SUPPORTING INFORMATION

Statistical mechanics for natural flocks of birds

W Bialek, A Cavagna, I Giardina, T Mora, E Silvestri, M Viale and AM Walczak

I. MAXIMUM ENTROPY APPROACH

The maximum entropy method [1] has a long history. Recent developments in experimental methods have renewed interest in this idea as a path for constructing statistical mechanics models of biological systems directly from real data, with examples drawn from networks of neurons [2–7], ensembles of amino acid sequences [8–11], biochemical and genetic networks [12, 13], and the statistics of letters in words [14]. Here we give a review of the basic ideas leading to Eq (1) of the main text, hoping to make the discussion accessible to a wider readership.

Imagine a system whose state at any one instant of time is described by a set of variables $\{x_1, x_2, \cdots, x_N\} \equiv \mathbf{x}$. For the moment we don't need to specify the nature of these variables—they could be positions or velocities of individual birds $\mathrm{i} = 1, 2, \cdots N$ in a flock, or more subtle parameters of body shape or instantaneous posture. Whatever our choice of variables, we know that when the number of elements in the system N (here, the number of birds in the flock) becomes large, the space $\mathbf{x}$ becomes exponentially larger. Thus there is no sense in which we can "measure" the distribution of states taken on by the system, because the number of possibilities is just too large. On the other hand, we can obtain reliable measurements of certain average quantities that are related to the state $\mathbf{x}$. To give a familiar example, we can't measure the velocity of every electron in a piece of wire, but certainly we can measure the average current that flows through the wire. Formally, there can be several such functions, $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, $\cdots$, $f_K(\mathbf{x})$, of the state $\mathbf{x}$. The minimally structured distribution for these data is the most random distribution $P(\mathbf{x})$ that is consistent with the observed averages of these functions $\{\langle f_\nu(\mathbf{x})\rangle_{\rm exp}\}$, where $\langle\cdots\rangle_{\rm exp}$ denotes an average measured experimentally.

To find the "most random" distribution, we need a measure of randomness. Another way to say this is that we want the distribution $P(\mathbf{x})$ to hide as much information about $\mathbf{x}$ as possible. One might worry that information and randomness are qualitative concepts, so that there would be many ways to implement this idea. In fact, Shannon proved that there is only one measure of randomness or available information that is consistent with certain simple criteria [15, 16], and this is the entropy

$$S[P] = -\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x}) \ . \tag{S1}$$

Thus we want to maximize $S[P]$ subject to the constraint that the expectation values computed with P match the experimentally measured ones, that is

$$\langle f_\mu(\mathbf{x})\rangle_{\rm exp} = \langle f_\mu(\mathbf{x})\rangle_P \equiv \sum_{\mathbf{x}} P(\mathbf{x}) f_\mu(\mathbf{x}) \tag{S2}$$

for all μ [1]. The distribution $P(\mathbf{x})$ must also be normalized, and it is convenient to think of this as the statement that the average of the "function" $f_0(\mathbf{x}) = 1$ must equal the "experimental" value of 1. Our constrained optimization problem can be solved using the method of the Lagrange multipliers [17]: we introduce a generalized entropy function,

$$\mathcal{S}[P; \{\lambda_\nu\}] = S[P] - \sum_{\mu=0}^{K} \lambda_\mu \left[\langle f_\mu(\mathbf{x})\rangle_P - \langle f_\mu(\mathbf{x})\rangle_{\rm exp}\right] \ , \tag{S3}$$

where a multiplier λ_μ appears for each constraint to be satisfied, and then we maximize $\mathcal{S}$ with respect to the probability distribution $P(\mathbf{x})$ and optimize it with respect to the parameters $\{\lambda_\nu\}$.

