $n = 3$. Let $\Pi_1 = \{K_1\}$ and $\Pi_2 = \{K_2, K_3\}$, where*

$$
\begin{aligned}
K_1 &= (\emptyset; \{1\}, \{2\}, \{3\}) \\
K_2 &= (\emptyset; \{1, 2\}, \{3\}) \\
K_3 &= (\{3\}; \{1\}, \{2\}).
\end{aligned}
$$

*Then $\Pi_1 \neq \Pi_2$ but $Im(\Pi_1) = Im(\Pi_2)$.*

It was shown in [10, 12] that full conditional (mutual) independence are axiomatizable. This can be regarded as an alternative characterization of FCMIs, which however is not in closed form.

## 2.3 Markov Random Field

In the definition of an MRF, each cutset $U$ in $G$ specifies an FCMI on $X_1, X_2, \cdots, X_n$, denoted by $[U]$. Formally,

$$[U]: \ X_{V_1(U)}, \cdots, X_{V_{s(U)}(U)} \text{ are mutually independent}$$

$$\text{conditioning on } X_U.$$

Then in light of (6), for $A \in \mathcal{A}_n$ such that $s(U_A) > 1$, $[U_A]$ is the FCMI induced by the cutset $U_A$. It follows that $X_1, X_2, \cdots, X_n$ form a Markov graph $G$ if and only if

$$\bigwedge_{A \in \mathcal{A}_n : s(U_A) > 1} [U_A] \overset{\triangle}{=} [U_G], \tag{8}$$

where '$\wedge$' denotes 'logical AND'. This is the collection of FCMIs induced by graph $G$.

**Definition 9.** *Let $G = (V, E)$ be a graph. For an atom $A$ of $\mathcal{F}_V$, if $s(U_A) = 1$, i.e., $G \backslash U_A$ is connected, then $A$ is a Type I atom of $G$, otherwise, i.e., $s(U_A) > 1$, $A$ is a Type II atom of $G$. The sets of all Type I and Type II atoms of $G$ are denoted by $\mathcal{T}_{\mathrm{I}}(G)$ and $\mathcal{T}_{\mathrm{II}}(G)$, respectively.*

**Definition 10.** *For a graph $G = (V, E)$, the image of $G$ is defined by*

$$Im(G) = Im([U_G]).$$

**Theorem 2.** *(cf. [18, Theorem 12.25]) $Im(G) = \mathcal{T}_{\mathrm{II}}(G)$.*

The above theorem gives a precise characterization of $Im(G)$. It follows from the discussion in Section 2.2 that $X_1, X_2, \cdots, X_n$ form a Markov graph $G$ if and only if $\mu^*(A) = 0$ for all $A \in \mathcal{T}_{\mathrm{II}}(G)$, i.e., $\mu^*$ vanishes on all the Type II atoms of $G$.

**Example 6.** *For the cycle graph $G$ in Fig. 2, $\mathcal{T}_{\mathrm{II}}(G) = \{1\bar{2}3\bar{4}, \bar{1}2\bar{3}4\}$. Random variables $X_1, X_2, X_3$ and $X_4$ are represented by $G$ if and only if $\mu^*(1\bar{2}3\bar{4}) = \mu^*(\bar{1}2\bar{3}4) = 0$.*

A graph $G = (V, E)$ and the collection $[U_G]$ of FCMIs it induces uniquely specify each other, because for distinct $u, v \in V$, $\{u, v\} \in E$ if and only if the FCMI $(V - \{u, v\}; \{u\}, \{v\})$ is not in $[U_G]$. This can be seen as follows. If $\{u, v\} \in E$, then $V - \{u, v\}$ is not a cutset in $G$, and so

11

$(V - \{u, v\}; \{u\}, \{v\}) \notin [U_G]$. On the other hand, if $(V - \{u, v\}; \{u\}, \{v\}) \notin [U_G]$, then $V - \{u, v\}$ is not a cutset in $G$, which implies $\{u, v\} \in E$.

Although a collection of FCMIs is in general not uniquely determined by its image (cf. Example 5), the following proposition asserts that a graph $G$ (and hence $[U_G]$) is uniquely determined by its image $Im(G)$.

**Proposition 4.** *For a graph $G = (V, E)$, $\{u, v\} \in E$ if and only if the atom*

$$\tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V - \{u,v\}} \tag{9}$$

*is not in $Im(G)$.*

**Proof** Denote the atom in (9) by $A$. If $\{u, v\} \notin E$, then $G$ induces the FCMI $K = (V - \{u, v\}; \{u\}, \{v\})$. Obviously, $Im(K) = \{A\}$, and hence $A \in Im(K) \subset Im(G)$.

To prove the converse, assume that atom A is in $Im(G)$, and specifically in some $Im([U_{A'}])$ such that $s(U_{A'}) > 1$. It follows from Definition 8 that in order for $A$ to be in $Im([U_{A'}])$, it is necessary for $u$ and $v$ to be in different sets in $V_1(U_{A'}), V_2(U_{A'}), \cdots, V_{s(U_{A'})}(U_{A'})$. This implies that $V - \{u, v\}$ is a cutset in $G$, and hence $\{u, v\} \notin E$. $\qquad\square$

With this proposition, a graph $G$ can be recovered from $Im(G)$ as follows. Start with the complete graph $K_n$. If there exists an atom in $Im(G)$ as prescribed by (9) for some distinct $u, v \in V$, then remove edge $\{u, v\}$ from the graph. Repeat this step until no more edges can be removed. Note that this algorithm produces a unique graph, i.e., $G$. As a corollary, the uniqueness of the Markov graph induced by $Im(G)$ is proved, i.e., for two graphs $G = (V, E)$ and $G' = (V, E')$ where $E \neq E'$, $Im(G) \neq Im(G')$.

## 2.4   Smallest Graph Representation

As discussed in Section 1, we are interested in the "smallest" graph that can represent a given set of random variables $X_i, i \in V$. To fix ideas, we first give a formal definition of this notion.

**Definition 11.** *A graph $G = (V, E)$ is the smallest graph representation for a set of random variables $X_i, i \in V$ if $G$ is a representation for $X_i, i \in V$ and is a subgraph of any representation $G'$ for $X_i, i \in V$.*

We know from Section 2.3 that a graph $G = (V, E)$ can represent $X_i, i \in V$ if and only if $\mu^*$ vanishes on all the atoms in $Im(G)$. Note that if $G$ is a subgraph of $G'$, then a cutset in $G'$ is also a cutset in $G$. It follows that $[U_{G'}]$ is a "subset" of $[U_G]$, and hence $Im(G') \subset Im(G)$.

For the given set of random variables $X_i, i \in V$, let $\mathcal{A}_{\mathrm{II}}$ be the set of nonempty atoms of $\mathcal{F}_V$ on which $\mu^*$ vanishes. Following the last paragraph, if $G$ is the smallest representation for $X_i, i \in V$, then $Im(G) \subset \mathcal{A}_{\mathrm{II}}$ and $Im(G') \subset Im(G)$ for any representation $G'$ for $X_i, i \in V$. The next theorem gives a characterization of such a graph if it exists.

**Theorem 3.** *For a given set of random variables $X_i, i \in V$, let $\mathcal{A}_{\mathrm{II}}$ be the set of nonempty atoms of $\mathcal{F}_V$ on which $\mu^*$ vanishes. Let $\hat{G} = (V, \hat{E})$ be such that $\{u, v\} \in \hat{E}$ if and only if the atom in (9) is not in $\mathcal{A}_{\mathrm{II}}$. Then if the smallest graph representation for $X_i, i \in V$ exists, it is equal to $\hat{G}$.*

We first prove the following two lemmas.

**Lemma 1.** *Every graph that can represent $X_i, i \in V$ contains $\hat{G}$ as a subgraph.*

**Proof** Let $G' = (V, E')$ be any graph that can represent $X_i, i \in V$. Consider any edge $\{u, v\}$ in $\hat{G}$, i.e., $\{u, v\} \in \hat{E}$. By construction, the atom in (9) is not in $\mathcal{A}_{\mathrm{II}}$. Then $\{u, v\} \in E'$, otherwise the FCMI $(V - \{u, v\}; \{u\}, \{v\})$ holds, i.e.,

$$I(X_u; X_v | X_{V-\{u,v\}}) = \mu^*(\tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V-\{u,v\}}) = 0,$$

which is a contradiction because the atom in (9) is not in $\mathcal{A}_{\mathrm{II}}$. Thus if $G'$ can represent $X_i, i \in V$, then $G'$ contains $\hat{G}$ as a subgraph. $\qquad \square$

**Lemma 2.** *If $\{u, v\}$ is an edge in every graph that can represent $X_i, i \in V$, then $\{u, v\}$ is an edge in $\hat{G}$.*

**Proof** Let $\{u, v\}$ be an edge in every graph that can represent $X_i, i \in V$. If a graph does not contain $\{u, v\}$, then it cannot represent $X_i, i \in V$. In particular, the graph $K_n \backslash \{u, v\}$ obtained by removing $\{u, v\}$ from the complete graph $K_n$ cannot represent $X_i, i \in V$. Since the only FCMI imposed by $K_n \backslash \{u, v\}$ is $[\{u, v\}]$ (i.e., $X_u$ and $X_v$ are independent conditioning on $X_{V-\{u,v\}}$), this means that $X_u$ and $X_v$ are not independent conditioning on $X_{V-\{u,v\}}$, or $\mu^*(\tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V-\{u,v\}}) > 0$. In other words, the atom $\tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V-\{u,v\}}$ is not in $\mathcal{A}_{\mathrm{II}}$, which implies that $\{u, v\}$ is an edge in $\hat{G}$. $\qquad \square$

**Proof of Theorem 3** Assume the smallest graph representation for $X_i, i \in V$ exists and let it be $\tilde{G}$. By Lemma 1, $\hat{G}$ is a subgraph of $\tilde{G}$. On the other hand, since $\tilde{G}$ is a subgraph of every graph that can represent $X_i, i \in V$, Lemma 2 implies that $\tilde{G}$ is a subgraph of $\hat{G}$. Hence, $\tilde{G} = \hat{G}$. $\qquad\square$

**Corollary 1.** *The smallest graph representation for $X_i, i \in V$ exists if and only if $\hat{G}$ is a representation for $X_i, i \in V$.*

**Proof** Assume that the smallest graph representation for $X_i, i \in V$ exists. By Theorem 3, it is equal to $\hat{G}$, and so $\hat{G}$ is a representation for $X_i, i \in V$. Conversely, if $\hat{G}$ is a representation for $X_i, i \in V$, then by Lemma 1, it is the smallest representation for $X_i, i \in V$. $\qquad\square$

**Example 7.** *Let $n = 3$ and consider $\mu^*$ such that*

$$\mathcal{A}_{\text{II}} = \{12\bar{3}, 1\bar{2}3\}. \tag{10}$$

*Accordingly, the graph $\hat{G}$ defined in Proposition 3 is illustrated in Fig. 3. However,*

$$Im(\hat{G}) = \{12\bar{3}, 1\bar{2}3, 123\} \not\subset \mathcal{A}_{\text{II}},$$

*i.e., $\hat{G}$ cannot represent $X_1, X_2$, and $X_3$. Then by Corollary 1, there does not exist a smallest graph representation of $X_1, X_2$, and $X_3$.*

The above example shows that the smallest graph representation may not exist for a given set of random variables. However, if $\mathcal{A}_{\text{II}} = Im(G)$ for some graph $G$, then $G$ is in fact the smallest graph representation for $X_i, i \in V$. This is proved in the next proposition.

**Proposition 5.** *If $\mathcal{A}_{\text{II}} = Im(G)$ for some graph $G$, then $G$ is the smallest graph representation for $X_i, i \in V$.*

**Proof** We see that a graph $G$ can be recovered from its image $Im(G)$ using the algorithm described at the end of Section 2.3, and in fact $G = \hat{G}$. Therefore, $Im(\hat{G}) = Im(G) = \mathcal{A}_{\text{II}}$, which implies that $Im(\hat{G}) \subset \mathcal{A}_{\text{II}}$. Hence, $\hat{G}$ is a graph representation for $X_i, i \in V$. It then follows from Theorem 3 that $\hat{G}$, i.e., $G$, is the smallest graph representation for $X_i, i \in V$. $\qquad\square$

To our knowledge, Corollary 1 is new. A related result can be found in [8], where it was proved that if the underlying probability measure $p$ is strictly positive, then the smallest graph representation for $X_i, i \in V$ always exists and is equal to $\hat{G}$.

**Example 8.** *In Example 7, the constraint (10) is equivalent to $X_1 \perp\!\!\!\perp X_2 \mid X_3$ and $X_1 \perp\!\!\!\perp X_3 \mid X_2$, while $[U_{\hat{G}}] = [U_\emptyset, U_{\{2\}}, U_{\{3\}}]$ consists of the FCMIs*

$$X_1 \perp\!\!\!\perp (X_2, X_3), \ X_1 \perp\!\!\!\perp X_3 \mid X_2, \ X_1 \perp\!\!\!\perp X_2 \mid X_3.$$

*We have shown in Example 1 that*

$$\left. \begin{array}{l} X_1 \perp\!\!\!\perp X_2 \mid X_3 \\ X_1 \perp\!\!\!\perp X_3 \mid X_2 \end{array} \right\} \ \Rightarrow \ X_1 \perp\!\!\!\perp (X_2, X_3),$$

*holds if the underlying probability distribution $p$ is strictly positive, or $p > 0$, but does not hold in general. This means that if $p > 0$, then $\hat{G}$ represents $X_1, X_2$, and $X_3$, but in general it does not. These conclusions are consistent with the result in [8] and the discussion in Example 7, respectively.*

## 3 Subfield of a Markov Random Field

Let $X_i, i \in V$ form an MRF represented by some graph $G = (V, E)$. Note that such a graph $G$ can always be found, because $K_n$ is always a representation of $X_i, i \in V$. Let $V'$ be a subset of $V$. In this section, we seek the smallest graph that can always represent the subfield $X_i, i \in V'$.

**Definition 12.** *Let $G = (V, E)$ and $G' = (V', E')$ where $V' \subset V$. If $[U_G] \Rightarrow [U_{G'}]$, we write $G \Rightarrow G'$.*

Let $X_i, i \in V$ form an MRF represented by a graph $G$. Following the definition above, if $G \Rightarrow G'$, then $X_i, i \in V'$ form an MRF represented by $G'$.

**Definition 13.** *Let $G = (V, E)$, and let $V' \subset V$. Let $G^*(V') = (V', E')$ be such that for distinct $u, v \in V'$, $\{u, v\} \in E'$ if and only if there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V'$.*

Obviously, $G^*(V) = G$. We will prove in Theorem 8, the main theorem of this section, that $G^*(V')$ is the smallest $G'$ such that $G \Rightarrow G'$.

**Example 9.** *Consider an MRF represented by the graph G in Fig. 4, which indeed is a Markov chain. Let $V' = \{1, 3, 5, 6\}$. Then $G^*(V')$ is illustrated as the overlay graph in grey.*

**Example 10.** *Consider an MRF represented by the more elaborate graph G in Fig. 5. Let $V' = \{1, 2, 5, 6, 8, 9\}$. Then $G^*(V')$ is illustrated as the overlay graph in grey.*

Consider $V'' \subset V' \subset V$. The next proposition asserts that $G^*(V'')$ can be obtained in two steps. First obtain $G^*(V')$ from $G$ by applying Definition 13. Then obtain $G^*(V'')$ from $G^*(V')$ by applying Definition 13 again with $G^*(V')$ in place of $G$.

