
Marginal Densities, Factor Graph Duality, and High-Temperature Series Expansions

Mehdi Molkaiaie

Department of Statistical Sciences
University of Toronto

Abstract

We prove that the marginal densities of a global probability mass function in a primal normal factor graph and the corresponding marginal densities in the dual normal factor graph are related via local mappings. The mapping depends on the Fourier transform of the local factors of the models. Details of the mapping, including its fixed points, are derived for the Ising model, and then extended to the Potts model. By employing the mapping, we can transform *simultaneously* all the estimated marginal densities from one domain to the other, which is advantageous if estimating the marginals can be carried out more efficiently in the dual domain. An example of particular significance is the ferromagnetic Ising model in a positive external field, for which there is a rapidly mixing Markov chain (called the subgraphs-world process) to generate configurations in the dual normal factor graph of the model. Our numerical experiments illustrate that the proposed procedure can provide more accurate estimates of marginal densities in various settings.

1 Introduction

In any probabilistic inference problem, one of the main objectives is to compute the local marginal densities of a global probability mass function (PMF). Such a computation in general requires a summation with an exponential number of terms, which makes its exact computation intractable [Dagum and Luby, 1993].

Our approach for estimating marginal densities hinges

on the notions of the normal realization (in which there is an edge for every variable) [Forney, 2001], the normal factor graph (NFG), and the dual NFG. The NFG duality theorem states that the partition function of a primal NFG and the partition function of its dual are equal up to some known scale factor [Al-Bashabsheh and Mao, 2011, Forney, 2011]. It has been demonstrated that, in the low-temperature regime, Monte Carlo methods for estimating the partition function converge faster in the dual NFG than in the primal NFG of the two-dimensional (2D) Ising model [Molkaiaie and Loeliger, 2013] and of the q -state Potts model [Al-Bashabsheh and Mao, 2014, Molkaiaie and Gómez, 2018].

In this paper, we prove that marginal densities of a global PMF of a primal NFG and the corresponding marginals of the dual NFG are related via local mappings. Remarkably, the mapping is independent of the size of the model, of the topology of the graph, and of any assumptions on the parameters of the model.

Each marginal density can of course be expressed as a ratio of two partition functions. In non-homogeneous models, each ratio needs to be estimated *separately* via variational inference algorithms or via Monte Carlo methods. However, our proposed mapping allows a *simultaneous* transformation of estimated marginal densities from one domain to the other.

The mapping is practically advantageous if computing such estimates can be done more efficiently in the dual NFG than in the primal NFG. Indeed, for the ferromagnetic Ising model in a positive external field there is a rapidly mixing Markov chain (called the subgraphs-world process) to generate configurations in the dual NFG of the Ising model. As models, we mainly focus on binary models with symmetric pairwise interactions (e.g., the Ising model). However, we will briefly discuss extensions of the proposed mappings to non-binary models (e.g., the q -state Potts model).

Next, we will describe our models in the primal and in the dual domains.

2 THE PRIMAL MODEL

Suppose variables X_1, X_2, \dots, X_N are associated with the vertices (sites) of a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = N$ vertices and $|\mathcal{E}|$ edges (bonds). Two variables (X_i, X_j) interact if their corresponding vertices are connected by an edge in \mathcal{G} . Each variable takes values in $\mathcal{A} = \mathbb{Z}/2\mathbb{Z}$, i.e., the set of integers modulo two. We will mainly view \mathcal{A} as a group with respect to addition.

In the primal domain, the probability of a configuration $\mathbf{x} \in \mathcal{A}^N$ is given by

$$\pi(\mathbf{x}) \propto \prod_{(i,j) \in \mathcal{E}} \psi_{i,j}(x_i, x_j) \prod_{v \in \mathcal{V}} \phi_v(x_v). \quad (1)$$

Furthermore, we assume that each pairwise potential factor $\psi_{i,j}(\cdot)$ is only a function of $y_{i,j} = x_i - x_j$. To lighten notations we denote the index pair $(i, j) \in \mathcal{V}^2$ by a single index $e \in \mathcal{E}$. In the primal domain, we express the global probability mass function (PMF) as

$$\pi_p(\mathbf{x}) = \frac{1}{Z_p} \prod_{e \in \mathcal{E}} \psi_e(y_e) \prod_{v \in \mathcal{V}} \phi_v(x_v). \quad (2)$$

Here, the normalization constant Z_p is the partition function, $\{\psi_e: \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}, e \in \mathcal{E}\}$ are the *edge-weighting factors*, and $\{\phi_v: \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}, v \in \mathcal{V}\}$ are the *vertex-weighting factors* [Molkaia, 2017, Forney, 2018].

The factorization in (2) can be represented by an NFG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where vertices represent the factors and edges represent the variables. The edge that represents some variable y_e is connected to the vertex representing the factor $\psi_e(\cdot)$ if and only if y_e is an argument of $\psi_e(\cdot)$. If a variable appears in more than two factors, it is replicated using an equality indicator factor [Forney, 2001].

For a 2D lattice, the NFG of (2) is depicted in Fig. 1, in which the unlabeled boxes represent $\psi_e(\cdot)$, small unlabeled boxes represent $\phi_v(\cdot)$. In Fig. 1, boxes labeled “+” are instances of zero-sum indicator factors $I_+(\cdot)$, which impose the constraint that all their incident variables sum to zero, and boxes labeled “=” are instances of equality indicator factors $I_-(\cdot)$, which impose the constraint that all their incident variables are equal.

E.g., the equality indicator factor involving x_1, x'_1 , and x''_1 is given by

$$I_-(x_1, x'_1, x''_1) = \delta(x_1 - x'_1) \cdot \delta(x_1 - x''_1) \quad (3)$$

and the zero-sum indicator factor involving x_1, x_2 , and y_1 is as in

$$I_+(y_1, x_1, x_2) = \delta(y_1 + x_1 + x_2), \quad (4)$$

where $\delta(\cdot)$ is the Kronecker delta function. (Note that all arithmetic manipulations are modulo two.)

