Entropic effects in a nonequilibrium system: Flocks of birds

Michele Castellana, William Bialek, Andrea Cavagna, and Irene Giardina

Phys. Rev. E **93**, 052416 — Published 23 May 2016

DOI: 10.1103/PhysRevE.93.052416

Entropic effects in a non-equilibrium system: Flocks of birds

Michele Castellana,[1,2,3] William Bialek,[1,4] Andrea Cavagna,[4,5] and Irene Giardina[4,5]

[1] *Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, United States*

[2] *Laboratoire Physico Chimie Curie, Institut Curie, PSL Research University, CNRS UMR 168, 75005 Paris, France*

[3] *Sorbonne Universités, UPMC Paris 06, 75005 Paris, France*

[4] *Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, 365 Fifth Ave., New York, New York 10016, United States*

[5] *Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, Rome, Italy and Dipartimento di Fisica, Università Sapienza, Rome, Italy*

When European starlings come together to form a flock, the distribution of their individual velocities narrows around the mean velocity of the flock. We argue that, in a broad class of models for the joint distribution of positions and velocities, this narrowing generates an entropic effect that opposes the cohesion of the flock. The strength of this effect depends strongly on the nature of the interactions among birds: If birds are coupled to a fixed number of neighbors, the entropic forces are weak, while if they couple to all other birds within a fixed distance, the entropic effects are sufficient to tear a flock apart.

I. INTRODUCTION

Entropic forces are a familiar concept in equilibrium statistical mechanics. From the ideal gas to the elasticity of random polymers and the effective forces between molecules in solution, we know that changing the entropy of a system generates an emergent, phenomenological effect analogous to the forces that result from changes in energy [1–3]. Does this intuition carry over into complex, non-equilibrium systems?

Consider a flock of European starlings, *Sturnus vulgaris*. As the birds come close to one another, they interact in ways that may cause their velocities to align [4]. If we imagine constructing the joint distribution of velocities for all the birds in the flock, alignment means that the entropy of this distribution goes down. Is there an entropic effect that might push the birds apart, allowing the entropy to increase? If this were an equilibrium system, the answer would be yes. But this is not an equilibrium system, by any means. Are there still entropic effects?

In recent years, there has been renewed interest in the use of maximum-entropy methods to describe the collective behavior of biological networks, with applications spanning scales from the network of amino acids in a family of proteins [5–9], to biochemical and genetic networks [10, 11], networks of neurons [12–20], and flocks of birds [21, 22]. The idea of the maximum-entropy method is to construct the least structured model of a system that is consistent with certain measured average properties [23]. In particular, if the only quantity that we measure is the energy, then constructing the maximum-entropy distribution is exactly the construction of the thermal equilibrium, leading to the Boltzmann distribution [24].

By using the maximum-entropy framework, we will show that even a non-equilibrium system can be subject to entropic effects. In the context of flocking [25, 26], this means that birds are subject to a repulsive effect due to the loss of entropy associated with their mutual orientation. But, in detail, we will see that this effect depends dramatically on the nature of the interactions and ordering in the system. If we imagine that the flight direction of individual birds maintains a certain level of interaction with the average direction of its n_c nearest neighbors ("topological interactions" [27]), then these entropic effects are weak and, in a sense that we will make precise, flocks can cohere even without explicit forces holding them together. On the other hand, if flocks are characterized by interaction between a bird and its neighbors within some characteristic distance r_c ("metric interactions"), then the entropic effects are strong enough that almost all realistic flocks will be broken into multiple disconnected pieces unless there are other explicitly cohesive forces. Although such forces surely exist, recent observations on real flocks of starlings suggest that the positional correlations among birds are weak [28], so if there is a strong repulsion from the entropy of flight directions, these would have to be finely balanced by other attractive interactions. Such fine tuning is unnecessary in the case of topological interactions.

Although we can make analytic progress on evaluating the entropic impact that results from directional ordering in the flock, computing its effect on the distribution of the birds' positions must be done numerically. This becomes challenging for large flocks, and so we have formulated a simpler problem where we do not keep track of the full configuration of birds' positions in space, but only of the graph that results from their network of mutual interactions. We carry out Monte Carlo (MC) simulations in

the space of these graphs, and we can see that the two problems have similar structures: In particular, the differences between metric and topological interactions arise in both cases, and the graph model allows us to follow these differences out to larger systems.

II. ENTROPIC EFFECTS IN MAXIMUM-ENTROPY MODELS

The essential intuition that we use in building maximum-entropy models for flocks of birds is that the dominant interactions are local, thus if we want to characterize the nature of order in the flock we should measure the degree of correlation between the flight velocities $\vec{v}_i$ of birds and their near neighbors [21, 22]. To be concrete, we will neglect variations in speed and consider only normalized velocities, so each bird i is described by a unit vector $\vec{s}_i \equiv \vec{v}_i/|\vec{v}_i|$, and it is located at position $\vec{x}_i$. We define a neighborhood $\mathcal{N}_i$ around bird i, and within $\mathcal{N}_i$ there are n_i neighbors—details of how this neighborhood is defined are discussed below. To measure the correlation of each bird's flight direction with the average of its neighbors, we compute

$$C_{\rm int} = \frac{1}{N}\sum_{i=1}^{N}\left\langle \vec{s}_i\cdot\left(\frac{1}{n_i}\sum_{j\in\mathcal{N}_i}\vec{s}_j\right)\right\rangle, \tag{1}$$

where the sum over index i runs over birds with $n_i \neq 0$, and $\langle\ \rangle$ denotes the average with respect to the joint distribution of directions $s \equiv \{\vec{s}_i\}$. It will be crucial in what follows that, although $C_{\rm int}$ depends explicitly on flight directions, it depends also on the birds' positions $x \equiv \{\vec{x}_i\}$. We can make this explicit by defining an adjacency matrix $n_{ij}(x)$ such that $n_{ij} = 1$ if $j \in \mathcal{N}_i$, and zero otherwise. Then we have

$$C_{\rm int} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\langle \frac{n_{ij}(x)}{n_i(x)}\vec{s}_i\cdot\vec{s}_j\right\rangle, \tag{2}$$

$$n_i(x) = \sum_{j=1}^{N} n_{ij}(x). \tag{3}$$

If local correlations $C_{\rm int}$ characterize the nature of ordering in the flock, then a good approximation of the full, joint distribution of flight directions can be obtained by building the maximum-entropy distribution consistent with the value of $C_{\rm int}$ observed in real flocks. Indeed, recent studies have shown that the maximum-entropy distribution that matches $C_{\rm int}$ provides accurate, parameter-free predictions for the behavior of two- and four-point correlations as functions of distance [21].

We now focus on the maximum-entropy construction for both flight directions and positions, for which there are two different points of view that we can take. In the first view, the relative positions of the birds are fixed, and we construct the distribution of flight directions of a single, connected flock given these positions: This method has been used in a recent work by some of the authors [21], and it is particularly useful for studying how local correlations between flight directions result in a propagation of order out to length scales comparable to the size of the flock. The resulting maximum-entropy distribution is

$$P(s|x) = \frac{1}{Z(x)}\exp\left(J\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{n_{ij}(x)}{n_i(x)}\vec{s}_i\cdot\vec{s}_j\right), \tag{4}$$

where, as usual, the partition function is given by

$$Z(x) = \int ds\exp\left(J\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{n_{ij}(x)}{n_i(x)}\vec{s}_i\cdot\vec{s}_j\right), \tag{5}$$

and

$$\int ds \equiv \prod_{i=1}^{N}\int d\vec{s}_i\,\delta\left(|\vec{s}_i|^2-1\right). \tag{6}$$

The parameter J is determined by the condition that $C_{\rm int}$ computed from this distribution matches what we observe for the real flock, $C_{\rm int}^{\rm obs}$ [21], and this is equivalent to solving the equation

$$\frac{\partial \ln Z(x)}{\partial J} = NC_{\rm int}^{\rm obs}. \tag{7}$$

This description of flight directions given the relative positions is useful in part because the neighbor relations among birds in the flock change slowly compared to flight directions [29], and because fluctuations in $C_{\rm int}$ from moment to moment in a single flocking event are small.

