Covariance and entropy in Markov random fields

Matthew G. Reyes
mgreyes@umich.edu

***Abstract*—We consider families of Markov random fields (MRFs) on an undirected graph using the exponential family representation. In earlier work [13] we proved that if the statistic that defines a family of MRFs is positively correlated, then the entropy is monotone decreasing in the exponential parameters. In this paper we address the converse, specifically within the context of the Ising model. The statistic for an edge is viewed as positive or negative as it favors similar or dissimilar values at the endpoints of the edge. We show that for an acyclic Ising model with no self statistics, the statistic is positively correlated regardless of the polarity of the edges. We further show that for a cyclic Ising model, the statistic is positively correlated if and only if the statistic is not frustrated; and that the entropy is monotone decreasing in the exponential parameters, if and only if the statistic is not frustrated.**

I. Preamble

In this paper we pick up the discussion started in [13] and continued in [14]; namely, examining the relationship between the statistic defining a family of Markov random fields (MRFs) and the behavior of information-theoretic quantities within that family of MRFs. In particular, we address the question of whether monotonicity of entropy over the family of MRFs implies that the statistic is positively correlated. The converse was shown in [13]. The interest in such questions arises from both engineering [15] and social science [16], [7] concerns.

Let $G = (V, E)$ be a graph. A family of exponential distributions is specified by a vector statistic $t = (t_{ij})$ defined on the endpoints of the edges E of the graph.[1] That is, for a given image $\mathbf{x} = \{x_i : i \in V\}$ and each edge $\{i, j\} \in E$, the function $t_{ij} : \mathcal{X}_i \times \mathcal{X}_j \longrightarrow \mathbb{R}$ determines the contribution of the pair (x_i, x_j) to the probability of $\mathbf{x}$. We say that X is *Markov with respect to* G, in that conditioning on a cutset renders separated subsets independent of one another [17]. The entire family of MRFs based on t is generated by introducing an exponential parameter $\theta = (\theta_{ij})$ where for each edge $\{i, j\}$, θ_{ij} scales the sensitivity of the distribution $p(\mathbf{x}) = p(\mathbf{x}; \theta)$ to the function t_{ij}. Specifically, if X is an MRF based on t with exponential parameter θ, the probability of an image $\mathbf{x}$ is

$$\begin{aligned} p(\mathbf{x}; \theta) &= \exp\{ \sum_{\{i,j\} \in E} \theta_{ij} t_{ij}(x_i, x_j) - \Phi(\theta)\} \\ &= \exp\{\langle \theta, t(\mathbf{x}) \rangle - \Phi(\theta)\}, \end{aligned}$$

where where

$$\Phi(\theta) = \log \left[\sum_{\mathbf{x} \in \mathcal{X}} \exp\{\langle \theta, t(\mathbf{x}) \rangle\} \right]$$

is the log-partition function. The set

$$\Theta = \{\theta \in \mathbb{R}_+^{|E|} \mid \Phi(\theta) < \infty\}$$

is the set of *admissable* exponential parameters[2], while $\mathcal{F} = \{p(\cdot; \theta) \mid \theta \in \Theta\}$ is the family of all MRFs based on t.

We denote by $H(\theta)$ the entropy of the MRF induced (indexed) by exponential parameter θ. Straightforward calculation reveals that

$$H(\theta) = \Phi(\theta) - \sum_{e \in E} \theta_e \mu_e,$$

where the sum is over edges of the graph, and where $\mu_e = \mathbb{E}[t_e]$ is the expected value of statistic t_e under $p \sim \theta$. Taking partial derivatives yields

$$\frac{\partial}{\partial \theta_{ij}} H(X; \theta) = - \sum_{\{k,l\} \in E} \theta_{kl} \sigma_{ij,kl}^2, \quad (1)$$

where $\sigma_{ij,m}^2 = \text{cov}\ (t_{ij}, t_m)$, $\sigma_{ij,kl}^2 = \text{cov}\ (t_{ij}, t_{kl})$, $\sigma_{ij,ij}^2 = \text{cov}\ (t_{ij}, t_{ij})$. The statistic t is termed *positively correlated* if all covariance terms $\sigma_{ij,kl}^2$ are positive. This condition was shown by Griffiths [8] to hold for a standard Ising model and was shown in [13] to imply monotonicity of entropy.