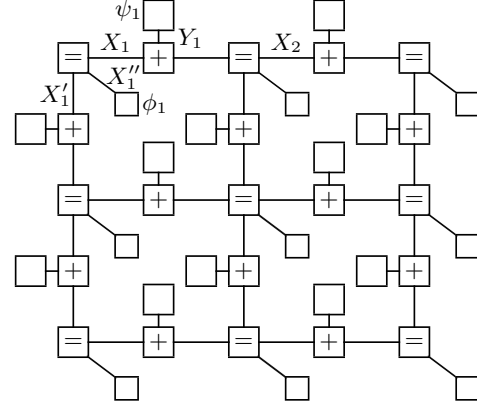


Figure 1: Primal NFG of the factorization (2).

In the primal NFG, variables include $\mathbf{X} = \{X_v: v \in \mathcal{V}\}$ and $\mathbf{Y} = \{Y_e: e \in \mathcal{E}\}$. However, these variables are not independent. Indeed, we can freely choose \mathbf{X} and therefrom fully determine \mathbf{Y} [Molkaia and Gómez, 2018, Forney, 2018]. E.g., if we take \mathcal{G} to be a d -dimensional lattice, we can compute each component Y_e of \mathbf{Y} by adding two components of \mathbf{X} that are incident to the corresponding zero-sum indicator factor (see Fig. 1).

The number of configurations in the primal domain is thus $|\mathcal{A}|^N$, and

$$Z_p = \sum_{\mathbf{x} \in \mathcal{A}^N} \prod_{e \in \mathcal{E}} \psi_e(y_e) \prod_{v \in \mathcal{V}} \phi_v(x_v). \quad (5)$$

The Ising model can be easily formulated via (2). In an Ising model the energy of a configuration \mathbf{x} is given by the Hamiltonian¹

$$\mathcal{H}(\mathbf{x}) = - \sum_{(i,j) \in \mathcal{E}} J_{i,j} \cdot (2\delta(x_i - x_j) - 1) - \sum_{v \in \mathcal{V}} H_v \cdot (2\delta(x_v) - 1), \quad (6)$$

which can be expressed as

$$\mathcal{H}(\mathbf{x}) = - \sum_{e \in \mathcal{E}} J_e \cdot (2\delta(y_e) - 1) - \sum_{v \in \mathcal{V}} H_v \cdot (2\delta(x_v) - 1). \quad (7)$$

Here J_e is the coupling parameter associated with the bond $e \in \mathcal{E}$ and H_v is the external field at site $v \in \mathcal{V}$. The model is called *homogeneous* if couplings are constant and *ferromagnetic* if $J_e \geq 0$ for all $e \in \mathcal{E}$.

The probability of \mathbf{x} is given by the Gibbs-Boltzmann distribution [Yeomans, 1992]

$$\pi_B(\mathbf{x}) \propto e^{-\beta \mathcal{H}(\mathbf{x})}, \quad (8)$$

¹In the bipolar case (i.e., when $\mathcal{X} = \{-1, +1\}$), the Hamiltonian is $\mathcal{H}(\mathbf{x}) = - \sum_{(i,j) \in \mathcal{E}} J_{i,j} x_i x_j - \sum_{1 \leq i \leq N} H_i x_i$.

where $\beta \in \mathbb{R}_{\geq 0}$ denotes the inverse temperature.

From (7) and (8), it is straightforward to obtain the edge-weighting factors of the Ising model as

$$\psi_e(y_e) = \begin{cases} e^{\beta J_e}, & \text{if } y_e = 0 \\ e^{-\beta J_e}, & \text{if } y_e = 1 \end{cases} \quad (9)$$

and the vertex-weighting factors as

$$\phi_v(x_v) = \begin{cases} e^{\beta H_v}, & \text{if } x_v = 0 \\ e^{-\beta H_v}, & \text{if } x_v = 1. \end{cases} \quad (10)$$

The Gibbs-Boltzmann distribution in (8) can therefore be expressed via the factorization (2).

3 THE DUAL MODEL

The dual NFG has the same topology as the primal NFG, but with factors replaced by the discrete Fourier transform (DFT) or the inverse DFT of corresponding factors in the primal NFG.

We can obtain the dual NFG of our binary models by replacing factors by their one-dimensional (1D) DFT, equality indicator factors by zero-sum indicator factors, and zero-sum indicator factors by equality indicator factors [Al-Bashabsheh and Mao, 2011, Molkaeraie and Loeliger, 2013, Molkaeraie, 2016].

We will use the tilde symbol to denote variables in the dual NFG, which also take values in \mathcal{A} .

The dual NFG of Fig. 1 is illustrated in Fig. 2, in which the unlabeled boxes represent $\tilde{\psi}_e: \mathcal{A} \rightarrow \mathbb{R}$, the 1D DFT of $\psi_e(\cdot)$, given by

$$\tilde{\psi}_e(\tilde{y}_e) = \begin{cases} \psi_e(0) + \psi_e(1), & \text{if } \tilde{y}_e = 0 \\ \psi_e(0) - \psi_e(1), & \text{if } \tilde{y}_e = 1 \end{cases} \quad (11)$$

and for $v \in \mathcal{V}$ the small unlabeled boxes are $\tilde{\phi}_v: \mathcal{A} \rightarrow \mathbb{R}$, the 1D DFT of $\phi_v(\cdot)$, as in

$$\tilde{\phi}_v(\tilde{x}_v) = \begin{cases} \phi_v(0) + \phi_v(1), & \text{if } \tilde{x}_v = 0 \\ \phi_v(0) - \phi_v(1), & \text{if } \tilde{x}_v = 1. \end{cases} \quad (12)$$

The set of variables in the dual domain consist of $\tilde{\mathbf{Y}} = \{\tilde{Y}_e: e \in \mathcal{E}\}$ and $\tilde{\mathbf{X}} = \{\tilde{X}_v: v \in \mathcal{V}\}$. Again, these variables are not independent as we can freely choose $\tilde{\mathbf{Y}}$ and therefrom fully determine $\tilde{\mathbf{X}}$. E.g., if we take \mathcal{G} to be a d -dimensional lattice and assume periodic boundaries, each component \tilde{X}_v of $\tilde{\mathbf{X}}$ can be computed by adding $2d$ components of $\tilde{\mathbf{Y}}$ that are incident to the corresponding zero-sum indicator factor (see Fig. 2).

