Properties of Kikuchi approximations constructed from clique based decompositions

Roberto Santana, Pedro Larrañaga, Jose A. Lozano
Intelligent System Group
Department of Computer Science and Artificial Intelligence
University of the Basque Country
P.O. Box 649, 20080 San Sebastián - Donostia, Spain

Kikuchi approximations constructed from clique-based decompositions can be used to calculate suitable approximations of probability distributions. They can be applied in domains such as probabilistic modeling, supervised and unsupervised classification, and evolutionary algorithms. This paper introduces a number of properties of these approximations. Pairwise and local Markov properties of the Kikuchi approximations are proved. We prove that, even if the global Markov property is not satisfied in the general case, it is possible to decompose the Kikuchi approximation in the product of local Kikuchi approximations defined on a decomposition of the graph. Partial Kikuchi approximations are introduced. Additionally, the paper clarifies the place of clique-based decompositions in relation to other techniques inspired by methods from statistical physics, and discusses the application of the results introduced in the paper for the conception of Kikuchi approximation learning algorithms.

1 Introduction

Belief propagation [1, 25] is a well-known technique used in statistical inference to obtain *a posteriori* marginal probabilities in graphical models. Generalized belief propagation enables the class of models to be extended where these inference algorithms can be applied. Recent work on generalized belief propagation methods [38] have revealed the applicability that results achieved in the statistical physical domain, in the approximation of energy and entropy measures have in the machine learning domain. One of the contributions from the field of statistical physics to inference algorithms comes from the use of region-based decompositions, like the Bethe [6] and Kikuchi [14, 21] approximations, in the context of generalized belief propagation.

Basically, a region-based decomposition can be seen as a function defined on the variables associated to the vertices of a graph. The global function is formed by the composition of local subfunctions defined in those variables grouped in each of the regions. Common ways of composition are the sum and product of the local functions. For instance, in the free energy approximation, regions serve as the basic units to define the local energies, which are combined to give the global free energy function. Region-based decompositions can be used for the approximation of other measures; in this paper we use them to calculate suitable approximations of probability distributions. In this context, an essential question is how to determine a convenient region-based decomposition that maximizes the accuracy of the approximation. There are algorithms that serve to calculate these decompositions.

One particular case of such algorithms is the cluster variation method (CVM) [14, 21], originally introduced to obtain Kikuchi approximations of the free energy in statistical physics. Starting from an initial set of regions defined on a graph, the CVM determines a way to obtain a whole set of regions where the free energy is decomposed. The CVM does not specify any particular choice for the initial regions. Nevertheless, the Kikuchi approximation clearly depends on this choice.

Some approaches that try to cope with the problem of selecting an approppriate set of initial regions have been published. This interest highlights the relevance of this problem for the field. The methods introduced have been proposed in the context of belief propagation algorithms [32], and are also related to work done on structural mean field methods [34, 35]. The clique-based decomposition of the graph has been introduced in [26] as a particular way of selecting the initial regions for the CVM. The method was used for optimization by means of estimation of distribution algorithms (EDAs) [17] that employ Kikuchi approximations of the probability.

In this paper, we focus on the clique-based decomposition as the method for selecting the initial regions of the Kikuchi approximation. We provide a number of theoretical properties that are satisfied by this approximation. The properties proved in this paper are useful for tasks such as evaluating the quality of the approximation, designing of algorithms to learn Kikuchi approximations from data, and designing sampling algorithms. Our work can be seen in the more general context of setting the theoretical basis for the application of Kikuchi approximations as a practical approximation method in machine learning. In order to achieve this purpose, we investigate Markov and decomposability properties of the Kikuchi approximation constructed from clique-based decompositions of the graphs.

Additionally, we pay attention to recent developments in the application of region-based decompositions in machine learning. Our analysis intends to throw some light on the way these approximations are being conceived and applied in the field. We clarify the place of clique-based decompositions in relation to other techniques inspired by methods from statistical physics.

The rest of this paper is presented as follows. In Section 2, an introduction to clique-based decompositions and the Kikuchi approximation of the probability distribution are presented. Section 3 proves that the clique-based Kikuchi approximation satisfies the local and pairwise properties with respect to the independence graph. We show that the global Markov property is not satisfied. Section 4 shows that the Kikuchi approximation can be factorized in the irreducible components of the graph. In Section 5 we argue that our work is part of a current research trend that benefits from the cross-fertilization between machine learning and statistical physics. Section 6 discusses possible applications and presents the conclusions of the paper.

2 Kikuchi approximation: recapitulation

Kikuchi approximations of the energy [14] are region-based decompositions of the energy that satisfy certain constraints. The *Kikuchi approximation of a probability distribution from a clique-based decomposition of an independence graph* [26] is a particular type of factorization that use marginal distributions. The marginals in the factorization are completely determined by the independence graph. Given this graph, the clique-based decomposition is formed by the maximal cliques of the graphs and their intersections. All these cliques are called regions.

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a vector of discrete random variables. We will use $\mathbf{x} = (x_1, \ldots, x_n)$ to denote an assignment to the variables. S will denote a set of indices in $N = \{1, \ldots, n\}$, and $\mathbf{X}_S$ (respectively $\mathbf{x}_S$) a subset of the variables of $\mathbf{X}$ (respectively a subset of values of $\mathbf{x}$) determined by the indices in S. We will work with positive probability distributions denoted by $p(\mathbf{x})$. Similarly, $p(\mathbf{x}_S)$ will denote the marginal probability distribution for $\mathbf{X}_S$. We use $p(x_i \mid x_j)$ to denote the conditional probability distribution of X_i given $X_j = x_j$.

An undirected graph $G = (V, E)$ is defined by a non-empty set of vertices V, and a set of edges E. An edge between vertices i and j will be represented by $i \sim j$. Given a probability distribution $p(\mathbf{x})$, its independence graph is a graph $G = (V, E)$ that associates one vertex with every variable of $\mathbf{X}$, and where two vertices are connected if the corresponding variables are conditionally dependent given the rest of the variables.

We define a region R of the independence graph $G = (V, E)$ of a probability distribution $p(\mathbf{x})$ as a subset of V. A graph region-based decomposition $(\mathcal{R}, U)$, is a set of regions $\mathcal{R}$ that covers V, and an associated set of *overcounting numbers* U which is formed by assigning one overcounting number c_R for each $R \in \mathcal{R}$. c_R will always be an integer, and might be zero or negative for some R. There are different methods that find region-based decompositions [6, 14, 2, 36], among them the CVM that learns Kikuchi approximations. In the CVM, $\mathcal{R}$ is formed recursively by an initial set of regions $\mathcal{R}_0$ such that each node is in at least one region of $\mathcal{R}_0$, and any other region in $\mathcal{R}$ is the intersection of one or more of the regions in $\mathcal{R}$. The set of regions $\mathcal{R}$ is closed under the intersection operation, and can be ordered as a partially ordered set [21].

To be valid, a decomposition must satisfy a number of constraints that relate $\mathcal{R}$ and U. Inspired by the work by Yedidia et al. [37] we call this sub-problem as that of *finding a valid region-based decomposition of a graph.* A set of regions $\mathcal{R}$ and overcounting numbers U give a valid region-based graph decomposition [37] when for every variable

X_i:

$$\sum_{\substack{R\in\mathcal{R}\\ X_i\subseteq\mathbf{X}_R}} c_R = 1 \tag{1}$$

Equation (1) states that for any variable X_i the sum of the overcounting numbers of regions that contain X_i is 1. Equation (2) can be obtained extending the previous constraint to every subset of variables $\mathbf{X}_S$ in the following way:

$$\sum_{\substack{R\in\mathcal{R}\\ \mathbf{X}_S\subseteq\mathbf{X}_R}} c_R = 1 \tag{2}$$

where the sum of the overcounting numbers of regions that contain $\mathbf{X}_S$ is also 1. Pakzad and Anantharam [24] call conditions represented by equations (1) and (2) as the balanced and totally balanced conditions respectively. These authors prove that any collection of regions that is closed under intersection fulfills the balanced and totally balanced conditions [24].