Maximizing with respect to $P(\mathbf{x})$ give us

$$\begin{aligned} 0 &= \frac{\partial \mathcal{S}[P; \{\lambda_\nu\}]}{\partial P(\mathbf{x})} \\ &= \frac{\partial S[P]}{\partial P(\mathbf{x})} - \sum_{\mu=0}^{K} \lambda_\mu \frac{\partial \langle f_\mu(\mathbf{x})\rangle_P}{\partial P(\mathbf{x})} \\ &= -\ln P(\mathbf{x}) - 1 - \sum_{\mu=0}^{K} \lambda_\mu f_\mu(\mathbf{x}), \end{aligned} \tag{S4}$$

$$\Rightarrow P(\mathbf{x}) = \frac{1}{Z(\{\lambda_\nu\})} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x})\right], \tag{S5}$$

where $Z(\{\lambda_\nu\}) = \exp(-\lambda_0 - 1)$. Since optimizing with respect to λ_0 will enforce normalization of the distribution, we can write, explicitly,

$$Z(\{\lambda_\nu\}) = \sum_{\mathbf{x}} \exp\left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x})\right]. \tag{S6}$$

Optimizing with respect to $\{\lambda_\nu\}$ gives us a set of K simultaneous equations

$$\begin{aligned} 0 &= \frac{\partial \mathcal{S}[P; \{\lambda_\nu\}]}{\partial \lambda_\mu} \\ &= \langle f_\mu(\mathbf{x})\rangle_{\rm exp} - \langle f_\mu(\mathbf{x})\rangle_P \end{aligned}$$

$$\Rightarrow \langle f_\mu(\mathbf{x})\rangle_{\rm exp} \;=\; \frac{1}{Z(\{\lambda_\nu\})}\sum_{\mathbf{x}} f_\mu(\mathbf{x}) \exp\left[-\sum_{\mu=1}^{K}\lambda_\mu f_\mu(\mathbf{x})\right] . \tag{S7}$$

Thus, when we optimize $\mathcal{S}$ with respect to the parameters $\{\lambda_\nu\}$ we are enforcing that the expectation values of the $\{f_\mu(\mathbf{x})\}$ agree with their experimental values, which is the starting point of the maximum entropy construction. Note also that, if we substitute Eq (S5) back into Eq (S3), we obtain

$$\begin{aligned}\mathcal{S}\left[P;\{\lambda_\nu\}\right] &= \ln Z(\{\lambda_\nu\}) + \sum_{\mu=0}^{K}\lambda_\mu \langle f_\mu(\mathbf{x})\rangle_{\rm exp} \\ &= -\langle \log P(\mathbf{x})\rangle_{\rm exp}, \end{aligned}\tag{S8}$$

which is minus the log probability, or likelihood, that the model generates the observed data. The optimal values of $\{\lambda_\nu\}$ correspond to minima of $\mathcal{S}$, as can be checked by considering the second derivatives. Therefore, the maximum entropy approach also corresponds to *maximizing* the likelihood that the model in Eq (S6) generates the observed data.

The maximum entropy distributions are familiar from statistical mechanics. Indeed we recall that a system in thermal equilibrium is described by a probability distribution that has the maximum possible entropy consistent with its average energy. If the system has states described by a variable $\mathbf{x}$, and each state has an energy $E(\mathbf{x})$, then this equilibrium distribution is

$$P(\mathbf{x}) = \frac{1}{Z(\beta)} e^{-\beta E(\mathbf{x})}, \tag{S9}$$

where $\beta = 1/k_B T$ is the inverse temperature, and the partition function $Z(\beta)$ normalizes the distribution,

$$Z(\beta) = \sum_{\mathbf{x}} e^{-\beta E(\mathbf{x})}. \tag{S10}$$

In this view, the temperature is just a parameter we have to adjust so that the average value of the energy agrees with experiment. The fact that equilibrium statistical mechanics is the prototype of maximum entropy models encourages us to think that the maximum entropy construction defines an effective "energy" for the system. Comparing Eq's (S5) and (S9) gives us

$$E(\mathbf{x}) = \sum_{\mu=1}^{K}\lambda_\mu f_\mu(\mathbf{x}), \tag{S11}$$

and an effective temperature $k_B T = 1$. This is a mathematical equivalence, not an analogy, and means that we can carry over our intuition from decades of theoretical work on statistical physics.