In the second approach, which is the one used in this study, we imagine that we observe a flock for a time long enough that the birds' relative positions and nearest-neighbor relations rearrange substantially, and the flock possibly disconnects into multiple subflocks. It follows that when we compute the average involved in defining $C_{\rm int}$, Eq. (2), we are averaging not just over flight directions, but also over positions. Now we can ask for the joint maximum-entropy distribution of positions and flight directions that is consistent with the $C_{\rm int}^{\rm obs}$, and the answer is

$$P(x,s) = \frac{1}{Z_0}\exp\left(J\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{n_{ij}(x)}{n_i(x)}\vec{s}_i\cdot\vec{s}_j\right). \tag{8}$$

Of course we might know more about the flock than just $C_{\rm int}^{\rm obs}$. For example, we might have some information about the distribution of pairwise distances between birds, in which case the maximum-entropy distribution becomes

$$P(x,s) = \frac{1}{Z_1} \exp\left(J \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{n_{ij}(x)}{n_i(x)} \vec{s}_i \cdot \vec{s}_j - \sum_{i=1}^{N} \sum_{j=1}^{N} V(|\vec{x}_i - \vec{x}_j|) \right), \tag{9}$$

where the effective potential $V(r)$ must be tuned to match the distribution of pairwise distances.

Once we have a model for the joint distribution of positions and velocities, we can integrate out the velocities to give the distribution of positions alone. We will refer to this as the "motional distribution", $P_{\rm mot}(x)$, because if we start in the simplest case of Eq. (8) all the nontrivial structure of this distribution arises from the motion of the birds. We have

$$P_{\rm mot}(x) \equiv \int ds P(x,s) = \frac{Z(x)}{Z_0}. \tag{10}$$

Thus, allowing ourselves the usual language of statistical mechanics, the free energy $F(x) = -\ln Z(x)$ acts as an effective potential for the flock,

$$P_{\rm mot}(x) \propto e^{-F(x)}. \tag{11}$$

Notice that if the flock is perfectly ordered, so that all $\vec{s}_i$ are equal, then the exponential in Eq. (5) is just JN, independent of x. In fact, real flocks are highly polarized, and we can compute $Z(x)$ with an expansion around this perfectly ordered state [21, 22] by means of the spin-wave approximation—a method used in solid-state physics to study perturbations in fully ordered ferromagnetic states [30]. In this approximation, the free energy $F(x)$ is dominated by the entropy of the fluctuations in the flight directions, so that gradients in this free energy result in an entropic effect on the birds' spatial configuration.

We conclude this discussion with a cautionary remark about the interpretation of maximum-entropy models, and their relation to equilibrium statistical physics. When we look at Eq. (11), it is tempting to note the equivalence with a Boltzmann distribution and interpret $F(x)$ as the Hamiltonian of the system, and we will sometimes lapse into this language ourselves. In more biologically motivated models of flocks, one speaks of "social forces" that drive cohesion and orientational ordering [31–35], and one might tempted to identify these social forces with derivatives of the terms in the effective Hamiltonian $F(x)$, but this need not be correct. Indeed, the maximum-entropy construction does not imply that $F(x)$ is really the energy of the system, nor does it even mean that the dynamics of the system correspond to the Brownian motion in the potential $F(x)$. As is well known, there are infinitely many dynamical processes that can give rise to the same stationary distribution [36]. The maximum-entropy method aims to characterize this distribution directly, incorporating only the minimal structure needed to match a small set of empirical observations. The choice of which observations to match is based on the phenomenology: for example, in the case of bird flocking the marked polarization of flight directions suggests that a natural quantity to match is the correlation between directions of flight [21]. Such choice of matched observations must be tested by checking that the resulting maximum-entropy distributions provide an accurate description of the system, as in Ref. [21]. Finally, the equivalence between the maximum-entropy distribution and equilibrium statistical physics models means that we can carry over much of what we know about expectation values, correlation functions and, as we have seen, entropic effects. But we cannot jump from this probabilistic description back to a model of the underlying dynamics.

III. THE MOTIONAL FREE ENERGY

Our task now is to compute the partition function in Eq. (5). It is useful to note that this can be rewritten more symmetrically, as

$$Z(x) = \int ds \exp\left(\sum_{i,j=1}^{N} J_{ij}(x)\, \vec{s}_i \cdot \vec{s}_j \right), \tag{12}$$

$$J_{ij}(x) = \frac{J}{2}\left(\frac{n_{ij}(x)}{n_i(x)} + \frac{n_{ji}(x)}{n_j(x)} \right). \tag{13}$$

For future reference, we observe that a configuration x defines a graph $G(x)$ with N vertices, where each nonzero element of the adjacency matrix $n_{ij}(x)$ corresponds to an edge between vertices i and j in G—see below for details. We denote by k the number of connected components in G, and by N_l the number of birds in the l-th connected component, with $l = 1, \ldots, k$. We can relabel the birds so that the matrix J_{ij} consists of k uncoupled blocks.

To evaluate $Z(x)$, we will use the spin-wave approximation, which is valid at large J, such that each connected component of the flock is strongly polarized. For each block we define the net polarization

$$\vec{S}_l \equiv \frac{1}{N_l} \sum_{i \in N_l} \vec{s}_i \equiv S_l \hat{n}_l, \tag{14}$$

where S_l, $\hat{n}_l$ are the norm and the direction of $\vec{S}_l$, respectively, and the sum for $i \in N_l$ runs over all birds in the l-th connected component. We now decompose the

velocity $\vec{s}_i$, with $i \in N_l$, into components parallel and perpendicular to $\vec{S}_l$

$$\vec{s}_i = s_i^L \hat{n}_l + \vec{\pi}_i. \tag{15}$$

Substituting into Eq. (12), we obtain

$$Z(x) = \prod_{l=1}^{k} \int d\vec{S}_l \, d^{N_l} s^L \, d^{N_l}\vec{\pi} \prod_{i \in N_l} \frac{\delta\left(s_i^L - \sqrt{1-|\vec{\pi}_i|^2}\right)}{2\sqrt{1-|\vec{\pi}_i|^2}} \exp\left(\sum_{i,j \in N_l}^{N} J_{ij}(x)(s_i^L s_j^L + \vec{\pi}_i \cdot \vec{\pi}_j)\right) \delta\left(\vec{S}_l - \frac{1}{N_l}\sum_{i \in N_l} \vec{s}_i\right), \tag{16}$$

where in Eq. (16) the first Dirac delta results from Eq. (6), we used the fact that J_{ij} is a block matrix, we inserted a factor of unity

$$\int d\vec{S}_l \, \delta\left(\vec{S}_l - \frac{1}{N_l}\sum_{i \in N_l} \vec{s}_i\right) = 1,$$

and we rewrote the dot product $\vec{s}_i \cdot \vec{s}_j$ in terms of the $s^L, \vec{\pi}$ coordinates by using Eq. (15).

Guided by the experimental observation that birds in a block fly in directions mostly parallel to $\vec{S}_l$ [21], we assume that the perpendicular velocity components $\vec{\pi}_i$ are small, and we will thus expand the right-hand side of Eq. (16) in powers of $\vec{\pi}_i$ with the spin-wave approximation. Specifically, we manipulate Eq. (16) as follows: We rewrite the Dirac delta function in Eq. (16) in terms of the $s^L, \vec{\pi}$ coordinates as

$$\delta\left(\vec{S}_l - \frac{1}{N_l}\sum_{i \in N_l} \vec{s}_i\right) = \delta\left(S_l - \frac{1}{N_l}\sum_{i \in N_l} s_i^L\right) \delta\left(\frac{1}{N_l}\sum_{i \in N_l} \vec{\pi}_i\right), \tag{17}$$

we integrate with respect to $\{s_i^L\}$, we expand the term in parentheses in the exponent to leading order as

$$s_i^L s_j^L = \sqrt{1-|\vec{\pi}_i|^2}\sqrt{1-|\vec{\pi}_j|^2} \approx 1 - \frac{|\vec{\pi}_i|^2}{2} - \frac{|\vec{\pi}_j|^2}{2},$$

and we perform the integration with respect to $\vec{S}_l$ in spherical coordinates. We obtain

$$Z(x) = \prod_{l=1}^{k} 4\pi \int d^{N_l}\vec{\pi} \prod_{i \in N_l} \frac{1}{2\sqrt{1-|\vec{\pi}_i|^2}} \exp\left[\sum_{ij \in N_l} J_{ij}(x)\left(1 - \frac{|\vec{\pi}_i|^2}{2} - \frac{|\vec{\pi}_j|^2}{2} + \vec{\pi}_i \cdot \vec{\pi}_j\right)\right] \delta\left(\frac{1}{N_l}\sum_{i \in N_l} \vec{\pi}_i\right). \tag{18}$$

Note that the factors of 4π in Eq. (18) arise from the angular integration over all possible directions of the mean velocity $\vec{S}_l$ of the l-th connected component: these are explicitly entropic terms.

