Inferring couplings in networks across order-disorder phase transitions

Vudtiwat Ngampruetikorn,[1,*,†] Vedant Sachdeva,[2,*] Johanna Torrence,[2,‡] Jan Humplik,[3] David J. Schwab,[1,*] and Stephanie E. Palmer[2,*]

[1]*Initiative for the Theoretical Sciences, The Graduate Center, CUNY, New York, New York 10016, USA*

[2]*Department of Organismal Biology and Anatomy and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*

[3]*Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria*

(Received 14 October 2021; accepted 31 May 2022; published 24 June 2022)

Statistical inference is central to many scientific endeavors, yet how it works remains unresolved. Answering this requires a quantitative understanding of the intrinsic interplay between statistical models, inference methods, and the structure in the data. To this end, we characterize the efficacy of direct coupling analysis (DCA)—a highly successful method for analyzing amino acid sequence data—in inferring pairwise interactions from samples of ferromagnetic Ising models on random graphs. Our approach allows for physically motivated exploration of qualitatively distinct data regimes separated by phase transitions. We show that inference quality depends strongly on the nature of data-generating distributions: optimal accuracy occurs at an intermediate temperature where the detrimental effects from macroscopic order and thermal noise are minimal. Importantly our results indicate that DCA does not always outperform its local-statistics-based predecessors; while DCA excels at low temperatures, it becomes inferior to simple correlation thresholding at virtually all temperatures when data are limited. Our findings offer insights into the regime in which DCA operates so successfully, and more broadly, how inference interacts with the structure in the data.

DOI: 10.1103/PhysRevResearch.4.023240

I. INTRODUCTION

A quantitative understanding of the limitations and biases of inference methods is critical for developing high performing and trustworthy approaches to data analyses. While emerging, such an understanding is incomplete, not least because it requires a thorough investigation of the intertwined nature of statistical models, inference methods, and the structure in the data [1]. Statistical physics models are ideally suited for this investigation for three main reasons. First, they often encompass the statistical models used in practice; take, for example, the Potts model in direct coupling analysis (DCA) [2,3]. Second, they enjoy a number of well-studied inference methods owing to a long history of inverse statistical physics problems [4–6]. Third, they provide a controlled and physically motivated way to alter data-generating distributions across qualitatively distinct regimes. Adopting a statistical physics approach, we characterize the performance of DCA, one of the most oft-used tools in biological sequence analyses, and highlight the importance of the structure in the data in quantifying the performance of inference methods.

DCA has proved successful as a technique for inferring the physical interactions that underpin the structure of biological molecules from amino acid sequence data [2,3]. This success has led to new insights into the protein folding problem [7] and how RNAs obtain their structures [8–10]. The essence of DCA is to draw a distinction between direct and indirect correlations—those originating from direct physical interactions between two sites in a sequence and those mediated via other sites—by fitting a global statistical model to sequence data. But while DCA supersedes its local-statistics-based predecessors in virtually all applications, relatively little is known about the conditions that underlie its success [11].

The statistical model in DCA, well known in physics as the Potts model [12], captures a phase transition that results from a competition between disorder-promoting thermal noise and order-promoting interactions. The disordered phase, which prevails at high temperatures, describes a system whose constituents (e.g., residues in a sequence) are largely uncorrelated; on the other hand, a macroscopic number of such constituents assume the same state in the low-temperature ordered phase. Both phases make for difficult inference: the data are noisy in the disordered phase and macroscopic ordering leads to strong indirect correlations in the ordered phase [13]. A question arises as to the regime in which DCA operates so successfully and more broadly how the nature of data-generating distributions affects inference (see also Ref. [14]).

Recent work suggests that sequence data are drawn from distributions poised at the onset of order [15,16]. This regime sits at the boundary of the two phases, thus minimizing the detrimental effects from thermal noise while avoiding precipitation of macroscopic order. In fact,

*These authors contributed equally to this work.

†Corresponding author: vngampruetikorn@gc.cuny.edu

‡Present address: ShopRunner, Inc., 350 N. La Salle Dr., Chicago, IL.

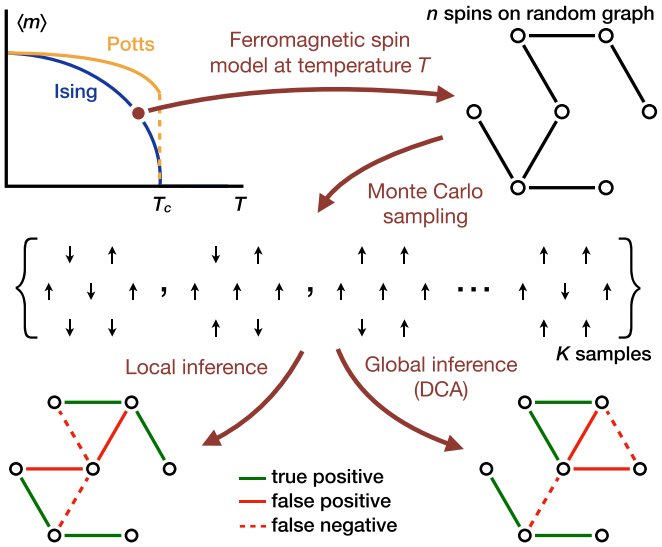

FIG. 1. Data generation and inference. We generate samples from a ferromagnetic spin model on an Erdős-Rényi random graph and evaluate inference methods on the data at different model temperatures across order-disorder phase transitions. Direct coupling analysis ranks the likelihood of an interaction by leveraging global statistics, whereas local inference uses pairwise statistics such as empirical correlations. We obtain predictions by thresholding the likelihood scores. In general, local and global inference methods result in different predictions.

signatures of criticality—a defining property of a type of phase transitions—appear ubiquitous across a wide variety of biological systems [17,18], including antibody diversity [19], genetic regulations [20,21], neural networks [22–28], behaviors of individuals [29], and those of groups [30,31]. This apparent ubiquity has inspired a search for the origin of this behavior [32–35] as well as work that attempts to uncover its function [36]. However, the structure of data distributions alone cannot capture the complete phenomenology of inference and as such cannot explain the success of DCA relative to local-statistics-based methods.

The use of the Potts model to capture correlations among constituents of a system is neither unique to DCA nor limited to analyzing sequence data. Indeed, this approach is applicable to a range of biological systems from neural activity [37,38] to flocks of birds [39]. In addition, the Potts model is closely related to probabilistic graphical models and Markov random fields in probability theory, statistics,and machine learning with applications including inferring interactions among genetic transcription factors [40] and computer vision [41]. Understanding what affects the performance of DCA and when it outperforms local statistical inference is relevant to a large class of problems beyond the application of DCA in structural biology.

Here we investigate the efficacy of DCA in inferring pairwise couplings from samples drawn from ferromagnetic spin models on random graphs at different temperatures across order-disorder phase transitions (see Fig. 1). We demonstrate that the inference quality depends on data-generating distributions; in particular, better inference methods need not be more elaborate nor computationally more expensive. We show that a simple method based on thresholding pairwise correlations can easily outperform DCA at all temperatures in the undersampled regime—a condition applicable to nearly all amino acid sequence datasets. We find further that more data improve DCA most significantly in the ordered phase where strong indirect correlations limit the performance of local methods. Interestingly, we do not observe direct effects of criticality despite its association with diverging Fisher information [42–46]. Instead, we attribute the accuracy maximum at an intermediate temperature to the competition between the emergence of macroscopic order at low temperatures and high thermal noise level at high temperatures. Our work underscores the necessity to characterize the role of data-generating distributions when evaluating inference methods and offers a first step towards a deeper understanding of the intertwined nature of inference, models, and the structure in the data.