In the dual NFG, the number of configurations is $|\mathcal{A}|^{|\mathcal{E}|}$, and its the partition function Z_d is given by

$$Z_d = \sum_{\tilde{\mathbf{y}} \in \mathcal{A}^{|\mathcal{E}|}} \prod_{e \in \mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v). \quad (13)$$

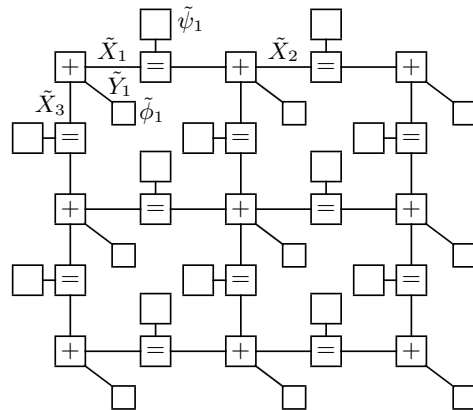


Figure 2: The dual of the NFG in Fig. 1.

On condition that factors (11) and (12) are nonnegative, we can define the global PMF in the dual NFG as

$$\pi_d(\tilde{\mathbf{Y}}) = \frac{1}{Z_d} \prod_{e \in \mathcal{E}} \tilde{\psi}_e(\tilde{y}_e) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v). \quad (14)$$

The dual Ising model can be expressed via (14). Indeed

$$\tilde{\psi}_e(\tilde{y}_e) = \begin{cases} 2 \cosh(\beta J_e), & \text{if } \tilde{y}_e = 0 \\ 2 \sinh(\beta J_e), & \text{if } \tilde{y}_e = 1, \end{cases} \quad (15)$$

in agreement with (9) and (11), and

$$\tilde{\phi}_v(\tilde{x}_v) = \begin{cases} 2 \cosh(\beta H_v), & \text{if } \tilde{x}_v = 0 \\ 2 \sinh(\beta H_v), & \text{if } \tilde{x}_v = 1, \end{cases} \quad (16)$$

in agreement with (10) and (12).

If the model is ferromagnetic (i.e., $J_e \geq 0$) and in a nonnegative external field (i.e., $H_v \geq 0$), factors (15) and (16) will be nonnegative. In this case, the global PMF of the dual Ising model is given by (14).

Throughout this paper, we assume that each edge connects two distinct vertices of the NFG (i.e., there are no dangling edges with one end attached to a vertex and the other end free). In this setting, according to the NFG duality theorem, the partition functions Z_p and Z_d are equal up to some scale factor $\alpha(\mathcal{G})$. Indeed

$$Z_d = \alpha(\mathcal{G}) \cdot Z_p, \quad (17)$$

where $\alpha(\mathcal{G})$ only depends on the topology of \mathcal{G} .

For more details, see [Al-Bashabsheh and Mao, 2011], [Molkaeraie, 2017, Appendix], [Forney, 2018, Thm 8].

4 THE DUAL ISING MODEL AND HIGH-TEMPERATURE SERIES EXPANSIONS

In [Jerrum and Sinclair, 1993], the authors proposed a rapidly mixing Markov chain (called the subgraphs-world process) which evaluates the partition function

of an arbitrary ferromagnetic Ising model in a positive external field to any specified degree of accuracy.

The mixing time of the process is polynomial in the size of the model at *all* temperatures. Indeed, the expected running time of the generator of the subgraphs-world process is $\mathcal{O}(|\mathcal{E}|^2 N^8 (\log \delta^{-1} + |\mathcal{E}|))$, where δ is the confidence parameter. For more details, see [Jerrum and Sinclair, 1993, Section 4].

The subgraphs-world process employs the following expansion of Z defined on the set of edges $\mathcal{W} \subseteq \mathcal{E}$ in powers of $\tanh(H)$ and $\tanh(J_e)$ as

$$Z \propto \sum_{\mathcal{W} \subseteq \mathcal{E}} \tanh(H)^{|\text{odd}(\mathcal{W})|} \prod_{e \in \mathcal{W}} \tanh(J_e), \quad (18)$$

where $\text{odd}(\mathcal{W})$ denotes the set of all odd-degree vertices in the subgraph of \mathcal{E} induced by \mathcal{W} . The expansion (18) is known as the high-temperature series expansion of the partition function [Newell and Montroll, 1953, Yeomans, 1992, Grimmett and Janson, 2009].

Proposition 1. The configurations that arise in the high-temperature series expansion of the partition function (which are the configurations of the subgraphs-world process) coincide with the valid configurations in the dual NFG of the Ising model.

See [Molkaraie and Gómez, 2018, Section VIII] and [Forney, 2018, Section III-E] for the proof.

Following Proposition 1, we can employ the subgraphs-world process (as a generator for the subgraphs-world configurations) to generate configurations in the dual NFG of the Ising model. The process is rapidly mixing and therefore converges in polynomial time. However, under reasonable complexity assumptions, there is no generalization of this approximation scheme to the (non-binary) Potts model or to spin glasses. For more details, see [Goldberg and Jerrum, 2012, Galanis et al., 2016].

Next, we will present local (edge-based) mappings that transform marginal densities from the dual NFG to the primal NFG, or vice versa. The mappings depend on the DFT of the local factors of the models.

5 MARGINAL DENSITIES IN THE PRIMAL AND DUAL DOMAINS

The edge marginal PMF of $e \in \mathcal{E}$ in the primal NFG can be computed as

$$\pi_{p,e}(a) = \frac{Z_{p,e}(a)}{Z_p}, \quad a \in \mathcal{A}, \quad (19)$$

where

$$Z_{p,e}(a) = \sum_{\mathbf{x} \in \mathcal{A}^N} \delta(y_e - a) \prod_{e' \in \mathcal{E}} \psi_{e'}(y_{e'}) \prod_{v \in \mathcal{V}} \phi_v(x_v).$$

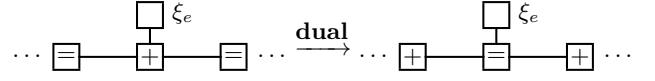


Figure 3: The edge $e \in \mathcal{E}$ in the intermediate primal NFG (left) and in the intermediate dual NFG (right). The unlabeled box (left) represents (22) and the unlabeled box (right) represents (23).