We now rewrite the square-root term in Eq. (18) as

$$\frac{1}{\sqrt{1-|\vec{\pi}_i|^2}} = \exp\left[-\frac{1}{2}\log\left(1-|\vec{\pi}_i|^2\right)\right] \approx \exp\left(\frac{|\vec{\pi}_i|^2}{2}\right), \tag{19}$$

we use Eq. (19) in Eq. (18), and we obtain

$$Z(x) = \prod_{l=1}^{k} 4\pi \int d^{N_l}\vec{\pi} \exp\left[\sum_{ij \in N_l} J_{ij}(x)\left(1 - \frac{|\vec{\pi}_i|^2}{2} - \frac{|\vec{\pi}_j|^2}{2} + \vec{\pi}_i \cdot \vec{\pi}_j\right) + \frac{1}{2}\sum_{i \in N_l} |\vec{\pi}_i|^2\right] \delta\left(\frac{1}{N_l}\sum_{i \in N_l} \vec{\pi}_i\right), \tag{20}$$

where we omitted a multiplicative factor independent of x. We now observe that the first term in the exponential of Eq. (20) is of the order of $J|\vec{\pi}|^2$, while the second term is of the order of $|\vec{\pi}|^2$: We assume that J is large enough [21] to neglect the second term, and we obtain

$$Z(x) = \prod_{l=1}^{k} 4\pi \int d^{N_l}\vec{\pi} \exp\left[\sum_{ij \in N_l} J_{ij}(x)\left(1 - \frac{|\vec{\pi}_i|^2}{2} - \frac{|\vec{\pi}_j|^2}{2} + \vec{\pi}_i \cdot \vec{\pi}_j\right)\right] \delta\left(\frac{1}{N_l}\sum_{i \in N_l} \vec{\pi}_i\right). \tag{21}$$

In what follows, we will drop the first addend in the exponential in Eq. (21), because this term gives rise to a multiplication factor e^{NJ} which is independent of the positional configuration x as noted above. Then, since $\vec{\pi}_i$ is a two-dimensional vector, the integral with respect to $\vec{\pi}_i$ can be rewritten as a product of two identical integrals

$$Z(x) = \prod_{l=1}^{k} 4\pi \left[\int d^{N_l}\pi \exp\left(-\sum_{ij\in N_l} \pi_i \Lambda_{ij}(x) \pi_j \right) \delta\left(\frac{1}{N_l} \sum_{i\in N_l} \pi_i \right) \right]^2, \tag{22}$$

where π_i denotes one component of $\vec{\pi}_i$, and we introduced the Laplacian

$$\Lambda_{ij}(x) \equiv \delta_{ij} \sum_{l=1}^{N} J_{li}(x) - J_{ij}(x); \tag{23}$$

since J_{ij} has a block structure, so does Λ_{ij}. We denote by $\{\lambda_p^l\}$ and $\{\vec{v}_p^l\}$ the eigenvalues and eigenvectors of the l-th block of the Laplacian, respectively. We will now make use of two known results from graph theory: First, all Laplacian eigenvalues $\{\lambda_p^l\}$ are non-negative [37]. Second, since the l-th component of the flock is connected, the l-th block of the Laplacian has only one zero eigenvalue [38], which we denote by λ_1^l. Therefore, we have

$$0 = \lambda_1^l < \lambda_2^l \leq \cdots \leq \lambda_{N_l}^l, \tag{24}$$

where we labelled the nonzero eigenvalues in increasing order. In addition, summing both sides of Eq. (23) with respect to $j \in N_l$, we find that the vector $\vec{v}_1^l = 1/\sqrt{N_l}(1, \cdots, 1)$ is the eigenvector with eigenvalue $\lambda_1^l = 0$. We now rewrite the Laplacian in terms of its eigenvalues and eigenvectors:

$$\Lambda_{ij} = \sum_{p=1}^{N_l} v_{p,i}^l \lambda_p^l v_{p,j}^l, \quad i,j \in N_l. \tag{25}$$

Setting $u_{l,p} \equiv \sum_{i=1}^{N_l} v_{p,i}^l \pi_i$, we finally obtain

$$\begin{aligned} Z(x) &= \prod_{l=1}^{k} 4\pi \left[\int d^{N_l} u \exp\left(-\sum_{p=2}^{N_l} \lambda_p^l(x) u_{l,p}^2 \right) \delta\left(\frac{u_{l,1}}{\sqrt{N_l}} \right) \right]^2 \\ &= \prod_{l=1}^{k} 4\pi N_l \prod_{p=2}^{N_l} \frac{\pi}{\lambda_p^l(x)}. \end{aligned} \tag{26}$$

Because of the delta function in the first line, Eq. (26) involves an integration with respect to the projection of the perpendicular velocity π on all Laplacian eigenmodes except the first one: as a result, the final expression for $Z(x)$ in the second line involves a product over all nonzero Laplacian eigenvalues.

Following Eqs. (10) and (26), we obtain

$$P_{\rm mot}(x) \propto Z(x) = \prod_{l=1}^{k} 4\pi N_l \prod_{p=2}^{N_l} \frac{\pi}{\lambda_p^l(x)}. \tag{27}$$

Equation (27) tells us that if the flock with N birds consists of one connected cluster, then there is an overall factor of $4\pi N$, and $N-1$ powers of J/π in the denominator of $P_{\rm mot}$, since all $\lambda_p^l \propto J$. If the flock is cut in two halves, then we lose one power of J/π, and the factor $4\pi N \to (4\pi N/2)^2$; the net result is to multiply $P_{\rm mot}$ by NJ. This factor is essentially the increase in entropy associated with the creation of a new zero mode in the joint distribution; evidently for large N and large J, it favors breaking the flock in half. There is some subtlety, however, since when we break the flock in half we also shift the eigenvalues λ_p^l in each half, and it is not clear how this balances against the zero mode. We will see that the answer depends on the nature of the interactions between birds.

IV. METRIC VS TOPOLOGICAL INTERACTIONS

In what follows we will consider a flock of N birds in a volume V, hence at mean density $\rho = N/V$. We can imagine the interaction between birds having two very different forms [27]. In the first one, birds interact with