In this paper, we discuss the case where the pairwise correlations $\langle \vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\rangle$ are measured experimentally. Thus we can use the general maximum entropy formulation, identifying $\mathbf{x} = \{\vec{s}_i\}$ and $f_\mu(\mathbf{x}) = \vec{s}_{\rm i}\cdot\vec{s}_{\rm j}$. Since the quantities that will be measured refer to pairs, it is useful to set $\lambda_\mu = -J_{\rm ij}$, and we obtain Eq (1) of the main text, i.e.

$$P(\{\vec{s}_{\rm i}\}) = \frac{1}{Z(\{J_{\rm ij}\})}\exp\left[\frac{1}{2}\sum_{\rm i=1}^{N}\sum_{\rm j=1}^{N} J_{\rm ij}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\right]. \tag{S12}$$

As before, the parameters $\{J_{\rm ij}\}$ must be adjusted so that $\langle \vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\rangle_P = \langle \vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\rangle_{\rm exp}$.

The model defined by Eq (S12) is identical to a well known model for magnetism, the Heisenberg model. In that case, the model describes individual spins, which tend to mutually align according to the interactions $J_{\rm ij}$. In this context, the effective energy is

$$E(\{\vec{s}_{\rm i}\}) = -\frac{1}{2}\sum_{\rm i=1}^{N}\sum_{\rm j=1}^{N} J_{\rm ij}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}. \tag{S13}$$

For $J_{\rm ij} > 0$, the energy is lowered when the vectors $\vec{s}_{\rm i}$ and $\vec{s}_{\rm j}$ are parallel.

Another case, which is relevant for our analysis, is the one where the function measured experimentally is not the full pairwise correlation matrix, but a restricted local measure of correlation. We can for example consider the average correlation among pairs within a neighborhood of size n_c

$$C_{\rm int} = \frac{1}{N}\sum_{\rm i}\frac{1}{n_c}\sum_{\rm j\in n_c^i}\langle \vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\rangle \tag{S14}$$

The maximum entropy model consistent with $C_{\rm int}$ can be found setting $\mathbf{x} = \{\vec{s}_i\}$ and $f_\mu(\mathbf{x}) = (1/Nn_c)\sum_{\rm i}\sum_{\rm j\in n_c^i}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}$. Since the measured quantity is a single scalar, $C_{\rm int}$, there is only one Lagrange multiplier, $\lambda = -J$. In this way, we immediately get Eq (4) of the main text, i.e.

$$P(\{\vec{s}_{\rm i}\}) = \frac{1}{Z(J,n_c)}\exp\left[\frac{J}{2}\sum_{\rm i=1}^{N}\sum_{\rm j\in n_c^i}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\right], \tag{S15}$$

As before, for any given value of n_c, the parameter J must be adjusted so that the expected value of $C_{\rm int}$ computed with the distribution (S15) be equal to the experimentally measured one.

II. THE SPIN WAVE APPROXIMATION

The most demanding step in evaluating the probability distribution in Eq (S12) is the computation of the partition function

$$Z(\{J_{\rm ij}\}) = \int d^N\vec{\boldsymbol{s}}\;\exp\left[\frac{1}{2}\sum_{\rm i=1}^{N}\sum_{\rm j=1}^{N} J_{\rm ij}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\right], \tag{S16}$$

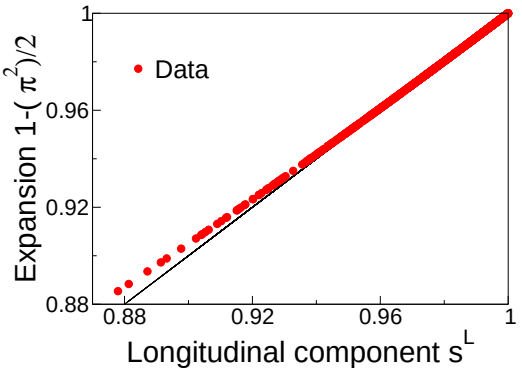

FIG. S1: Longitudinal components of the flight directions vs. prediction of the spin wave expansion, for all individuals in the snapshot of Fig 1 of the main text. The black line has slope 1. Note that 95% of the birds have $s_{\rm i}^{\rm L} > 0.94$, and lie well on the line.

where we recall that the $\{\vec{s}_i\}$ are real, three dimensional vectors of unit length and $d^N\vec{\boldsymbol{s}} = \prod_{\rm i} d\vec{s}_{\rm i}$.