II. DATA GENERATION

To highlight the role of a phase transition, we consider the problem of reconstructing the interaction matrix of an Ising model on a random graph. A limiting case of the Potts model, the Ising model is one of the simplest models that captures a phase transition. It describes a system of n spins, $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, each of which is a binary variable $\sigma_i \in \{\pm 1\}$. The spins interact via the Hamiltonian

$$\mathcal{H}(\vec{\sigma}) = -\sum_{i=1}^{n}\sum_{j=i+1}^{n} J_{ij}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i, \tag{1}$$

where J_{ij} denotes the interaction between spins i and j, and h_i the bias field on spin i. The probability distribution of this system is given by

$$P(\vec{\sigma}) = \frac{e^{-\beta\mathcal{H}(\vec{\sigma})}}{\sum_{\vec{\sigma}'} e^{-\beta\mathcal{H}(\vec{\sigma}')}}, \tag{2}$$

where $\beta = 1/T$ is the inverse temperature and the summation is over all spin configurations.

Figure 1 provides an overview of our work. We generate samples from a uniform-interaction ferromagnetic Ising model on an Erdős-Rényi random graph,

$$\mathcal{H}^{\text{data}}(\vec{\sigma}) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j \quad \text{with} \quad J_{ij} \sim \text{Bern}(\lambda/n) \tag{3}$$

for a graph with n vertices and mean degree λ. Each interaction is drawn from a Bernoulli distribution with parameter $p = \lambda/n$, i.e., an interaction is present ($J_{ij} = 1$) with probability p and absent ($J_{ij} = 0$) with probability $1 - p$. In the thermodynamic limit $n \to \infty$, a sharp transition exists between the high-temperature disordered phase and the low-temperature ordered phase. This phase transition is characterized by the order parameter $\Delta \equiv \frac{1}{n}|\langle\sum_i \sigma_i\rangle|$, which vanishes in the disordered phase and grows continuously with decreasing temperature in the ordered phase. A standard mean-field approximation yields the critical temperature $T_c = \lambda$ with the order parameter given by the largest root of the equation $\Delta = \tanh(\lambda\Delta/T)$. As a result, when the mean degree is relatively high, the effect of a change in λ is completely captured by critical temperature rescaling [see also Eq. (C5)]. Our results are

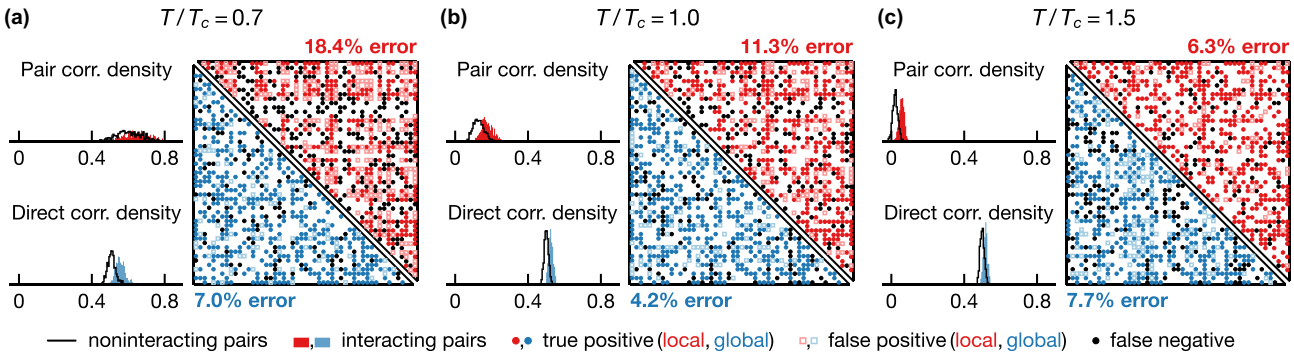

FIG. 2. Local statistical modeling outperforms mean-field DCA in the disordered phase. We show density histograms of empirical and direct pair correlations—$\langle\sigma_i\sigma_j\rangle_{\text{data}}$ and $\langle\sigma_i\sigma_j\rangle_{\text{dir}}$ [see Eq. (5)]—for interacting (filled) and noninteracting (line) pairs of spins at $T/T_c = 0.7, 1.0, 1.5$ [(a)–(c), respectively]. The predictions of pairwise interactions are depicted in a contact map for local (upper half) and global (lower half) inference. The discrimination threshold is chosen such that the number of positive predictions is equal to the number of real interactions, and false positives and false negatives are equal (see legend). In general, both empirical and direct pair correlations are higher among interacting spins and are thus informative of interactions. For local inference, the prediction error decreases with temperature and is smaller than that of global inference at $T/T_c = 1.5$ (c). Global inference error exhibits nonmonotonic temperature dependence and is minimal at an intermediate temperature $T/T_c = 1.0$ (b). Shown results are based on 5×10^3 samples drawn from Ising models on an Erdős-Rényi graph with 50 vertices and mean degree 20.

based on samples generated with exact Monte Carlo sampling [47].

III. MEAN-FIELD INVERSION

While several methods exist for the inverse Ising problem [5], we focus on the so-called naive mean-field inversion which forms the basis for a number of practically relevant algorithms [3,4,7,48]. Derived from a mean-field theory and the linear response theorem [49,50] (see Appendix B), the naive mean-field inversion expresses interactions J_{ij} in terms of empirically accessible connected correlation matrix C,

$$\beta J_{ij} = -(C^{-1})_{ij} \quad \text{for } i < j, \tag{4}$$

where $C_{ij} \equiv \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$. In the following, global statistical inference refers to the naive mean-field inversion.

IV. RESULTS

A. Discriminability of interactions

One measure of inference quality is the ability to discriminate directly interacting spin pairs from those that interact only via other spins. Figure 2 visualizes this discrimination based on local and global statistical inference. For each spin pair, we assign a score that ranks the likelihood of an interaction being present; here, we use empirical correlations $\langle\sigma_i\sigma_j\rangle_{\text{data}}$ and direct correlations $\langle\sigma_i\sigma_j\rangle_{\text{dir}}$ in local and global inference, respectively. The average $\langle\cdots\rangle_{\text{data}}$ is taken with respect to the empirical distribution and $\langle\cdots\rangle_{\text{dir}}$ to the direct pairwise distribution [2],

$$\hat{P}^{\text{dir}}_{ij}(\sigma_i,\sigma_j) \equiv \frac{\exp(\beta\hat{J}_{ij}\sigma_i\sigma_j + \tilde{h}_i\sigma_i + \tilde{h}_j\sigma_j)}{\sum_{\sigma_i',\sigma_j'}\exp(\beta\hat{J}_{ij}\sigma_i'\sigma_j' + \tilde{h}_i\sigma_i' + \tilde{h}_j\sigma_j')}, \tag{5}$$

where $\hat{J}_{ij}$ denotes the inferred interactions from naive mean-field inversion and the fields $\tilde{h}_i$ and $\tilde{h}_j$ are chosen such that the marginal distributions coincide with empirical single-spin distributions. In Fig. 2, we see that on average both empirical and direct correlations are higher among interacting pairs and are thus predictive of true interactions. To turn the likelihood scores into concrete predictions, we need to define a threshold which separates positive and negative predictions. We choose a discrimination threshold that equates the number of positive predictions to the number of true interactions and display inference predictions and errors as a contact map [Figs. 2(a)–2(c)]. The accuracy of the global approach exhibits nonmonotonic temperature dependence with higher error rates at temperatures above and below T_c. In contrast, the accuracy of local inference increases with temperature over the range shown in Fig. 2. (But note that the accuracy must eventually go down at adequately high temperatures; see Fig. 3.) While the error rate of global inference is less than half that of local inference at low temperatures [Figs. 2(a) and 2(b)], a local statistical approach outperforms global inference at high temperature [Fig. 2(c); see also Fig. 3].