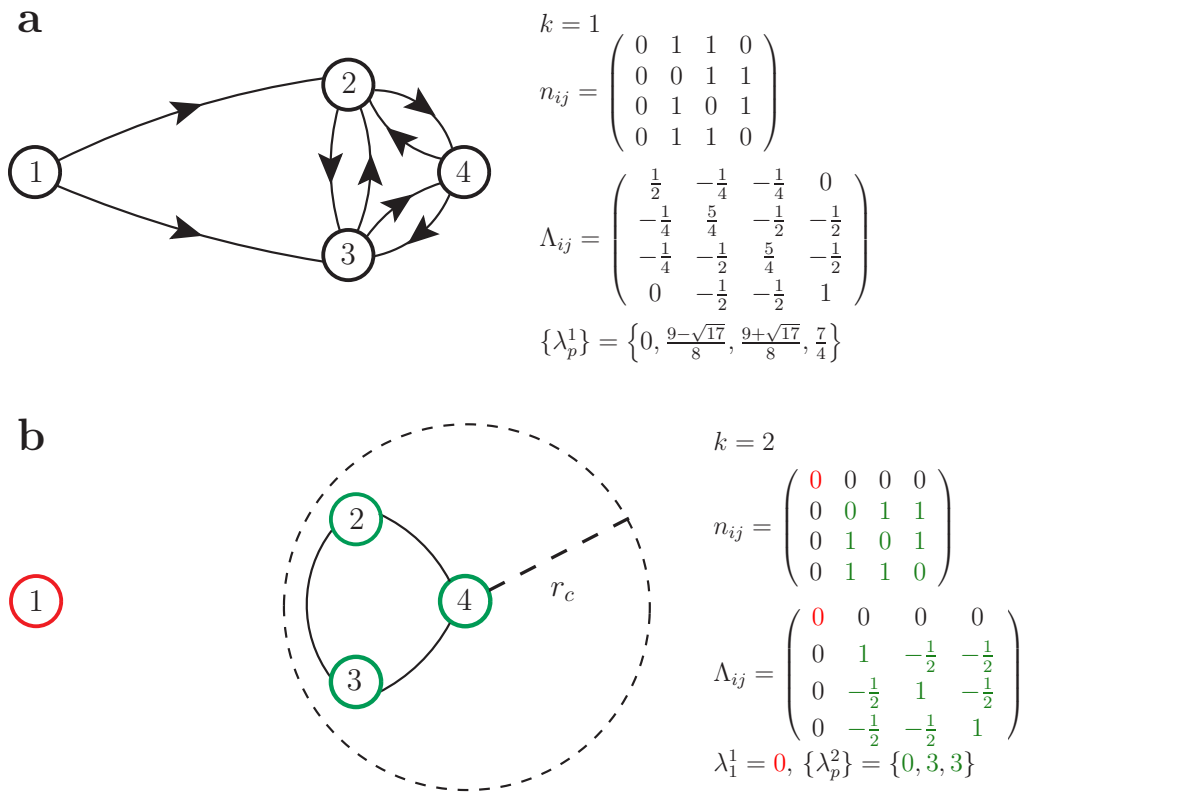

FIG. 1: Graphs corresponding to a a two-dimensional flock with $N = 4$ birds and $J = 1$. The number of connected components, the adjacency and Laplacian matrices, and the Laplacian eigenvalues are also shown. (a) Topological case: Here $n_c = 2$, and the graph G is a directed graph with one connected component. The Laplacian has only one null eigenvalue, corresponding to one connected component in the graph [38]. (b) Metric case: r_c is the radius of the dashed circle centered around bird 4, and the graph G is an undirected graph with two connected components. The first connected component is given by bird 1 (in red), and the second connected component is given by birds $2, 3, 4$ (in green). The adjacency and Laplacian matrices are block matrices composed of two blocks corresponding to the first and second connected component depicted in red and green, respectively. The Laplacian has two null eigenvalues, corresponding to two connected components in G.

their n_c nearest neighbors, independently of distance. In the second one, birds interact with other birds within a characteristic distance r_c. The first model is referred to as a "topological" interaction, while the second is summarized as a "metric" interaction [21, 27]. We can compare the two models by equating the mean number of interacting neighbors in the topological case with the fixed number of neighbors in the metric case:

$$n_c = \rho \frac{4\pi r_c^3}{3}. \tag{28}$$

In both models this is the only relevant dimensionless parameter at large N and V.

For a topological interaction network, the adjacency matrix is

$$n_{ij}(x) \equiv \begin{cases} 1 & \text{If } j \neq i \text{ is amongst the first } n_c \\ & \text{nearest neighbors of } i \\ 0 & \text{Otherwise} \end{cases}. \tag{29}$$

Here we have $n_i = n_c$ for all i, and Eq. (13) takes the simple form

$$J_{ij}(x) = \frac{J}{2n_c}(n_{ij}(x) + n_{ji}(x)). \tag{30}$$

For topological interactions, the graph $G(x)$ which corresponds to a configuration x is a directed graph, and it can be constructed by considering all vertex pairs i, j and inserting a directed edge from i to j if j is amongst the first n_c nearest neighbors of i—see Fig. 1a. In what follows, we will consider a directed graph as connected if replacing all of its directed edges with undirected edges produces a connected undirected graph.

For a metric interaction network, the adjacency matrix is

$$n_{ij}(x) \equiv \begin{cases} 1 & \text{If } j \neq i \text{ and } |\vec{x}_j - \vec{x}_i| \leq r_c \\ 0 & \text{Otherwise} \end{cases}, \tag{31}$$

and the interaction matrix J_{ij} is given by Eq. (13). In the metric case, the graph $G(x)$ which corresponds to a configuration x is an undirected graph constructed by inserting an edge between all vertex pairs i, j such that $|\vec{x}_j - \vec{x}_i| \leq r_c$ as shown in Fig. 1b.

In the first models for collective behavior, the metric structure of interactions seemed obvious. Later on, quantitative studies on flocks of starlings [27] led to the idea of topological interactions, and this has been supported by further analyses of these data [21]. In particular, the evidence for topological vs. metric interactions in previous studies is based on the comparison of multiple flocking events with significantly different densities, where each event is given by a single, connected flock with fixed birds' relative positions [21]: across many such events, the data are described very accurately by a fixed number of neighbors n_c, rather than by a fixed interaction range r_c. Here, we will allow model flocks to explore a much wider range of configurations where birds significantly rearrange their relative positions, and the flock possibly disconnects driven by the entropic effects above, and so we expect the distinction between topological and metric interactions to emerge more distinctly.

V. MONTE CARLO SAMPLING OF THE MOTIONAL DISTRIBUTION

We are interested in exploring the entropic effects generated by the alignment of the birds, as described by Eq. (27). Even though we can write the distribution of positions explicitly, computing analytically averages with the distribution (27) is not an easy task, it is thus natural to generate samples from this distribution with MC simulations. We will do this in Section V A, for both metric and topological interactions. Although there are clear results, we find that it is difficult to push these simulations to large N, because calculating the eigenvalues of the Laplacian requires $O(N^3)$ operations. To make progress, in Section V B we introduce a simpler version of the problem, in which we sample interaction graphs rather than the underlying positions. While not quite the same problem, we will see consistent results that extend to much larger N. The essential point to emerge from both analyses is that topological interactions lead to flocks that stay connected for realistic values of the number of neighbors, while in metric networks the repulsive entropic effects are strong enough to rip the flock apart into multiple components for all sensible values of the metric interaction range.

A. Monte Carlo on positions

Equation (27) gives us the motional distribution of birds' positions in a flock with only two parameters: The strength of the interactions J and their range n_c. Before we study the effects of these interactions on the spatial configurations of the flock, we should start by asking what happens if the N birds are simply in random positions drawn uniformly throughout a box of volume V, which will be set equal to unity in what follows. Even in this simple case there is a question about whether the resulting network of interactions—topological or metric—supports a single, connected cluster of birds, thus allowing for the possibility of coherent flocking behavior.

Concretely, for any spatial configuration we can construct the adjacency matrix $n_{ij}(x)$, see Eq. (29) or (31), the resulting interaction matrix $J_{ij}(x)$ in Eq. (13), and finally the Laplacian $\Lambda_{ij}(x)$ in Eq. (23). Then we count the zero modes of the Laplacian: if there is just one, then the flock is connected; if there is more than one, then the flock has broken into disconnected pieces.

In Fig. 2 we show results on the probability p_c of finding a single connected cluster for random configurations, which is computed as the fraction of configuration samples that are given by a connected interaction network. We plot p_c as a function of n_c for topological and metric interactions and we see that, for topological interactions, connectedness is guaranteed by very modest values of n_c. Perhaps surprisingly, this is not the case for metric interactions. When interactions are limited to a fixed distance, even random fluctuations are enough to prevent the formation of a single connected cluster, unless the range of interactions is very large. This is a hint that metric and topological interactions really are very different.

To see the effects of entropic forces, we need to draw samples from Eq. (27). We will view n_c as a parameter to be varied as for random positions, but we would like to set the strength of interactions to some reasonable value. To do so, we will use the value $J/n_c \sim 16$ which has been obtained from maximum-entropy models built for real, connected flocks with fixed relative positions [21] by matching the observed values of $C_{\rm int}$ [40]. As we draw samples from $P_{\rm mot}(x)$, we focus on the probability p_c that these spatial configurations correspond to a single connected cluster; results are shown in Fig. 2. We see that the entropic forces are repulsive, so that flocks which interact through the motional potential are less likely to be connected than random flocks, and this is true at all the values of N and n_c that we can access, for both metric and topological interactions. But for topological interactions, once $n_c \gtrsim 7$ we find that p_c is essentially equal to one. In contrast, for the metric interactions there is a very gradual dependence of p_c on n_c that sets in only at large n_c. This effect is dramatic because—in the range of N we can study—the entropic forces are ripping the flock apart even when the range of metric interactions is so large that $n_c > N$.