**Proposition 6.** *Let $G = (V, E)$ and $V'' \subset V' \subset V$. Then $G^*(V'') = (G^*(V'))^*(V'')$.*

**Proof** See Appendix A.

Consider $G \backslash (V - V') = (V', E'')$, where

$$E'' = \{\{v, w\} : v, w \in V' \text{ and } \{v, w\} \in E\}.$$

For distinct $v, w \in V'$, if $\{v, w\} \in E''$, then $\{v, w\} \in E'$ by the definition of $G^*(V')$. In other words, $G^*(V')$ always contains $G \backslash (V - V')$ as a subgraph. However, $G^*(V') \neq G \backslash (V - V')$ in general. In other words, $G^*(V')$ is not necessarily a subgraph of $G$. The following proposition gives the condition for $G^*(V')$ to be exactly equal to $G \backslash (V - V')$.

**Proposition 7.** *Let $G = (V, E)$, and let $V' \subset V$. Let $\rho(V')$ be the set of elements of $V'$ such that some of their neighbors are in $V - V'$, i.e.,*

$$\rho(V') = \{v \in V' : \{u, v\} \in E \text{ for some } u \in V - V'\}. \tag{11}$$

*Then $G^*(V') = G \backslash (V - V')$ if and only if for distinct $v, w \in \rho(V')$, if $\{v, w\}$ is not an edge in $G \backslash (V - V')$, then there exists no path between $v$ and $w$ in $G$ on which all the vertices other than $v$ and $w$ are in $V - V'$.*

**Proof** Note that $G^*(V') = G \backslash (V - V')$ is equivalent to $E' = E''$. We already have proved that $E'' \subset E'$ always holds, so we only need to prove that the condition in the proposition for $G^*(V') = G \backslash (V - V')$ is necessary and sufficient for $E' \subset E''$.

For any distinct $v, w, \in V'$, consider two cases. If either $v$ or $w$ is not in $\rho(V')$, then $\{v, w\} \in E'$ implies $\{v, w\} \in E''$. If both $v$ and $w$ are in $\rho(V')$, then the condition in the proposition for $G^*(V') = G\backslash(V - V')$ is necessary and sufficient for $\{v, w\} \in E'$ to imply $\{v, w\} \in E''$. The proposition is proved. $\qquad\square$

**Example 11.** *Consider the graph G in Fig. 6 and let $V' = \{2, 3, 4\}$. Here $\rho(V') = V'$ because each vertex in $V'$ is connected to some vertex in $V - V'$. Now $\{2, 4\}$ is the only pair of vertices that is not an edge in $G\backslash(V - V')$. Since there exists no path between vertices 2 and 4 on which all the vertices other than 2 and 4 are in $V - V' = \{1, 5\}$, by Proposition 3, $G^*(V') = G\backslash(V - V')$, which is illustrated as the overlay graph in grey.*

**Corollary 2.** *Let $V = \{1, 2, \cdots, n\}$ and $V' = V - \{n\}$, where $n \geq 2$. Let $X_i, i \in V$ be represented by a graph $G = (V, E)$ such that $\{n - 1, n\} \in E$ and $n - 1$ is the only neighbor of n. Then $X_i, i \in V'$ is represented by $G\backslash\{n\}$.*

**Proof** This is a special case of Proposition 7 with $\rho(V') = \{n - 1\}$. $\qquad\square$

As discussed above, $G^*(V')$ always contains $G\backslash(V - V')$ as a subgraph. The next theorem gives an alternative characterization of $G^*(V')$ that describes the relation between $G^*(V')$ and $G\backslash(V - V')$ more explicitly. For $U \subset V$, let

$$\phi(U) = \{v \in V - U : \{v, w\} \in E \text{ for some } w \in U\}$$

be the set of neighbors of $U$ in graph $G$,[6] and

$$\kappa(U) = \{\{u, v\} : u, v \in U\}$$

be the set of edges of the clique formed by the vertices in $U$.

**Theorem 4.** *Let $G = (V, E)$. For $V' \subset V$, let $G^*(V') = (V', E')$ and $G\backslash(V - V') = (V', E'')$. Then*

$$E' = E'' \cup \bigcup_{i=1}^{s(V')} \kappa(\phi(V_i(V'))), \qquad (12)$$

*where $V_1(V'), V_2(V'), \cdots, V_{s(V')}(V')$ are the components of $G\backslash V'$.*

---

[6]Note that $\phi(U) = \rho(V - U)$, where $\rho$ is defined in (11).

**Proof** To facilitate our discussion, let $\tilde{E}$ denote the set on the right hand side of (12). We first prove that $E' \subset \tilde{E}$. By Definition 13, if $\{u, v\} \in E'$, then there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V'$. Denote this set of vertices in $V - V'$ by $S'$. If $S' = \emptyset$, then we have $\{u, v\} \in E''$. Otherwise, since the vertices in $S'$ are connected in $G \backslash V'$, $S'$ is a subset of $V_i(V')$ for some $1 \le i \le s(V')$. As such, $u, v \in \phi(V_i(V'))$ and hence $\{u, v\} \in \kappa(\phi(V_i(V')))$. This completes the proof for $E' \subset \tilde{E}$.

It remains to prove that $\tilde{E} \subset E'$. Let $\{u, v\} \in \tilde{E}$. If $\{u, v\} \in E''$, then $\{u, v\} \in E'$ because $E'' \subset E'$ as discussed. If $\{u, v\} \in \kappa(\phi(V_i(V')))$ for for some $1 \le i \le s(V')$, then $u, v \in \phi(V_i(V'))$, i.e., there exists $u', v' \in V_i(V')$ ($u'$ and $v'$ are not necessarily distinct) such that $\{u, u'\}, \{v, v'\} \in E$. Since $u'$ and $v'$ are in the same component of $G \backslash V'$, namely $V_i(V')$, they are connected and it follows that there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V'$. Therefore, $\{u, v\} \in E'$ and we conclude that $\tilde{E} \subset E'$. The theorem is proved. $\qquad \square$

**Corollary 3.** *In Theorem 4, if $V' = V - \{n\}$, then*

$$E' = E'' \cup \kappa(\phi(\{n\})).$$

**Proof** If suffice to observe that $\{n\}$ forms the only component of $G \backslash V'$. $\qquad \square$

**Example 12.** *Refer to Example 10 and Fig. 5. Here $V - V' = \{3, 4, 7\}$. The components of $G \backslash V'$ are $\{3, 4\}$ and $\{7\}$, and $\phi(\{3, 4\}) = \{1, 2, 5, 6\}$ and $\phi(\{7\}) = \{2, 5, 8, 9\}$. Then*

$$E' = E'' \cup \kappa(\{1, 2, 5, 6\}) \cup \kappa(\{2, 5, 8, 9\}).$$

**Theorem 5.** *If $G \Rightarrow G' = (V', E')$, then $\{u, v\} \in E'$ if there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V'$.*

**Proof** Consider distinct $u, v \in V'$ such that there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V'$. Denote this set of vertices in $V - V'$ by $S'$. Consider

$$\tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V'-\{u,v\}} = \bigcup_{S \subset V-V'} \left( \tilde{X}_u \cap \tilde{X}_v \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{V-S-\{u,v\}} \right). \tag{13}$$

18

Since $S' \subset V - V'$, we see that

$$A' = \tilde{X}_u \cap \tilde{X}_v \cap \left( \bigcap_{t \in S'} \tilde{X}_t \right) - \tilde{X}_{V - S' - \{u,v\}}$$

is one of the atoms in the union in (13). Note that $s(U_{A'}) = 1$ because $u$, $v$, and the vertices in $S'$ form a path in $G$. Thus $A'$ is a Type I atom for $G$.

Now construct $X_i, i \in V$ by letting

$$X_i = \begin{cases} Z & \text{if } i \in S' \cup \{u, v\} \\ \text{constant} & \text{otherwise,} \end{cases}$$

where $Z$ is a random variable such that $0 < H(Z) < \infty$. Then by the proof of Theorem 3.11 in [18], for all $A \in \mathcal{A}_V$,

$$\mu^*(A) = \begin{cases} H(Z) & \text{if } A = A' \\ 0 & \text{otherwise.} \end{cases}$$

Now for $X_i, i \in V$ so constructed, $\mu^*$ vanishes on all the Type II atoms of $G$ because $A'$, the only atom on which $\mu^*$ does not vanish, is a Type I atom. Then from the discussion following Theorem 2, we see that $X_i, i \in V$ satisfy $[U_G]$. On the other hand, in light of (13), we have

$$\begin{aligned} \mu^* \left( \tilde{X}_u \cap \tilde{X}_v - \tilde{X}_{V' - \{u,v\}} \right) &= \sum_{S \subset V - V'} \mu^* \left( \tilde{X}_u \cap \tilde{X}_v \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{V - S - \{u,v\}} \right) \\ &= H(Z) \\ &> 0, \end{aligned}$$

i.e., $X_u$ and $X_v$ are not independent conditioning on $X_{V' - \{u,v\}}$. Hence, for any $G' = (V', E')$, if $G \Rightarrow G'$, then $V' - \{u, v\}$ is not a cutset in $G'$, which implies that $\{u, v\} \in E'$. The theorem is proved. $\square$

The next theorem is a rephrase of Theorem 5 in light of the definition of $G^*(V')$ (Definition 13).

**Theorem 6.** *If $G \Rightarrow G'$, then $G'$ contains $G^*(V')$ as a subgraph.*

**Theorem 7.** $G \Rightarrow G^*(V')$.

19

**Proof** Let $X_i, i \in V$ be any set of random variables which satisfy $[U_G]$. We need to prove that $X_i, i \in V'$ satisfy $[U_{G^*(V')}]$. For a fixed cutset $T \subset V'$ in $G^*(V')$, let $k$ be the number of components in $G^*(V') \backslash T$ and denote these components by $Q_1, Q_2, \cdots, Q_k$. To prove that $X_i, i \in V'$ satisfy $[U_{G^*(V')}]$, it suffices to prove that for every cutset $T$ in $G^*(V')$, $X_{Q_1}, X_{Q_2}, \cdots, X_{Q_k}$ are mutually independent conditioning on $X_T$.

Note that $\{T, Q_1, Q_2, \cdots, Q_k\}$ is a partition of $(T \cup (\bigcup_i Q_i)) \subsetneq V$. Following the discussion immediately after Theorem 1, we see that it suffices to prove that $\mu^*$ vanishes on the sets prescribed in (5). The atoms of $\mathcal{F}_V$ contained in a set prescribed in (5) have the form

$$\left( \bigcap_{i=1}^{k} \bigcap_{j \in W_i} \tilde{X}_j \right) \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S)}, \tag{14}$$

where $S \subset V - V'$, $W_i \subset Q_i$, $1 \le i \le k$, and there exist at least two $i$ such that $W_i \ne \emptyset$.

We will prove that every atom prescribed in (14) is a Type II atom of $G$. Since $X_i, i \in V$ satisfy $[U_G]$, $\mu^*$ vanishes on these atoms. It then follows that

$$
\begin{aligned}
&\mu^* \left( \left( \bigcap_{i=1}^{k} \bigcap_{j \in W_i} \tilde{X}_j \right) - \tilde{X}_{T \cup (\bigcup_{i=1}^{k}(Q_i - W_i))} \right) \\
&= \sum_{S \subset V - V'} \mu^* \left( \left( \bigcap_{i=1}^{k} \bigcap_{j \in W_i} \tilde{X}_j \right) \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S)} \right) \\
&= \sum_{S \subset V - V'} 0 \\
&= 0,
\end{aligned}
$$

i.e., $\mu^*$ vanishes on the sets prescribed in (5), as is to be proved.

To prove that the atom in (14) is a Type II atom of $G$, we need to show that $(T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$ is a cutset in $G$. Now in (14), let $1 \le i' < i'' \le k$ be such that $W_{i'} \subset Q_{i'}$ and $W_{i''} \subset Q_{i''}$ are nonempty, and let $u \in W_{i'}$ and $v \in W_{i''}$. We claim that $u$ and $v$ are disconnected in $G \backslash (T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$. Assume the contrary is true, i.e., there exists a path between $u$ and $v$ in $G \backslash (T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$. First of all, both $u$ and $v$ are in $V' - T$ and they belong to different components in $G^*(V') \backslash T$. Since $T \subset V' \subset V$, the vertices between $u$ and $v$ on this path are either in $V' - T$ or $V - V'$. Then on this path (including $u$ and $v$) there exists two distinct vertices $w$ and $z$ in $V' - T$ such that

20

1) $w$ and $z$ are in different components in $G^*(V')\backslash T$;

2) all the vertices between $w$ and $z$ on the path are in $V - V'$

(it is possible that $w = u$ and $z = v$). Then 2) above implies that $\{w, z\}$ is an edge in $G^*(V')$ (cf. Definition 13), which is a contradiction to 1). Therefore, we conclude that $u$ and $v$ are disconnected in $G\backslash(T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$. Hence $G\backslash(T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$ has at least two components and $(T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$ is a cutset in $G$. This completes the proof of the theorem. $\square$

The following corollary gives a structural property of $G^*(V')$.

**Corollary 4.** *If $T$ is a cutset in $G^*(V')$, then $T$ is also a cutset in $G$.*

**Proof** In the proof of Theorem 7, we have proved that if $T$ is a cutset in $G^*(V')$, then $(T \cup (\bigcup_{i=1}^{k}(Q_i - W_i)) \cup (V - V' - S))$ is a cutset in $G$. By setting $S = V - V'$ and $W_i = Q_i$ for all $i$, this cutset becomes $T$. This proves the corollary. $\square$

Combining Theorem 6 and Theorem 7, we have proved the main result of this section.

**Theorem 8.** *Let $G = (V, E)$, and let $V' \subset V$. Then $G^*(V')$ is the smallest $G'$ such that $G \Rightarrow G'$.*

We end this section with a discussion. There has been much research along the line of MRFs in the field of graphical models [25]. In particular, classes of graphical models that contain undirected graph as a special case were defined in [26, 27], where a separation criterion was provided for which the class of graphical models is stable under marginalization. In the context of the present paper, their result can be described as follows. Let $G = (V, E)$ be an undirected graph and $V' \subset V$. In [26, 27], an algorithm is provided that takes $G$ as the input and produces a graph as the output which is essentially the same as $G^*(V')$, and it was shown that if $X_i, i \in V$ satisfy only those conditional independencies induced by $G$ (i.e., $X_i, i \in V$ satisfy the conditional independencies induced by $G$ and no more), then $X_i, i \in V'$ satisfy only those conditional independencies induced by $G^*(V')$. This implies that if $G \Rightarrow G'$, then $G'$ cannot be a subgraph of $G^*(V')$, i.e., Theorem 8.

21

Although the graph produced by the algorithm in [27] is essentially the same as $G^*(V')$, it is not given in closed form. By contrast, our closed-form characterizations of $G^*(V')$ (Definition 13, Theorem 4, and Corollary 3) facilitate the development of further results, including Proposition 6 and the recursive approach for constructing information diagrams for MRFs to be discussed in Section 6.