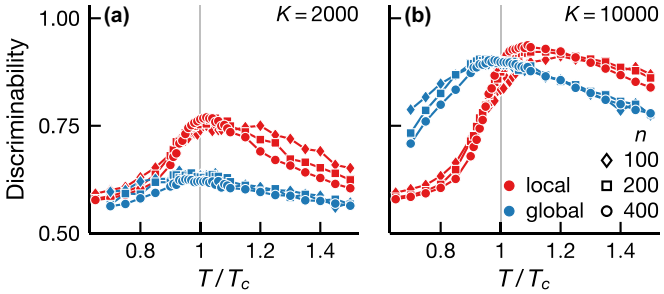

FIG. 3. Local inference is more data efficient but more severely affected by macroscopic order. We depict the local (red) and global (blue) inference discriminability of interactions (area under the ROC curve) for Ising models on Erdős-Rényi graphs with mean degree 40 and different number of vertices n (see legend) for sample sizes $K = 2\times10^3$ and 10^4 [(a) and (b), respectively]. Both local and global inference exhibits discriminability maximum near T_c. Local inference is more discriminating at all temperatures when the data are limited (a). But global inference performs better in the ordered phase when more data are available (b).

Although specifying a discrimination threshold allows us to make concrete predictions, its choice is often arbitrary. We now consider a more general measure of discriminability grounded in receiver operating characteristic (ROC) analysis. ROC analysis constructs a curve that traces the true and false positive rates as the discrimination threshold varies. In the following, we identify discriminability with the area under the ROC curve which is equal to the probability that a real positive scores higher than a real negative.

Local and global statistical inference exhibits qualitatively different sample size dependence (see Fig. 3). At low samples, local inference is more discriminating than naive mean-field inversion at all temperatures [Fig. 3(a)]. This behavior is a result of the distinct natures of local and global approaches. Global inference requires a good estimate of the full joint distribution, whereas local inference relies only on pairwise distributions which are much easier to estimate, especially with limited samples. An increase in samples improves both local and global inference but this improvement diminishes for local inference at low temperatures [Fig. 3(b)]. This results from the fact that the entropy of the model increases with temperature and thus, given a fixed number of samples, a low-temperature model is better sampled. In Fig. 3(a), pairwise distributions are already well sampled at low temperatures and more samples do not lead to higher accuracy for local inference [Fig. 3(b)]. However, well-sampled pairwise distributions do not imply a good estimate of the full distribution; indeed, more samples improve the discriminability of global inference in the low-temperature regimes, i.e., the blue points in Fig. 3(b) are higher than in Fig. 3(a) below T_c.

Inference performance depends not only on well-measured probability distributions but also the structure of the distributions. Despite having lower entropy and being better sampled, low-temperature models are more difficult to infer compared to those in the vicinity of the phase transition (see Fig. 3). This feature is a consequence of macroscopic ordering below T_c. In the ordered phase, two spins are likely to align regardless of the presence of an interaction and therefore pair correlations become less discriminating. While the decrease in discriminability affects both local and global inference, its effect is less severe for global inference (Fig. 3). The use of global statistics—statistical quantities that require measurements of the entire system such as the inverse connected correlation matrix—helps avoid direct comparisons between spin pairs in dense clusters of the interaction graph and those in sparser parts.

B. The effects of local interaction networks on inference

Indeed, local inference is more likely to misclassify well-connected noninteracting spin pairs. To illustrate this point, we randomly divide all of the spin pairs into two disjoint sets for validation and testing. We use the validation set to determine a discrimination threshold and report inference quality on the test set. In Fig. 4 we use 20% of pairs in validation and choose the discrimination threshold such that the resulting true and false positive rates are closest to that of ideal classifiers, as measured by the Euclidean distance in the ROC plane [panel (a)]. Note that while the Euclidean distance is not the only possibility, the concavity of the ROC curve means our results remain qualitatively the same for any metric based on ℓ_p norm with $p \geqslant 1$. Figures 4(b) and 4(c) show that the quality of local inference deteriorates faster as temperature decreases below T_c; i.e., decreasing true positive rate, increasing false positive rate, and more overprediction (excess positive predictions compared to ground truth).

We characterize the false positives (misclassified noninteracting pairs) by the number of shortest paths between spins in each pair [Fig. 4(d)]. Here we focus only on pairs with a graph distance of two (less than 2% of pairs have distance greater than two for this particular graph). At high temperatures the distribution of the number of shortest paths among false positives is the same as that for noninteracting pairs; that is, any noninteracting pair is equally likely to be misclassified. As temperature lowers to around T_c, the false positives from local inference contain a disproportionately large fraction of pairs that are connected by more paths. This behavior is a direct consequence of the emergence of order which generates strong correlations, especially among pairs in denser parts of the graph. At very low temperatures, macroscopic order proliferates and pair correlations are strong regardless of the number of paths or physical interactions. While this effect reduces the disproportionate misclassification among better connected pairs, it increases the discrepancy between the predicted and actual positive rates [Fig. 4(c)]. In fact, the positive rate of $\sim$50% results from the fact that any pair leads to a positive prediction with probability $\frac{1}{2}$. We see that in contrast to local inference, mean-field DCA is less likely to confound path multiplicity with interactions, especially close to the onset of order. In addition it suffers less from strong indirect correlations as evidenced by smaller overprediction rates at low temperatures. In sum, leveraging global statistics helps DCA draw a better distinction between direct and indirect correlations, thus making it more accurate at low temperatures.

C. Root-mean-square error of inferred couplings

While a useful characterization of discriminability, ROC analysis is agnostic about the magnitude of the inferred interactions. We now show that the root-mean-square (rms) error of the interactions inferred by naive mean-field inversion exhibits similar temperature dependence to discriminability. In Fig. 5(a), we see that the rms error is smallest at a temperature slightly below T_c for a range of sample sizes. Figure 5(b) reveals the origin of this temperature dependence. On average mean-field inversion correctly predicts the interactions—$J_{ij} \in \{0, 1\}$ depending on whether an interaction is present—but the prediction variance is minimum around T_c. Above T_c, an increase in temperature leads to a model with higher entropy, thus requiring a larger number of samples to maintain inference accuracy. Below T_c, macroscopic order interferes with inference by generating strong indirect correlations among noninteracting pairs.