The first and arguably the most studied MRF is the Ising model[3], introduced as a model for the interactions of iron atoms and used to analyze spontaneous magnetization within collections of such atoms [10]. For each node $i \in V$, the random variable X_i assumes a value in the set $\{-1, 1\}$, in the initial context indicating which of two possible spins the atom could assume. For an edge $\{i, j\} \in E$, the statistic t_{ij} can be either a *positive* $t_{ij}(x_i, x_j) = x_i x_j$, or *negative* $t_{ij}(x_i, x_j) = -x_i x_j$. Likewise, for node $i \in V$, the statistic can be positive or negative, as $t_i(x_i) = x_i$ or $t_i(x_i) = -x_i$, respectively. In Section we consider Ising models with only edge statistics. However, in Section V we see an example of an Ising model on a chain with conflicting self statistics. A positive edge statistic favors similar values on the endpoints of

[1] Properly, this is a *pairwise* MRF with no self statistics. Generalizations to other MRFs are straightforward.

[2] The sign restriction on θ is used since as discussed below we allow the statistic to be positive or negative.

[3] Strictly speaking, a multivariate Gaussian distribution is an MRF, the graph in question determined by the nonzero elements of the inverse covariance matrix, so one could argue that this is the most studied MRF.

the edge, while negative edge statistics favor dissimilar values on the endpoints. [4].

For an Ising model on an acyclic graph and with no self statistics, the model is *cooperative* in the sense that there exists a configuration that simultaneously maximizes the value of each component of the statistic. This can be seen by starting at a given node and assigning to it either a -1 or +1. For argument's sake, let's say +1. Then, for each neighbor $j \in \partial i$, if the edge $\{i, j\}$ is positive, assigning +1, and if the edge is negative, assigning -1 to node j. By iteratively applying the same procedure until the leaves are reached, an assignment is made to all nodes that maximize all statistics. In sense, then, the topology cooperates with the statistic for this maximization. The term *frustrated* means that there is no configuration that can be assigned to the nodes that will simultaneously maximize the value of each component of the statistic. Frustration is commonly discussed in connection with cycles, as the return of a path to its starting node introduces a consistency constraint whereby the last step of the path, the one to close the cycle, is already determined.

In this paper we show that for an Ising model on a cycle, the statistic t is positively correlated if and only if t is not frustrated, i.e. cooperative. Likewise, we show that entropy is monotone decreasing over $\mathcal{F}$ if and only if t is non-frustrated. This comes about by showing that for a frustrated statistic, all covariance terms are negative, while for a cooperative statistic, covariances are of the same magnitude, but positive. It remains to generalize these correlation inequalities to the standard setting, i.e. [6].

An initial motivation for considering monotonicity of entropy is the recently introduced method for performing lossy compression of an Ising model [15]. There, a cutset $U \subset V$ of sites was losslessly encoded, and the remaining sites estimated conditioned on the values of the cutset. We showed in [13] that the (marginal) entropy of the subset X_U is upper bounded by the entropy of the reduced MRF on the subgraph $G(U)$ induced by U, with exponential coordinates inherited from the original θ. Furthermore, and as part of the proof of the above, we showed that the entropy is monotone decreasing in the exponential parameters, therefore allowing us to obtain possibly tractable upper bounds to that of the reduced MRF by removing more edges.