It is also worth pointing out that our proof of Theorem 8, which is information-theoretic, is interesting on its own because it is developed upon the view that an MRF is a collection of FCMIs. As such, some of the results in this paper can potentially be generalized for general collections of FCMIs, which is beyond the scope of graphical models.

# 4 Markov Tree

Suppose $X_i, i \in V$ are represented by a graph $G$. If $G$ is a tree, then $X_i, i \in V$ form a Markov tree. If $G^*(V')$ is also a tree, we say that $X_i, i \in V'$ form a Markov subtree. For the special case when $G$ is a path, it is easy to see that $G^*(V')$ is always a path (see Example 9 for instance). In other words, if $X_i, i \in V$ form a Markov chain, then for any $V' \subset V$, $X_i, i \in V'$ always form a Markov subchain.

However, if $X_i, i \in V$ form a Markov tree, for an arbitrary subset $V'$ of $V$, $X_i, i \in V'$ may or may not form a Markov subtree. The following theorem, which is an application of Theorem 8, gives a necessary and sufficient condition for $X_i, i \in V'$ to form a Markov subtree.

**Theorem 9.** *Let $X_i, i \in V$ form an MRF represented by a tree $G = (V, E)$. For $V' \subset V$, $G^*(V')$ is a tree if and only if there do not exist $u \in V - V'$ and $v_1, v_2, v_3 \in V'$ such that for $i = 1, 2, 3$, all the vertices on the path between $u$ and $v_i$ except for $v_i$ are in $V - V'$.*

**Proof** We first prove the "only if" part. Assume that $G^*(V') = (V', E')$ is a tree and there exist $u \in V - V'$ and $v_1, v_2, v_3 \in V'$ and such that for $i = 1, 2, 3$, all the vertices on the path between $u$ and $v_i$ except for $v_i$ are in $V - V'$. By Definition 13, the edges $(v_1, v_2)$, $(v_2, v_3)$, and $(v_1, v_3)$ are in $E'$. Hence $v_1, v_2, v_3$ form a cycle in $G^*(V')$, a contradiction to the assumption that $G^*(V')$ is a tree.

We now prove the "if" part. Assume that $G^*(V') = (V', E')$ is not a tree. Then there exists a cycle $w_0, w_1, \cdots, w_{m-1}, w_0$ in $G^*(V')$, where $m \geq 3$ and $w_0, w_1, \cdots, w_{m-1} \in V'$ are distinct. By

22

Definition 13, for each $0 \leq i \leq m - 1$, there exists a path between $w_i$ and $w_{i+1}$ in $T$ on which all the intermediate vertices are in $V - V'$, where '+' in the subscript denotes modulo $m$ addition. This path is in fact unique because $G$ is a tree, so we denote it by $\text{Path}(w_i, w_{i+1})$.

If all the vertices on the collection of paths $\text{Path}(w_i, w_{i+1}), 0 \leq i \leq m-1$, except for the endpoints, are distinct, since $w_0, w_1, \cdots, w_{m-1}$ are distinct, these paths together form a cycle in $T$ which is a contradiction because $T$ is a tree. Otherwise, there exists a vertex $u \in V - V'$ which is on both $\text{Path}(w_i, w_{i+1})$ and $\text{Path}(w_j, w_{j+1})$ for some $0 \leq i < j \leq m - 1$. Note that $|\{w_i, w_{i+1}\} \cup \{w_j, w_{j+1}\}| \geq 3$, with equality if and only if $j = i+1 \mod m$. Then there exist $v_1, v_2, v_3 \in \{w_i, w_{i+1}\} \cup \{w_j, w_{j+1}\} \subset V'$ such that for $i = 1, 2, 3$, all the vertices on the path between $u$ and $v_i$ except for $v_i$ are in $V - V'$. The theorem is proved. $\qquad \square$

**Example 13.** *Consider a Markov tree represented by the tree G in Fig. 7 and let $V' = \{1, 4, 8, 9, 12\}$. The graph $G^*(V')$, illustrated as the overlay graph in grey, is evidently a tree. We call $G^*(V')$ a Markov subtree. It can be checked that the condition in Theorem 9 is satisfied.*

*However, if $V'$ also includes vertex 7, then $G^*(V')$ as shown in Fig. 8 is not a tree. By letting $u = 6$, $v_1 = 4$, $v_2 = 7$, and $v_3 = 8$, we see that the condition in Theorem 9 is violated because u is connected to each of $v_1$, $v_2$, and $v_3$ by an edge in $V - V'$.*

## 5  Markov Chain

A Markov chain is a special case of a Markov tree. However, there are certain properties that are possessed by a Markov chain but not by a Markov tree in general. Consider the graph $P_n = (V, E)$, where $V = \{1, 2, \cdots, n\}$ and the edges in $E$ are $\{i, i + 1\}$ for $i = 1, 2, \cdots, n - 1$. Evidently, $P_n$ is a path. If $X_i, i \in V$ is represented by $P_n$, then $X_i, i \in V$ form the Markov chain $X_1 \to X_2 \to \cdots \to X_n$. The following properties of a (finite-length) Markov chain were proved in [11]:

(C1)  An atom $A$ of $\mathcal{F}_V$ is a Type I atom if and only if

$$U_A = V - \{l, l + 1, \cdots, u\} \tag{15}$$

where $1 \leq l \leq u \leq n$, i.e., the indices of the set variables in $A$ that are not complemented are consecutive.

(C2) The values of $\mu^*$ on all the Type I atoms are nonnegative.

(C3) $\mu^*$ vanishes on all the Type II atoms.

Since $\mu^*$ vanishes on all the Type II atoms and is nonnegative on all the Type I atoms, it is a measure on $\mathcal{F}_V$. Also, the $I$-Measure $\mu^*$ of a finite-length Markov chain can be represented by a 2-dimensional information diagram as in Fig. 9, in which all the Type II atoms are suppressed.

Subsequently, (C3) was generalized for arbitrary finite undirected graphs [15]. However, the nonnegativity of the $I$-Measure does not hold even for the simplest Markov tree that is not a Markov chain [15].

**Example 14.** *Let $Z_1$ and $Z_2$ be i.i.d. random variables each distributed uniformly on $\{0, 1\}$. Let $X_1 = Z_1$, $X_2 = Z_2$, $X_3 = Z_1 + Z_2 \mod 2$, and $X_4 = (Z_1, Z_2)$. Since $X_1$, $X_2$, and $X_3$ are functions of $X_4$, they are mutually independent conditioning on $X_4$. Thus $X_1, X_2, X_3$, and $X_4$ form a Markov tree represented by the "star" in Figure 10. It is not difficult to show that (see [18, Example 3.10])*

$$\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) = -1,$$

*and hence $\mu^*$ is not nonnegative.*

Before explaining the significance of the nonnegativity of $\mu^*$ for Markov chains, we first review the following result in [11] which is instrumental in proving the nonnegativity of $\mu^*$ for a Markov chain. Prior to [11], the same result (and also the converse) was proved in [1] for the special case $U_A = \emptyset$.

**Lemma 3.** *If $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ form a Markov chain, then for a Type I atom with $U_A$ defined in (15),*

$$\mu^*(A) = \mu^* \left( \tilde{X}_l \cap \tilde{X}_{l+1} \cap \cdots \cap \tilde{X}_u - \tilde{X}_{U_A} \right) = \mu^* \left( \tilde{X}_l \cap \tilde{X}_u - \tilde{X}_{U_A} \right). \tag{16}$$

Note that the first equality above follows directly from the definition of $U_A$, and the quantity on the right hand side is equal to $I(X_l; X_u | X_{U_A})$ which is always nonnegative. In other words, Lemma 3 asserts that the values of $\mu^*$ on all the Type I atoms are nonnegative. Therefore, $\mu^*$ is a measure.

As mentioned in Section 2.1, for all $A \in \mathcal{A}_V$, $\mu^*(A)$ is a linear combination of $H(X_B)$ for nonempty subsets $B$ of $V$. Then if $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ forms a Markov chain, any linear information inequality involving $X_i, i \in V$ can be expressed in the form

$$\sum_{A \in \mathcal{T}_I} c_A \mu^*(A) \geq 0,$$

where $c_A \in \mathbb{R}$. The following theorem gives a complete characterization of such inequalities that always holds.

**Theorem 10.** *If $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ forms a Markov chain, then*

$$\sum_{A \in \mathcal{T}_I} c_A \mu^*(A) \geq 0 \tag{17}$$

*always holds if and only if $c_A \geq 0$ for all $A \in \mathcal{T}_I$.*

**Proof** If $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ forms a Markov chain, then $\mu^*(A) \geq 0$ for all $A \in \mathcal{T}_I$. If $c_A \geq 0$ for all $A \in \mathcal{T}_I$, then evidently (17) always holds.

To prove the converse, assume that $c_{A'} < 0$ for an atom $A' \in \mathcal{T}_I$. Now construct $X_i, i \in V$ by letting

$$X_i = \begin{cases} Z & \text{if } i \in U_{A'} \\ \text{constant} & \text{otherwise,} \end{cases}$$

where $Z$ is a random variable such that $0 < H(Z) < \infty$. Then by the proof of Theorem 3.11 in [18], for all $A \in \mathcal{T}_I$,

$$\mu^*(A) = \begin{cases} H(Z) & \text{if } A = A' \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\sum_{A \in \mathcal{T}_I} c_A \mu^*(A) = c_{A'} \mu^*(A') < 0$$

since $c_{A'} < 0$ and $\mu^*(A') > 0$. Hence, (17) does not always hold and the converse is proved. $\square$

**Remark** Let $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$ form a Markov chain and consider any inequality of the form (17) that always holds. Theorem 10 asserts that the left hand side of (17) must be a conic combination of $\mu^*(A), A \in \mathcal{T}_I$. Since $\mu^*(A)$ is a Shannon's information measure for all $A \in \mathcal{T}_I$,

25

we see that (17) is implied by the nonnegativity of Shannon's information measures and hence is a *Shannon-type information inequality* (see [18, Ch. 14]). Therefore, we conclude that there exist no *non-Shannon-type* information inequalities for a Markov chain.

Fix a graph $G$ and let $\mathcal{P}_G$ be the class of probability measures $P$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ such that $P$ forms an MRF represented by $G$. In the rest of this section, we prove that the $I$-Measure $\mu^*$ of every $P \in \mathcal{P}_G$ is nonnegative if and only if $G$ is either a path or a forest of paths. In other words, the MRF represented by such a graph $G$ is either a Markov chain or a collection of mutually independent Markov chains. In this sense we say that the Markov chain is the only MRF for which the $I$-Measure is always nonnegative.

In the following, we present a theorem which is a generalization of Lemma 3. Unlike Lemma 3 that applies only to Markov chains, this theorem applies to all MRFs.

**Theorem 11.** *Let $X_1, X_2, \cdots, X_n$ form a Markov graph $G = (V, E)$. For a Type I atom $A$ of $G$ with $|U_A| \leq n - 2$,*

$$\mu^*(A) = \mu^* \left( \bigcap_{k \in B} \tilde{X}_k - \tilde{X}_{U_A} \right), \tag{18}$$

*where*

$$B = \{k \in V - U_A : s(U_A \cup \{k\}) = 1\},$$

*i.e., a vertex $k \notin U_A$ is in $B$ if and only if upon removing all the vertices in $U_A$ and vertex $k$, the graph remains connected.*

**Example 15.** *Consider an MRF represented by the graph in Fig. 11. For the Type I atom $\bar{1}2\bar{3}45678$, using Theorem 11, $B = \{2, 7, 8\}$, and so*

$$\mu^* \left( \tilde{X}_2 \cap \tilde{X}_4 \cap \tilde{X}_5 \cap \tilde{X}_6 \cap \tilde{X}_7 \cap \tilde{X}_8 - \tilde{X}_{\{1,3\}} \right) = \mu^* \left( \tilde{X}_2 \cap \tilde{X}_7 \cap \tilde{X}_8 - \tilde{X}_{\{1,3\}} \right).$$

*For the Type I atom $\bar{1}\bar{2}3454 67\bar{8}$, $B = \{3, 4, 6, 7\}$, and so*

$$\mu^* \left( \tilde{X}_3 \cap \tilde{X}_4 \cap \tilde{X}_5 \cap \tilde{X}_6 \cap \tilde{X}_7 - \tilde{X}_{\{1,2,8\}} \right) = \mu^* \left( \tilde{X}_3 \cap \tilde{X}_4 \cap \tilde{X}_6 \cap \tilde{X}_7 - \tilde{X}_{\{1,2,8\}} \right).$$

To gain insight into Theorem 11, we first state the next lemma. This lemma and the technical lemma that follows will be proved in Appendix C.

26

**Lemma 4.** *In Theorem 11, $|B| \geq 2$.*

**Remark** When $|B| = 2$, the term on the right hand side of (18) becomes a (conditional) mutual information, which is always nonnegative.

The following lemma will be used in the proof of Theorem 11.

**Lemma 5.** *In Theorem 11, let $W = V - U_A - B$. For any $S \subsetneq W$, $s(U_A \cup (W - S)) > 1$.*

**Proof of Theorem 11** Let $W = V - U_A - B$. Consider

$$
\mu^* \left( \bigcap_{k \in B} \tilde{X}_k - \tilde{X}_{U_A} \right) = \mu^* \left( \bigcup_{S \subset W} \left( \left( \bigcap_{k \in B} \tilde{X}_k \right) \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{U_A \cup (W - S)} \right) \right)
$$

$$
= \sum_{S \subset W} \mu^* \left( \left( \bigcap_{k \in B} \tilde{X}_k \right) \cap \left( \bigcap_{t \in S} \tilde{X}_t \right) - \tilde{X}_{U_A \cup (W - S)} \right).
$$

In the above summation, for $S \subsetneq W$, $s(U_A \cup (W - S)) > 1$ by Lemma 5. Therefore, except for the atom corresponding to $S = W$, i.e., $A$, all the atoms are Type II atoms of $G$. It then follows that

$$
\mu^* \left( \bigcap_{k \in B} \tilde{X}_k - \tilde{X}_{U_A} \right) = \mu^* \left( \left( \bigcap_{k \in B} \tilde{X}_k \right) \cap \left( \bigcap_{t \in W} \tilde{X}_t \right) - \tilde{X}_{U_A} \right)
$$

$$
= \mu^* \left( \left( \bigcap_{k \in V - U_A} \tilde{X}_k \right) - \tilde{X}_{U_A} \right)
$$

$$
= \mu^*(A).
$$

The theorem is proved. □

Theorem 11 can be applied to identify atoms on which the value of $\mu^*$ is always nonnegative, because when $|B| = 2$, the term on the right hand side of (18) corresponds to a (conditional) mutual information.

Consider the graph $G = (V, E)$, where $V = \{1, 2, \cdots, n\}$ and the edges in $E$ are $\{i, i + 1\}$ for $i = 1, 2, \cdots, n - 1$ and $\{1, n\}$. Evidently, $G$ is a *cycle graph*, and if random variables $X_i, i \in V$ are represented by $G$, they form a *Markov ring*. Then $A$ is a Type I atom of $G$ if and only if $U_A = \emptyset$ or $U_A$ is a consecutive subset of $V$ in the cyclic sense (e.g., $\{1, n\}$ is regarded as a consecutive subset of $V$). An application of Theorem 11 reveals that $\tilde{X}_1 \cap \tilde{X}_2 \cap \cdots \cap \tilde{X}_n$ (i.e., $U_A = \emptyset$) is the only

27

atom on which $\mu^*$ may take a negative value, because the value of $\mu^*$ on any other Type I atom is a conditional mutual information. This observation is instrumental in the proof of the next theorem, the main result in this section.