We will apply the CVM making a particular choice of the initial regions. We will form set $\mathcal{R}_0$ by taking one region for each maximal clique in G. As a result, all the regions $R \in \mathcal{R}$ will be cliques because they are the intersection of two or more cliques. We call this type of region-based decomposition of undirected graphs a *clique-based decomposition.*

We define the Kikuchi approximation of the probability distribution $p(\mathbf{x})$ associated with a clique-based decomposition, $k(\mathbf{x})$ as:

$$k(\mathbf{x}) = \prod_{R\in\mathcal{R}} p(\mathbf{x}_R)^{c_R}, \tag{3}$$

where $\mathcal{R}$ comes from a clique-based decomposition and the overcounting numbers c_R are calculated using the following recursive formula:

$$c_R = 1 - \sum_{\substack{S\in\mathcal{R}\\ R\subset S}} c_S \tag{4}$$

where c_S is the overcounting number of any region S in $\mathcal{R}$ such that S is a superset of R. c_R values corresponding to the initial maximal cliques are equal to 1. If c_R is different from zero, the region is included in the clique-based decomposition.

From now on, when we refer to a Kikuchi approximation, we imply a Kikuchi approximation obtained from a clique-based decomposition.

Example 1 *Kikuchi approximation corresponding to the independence graph shown in Figure 1.*

$$k(\mathbf{x}) = \frac{p(x_A,x_T)p(x_T,x_E)p(x_E,x_X)p(x_E,x_L)p(x_L,x_S)p(x_S,x_B)p(x_B,x_D)p(x_E,x_D)}{p(x_T)p(x_E)^3p(x_L)p(x_S)p(x_B)p(x_D)} \tag{5}$$

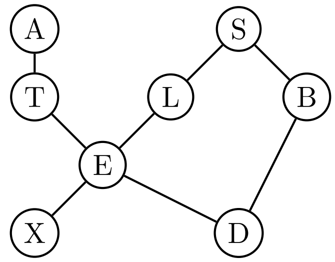

Figure 1: Undirected graph associated with the Asia network

In example 1, the Kikuchi approximation of the probability corresponding to the graph is included. Notice that the factors of the decomposition correspond to the eight maximal cliques of the graph, and their overlappings. The factor corresponding to variable X_E has an overcounting number 3 because it is included in four original cliques.

3 Markov properties of the Kikuchi approximation

Graphical models associated with probability distributions display a number of conditional and marginal independence properties that are stated by the Markov properties defined on the graph. These independence properties can be used for more efficient storing and sampling of probability distributions. Therefore, it is an important question to investigate which information about the properties of the Kikuchi approximation can be extracted from the graphical model where it is defined. In this section we show that certain independence properties of the Kikuchi approximation can be deduced from the graph structure.

First, we prove that the Kikuchi approximation satisfies the local and pairwise Markov properties with respect to the independence graph. It is also proved that the global Markov property is not fulfilled. Instead, we present the Kikuchi decomposition property that will be an important building block of the decomposability results shown in the next section.

3.1 Notation

Given an undirected graph G, the boundary and closure of a set of variables (respectively a set of values) are respectively defined as:

Definition 1 *The boundary of a set of vertices, $\mathbf{X}_S \subseteq \mathbf{X}$, is the set of vertices in $\mathbf{X} \setminus \mathbf{X}_S$ that neighbors at least one vertex in $\mathbf{X}_S$. The boundary of $\mathbf{X}_S$ is denoted $bd(\mathbf{X}_S)$.*

Definition 2 *The closure of a set of vertices, $\mathbf{X}_S \subseteq \mathbf{X}$ is the set of vertices $cl(\mathbf{X}_S) = \mathbf{X}_S \cup bd(\mathbf{X}_S)$.*

The marginal and conditional functions of the Kikuchi approximation are defined as:

$$k(\mathbf{x}_S) = \sum_{\mathbf{x}' | \mathbf{x}'_S = \mathbf{x}_S} k(\mathbf{x}') \tag{6}$$

$$k(\mathbf{x}_A \mid \mathbf{x}_B) = \frac{k(\mathbf{x}_{\{A,B\}})}{k(\mathbf{x}_B)} \tag{7}$$

where $\{A, B\}$ is a simplified notation for $\{A \cup B\}$. As $k(\mathbf{x})$ is not necessarily a probability distribution, neither are $k(\mathbf{x}_S)$ and $k(\mathbf{x}_A \mid \mathbf{x}_B)$. Nevertheless, they can respectively be used as approximations of $p(\mathbf{x}_S)$ and $p(\mathbf{x}_A \mid \mathbf{x}_B)$.

Given any region A, $K(\mathbf{x}, A) = \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_A \neq \emptyset}} p(\mathbf{x}_R)^{c_R}$, and $\bar{K}(\mathbf{x}, A) = \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_A = \emptyset}} p(\mathbf{x}_R)^{c_R}$, then $k(\mathbf{x})$ can be expressed in the following way:

$$\begin{aligned} k(\mathbf{x}) &= \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_A \neq \emptyset}} p(\mathbf{x}_R)^{c_R} \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_A = \emptyset}} p(\mathbf{x}_R)^{c_R} \\ &= K(\mathbf{x}, A) \bar{K}(\mathbf{x}, A) \end{aligned} \tag{8}$$

where $K(\mathbf{x}, A)$ and $\bar{K}(\mathbf{x}, A)$ have been introduced for notational convenience, to represent $k(\mathbf{x})$ more concisely.

The non-standard notation $\sim \{\mathbf{x}_S\}$ will be used to represent the summation over all variables except $\mathbf{X}_S$, when $\mathbf{X}_S = \mathbf{x}_S$, obtaining:

$$p(\mathbf{x}_S) = \sum_{\sim\{\mathbf{x}_S\}} p(\mathbf{x}') = \sum_{\mathbf{x}' | \mathbf{x}'_S = \mathbf{x}_S} p(\mathbf{x}') \tag{9}$$

Notice the use of this notation in equation (10) which shows two different ways of calculating the marginal probabilities of $bd(\mathbf{X}_A)$.

$$\sum_{\sim cl(\mathbf{x}_A)} \sum_{\sim\{\mathbf{x} \setminus \mathbf{x}_A\}} p(\mathbf{x}') = \sum_{\sim bd(\mathbf{x}_A)} p(\mathbf{x}') \tag{10}$$

Equation (10), together with implications (11) and (12) below, are used in our proofs.

$$\mathbf{X}_R \supseteq X_i \subset \mathbf{X}_A \Rightarrow \mathbf{X}_R \subseteq cl(\mathbf{X}_A) \tag{11}$$

$$\mathbf{X}_R \not\supseteq X_i \subset \mathbf{X}_A \Rightarrow \mathbf{X}_R \subseteq \mathbf{X} \setminus \mathbf{X}_A \tag{12}$$

Implication (11) derives from the fact that in the clique-based decomposition any region is a clique. Therefore, any set of variables in the same region as $\mathbf{X}_A$ belongs to its closure. On the other hand, implication (12) represents the fact that if none of the variables that are in a region $\mathbf{X}_R$ belongs to $\mathbf{X}_A$, then the whole region $\mathbf{X}_R$ belongs to $\mathbf{X} \setminus \mathbf{X}_A$.