Another motivation beyond the use of MRFs in engineering-centric objectives is the use of MRFs to model dynamic systems of interacting nodes/agents/players. A branch of game theory looks at games played on a graph, i.e., games are played iteratively over time between adjacent nodes of the graph. The statistics of the Ising model match the payoff structure of a class of games referred to as *coordination* games, in that they are used to model the coordination of behaviors or preferences over a network of interacting individuals. Such models have been used in the social, political, and economic sciences for some time [16], and owing to the growing prominence of social networking, there has been a surge of interest in developing and understanding quantitative models of network behavior [12], [9], [1]. This is due principally to the abundance of data that is available, not only in terms of network structure, but also the expression of individual preferences in the form of online purchases, as an example. Furthermore, the potential for quantitative models to make meaningful predictions is naturally of interest to those who seek to influence the expression of preferences on the network, for instance those seeking social and political change, as well as retailers [7].

A Markov random field is a natural tool toward such endeavors due to its amenable form and the plethora of algorithms that exist for such models [17], as well from a statistical modeling perspective [11]. More precisely, the positive statistic of the Ising model matches the payoff structure of a *coordination* game, while the negative statistic matches the payoff for an *anti-coordination* game. Therefore, if we consider a network of individuals expressing preferences while playing coordination or anti-coordination games, then given *a priori* knowledge about expected payoffs (statistics) on the edges, the maximum entropy equilibrium probability distribution is the Markov (here, Ising) random field with exponential coordinates chosen to enforce the moment constraints.

The paper is organized as follows. Section II covers background material. Section III computes the log partition function and moments for the acyclic and cyclic Ising models. Section IV covers frustration and cooperation, Section V shows a few simple examples, Section VI includes some concluding remarks, and the Appendix contains the proof of Thm. 4.3.

II. Nuts and Bolts

A random field is a finite collection of random variables $\mathbb{X} = (X_1, \cdots, X_N)$ indexed by V. For each $i \in V$, X_i takes a value x_i in discrete state space $\mathcal{X}_i$. An instantiation $\mathbf{x}$ of $\mathbb{X} = X_V$ is called an *image*, where $\mathcal{X} = \mathcal{X}_V$ is the space of all images. We let E be a subset of all pairs of nodes in V. Then $G = (V, E)$ is a graph with nodes V and undirected edges E. A random field X is said to be *Markov* with respect to G if for every pair of nodes i and j not joined by an edge, the random variables X_i and X_j are conditionally independent given the remaining variables $X_{V\setminus\{i,j\}}$ [17]. By the Hammersley-Clifford theorem [17], the probability distribution for a positive MRF X can be expressed as the product of factors defined on subsets of the random variables, i.e., as an exponential distribution.

The statistic t is said to provide a *minimal representation* of $\mathcal{F}$ if the $\{t_{ij}\}$ are affinely independent. In this case, it is well-known that the mapping $p : \Theta \longrightarrow \mathcal{F}$ defined by $\theta \longmapsto p(\cdot;\theta)$ is one-to-one. It is clear that the map $p(\cdot;\theta)$ is infinitely differentiable. Therefore if Θ is an open subset of $\mathbb{R}^{|E|}$, then $\mathcal{F}$ is a statistical manifold with coordinate system Θ [2], [3]. Thus a particular element θ indexes an MRF $X \sim p(\cdot;\theta)$. We assume throughout that t is minimal.

For each $\theta \in \Theta$, the expected value of the statistic t is the vector $\mathbb{E}_\theta[t] = \mu$, which is referred to as the moments of the MRF under θ. The set of all moments corresponding to MRFs based on t is

$$\mathcal{M} = \{\mu \in \mathbb{R}^{|E|} \mid \exists \theta \in \Theta, \mathbb{E}_\theta[t] = \mu\}.$$

The *mean parameter map* $\Lambda : \Theta \longrightarrow \mathcal{M}$ is defined as $\Lambda(\theta) = \nabla\Phi(\theta)$. It can be shown that if t is a minimal representation, then Λ is a bijection [3]. Hence, there exists a bijection between the set of mean parameters $\mathcal{M}$ and the manifold $\mathcal{F}$. Thus the set of moments $\mathcal{M}$ is also a coordinate system for $\mathcal{F}$. Therefore an MRF based on t can be indexed by an exponential parameter θ as $p(\cdot;\theta)$ or the corresponding moment parameter μ as $p(\cdot;\mu)$. *Information geometry* studies the structure of $\mathcal{F}$ as a statistical manifold parameterized by the dual coordinate systems Θ and $\mathcal{M}$ [2], and is a powerful tool in the study of MRFs.