Hence

$$Z_{p,e}(a) = \psi_e(a) S_e(a), \quad (20)$$

with

$$S_e(a) = \sum_{\mathbf{x} \in \mathcal{A}^N} \delta(y_e - a) \prod_{e' \in \mathcal{E} \setminus e} \psi_{e'}(y_{e'}) \prod_{v \in \mathcal{V}} \phi_v(x_v). \quad (21)$$

Here, $Z_{p,e}(a) \geq 0$ and $Z_p = \sum_{a \in \mathcal{A}} Z_{p,e}(a) = \sum_{a \in \mathcal{A}} \psi_e(a) S_e(a)$, hence (19) is a valid PMF over \mathcal{A}^N .

In coding theory terminology, $\{\psi_e(a), a \in \mathcal{A}\}$ is called the *intrinsic* message vector and $\{S_e(a), a \in \mathcal{A}\}$ is called the *extrinsic* message vector at edge $e \in \mathcal{E}$. According to the sum-product message passing update rule, the edge marginal PMF vector is computed as the dot product of the intrinsic and extrinsic message vectors up to scale. The scale factor is equal to the partition function Z_p [Forney, 2001, Kschischang et al., 2001].

In our setup, $S_e(a)$ is the partition function of an intermediate primal NFG with all factors as in the primal NFG, excluding the factor $\psi_e(y_e)$, which is replaced by

$$\xi_e(y_e; a) = \delta(y_e - a). \quad (22)$$

Fig. 3 (left) shows the corresponding edge in the intermediate primal NFG. The intermediate dual NFG is shown in Fig. 3 (right), in which the factor $\tilde{\psi}_e(\tilde{y}_e)$ is replaced by

$$\tilde{\xi}_e(\tilde{y}_e; a) = \begin{cases} \delta(a) + \delta(1 - a), & \text{if } \tilde{y}_e = 0 \\ \delta(a) - \delta(1 - a), & \text{if } \tilde{y}_e = 1, \end{cases} \quad (23)$$

which is the 1D DFT of (22). According to the NFG duality theorem (17), the partition function of the intermediate dual NFG is $\alpha(\mathcal{G}) \cdot S_e(a)$.

Similarly, in the dual NFG the edge marginal PMF of $e \in \mathcal{E}$ is

$$\pi_{d,e}(a') = \frac{Z_{d,e}(a')}{Z_d}, \quad a' \in \mathcal{A}. \quad (24)$$

Hence

$$\begin{aligned} Z_{d,e}(a') &= \tilde{\psi}_e(a') \cdot \\ &\left(\sum_{\tilde{\mathbf{y}} \in \mathcal{A}^{|\mathcal{E}|}} \delta(\tilde{y}_e - a') \prod_{e' \in \mathcal{E} \setminus e} \tilde{\psi}_{e'}(\tilde{y}_{e'}) \prod_{v \in \mathcal{V}} \tilde{\phi}_v(\tilde{x}_v) \right) \\ &= \tilde{\psi}_e(a') \tilde{S}_e(a'). \end{aligned} \quad (25)$$

Proposition 2. The vectors $\{S_e(a), a \in \mathcal{A}\}$ and $\{\tilde{S}_e(a'), a' \in \mathcal{A}\}$ are DFT pairs.

Proof. For $a \in \mathcal{A}$, the partition function of the intermediate dual NFG is the dot product of message vectors $\{\tilde{\xi}_e(a'; a), a' \in \mathcal{A}\}$ and $\{\tilde{S}_e(a'), a' \in \mathcal{A}\}$. Thus

$$\alpha(\mathcal{G}) \cdot S_e(a) = \sum_{a' \in \mathcal{A}} \tilde{\xi}_e(a'; a) \tilde{S}_e(a'), \quad (26)$$

which gives

$$\alpha(\mathcal{G}) \cdot S_e(a) = (\tilde{S}_e(0) + \tilde{S}_e(1)) \cdot \delta(a) + (\tilde{S}_e(0) - \tilde{S}_e(1)) \cdot \delta(1 - a). \quad (27)$$

After setting $a = 0$ and $a = 1$ in (27), we obtain

$$\begin{bmatrix} S_e(0) \\ S_e(1) \end{bmatrix} = \frac{1}{\alpha(\mathcal{G})} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \tilde{S}_e(0) \\ \tilde{S}_e(1) \end{bmatrix} \quad (28)$$

which is an instance of the two-point DFT. \blacksquare

Proposition 3. The vectors $\{\pi_{p,e}(a)/\psi_e(a), a \in \mathcal{A}\}$ and $\{\pi_{d,e}(a')/\tilde{\psi}_e(a'), a' \in \mathcal{A}\}$ are DFT pairs.

Proof. From (19) and (21) we have

$$S_e(a) = Z_p \cdot \frac{\pi_{p,e}(a)}{\psi_e(a)}, \quad a \in \mathcal{A}. \quad (29)$$

But (17), (24), and (25) yield

$$\tilde{S}_e(a') = Z_d \cdot \frac{\pi_{d,e}(a')}{\tilde{\psi}_e(a')} \quad (30)$$

$$= \alpha(\mathcal{G}) \cdot Z_p \cdot \frac{\pi_{d,e}(a')}{\tilde{\psi}_e(a')}, \quad a' \in \mathcal{A}. \quad (31)$$

Putting (31) and (29) in (28), and after a little rearranging, we obtain the following mapping

$$\begin{bmatrix} \pi_{p,e}(0)/\psi_e(0) \\ \pi_{p,e}(1)/\psi_e(1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \pi_{d,e}(0)/\tilde{\psi}_e(0) \\ \pi_{d,e}(1)/\tilde{\psi}_e(1) \end{bmatrix} \quad (32)$$

in matrix-vector format via the two-point DFT. \blacksquare

By virtue of Proposition 3, it is possible to estimate edge marginal densities in one domain, and then transform them to the other domain all together. The mapping is fully local, and is independent of the size of the graph N and of the topology of \mathcal{G} . (Indeed, the relevant information regarding the rest of the graph is incorporated in the estimated edge marginal densities.)