D. The role of data-generating models

Since inference quality is intrinsically a combined property of inference methods and data distributions, it is *a priori* unclear whether the observed nonmonotonic

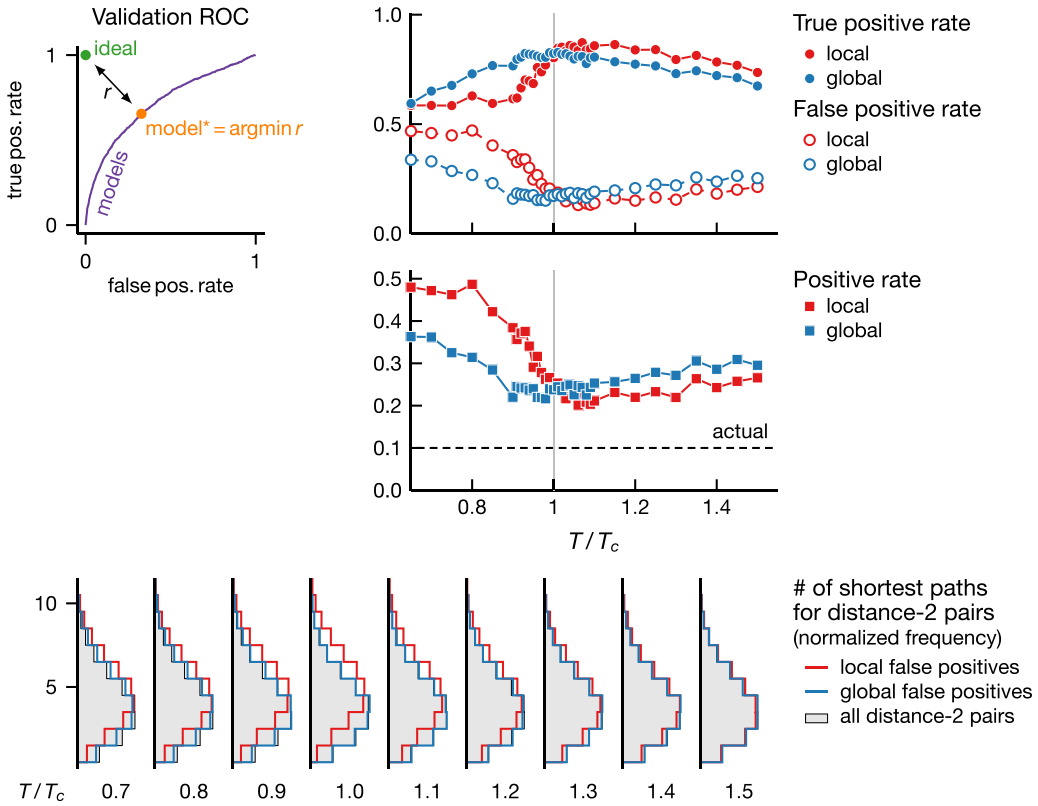

FIG. 4. Local inference is more likely to misclassify well-connected noninteracting pairs. We use 20% of pairs chosen at random (validation set) to compute the discrimination threshold (a) and report inference properties on the rest [test set, (b)–(d)]. (a) Typical ROC curve for the validation set. We choose a threshold such that the resulting model is closest to the ideal model, as measured by the Euclidean distance in the ROC space. (b) True and false positive rates vs temperature. Both local and global methods are most accurate at a temperature close to T_c but local inference worsens faster at low temperatures. (c) Temperature dependence of the positive rate (the ratio between positive predictions and all pairs). Overprediction is most acute for local inference at low temperatures. (d) Distribution of the number of shortest paths among false positive pairs with graph distance two at different temperatures. At low temperatures the false positives from local inference contain a larger fraction of highly connected pairs, compared to all pairs with distance two (gray) as well as to the false positives from global inference. Thus noninteracting pairs in denser parts of the graph are likelier to be misclassified than those in sparser parts. Shown results are based on 10^4 samples from an Ising model on an Erdős-Rényi graph with 400 vertices and mean degree 40.

temperature dependence (Figs. 3 and 5) originates from the inductive bias in inference methods or the structure in the data. To isolate the role of data-generating models, we consider the response of data distributions to a change in model parameters as a proxy for how informative a data point is about model parameters. We quantify the distributional response by the f divergence, an information-theoretic distance between two distributions, defined via $D_f(P_X\|Q_X) \equiv \langle f(P_X/Q_X)\rangle_{X\sim Q_X}$ where $f:[0,\infty)\to(-\infty,\infty)$ is convex and $f(1)=0$. The f divergence between two zero-field Ising models on different graphs, parametrized by J and J', reads [see Eqs. (2) and (3)]

$$D_f(J',J) = \left\langle f\left(\frac{e^{\beta\sum_{i<j}\Delta J_{ij}\sigma_i\sigma_j}}{\left\langle e^{\beta\sum_{i<j}\Delta J_{ij}\sigma'_i\sigma'_j}\right\rangle_{\vec\sigma'\sim\mathcal{H}_J}}\right)\right\rangle_{\vec\sigma\sim\mathcal{H}_J}, \tag{6}$$

where $\Delta J = J' - J$ and the average $\langle\cdots\rangle$ is with respect to the model on the graph J.

Before we discuss the numerical results, it is instructive to derive an expression for the f divergence in a mean-field approximation. Expanding Eq. (6) around $\beta = 0$ and taking $P(\vec\sigma) = \prod_i \frac{1}{2}(1+\sigma_i\Delta)$ yield

$$D_f^{\rm mf}(J',J) = \frac{1}{2}f''(1)\|\Delta J\|_1\frac{1-\Delta(T)^4}{T^2}, \tag{7}$$

where $\Delta(T)$ is the mean-field order parameter and the ℓ_1 norm $\|\Delta J\|_1$ counts the number of different edges in J and J'. Note that the elements of J and J' are either zero or one and we set $J_{ij}=0$ for $i\geqslant j$ as they do not enter the model [see Eq. (3)]. In the disorder phase $T > T_c$, high noise level makes models less dependent on the parameters and the f divergence decays as T^{-2}. The dependence on the order parameter means different parameters also result in more similar models at low temperatures [since $\Delta(T)\to 1$ as $T\to 0$]. Indeed, the competition between thermal noise and macroscopic order leads to a maximum at $T/T_c \approx 0.83$. Figure 6 illustrates the temperature dependence of the f divergence between two Ising models. Here we adopt the Jensen-Shannon (JS) divergence which is an f divergence defined with $f(t) = (t+1)\log_2\frac{2}{t+1} + t\log_2 t$. We compute the divergence $D_{JS}(J',J)$ from data using Eq. (6) for a fixed Erdős-Rényi graph J and we generate J' by randomly deleting and adding edges

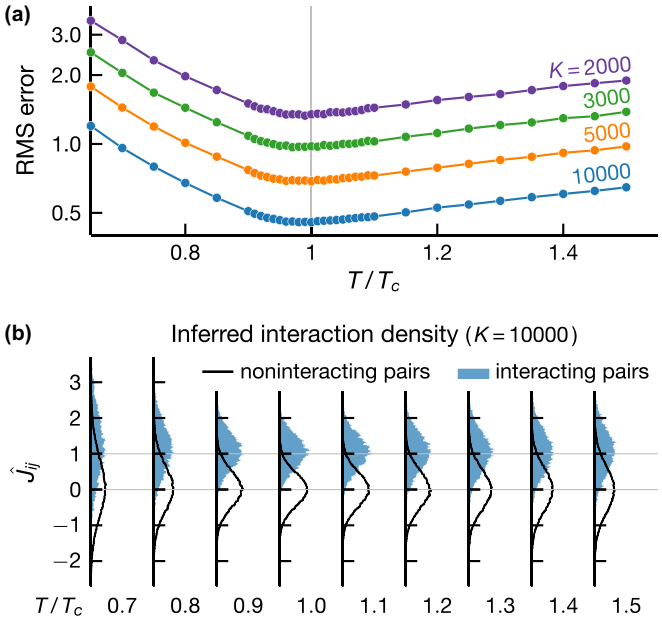

FIG. 5. Interactions inferred from mean-field DCA are statistically unbiased with smallest variances around phase transitions. (a) Root-mean-square error of inferred interactions as a function of temperature at different sample sizes K (see legend). (b) Density histograms of inferred interactions for noninteracting and interacting pairs whose true interactions are one and zero, respectively. Shown results are for an Ising model on an Erdős-Rényi graph with 400 vertices and mean degree 40.