In presence of strong ordering, we can use the "spin wave" approximation [18] to compute analytically the partition function of the Heisenberg model, Eq (S16). Let us call $\vec{S} = (1/N)\sum_i \vec{s}_{\rm i} = S\vec{n}$ the global order parameter, or polarization, measuring the degree of collective alignment, where $\vec{n}$ is a unit vector. Individual orientations can be rewritten in terms of a longitudinal and a perpendicular component with respect to $\vec{n}$,

$$\vec{s}_{\rm i} = s_{\rm i}^{\rm L}\vec{n} + \vec{\pi}_{\rm i} \ , \tag{S17}$$

where, by construction, $\sum_{\rm i} s_{\rm i}^{\rm L} = SN$, $\vec{\pi}_{\rm i}\cdot\vec{n} = 0$, and $\sum_{\rm i}\vec{\pi}_{\rm i} = 0$. The partition function then reads

$$Z(\{J_{\rm ij}\}) = \int d^N\mathbf{s}^{\mathbf{L}}\, d^N\vec{\boldsymbol{\pi}} \ \prod_{\rm i}\delta\left((s_{\rm i}^{\rm L})^2 + |\vec{\pi}_{\rm i}|^2 - 1\right) \delta\left(\sum_{\rm i}\vec{\pi}_{\rm i}\right)\exp\left[\frac{1}{2}\sum_{\rm i=1}^{N}\sum_{\rm j=1}^{N} J_{\rm ij}\left(s_{\rm i}^L s_{\rm j}^L + \vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j}\right)\right] \tag{S18}$$

where $d^N\mathbf{s}^{\mathbf{L}} = \prod_{\rm i} ds_{\rm i}^L$ and $d^N\vec{\boldsymbol{\pi}} = \prod_{\rm i} d\vec{\pi}_{\rm i}$. The delta functions implement the constraint on the length of each vector $\vec{s}_{\rm i}$ and the global constraint on the $\vec{\pi}_{\rm i}$. Note that since the $\{\vec{\pi}_{\rm i}\}$'s belong to the subspace perpendicular to $\vec{n}$, in Eq (S19) there are only two independent degrees of freedom for each integration variable.

If the system is highly polarized, $S \sim 1$ and $|\vec{\pi}_{\rm i}| \ll 1$. The constraint on the norm of the vectors can then be written as $s_{\rm i}^L \sim 1 - |\vec{\pi}_{\rm i}|^2/2$. Note that indeed flocks are very polarized groups (see Table S1) and this expression is very well satisfied by the data, as shown in Fig S1. Using this expansion the longitudinal components can be integrated out easily. The partition function then becomes, to leading order in the $\vec{\pi}$'s,

$$Z(\{J_{\rm ij}\}) = \int d^N\vec{\boldsymbol{\pi}}\ \left[\prod_{\rm i}\frac{1}{\sqrt{1-|\vec{\pi}_{\rm i}|^2}}\right]\ \delta\Big(\sum_{\rm i}\vec{\pi}_{\rm i}\Big) \exp\left[-\frac{1}{2}\sum_{\rm i,j=1}^{N} A_{\rm ij}\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j} + \frac{1}{2}\sum_{\rm i,j=1}^{N} J_{\rm ij}\right] \tag{S19}$$

with

$$A_{\rm ij} = \sum_k J_{\rm ik}\delta_{\rm ij} - J_{\rm ij} \ . \tag{S20}$$

The product over $1/\sqrt{1-|\vec{\pi}_{\rm i}|^2}$ in Eq (S19) is the Jacobian coming from the integration over the $s_{\rm i}{}^{\rm L}$. This term gives rise to sub–leading contributions in the spin wave approximation, and we shall drop it. We have checked in our computations that the corrections due to this term are indeed negligible.