3.2 Markov properties

The Markov properties of a probability distribution $p(\mathbf{x})$ given its independence graph G are:

(i) Pairwise Markov property: for all non-adjacent vertices X_i and X_j in G,

$$p(x_i, x_j \mid \mathbf{x} \setminus (x_i, x_j)) = p(x_i \mid \mathbf{x} \setminus (x_i, x_j))p(x_j \mid \mathbf{x} \setminus (x_i, x_j))$$

(ii) Local Markov property: for every vertex X_i in G,

$$p(x_i, \mathbf{x} \setminus cl(x_i) \mid bd(x_i)) = p(x_i|bd(x_i))p(\mathbf{x} \setminus cl(x_i)|bd(x_i))$$

(iii) Global Markov property: for all disjoint subsets $\mathbf{X}_A$, $\mathbf{X}_B$, and $\mathbf{X}_C$, whenever $\mathbf{X}_B$ and $\mathbf{X}_C$ are separated by $\mathbf{X}_A$ in the graph, in the sense that all paths from $\mathbf{X}_B$ to $\mathbf{X}_C$ go through $\mathbf{X}_A$, then:

$$p(\mathbf{x}_B, \mathbf{x}_C \mid \mathbf{x}_A) = p(\mathbf{x}_B \mid \mathbf{x}_A)p(\mathbf{x}_C \mid \mathbf{x}_A)$$

We begin the presentation of results by proving an important theorem that will be used in the demonstration of the Markov properties satisfied by the Kikuchi approximation.

Theorem 1 *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, and a set of variables $\mathbf{X}_A$,*

$$k(\mathbf{x}_A|\mathbf{x} \setminus \mathbf{x}_A) = k(\mathbf{x}_A|bd(\mathbf{x}_A))$$

Proof:

$$\begin{aligned}
&k(\mathbf{x}_A \mid \mathbf{x} \setminus \mathbf{x}_A) \\
&= \frac{k(\mathbf{x})}{k(\mathbf{x} \setminus \mathbf{x}_A)} \\
&= \frac{K(\mathbf{x}, A)\bar{K}(\mathbf{x}, A)}{\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A)} \\
&= \frac{K(\mathbf{x}, A)\bar{K}(\mathbf{x}, A)}{\bar{K}(\mathbf{x}, A)\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} K(\mathbf{x}', A)} \\
&= \frac{K(\mathbf{x}, A)}{\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} K(\mathbf{x}', A)} \frac{\sum_{\sim\{cl(\mathbf{x}_A)\}} \bar{K}(\mathbf{x}', A)}{\sum_{\sim\{cl(\mathbf{x}_A)\}} \bar{K}(\mathbf{x}', A)} \\
&= \frac{\sum_{\sim\{cl(\mathbf{x}_A)\}} K(\mathbf{x}, A)\bar{K}(\mathbf{x}', A)}{\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} \sum_{\sim\{cl(\mathbf{x}_A)\}} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A)} \\
&= \frac{k(cl(\mathbf{x}_A))}{k(bd(\mathbf{x}_A))} \\
&= k(\mathbf{x}_A \mid bd(\mathbf{x}_A))
\end{aligned}$$

□

Theorem 2 (Local Markov property) *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, and a variable X_i,*

$$k(x_i, \mathbf{x} \setminus cl(x_i) \mid bd(x_i)) = k(x_i|bd(x_i))k(\mathbf{x} \setminus cl(x_i)|bd(x_i))$$

Proof:

We start from a particular case of theorem 1 when $A = i$.

$$\begin{aligned}
& k(x_i|\mathbf{x} \setminus x_i) = k(x_i|bd(x_i)) \\
& \Rightarrow \frac{k(\mathbf{x})}{k(\mathbf{x} \setminus x_i)} = \frac{k(cl(x_i))}{k(bd(x_i))} \\
& \Rightarrow \frac{k(\mathbf{x})}{k(bd(x_i))} = \frac{k(cl(x_i))}{k(bd(x_i))} \frac{k(\mathbf{x} \setminus x_i)}{k(bd(x_i))} \\
& \Rightarrow k(x_i, \mathbf{x} \setminus cl(x_i) \mid bd(x_i)) = k(x_i|bd(x_i))k(\mathbf{x} \setminus cl(x_i)|bd(x_i))
\end{aligned}$$

□

Theorem 3 (Conditional independence between disconnected sets) *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, two sets of vertices $\mathbf{X}_A$ and $\mathbf{X}_B$, such that there is not any edge between a vertex in $\mathbf{X}_A$ and a vertex in $\mathbf{X}_B$ then:*

$$k(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B)) = k(\mathbf{x}_A \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))k(\mathbf{x}_B \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))$$

Proof:

First, we propose a suitable factorization of the Kikuchi approximation in cliques that contain vertices in $\mathbf{X}_A$, $\mathbf{X}_B$, and $\mathbf{X} \setminus (\mathbf{X}_A, \mathbf{X}_B)$. As there is not any clique that contains one vertex from $\mathbf{X}_A$ and another one from $\mathbf{X}_B$, these three sets determine a non-overlapping partition of all the cliques and a factorization of $k(\mathbf{x})$. Using the notation introduced in (8), we obtain:

$$\begin{aligned}
k(\mathbf{x}) &= \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_A \neq \emptyset}} p(\mathbf{x}_R)^{c_R} \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap \mathbf{x}_B \neq \emptyset}} p(\mathbf{x}_R)^{c_R} \prod_{\substack{R \in \mathcal{R} \\ \mathbf{x}_R \cap (\mathbf{x}_A \cup \mathbf{x}_B) = \emptyset}} p(\mathbf{x}_R)^{c_R} \\
&= K(\mathbf{x}, A)K(\mathbf{x}, B)\bar{K}(\mathbf{x}, (A, B)) \qquad (13)
\end{aligned}$$

On the other hand, applying the definition given in Section 3.1 for a conditional function of the Kikuchi approximation, we have:

$$
\begin{aligned}
&k(\mathbf{x}_A \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))k(\mathbf{x}_B \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B)) \\
&= \frac{k(\mathbf{x} \setminus \mathbf{x}_B)k(\mathbf{x} \setminus \mathbf{x}_A)}{k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))} \\
&= \frac{\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_B\}} K(\mathbf{x}, A)K(\mathbf{x}, B)\bar{K}(\mathbf{x}, (A, B))}{\sum_{\sim\{\mathbf{x}\setminus(\mathbf{x}_A,\mathbf{x}_B)\}} K(\mathbf{x}, A)K(\mathbf{x}, B)\bar{K}(\mathbf{x}, (A, B))} \frac{\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} K(\mathbf{x}, A)K(\mathbf{x}, B)\bar{K}(\mathbf{x}, (A, B))}{k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))} \\
&= \frac{(\bar{K}(\mathbf{x}, (A, B)))^2(K(\mathbf{x}, A)K(\mathbf{x}, B))(\sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_B\}} K(\mathbf{x}, B) \sum_{\sim\{\mathbf{x}\setminus\mathbf{x}_A\}} K(\mathbf{x}, A))}{\bar{K}(\mathbf{x}, (A, B))(\sum_{\sim\{\mathbf{x}\setminus(\mathbf{x}_A,\mathbf{x}_B)\}} K(\mathbf{x}, A)K(\mathbf{x}, B))k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))} \\
&= \frac{\bar{K}(\mathbf{x}, (A, B))K(\mathbf{x}, A)K(\mathbf{x}, B))}{k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))} \\
&= \frac{k(\mathbf{x})}{k(\mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))} \\
&= k(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x} \setminus (\mathbf{x}_A, \mathbf{x}_B))
\end{aligned}
$$

□

Theorem 4 (Pairwise Markov property) *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, and two variables X_i and X_j, if the corresponding vertices are not joined in G:*

$$k(x_i, x_j \mid \mathbf{x} \setminus (x_i, x_j)) = k(x_i \mid \mathbf{x} \setminus (x_i, x_j))k(x_j \mid \mathbf{x} \setminus (x_i, x_j))$$

Proof:

This property is a particular case of theorem (3) when $\mathbf{X}_A = X_i$ and $\mathbf{X}_B = X_j$. □

We show that, in general, the Kikuchi approximation does not fulfill the global Markov property.