How does a flock disconnect for small values of the interaction range? To clarify this point, in Fig. 3 we plot the MC average of the fraction of birds in the largest connected component, which we will denote by ϕ. Comparing Fig. 3 with Fig. 2, we see that even for values of n_c so small that the connection probability is $\sim 50\%$, there is a significant fraction of birds in the largest connected component, and this is true for both topological

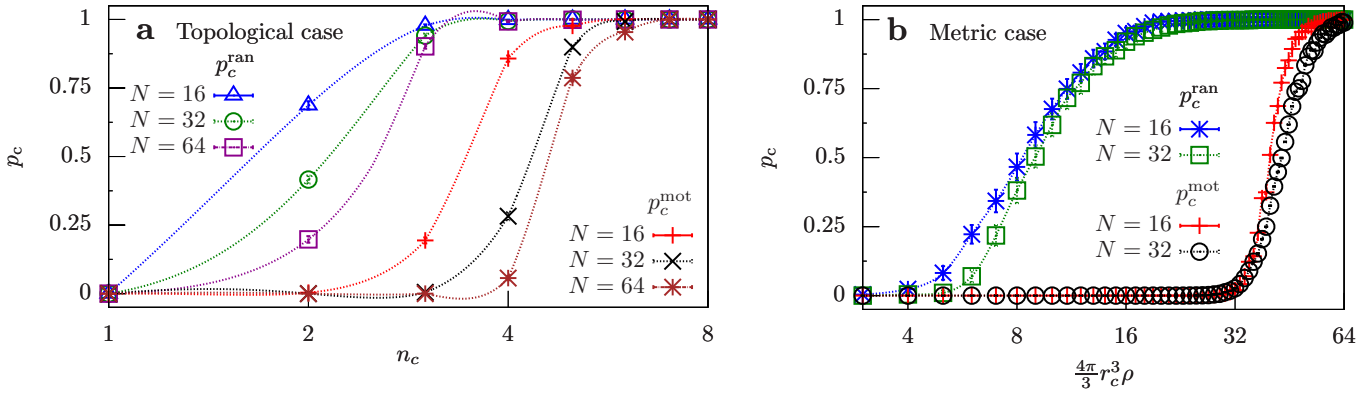

FIG. 2: Probability of finding a connected flock. Connection probabilities $p_c^{\rm ran}$, $p_c^{\rm mot}$, from random positions and from the motional potential, respectively, as functions of the number of neighbors n_c, both in the topological case (a) and in the metric case (b), where the continuous curves are shown to guide the eye. In the metric case, the average number of neighbors n_c is obtained from the interaction range r_c from the relation $n_c = \rho \frac{4\pi}{3} r_c^3$, where ρ is the average density of the flock. Error bars have been estimated with the bootstrap method [39].

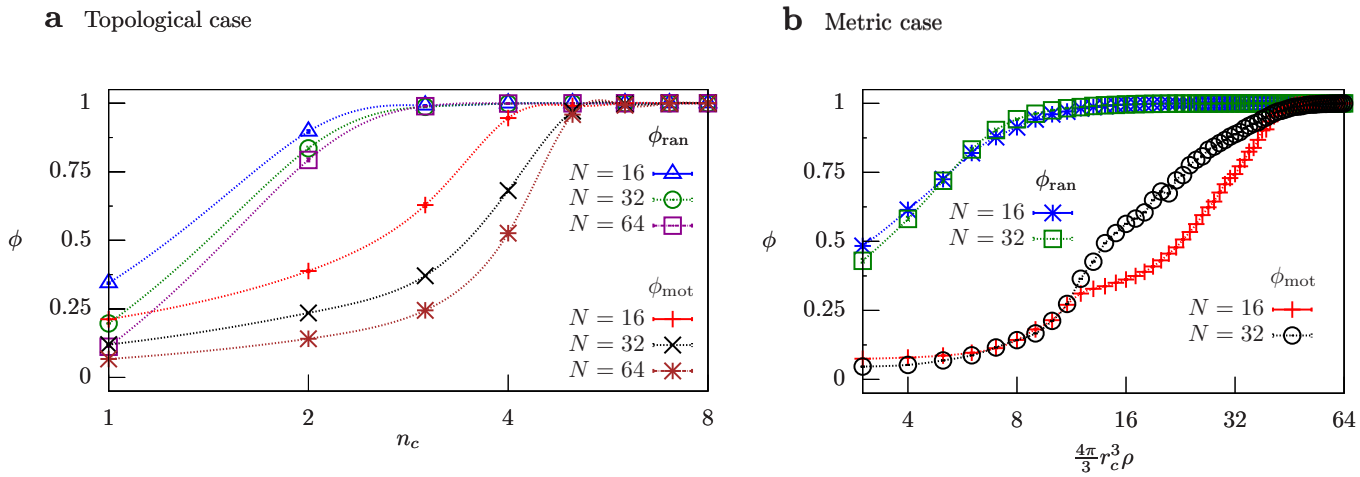

FIG. 3: Fraction of birds in the largest connected component of a flock. The fraction ϕ from the motional potential and from random positions is plotted as a function of the number of neighbors n_c, both in the topological case (a) and in the metric case (b). In the metric case, the average number of neighbors n_c is obtained as in Fig. 2 .

and metric interactions, and for configurations from both the motional potential and random positions. It follows that, as n_c is lowered, the flock disconnection is characterized by the appearance of multiple, small connected components, each of which contains a few individuals, while the vast majority of birds are still part of a giant component with $\sim N$ birds. In addition, Fig. 3 shows that the curves drawn from the motional potential in the metric case display an elbow at $n_c \approx 11$ and $n_c \approx 16$ for $N = 16$ and $N = 32$, respectively. These values of n_c correspond, through Eq. (28), to a characteristic distance r_c of the order of the linear size of the volume V: the elbows above could thus be interpreted in terms of a crossover between a regime $r_c \lesssim V^{1/3}$ where each birds interacts with individuals in a fraction of V, to a regime $r_c \gtrsim V^{1/3}$ where the interaction spans the whole volume.

B. Monte Carlo on graphs

The adjacency matrix $n_{ij}(x)$ defines a graph $G(x)$, as shown in Section IV. In the topological case, any positional configuration x corresponds to a directed graph $G(x)$ with a fixed number n_c of outgoing edges per vertex and thus a fixed total number of edges. Conversely, in the metric case the number of edges connected to a vertex may vary depending on the positional configuration x, but the total number of edges M is almost fixed, with

$$M = \sum_{i<j=1}^{N} n_{ij} = \frac{N}{2} \left(\frac{1}{N} \sum_{i=1}^{N} n_i \right) \approx \frac{N}{2} n_c. \qquad (32)$$

In the last step we set $(1/N) \sum_{i=1}^{N} n_i \to n_c$, which is valid at large N if there are no long-range correlations among

the fluctuations in n_i. If we could neglect fluctuations in n_i all together, then Eqs. (13) and (23) would show that the Laplacian matrix Λ_{ij} is proportional to a sparse matrix with integer coefficients, allowing for a much more efficient computation, as explained below.

These observations suggest a natural way of simplifying MC simulations. Instead of sampling the positional configurations x, in the topological case we sample all directed graphs with n_c outgoing edges from each vertex, while in the metric case we sample all undirected graphs with $Nn_c/2$ total edges, assigning to a graph G the probability

$$\mathcal{P}_{\rm mot}(G) = C \prod_{l=1}^{k} 4\pi N_l \prod_{p=2}^{N_l} \frac{\pi}{\lambda_l^p(G)}, \qquad (33)$$

where the λ_l^p are calculated in the approximation that $n_i \to n_c$ in Eq. (13). Since the total number of edges is constant both in the topological and metric case, a MC move is now given by an edge insertion and deletion, and the impact of these moves on the eigenvalues of the Laplacian can be computed in $O(N^2)$ operations at the most by using the LDL matrix factorization method [41, 42]—see the Appendix for details. As a result, we will be able to study values of N comparable to those of natural flocks, i.e. $N \sim 1000$.

As before, we focus on the probability that the flock is in a single connected graph, which will be denoted by p_c; results are shown in Fig. 4. In the topological case, this probability is one for all $n_c > 1$, both in the case where graphs are chosen at random ($p_c^{\rm ran}$) and when the graphs are chosen from the motional distribution ($p_c^{\rm mot}$) [43]. In contrast, for the metric case the connection probability for random graphs $p_c^{\rm ran}$ is close to one for only for $n_c \gtrsim 10$, while even larger values of $n_c \gtrsim 15$ are needed for the connection probability from the motional potential to be close to one.