The matrix A is, by construction, a positive semi–definite matrix. We can find eigenvalues $a_{\rm k}$ and eigenvectors $\mathbf{w}^{\rm k}$ as usual through

$$\sum_{\rm j} A_{\rm ij} w_{\rm j}^{\rm k} = a_{\rm k} w_{\rm j}^{\rm k}. \tag{S21}$$

There is one zero eigenvalue, $a_1 = 0$, corresponding to the constant eigenvector $\vec{w}^1 = (1/\sqrt{N}, 1/\sqrt{N}, \cdots, 1/\sqrt{N})$:

$$\sum_{\rm j} A_{\rm ij} w_{\rm i}^1 = \frac{1}{\sqrt{N}}\sum_{\rm j} A_{\rm ij} = 0. \tag{S22}$$

The argument of the delta function in Eq (S19) is related only to the projection of the $\{\vec{\pi}_{\rm i}\}$ onto this zero mode. We note that in a system with translation invariance, the eigenvectors are Fourier modes, or plane waves, and these are called spin waves in the theory of magnetism. The zero eigenmode is related to the spontaneous breaking of symmetry when the flock chooses a consensus direction of flight—all directions $\vec{n}$ are equally probable, a priori, and hence have equal probability or energy, and the zero mode is the remanent of this symmetry; in physics this is the Goldstone mode.

We can now rewrite Eq (S19) in the orthonormal basis defined by $\{\vec{w}^{\rm k}\}$:

$$Z(\{J_{\rm ij}\}) = \int d^N\vec{\boldsymbol{\pi}}'\ \delta\left(\vec{\pi}_1'\right)\exp\left[-\frac{1}{2}\sum_{\rm k=1}^{N} a_{\rm k}|\vec{\pi}_{\rm k}'|^2 + \frac{1}{2}\sum_{\rm i,j=1}^{N} J_{\rm ij}\right], \tag{S23}$$

where $\vec{\pi}_k' = \sum_{\rm i} w_{\rm i}^k\vec{\pi}_{\rm i}$. Remembering that $\vec{\pi}$ is a two-dimensional vector, this leads to

$$\log Z(\{J_{\rm ij}\}) = -\sum_{\rm k>1}\log(a_{\rm k}) + \frac{1}{2}\sum_{\rm i,j=1}^{N} J_{\rm ij} \ , \tag{S24}$$

where we drop constant terms independent of J_{ij}.

Let us now proceed, at a formal level, with the maximum entropy approach. The parameters J_{ij} are fixed by requiring that $\langle \vec{s}_i \cdot \vec{s}_j \rangle_P = \langle \vec{s}_i \cdot \vec{s}_j \rangle_{exp}$. If we focus on the perpendicular part of the correlation, this implies

$$\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle_{exp} = 2 \sum_{k>1} \frac{w_i^k w_j^k}{a_k} \ , \tag{S25}$$

where the right hand side, the expectation value $\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle_P$, can be obtained from Eq (S23) using Gaussian integration rules, the factor 2 coming from the two independent degrees of freedom of each $\vec{\pi}_i$. According to this equation, the matrix A_{ij}—and therefore the interaction matrix J_{ij}—is easily obtained by taking the inverse of the experimental perpendicular correlation function (once we take away the zero mode due to symmetry). But, to be invertible, the experimental correlation matrix must have $N-1$ nonzero eigenvalues. This can only be achieved by performing a huge number of experiments, i.e. evaluating the experimental average over a number of independent samples larger than the number of birds in the flock. As discussed in the main text, the interaction network in a flock changes continuously in time, since individuals move and change their neighbors. But the average over many independent realizations of $\langle \vec{s}_i \cdot \vec{s}_j \rangle$ would require birds to stay still at some fixed positions, while updating and realigning their velocities, which is definitely not the case. In other terms, different experimental samples (i.e. snapshots) correspond to different networks J_{ij} and cannot be averaged together. Thus, in our case, the maximum entropy model must be solved independently at each time step, for which we have only one experimental sample. Unfortunately, if we compute the correlation *matrix* from a single snapshot, it has rank two and cannot be inverted. In other words, a single sample does not provide us with a reasonable experimental estimate of the entire correlation matrix. This motivates, as discussed in the text, the analysis of a more restricted problem, where we consider the average local correlations C_{int} defined in Eq (S14). Indeed we note that, in large flocks, due to law of large numbers, we have

$$C_{int} = \frac{1}{N} \sum_i \frac{1}{n_c} \sum_{j \in n_c^i} \langle \vec{s}_i \cdot \vec{s}_j \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j \in n_c^i} \vec{s}_i \cdot \vec{s}_j \ . \tag{S26}$$

In other terms, since C_{int} is a spatial average of a local quantity (the correlation of a given bird with its interacting neighbors), it can be estimated also from a single snapshot.