We state without proof that

Proposition 4. The vectors $\{\pi_{p,v}(a)/\phi_v(a), a \in \mathcal{A}\}$ and $\{\pi_{d,v}(a')/\tilde{\phi}_v(a'), a' \in \mathcal{A}\}$ are DFT pairs.

6 DETAILS OF THE MAPPING FOR THE ISING MODEL

For the general Ising model substituting factors (9) and (15) in (32) yields

$$\begin{bmatrix} \pi_{p,e}(0) \\ \pi_{p,e}(1) \end{bmatrix} = \begin{bmatrix} \frac{e^{\beta J_e}}{2 \cosh(\beta J_e)} & \frac{e^{\beta J_e}}{2 \sinh(\beta J_e)} \\ \frac{e^{-\beta J_e}}{2 \cosh(\beta J_e)} & -\frac{e^{-\beta J_e}}{2 \sinh(\beta J_e)} \end{bmatrix} \cdot \begin{bmatrix} \pi_{d,e}(0) \\ \pi_{d,e}(1) \end{bmatrix} \quad (33)$$

for $\beta J_e \neq 0$.

Let us consider a homogeneous and ferromagnetic Ising model. A straightforward calculation shows that the fixed points of the mapping (33) are given by

$$\begin{bmatrix} \pi_{p,e}^*(0) & \pi_{p,e}^*(1) \end{bmatrix} = \begin{bmatrix} \frac{e^{\beta J} \cosh(\beta J)}{1 + \sinh(2\beta J)} & \frac{e^{-\beta J} \sinh(\beta J)}{1 + \sinh(2\beta J)} \end{bmatrix} \quad (34)$$

Fig. 4 shows the fixed points (34) as a function of βJ .

Proposition 5. The min of $\pi_{p,e}^*(0)$ and the max of $\pi_{p,e}^*(1)$ are attained at the criticality of the 2D homogeneous Ising model without an external field.

Proof. In the thermodynamic limit (i.e., as $N \rightarrow \infty$) the 2D Ising model undergoes a phase transition at $\beta J_c = \ln(1 + \sqrt{2})/2 \approx 0.44$ [Onsager, 1944].

In the absence of an external field and for $\beta J = 1$, the Hamiltonian (7) can be expressed as

$$\mathcal{H}(\mathbf{y}) = - \sum_{e \in \mathcal{E}} (2\delta(y_e) - 1) \quad (35)$$

$$= - \sum_{e \in \mathcal{E}} (1 - 2y_e), \quad (36)$$

where $y_e = x_i - x_j$ for $e = (i, j) \in \mathcal{E}$.

The average energy is equal to

$$\bar{\mathcal{H}}(\mathbf{y}) = \sum_{\mathbf{y} \in \mathcal{A}} \pi_{\mathbf{p}}(\mathbf{y}) \mathcal{H}(\mathbf{y}) \quad (37)$$

$$= -|\mathcal{E}| \cdot (1 - 2\mathbb{E}[Y_e]) \quad (38)$$

$$= -|\mathcal{E}| \cdot (1 - 2\pi_{p,e}(1)). \quad (39)$$

In the 2D Ising model with periodic boundaries $|\mathcal{E}| = 2N$, thus the average energy per site is given by

$$\bar{\mathcal{H}}(\mathbf{y})/N = -2(1 - 2\pi_{p,e}(1)). \quad (40)$$

From Onsager's closed-form solution

$$\lim_{N \rightarrow \infty} \frac{\ln Z}{N} = \frac{1}{2} \ln(2 \cosh^2 2\beta J) + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \ln(1 + \sqrt{1 - \kappa^2 \sin^2 \theta}) d\theta \quad (41)$$

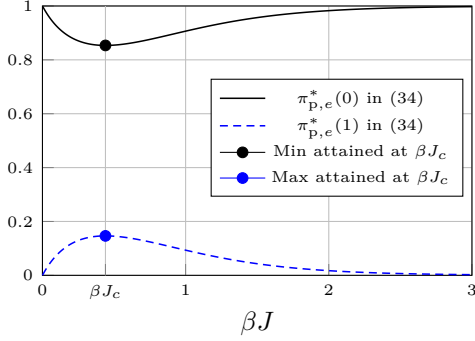


Figure 4: The fixed points (34) as a function of βJ . The filled circles show the fixed points at criticality of the 2D Ising model given by (46).

and the average (internal) energy per site is given by

$$U(\beta J) = \lim_{N \rightarrow \infty} -\frac{1}{N} \cdot \frac{\partial \ln Z}{\partial \beta J} \quad (42)$$

$$= -\coth(2\beta J).$$

$$\left(1 - \frac{1}{2\pi} (1 - \kappa \sinh 2\beta J) \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1 - \kappa^2 \sin^2 \theta}}\right) \quad (43)$$

with

$$\kappa(\beta J) = \frac{2 \sinh 2\beta J}{\cosh^2 2\beta J}. \quad (44)$$

See [Onsager, 1944], [Baxter, 2007, Chapter 7] for more details.

A routine calculation shows that $\kappa(\beta J_c) = 1$, thus

$$U(\beta J_c) = -\sqrt{2}. \quad (45)$$

From (40) and (45), we obtain $\pi_{p,e}^*(1) = (2 - \sqrt{2})/4$. Therefore, at criticality

$$\left[\pi_{p,e}^*(0) \quad \pi_{p,e}^*(1) \right] = \left[(2 + \sqrt{2})/4 \quad (2 - \sqrt{2})/4 \right], \quad (46)$$

which coincides with the min of $\pi_{p,e}^*(0)$ and the max of $\pi_{p,e}^*(1)$. We emphasize that in the 2D homogeneous Ising model in zero field and in the thermodynamic limit (i.e., as $N \rightarrow \infty$), edge marginal densities in the primal and dual domains are equal at criticality. ■

The fixed points $\pi_{p,e}^*(0)$ and $\pi_{p,e}^*(1)$ at criticality in (46) are illustrated by filled circles in Fig. 4.