We note that our simplified MC method includes graphs which do not have a three-dimensional layout: While any configuration x can be mapped onto a graph G (Fig. 1), not all graphs G correspond to a configuration x. If we break the graph into disconnected pieces, it is always easier to find a mapping into a configuration x, and so we expect that sampling graphs rather than positions will overestimate the probability that flocks are connected. This is confirmed by a detailed comparison between Figs. 2 and 4. Despite this quantitative difference, the MC on graphs confirms the qualitative scenario from the MC on positions: the metric potential has a strong repulsive effect which rips the flock apart into multiple components.

Figure 2 gives a hint that, with metric interactions, larger flocks have a lower probability of being connected, and this trend continues to larger N for the simulation on graphs in Fig. 4. Put another way, larger flocks require a larger range of interaction n_c in order to stay connected. Marking the crossover n_c^* between connected and disconnected regimes by $p_c(n_c^*, N) = 1/2$ for both random graphs and graphs drawn from the motional distribution, we see in Fig. 5 that n_c^* increases with N along an approximately logarithmic trajectory, such that a factor of two increase in N requires an extra contact to maintain coherence.

Proceeding along the lines of Section V A, we show how the flocks disconnect for small n_c by considering the fraction of birds in the largest connected component averaged with the MC method on graphs, which we will denote by φ. In Fig. 6, we show φ as a function of n_c. The results are along the lines of Fig. 3 obtained with the MC on positions. Namely, φ is close to one even for relatively small values of the connection probability, thus the flock disconnection is characterized by a giant connected component with $\sim N$ birds and multiple, small components containing a few birds each: Interestingly, the curves in Fig. 6 overlap reasonably well for different values of N, suggesting that this result holds also for larger flocks.

VI. DISCUSSION

The past decade has seen considerable interest in the use of maximum-entropy models to describe biological networks, from single protein molecules up to groups of organisms. The maximum-entropy method has deep connections to equilibrium statistical mechanics. Although these connections are a source of intuition, they also create opportunities for confusion. Namely, the maximum-entropy distribution that is consistent with pairwise correlations among the variables in a network, e.g. the normalized velocities of the birds, has the form of a Boltzmann distribution in which the "energy" is built out of pairwise interactions among these variables. This is not, of course, the actual energy, and there is no reason to think that the interactions out of which the energy is built correspond to physical interactions.

In a flock of birds such as European starlings, there is no part of the problem that is in thermal equilibrium, but nonetheless we can write a maximum-entropy approximation to the joint distribution of positions and velocities for all the birds in the flock, as in Eqs. (8) or (9). Once we integrate out the velocities, the resulting motional distribution of positions has a term in the exponential that is exactly the logarithm of the partition function for the velocities at fixed positions—the free energy. For flocks that are strongly polarized, as in real flocks, this free-energy potential on positions is dominated by the entropy of the birds' velocities, and in this sense the birds' positions are subject to entropic effects. We expect these effects to be repulsive, since disconnected groups of birds have more freedom to reorient their flight directions and thus satisfy their tendency to increase the entropy, and this intuition is borne out by detailed simulations. The

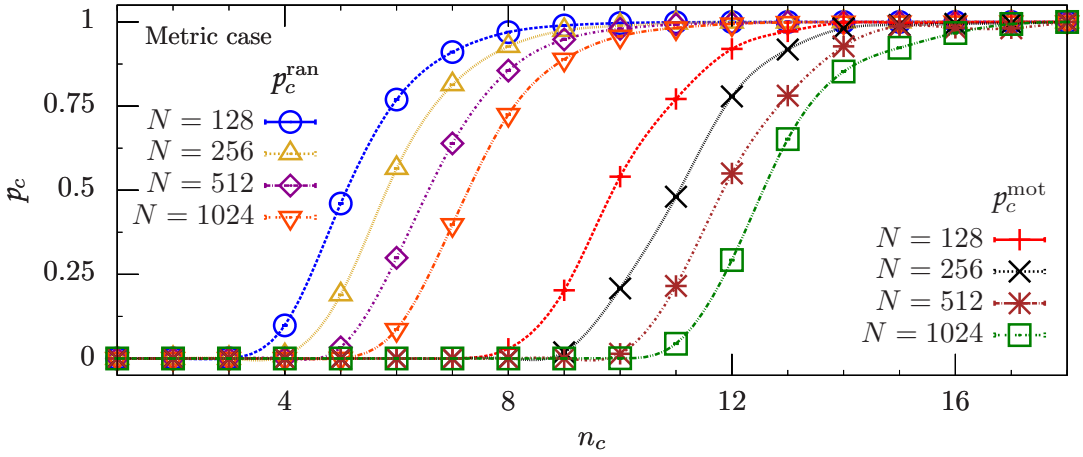

FIG. 4: Probability of finding a connected flock with the simplified Monte Carlo method on graphs in the metric case. Connection probabilities $p_c^{\rm ran}$, $p_c^{\rm mot}$ from random graphs and from the motional potential, respectively, as functions of the number of neighbors n_c. In the topological case, $p_c^{\rm mot}$ is equal to one for all $n_c > 1$, and it is not shown here.

surprise is that the strength of this repulsion depends dramatically on the form of the interactions that we constrain. If we imagine that the essential interactions are between a bird and its n_c nearest neighbors (topological interactions), then the entropic effects are quite weak, and they leave flocks fully connected with high probability at reasonable values of n_c. In contrast, if the essential interactions are between a bird and all the other birds within fixed distance r_c (metric interactions), then the entropic effects are so strong to tear the flock apart into multiple components.

Given the complicated form of the motional distribution, Monte Carlo (MC) simulations are computationally demanding, and they are thus limited to small flocks. To address this problem, we explored a slightly different formulation in which we sample graphs of bird-bird interactions rather than the birds' positions themselves; this allows for using an efficient MC update algorithm based on LDL matrix factorization [41], with which we could analyze realistically large flocks. Our main result is in line with the one obtained by sampling the positions: For a topological network, the configurations generated by the motional distribution are connected for all values of the number of nearest neighbors. On the other hand, for a metric network the positional configurations are disconnected with high probability unless the metric interaction range is increased to unrealistically large values.

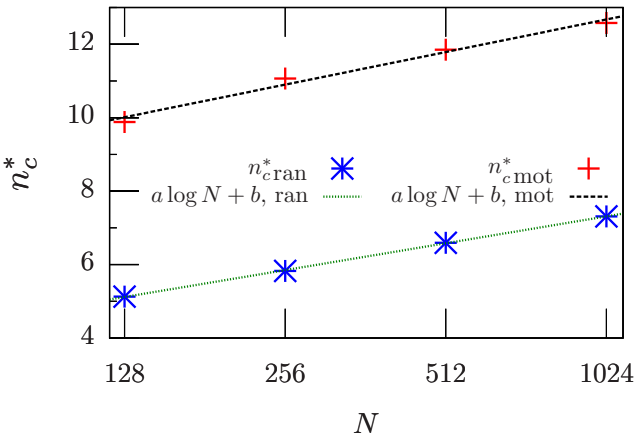

FIG. 5: Crossover values n_c^* of the number of neighbors in the metric case, given by $p_c(n_c^*, N) = 1/2$, for random graphs and for graphs drawn from the motional potential, as functions of the flock size N. The fitting functions $n_c^* = a \log N + b$ are shown to guide the eye.

In real flocks, the absence of large local density fluctuations means that the distinction between topological and metric interactions must be based on comparisons across different flocking events [21]. Our analysis of entropic effects provides a different path to comparing these two kinds of interactions. In particular, an important conclusion of our analysis is that, while the strongly repulsive entropic effects of metric interactions could be compensated by explicit cohesive forces, the fact that positional correlations in flocks are weak [28] means that these strong opposing forces would have to be carefully balanced. Such fine tuning is not needed if the interactions are topological rather than metric.

Overall, the maximum-entropy analysis presented here may have interesting applications for studying entropic effects in systems of biological and physical interest. For example, this maximum-entropy model could be extended to reproduce not only the motional distribution of birds' positions, but the actual, full positional distribution. Our method would then provide a direct estimate of the maximum size that a flock could reach before it breaks into multiple components, thus providing a prediction that could be directly tested in field studies.