III. COMPUTATION WITH FREE BOUNDARIES

Let us now address more in details the reduced model (S15) (Eq (4) in main text), where each individual interacts with constant strength with its first n_c neighbors. This model can be seen as a specific case of Eq (S12), where the J_{ij}'s have a particularly simple form:

$$J_{ij} = J \ n_{ij} \tag{S27}$$

with

$$n_{ij} = \begin{cases} 1 & \text{if } j \in n_c^i \text{ and } i \in n_c^j \, , \\ \frac{1}{2} & \text{if } j \in n_c^i \text{ and } i \notin n_c^j \, , \text{ or vice versa, and} \\ 0 & \text{otherwise.} \end{cases} \tag{S28}$$

Here, J indicates the strength of the interaction and n_c^i indicates the set of the first n_c neighbors of bird i. Since we know the spatial coordinates of all the birds in the flock, once the parameter n_c is fixed, we can compute all the neighborhoods and determine the matrix n_{ij}. In the spin wave expansion (S19) therefore $A_{ij} = J\tilde{A}_{ij}$, where $\tilde{A} = \delta_{ij} \sum_k n_{ik} - n_{ij}$ only depends on the neighborhood relations.

Before proceeding with the full computation with fixed boundary conditions, let us briefly look at the simplest case, where we allow all the $\vec{\pi}_i$'s to freely fluctuate according to Eq (S23). The result can be read directly from Eq (S24), giving

$$\log Z(J, n_c) = -\sum_{k>1} \log(J\lambda_k) + \frac{NJn_c}{2} \ , \tag{S29}$$

where the λ_k are the eignevalues of $\tilde{A}$. Similarly, we can compute the correlation functions,

$$\begin{aligned} \langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle &= \frac{2}{J} \sum_{k>1} \frac{w_i^k w_j^k}{\lambda_k} \\ \langle s_i^L s_j^L \rangle &= 1 - \frac{1}{J} \sum_{k>1} \frac{(w_i^k)^2 + (w_j^k)^2}{\lambda_k} \end{aligned} \tag{S30}$$

where λ_k and $\mathbf{w}^k$ are, again, the eigenvalues and the eigenvectors of the matrix $\tilde{A}$ and depend only on n_c.

To build the maximum entropy model, we need to find the appropriate values for J and n_c. We have shown that the reduced model (S15) is the maximum entropy model consistent with the quantity C_{int}, i.e. the degree of correlation up to the interaction range n_c. The parameter J is therefore fixed by requiring that

$$C_{int}(J, n_c) = C_{int}^{exp} \tag{S31}$$

where $C_{int}(J, n_c)$ indicates the value of C_{int} computed with model (S15) (with given values of J and n_c) and C_{int}^{exp} is the experimental value of C_{int} in a single snapshot (see Eq (S26)). As explained previously, this is mathematically equivalent to maximizing the log–likelihood of the experimental data, given the model. This can be