Other than the exponential coordinates themselves, the most important quantities for an MRF are the log partition function and its derivatives. In this paper we consider acyclic Ising models and Ising models on a single cycle. In both instances, there is a natural univariate ordering of the edges that corresponds in a one-to-one or near one-to-one fashion with a natural univariate ordering of the nodes. In Sections and such univariate indexing is used, while in Section , bivariate labeling of the statistics and exponential parameters is used. The moments and covariances are computed as follows.

$$\frac{\partial}{\partial\theta_i} \ln\Phi(\theta) = \mu_i \tag{2}$$

$$\frac{\partial^2}{\partial\theta_i\theta_j} \ln\Phi(\theta) = \sigma^2_{i,j} \tag{3}$$

III. Partition function and moments

A useful tool in the analysis of Ising models [4] is the *transfer matrix*. For an edge with positive statistic, the transfer matrix is

$$\Psi_j = \begin{bmatrix} e^{\theta_j} & e^{-\theta_j} \\ e^{-\theta_j} & e^{\theta_j} \end{bmatrix}$$

while for the negative statistic it is

$$\Psi_j = \begin{bmatrix} e^{-\theta_j} & e^{\theta_j} \\ e^{\theta_j} & e^{-\theta_j} \end{bmatrix}$$

The eigenvalues of the positive transfer matrix are $\cosh\theta_j$ and $\sinh\theta_j$, with corresponding eigenvectors $[11]^T$ and $[1-1]^T$. For the negative transfer matrix, the eigenvalues are $\cosh\theta_j$ and $\sinh -\theta_j$, again with eigenvectors $[11]^T$ and $[1-1]^T$. Thus the spectral decomposition of the two transfer matrices are

$$\Psi_j = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C_j & 0 \\ 0 & S_j \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and

$$\Psi_j = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C_j & 0 \\ 0 & -S_j \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

where $C_j = \cosh\theta_j$ and $S_j = \sinh\theta_j$. Because the hyperbolic functions $\cosh\theta$ and $\sinh\theta$ appear so often in the forthcoming, it is convenient to introduce shorthand. We let $C_i = \cosh\theta_i$ and $S_i = \sinh\theta_i$. Since the polarity of the edges factors into the partition function and hence the moments, we will also use $C'_i = \cosh s(t_i)\theta_i$ and $S'_i = \sinh s(t_i)\theta_i$. Differentiating these shows that we will also need recourse to $\bar{C}_j = s(t_j)\cosh s(t_j)\theta_j$ and $\bar{S}_j = s(t_j)\sinh s(t_j)\theta_j$.

Using methods from [4] we can compute the partition function for the following cases: acyclic with no self statistics; chain with conflicting self statistics at the end nodes but no other self statistics; cycle with only self statistics. It can be shown that Ising model on an acyclic graph with no self statistics has a partition function that can be computed as

$$\begin{aligned} &Z(\theta) = \\ &[1\ 1] \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \prod_i C_i & 0 \\ 0 & \prod_i S_i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 2^N \prod_i C_i. \end{aligned}$$

For an Ising cycle with no self statistics, the partition function can be computed as

$$\begin{aligned} &Z(\theta) = \\ &\operatorname{tr}\left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \prod_i C_i & 0 \\ 0 & \prod_i S_i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \\ &= 2^N \left(\prod_i C_i + \prod_i S_i \right). \end{aligned}$$

For an Ising chain with conflicting self statistics, the partition function is computed as

$$\begin{aligned} &Z(\theta) \\ &[1\ 1] \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \prod_i C_i & 0 \\ 0 & \prod_i S_i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 2^N \left(\prod_i C_i + \prod_i S_i \right). \end{aligned}$$

Note that the cycle with no self statistics and the chain with self statistics have the same partition function. Since the moments and covariances are simply partial derivatives of the log partition function, we can already see that the chain with self statistics and a cycle with no self statistics are equivalent with respect to monotonicity of entropy. Now that we have the partition function for the different cases under consideration, we can compute moments, in particular covariances, by taking partial derivatives.