**Theorem 12.** *Let G be a connected graph. Then $\mu^*$ is nonnegative for every $P \in \mathcal{P}_G$ if and only if G is a path.*

The 'if' part of Theorem 12 is immediate because the *I*-Measure for a Markov chain is always nonnegative. Toward proving the 'only if' part, we first classify all connected graphs into the following two classes:

$K1$: there exists a vertex whose degree is at least 3;

$K2$: all the vertices have degree less than or equal to 2.

We further classify the graphs in $K2$ into two subclasses:

$K2$-$a$: all the vertices have degree 2;

$K2$-$b$: some vertices have degree 1.

It is easy to see that a graph belonging to subclass $K2$-$a$ is a cycle graph, and a graph belonging to subclass $K2$-$b$ is a path. Thus in order to establish Theorem 12, it suffices to prove Theorem 13 and Theorem 14 below which assert that $\mu^*$ is not always nonnegative if $X_i, i \in V$ are represented by a graph belonging to $K1$ and $K2$-$a$, respectively.

**Theorem 13.** *The I-Measure $\mu^*$ for an MRF represented by a graph G belonging to $K1$ is not always nonnegative.*

**Proof** Consider a graph $G = (V, E)$ in $K1$. Let $u \in V$ be a vertex whose degree is at least 3, and let $\{u, v_l\} \in E$, where $l = 1, 2, 3$ and $v_1, v_2$, and $v_3$ are distinct. Let $Z$ and $T$ be independent fair bits.

Define random variables $X_i, i \in V$ as follows:

$$
X_i = \begin{cases}
Z & i = v_1 \\
T & i = v_2 \\
Z + T \bmod 2 & i = v_3 \\
(Z, T) & i = u \\
\text{constant} & \text{otherwise}
\end{cases}
$$

Consider any cutset $U$ of $G$:

1. If $u \notin U$, then $u$ and $v$ for all $v \in \{v_1, v_2, v_3\} \backslash U$ are in the same component of $G \backslash U$ because $u$ and $v$ are connected by an edge in $G$. Since $X_i = $ constant for all $i \neq u, v_1, v_2, v_3$, it is readily seen that $X_{V_1(U)}, X_{V_2(U)}, \cdots, X_{V_{s(U)}(U)}$ are mutually independent conditioning on $X_U$.

2. If $u \in U$, since $X_{v_l}, l = 1, 2, 3$ are functions of $X_u$ and $X_i = $ constant for all $i \neq u, v_1, v_2, v_3$, it is readily seen that $X_{V_1(U)}, X_{V_2(U)}, \cdots, X_{V_{s(U)}(U)}$ are mutually independent conditioning on $X_U$.

Thus in either case $X_i, i \in V$ are represented by $G$. Then

$$
\mu^* \left( \tilde{X}_u \cap \tilde{X}_{v_1} \cap \tilde{X}_{v_2} \cap \tilde{X}_{v_3} - \bigcup_{i \neq u, v_1, v_2, v_3} \tilde{X}_i \right) = \mu^* (\tilde{X}_u \cap \tilde{X}_{v_1} \cap \tilde{X}_{v_2} \cap \tilde{X}_{v_3})
$$
$$
= -1,
$$

where the first equality can be seen by expanding $\mu^* \left( \tilde{X}_u \cap \tilde{X}_{v_1} \cap \tilde{X}_{v_2} \cap \tilde{X}_{v_3} - \bigcup_{i \neq u, v_1, v_2, v_3} \tilde{X}_i \right)$ using [18, Theorem 3.19] into a linear combination of $H(\cdot \mid \tilde{X}_i, i \neq u, v_1, v_2, v_3) = H(\cdot)$, and the second equality can easily be verified (cf. Problem 5, Ch. 12 in [18]). Hence, $\mu^*$ for $X_i, i \in V$ represented by a graph $G$ belonging to $K1$ is not always nonnegative.

**Theorem 14.** *The I-Measure $\mu^*$ for an MRF represented by a graph $G$ belonging to K2-a is not always nonnegative.*

**Proof** Consider a graph $G = (V, E)$ in *K2-a*, i.e., $G$ is a cycle graph. For convenience, let $V = \{0, 1, \cdots, n-1\}$. The edge set $E$ is specified by $\{u, v\} \in E$ if and only if $|u - v| = 1$, where "$-$" denotes modulo $n$ subtraction. Let $F$ denote a finite field containing at least $n - 1$ elements. Let $Z$ and $T$ be

29

independent random variable, each taking values in $F$ according to the uniform distribution. Now define random variables $X_i, i \in V$ as follows:

$$
X_i = \begin{cases}
Z & i = 0 \\
T & i = 1 \\
Z + \alpha_i T & i = 2, 3, \cdots, n - 1
\end{cases}
$$

where $\alpha_i, i = 2, 3, \cdots, n-1$ are distinct nonzero elements of $F$. It is evident that $X_i, i = 0, 1, \cdots, n-1$ are pairwise independent but not mutually independent, and that for any distinct $i, i', i''$, we have $X_{i''}$ being a function of $(X_i, X_{i'})$.

We now show that $X_i, i \in V$ is represented by $G$. Since $G$ is a cycle graph, for any $U \subset V$, if the vertices in $U$ are connected in $G$, the vertices in $V - U$ are also connected in $G$. Therefore, if $U$ is a cutest in $G$, the vertices in $U$ are not connected in $G$. This implies that $|U| \geq 2$. From the foregoing, $X_{V-U}$ is a function of $X_U$. Then we see that $X_{V_1(U)}, X_{V_2(U)}, \cdots, X_{V_{s(U)}(U)}$ are mutually independent conditioning on $X_U$. Therefore, $X_i, i \in V$ is represented by $G$.

It remains to show that $\mu^*$ is not nonnegative. For the sake of convenience, assume the logarithms defining entropy are in the base $|F|$. Then for $B \subset V$ such that $B \neq \emptyset$,

$$
H(X_B) = \begin{cases}
1 & \text{if } |B| = 1 \\
2 & \text{if } 2 \leq |B| \leq n.
\end{cases} \tag{19}
$$

We will show that $\mu^*$ is given by

$$
\mu^* \left( \bigcap_{i \in W} \tilde{X}_i - \bigcup_{j \in V-W} \tilde{X}_j \right) = \begin{cases}
0 & \text{if } 1 \leq |W| \leq n - 2 \\
1 & \text{if } |W| = n - 1 \\
-(n-2) & \text{if } |W| = n
\end{cases} \tag{20}
$$

for $W \subset V$. Toward this end, owing to the uniqueness of $\mu^*$, we only need to verify that $\mu^*$ as prescribed by (20) satisfies (19). The details are given in Appendix B. Then the theorem is proved because $\mu^*$ is not nonnegative. $\qquad \square$

**Theorem 15.** *Let $G$ be a graph with at least two components. Then $\mu^*$ is nonnegative for every $P \in \mathcal{P}_G$ if and only if $G$ is a forest of paths.*

**Proof** We first prove the 'only if' part. Assume that $G$ is not a forest of paths, i.e., there exists a component of $G$ which is not a path. Denote the vertices of this component by $V'$ and let $X_i, i \in V \backslash V'$ be constant. Then by Theorem 12, we can construct $X_i, i \in V$ such that $\mu^*(S) < 0$ for some $S \subset \mathcal{F}_{V'} \subset \mathcal{F}_V$, where $\mathcal{F}_{V'}$ is the $\sigma$-field generated by $\{\tilde{X}_i, i \in V'\}$. Hence $\mu^*$ is not nonnegative, and the 'only if' part is proved.

To prove the 'if' part, we need to prove that if an MRF is represented by a graph $G$ which is a forest of paths, then $\mu^*$ is always nonnegative. Let $m$ be the number of components of $G$, where $m \geq 2$, and denote the sets of vertices of these components by $V_1, V_2, \cdots, V_m$. Without loss of generality, assume that the indices in each $V_i$ are consecutive.

Now observe that a nonempty atom $A$ of $\mathcal{F}_V$ is a Type I atom of $\mathcal{F}_V$ if and only if $U_A$ has the form (15) and $\{l, l+1, \cdots, u\} \subset V_i$ for some $1 \leq i \leq m$. If $l = u$, then

$$\mu^*(A) = H(\tilde{X}_l | \tilde{X}_{V-\{l\}}) \geq 0.$$

If $l < u$, then by Theorem 11,

$$\mu^*(A) = \mu^*\left(\bigcap_{l \leq k \leq u} \tilde{X}_k - \tilde{X}_{U_A}\right) = \mu^*(\tilde{X}_l \cap \tilde{X}_k - \tilde{X}_{U_A}) \geq 0.$$

Hence $\mu^*$ is nonnegative, and the theorem is proved. □

# 6 Information Diagrams for Markov Random Fields

As discussed in Section 5, the $I$-Measure $\mu^*$ of a finite-length Markov chain can be represented by a 2-dimensional information diagram as in Fig. 9. Such an information diagram is a "correct" representation in the sense that the closed curves representing the set variables intersect with each other in such a way that

1. all the Type I atoms are nonempty (not suppressed);

2. all the Type II atoms are empty (suppressed).

We call Fig. 9 an information diagram (customized) for a Markov chain, or more specifically an information diagram for the path $P_n$ (as discussed in Section 5). With such an information diagram, it

31

is relatively easy to discover information inequalities and identities pertaining to a Markov chain by visualization, which may be difficult otherwise. A notable such example is an information identity for a Markov chain of five random variables that was useful in proving an outer bound for multiple descriptions [16] (see also [18, Example 3.18]).

Owing to its simple and regular structure, it is possible to construct an information diagram for a Markov chain by trial and error. However, constructing an information diagram for a general MRF requires a more systematic approach. In the rest of this section, we develop a method for this purpose by using the characterization of a subfield of an MRF in Section 4.

To simplify notation, we use $N_n$ to denote $\{1, 2, \cdots, n\}$. Consider $X_i, i \in V$ forming a Markov graph $G = (V, E)$ with $V = N_n$. Using Corollary 3 as the recipe, we can construct $G^*(N_{n-1})$. Then by repeating this step with $G^*(N_{n-1})$ in place of $G$, we can construct $(G^*(N_{n-1}))^*(N_{n-2})$, which from Proposition 6 is in fact equal to $G^*(N_{n-2})$. In the same fashion, we can construct the graphs $G^*(N_{n-3}), \cdots, G^*(N_1)$ recursively.

In our method for constructing an information diagram for $G$, we construct a sequence of information diagrams for $G^*(N_1)$, $G^*(N_2), \cdots, G^*(N_n) = G$ recursively, with the last one being the desired information diagram. Denote these information diagrams by $\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_n$. For the convenience of discussion, denote the closed curve representing $\tilde{X}_m$ by $C_m$ for $1 \le m \le n$.

Now the graph $G^*(N_1)$ consists of the single vertex 1 and no edge. Then an information diagram consisting of any closed curve representing $\tilde{X}_1$ would be a correct representation for $G^*(N_1)$. Call this information diagram $\mathcal{D}_1$.

We observe that for $m = 2, 3, \cdots, n$, an atom $A \in \mathcal{A}_{N_{m-1}}$ generates the two atoms $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ in $\mathcal{A}_{N_m}$, and there is the extra atom $\tilde{X}_1^c \cap \tilde{X}_2^c \cap \cdots \cap \tilde{X}_{m-1}^c \cap \tilde{X}_m$ in $\mathcal{A}_{N_m}$ that is not generated by any atom in $\mathcal{A}_{N_{m-1}}$.

For $m = 2, 3, \cdots, n$, in constructing $\mathcal{D}_m$ from $\mathcal{D}_{m-1}$, we add the closed curve $C_m$ to the former in a suitable way. In order for this recursive approach to work, we need to ensure that a Type II atom of $G^*(N_{m-1})$ that is suppressed in $\mathcal{D}_{m-1}$ would not generate a Type I atom of $G^*(N_m)$ which is not to be suppressed in $\mathcal{D}_m$. This is proved in the next theorem.

**Theorem 16.** *For $m = 2, 3, \cdots, n$, if $A$ is a Type II atom of $G^*(N_{m-1})$, then both $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$*

*are Type II atoms of* $G^*(N_m)$.

**Proof** Assume that $A$ is a Type II atom of $G^*(N_{m-1})$. We first prove that $A \cap \tilde{X}_m^c$ is a Type II atom of $G^*(N_m)$. From the discussion following Proposition 6, we know that $G^*(N_m)\backslash\{m\}$ is a subgraph of $G^*(N_{m-1})$. As such, upon removing all the vertices in $U_A$ in both graphs, we see that $G^*(N_m)\backslash(\{m\} \cup U_A)$ is subgraph of $G^*(N_{m-1})\backslash U_A$, where the latter is disconnected because $A$ is a Type II atom of $G^*(N_{m-1})$. It then follows that $G^*(N_m)\backslash(\{m\}\cup U_A)$ is also disconnected. Upon noting that $U_{A\cap\tilde{X}_m^c} = U_A \cup \{m\}$, we see that $G^*(N_m)\backslash U_{A\cap\tilde{X}_m^c} = G^*(N_m)\backslash(\{m\} \cup U_A)$ which is disconnected. Therefore, $A \cap \tilde{X}_m^c$ is a Type II atom of $G^*(N_m)$.

We now prove that $A \cap \tilde{X}_m$ is a Type II atom of $G^*(N_m)$. Let

$$\gamma_m = \{\, j : \{j, m\} \text{ is an edge in } G^*(N_m)\,\} \tag{21}$$

be the set of neighbors of vertex $m$ in $G^*(N_m)$. Note that $\gamma_m \subset N_{m-1}$. Let $\tilde{E}_m$ and $\tilde{E}_{m-1}$ be the sets of edges of $G^*(N_m)\backslash U_A$ and $G^*(N_{m-1})\backslash U_A$, respectively. Evidently,

$$\tilde{E}_m - \tilde{E}_{m-1} = \{\{m, j\} : j \in \gamma_m - U_A\}, \tag{22}$$

where $\gamma_m - U_A$ is the set of neighbors of vertex $m$ in $G^*(N_m)\backslash U_A$. We consider two cases for $\gamma_m - U_A$.

$\underline{\gamma_m - U_A = \emptyset}$

This is the case when vertex $m$ has no neighbor in $G^*(N_m)\backslash U_A$. Since $A$ is a Type II atom of $G^*(N_{m-1})$, $G^*(N_{m-1})\backslash U_A$ is disconnected. From (22), we have $\tilde{E}_m - \tilde{E}_{m-1} = \emptyset$, so that $\tilde{E}_m \subset \tilde{E}_{m-1}$. This implies that $G^*(N_m)\backslash U_A$ is also disconnected. Upon noting that $U_{A\cap\tilde{X}_m} = U_A$, we have $G^*(N_m)\backslash U_{A\cap\tilde{X}_m} = G^*(N_m)\backslash U_A$ which is disconnected. Therefore, $A \cap \tilde{X}_m$ is a Type II atom of $G^*(N_m)$.