Proposition 6. In an arbitrary ferromagnetic Ising model in a nonnegative external field, it holds that

$$\pi_{p,e}(0) \geq \frac{1}{1 + e^{-2\beta J_e}} \quad (47)$$

and

$$\pi_{d,e}(0) \geq \frac{1 + e^{-2\beta J_e}}{2}. \quad (48)$$

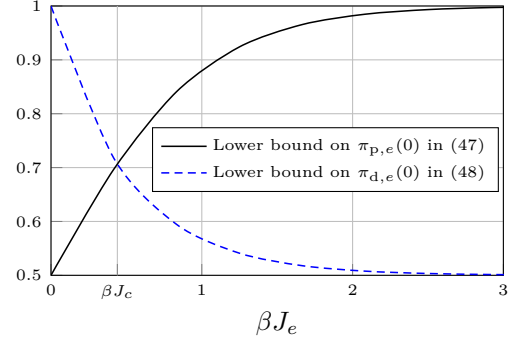


Figure 5: For a ferromagnetic Ising model in a nonnegative external field, the solid black plot and the dashed blue plot show the lower bound on $\pi_{p,e}(0)$, given by (47), and the lower bound on $\pi_{d,e}(0)$, given by (48), as a function of βJ_e , respectively. The lower bounds intersect at the criticality of the 2D homogeneous Ising model in zero field, denoted by βJ_c .

Proof. Since the Ising model is ferromagnetic and in a nonnegative external field, we can define the global PMF $\pi_{d,e}(\cdot)$ in the dual domain as in (14). From (33), we have

$$\frac{\pi_{p,e}(0)}{e^{\beta J_e}} = \frac{\pi_{d,e}(0)}{2 \cosh(\beta J_e)} + \frac{\pi_{d,e}(1)}{2 \sinh(\beta J_e)} \quad (49)$$

$$= \frac{1}{2 \sinh(\beta J_e)} - \frac{e^{-\beta J_e}}{\sinh(2\beta J_e)} \pi_{d,e}(0). \quad (50)$$

We conclude from (50) that $\pi_{p,e}(0)$ achieves its minimum when $\pi_{d,e}(0) = 1$. After substituting $\pi_{d,e}(0) = 1$ in (50), and after a little rearranging, we obtain

$$\pi_{p,e}(0) \geq \frac{e^{\beta J_e}}{2 \cosh(\beta J_e)} \quad (51)$$

$$= \frac{1}{1 + e^{-2\beta J_e}}. \quad (52)$$

The proof of (48) follows along the same lines. ■

Proposition 6 is valid for arbitrary ferromagnetic Ising models in a nonnegative external magnetic field, i.e., the bonds do not depend on N (the size of the graphical model \mathcal{G}) and on the topology of \mathcal{G} .

Fig. 5 shows the lower bounds in (47) and (48) as a function of βJ_e . The lower bounds intersect at βJ_c , i.e., at the criticality of the 2D homogeneous Ising model in the absence of an external field.

Remark 1. From (47) and (48) we conclude that in an arbitrary ferromagnetic Ising model in a nonnegative external field

$$\pi_{p,e}(0) \pi_{d,e}(0) \geq \frac{1}{2}, \quad (53)$$

which is in the form of an uncertainty principle.

7 GENERALIZATION TO NON-BINARY MODELS

We briefly discuss extensions of our mapping to non-binary models, in particular to the q -state Potts model [Wu, 1982, Baxter, 2007]. Accordingly, we let $\mathcal{A} = \mathbb{Z}/q\mathbb{Z}$ for some integer $q \geq 2$. (The binary Ising model is recovered as the special case $q = 2$.)

In the absence of an external field, the Hamiltonian of the model is given by

$$\mathcal{H}(\mathbf{x}) = - \sum_{e \in \mathcal{E}} J_e \cdot \delta(y_e). \quad (54)$$

From (8) and (54), we obtain that in the primal NFG

$$\psi_e(y_e) = \begin{cases} e^{\beta J_e}, & \text{if } y_e = 0 \\ 1, & \text{otherwise,} \end{cases} \quad (55)$$

and in the dual NFG, factors are equal to the 1D DFT of (55) given by

$$\tilde{\psi}_e(\tilde{y}_e) = \begin{cases} e^{\beta J_e} - 1 + q, & \text{if } \tilde{y}_e = 0 \\ e^{\beta J_e} - 1, & \text{otherwise,} \end{cases} \quad (56)$$

which is nonnegative if the model is ferromagnetic (i.e., $J_e \geq 0$). See [Al-Bashabsheh and Mao, 2014, Molkaeraie and Gómez, 2018] for more details on constructing the primal and the dual NFG of the Potts model, with or without an external field.

A straightforward generalization Proposition (3) gives the mapping between $\{\pi_{p,e}(a)/\psi_e(a), a \in \mathcal{A}\}$ and $\{\pi_{d,e}(a')/\tilde{\psi}_e(a'), a' \in \mathcal{A}\}$ via $W_q = \{w_{k,\ell}, k, \ell \in \mathcal{A}\}$ with $w_{k,\ell} = e^{\frac{-2\pi i}{|\mathcal{A}|} k\ell}$, where W_q is the q -point DFT matrix (i.e., the Vandermonde matrix for the roots of unity) and $i = \sqrt{-1}$ (see [Bracewell, 1999]).

However, due to symmetry in the factors of the primal (55) and the dual Potts model (56), we have

$$\pi_{p,e}(1)/\psi_e(1) = \dots = \pi_{p,e}(q-1)/\psi_e(q-1) \quad (57)$$

and

$$\pi_{d,e}(1)/\tilde{\psi}_e(1) = \dots = \pi_{d,e}(q-1)/\tilde{\psi}_e(q-1). \quad (58)$$

Thus, e.g., for $q = 3$, the mapping yields

$$\begin{aligned} \frac{\pi_{p,e}(1)}{\psi_e(1)} &= \frac{\pi_{d,e}(0)}{\tilde{\psi}_e(0)} + \frac{\pi_{d,e}(1)}{\tilde{\psi}_e(1)} e^{-\frac{2\pi i}{3}} + \frac{\pi_{d,e}(2)}{\tilde{\psi}_e(2)} e^{-\frac{4\pi i}{3}} \\ &= \frac{\pi_{d,e}(0)}{\tilde{\psi}_e(0)} - \frac{\pi_{d,e}(1)}{\tilde{\psi}_e(1)}, \end{aligned} \quad (59)$$

which is real-valued.