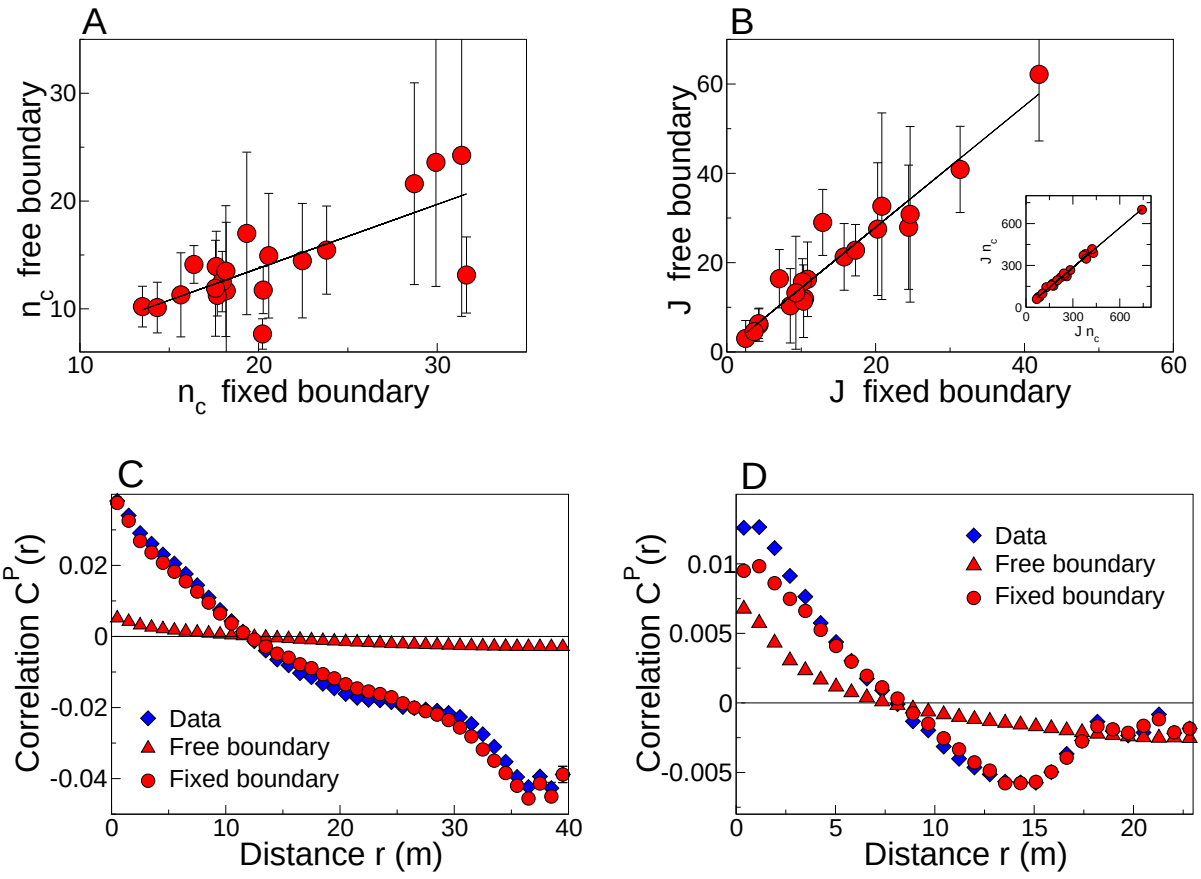

FIG. S2: Computation with free boundary conditions vs computation with fixed flight directions on the border. (A) Values of the parameter n_c for all the flocking events; the black line is the linear regression. Error bars are standard deviations across multiple snapshots of the same flock. (B) Values of the parameter J for all the flocking events. Inset: The product Jn_c computed with free boundary vs. Jn_c computed with fixed boundary; now the slope is almost unity. (C) and (D) Perpendicular correlation as a function of distance for event 28–10 (as in Fig 1; $N = 1246$ birds) and event 32–06 ($N = 809$ birds). Different symbols correspond to the correlation measured in experiments, the correlation computed with free boundary conditions and the one computed with fixed boundary conditions. Taking into account the flight directions of individuals on the border significantly improves the prediction for the correlation.

written simply as

$$\left\langle \log P\left(\{\vec{s}_i\}\right) \right\rangle_{\rm exp} = -\log Z(J, n_c) + \frac{1}{2} JNn_c C_{\rm int}^{\rm exp} \ , \tag{S32}$$

Maximizing Eq (S32) with respect to J (or, equivalently, solving Eq (S31)) gives

$$\frac{1}{J} = \frac{n_c}{2}\left(1 - C_{\rm int}^{\rm exp}\right) \ . \tag{S33}$$

This equation provides an explicit expression of J as a function of n_c. At this point, we are only left with one parameter to be fixed. To select the n_c that best explains the data, we use the principle of maximum likelihood and maximize the likelihood (S32) also with respect to n_c. Substituting $J(n_c)$ into Eq (S32), the likelihood becomes a function of n_c only, and its maximum can be found numerically.