Conjecture 1 (Global Markov property) *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, for all disjoint subsets $\mathbf{X}_A$, $\mathbf{X}_B$, and $\mathbf{X}_C$, whenever $\mathbf{X}_B$ and $\mathbf{X}_C$ are separated by $\mathbf{X}_A$ in the graph, then:*

$$k(\mathbf{x}_B, \mathbf{x}_C \mid \mathbf{x}_A) = k(\mathbf{x}_B \mid \mathbf{x}_A)k(\mathbf{x}_C \mid \mathbf{x}_A)$$

We present a counterexample of conjecture 1. It corresponds to the simplest case, when $\mathbf{X}_B = X_i$, $\mathbf{X}_C = X_j$, X_i and X_j are not connected in the graph, and $\mathbf{X}_A = \emptyset$. For this case, the global Markov property states that:

$$k(x_i, x_j) = k(x_i)k(x_j) \tag{14}$$

We use the symbol $\leftrightarrow$ to represent that two variables are connected by at least one path in the graph. Let $\mathbf{X}_A = \{X_k \mid X_k \leftrightarrow X_i\}$. $\mathbf{X}_A, \mathbf{X}_{N\setminus A}$ forms a partition of the graph such that there is not any edge that joins variables from both components. Although

$\mathbf{X}_A$ depends on X_i, for a more compact notation we do not represent this dependence in the notation. Using the factorization of the Kikuchi approximation (8), we obtain an expression for the bivariate marginal of the Kikuchi approximation for (X_i, X_j).

$$
\begin{aligned}
& k(x_i, x_j) \\
&= \sum_{\mathbf{x}' | \mathbf{x}'_{i,j} = \mathbf{x}_{i,j}} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A) \\
&= \sum_{\mathbf{x}'_A | x'_i = x_i} \sum_{\mathbf{x}'_{N\setminus A} | x'_j = x_j} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A) \\
&= \sum_{\mathbf{x}'_A | x'_i = x_i} K(\mathbf{x}', A) \left(\sum_{\mathbf{x}'_{N\setminus A} | x'_j = x_j} \bar{K}(\mathbf{x}', A) \right) \\
&= \sum_{\mathbf{x}'_A | x'_i = x_i} K(\mathbf{x}', A) \sum_{\mathbf{x}'_{N\setminus A} | x'_j = x_j} \bar{K}(\mathbf{x}', A) \\
&= k_A(x_i) k_{N\setminus A}(x_j)
\end{aligned}
\tag{15}
$$

In the factorization, k_A refers to the Kikuchi factorization constructed from the subgraph that includes vertices and edges that contain variables in $\mathbf{X}_A$. The factorization is possible since the partition of the graph in two disconnected components similarly enables the cliques of the graph to be partitioned in only two sets. Likewise, we obtain an expression for the product of the univariate marginals of the Kikuchi approximation.

For the proof we use the following two equalities:

$$
\sum_{\mathbf{x}'_{N\setminus A} | x'_i = x_i} \bar{K}(\mathbf{x}', A) = \sum_{\mathbf{x}'_{N\setminus A}} \bar{K}(\mathbf{x}', A) = k_{N\setminus A}(\mathbf{x}_{N\setminus A})
$$

$$
\sum_{\mathbf{x}'_A | x'_j = x_j} K(\mathbf{x}', A) = \sum_{\mathbf{x}'_A} K(\mathbf{x}', A) = k_A(\mathbf{x}_A)
$$

The product of the univariate Kikuchi approximations for variables X_i and X_j results in:

$$
\begin{aligned}
& k(x_i) k(x_j) \\
&= \sum_{\mathbf{x}' | x'_i = x_i} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A) \sum_{\mathbf{x}' | x'_j = x_j} K(\mathbf{x}', A)\bar{K}(\mathbf{x}', A) \\
&= \sum_{\mathbf{x}'_A | x'_i = x_i} K(\mathbf{x}', A) \sum_{\mathbf{x}'_{N\setminus A} | x'_i = x_i} \bar{K}(\mathbf{x}', A) \sum_{\mathbf{x}'_A | x'_j = x_j} K(\mathbf{x}', A) \sum_{\mathbf{x}'_{N\setminus A} | x'_j = x_j} \bar{K}(\mathbf{x}', A) \\
&= k_A(x_i) k_{N\setminus A}(x_j) \sum_{\mathbf{x}'_A} k_A(\mathbf{x}'_A) \sum_{\mathbf{x}'_{N\setminus A}} k_{N\setminus A}(\mathbf{x}'_{N\setminus A})
\end{aligned}
\tag{16}
$$

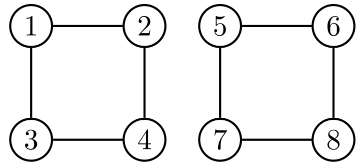

Figure 2: Independence graph with two disconnected components

The left (equation (15)) and right (equation (16)) parts of equation (14) are only equal, when

$$\sum_{\mathbf{x}'_A} k_A(\mathbf{x}'_A) \sum_{\mathbf{x}'_{N\setminus A}} k_{N\setminus A}(\mathbf{x}'_{N\setminus A}) = 1 \tag{17}$$

A sufficient condition for equation (17) to be fulfilled is for $k_A(\mathbf{x}_A)$ and $k_{N\setminus A}(\mathbf{x}_{N\setminus A})$ to be probability distributions. However, this is not a necessary condition.

Example 2 describes a situation where the Kikuchi approximations of $k_A(\mathbf{x}_A)$ and $k_{N\setminus A}(\mathbf{x}_{N\setminus A})$ are not probability distributions in general, and the global Markov property is clearly not satisfied.

Example 2 *The graph shown in Figure 2 comprises two disconnected components. For an arbitrarily chosen probability $p(\mathbf{x})$ that satisfies the conditional independence properties described by the graph it holds $p(x_1, x_5) = p(x_1)p(x_5)$. We can construct the Kikuchi approximation according to the graph. Let $k(\mathbf{x})$ be such an approximation; then, it may occur that $k(x_1, x_5) \neq k(x_1)k(x_5)$. The reason is that, as $k(\mathbf{x})$ is not a probability distribution, the expressions $\sum_{x'_1,x'_2,x'_3,x'_4} k_{(1,2,3,4)}(x'_1, x'_2, x'_3, x'_4)$ and $\sum_{x'_5,x'_6,x'_7,x'_8} k_{(5,6,7,8)}(x'_5, x'_6, x'_7, x'_8)$ can both be different to one. Thus, equation (17) might not be satisfied.*

Although the global Markov property is not fulfilled, equation (15) points to the fact that the Kikuchi approximation can be factorized in the product of Kikuchi approximations calculated from the components of an independence graph partition. The following theorem formalizes this observation.