$\underline{\gamma_m - U_A \neq \emptyset}$

This is the case when vertex $m$ has at least one neighbor in $G^*(N_m)\backslash U_A$. For any distinct vertices $i_1$ and $i_2$ that are both neighbors of vertex $m$ in $G^*(N_m)\backslash U_A$ (and therefore also neighbors of vertex $m$ in $G^*(N_m)$), according to Corollary 3, $\{i_1, i_2\}$ is an edge in $G^*(N_{m-1})$ and hence $\{i_1, i_2\} \in \tilde{E}_{m-1}$ (because $i_1, i_2 \notin U_A$), which implies that $i_1$ and $i_2$ belong to the same component in $G^*(N_{m-1})\backslash U_A$. Equivalently, if $i_1$ and $i_2$ belong to different components in $G^*(N_{m-1})\backslash U_A$, then $i_1$ and $i_2$ cannot both

33

be neighbors of $m$ in $G^*(N_m)\backslash U_A$. Therefore, the neighbors of $m$ in $G^*(N_m)\backslash U_A$ all belong to the same component in $G^*(N_{m-1})\backslash U_A$, and we denote this component by $V_1$.

Since $G^*(N_{m-1})\backslash U_A$ is disconnected, there exists another component $V_2$ in $G^*(N_{m-1})\backslash U_A$. Now consider any $i_1 \in V_1$ and $i_2 \in V_2$. Since $i_1$ and $i_2$ are in different components in $G^*(N_{m-1})\backslash U_A$, we have $\{i_1, i_2\} \notin \tilde{E}_{m-1}$. Then we see from the discussion in the last paragraph that $i_1$ and $i_2$ cannot both be neighbors of $m$ in $G^*(N_m)\backslash U_A$. By Corollary 3, $\{i_1, i_2\}$ is not an edge in $G^*(N_{m-1})$ and hence not an edge in $G^*(N_m)\backslash U_A$ because $G^*(N_m)\backslash U_A$ is a subgraph of $G^*(N_{m-1})$. Therefore, $(i_1, i_2) \notin \tilde{E}_m$. Also, since all the neighbors of vertex $m$ in $G^*(N_m)\backslash U_A$ are in $V_1$, $i_2$ is not a neighbor of vertex $m$ in $G^*(N_m)\backslash U_A$, and therefore $\{m, i_2\} \notin \tilde{E}_m$.

Summarizing the above, we have proved that for any $i \in V_1 \cup \{m\}$ and $i_2 \in V_2$, $\{i, i_2\} \notin \tilde{E}_m$. Hence, $V_1 \cup \{m\}$ and $V_2$ are distinct components in $G^*(N_m)\backslash U_A$, so that $G^*(N_m)\backslash U_A$ is disconnected. Finally, upon noting that $G^*(N_m)\backslash U_{A \cap \tilde{X}_m} = G^*(N_m)\backslash U_A$, we see that $A \cap \tilde{X}_m$ is a Type II atom of $G^*(N_m)$. $\qquad\square$

When we construct $\mathcal{D}_m$ by adding $\tilde{X}_m$ to $\mathcal{D}_{m-1}$, for each Type I atom $A$ of $G^*(N_{m-1})$, the closed curve $C_m$ is required to

(B1) *Split $A$ into two regions* if both $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ are Type I atoms of $G^*(N_m)$, so that both $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ are not suppressed in $\mathcal{D}_m$;

(B2) *Include $A$ in $\tilde{X}_m$* if $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ are Type I and Type II atoms of $G^*(N_m)$, respectively, so that $A \cap \tilde{X}_m$ is not suppressed and $A \cap \tilde{X}_m^c$ is suppressed in $\mathcal{D}_m$; or

(B3) *Exclude $A$ from $\tilde{X}_m$* if $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ are Type II and Type I atoms of $G^*(N_m)$, respectively, so that $A \cap \tilde{X}_m$ is suppressed and $A \cap \tilde{X}_m^c$ is not suppressed in $\mathcal{D}_m$.

However, if both $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ are Type II atoms of $G^*(N_m)$, there is no way the closed curve $C_m$ can be drawn such that both of these atoms are suppressed in $\mathcal{D}_m$. Under this situation, our recursive approach for constructing an information diagram for $G$ would not work. The following theorem (see Corollary 5) precludes this possibility.

**Theorem 17.** *If A is a Type I atom of $G^*(N_{m-1})$, then*

*i) if $|\gamma_m - U_A| = 0$, then A belongs to (B3);*

*ii) if $|\gamma_m - U_A| = 1$, then A belongs to (B1);*

*iii) if $|\gamma_m - U_A| \geq 2$, then A belongs to either (B1) or (B2).*

**Corollary 5.** *If A is a Type I atom of $G^*(N_{m-1})$, then at least one of $A \cap \tilde{X}_m$ and $A \cap \tilde{X}_m^c$ is a Type I atom of $G^*(N_m)$.*

**Proof of Theorem 17** Let $E'_m$ and $E''_m$ be the edge sets of $G^*(N_{m-1})$ and $G^*(N_m)\backslash\{m\}$, respectively. By Corollary 3, we have

$$E'_m = E''_m \cup \kappa(\gamma_m), \tag{23}$$

which implies

$$E'_m - E''_m \subset \kappa(\gamma_m). \tag{24}$$

Let $\tilde{E}'_m$ and $\tilde{E}''_m$ be the edge sets of $G^*(N_{m-1})\backslash U_A$ and $G^*(N_m)\backslash(\{m\} \cup U_A)$, respectively. We see that $\tilde{E}'_m$ and $\tilde{E}''_m$ can be obtained respectively from $E'_m$ and $E''_m$ by removing the edges joining at least one vertex in $U_A$ from these sets. Then upon removing these edges from every set in (24), we obtain

$$\tilde{E}'_m - \tilde{E}''_m \subset \kappa(\gamma_m - U_A). \tag{25}$$

Assume A is a Type I atom of $G^*(N_{m-1})$. We now prove the theorem for each of the three cases.

i) $|\gamma_m - U_A| = 0$

We first prove that $A \cap \tilde{X}_m^c$ is a Type I atom of $G^*(N_m)$. Since $|\gamma_m - U_A| = 0$, i.e., $\gamma_m - U_A = \emptyset$, it follows from (25) that $\tilde{E}'_m - \tilde{E}''_m = \emptyset$, or $\tilde{E}'_m \subset \tilde{E}''_m$. In fact,

$$\tilde{E}'_m = \tilde{E}''_m \tag{26}$$

because $\tilde{E}''_m \subset \tilde{E}'_m$ in general (cf. (23)). Now, if A is a Type I atom of $G^*(N_{m-1})$, then $G^*(N_{m-1})\backslash U_A$ is connected. In view of (26), we see that $G^*(N_m)\backslash(\{m\} \cup U_A) = G^*(N_m)\backslash U_{A \cap \tilde{X}_m^c}$ is also connected. Therefore, $A \cap \tilde{X}_m^c$ is a Type I atom of $G^*(N_m)$.

We now prove that $A \cap \tilde{X}_m$ is a Type II atom of $G^*(N_m)$. From the last paragraph, $G^*(N_m)\backslash(\{m\} \cup U_A)$ is connected. Then all the vertices in $N_{m-1} - U_A$ are connected in $G^*(N_m)\backslash(\{m\} \cup U_A)$, and

35

they remain connected in $G^*(N_m)\backslash U_A$ because $G^*(N_m)\backslash(\{m\} \cup U_A)$ is a subgraph of $G^*(N_m)\backslash U_A$.

On the other hand, since $|\gamma_m - U_A| = 0$, $\{m, i\}$ is not an edge in $G^*(N_m)$ for all $i \in N_{m-1} - U_A$, and so in in $G^*(N_m)$, vertex $m$ is not connected with any vertex in $N_{m-1} - U_A$. Since $G^*(N_m)\backslash U_A$ is a subgraph of $G^*(N_m)$, we see that in $G^*(N_m)\backslash U_A$, vertex $m$ is not connected with any vertex in $N_{m-1} - U_A$. Thus $G^*(N_m)\backslash U_A$ is disconnected. Hence, we conclude that $G^*(N_m)\backslash U_{A \cap \tilde{X}_m} = G^*(N_m)\backslash U_A$ is disconnected, i.e. $A \cap \tilde{X}_m$ is a Type II atom of $G^*(N_m)$.

ii) $|\gamma_m - U_A| = 1$

To prove that $A \cap \tilde{X}_m^c$ is a Type I atom of $G^*(N_m)$, we only have to observe that in (25), $\kappa(\gamma_m - U_A) = \emptyset$ when $|\gamma_m - U_A| = 1$. Then (26) holds and we can apply the same argument as in case i).

We now prove that $A \cap \tilde{X}_m$ is a Type I atom of $G^*(N_m)$. For the economy of presentation, we will give a proof that also covers case iii), i.e., $|\gamma_m - U_A| \geq 2$.

It follows from (25) that for $i, j \in N_{m-1} - U_A$, if $\{i, j\} \in \tilde{E}_m'$ and $\{i, j\} \notin \tilde{E}_m''$, then $\{i, j\} \in \kappa(\gamma_m - U_A)$, i.e., $i, j \in \gamma_m - U_A$. Equivalently, if $i \notin \gamma_m - U_A$ or $j \notin \gamma_m - U_A$, then $\{i, j\} \notin \tilde{E}_m'$ or $\{i, j\} \in \tilde{E}_M''$. In other words, if $i$ and $j$ are not both in $\gamma_m - U_A$ and $\{i, j\} \in \tilde{E}_m'$, then $\{i, j\} \in \tilde{E}_m''$.

Since $A$ is a Type I atom of $G^*(N_{m-1})$, $G^*(N_{m-1})\backslash U_A$ is connected. Then for any distinct $\alpha, \beta \in N_{m-1} - U_A$, there exists a path between vertices $\alpha$ and $\beta$ in $G^*(N_{m-1})\backslash U_A$.

If such a path does not contain an edge with both endpoints in $\gamma_m - U_A$, from the discussion in the second last paragraph, we see that all the edges on the path are in $G^*(N_m)\backslash(\{m\} \cup U_A)$ and hence are also in $G^*(N_m)\backslash U_A$ because $G^*(N_m)\backslash(\{m\} \cup U_A)$ is a subgraph of $G^*(N_m)\backslash U_A$. In other words, $\alpha$ and $\beta$ are connected in $G^*(N_m)\backslash U_A$.

If the path contains any edge with both endpoints in $\gamma_m - U_A$, we can construct another path between $\alpha$ and $\beta$ in $G^*(N_m)\backslash U_A$ by replacing every such an edge $\{u, v\}$ by edges $\{u, m\}$ and $\{m, v\}$ which are both in $G^*(N_m)\backslash U_A$. Then we see that $\alpha$ and $\beta$ are connected in $G^*(N_m)\backslash U_A$.

Finally, since $\gamma_m \not\subset U_A$, there exists a vertex $u \in \gamma_m - U_A$. Then for any $\alpha \in N_{m-1} - U_A$, we can readily see that vertices $\alpha$ and $m$ are connected in $G^*(N_m)\backslash U_A$ because $\alpha$ is connected with $u$ and $u$ is connected with $m$. Summarizing the above, for any distinct $\alpha, \beta \in N_m - U_A$, $\alpha$ and $\beta$ are connected in $G^*(N_m)\backslash U_A$. This implies that $G^*(N_m)\backslash U_A = G^*(N_m)\backslash U_{A \cap \tilde{X}_m}$ is connected, i.e., $A \cap \tilde{X}_m$ is a Type I atom of $G^*(N_m)$.

ii) $|\gamma_m - U_A| \geq 2$

To prove that $A$ belongs to either (B1) or (B2), we only need to prove that $A \cap \tilde{X}_m$ is a Type I atom of $G^*(N_m)$. This has already been proved in case ii). □

Now in light of Theorem 16 and Corollary 5, we can construct $\mathcal{D}_m$ from $\mathcal{D}_{m-1}$ according to the prescriptions in (B1)-(B3) for each Type I atom of $G^*(N_{m-1})$. Theorem 17 helps simplify the checking of which of (B1) to (B3) these atoms belong to, as we now explain. The three conditions for $A$ in Theorem 17 are

i) $|\gamma_m - U_A| = 0$: This means that $A \not\subset \tilde{X}_i$ for any $i \in \gamma_m$, or equivalently, $A \not\subset \tilde{X}_{\gamma_m}$. In this case, the closed curve $C_m$ excludes the atom $A$.

ii) $|\gamma_m - U_A| = 1$: This means that $A \subset \tilde{X}_i$ for only one $i \in \gamma_m$. In this case, the closed curve $C_m$ splits the atom $A$.

iii) $|\gamma_m - U_A| \geq 2$: This means that $A \subset \tilde{X}_i$ for more than one $i \in \gamma_m$. If $A$ belongs to (B1), the closed curve $C_m$ splits the atom $A$, otherwise ($A$ belongs to (B2)) it includes the atom $A$.

Note that if $A$ satisfies condition iii), we still need to check whether it belongs to (B1) or (B2), which is unavoidable. This is explained by the two examples in Figs. 12 and 13. In each of these examples, we have $n = 3$ and $\gamma_3 = \{1, 2\}$. Consider the Type I atom $A = 12$ of $G^*(N_2)$. Then $U_A = \emptyset$ and $|\gamma_3 - U_A| = 2$, i.e., $A$ satisfies condition iii).

**Example 16.** *We apply our method to construct an information diagram for the Markov chain $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$. From Corollary 2, we see that for $m = n, n - 1, \cdots, 1$, $G^*(N_m)$ is the path $P_n$ defined in Section 5.*

*The information diagram $\mathcal{D}_1$, which consists of only the set variable $\tilde{X}_1$, is completely trivial. For $2 \leq m \leq n$, when we construct $\mathcal{D}_m$ from $\mathcal{D}_{m-1}$, since $\gamma_m = \{m - 1\}$ (cf. (21)) is a singleton, every Type I atom $A \subset \tilde{X}_{m-1}$ in $\mathcal{D}_{m-1}$ satisfies the condition that $A \subset \tilde{X}_i$ for only one $i \in \gamma_m$. Then based on the discussion in the foregoing, the closed curve $C_m$ excludes every Type I atom in $\mathcal{D}_{m-1}$ that is not contained in $\tilde{X}_{m-1}$ and splits every Type I atom in $\mathcal{D}_{m-1}$ that is contained in $\tilde{X}_{m-1}$. This way, we can obtain the information diagram in Fig. 9.*

Our method is robust in the sense that it works for any labeling of the vertices. However, it may be more convenient to construct the information diagram for one labeling than another. This is illustrated in the following two examples.