in J, allowing J and J' to have different numbers of edges. We see that, as expected from the mean-field analysis, the f divergence decays as T^{-2} at high temperatures and peaks at a temperature below T_c with its scale controlled by the number of different interactions in J and J' [Fig. 6(a)]. In Fig. 6(b), we compare the empirical JS divergence to the mean-field approximation [Eq. (7)] and find good agreement for $T > T_c$. Below T_c, the mean-field result only captures the qualitative behavior due to large variance in the JS divergence (from

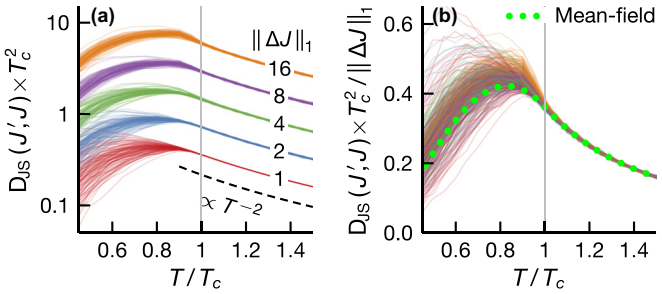

FIG. 6. Jensen-Shannon (JS) divergence between two Ising models vs temperature. (a) JS divergences computed from 10^4 samples using Eq. (6) for a fixed graph J and many realizations of J' generated by randomly deleting and adding edges to J. The curves are grouped by the number of different edges in J and J' (see legend). (b) Empirical JS divergences compared to a mean-field prediction, Eq. (7), showing good agreement for $T > T_c$ [same color code as in (a)]. Here J is an Erdős-Rényi graph with 400 vertices and mean degree 40.

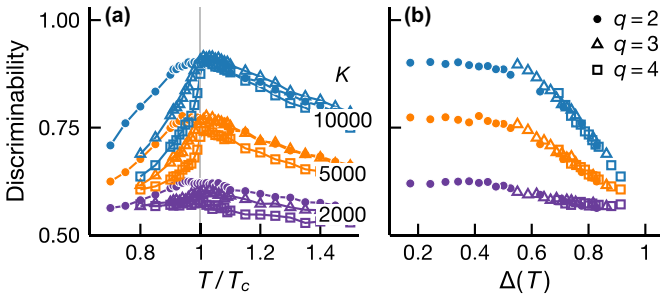

FIG. 7. Interaction discriminability for Ising and Potts models. Discriminability maximum results from the competition between thermal noise and macroscopic ordering but is not a signature of criticality associated with second-order phase transitions. We show DCA discriminability at different sample sizes K (see legend) as a function of temperature (a) and mean-field order parameter Δ (b). In all cases, discriminability peaks at an intermediate temperature and displays similar temperature dependence above T_c. By plotting discriminability as a function of Δ for $T < T_c$, we see that a different temperature dependence for Ising and Potts models at $T < T_c$ originates from the fact that macroscopic order forms more rapidly in Potts models which admit first-order phase transitions. This highlights the detrimental effect of macroscopic order on inference quality. Shown results are based on the same Erdős-Rényi interaction graph with 400 vertices and mean degree 40.

different realizations of J'). This is an expected result since the locations where macroscopic order nucleates depend on graph structure and a change to which can yield a range of divergences.

E. Inference discriminability for Potts models

It is tempting to view the inference quality maximum as a manifestation of critical phenomena, not least because the Fisher information (magnetic susceptibility) diverges at T_c [42–46]. However, criticality does not seem to play an important role in inferring the interaction graph. Indeed, Fig. 6 illustrates that the distance between two models on different graphs varies smoothly across the critical temperature.

To elaborate this point further, we consider q-state Potts models on an Erdős-Rényi random graph which generalizes the binary spins in Ising models to q states. Unlike the Ising model, a q-state Potts model with $q > 2$ exhibits a discontinuous phase transition which does not display critical behaviors and at which the susceptibility remains finite. Figure 7 compares the inference discriminability for three- and four-state Potts models with that for Ising models ($q = 2$). We use the naive mean-field inversion, generalized to Potts models [3] for both Ising and Potts models (see Appendix B). In Fig. 7, we see that, in the disordered phase, the discriminability for Potts and Ising models shows similar dependence on sample size and temperature. In the ordered phase, the inference quality decreases with temperature and worsens with increasing q. This q dependence results from the fact that macroscopic order forms more rapidly for larger q with order parameter discontinuity growing with q (see Appendix C [Eq. (C7)]). In fact, Fig. 7(b) illustrates that the inference discriminability for Potts and Ising cases displays similar dependence on the mean-field order parameter (for a mean-field analysis of the

Potts model, see Ref. [12] and Appendix A), thus suggesting that macroscopic ordering rather than criticality is an important determinant of inference performance.

V. DISCUSSION

Despite being more elaborate and computationally more expensive than local statistical approaches, mean-field DCA does not always lead to better inference quality. Indeed, we show that local statistical methods can be more accurate when data are limited. More generally, although global statistics encode more information that could potentially improve inference, they are more difficult to estimate in the undersampled regime. Inference quality depends not only on sample size but also on the nature of data distributions. A low-temperature generative model, while better sampled due to lower entropy, is more difficult to infer, compared to higher-temperature models around the phase transition. This feature highlights how macroscopic ordering, and more broadly data distributions, can interfere with inference. For models exhibiting an order-disorder phase transition, we find that DCA provides the most advantage over local statistical modeling in the ordered phase and when the systems are relatively well sampled. Our results highlight the fact that inference quality can only be quantified with respect to the structure in the data and illustrate the central role of data-generating distributions in understanding inductive biases of inference methods [51]. Finally, our work lays a foundation for future investigations seeking to provide a prescription for inference method selections based on the structure in the data.

While we consider ferromagnetic models on relatively dense interaction networks, our analysis yields qualitative insights applicable to models with sparser interactions. In particular, we expect better performance from local inference as each spin pair becomes less connected (see Sec. IV B). In addition, the increased probability of isolated spins means that the connected correlation matrix is more likely to be singular, thus making naive mean-field inversion ill-defined without regularization. A quantitative study of inference for models on sparse networks is an interesting research direction, not least because of the important role of fluctuations in such models.

Although we base our analysis on naive mean-field inversion, a number of methods exist for inferring pairwise interactions (see, e.g., Ref. [5]). The general conclusion of our work also applies to these methods; the inference quality must depend on the structure in the data-generating distribution as well as the number of available observations. Revealing the optimal setting for each of these methods is likely to require generative models that capture different types of correlations in the system, and is a promising avenue for future research. For the ferromagnetic model considered here, we expect that our qualitative results hold for other inference methods, not least because the inference performance maximum near the phase transition stems from the property of the generative model (see Sec. IV D).

To isolate the role of a phase transition, we specialize our analysis to uniform-interaction models on Erdős-Rényi random graphs which tend to be less structured than interaction graphs of real systems. For example, the structural organization of proteins leads to a hierarchy of sectors of strongly interacting amino acids [52]. Spin models on hierarchical random graphs also capture order-disorder phase transitions [53] and it would be interesting to investigate how such a structure affects inference. Another promising future direction is to extend our analysis beyond ferromagnetic models to systems with richer phase diagrams such as spin-glass models and sparse Hopfield networks.

ACKNOWLEDGMENTS

This work was supported in part by the Alfred P. Sloan Foundation, the Simons Foundation, the National Institutes of Health under Award No. R01EB026943, and the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030).