Theorem 5 (Kikuchi decomposition property) *Given a Kikuchi approximation $k(\mathbf{x})$ defined on a graph G, such that $\mathbf{X} = \mathbf{X}_A \cup \mathbf{X}_B \cup \mathbf{X}_C$, and $\mathbf{X}_A$ is a separator of $\mathbf{X}_B$ and $\mathbf{X}_C$, then:*

$$k(\mathbf{x}) = \frac{k_{AB}(\mathbf{x}_A, \mathbf{x}_B) k_{AC}(\mathbf{x}_A, \mathbf{x}_C)}{k_A(\mathbf{x}_A)} \tag{18}$$

Proof:

First we propose a suitable factorization of the Kikuchi approximation in three sets of cliques:

$$k(\mathbf{x}) = \frac{\prod_{R\in\mathcal{R}\mathbf{X}_R\subseteq\mathbf{X}_{A\cup B}} p(\mathbf{x}_R)^{c_R} \prod_{R\in\mathcal{R}\mathbf{X}_R\subseteq\mathbf{X}_{A\cup C}} p(\mathbf{x}_R)^{c_R}}{\prod_{R\in\mathcal{R}\mathbf{X}_R|\mathbf{X}_R\subseteq\mathbf{X}_A} p(\mathbf{x}_R)^{c_R}}$$

$$= \frac{K(\mathbf{x},(A,B))K(\mathbf{x},(A,C))}{K(\mathbf{x},A)} \tag{19}$$

It is clear that every region belongs to $\mathbf{X}_{A\cup B}$, to $\mathbf{X}_{A\cup C}$, or to both regions. In the latter case, the region belongs to $\mathbf{X}_A$. As cliques in $\mathbf{X}_A$ are counted twice, the factorization includes the division by $K(\mathbf{x}, A)$. To prove (18) we only need to show that $k_{AB}(\mathbf{x}_A, \mathbf{x}_B) = K(\mathbf{x},(A,B))$, $k_{AC}(\mathbf{x}_A, \mathbf{x}_C) = K(\mathbf{x},(A,C))$ and $k_A(\mathbf{x}_A) = K(\mathbf{x}, A)$.

Let us take the case of $k_{AB}(\mathbf{x}_A, \mathbf{x}_B)$. The regions in this Kikuchi approximation are precisely those in $K(\mathbf{x},(A,B))$. To see that the overcounting values coincide in every equation we take a region R, and decompose the expression of the overcounting value in the following way:

$$c_R = 1 - \sum_{\substack{S\in\mathcal{R}\\ S\supset R}} c_S$$

$$= 1 - \left(\sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup B\cup C), S\supset R}} c_S \right)$$

$$= 1 - \left(\sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup B), S\supset R}} c_S + \sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup C), S\supset R}} c_S - \sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq A, S\supset R}} c_S \right) \tag{20}$$

The last term in equation (20) has been substracted because regions contained in $\mathbf{X}_A$ have been counted twice, in $\mathbf{X}_{A\cup B}$ and $\mathbf{X}_{A\cup C}$.

If $R \cap B \neq \emptyset$, $c_R = 1 - (\sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup B), S\supset R}} c_S)$, is the overcounting value that can be calculated from cliques in $\mathbf{X}_{A\cup B}$. A similar case takes places if $R \cap C \neq \emptyset$, $c_R = 1 - (\sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup C), S\supset R}} c_S)$. In the particular case in which $R \cap A \neq \emptyset$, two of the factors simplify and $c_R = 1 - (\sum_{\substack{S\in\mathcal{R}\\ S|S\subseteq(A\cup B), S\supset R}} c_S)$. The same analysis is valid for $k_{AC}(\mathbf{x}_A, \mathbf{x}_C)$ and $k_A(\mathbf{x}_A)$. □

To summarize the results proved in this section, we have shown that some of the properties of the independence graph are translated into the Kikuchi approximation. Pairwise and local Markov properties are fulfilled. However, the global Markov property is not satisfied in the general case. Instead, we have proved that it is possible to decompose the Kikuchi approximation in the product of local Kikuchi approximations defined from a decomposition of the graph.

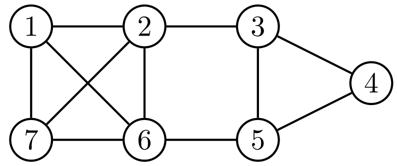

Figure 3: Independence graph with one incomplete maximal irreducible component and two complete ones

4 Decomposability of the Kikuchi approximation

Decomposability is essential to handle feasible approximations of a probability. In this section, we show how the Kikuchi decomposition property will permit the definition of Kikuchi approximations in which each factor itself is a Kikuchi approximation corresponding to a subgraph of the original independence graph. We go one step further and define the class of partial Kikuchi approximations, in which some of the Kikuchi approximation components correspond to exact marginal probability distributions. First, we introduce a number of definitions, propositions and theorems, taken from Whittaker (1991), that lead to a factorization of the Kikuchi approximation based on the irreducible components of the independence graph.

4.1 Definitions

Definition 3 *There exists a decomposition of the random vector $\mathbf{X}$ with respect to a probability distribution $p(\mathbf{x})$ or equivalently, $\mathbf{X}$ is reducible, if and only if there exists a partition $\mathbf{X}$ into $(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C)$ such that:*

(i) $p(\mathbf{x}_B, \mathbf{x}_C|\mathbf{x}_A) = p(\mathbf{x}_B|\mathbf{x}_A)p(\mathbf{x}_C|\mathbf{x}_A)$ *and neither B nor C are empty; and*

(ii) the subgraph on A, in the independence graph of $\mathbf{X}$ is complete.

If so, the components of $\mathbf{X}$ are $\mathbf{X}_{AB} = (\mathbf{X}_A, \mathbf{X}_B)$ and $\mathbf{X}_{AC} = (\mathbf{X}_A, \mathbf{X}_C)$. If such a decomposition does not exist $\mathbf{X}$ is said to be irreducible.

Example 3 *Consider the graph in Figure 3 and the partition defined by sets $A = \{2, 6\}$, $B = \{1, 7\}$ and $C = \{3, 4, 5\}$. Then, $\mathbf{X}_{AB}$ and $\mathbf{X}_{AC}$ are the components that form a decomposition of the graph. $\mathbf{X}_{AB}$ is an irreducible component, and $\mathbf{X}_{AC}$ is a reducible one because it can be decomposed according to definition 3.*

Definition 4 *The random vectors $\mathbf{X}_{D_1}, \mathbf{X}_{D_2}, \ldots, \mathbf{X}_{D_m}$ are the maximal irreducible components of $\mathbf{X}$ if and only if:*

(i) each vector $\mathbf{X}_{D_i}$ is an irreducible component of $\mathbf{X}$;

(ii) no subset D_i, is a proper subset of any other, D_j; and

(iii) $\mathbf{X} = \mathbf{X}_{D_1} \cup \mathbf{X}_{D_2} \cup \cdots \cup \mathbf{X}_{D_m}$

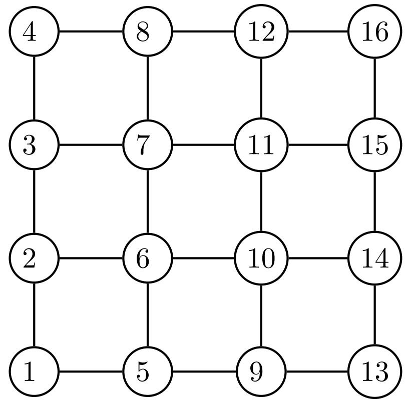

Figure 4: The grid lattice for the Ising and sping glass models is an irreducible incomplete component

An irreducible component is said to be complete if it is a clique. Otherwise it is called an incomplete irreducible component.