Due to (2) and (3), to compute the covariances of the statistic for a given exponential parameter θ, we need to compute first and second order partial derivatives of the partition function, and therefore first and second order partial derivatives of the signed hyperbolic function C' and S'. Because allowing negative as well as positive statistics means that the eigenvalues of an edge's transfer matrix can potentially be $\cosh(-\theta)$ and $\sinh(-\theta)$. Obviously $\cosh(\theta) = \cosh\theta$ so this makes no difference, but with $\sinh$ it does, so we use C',S' to keep it general, then $\bar{C}$, $\bar{S}$ appear when we take derivatives, which we do to compute the moments.

Proposition 3.1: Let X be an Ising model on an acyclic graph. Then,

$$\begin{aligned} \mu_{ij} &= \tanh\theta_{ij} \\ \sigma^2_{ij,kl} &= 0 \\ \sigma^2_{ij} &= \text{sech}\theta_{ij} \end{aligned}$$

Proof: Straightforward calculation. ■

Let G denote a graph with M nodes $\{1, 2, \ldots, M\}$ connected successively in a cycle. As well, the exponential statistics for the edges will be indexed $\theta_1, \theta_2, \ldots, \theta_M$, likewise with the statistics and moments.

Proposition 3.2: For an Ising cycle with no self statistics, the moments and covariances are

$$\mu_i = \frac{\bar{C}_i \prod_{j\neq i} S'_j + \bar{S}_i \prod_{j\neq i} C'_j}{2\left[\prod_{i,i+1} C'_i + \prod_{i,i+1} S'_i\right]},$$

and

$$\begin{aligned} \sigma^2_{jl} &= \frac{\left[\bar{C}_j\bar{C}_l \prod_{i\neq j} S'_j + \bar{S}_j\bar{S}_l \prod_{i\neq j} C'_j\right]}{\left[\prod_i C'_i + \prod_i S'_i\right]} \\ &\quad - \frac{\left[\bar{C}_j \prod_{i\neq j} S'_i + \bar{S}_j \prod_{i\neq j} C'_i\right]\left[\bar{C}_l \prod_{i\neq l} S'_i + \bar{S}_l \prod_{i\neq l} C'_i\right]}{\left[\prod_i C'_i + \prod_i S'_i\right]^2} \end{aligned}$$

and

$$\sigma^2_j = 1 - \frac{\left[\bar{C}_j \prod_{i\neq j} S'_i + \bar{S}_j \prod_{i\neq j} C'_i\right]^2}{\left[\prod_i C'_i + \prod_i S'_i\right]^2}$$

Proof: Straightforward calculation. ■

IV. Frustration and cooperation

It is also straightforward to show that for an acyclic Ising model, the statistic is positively correlated.

Proposition 4.1: Let t be a statistic for a family of acyclic Ising models. If there are no self statistics, then t is cooperative and t is positively correlated.

Proof: To show that t is cooperative, select an arbitrary node, call it i. Assign it a value, say x_i. For each neighbor $j \in \partial i$, set $X_j = x_i$ if edge $\{i, j\}$ is positive, otherwise set $X_j = -x_i$ if edge $\{i, j\}$ is negative. Repeating this procedure for $k \in \partial j\backslash i$, all $j \in \partial i$, and so on until the leaves of the graph are reached, one achieves a configuration on G that maximizes each statistic.