For a homogeneous and ferromagnetic Potts model, the fixed points of $\pi_{p,e}^*(0)$ can be derived as

$$\pi_{p,e}^*(0) = \frac{e^{\beta J}(e^{\beta J} - 1 + q)}{e^{2\beta J} - 2(1 - q)e^{\beta J} + 1 - q}. \quad (60)$$

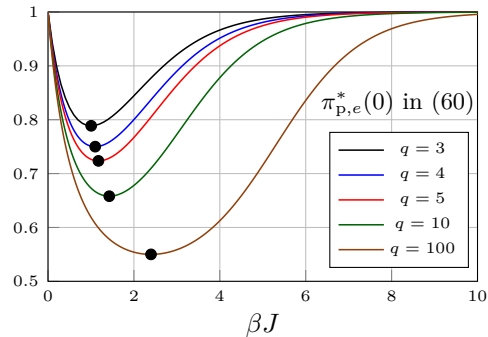


Figure 6: The fixed points (60) as a function of βJ for different values of q . The filled circles show the fixed points at criticality of the 2D Potts model located at $\beta J_c = \ln(1 + \sqrt{q})$.

Fig. 6 shows the fixed point (60) as a function of βJ . Like the Ising model, the minimum of $\pi_{p,e}^*(0)$ is attained at the criticality of the 2D Potts model located at $\beta J_c = \ln(1 + \sqrt{q})$ [Wu, 1982]. In a similar way, one can obtain the fixed points of $\pi_{p,e}^*(t)$, for $t \in \{1, 2, \dots, q-1\}$.

Remark 2. Transforming marginals from one domain to the other requires a matrix-vector multiplication with computational complexity $\mathcal{O}(|\mathcal{A}|^2)$. However, when there is symmetry in the factors, as in (9) and (55), the complexity can be reduced to $\mathcal{O}(|\mathcal{A}|)$.

Remark 3. In binary models, factors in the dual NFG can in general take negative values, and in nonbinary models, the factors can be complex-valued. In such cases a valid PMF can no longer be defined in the dual domain. The mappings remain nevertheless valid; but for *marginal functions* (instead of marginal densities) of a global function with a factorization given by (14).

8 NUMERICAL EXPERIMENTS

In both domains estimates of marginal densities can be obtained via Monte Carlo methods or via variational algorithms [Robert and Casella, 2004, Murphy, 2012]. We only consider the subgraphs-world process (SWP) and two variational algorithms, the belief propagation (BP) and the tree expectation propagation (TEP), for the Ising model. Estimated marginals in the dual domain are then transformed all together to the primal domain via (32) and (59). In all experiments, the exact marginal densities are computed via the junction tree algorithm implemented in [Mooij, 2010].

The choice of methods and the models is far from exhaustive – our goal is to show the advantage of using the mappings in approximating marginal densities in similar settings. Additional numerical experiments are reported in the Supplementary Material.

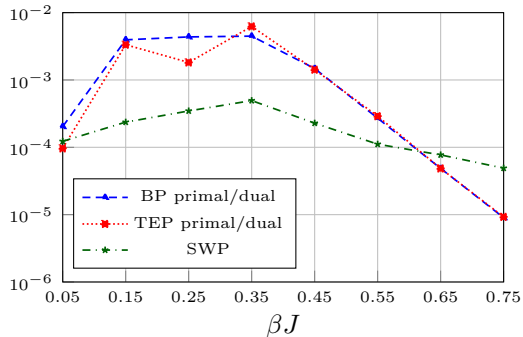


Figure 7: Relative error as a function of βJ in estimating $\pi_{p,e}(0)$ of a homogeneous Ising model in a constant external field $\beta H = 0.15$, with periodic boundaries, and with size $N = 6 \times 6$.

In our first experiment, we consider a 2D homogeneous Ising model, in a constant external field $\beta H = 0.15$, with periodic boundaries, and with size $N = 6 \times 6$. For this model, BP and TEP in the primal and in the dual domains give virtually indistinguishable approximations. We also apply SWP using 10^5 samples. Fig. 7 shows the relative error in estimating $\pi_{p,e}(0)$ as a function of βJ , where SWP (which operates in the dual NFG) gives good estimates in the whole range.

Compared to variational algorithms, convergence of the SWP is slow; moreover, SWP is only applicable to ferromagnetic Ising models in a positive field. (Indeed, in order to have an irreducible Markov chain in the SWP the external field needs to be non-zero [Welsh, 1993, Chapter 8]). In the rest of the experiments, we consider Ising models in the absence of an external field, and only compare the efficiency of variational algorithms employed in the primal and in the dual domains.

In the second experiment, we consider a 2D Ising model with size $N = 6 \times 6$ and with periodic boundaries. Couplings are chosen randomly according to a half-normal distribution, i.e., $\beta J_e = |\beta J'_e|$ with $\beta J'_e \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. Fig. 8 shows the average relative error in estimating the marginal density $\pi_{p,e}(0)$ as a function of σ^2 , where the results are averaged over 200 independent realizations. We consider a fully-connected Ising model with $N = 10$ in our last experiment. Couplings are chosen randomly according to $\beta J_e \stackrel{i.i.d.}{\sim} \mathcal{U}[0.05, \beta J_x]$, i.e., uniformly between 0.05 and βJ_x denoted by the x -axis. The average relative error over 50 independent realizations is illustrated in Fig. 9.

In both experiments, BP and TEP provide close approximations in the dual domain, therefore only BP results are reported. Figs. 8 and 9 show that for $\sigma^2 > 0.25$ and $\beta J_x > 0.20$, BP in the dual NFG can significantly improve the quality of estimates – even by more than two orders of magnitude in terms of relative error.