APPENDIX A: GRAPHICAL POTTS MODELS

Potts models describe a system of q-state spins $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ with $\sigma_i \in \{1, 2, \ldots, q\}$, interacting via the Hamiltonian,

$$\mathcal{H}(\vec{\sigma}) = -\sum_{i=1}^{n}\sum_{j=i+1}^{n} J_{ij}(\sigma_i, \sigma_j) - \sum_{i=1}^{n} h_i(\sigma_i). \tag{A1}$$

The probability distribution of this system is given by

$$P(\vec{\sigma}) = \frac{e^{-\beta\mathcal{H}(\vec{\sigma})}}{\sum_{\vec{\sigma}'} e^{-\beta\mathcal{H}(\vec{\sigma}')}}. \tag{A2}$$

This measure is invariant under the gauge transformation,

$$h_i(\mu) \to h_i(\mu) + \phi_i + \sum_{j}^{j\neq i} \Lambda_{ij}(\mu),$$

$$J_{ij}(\mu, \nu) \to J_{ij}(\mu, \nu) - \Lambda_{ij}(\mu) - \Lambda_{ji}(\nu) + \psi_{ij} \tag{A3}$$

for any ϕ_i, ψ_{ij}, and $\Lambda_{ij}(\mu)$. This gauge symmetry means that the Potts measure is characterized by $\binom{n}{2}(q-1)^2 + n(q-1)$ independent parameters, which is the same number of independent parameters in single- and two-spin distributions, $P(\sigma_i)$ and $P(\sigma_i, \sigma_j)$ (see, e.g., Ref. [3]). Indeed, for specified $P(\sigma_i)$ and $P(\sigma_i, \sigma_j)$ the Potts measure is the unique maximum-entropy model [3]. Another consequence of the gauge invariance is that a family of model parameters (J, h) can result in the same measure. As a result, inference methods that produce a unique set of parameters must invoke gauge fixing conditions (either explicitly or via implicit regularization).

APPENDIX B: MEAN-FIELD INVERSION

For completeness, we reproduce the derivation of the mean-field inversion method for Potts models from Ref. [3]. We define the free energy

$$\mathcal{F} = \mathcal{F}(J, h) = -\ln \sum_{\vec{\sigma}} e^{-\beta\mathcal{H}(\vec{\sigma})}. \tag{B1}$$

It follows that the first- and second-order derivatives of this free energy are related to the single-spin and pairwise distributions via

$$\frac{\partial\mathcal{F}}{\partial h_{i\mu}} = -P_{i\mu} \quad \text{and} \quad \frac{\partial^2\mathcal{F}}{\partial h_{i\mu}\partial h_{j\nu}} = -P_{i\mu,j\nu} + P_{i\mu}P_{j\nu}, \tag{B2}$$

where we introduce the shorthand notations

$$h_{i\mu} = h_i(\sigma_i = \mu), \qquad J_{i\mu,j\nu} = J_{ij}(\sigma_i = \mu, \sigma_j = \nu),$$
$$P_{i\mu} = \sum_{\vec{\sigma}} \delta_{\sigma_i,\mu} P(\vec{\sigma}), \quad P_{i\mu,j\nu} = \sum_{\vec{\sigma}} \delta_{\sigma_i,\mu}\delta_{\sigma_j,\nu} P(\vec{\sigma}).$$

Equation (B2) also implies

$$\frac{\partial P_{i\mu}}{\partial h_{j\nu}} = P_{i\mu,j\nu} - P_{i\mu}P_{j\nu} \equiv C_{i\mu,j\nu}, \tag{B3}$$

where $C_{i\mu,j\nu}$ denotes the connected correlation matrix.

1. Gauge fixing

To infer a unique set of model parameters, we adopt the lattice-gas gauge which explicitly limits the model parameters to those that are independent [see Eq. (A3) and the text around it]. In this gauge each spin has a gauge state, c_i for spin i, for which the pairwise coupling and local field vanish, i.e.,

$$\forall \vec{\sigma}, i, j : J_{ij}(\sigma_i, c_j) = J_{ij}(c_i, \sigma_j) = h_i(c_i) = 0. \tag{B4}$$

We assume this gauge in the following analysis unless specified otherwise.

2. Legendre transformation

Since the local field $h_{i\mu}$ is conjugate to the single-spin distributions $P_{i\mu}$ [see Eq. (B2)], we can define a Legendre transform of the free energy

$$\mathcal{G} = \mathcal{F} + \sum_{i\mu} h_{i\mu} P_{i\mu}. \tag{B5}$$

Note that $\mathcal{G}$ does not depend explicitly on the probability of the gauge state P_{ic_i}; it is left out of the summation by the gauge condition $h_{ic_i} = 0$ [Eq. (B4)]. In this ensemble the local fields are given by

$$h_{i\mu} = \frac{\partial \mathcal{G}}{\partial P_{i\mu}}. \tag{B6}$$

Taking the derivative of the above equation yields

$$\frac{\partial h_{i\mu}}{\partial P_{j\nu}} = \frac{\partial^2 \mathcal{G}}{\partial P_{i\mu}\partial P_{j\nu}} = (C^{-1})_{i\mu,j\nu}, \tag{B7}$$

where the last equality follows from Eq. (B3) and the fact that the first-order derivatives of a function and its Legendre transform are inverse functions of one another. Note that the indices $(i\mu, j\nu)$ in Eqs. (B6) and (B7) do not include the gauge states.

3. Small-coupling expansion

To derive the mean-field inversion, we consider a systematic expansion around the noninteracting Hamiltonian, treating the coupling term as a perturbation [54,55],

$$-\beta \mathcal{H}_\alpha(\vec{\sigma}) = \alpha \sum_{i<j} J_{ij}(\sigma_i, \sigma_j) + \sum_i h_i(\sigma_i), \tag{B8}$$

where the parameter α tunes the interaction strength: $\mathcal{H}_0$ corresponds to the noninteracting case and $\mathcal{H}_1$ to the original Hamiltonian. Expanding $\mathcal{G}$ as a power series in α yields

$$\mathcal{G}_\alpha = \mathcal{G}_0 + \mathcal{G}_0'\alpha + \tfrac{1}{2}\mathcal{G}_0''\alpha^2 + O(\alpha^3), \tag{B9}$$

where $\mathcal{G}_\alpha' = d\mathcal{G}_\alpha/d\alpha$ and $\mathcal{G}_\alpha'' = d^2\mathcal{G}_\alpha/d\alpha^2$. Substituting the above expression in Eqs. (B6) and (B7) gives

$$h_{i\mu} = \frac{\partial \mathcal{G}_0}{\partial P_{i\mu}} + \frac{\partial \mathcal{G}_0'}{\partial P_{i\mu}}\alpha + O(\alpha^2),$$
$$(C^{-1})_{i\mu,j\nu} = \frac{\partial \mathcal{G}_0}{\partial P_{i\mu}\partial P_{j\nu}} + \frac{\partial \mathcal{G}_0'}{\partial P_{i\mu}\partial P_{j\nu}}\alpha + O(\alpha^2) \tag{B10}$$

for $i\mu \neq ic_i$ and $j\nu \neq jc_j$.