Example 4 $\mathbf{X}_{1,2,6,7}$*,* $\mathbf{X}_{2,3,5,6}$*, and* $\mathbf{X}_{3,4,5}$ *are the maximal irreducible components of the graph shown in Figure 3.* $\mathbf{X}_{1,2,6,7}$ *and* $\mathbf{X}_{3,4,5}$ *are complete irreducible components.* $\mathbf{X}_{2,3,5,6}$ *is an irreducible component that is not complete.*

Proposition 1 ((Whittaker [33], pp. 385) Irreducible component factorization) *The maximal irreducible components of* $\mathbf{X}$ *corresponding to the subsets* $\{\mathbf{X}_{D_1}, \mathbf{X}_{D_2}, \ldots, \mathbf{X}_{D_m}\}$ *are unique and the density function of* $\mathbf{X}$ *factorizes uniquely into* $f(\mathbf{x}) = \frac{f_{\mathbf{x}_{D_1}} f_{\mathbf{x}_{D_2}} \ldots f_{\mathbf{x}_{D_m}}}{g}$*, where function* g *is a product of marginal density functions,* $g = \prod f_{\mathbf{x}_A}$*, in which each subset* $\mathbf{X}_A$ *is an intersection of irreducible components, and it is complete.*

Definition 5 (Whittaker [33], pp. 389) *An* n*-dimensional random vector* $\mathbf{X}$*, or its density function, is decomposable if and only if there exists a sequence of decompositions to complete irreducible components.*

Independence graphs can be decomposed into irreducible components. The problem is known as decomposition by clique separators [27] or maximal prime subgraph decomposition [23] and it may solved by certain algorithms [27]. Maximal prime subgraph decompositions have been proposed in Bayesian networks as a computational structure for lazy propagation [23]. However, in the case of Bayesian networks, this type of decompositions has also been criticized as a very limited representation of the independence relationships of this class of models [7].

One important remark is that irreducible incomplete components can be very large and, in fact, a graph can be formed by a unique irreducible component (e.g. the graph shown in Figure 4).

4.2 Decomposability of the Kikuchi approximation

We study how to extend the results achieved for the factorization of distributions to the Kikuchi approximation.

Theorem 6 *Given the independence graph G of $\mathbf{X}$, and the Kikuchi approximation $k(\mathbf{x})$ defined on G, if there exists a partition $\mathbf{X}$ into $(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C)$ such that the components of $\mathbf{X}$ are $\mathbf{X}_{AB} = (\mathbf{X}_A, \mathbf{X}_B)$ and $\mathbf{X}_{AC} = (\mathbf{X}_A, \mathbf{X}_C)$, then:*

$$k(\mathbf{x}) = \frac{k_{AB}(\mathbf{x}_{AB})k_{AC}(\mathbf{x}_{AC})}{k_A(\mathbf{x}_A)} \tag{21}$$

Proof:

This is a particular case of theorem 5 when the separator $\mathbf{X}_A$ is a clique and, therefore, $k_A(\mathbf{x}_A) = p(\mathbf{x}_A)$. □

Theorem 7 *Given the independence graph G, a factorization of $\mathbf{X}$ in irreducible components $p(\mathbf{x}) = \frac{p_{\mathbf{x}_{D_1}} p_{\mathbf{x}_{D_2}} \cdots p_{\mathbf{x}_{D_m}}}{\prod_{i=1}^{m-1} p_{\mathbf{x}_{d_i}}}$, in which $\mathbf{x}_{d_1}, \mathbf{x}_{d_2}, \ldots, \mathbf{x}_{d_{m-1}}$ is the (possibly empty) set of complete irreducible components, intersection of the irreducible components, then the Kikuchi approximation $k(\mathbf{x})$ defined on G can be decomposed as:*

$$k(\mathbf{x}) = \frac{k_{D_1}(\mathbf{x}_{D_1})k_{D_2}(\mathbf{x}_{D_2}) \cdots k_{D_m}(\mathbf{x}_{D_m})}{\prod_{i=1}^{m-1} p_{\mathbf{x}_{d_i}}}$$

Proof:

The proof will be done by induction on the number of components, and using theorem 6.

If $m = 1$, there is only one component $\mathbf{X}_{D_1}$ and no separators. In this case, $k(\mathbf{x}) = k(\mathbf{x}_{D_1})$.

Let us suppose that, for the component $\mathbf{X}_{(D_1, D_2, \ldots, D_{i-1})}$, the theorem holds, i.e.

$$k(\mathbf{x}_{(D_1, D_2, \ldots, D_{i-1})}) = \frac{k_{D_1}(\mathbf{x}_{D_1})k_{D_2}(\mathbf{x}_{D_2}) \cdots k_{D_{i-1}}(\mathbf{x}_{D_{i-1}})}{\prod_{j=1}^{i-2} p_{\mathbf{x}_{d_j}}}$$

Now we prove that, for $\mathbf{X}_{(D_1, D_2, \ldots, D_i)}$, the theorem is satisfied.

Let $A = d_{i-1}$, $B = \{D_1, D_2, \ldots, D_{i-1}\}$, and $C = D_i$. $(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C)$ is a partition of $\mathbf{X}_{(D_1, D_2, \ldots, D_i)}$ that satisfies the conditions of theorem 6. Therefore,

$$\begin{aligned}
&k(\mathbf{x}_{(D_1, D_2, \ldots, D_i)}) \\
&= \frac{k(\mathbf{x}_{(D_1, D_2, \ldots, D_{i-1})})k(\mathbf{x}_{D_i})}{k(\mathbf{x}_{d_i})} \\
&= \frac{k_{D_1}(\mathbf{x}_{D_1})k_{D_2}(\mathbf{x}_{D_2}) \cdots k_{D_{i-1}}(\mathbf{x}_{D_{i-1}})}{\prod_{j=1}^{i-2} p_{\mathbf{x}_{d_j}}} \frac{k(\mathbf{x}_{D_i})}{k(\mathbf{x}_{d_i})} \\
&= \frac{k_{D_1}(\mathbf{x}_{D_1})k_{D_2}(\mathbf{x}_{D_2}) \cdots k_{D_i}(\mathbf{x}_{D_i})}{\prod_{j=1}^{i-1} p_{\mathbf{x}_{d_j}}}
\end{aligned}$$

For all i, $\mathbf{X}_{d_i}$ is a complete irreducible component (i.e. clique). Thus, $k_{d_i}(\mathbf{x}_{d_i}) = p(\mathbf{x}_{d_i})$.

We have reached the decomposition of the Kikuchi approximation of $\mathbf{X}_{(D_1,D_2,\ldots,D_i)}$. To complete the proof, notice that $\mathbf{X} = \mathbf{X}_{(D_1,D_2,\ldots,D_m)}$. □

Example 5 *We analyze the independence graph shown in Figure 3. The Kikuchi approximation of this graph is:*

$$k(\mathbf{x}) = \frac{p(x_1,x_2,x_6,x_7)p(x_2,x_3)p(x_5,x_6)p(x_3,x_4,x_5)}{p(x_2)p(x_3)p(x_5)p(x_6)} \tag{22}$$

The factorization of $k(\mathbf{x})$ based on the irreducible components of the graph is:

$$k(\mathbf{x}) = \frac{k(x_1,x_2,x_7,x_6)k(x_2,x_3,x_5,x_6)k(x_3,x_4,x_5)}{k(x_2,x_6)k(x_3,x_5)} \tag{23}$$

The Kikuchi approximation of the incomplete irreducible component (X_2,X_3,X_5,X_6) is:

$$k_{(2,3,5,6)}(x_2,x_3,x_5,x_6) = \frac{p(x_2,x_3)p(x_3,x_5)p(x_5,x_6)p(x_2,x_6)}{p(x_2)p(x_3)p(x_5)p(x_6)} \tag{24}$$

Substituting equation (24) *in* (23)*, and considering that the other factors are calculated from complete irreducible components, and that therefore, the Kikuchi marginals coincide with the probability marginals, we obtain the original Kikuchi approximation shown in equation* (22)*.*

Now we highlight another aspect related to the Kikuchi approximation of an independence graph: by identifying the regions of the independence graph where the Kikuchi approximation is localized (the incomplete components), we can estimate to what extent the Kikuchi approximation is used for the approximation of the distribution associated with a given graph, and therefore we can obtain a measure of the approximation accuracy. Many irreducible components will imply more components approximated with the Kikuchi approximation.

Furthermore, we can constrain the use of the Kikuchi approximation to certain areas of the graph.