**Example 17.** *Consider the graph G in Fig. 14. Upon noting that $\gamma_3$ and $\gamma_4$ are singletons (both equal to $\{2\}$), by using the technique in Example 16, we can readily construct the information diagram in Fig. 15.*

**Example 18.** *Consider the graph G in Fig. 10, which is the same as the one in Fig. 14 except that the vertices are labelled differently. Here, $G^*(N_3)$ is the complete graph $K_3$, and so there is no Type II atom in $\mathcal{D}_3$. Now $\gamma_4 = \{1, 2, 3\}$. We have to consider all the atoms that are subsets of $\tilde{X}_{\gamma_4} = \tilde{X}_1 \cup \tilde{X}_2 \cup \tilde{X}_3$, namely all the atoms in $\mathcal{D}_3$ because $N_3 - \gamma_4 = \emptyset$. For the atoms $1\bar{2}\bar{3}$, $\bar{1}2\bar{3}$, and $\bar{1}\bar{2}3$ in $\mathcal{D}_3$, since each of them is a subset of $\tilde{X}_i$ for only one $i \in \gamma_4$, the closed curve $C_4$ splits each of these atoms. For an atom A equal to $\bar{1}23$, $1\bar{2}3$, or $12\bar{3}$, we can check that $A \cap \tilde{X}_4 \in \mathcal{T}_I(G)$ and $A \cap \tilde{X}_4^c \in \mathcal{T}_{II}(G)$. Therefore, A belongs to (B2). For the atom 123, we can check that $1234 \in \mathcal{T}_I(G)$ and $123\bar{4} \in \mathcal{T}_{II}(G)$, so this atom also belongs to (B2). As such, the closed curve $C_4$ includes all these four atoms, and we can construct the information diagram in Fig. 16. It can readily to checked that this information diagram is equivalent to the one in Fig. 15 constructed in Example 17.*

**Example 19.** *Figs. 17 and 18 show a graph G and the corresponding information diagram, respectively. In light of Corollary 2, we see that $G^*(N_4)$ is the path $P_4$ defined in Section 5. Then we can use the information diagram for $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$ as $\mathcal{D}_4$ and add to it the closed curves $C_5$ and $C_6$ using the technique in Example 16 to obtain the information diagram in Fig. 18.*

# 7 Conclusion

The theory of *I*-Measure proves to be a very useful tool for characterizing full conditional independence structures and MRFs [15], because with the *I*-Measure, the fundamental set-theoretic structure of the problem is revealed. In this paper, we apply this tool to obtain three main results related to MRFs.

For an MRF represented by an undirected graph, a subfield is a subset of the random variables forming the MRF . We have determined the smallest undirected graph that can always represent a subfield as an MRF. This is our first main result. As an application of this result, we have obtained a necessary and sufficient condition for a subfield of a Markov tree to be also a Markov tree.

A Markov chain can be regarded a special case of an MRF. It was previously known that the $I$-Measure of a Markov chain is always nonnegative [11]. Here, we have proved that the $I$-Measure is nonnegative for every MRF represented by a given undirected graph if and only if the graph is a forest of paths, i.e., the Markov random field is a collection of independent Markov chains. This means that Markov chains are essentially the only MRFs such that the $I$-Measure is always nonnegative. This is our second main result. In the course of proving this result, we have obtained some interesting properties of the $I$-Measure pertaining to an MRF.

Our third main result is a nontrivial application of our first main result. In [11], a construction of an information diagram for a Markov chain was presented. By applying our first main result, we have developed a recursive approach for constructing information diagrams for MRFs. Such diagrams not only reveal the special structure of the $I$-Measure for an MRF, but they also are very useful for identifying information identities and inequalities pertaining to an MRF.

The work in our paper is based upon the the view that an MRF is a collection of full conditional mutual independencies. As such, some of our results can potentially be generalized for Markov structures beyond MRFs.

## A   Proof of Proposition 6

Let

$$
\begin{aligned}
G^*(V') &= (V', E'), \\
G^*(V'') &= (V'', E''),
\end{aligned}
$$

and

$$
(G^*(V'))^*(V'') = (V'', \tilde{E}'').
$$

Consider any $\{u, v\} \in E''$. By the definition of $G^*(V'')$, there exists a path between $u$ and $v$ in $G$ on which all the intermediate vertices are in $V - V''$. Now on this path, let $w_1, w_2, \cdots, w_k$ be the vertices in $V' - V''$ in the direction from $u$ to $v$. Then the vertices between $u$ and $w_1$, the vertices between $w_i$ and $w_{i+1}$ for $1 \le i \le k-1$, and the vertices between $w_k$ and $v$ are all in $(V - V'') - (V' - V'') = V - V'$, because $V'' \subset V' \subset V$. By the definition of $G^*(V')$, the edges $\{u, w_1\}$, $\{w_i, w_{i+1}\}$ for $1 \le i \le k - 1$, and $\{w_k, v\}$ are in $E'$. In other words, there is a path between $u$ and $v$ in $G^*(V')$ on which all the intermediate vertices are in $V' - V''$. Then by the definition of $(G^*(V'))^*(V'')$, $\{u, v\} \in \tilde{E}''$. This proves that $E'' \subset \tilde{E}''$.

On the other hand, consider any $\{u, v\} \in \tilde{E}''$. By the definition of $(G^*(V'))^*(V'')$, there exists a path between $u$ and $v$ in $G^*(V')$ on which all the intermediate vertices are in $V' - V''$. Let the edges on this path be $\{u, w_1\}$, $\{w_1, w_2\}$, $\cdots$, $\{w_{k-1}, w_k\}$, and $\{w_k, v\}$, where $w_i, 1 \le i \le k$ are in $V' - V''$. Since $\{u, w_1\} \in E'$, by the definition of $G^*(V')$, there exists a path between $u$ and $w_1$ on which all the intermediate nodes are in $V - V'$. Similarly, there exists a path between $w_i$ and $w_{i+1}$ for $1 \le i \le k$, and a path between $w_k$ and $v$, on which all the intermediate nodes are in $V - V'$. Thus there exists a path between $u$ and $v$ in $G$ on which all the intermediate nodes are in $(V - V') \cup (V' - V'') = V - V''$, because $V'' \subset V' \subset V$. Then by the definition of $G^*(V'')$, $\{u, v\} \in E''$. This proves that $\tilde{E}'' \subset E''$.

Hence, we conclude that $\tilde{E}'' = E''$, i.e., $(G^*(V'))^*(V'') = G^*(V'')$. The proposition is proved.

## B  Verification of $\mu^*$ in the Proof of Theorem 14

In this appendix, we verify that $\mu^*$ as prescribed by (20) satisfies (19). First, for $i \in V$, consider

$$
\begin{aligned}
\mu^*(\tilde{X}_i) &= \mu^*\left( \bigcup_{S \subset V - \{i\}} \left( \tilde{X}_i \cap \tilde{X}_S - \tilde{X}_{V - S - \{i\}} \right) \right) \\
&= \sum_{S \subset V} \mu^*\left( \tilde{X}_i \cap \tilde{X}_S - \tilde{X}_{V - S - \{i\}} \right).
\end{aligned}
$$

From (20), we see that $\mu^*(\cdot)$ in the above summation vanishes if $|S| \le n - 3$, and so

$$
\begin{aligned}
\mu^*(\tilde{X}_i) &= \sum_{j \ne i} \mu^*\left( \bigcap_{k \ne j} \tilde{X}_k - \tilde{X}_j \right) + \mu^*\left( \tilde{X}_0 \cap \tilde{X}_1 \cap \cdots \cap \tilde{X}_{n-1} \right) \\
&= (n - 1) \cdot 1 - (n - 2) \\
&= 1.
\end{aligned}
\tag{27}
$$

This verifies (19) for the case $|B| = 1$. Next, for $0 \le i < j \le n - 1$, consider

$$
\begin{aligned}
\mu^*(\tilde{X}_i \cap \tilde{X}_j) &= \sum_{S \subset V-\{i,j\}} \mu^* \left( \tilde{X}_i \cap \tilde{X}_j \cap \left( \bigcap_{k \in S} \tilde{X}_k \right) - \left( \bigcup_{l \in V-\{i,j\}-S} \tilde{X}_l \right) \right) \\
&= \sum_{\substack{S \subset V-\{i,j\} \\ |S| \ge n-3}} \mu^* \left( \tilde{X}_i \cap \tilde{X}_j \cap \left( \bigcap_{k \in S} \tilde{X}_k \right) - \left( \bigcup_{l \in V-\{i,j\}-S} \tilde{X}_l \right) \right) \\
&= \sum_{m \ne i,j} \mu^* \left( \bigcap_{r \in V-\{m\}} \tilde{X}_r - \tilde{X}_m \right) + \mu^* \left( \bigcap_{s=1}^{n} \tilde{X}_s \right) \\
&= (n-2) \cdot 1 - (n-2) \\
&= 0. \tag{28}
\end{aligned}
$$

It follows from (27) and (28) that

$$
\begin{aligned}
\mu^*(\tilde{X}_i \cup \tilde{X}_j) &= \mu^*(\tilde{X}_i) + \mu^*(\tilde{X}_j) - \mu^*(\tilde{X}_i \cap \tilde{X}_j) \tag{29} \\
&= 1 + 1 - 0 \tag{30} \\
&= 2. \tag{31}
\end{aligned}
$$

This verifies (19) for the case $|B| = 2$. Now for $1 \le i < j < k \le n$, consider

$$
\begin{aligned}
&\mu^*(\tilde{X}_k - (\tilde{X}_i \cup \tilde{X}_j)) \\
&= \sum_{S \subset V-\{i,j,k\}} \mu^* \left( \tilde{X}_k \cap \left( \bigcap_{l \in S} \tilde{X}_l \right) - \left( \tilde{X}_i \cup \tilde{X}_j \cup \bigcup_{m \in V-\{i,j,k\}-S} \tilde{X}_m \right) \right).
\end{aligned}
$$

In the above, since $|k \cup S| \le n-2$, we see from (20) that every term in the above summation vanishes, and so

$$
\mu^*(\tilde{X}_k - (\tilde{X}_i \cup \tilde{X}_j)) = 0. \tag{32}
$$

Finally, consider $B \subset V$ such that $3 \le |B| \le n$ and let $i, j$ be two arbitrary elements of $B$. Then in light of (31) and (32), we have

$$
\begin{aligned}
\mu^*(\tilde{X}_B) &= \mu^*(\tilde{X}_i \cup \tilde{X}_j) + \mu^* \left( \tilde{X}_{B-\{i,j\}} - (\tilde{X}_i \cup \tilde{X}_j) \right) \\
&\le \mu^*(\tilde{X}_i \cup \tilde{X}_j) + \sum_{k \in B-\{i,j\}} \mu^*(\tilde{X}_k - (\tilde{X}_i \cup \tilde{X}_j)) \\
&= 2 + 0 \\
&= 2,
\end{aligned}
$$

41

where the inequality above is justified by the union bound because $\mu^*$ is nonnegative on all the atoms in $\tilde{X}_{B-\{i,j\}} - (\tilde{X}_i \cup \tilde{X}_j)$ (cf. (20)). On the other hand, we have

$$
\begin{aligned}
\mu^*(\tilde{X}_B) &= \mu^*(\tilde{X}_i \cup \tilde{X}_j) + \mu^* \left( \tilde{X}_{B-\{i,j\}} - (\tilde{X}_i \cup \tilde{X}_j) \right) \\
&\geq \mu^*(\tilde{X}_i \cup \tilde{X}_j) \\
&= 2,
\end{aligned}
$$

again because $\mu^*$ is nonnegative on all the atoms in $\tilde{X}_{B-\{i,j\}} - (\tilde{X}_i \cup \tilde{X}_j)$. Therefore, $\mu^*(\tilde{X}_B) = 2$, verifying (19) for the case $3 \leq |B| \leq n$.

## C    Proof of Lemmas 4 and 5

In this appendix, we prove Lemmas 4 and 5 via the following lemma.

**Lemma 6.** *Let $G = (V, E)$ be a connected undirected graph and $B = \{v \in V : s(\{v\}) = 1\}$. Then*

    *a) for any $k \in V - B$, we have $B \cap J_i \neq \emptyset$ for all $1 \leq i \leq s_k$, where $J_1, J_2, \cdots, J_{s_k}$ ($s_k \geq 2$) are the components of $G\backslash\{k\}$;*

    *b) $s(R) > 1$ for all nonempty subset $R$ of $V - B$.*

**Proof** We assume that $B \neq V$, since otherwise $V - B = \emptyset$ and the lemma has no assertion.

We first prove a). Let $k \in V - B$, and by the definition of $B$, we have $s_k \geq 2$. Consider any spanning tree $T$ of $G$. Note that $T$ must contain at least one edge connecting $k$ and each $J_i$ ($1 \leq i \leq s_k$) because $G$ is connected. For any fixed $i$, consider such an edge and call it $e$. Upon removing $e$, $T$ is disconnected with one component being a subtree containing $k$ and the other component being a subtree not containing $k$. For the latter subtree, all the vertices are in $J_i$, otherwise there exists an edge connecting $J_i$ and $J_{i'}$ ($i' \neq i$), which is a contradiction because $J_1, J_2, \cdots, J_{s_k}$ are the components of $G\backslash\{k\}$. Then this subtree must have at least one leaf in $J_i$, say $l$. Note that $\{l\}$ is not a cutset in $T$ and hence not a cutset in $G$. Therefore $B \cap J_i \neq \emptyset$, proving a).

We now prove b). Consider any nonempty subset $R$ of $V - B$ and fix $k \in R$. Since $R \subset V - B$, we have $R \cap B = \emptyset$. By a), $B \cap J_i \neq \emptyset$ for all $1 \leq i \leq s_k$, with $s_k \geq 2$. Then the vertices in $B$ are not all

42

connected in $G\backslash\{k\}$, and hence not all connected in $G\backslash R$ because $k \in R$ and $R \cap B = \emptyset$. This implies that $G\backslash R$ is not connected, or $s(R) > 1$. The lemma is proved. $\qquad\square$

Lemmas 4 and 5 can now be obtained as follows. In Theorem 11, $U_A$ is a Type I atom, and so $G\backslash U_A$ is connected. The same holds for Lemmas 4 and 5. Lemma 4 is trivial for $|U_A| = n - 2$. For $|U_A| < n - 2$, it can be obtained by applying Part a) of Lemma 6 to $G\backslash U_A$. Finally, by applying Part b) of Lemma 6 to $G\backslash U_A$ with $R = W - S$, we obtain Lemma 5.

## Acknowledgment

## References

[1] Hu Guo Ding, "On the amount of information," *Teor. Veroyatnost. i Primenen.*, 4: 447-455, 1962 (in Russian).

[2] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," (unpublished) 1971.

[3] F. Spitzer, "Random fields and interacting particle systems," M.A.A. Summer Seminar Notes, 1971.

[4] J. Moussouris, "Gibbs and Markov random systems with constraints," *J. Stat. Physics*, vol. 10, pp. 11-33, 1974.

[5] C. Preston, *Random Fields*, Springer-Verlag, 1974.

[6] R. Kindermann and J. L. Snell, *Markov Random Fields and their Applications,* American Mathematical Society, 1980.

[7] S. Asmussen and D. Edwards, "Collapsibility and response variables in contingency tables," *Biometrika*, vol. 70, no. 3, pp. 567-578, Dec 1983.

[8] J. Pearl and A. Paz, "GRAPHOIDS: A graph-based logic for reasoning about relevance relations," Technical Report 850038 (R-53-L), Cognitive Systems Laboratory, University of California, Los Angeles, 1985. Short version in *Advances in Artificial Intelligence – II*, ed. B. Du Boulay, D. Hogg, and L. Steels, pp. 357-363, Amsterdam: North Holland, 1987.

[9] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466-474, May 1991.

[10] F. M. Malvestuto, "A unique formal system for binary decompositions of database relations, probability distributions, and graphs," *Information Sciences* **59** (1992) 21-52; with Comment by F. M. Malvestuto and M. Studený, *Information Sciences* **63** (1992), 1-2.

[11] T. Kawabata and R. W. Yeung, "The structure of the *I*-Measure of a Markov chain," *IEEE Transactions on Information Theory,* vol. 38, pp. 1146-1149, May 1992.