For distinct edges j and l, we have

$$\begin{aligned} \sigma^2_{j,l} &= \frac{\partial}{\partial\theta_l}\mu_j \\ &= \frac{\partial}{\partial\theta_l}\tanh\theta_j \\ &= 0. \end{aligned}$$

For edge j, we have

$$\begin{aligned} \sigma^2_j &= \frac{\partial}{\partial\theta_j}\mu_j \\ &= \frac{\partial}{\partial\theta_j}\tanh\theta_j \\ &= \text{sech}\theta_j \\ &> 0, \end{aligned}$$

from which the conclusion follows. ■

It is straightforward to see that if one now includes self statistics on individual nodes, that dependence between the edge statistics arises, owing to the influence of the "external field". In Section we show covariances and entropy as functions of the exponential coordinate on one of the edges in a chain Ising model. The model has conflicting self statistics on the endpoints of the chain.

For a cycle, we find that frustration is tantamount to positive correlation of the statistic. As such and as well, monotonicity of entropy is likewise synonymous. The proof of the following theorem is given in the Appendix.

Theorem 4.2: Let t be the statistic, with no self statistic components, for an Ising model on a cycle G. The statistic t is positively correlated and entropy is monotone decreasing in θ if and only if t is not frustrated (is cooperative) on G.

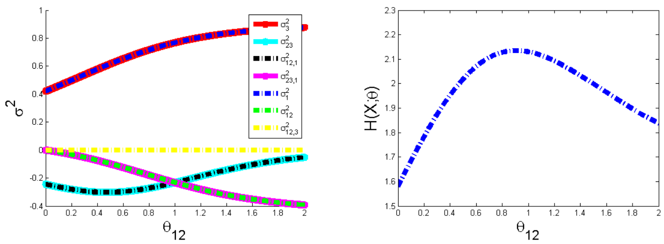
The following is the analogue for the chain with self statistics on the end nodes. The proof is identical to that of the previous theorem and is therefore not repeated.

Theorem 4.3: Let t be the statistic for an Ising model on a chain G with self statistics on the endpoints. The statistic t is positively correlated and entropy is monotone decreasing in θ if and only if t is not frustrated (is cooperative) on G.

V. Examples

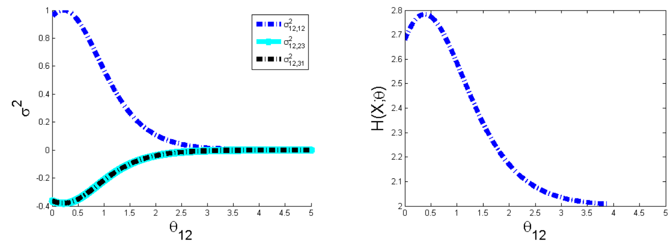
Consider Ising models on the graphs shown in Figure 4. In (a) we have a three node cycle where one of the statistics is negative. This is a frustrated model, and the covariances and entropy are plotted as functions of θ_{12} in Figure 3. We see that the entropy increases as θ_{12} is first increased from 0, then eventually becomes monotone decreasing.

In Figure 4 (b) we have a chain graph where the two endpoints have conflicting self-statistics. In this sense the model is frustrated as there is no configuration that will maximize each edge statistic as well as both self statistic. The entropy and covariance, as functions of θ_{12}, are shown



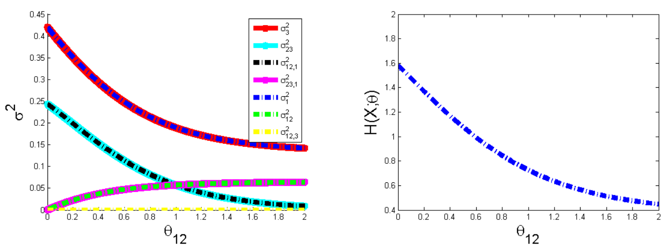
(a) (b)

Fig. 1. Plots for (a) covariances and (b) entropy, for the model depicted in Figure 4 (b), as a function of θ_{12}, the weight on a positive edge statistic.