Given the independence graph G of $\mathbf{X}$, a partial Kikuchi approximation of the probability is that where only a subset of all the irreducible incomplete components are approximated by the corresponding Kikuchi approximation of the components. The rest of components are calculated exactly.

The partial Kikuchi approximation admits the existence of components that are calculated exactly, i.e. they can be triangulated like methods for doing inference in graphical models usually do.

Let us suppose that the number of incomplete irreducible components in G is t. The number of partial Kikuchi approximation is $2^t - 2$, including the complete Kikuchi approximation. Hence, the total number of partial Kikuchi approximations is $2^t - 1$. Rationales for selecting one partial Kikuchi approximation rather than others might be related to the size of the induced triangulated clique (e.g. 4-sized cliques could be assumed to be triangulated). Another criterion can be the cardinality of the variables that are in the incomplete irreducible component (e.g. when the cardinality of the variables involved is high, the Kikuchi approximation is recomendable to diminish the number of parameters needed to approximate the model).

5 Related work on Kikuchi approximations

We have presented a number of properties fulfilled by the Kikuchi approximation constructed from clique-based decompositions. Now we relate these results to current research on similar topics.

5.1 Kikuchi approximations

Kikuchi approximations are an example of a panoply of methods that approach the problem of approximating a measure (i.e. entropy, energy, probability) using graph-based decompositions. Initial applications of Bethe and Kikuchi approximations were constrained to the field of statistical physics [14, 20, 21]. The purpose of finding a way to decompose the otherwise difficult to handle free energy of a system led to the use of these approximations in physics. The idea was developed later by Yedidia et al. [38] in the context of generalized belief propagation. This contribution expanded the scope of application of belief propagation, which has been traditionally used in tasks such as, obtaining *a posteriori* marginal probabilities in graphical models [25], computing the most probable global states or system configurations [22], and improving the efficiency of iterative proportional fitting (IPF) [5, 29].

5.2 Generalized Kirkwood superposition

There is another path that leads to the notion of Kikuchi approximation treated in this paper. The Kirkwood superposition [15] is an approximation for the three-body distribution of liquids introduced in liquid-state statistical mechanics. Since this approximation has been recently applied in the machine learning community [12], we elaborate on its relationship with the clique-based decomposition and the Kikuchi approximation.

The essence of the original Kirkwood superposition approximation is that all possible correlations in a system are expressed by binary correlations. Although the approach has been widely applied in the theory of liquids, its implications and range of applications are controversial (see Grouba et al. [9] for an extensive review on the subject). More relevant to our research is the derivation of the generalized Kirkwood superposition for the expansion of the information entropy in terms of correlation functions that have been proposed in Attard et al. [4].

In simple terms, this derivation proposes an approximation of the information entropy that includes successively higher-order correlations in a systematic fashion. The approach is extended by Matsuda [18] to calculate the higher order mutual information $In(\mathbf{X})$:

$$In(X_1,\ldots,X_n) = (-1)^n \sum_{x_1,\ldots,x_n} p(x_1,\ldots,x_n) ln \frac{p(x_1,\ldots,x_n)}{\bar{p}(x_1,\ldots,x_n)} \tag{25}$$

where $\bar{p}(\mathbf{x})$ is the generalized Kirkwood superposition defined as:

$$\bar{p}(x_1,\ldots,x_n) = \prod_{k=1}^{n-1} (-1)^{k+1} \prod_{\{i_1<\ldots<i_{n-k}\}} p(x_1,\ldots,x_{n-k}) \tag{26}$$

where $\prod_{\{i_1<\ldots<i_{n-k}\}}$ runs over all possible combinations $\{i_1,\ldots,i_{n-k}\} \subseteq \{1,\ldots,n\}$.

There is a clear relationship between the generalized Kirkwood superposition and the Kikuchi approximation constructed from a clique-based decomposition. The Kirkwood superposition corresponds to a situation in which a complete graph is considered and, instead of choosing the single maximal clique of size n, all cliques of size $n-1$ are chosen as the initial regions. Nevertheless, the higher-order contributions can be neglected from (26), obtaining approximations that may yield good results for weakly correlated examples [18].

5.3 Research trends in the application of region-based decompositions

Concerning the field of machine learning, we identify two main current research trends in the application of region-based decompositions.

(i) The use of region-based decompositions to design and improve inference methods, particularly, generalized belief propagation algorithms.

(ii) The use of region-based decompositions to find approximate factorizations of probability distributions based on marginal probability distributions.

The first research trend [3, 8, 10, 19, 24, 28, 30, 31, 38, 39] includes work on the determination of efficient message passing schemes in belief propagation, bounds on the accuracy of the inferred marginals, and conditions of convergence for the propagation algorithms. The second one [11, 12, 13, 26] focuses on the conception of measures to evaluate the accuracy of the learned approximations, algorithms to learn and sample these approximations from data, and the identification of significant interactions in data.

One common problem of both lines of research is the selection of the initial regions upon which the approximations are based. Some recent work on belief propagation algorithm addresses this problem using graph partition strategies [34, 35], sequential methods [32], and other approaches [24, 38].

Jakulin et al. [13] have proposed the used of Kikuchi approximations for supervised classification. A region-based decomposition learning algorithm is introduced with this objective. Region-based decompositions based on 2-way and 3-way interactions are tested. Higher order interactions are not considered.

Clique-based decompositions are a way to automatically determine the initial regions of the graph that can be used to construct region-based decompositions. The local

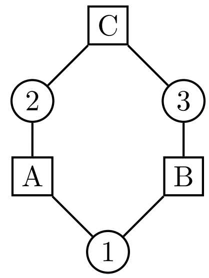

Figure 5: Factor graph

Markov property of the Kikuchi approximation was useful to define algorithms to learn and sample the approximations from data [26]. However, results presented in [26] did not provide any measure to evaluate the accuracy of the approximation. Furthermore, it was not clear whether Kikuchi approximations could be applied locally in the probability distribution approximations. The properties presented in this paper can be used to define decomposable accuracy measures that can help to create more sophisticated methods for learning the approximations.

Another difference between the belief propagation approach and the learning approach is the type of graph in which the construction of the Kikuchi approximation is based, and its interpretation. In the original work of Yedidia et al. [36] the Kikuchi approximation was calculated using pairwise or higher-order Markov random fields (MRFs). Regions in the graph comprised the variables and the sets in which the potential functions were defined. Recent work [37] focuses on models defined on factor graphs [16]. Factor graphs have variable and factor nodes. There is a variable node for each variable of the problem and a factor node for each node, with an edge connecting variable node i to factor node a if X_i is an argument of f_a.

In factor graphs, functions can represent some sort of interaction among their argument variables, but there is not any requirement concerning the type and strength of these interactions. The validity condition of the region-based decomposition used by Yedidia et al. (2002) [37] establishes that every variable and factor node is counted once in the approximation (using the c_R values in the sum) but, apart from this requirement, the choice of the initial regions is arbitrary. There are other different ways of representing the region graph decompositions which include Hasse diagrams [24], region graphs [37], and hypergraphs [31].

The work presented in this paper can be extended to include different graph representations. The definition of the Kikuchi approximation given in Section 2 is based on an independence graph that is interpreted as a graphical model. However, the Kikuchi approximation constructed from a clique-based decomposition can be defined not only on undirected graphical models but also on a larger class of models equivalent to hierarchical loglinear models. These models can be graphically represented using factor graphs.

In the factor graph representation of a clique-based decomposition each maximal clique will have an associated factor. As in the case of hierarchical models, it is assumed that the existence of a clique means that all lower interactions are covered by the model. The factor graph coincides with the maximal representation of the hierarchical models.