[12] D. Geiger and J. Pearl, "Logical and algorithmic properties of conditional in- dependence and graphical models," *Annals of Statistics*, vol. 21, no. 4, pp. 2001-2021, 1993.

[13] S. L. Lauritzen, *Graphical Models*, Oxford Science Publications, Oxford, 1996.

[14] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1924-1934, Nov 1997.

[15] R. W. Yeung, T. T. Lee, and Z. Ye, "Information-theoretic characterizations of conditional mutual independence and Markov random fields," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1996-2011, Jul 2002.

[16] F. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2012-2021, Jul 2002.

[17] L. E. Doyle, A. C. Kokaram, S. J. Doyle, and T. K. Forde, "Ad hoc networking, Markov random fields, and decision making," *IEEE Signal Processing Magazine*, vol. 23, no. 5, pp. 63-73, Sept 2006.

[18] R. W. Yeung, *Information Theory and Network Coding*, Springer, 2008.

[19] S.-e. Jeon and C. Ji, "Randomized and distributed self-configuration of wireless networks: Two-layer Markov random fields and near-optimality," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4859-4870, Sept 2010.

[20] Y. O. Yazir, R. Farahbod, A. Guitouni, S. Ganti, and Y. Coady, "Adaptive routing in mobile ad hoc networks based on decision aid approach," 8th ACM International Workshop on Mobility Management and Wireless Access, Bodrum, Turkey, Oct 17-18, 2010.

[21] A. Blake, P. Kohli, and C. Rother (Ed.), *Markov Random Fields for Vision and Image Processing*, MIT Press, 2011.

[22] T. A. B. Snijders, "Statistical models for social networks," *Annual Review of Sociology*, vol. 37, pp. 131-153, 2011.

[23] T. Wang, H. Krim, and Y. Viniotis, "A generalized Markov graph model: Application to social network analysis," *IEEE J. Selected Topics in Signal Processing*, vol. 7, pp. 318-332, 2013.

[24] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, pp. 80-90, 2014.

[25] M. Studený, *Probabilistic Conditional Independence Structures*, Springer-Verlag 2005.

[26] K. Sadeghi, "Stable mixed graphs," *Bernoulli* 19, pp. 2330-2358, 2013.