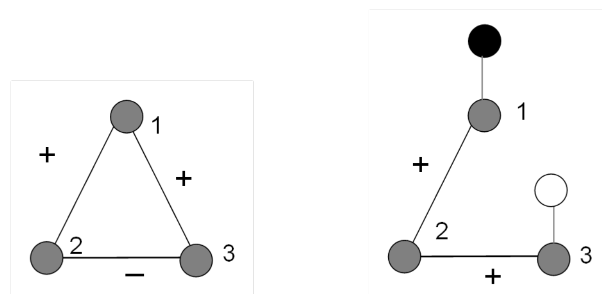
(a) (b)

Fig. 3. Plots for (a) covariances and (b) entropy, for the model depicted in Figure 4 (a), as a function of θ_{12}, the weight on the negative edge statistic.



(a) (b)

Fig. 2. Plots for (a) covariances and (b) entropy, for the model depicted in Figure 4 (b) with edge $\{1, 2\}$ now negative, as a function of θ_{12}.



(a) (b)

Fig. 4. Two frustrated models: (a) a cycle with an odd number of negative edges; and (b) an acyclic graph with conflicting self statistics.

in Figure 1, and are consistent with plots from the frustrated cycle. In 2 we plot the covariances and entropy for the case where the statistic t_{12} is now negative for the graph in Figure 4 (b). Here, the model is not frustrated, since the negative edge allows both self statistics as well as both edge statistics to be maximized.

VI. Discussion

In this paper we have taken a step towards better understanding the relationship between entropy and the statistic defining the family of Markov random fields. We have shown that positive correlation of the statistic and monotonicity of entropy are different ways of looking at cooperation of a statistic. It remains to show that this relationship holds for Ising models on more general topologies, as well as for more general classes of statistics, for instance Potts models. Furthermore, exploring the connection between classes of games used to model different types of interactions on networks and the corresponding MRFs could shed light on either the former, the latter, or both.

Also, we saw that an Ising model on an acyclic graph with conflicting self statistics is not only frustrated in the usual sense, it also exhibited negative covariances and non monotonic entropy as with the frustrated cycle.

Something worth exploring when looking at this problem on more general topologies is the pattern of covariances between the edges. For instance, is the covariance between statistics negative only if the edges in question lie on a frustrated cycle/path?

Acknowledgment

The author would like to thank Dave Neuhoff for pointing out errors in an earlier draft.

Appendix

A. Proof of Thm. 4.3

Lemma A.1: Let t be frustrated on an Ising cycle. Then,

$$\bar{C}_j\bar{C}_l \prod_{i\neq j,l} S'_i \quad = \quad -C_jC_l \prod_{i\neq j,l} S_i$$

Proof: Assuming t is frustrated means there are an odd number of negative edges. Either an odd or even number of $\{j, l\}$ are among these. If neither j nor l are negative, or if both j and l are negative, then an odd number of the factors in the product over $i \neq j$ are negative. Thus $\prod_i S'_i = -\prod_i S_i$, while $\bar{C}_j\bar{C}_l = C_jC_l$.

If one of j or l is among the negative edges, say j, then $\bar{C}_j\bar{C}_l = -C_jC_l$, while $\prod_{i\neq j,l} S'_i = \prod_{i\neq j,l} S_i$. ■

Lemma A.2: Let t be either frustrated or cooperative for an Ising cycle. Then,

$$\bar{S}_j\bar{S}_l \prod_{i\neq j,l} C'_i \quad = \quad S_jS_l \prod_{i\neq j,l} C_i$$

Proof: We see that

$$\begin{aligned}\bar{S}_j &= s(t_j)S'_j \\ &= s(t_j)\sinh s(t_j)\theta_j \\ &= \sinh\theta_j \\ &= S_i,\end{aligned}$$

and furthermore,

$$\begin{aligned}C'_i &= \cosh s(t_i)\theta_i \\ &= \cosh\theta_i \\ &= C_i,\end{aligned}$$

from which the conclusion immediately follows. ■

Lemma A.3: If t is frustrated for an Ising cycle, then

$$\bar{C}_j \prod_{i\neq j} S'_i = -C_j \prod_{i\neq j} S_i$$

Proof: If edge j is negative, then there are an even number of factors in product of S'_j that are negative. Thus the product is negative. If j is not negative, then there are an even number of negative factors in the product over S'_j, so again the overall product is negative. ■