4. Zeroth order

When $\alpha = 0$, the spins decouple and the free energy reads

$$\mathcal{F}_0 = -\sum_i \ln \sum_\nu e^{h_{i\nu}}. \tag{B11}$$

From Eq. (B2), we have $P_{i\mu} = e^{h_{i\mu}}/\sum_\nu e^{h_{i\nu}}$ and

$$\mathcal{G}_0 = \sum_{i\mu\neq ic_i} P_{i\mu}\ln P_{i\mu} + \sum_i \left(1 - \sum_{\nu\neq c_i} P_{i\nu}\right)\ln\left(1 - \sum_{\nu\neq c_i} P_{i\nu}\right). \tag{B12}$$

Taking the derivatives, we have

$$\frac{\partial \mathcal{G}_0}{\partial P_{i\mu}} = \ln\frac{P_{i\mu}}{P_{ic_i}} \quad \text{and} \quad \frac{\partial^2 \mathcal{G}_0}{\partial P_{i\mu}\partial P_{j\nu}} = \delta_{ij}\left(\frac{\delta_{\mu\nu}}{P_{i\mu}} + \frac{1}{P_{ic_i}}\right), \tag{B13}$$

where $P_{ic_i} = 1 - \sum_{\mu\neq c_i} P_{i\mu}$. We note that the pairwise coupling does not appear in the zeroth-order expansion.

5. First order

Differentiating the thermodynamic potential $\mathcal{G}_\alpha$ with respect to α gives

$$\mathcal{G}_\alpha' = -\sum_{\vec{\sigma}} \frac{e^{-\beta\mathcal{H}_\alpha(\vec{\sigma})}}{\sum_{\vec{\sigma}'} e^{-\beta\mathcal{H}_\alpha(\vec{\sigma}')}} \sum_{i<j} J_{ij}(\sigma_i, \sigma_j). \tag{B14}$$

Note that the expression for $\mathcal{G}_\alpha$ can be obtained from Eqs. (B1) and (B5) for the small-coupling Hamiltonian in Eq. (B8). In the limit $\alpha \to 0$, the Boltzmann weight becomes that of the noninteracting system and the above equation reduces to

$$\mathcal{G}_0' = -\sum_{i<j}\sum_{\mu\nu} P_{i\mu}P_{j\nu}J_{i\mu,j\nu}. \tag{B15}$$

Therefore we have

$$\frac{\partial \mathcal{G}_0'}{\partial P_{i\mu}} = -\sum_{j\nu}^{j\neq i} P_{j\nu}J_{i\mu,j\nu}. \tag{B16}$$

Here the gauge condition on J ensures that the single-spin probability of the gauge state does not appear on the right-hand side. Note that $J_{i\mu,j\nu}$ for $j < i$ does not enter the model and we let $J_{i\mu,j\nu} = J_{j\nu,i\mu}$ for convenience. Taking the derivative of Eq. (B16), we obtain

$$\frac{\partial^2 \mathcal{G}_0'}{\partial P_{i\mu}\partial P_{j\nu}} = -(1-\delta_{ij})J_{i\mu,j\nu}. \tag{B17}$$

Substituting Eq. (B13) and the above equation in Eq. (B10) gives

$$(C^{-1})_{i\mu,j\nu} \approx \begin{cases} \frac{\delta_{\mu\nu}}{P_{i\mu}} + \frac{1}{P_{ic_i}} & \text{if } i = j \\ -\alpha J_{i\mu,j\nu} & \text{if } j \neq i. \end{cases} \tag{B18}$$

Finally, we combine Eqs. (B10), (B13), and (B16) to obtain the self-consistent condition for the local fields

$$h_{i\mu} = \ln \frac{P_{i\mu}}{P_{ic_i}} - \alpha \sum_{j\nu}^{j\neq i} P_{j\nu} J_{i\mu,j\nu} + O(\alpha^2). \tag{B19}$$

The naive mean-field inversion method is based on Eqs. (B18) and (B19) which relate the model parameters to the empirically accessible connected correlation matrix.

APPENDIX C: PHASE TRANSITIONS IN POTTS MODELS ON HOMOGENEOUS RANDOM GRAPHS

Here we reproduce the mean-field analysis of Potts models (see, e.g., Ref. [12, Sec. I.C]). Consider a uniform-interaction ferromagnetic q-state Potts model on a graph,

$$\mathcal{H}(\vec{\sigma}) = -\sum_{(ij)\in\mathcal{E}} \delta_{\sigma_i,\sigma_j}, \tag{C1}$$

where $\delta_{\sigma_i,\sigma_j}$ denotes the Kronecker delta and the summation is over the graph's edges $\mathcal{E}$. In the mean-field approximation, all spins are identical and the internal energy and entropy of the system read

$$U = -|\mathcal{E}| \sum_{\mu=1}^{q} p_\mu^2 \quad \text{and} \quad S = -n \sum_{\mu=1}^{q} p_\mu \ln p_\mu, \tag{C2}$$

where p_μ is the fraction of spins in state μ, n the number of spins, and $|\mathcal{E}|$ the numbers of edges (interactions). To analyze the ferromagnetic transition, we consider the ansatz

$$p_\mu = \frac{1}{q}(1-\Delta) + \delta_{\mu,q}\Delta, \tag{C3}$$

where Δ is the order parameter and we chose the state q as the spin state of the ferromagnetic phase. This ansatz yields the free energy per spin

$$\beta[f(\Delta) - f(0)] = \frac{1+(q-1)\Delta}{q} \ln[1+(q-1)\Delta] + \frac{q-1}{q}(1-\Delta)\ln(1-\Delta) - \frac{q-1}{2q}\frac{\lambda}{T}\Delta^2, \tag{C4}$$

where $\lambda = 2|\mathcal{E}|/n$ is the mean coordination number. In the thermodynamic limit $n \to \infty$, a phase transition exists at the critical temperature

$$\frac{1}{T_c} = \frac{1}{\lambda} \times \begin{cases} q & \text{if } q \leqslant 2 \\ 2\frac{q-1}{q-2}\ln(q-1) & \text{if } q > 2. \end{cases} \tag{C5}$$

The free energy is minimized by $\Delta = 0$ for $T > T_c$ and by the largest root of the equation

$$e^{-\lambda\Delta/T} = \frac{1-\Delta}{1+(q-1)\Delta} \tag{C6}$$

for $T < T_c$. This phase transition is continuous for $q \leqslant 2$ and discontinuous for $q > 2$ in which the order parameter and internal energy per spin are discontinuous across the transition,

$$\Delta(T_c^-) - \Delta(T_c^+) = \frac{q-2}{q-1},$$
$$u(T_c^-) - u(T_c^+) = -\lambda\frac{(q-2)^2}{2q(q-1)}. \tag{C7}$$

Finally, we note that the above analysis is exact for complete graphs in which all spins in the system are truly (as opposed to statistically) identical.

[1] L. Zdeborová, Understanding deep learning is also a job for physicists, Nat. Phys. **16**, 602 (2020).

[2] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing, Proc. Natl. Acad. Sci. USA **106**, 67 (2009).

[3] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[4] Y. Roudi, E. Aurell, and J. Hertz, Statistical physics of pairwise probability models, Front. Comput. Neurosci. **3**, 22 (2009).

[5] H. C. Nguyen, R. Zecchina, and J. Berg, Inverse statistical problems: From the inverse Ising problem to data science, Adv. Phys. **66**, 197 (2017).

[6] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, Inverse statistical physics of protein sequences: A key issues review, Rep. Prog. Phys. **81**, 032601 (2018).

[7] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, Protein 3D structure computed from evolutionary sequence variation, PLoS One **6**, e28766 (2011).

[8] E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, and M. Weigt, Direct-coupling analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction, Nucleic Acids Res. **43**, 10444 (2015).

[9] C. Weinreb, A. J. Riesselman, J. B. Ingraham, T. Gross, C. Sander, and D. S. Marks, 3D RNA and functional interactions from evolutionary couplings, Cell **165**, 963 (2016).