Example 6 *Consider a probabilistic model with factors* $\mathbf{X}_{12}$*,* $\mathbf{X}_{23}$ *and* $\mathbf{X}_{13}$*. The Kikuchi approximation of this model would be*

$$k(\mathbf{x}) = \frac{p(x_1, x_2)p(x_2, x_3)p(x_1, x_3)}{p(x_1)p(x_2)p(x_3)}$$

but this approximation can not be recovered from any undirected graph because the original model is not graphical. Figure 5 shows the factor graph representation of this model.

6 Conclusions

In this paper, we have investigated a number of properties satisfied by the Kikuchi approximation constructed from the clique-based decomposition. We have shown that the Kikuchi approximation satisfies the local and pairwise Markov properties. We have proved that the global Markov property is not fulfilled. These results lay the foundations for the further development of algorithms that use Kikuchi approximations.

We have proposed a decomposition of the Kikuchi approximation according to the irreducible components of the graph. From this decomposition we have introduced the notion of partial Kikuchi approximations. A consequence of this result is that an initial measure of the complexity of the Kikuchi approximation can be given, based on the number of irreducible components and their complexity (number of nodes and of factors involved in the factorization). The results achieved indicate a way to investigate the accuracy of the Kikuchi and partial Kikuchi approximations.

This paper intends to go beyond the idea of employing heuristic approaches to construct region-based decompositions. Instead it proposes to construct and to use these decompositions taking into consideration the properties of the approximations that they determine.

Based on the results presented in this paper, it is possible to design algorithms that learn Kikuchi approximations from data by means of scoring-search techniques. These algorithms could be used for probability approximation, supervised classification, and optimization using EDAs.

The conception of more general region-based decompositions is a promising research area in probabilistic approximation. The work presented here is intended to be a modest step towards this goal.

7 Acknowledgments

This work was supported by the Saiotek S-PEO4UN25 and Etortek-Biolan projects from the Basque Government.

References

[1] S. M. Aji and R. J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, 2000.

[2] S. M. Aji and R. J. McEliece. The generalized distributive law and free energy minimization. In *Proceedings of the 39th Allerton Conference on Communication, Control, and Computing*, Allerton, Illinois, 2001.

[3] S. M. Aji and M. Yildirim. *Mathematical Systems Theory in Biology, Communications, Computation and Finance Series: The IMA Volumes in Mathematics and its Applications, Vol. 134*, chapter Belief propagation on partially ordered sets, pages 275–300. Springer Verlag, 2003.

[4] P. Attard, O. G. Jepps, and S. Marcelja. Information content of signals using correlation function expansions of the entropy. *Physical Review E*, 56:4052–4067, 1997.

[5] J. H. Badsberg and F. M. Malvestuto. An implementation of the iterative proportional fitting procedure by propagation trees. *Computational Statistics and Data Analysis*, 37(3):297–322, 2003.

[6] H. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London*, 150(871):552–575, 1935.

[7] C. Butz, Q. Hu, and X. Yang. Critical remarks on the maximal prime decomposition of Bayesian networks. In G. Wang, Q. Liu, Y. Yao, and A. Skowron, editors, *9th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC03)*, volume 2639 of *Lecture Notes in Computer Science*, pages 682–685. Springer Verlag, 2003.

[8] C. Crick and A. Pfeffer. Loopy belief propagation as a basis for communication in sensor networks. In *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2003)*, pages 159–166. Morgan Kaufmann Publishers, 2003.

[9] V. D. Grouba, A. V. Zorin, and I. A. Sevastianov. The superposition approximation: a critical review. *International Journal of Modern Physics B*, 18(1):1–44, 2004.

[10] T. Heskes. Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In *Advances in Neural Information Processing Systems, (NIPS-2002)*, volume 14, pages 343–350. MIT Press, 2003.

[11] R. Höns. *Estimation of Distribution Algorithms and Minimum Relative Entropy.* PhD thesis, University of Bonn, Bonn, Germany, 2005. Unpublished manuscript.

[12] A. Jakulin and I. Bratko. Testing the significance of attribute interactions. In *Proceedings of the 21th Conference on Machine Learning (ICML-2004)*, pages 409–416, Banff, Canada, 2004. ACM Press.

[13] A. Jakulin, I. Rish, and I. Bratko. Kikuchi-Bayes: Factorized models for approximate classification in closed form. Technical Report RC23314 (WO408-175), IBM, August 2004.

[14] R. Kikuchi. A theory of cooperative phenomena. *Physical Review*, 81(6):988–1003, 1951.

[15] J. G. Kirkwood and E. M. Boggs. The radial distribution function in liquids. *Journal of Chemical Physics*, 10:394–402, 1942.

[16] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[17] P. Larrañaga and J. A. Lozano, editors. *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation.* Kluwer Academic Publishers, Boston/Dordrecht/London, 2002.

[18] H. Matsuda. Physical nature of higher-order mutual information: intrinsic correlations and frustration. *Physical Review E*, 62(3):3096–3102, 2000.

[19] T. Minka. *A Family of Algorithms for Approximate Bayesian Inference.* PhD thesis, MIT Media Lab, Massachusetts Institute of Technology, Massachusetts, 2001.

[20] T. Morita. Cluster variation method for non-uniform Ising and Heisenberg models and spin-pair correlation function. *Progress of Theoretical Physics*, 85(2):243–255, 1991.

[21] T. Morita. Formal structure of the cluster variation method. *Progress of Theoretical Physics Supplements*, 115:27–39, 1994.

[22] D. Nilsson. An efficient algorithm for finding the M most probable configurations in probabilistic expert systems. *Statistics and Computing*, 2:159–173, 1998.

[23] K. G. Olesen and A. L. Madsen. Maximal prime subgraph decomposition of Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics: Part B*, 32(1):21–31, 2000.

[24] P. Pakzad and V. Anantharam. Estimation and marginalization using Kikuchi based methods. Technical report, Faculty of Electrical Engineering and Computer Science. University of California Berkeley, 2003.

[25] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[26] R. Santana. Estimation of distribution algorithms with Kikuchi approximations. *Evolutionary Computation*, 13(1):67–97, 2005.

[27] R. Tarjan. Decomposition by clique separators. *Discrete Mathematics*, 55:221–232, 1985.

[28] S. Tatikonda and M. I. Jordan. Loopy belief propagation and Gibbs measures. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, pages 493–500. Morgan Kaufmann Publishers, 2002.

[29] Y. Teh and M. Welling. On improving the efficiency of the iterative proportional fitting procedure. In *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, volume 9, 2003.

[30] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Tree-based reparameterization framework for analysis of belief propagation and related algorithms. Technical Report LIDS P-2510, Laboratory for Information and Decision Systems, MIT, 2001.

[31] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, Department of Statistics, University of California, Berkeley, September 2003.

[32] M. Welling. On the choice of regions for generalized belief propagation. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI-2004)*, pages 585–592, Banff, Canada, 2004. Morgan Kaufmann Publishers.

[33] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley Series in Probability and Mathematical Statistics, New York, 1991.

[34] E. Xing and M. I. Jordan. Graph partition strategies for generalized mean field inference. Technical Report CSD-03-1274, Division of Computer Science, University of California, Berkeley, 2003.

[35] E. Xing, M. I. Jordan, and S. Russell. Graph partition strategies for generalized mean field inference. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI-2004)*, pages 602–610, Banff, Canada, 2004. Morgan Kaufmann Publishers.

[36] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. Technical Report TR-2001-22, Mitsubishi Electric Research Laboratories, November 2001.

[37] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. Technical Report TR-2002-35, Mitsubishi Electric Research Laboratories, August 2002.

[38] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. Technical Report TR-2004-040, Mitsubishi Electric Research Laboratories, May 2004.

[39] A. Yuille. A double-loop algorithm to minimize the Bethe and Kikuchi free energies. *Neural Computation*, 14(6):1691–1722, 2001.