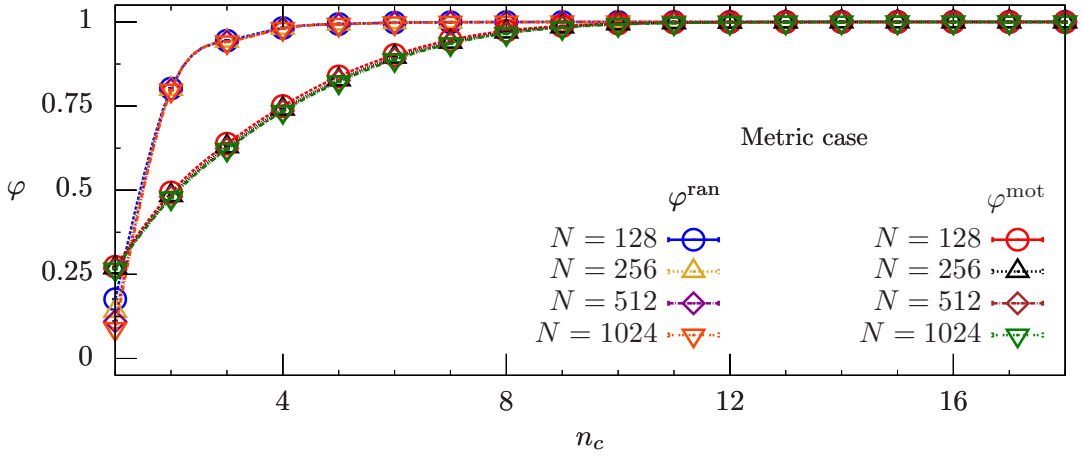

FIG. 6: Fraction of birds in the largest connected component of a flock with the simplified Monte Carlo method on graphs in the metric case. The fraction φ is plotted as a function of the number of neighbors n_c from the motional potential and from random positions, respectively.

Acknowledgments

MC is grateful to TA Davis, WW Hager and RB Israel for advice and support on LDL factorizations, and to M Tikhonov for helpful discussions. Work at Princeton and CUNY was supported in part by NSF Grants PHY–1305525 and CCF–0939370, by the Human Frontiers Science Program, and by the Simons and Swartz Foundations. Work in Rome was supported in part by IIT grant Seed Artswarm, ERC–StG Grant No. 257126, and US–AFOSR Grant No. FA95501010250 (through the University of Maryland). The simulations presented in this article were performed on computational resources supported by the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

Appendix

Here we discuss the MC simulations with the probability $\mathcal{P}_{\rm mot}(G)$ from Eq. (33). In Appendix 1 we show that $\mathcal{P}_{\rm mot}(G)$ can be related to the LDL factorization of the Laplacian matrix (23), and in Appendix 2 we show that a MC step can be performed efficiently by using a known update algorithm for LDL factorizations. For the sake of simplicity, we consider the metric case, and we assume that G is connected. The results below can then be easily extended to the topological case and to graphs with multiple connected components.

1. Relation between the motional probability and LDL factorization of the Laplacian

Since G has only one connected component, there is a single zero eigenvalue of Λ [38], which we denote by λ_1, and Eq. (33) shows that the probability $\mathcal{P}_{\rm mot}(G)$ is determined by the product of the nonzero eigenvalues of the Laplacian (23). Here, we will show that this product is related to the LDL factorization of Λ [42], which reads

$$A \equiv \Pi\Lambda\Pi^T = LDL^T, \tag{34}$$

where in Eq. (34) Π is a permutation matrix, L is a lower-triangular matrix with unit diagonal elements, D is a diagonal matrix whose diagonal elements will be denoted by $\{d_i\}$, and the matrix A has the same eigenvalues as Λ. Since Λ has one zero eigenvalue, there is a single zero diagonal entry in D, which we denote by d_1. We will now establish the connection between the spectrum of Λ and its LDL factorization by proving the following identity

$$\prod_{i=2}^{N} \lambda_i = N \prod_{i=2}^{N} d_i. \tag{35}$$

To prove Eq. (35), let us consider the characteristic polynomial of A: by using Sylvester's theorem [44], we have

$$f(\lambda) \equiv \det(A - \lambda I) = \det(DL^T L - \lambda I). \tag{36}$$

The characteristic polynomial (36) can be rewritten as

$$f(\lambda) = a_N - \lambda a_{N-1} + \cdots + (-1)^N \lambda^N, \tag{37}$$

where a_i is the sum of all diagonal minors of $DL^T L$ containing i rows and i columns [45]. Since $d_1 = 0$, the first row in $DL^T L$ is zero, thus the only nonzero diagonal minor with one row and one column is the one obtained by

deleting the first row and the first column from DL^TL. It follows that

$$a_{N-1} = \det(B), \tag{38}$$

where B is a $(N-1)\times(N-1)$ matrix with entries $B_{ij} = \sum_{l=1}^{N} D_{il}(L^TL)_{lj}$, $i,j = 2,\cdots,N$. Since D is diagonal, B is given by the product of the matrices obtained by removing the first row and the first column in D and L^TL respectively: hence, from Eq. (38) we obtain

$$a_{N-1} = \left(\prod_{i=2}^{N} d_i\right) \det(C_{1\,1}), \tag{39}$$

where $C_{1\,1}$ denotes the matrix obtained from L^TL by deleting the first row and the first column.

From Eq. (24), we have $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$, thus the eigenvalues of $A - \lambda I$ are $-\lambda, \lambda_2 - \lambda, \cdots, \lambda_N - \lambda$. It follows that the characteristic polynomial (36) reads

$$\begin{aligned} f(\lambda) &= (-\lambda)(\lambda_2 - \lambda)\cdots(\lambda_N - \lambda) \\ &= -\lambda \prod_{i=2}^{N} \lambda_i + O(\lambda^2). \end{aligned} \tag{40}$$

Comparing the coefficient of λ in the right-hand side of Eq. (37) with that in the right-hand side of Eq. (40) and using Eq. (39), we obtain

$$\prod_{i=2}^{N} \lambda_i = \det(C_{1\,1}) \prod_{i=2}^{N} d_i. \tag{41}$$

To derive Eq. (35), let us compute $\det(C_{1\,1})$. Equations (23) and (34) show that the vector $u \equiv (1,\cdots,1)$ is an eigenvector of A with eigenvalue zero. Setting $e_1 \equiv (1,0,\cdots,0)$, we have that $(L^T)^{-1}e_1$ is also an eigenvector of A with eigenvalue zero:

$$\begin{aligned} A(L^T)^{-1}e_1 &= LDL^T(L^T)^{-1}e_1 \\ &= LDe_1 \\ &= 0, \end{aligned} \tag{42}$$

where in the second line of Eq. (42) we have $De_1 = 0$ because $d_1 = 0$. Given that Λ is symmetric, the geometric multiplicity of λ_1 is equal to its algebraic multiplicity, the latter being equal to one. It follows that there is only one eigenvector of Λ with zero eigenvalue, thus only one eigenvector of A with zero eigenvalue: hence, u must be proportional to $(L^T)^{-1}e_1$. Also, $(L^T)^{-1}$ is upper triangular with unit diagonal entries, thus the first component of $(L^T)^{-1}e_1$ is equal to one, implying that

$$u = (L^T)^{-1}e_1. \tag{43}$$

To prove Eq. (35), we will relate the norm of u to L:

$$\begin{aligned} N &= u^Tu \\ &= e_1^TL^{-1}(L^T)^{-1}e_1 \\ &= e_1^T(L^TL)^{-1}e_1 \\ &= [(L^TL)^{-1}]_{1\,1}, \end{aligned} \tag{44}$$

where in the first line of Eq. (44) we used Eq. (43), and $[(L^TL)^{-1}]_{1\,1}$ denotes the entry in the first row and first column of $(L^TL)^{-1}$. By using Cramer's rule, the last line in Eq. (44) can be rewritten as

$$[(L^TL)^{-1}]_{1\,1} = \frac{\det(C_{1\,1})}{\det(L^TL)} = \det(C_{1\,1}), \tag{45}$$

where in Eq. (45) we used the identity $\det(L^TL) = [\det(L)]^2 = 1$. Equations (44) and (45) imply that $\det(C_{1\,1}) = N$. Substituting into Eq. (41), we obtain Eq. (35).