[27] K. Sadeghi, "Marginalization and conditioning for LWF chain graphs," *Annals of Statistics*, vol. 44, no. 4, pp. 1792-1816, 2016.
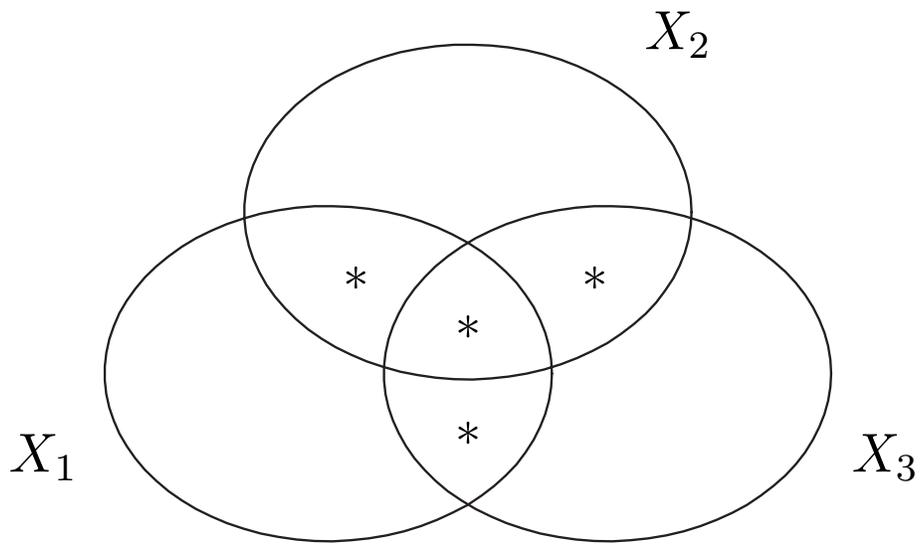
Figure 1: The information diagram for Example 3.
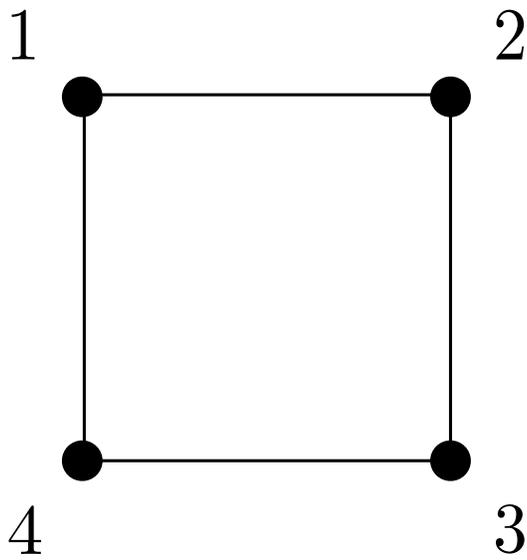
Figure 2: The graph $G$ in Example 6.
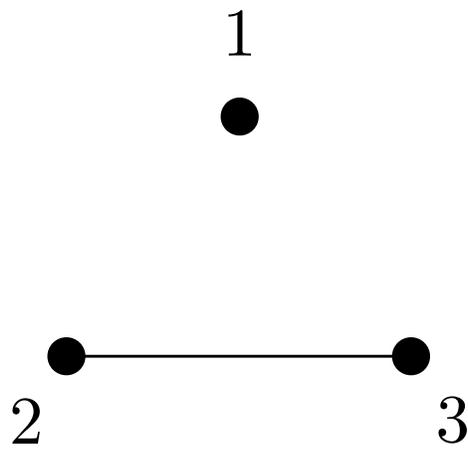
Figure 3: The graph $\hat{G}$ corresponding to the set $\mathcal{A}_{\mathrm{II}}$ in Example 7.
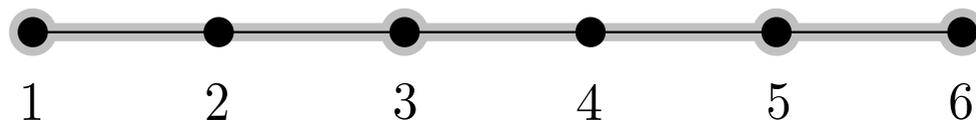
Figure 4: The graphs $G$ (black) and $G^*(V')$ (grey) in Example 9, with $V'\{1, 3, 5, 6\}$.
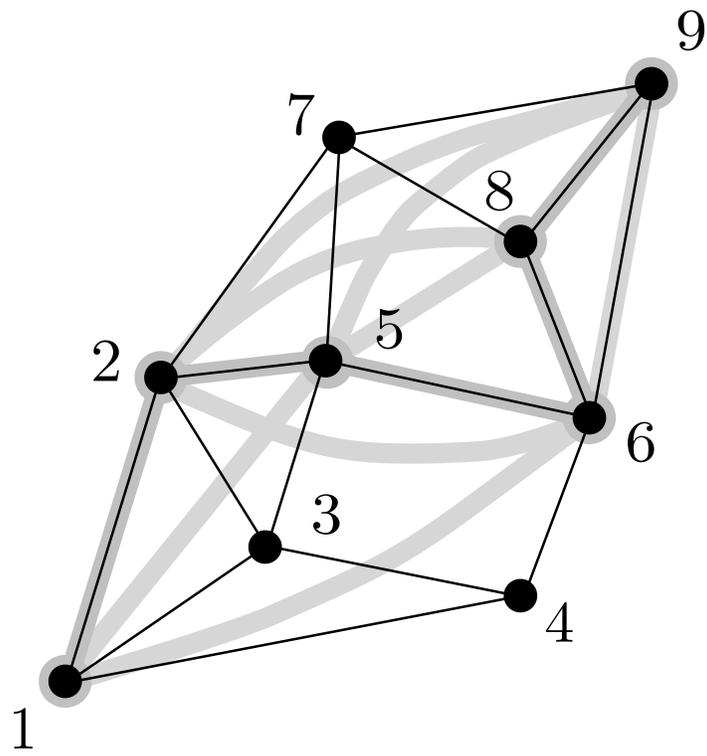
Figure 5: The graphs $G$ (black) and $G^*(V')$ (grey) in Example 10, with $V' = \{1, 2, 5, 6, 8, 9\}$.
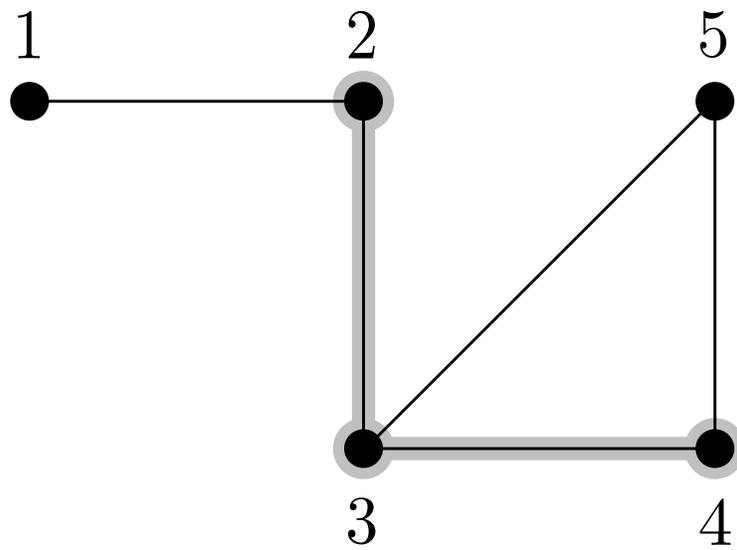
Figure 6: The graphs $G$ (black) and $G^*(V')$ (grey) in Example 11, with $V' = \{2, 3, 4\}$.
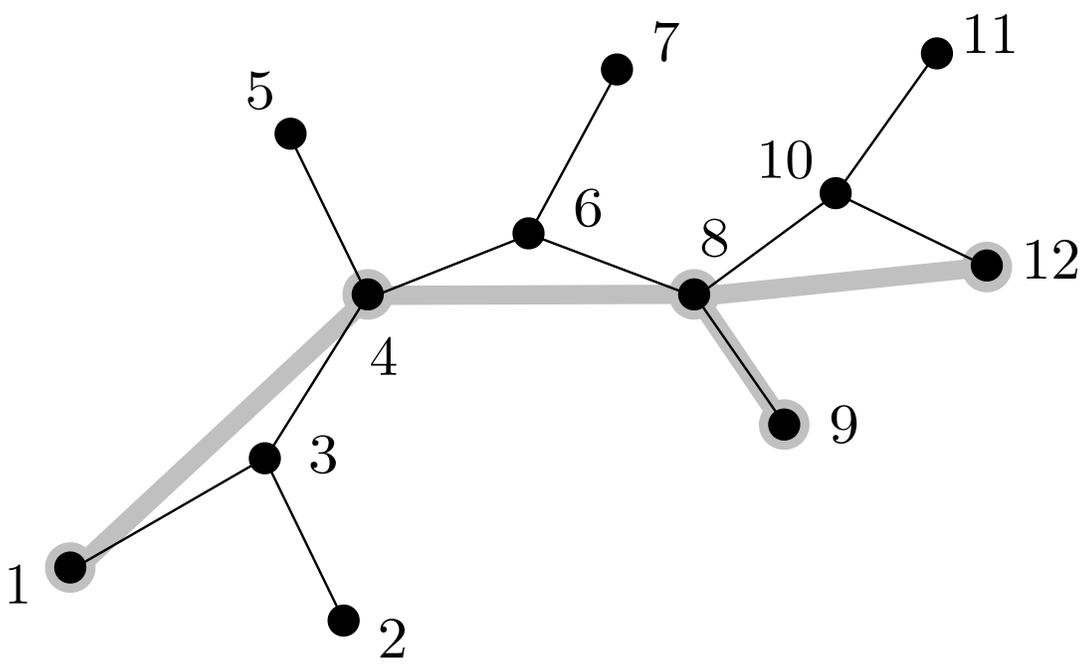
Figure 7: The trees $G$ (black) and $G^*(V')$ (grey) in Example 13, with $V' = \{1, 4, 8, 9, 12\}$.
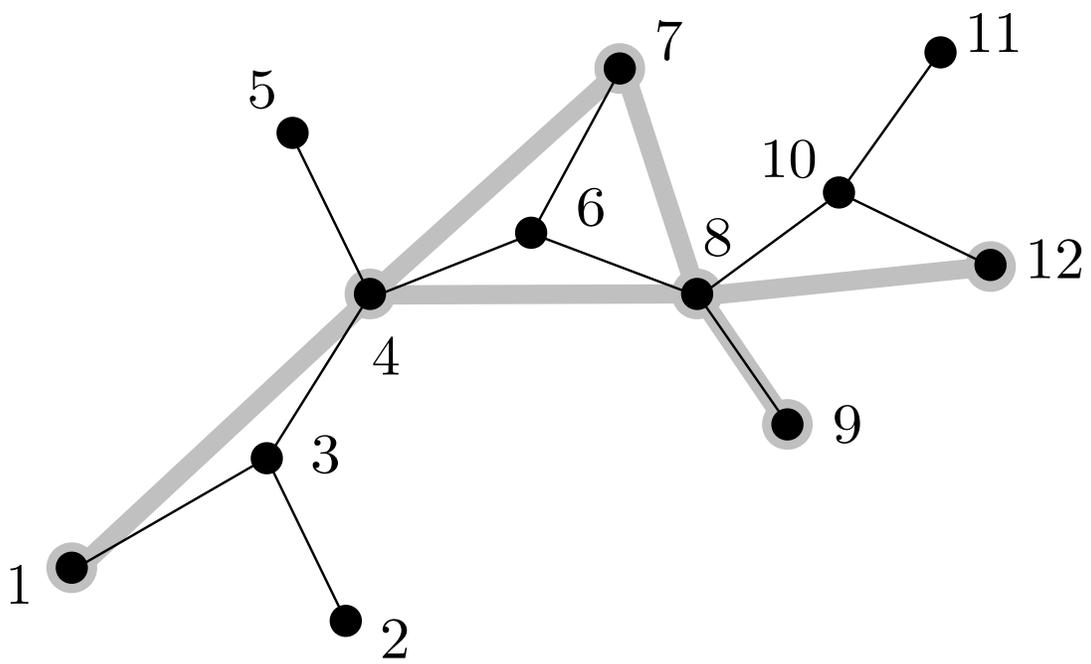
Figure 8: The tree $G$ (black) and the graph $G^*(V')$ (grey) in Example 13, with $V' = \{1, 4, 7, 8, 9, 12\}$.
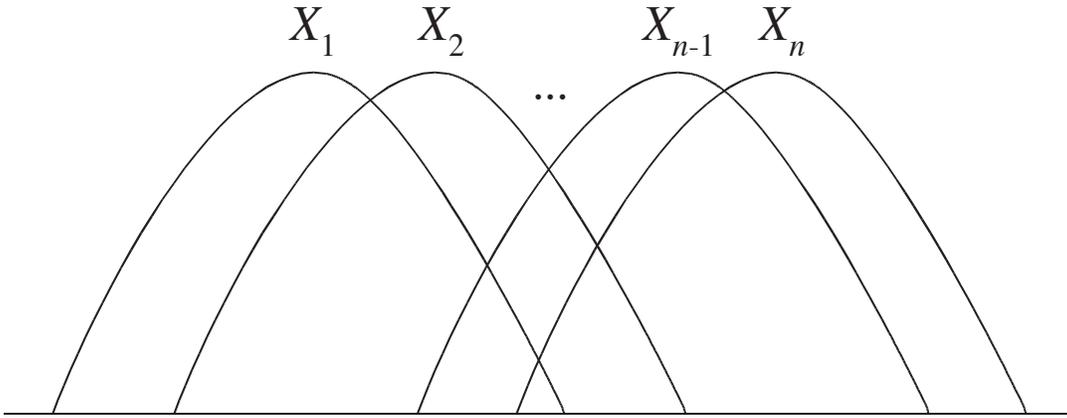
Figure 9: The information diagram for the Markov chain $X_1 \to X_2 \to \cdots \to X_n$.
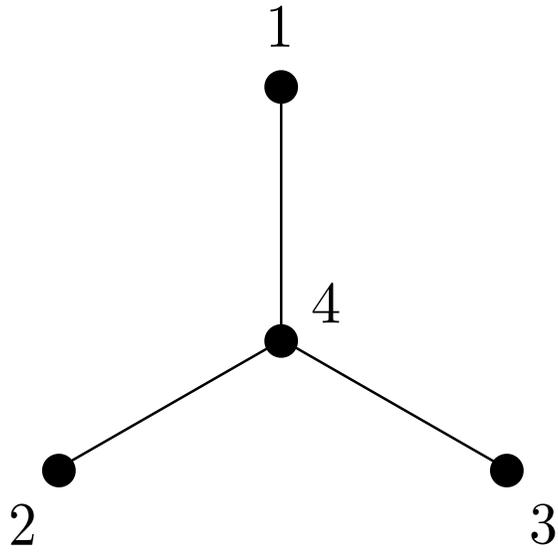
Figure 10: The "star" representing the Markov tree in Example 14.
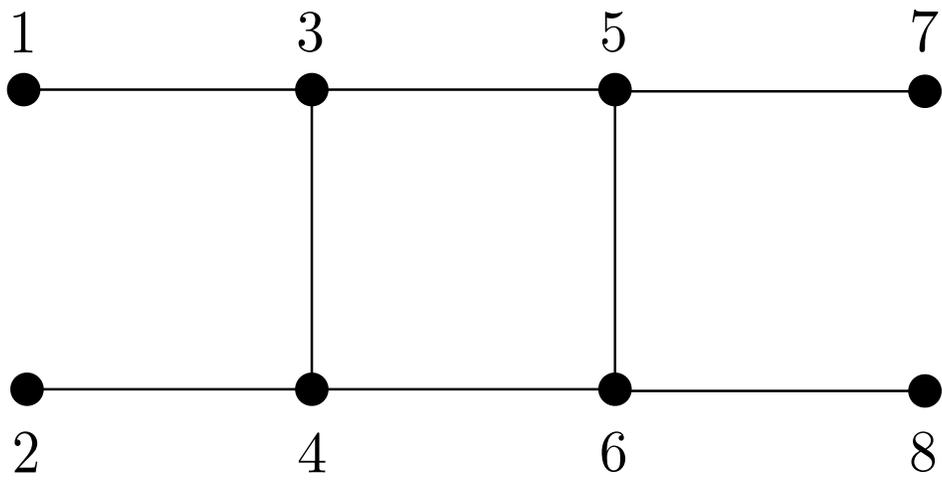
Figure 11: The graph in Example 15.
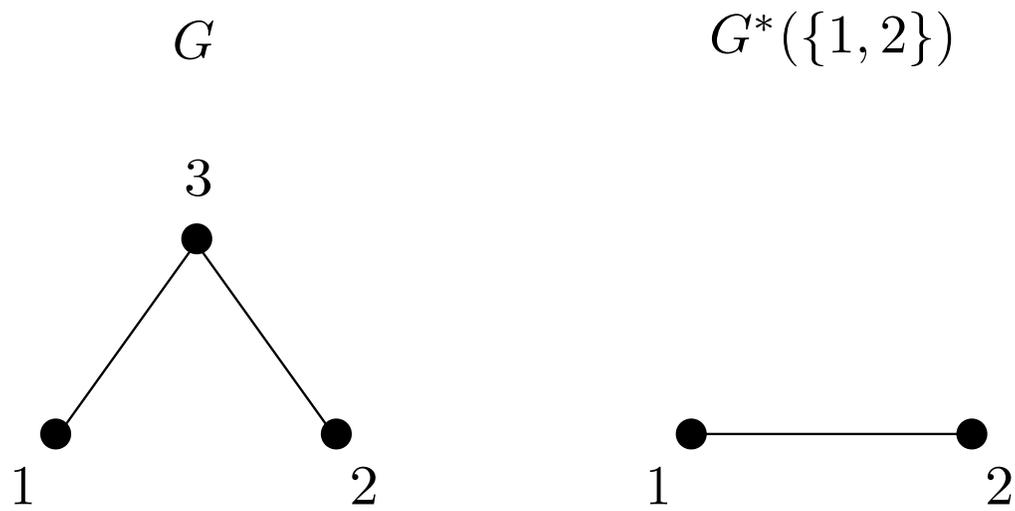
Figure 12: In this example, $A = 12 \in \mathcal{T}_{\mathrm{I}}(G^*(\{1,2\}))$, $A \cap \tilde{X}_3 = 123 \in \mathcal{T}_{\mathrm{I}}(G)$, and $A \cap \tilde{X}_3^c = 12\bar{3} \in \mathcal{T}_{\mathrm{II}}(G)$. Therefore, $A$ belongs to (B2).
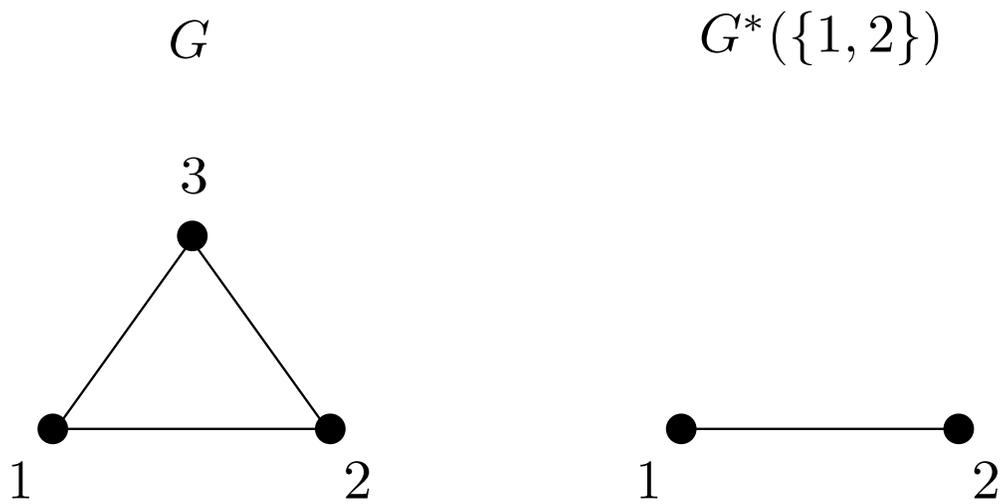
Figure 13: In this example, $A = 12 \in \mathcal{T}_I(G^*(\{1,2\}))$, $A \cap \tilde{X}_3 = 123 \in \mathcal{T}_I(G)$, and $A \cap \tilde{X}_3^c = 12\bar{3} \in \mathcal{T}_I(G)$. Therefore, $A$ belongs to (B1).
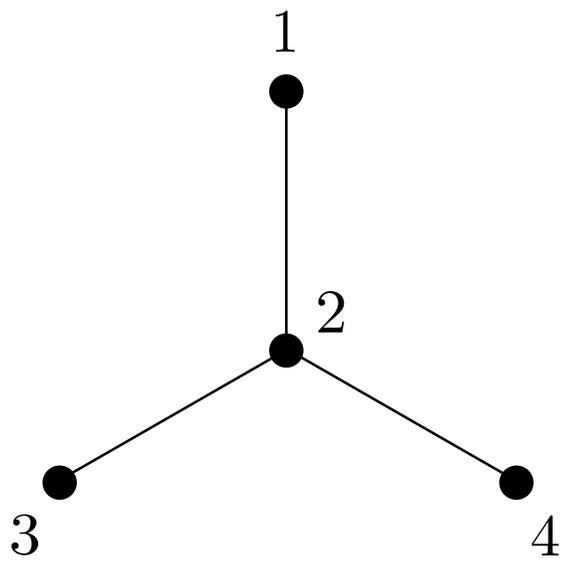
Figure 14: The graph $G$ in Example 17.
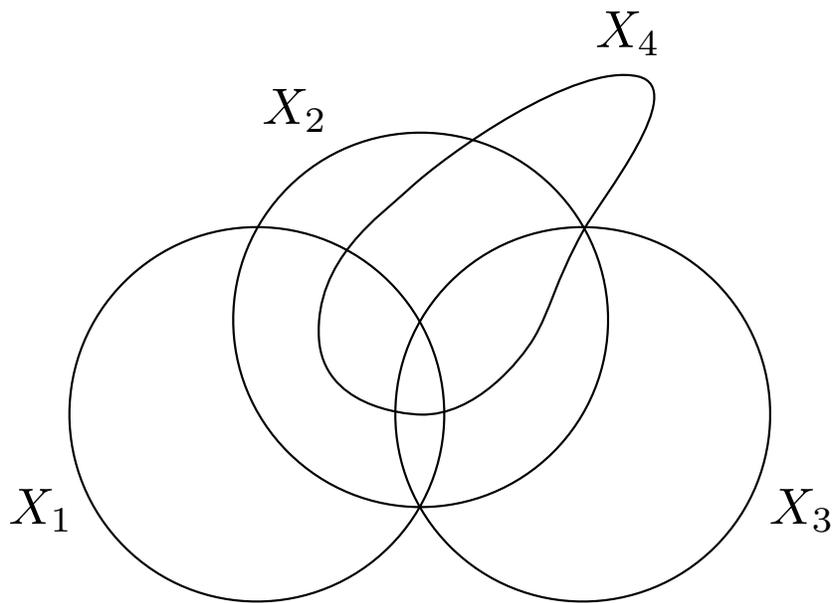
Figure 15: The information diagram in Example 17.
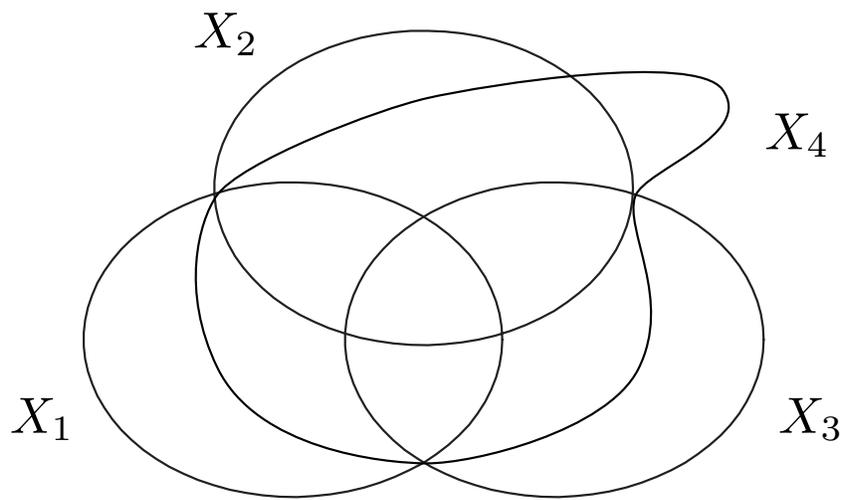
Figure 16: The information diagram in Example 18.
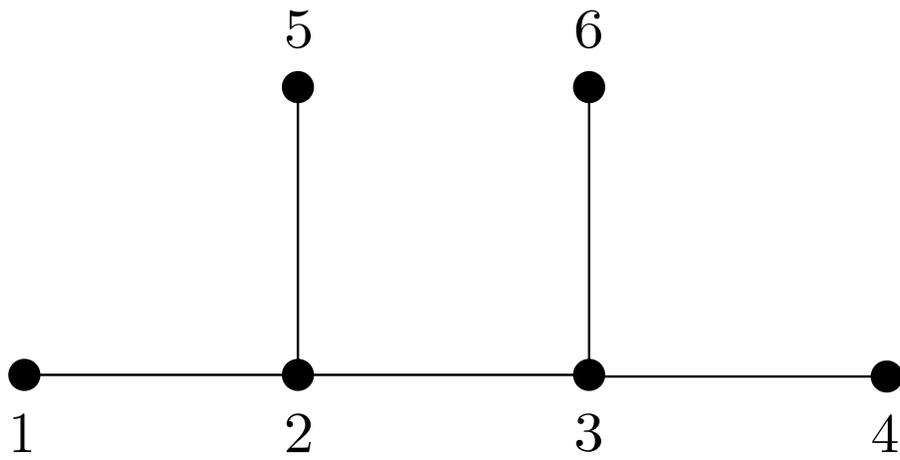
Figure 17: The graph $G$ in Example 19.
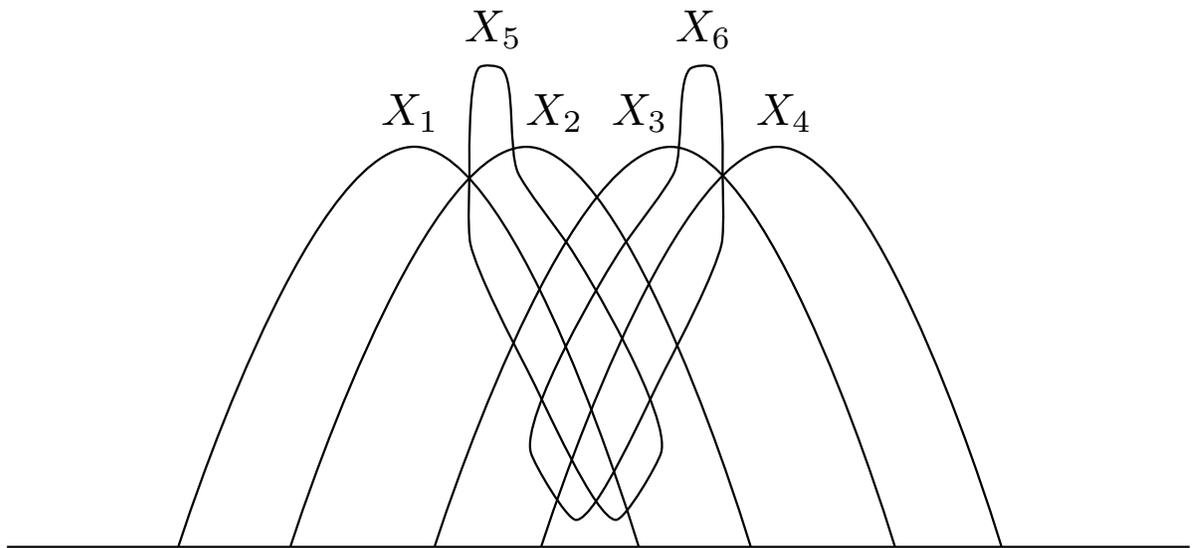
Figure 18: The information diagram in Example 19.