Lemma A.4: Whether t is cooperative or frustrated, we have

$$\bar{S}_j \prod_{i\neq j} C'_i = S_j \prod_{i\neq j} C_i$$

Proof: Since $\bar{S}_j = S_j$ and since $C'_i = C_i$, it follows immediately. ■

We now use the above lemmas to prove the theorem.

Proof: If t is frustrated, there are an odd number of negative edges. Applying the above lemmas to the formula for covariance, we get

$$\begin{aligned}\sigma^2_{jl} &= \frac{-\left(C_jC_l - S_jS_l\right)^2 \prod\limits_{i\neq j,l} C_iS_i}{\left(\prod\limits_i C_i + \prod\limits_i S_i\right)^2} \\ &< 0,\end{aligned}$$

so that t is clearly not positively correlated.

By continuity, one can choose $\theta_1, \theta_2, \ldots, \theta_n$ such that

$$\theta_j\sigma^2_j < \sum_{k\neq j}\theta_k\sigma^2_{jk},$$

which would indicate a positive partial derivative of $H(\theta)$ and therefore non monotonicity of entropy.

Let t be cooperative (i.e., non-frustrated). Then, there are an even number of negative edges, which Lemma implies that

$$\begin{aligned}\sigma^2_{jl} &= \frac{\left(C_jC_l - S_jS_l\right)^2 \prod\limits_{i\neq j,l} C_iS_i}{\left(\prod\limits_i C_i + \prod\limits_i S_i\right)^2} \\ &> 0.\end{aligned}$$

Hence, t is positively correlated, and from [13], $H(\theta)$ is monotone decreasing. ■

References

[1] H. Afrasiabi, R. Guerin, and S. Venkatesh, *Opinion formation in Ising networks*, ITA 2013, San Diego.

[2] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Oxford University Press, 2000.

[3] O.E. Barndorff-Nielson, *Information and Exponential Families*, Chichester, U.K.: Wiley, 1978.

[4] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*, New York: Academic, 1982.

[5] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 2005.

[6] C.M. Fortuin, P.W. Kasteleyn, J. Ginibre, "Correlation inequalities on some partially ordered sets", *Communications in Mathematical Physics*, 22 (89103), 1971.

[7] M. Gladwell, *The Tipping Point*, published by *Little, Brown, and Company*.

[8] R.B. Griffiths, "Correlations in Ising Ferromagnets I.", *Jrnl. Math Physics*, 8 (478), 1967.

[9] N. Immorlica, C. Brandt, G. Kamath, R. Kleinberg, *The Degree of Segregation in Social Networks*, ITA 2013, San Diego.

[10] E. Ising, "Beitrag sur Theorie des Ferromagnetismus", *Zeit. fur Physik*, 31 (253-258), 1925.

[11] E.T. Jaynes, *Information Theory and Statistical Mechanics*, The Physical Review, Vol 106, No 4, 620-630, May 1957.

[12] A. Montanari and A. Saberi, "The spread of innovations in social networks," *PNAS*, Nov. 2010.

[13] M. G. Reyes and D. L. Neuhoff, "Entropy Bounds for a Markov Random Subfield," ISIT 2009, Seoul, South Korea, July 2009.

[14] M. G. Reyes, "Cutset processing and compression of Markov random fields," *phd dissertation*, University of Michigan, 2011.

[15] M.G. Reyes, D. L. Neuhoff, and T.N. Pappas, "Lossy Cutset Coding of Bilevel Images Based on Markov Random Fields," submitted to IEEE Trans. Image Proc.

[16] T. Schelling, *Micromotives and Macrobehavior*, Norton 1978.

[17] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families and variational inference*, Technical Report 649, Sept. 2003.