2. Monte Carlo with LDL-factorization update

Since we intend to sample the space of graphs with a constant total number of edges, a MC move is given by one edge insertions and one edge deletion. We use Eqs. (23) and (30) to rewrite the Laplacian of G as

$$\Lambda_{ij} = \frac{J}{n_c}\left[\left(\sum_{l=1}^{N} n_{il}\right)\delta_{ij} - n_{ij}\right]. \tag{46}$$

We then take two vertices i and j in G that are not connected, and we insert an edge between them. As a result, we obtain a new graph G' with Laplacian Λ',

$$\Lambda' = \Lambda + \frac{J}{n_c}\, v\cdot v^T, \tag{47}$$

where the vector v is given by $v_l = \delta_{il} - \delta_{jl}$. Since Λ is sparse and Λ' is related to Λ by a transformation of the form (47), it can be shown [41] that the LDL factorization of the Laplacian $\Lambda' = L'D'L'^T$ can be computed from the LDL factorization of Λ in a number of steps proportional to the number of nonzero entries in L that change upon the update, which is bounded above by $O(N^2)$. The probability $\mathcal{P}_{\rm mot}(G')$ is then obtained from D' according to Eq. (35). An edge insertion can be thus performed with not more than $O(N^2)$ operations; by the same argument, an edge deletion—and thus a full MC move—can be also performed with $O(N^2)$ operations at the most.

Finally, it is important to point out that replacing the denominators n_i, n_j in Eq. (13) with their average value n_c—see Section V B—is crucial for this efficient update method to work: indeed, without this simplification the Laplacian would not have the simple form (46), and a Laplacian update upon edge insertion or deletion would not be of the form (47).

[1] N. Roos. Entropic forces in Brownian motion. *Am. J. Phys.*, 82(12):1161, 2014.

[2] G. W. Slater, S. J. Hubert, and G. I. Nixon. Construction of approximate entropic forces for finitely extensible nonlinear elastic (FENE) polymers. *Macromol. Theor. Simul.*, 3(4):695, 1994.

[3] J. C. Crocker, J. A. Matteo, A. D. Dinsmore, and A. G. Yodh. Entropic attraction and repulsion in binary colloids probed with a line optical tweezer. *Phys. Rev. Lett.*, 82(21):4352, 1999.

[4] J. Bostock and H. T. Riley. *Pliny the Elder, The Natural History. Book X, The natural history of birds. Chapter 35.* Taylor and Francis, London, 1855.

[5] W. Bialek and R. Ranganathan. Rediscovering the power of pairwise interactions. arXiv:0712.4397 [q-bio.QM], 2007.

[6] F. Seno, A. Trovato, J. R. Banavar, and A. Maritan. Maximum entropy approach for deducing amino acid interactions in proteins. *Phys. Rev. Lett.*, 100(7):078102, 2008.

[7] 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. *P. Natl. Acad. Sci. USA*, 106(1):67, 2009.

[8] 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(12):e28766, 2011.

[9] J. I. Sulkowska, F. Morocos, M. Weigt, T. Hwa, and J. N. Onuchic. Genomics–aided structure prediction. *P. Natl. Acad. Sci. USA*, 109(26):10340, 2012.

[10] T. R. Lezon, J. R. Banavar, M. Cieplak, A. Maritan, and N. V. Fedoroff. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *P. Natl. Acad. Sci. USA*, 103(50):19033, 2006.

[11] G. Tkačik. *Information Flow in Biological Networks.* PhD thesis, Princeton University, 2007.

[12] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007, 2006.

[13] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek. Ising models for networks of real neurons. arXiv:q-bio/0611072 [q-bio.NC], 2006.

[14] J. Shlens, G. D. Field, J. L. Gaulthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky. The structure of multi–neuron firing patterns in primate retina. *J. Neurosci.*, 26(32):8254, 2006.

[15] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J. Neurosci.*, 28(2):505, 2008.

[16] G. Tkacik, E. Schneidman, M. J. Berry II, and W. Bialek. Spin glass models for a network of real neurons. arXiv:0912.5409, 2009.

[17] I. E Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor. Sparse coding and higher–order correlations in fine–scale cortical networks. *Nature*, 466(7306):617, 2010.

[18] E. Ganmor, R. Segev, and E. Schniedman. Sparse low–order interaction network underlies a highly correlated and learnable neural population code. *P. Natl. Acad. Sci. USA*, 108(23):9679, 2011.

[19] E. Granot-Atedgi, G. Tkačik, R. Segev, and E. Schneidman. Stimulus–dependent maximum entropy models of neural population codes. *PLoS Comput. Biol.*, 9(3):e1002922, 2013.

[20] G. Tkačik, T. Mora, O. Marre, D. Amodei, M. J. Berry II, and W. Bialek. Thermodynamics for a network of neurons: Signatures of criticality. arXiv:1407.5946 [q-bio.NC], 2014.

[21] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak. Statistical mechanics for natural flocks of birds. *P. Natl. Acad. Sci. USA*, 109(13):4786, 2012.

[22] 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. *P. Natl. Acad. Sci. USA*, 111(20):7212, 2014.

[23] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620, 1957.

[24] G. Parisi. *Statistical Field Theory.* Perseus Books, New York, 1998.

[25] J. T. Emlen. Flocking behaviour in birds. *The Auk.*, 69(2):160, 1952.

[26] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Empirical investigation of starling flocks: A benchmark study in collective animal behaviour. *Anim. Behav.*, 76(1):201, 2008.

[27] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and Z. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *P. Nat. Acad. Sci. USA*, 105(4):1232, 2008.

[28] A. Cavagna, A. Cimarelly, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, R. Santagati, and F. Stefanini. New statistical tools for analyzing the structure of animal groups. *Math. Biosci.*, 214(1–2):32, 2008.

[29] A. Cavagna, S. M. D. Queiros, I. Giardina, F. Stefanini, and M. Viale. Diffusion of individual birds in starling flocks. *Proc. Roy. Soc. B-Biol. Sci.*, 280(1756):20122484, 2013.

[30] F. J. Dyson. General theory of spin-wave interactions. *Phy. Rev.*, 102(5):1217, 1956.

[31] I. Aoki. A simulation study on the schooling mechanism in fish. *Bull. Jpn. Soc. Sci. Fish*, 48(8):1081, 1982.

[32] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75(6):1226, 1995.

[33] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1):1, 2002.

[34] G. Grégoire, H. Chaté, and Y Tu. Moving and staying together without a leader. *Physica D–Nonlinear Phenomena*, 181(3–4):157, 2003.

[35] H. Hildenbrandt, C. Carere, and C. K. Hemelrijk. Self-organized aerial displays of thousands of starlings: a

model. *Behav. Ecol.*, 21(6):1349–1359, 2010.

[36] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Rev. Mod. Phys.*, 49(3):435, 1977.

[37] Mohar B. Some applications of Laplace eigenvalues of graphs. In Geňa H. and Gert S., editors, *Graph Symmetry*, volume 497 of *NATO ASI Series*, page 226. Springer Netherlands, 1997.

[38] J. S. Caughman and J. J. P. Veerman. Kernels of directed graph laplacians. *Electr. J. Comb.*, 13(1):R39, 2006.

[39] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999.

[40] In Eq. (4), with topological interactions we have $n_i(x) = n_c$ for all i, and hence

$$P(s|x) = \frac{1}{Z(x)} \exp\left(\frac{J}{n_c} \sum_{i=1}^{N} \sum_{j=1}^{N} n_{ij}(x) \vec{s}_i \cdot \vec{s}_j\right). \qquad (48)$$

Comparing with Ref. [21], we see that what was J in the previous work is J/n_c in the current formulation. In the metric case, we compare Eq. (4) with the flock configurations in Ref. [21] with a nearly homogeneous density ρ by setting $n_i(x) \approx n_c = 4\pi\rho\, r_c^3/3$, and we obtain the same relation between the interaction strengths.

[41] T. A. Davis and W. W. Hager. Modifying a sparse Cholesky factorization. *SIAM Journal on Matrix Analysis and Applications*, 20(3):606, 1999.

[42] T. A. Davis. User guide for CHOLMOD: a sparse Cholesky factorization and modification package. http://www.suitesparse.com.

[43] However, the graphs disconnect with a finite probability in the special case $n_c = 1$: For example, in the random case we have $p_c^{\rm ran} = 0.2756 \pm 0.00059$, 0.19864 ± 0.00057, $0.14822 \pm 0.00045, 0.10401 \pm 0.00041$ for $N = 128, 256, 512, 1024$, respectively.

[44] C. Pozrikidis. *An Introduction to Grids, Graphs, and Networks*. Oxford University Press, 2014.

[45] N. Jacobson. *Basic Algebra*, volume I. Freeman, 1985.