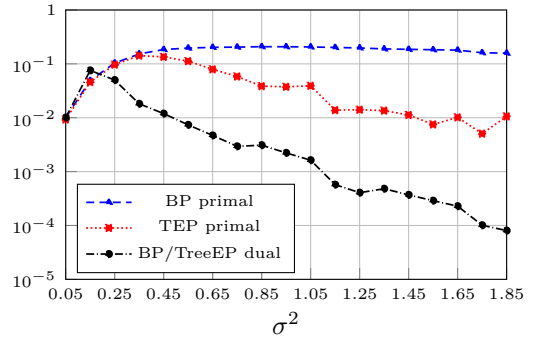


Figure 8: Average relative error in estimating $\pi_{p,e}(0)$ of an Ising model with periodic boundaries and with size $N = 6 \times 6$. Couplings are chosen randomly according to a half-normal distribution with variance σ^2 .

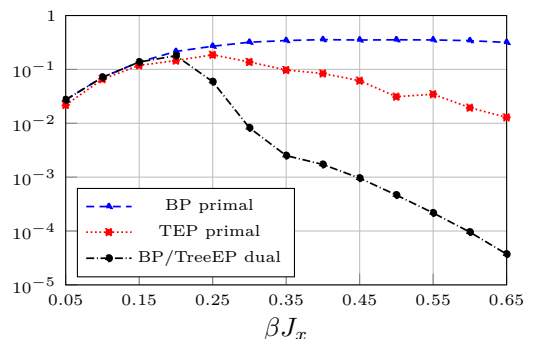


Figure 9: Average relative error in estimating $\pi_{p,e}(0)$ in a fully-connected Ising model with $N = 10$. Coupling parameters are chosen uniformly and independently between 0.05 and βJ_x denoted by the x -axis.

9 CONCLUSION

We proved that marginals densities of a primal NFG and the corresponding marginal densities of its dual NFG are related via local mappings. The mapping provides a simple procedure to transform simultaneously the estimated marginals from one domain to the other. Furthermore, the mapping relies on no assumptions on the size or on the topology of the graphical model. Our numerical experiments show that estimating the marginals in the dual NFG can sometimes significantly improve the quality of approximations in terms of relative error. In the special case of the ferromagnetic Ising model in a positive external field, there is indeed a rapidly mixing Markov chain (the subgraphs-world process) to generate configurations in the dual domain.

Acknowledgements

The author is extremely grateful to G. D. Forney, Jr., for his comments and for his continued support. The author also wishes to thank J. Dauwels, P. Vontobel, V. Gómez, and B. Ghojogh for their comments.

References

- [Al-Bashabsheh and Mao, 2011] Al-Bashabsheh, A. and Mao, Y. (2011). Normal factor graphs and holographic transformations. *IEEE Transactions on Information Theory*, 57:752–763.
- [Al-Bashabsheh and Mao, 2014] Al-Bashabsheh, A. and Mao, Y. (2014). On stochastic estimation of the partition function. *Proc. IEEE International Symposium on Information Theory*, pages 1504–1508.
- [Baxter, 2007] Baxter, R. J. (2007). *Exactly Solved Models in Statistical Mechanics*. Dover Publications.
- [Bracewell, 1999] Bracewell, R. N. (1999). *The Fourier Transform and Its Applications*. McGraw-Hill.
- [Dagum and Luby, 1993] Dagum, P. and Luby, M. (1993). Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153.
- [Forney, 2001] Forney, G. D. (2001). Codes on graphs: Normal realization. *IEEE Transactions on Information Theory*, 47:520–548.
- [Forney, 2011] Forney, G. D. (2011). Codes on graphs: Duality and MacWilliams identities. *IEEE Transactions on Information Theory*, 57:1382–1397.
- [Forney, 2018] Forney, G. D. (2018). Codes on graphs: Models for elementary algebraic topology and statistical physics. *IEEE Transactions on Information Theory*, 64:7465–7487.
- [Galanis et al., 2016] Galanis, A., Stefankovic, D., Vigoda, E., and Yang, L. (2016). Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45:2004–2065.
- [Goldberg and Jerrum, 2012] Goldberg, L. A. and Jerrum, M. (2012). Approximating the partition function of the ferromagnetic Potts model. *Journal of the ACM*, 59:1222–1239.
- [Grimmett and Janson, 2009] Grimmett, G. and Janson, S. (2009). Random even graphs. *The Electronic Journal of Combinatorics*, 16:R46.
- [Jerrum and Sinclair, 1993] Jerrum, M. and Sinclair, A. (1993). Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116.
- [Kschischang et al., 2001] Kschischang, F. R., Frey, B. J., and Loeliger, H.-A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47:498–519.
- [Molkaeraie, 2016] Molkaeraie, M. (2016). An importance sampling algorithm for the Ising model with strong couplings. *Proc. International Zurich Seminar on Communications*, pages 180–184.
- [Molkaeraie, 2017] Molkaeraie, M. (2017). The primal versus the dual Ising model. *Proc. 55th Annual Allerton Conference on Communication, Control, and Computing*, pages 53–60.
- [Molkaeraie and Gómez, 2018] Molkaeraie, M. and Gómez, V. (2018). Monte Carlo methods for the ferromagnetic Potts model using factor graph duality. *IEEE Transactions on Information Theory*, 59:7449–7464.
- [Molkaeraie and Loeliger, 2013] Molkaeraie, M. and Loeliger, H.-A. (2013). Partition function of the Ising model via factor graph duality. *Proc. IEEE International Symposium on Information Theory*, pages 2304–2308.
- [Mooij, 2010] Mooij, J. M. (2010). libdai: A free and open source C++ library for discrete approximate inference in graphical models. *The Journal of Machine Learning Research*, 11:2169–2173.
- [Murphy, 2012] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- [Newell and Montroll, 1953] Newell, G. F. and Montroll, E. W. (1953). On the theory of the Ising model of ferromagnetism. *Reviews of Modern Physics*, 25:353–389.
- [Onsager, 1944] Onsager, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65:117–149.
- [Robert and Casella, 2004] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.
- [Welsh, 1993] Welsh, D. J. A. (1993). *Complexity: Knots, Colourings and Countings*. Cambridge University Press.
- [Wu, 1982] Wu, F.-Y. (1982). The Potts model. *Reviews of Modern Physics*, 54:235–268.
- [Yeomans, 1992] Yeomans, J. M. (1992). *Statistical Mechanics of Phase Transitions*. Oxford University Press.