[10] J. Wang, K. Mao, Y. Zhao, C. Zeng, J. Xiang, Y. Zhang, and Y. Xiao, Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide–nucleotide interactions from direct coupling analysis, Nucleic Acids Res. **45**, 6299 (2017).

[11] Y. Kleeorin, W. P. Russ, O. Rivoire, and R. Ranganathan, Undersampling and the inference of coevolution in proteins, bioRxiv.

[12] F. Y. Wu, "The Potts model," Rev. Mod. Phys. **54**, 235 (1982).

[13] A. Montanari and J. Pereira, Which graphical models are difficult to learn? in *Advances in Neural Information Processing Systems*, edited by Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Curran Associates, Inc., Red Hook, NY, 2009), Vol. 22, pp. 1303–1311.

[14] W. Bialek, S. E. Palmer, and D. J. Schwab, What makes it possible to learn probability distributions in the natural world?, arXiv:2008.12279.

[15] Q.-Y. Tang, Y.-Y. Zhang, J. Wang, W. Wang, and D. R. Chialvo, Critical Fluctuations in the Native State of Proteins, Phys. Rev. Lett. **118**, 088102 (2017).

[16] Q.-Y. Tang and K. Kaneko, Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis, PLoS Comput. Biol. **16**, e1007670 (2020).

[17] T. Mora and W. Bialek, Are biological systems poised at criticality? J. Stat. Phys. **144**, 268 (2011).

[18] W. Bialek, Perspectives on theory at the interface of physics and biology, Rep. Prog. Phys. **81**, 012601 (2018).

[19] T. Mora, Jr., A. M. Walczak, W. Bialek, and C. G. Callan, Maximum entropy models for antibody diversity, Proc. Natl. Acad. Sci. USA **107**, 5405 (2010).

[20] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, Gene expression dynamics in the macrophage exhibit criticality, Proc. Natl. Acad. Sci. USA **105**, 1897 (2008).

[21] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, Morphogenesis at criticality, Proc. Natl. Acad. Sci. USA **111**, 3683 (2014).

[22] A. Levina, J. M. Herrmann, and T. Geisel, Dynamical synapses causing self-organized criticality in neural networks, Nat. Phys. **3**, 857 (2007).

[23] D. R. Chialvo, Emergent complex neural dynamics, Nat. Phys. **6**, 744 (2010).

[24] G. Tkačik, O. Marre, T. Mora, D. Amodei, M. J. Berry II, and W. Bialek, The simplest maximum entropy model for collective behavior in a neural network, J. Stat. Mech.: Theory Exp. (2013) P03011.

[25] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry II, and W. Bialek, Thermodynamics and signatures of criticality in a network of neurons, Proc. Natl. Acad. Sci. USA **112**, 11508 (2015).

[26] T. Mora, S. Deny, and O. Marre, Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons, Phys. Rev. Lett. **114**, 078105 (2015).

[27] X. Chen, F. Randi, A. M. Leifer, and W. Bialek, Searching for collective behavior in a small brain, Phys. Rev. E **99**, 052418 (2019).

[28] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Coarse Graining, Fixed Points, and Scaling in a Large Population of Neurons, Phys. Rev. Lett. **123**, 178103 (2019).

[29] A. C. Costa, T. Ahamed, and G. J. Stephens, Adaptive, locally linear models of complex dynamics, Proc. Natl. Acad. Sci. USA **116**, 1501 (2019).

[30] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, Social interactions dominate speed control in poising natural flocks near criticality, Proc. Natl. Acad. Sci. USA **111**, 7212 (2014).

[31] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale, Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms, Phys. Rev. Lett. **113**, 238102 (2014).

[32] D. J. Schwab, I. Nemenman, and P. Mehta, Zipf's Law and Criticality in Multivariate Data without Fine-Tuning, Phys. Rev. Lett. **113**, 068102 (2014).

[33] L. Aitchison, N. Corradi, and P. E. Latham, Zipf's law arises naturally when there are underlying, unobserved variables, PLoS Comput. Biol. **12**, e1005110 (2016).

[34] M. C. Morrell, A. J. Sederberg, and I. Nemenman, Latent Dynamical Variables Produce Signatures of Spatiotemporal Criticality in Large Biological Systems, Phys. Rev. Lett. **126**, 118302 (2021).

[35] P. Barrat-Charlaix, A. P. Muntoni, K. Shimagaki, M. Weigt, and F. Zamponi, Sparse generative modeling via parameter reduction of Boltzmann machines: Application to protein-sequence families, Phys. Rev. E **104**, 024407 (2021).

[36] M. Meijers, S. Ito, and P. R. ten Wolde, Behavior of information flow near criticality, Phys. Rev. E **103**, L010102 (2021).

[37] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature (London) **440**, 1007 (2006).

[38] G. Tkačik, E. Schneidman, M. J. Berry II, and W. Bialek, Ising models for networks of real neurons, arXiv:q-bio/0611072.

[39] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Statistical mechanics for natural flocks of birds, Proc. Natl. Acad. Sci. USA **109**, 4786 (2012).

[40] N. Friedman, Inferring cellular networks using probabilistic graphical models, Science **303**, 799 (2004).

[41] C. Wang, N. Komodakis, and N. Paragios, Markov random field modeling, inference & learning in computer vision & image understanding: A survey, Comput. Vis. Image. Und. **117**, 1610 (2013).

[42] D. Brody and N. Rivier, Geometrical aspects of statistical mechanics, Phys. Rev. E **51**, 1006 (1995).

[43] W. Janke, D. Johnston, and R. Kenna, in Proceedings of the XVIII Max Born Symposium "Statistical Physics outside Physics" [Information geometry and phase transitions, Physica A **336**, 181 (2004)].

[44] G. E. Crooks, Measuring Thermodynamic Length, Phys. Rev. Lett. **99**, 100602 (2007).

[45] I. Mastromatteo and M. Marsili, On the criticality of inferred models, J. Stat. Mech.: Theory Exp. (2011) P10012.

[46] M. Prokopenko, J. T. Lizier, O. Obst, and X. R. Wang, Relating Fisher information to order parameters, Phys. Rev. E **84**, 041116 (2011).

[47] J. G. Propp and D. B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Struct. Alg. **9**, 223 (1996).

[48] R. R. Stein, D. S. Marks, and C. Sander, Inferring pairwise interactions from biological data using maximum-entropy probability models, PLoS Comput. Biol. **11**, e1004182 (2015).

[49] H. J. Kappen and F. B. Rodríguez, Efficient learning in Boltzmann machines using linear response theory, Neural Comput. **10**, 1137 (1998).

[50] T. Tanaka, Mean-field theory of Boltzmann machine learning, Phys. Rev. E **58**, 2302 (1998).

[51] A. Das and I. R. Fiete, Systematic errors in connectivity inferred from activity in strongly recurrent networks, Nat. Neurosci. **23**, 1286 (2020).

[52] N. Halabi, O. Rivoire, S. Leibler, and R. Ranganathan, Protein sectors: Evolutionary units of three-dimensional structure, Cell **138**, 774 (2009).

[53] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Ising model on networks with an arbitrary distribution of connections, Phys. Rev. E **66**, 016104 (2002).

[54] A. Georges and J. S. Yedidia, How to expand around mean-field theory using high-temperature expansions, J. Phys. A: Math. Gen. **24**, 2173 (1991).

[55] J. S. Yedidia, An idiosyncratic journey beyond mean field theory, in *Advanced Mean Field Methods: Theory and Practice*, Neural Information Processing, edited by M. Opper and D. Saad (MIT Press, Cambridge MA, 2001), Chap. 3, pp. 21–36.