We have applied this procedure to our entire data-set: for a given flock at a given instant of time, we have computed the correlation $C_{\rm int}^{\rm exp}$ from the data and calculated J and n_c with the above free boundary computation. Then, we calculated mean and standard deviation of the interaction parameters for each flock over time. The obtained values of J and n_c are displayed in Fig S2, for all the flocking events we analyzed. They are strongly correlated to what we find with fixed boundary conditions (see next section): the value of n_c is slightly smaller, the value of J slightly larger, but the product Jn_c is approximately the same. On the contrary, the prediction for the perpendicular correlation as a function of distance (Fig S2, panels C and D) is less satisfactory: while the correlation length is correctly reproduced, the decay of the correlation with distance is significantly faster. Besides, the value of the perpendicular correlation near

$r = 0$ looks much smaller than the experimental value. To better understand this point we note that

$$C_{\rm int} = C_{\rm int}^{\rm P} + S + \left(1 - \frac{1}{N}\sum_{\rm i}\frac{1}{n_c}\sum_{\rm j\in n_c{}^i}\langle|\vec{\pi}_{\rm j}|^2\rangle\right) \ , \quad \text{(S34)}$$

where, we recall, S is the polarization. The first term in this decomposition of $C_{\rm int}$ represents the perpendicular part of the correlation up to scale n_c, while the last term is a 'local' polarization getting contributions only from individuals on a scale n_c. The maximum entropy model, by construction, reproduces correctly the experimental value of $C_{\rm int}$. What happens in the computation with free boundaries is that the model underestimates the contribution on short scales ($n < n_c$, corresponding to spatial scales of a few meters) from the perpendicular part of the correlation, and compensates by overestimating the polarization. The effect is more or less strong in different flocks, as seen in Fig S2, panels C and D.

As discussed in the main text, there are good reasons to think that birds on the edge of the flock should be described differently from those in the bulk; Fig S2-C is evidence that if we ignore these differences we really do fail to predict correctly the correlation structure of the flock as a whole.

IV. COMPUTATION WITH FIXED BOUNDARIES

To improve our approach, we need to consider more appropriate boundary conditions. As discussed in the main text, birds on the border of the flock are likely to behave differently from birds in the interior of the flock. This occurs because they experience a different kind of neighborhood, part of the space around them being devoid of neighbors. Besides, these birds are continuously exposed to external stimuli and their dynamics may be strongly influenced by environmental factors (approaching predators, obstacles, nearby roosts, ...). Thus, modelling birds on the border might require taking into account other ingredients than the interactions between individuals. Rather than trying to making a model of these (largely unknown) factors, we can take the velocities of these border birds as given, and ask that our model of interactions predict the propagation of order throughout the bulk of the flock.

If we consider the flight directions of birds on the border as given, the computation of the partition function becomes more complicated. The starting point is analogous to Eq (S16), but integration must be performed with respect to internal variables only. It is then convenient to separate, in the exponent of Eq (S16), contributions coming from internal and external birds. Let us call $\mathcal{I}$ and $\mathcal{B}$ the subsets of internal and border individuals, respectively. Then, in the spin wave approximation, we find an expression similar to Eq (S19):

$$Z(\{J_{\rm ij}\};\mathcal{B}) = \int d^{\mathcal{I}}\vec{\boldsymbol{\pi}}\ \delta\left(\sum_{\rm i}\vec{\pi}_{\rm i}\right)\exp\left[-\frac{1}{2}\sum_{\rm i,j\in\mathcal{I}}A_{\rm ij}\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j} + \sum_{\rm i\in\mathcal{I}}\vec{\pi}_{\rm i}\cdot\vec{h}_{\rm i} + \frac{1}{2}\sum_{\rm i,j\in\mathcal{I}}J_{\rm ij} + \frac{1}{2}\sum_{\rm i\in\mathcal{I}}h_{\rm i}^{\rm L} + \frac{1}{2}\sum_{\rm i,j\in\mathcal{B}}J_{\rm ij}\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\right], \quad \text{(S35)}$$

where

$$\vec{h}_{\rm i} \ = \ \sum_{\rm l\in\mathcal{B}}J_{\rm il}\vec{s}_{\rm l} = \sum_{\rm l\in\mathcal{B}}J_{\rm il}\left(s_{\rm l}^{\rm L}\vec{n} + \vec{\pi}_{\rm l}\right) = h_{\rm i}^{\rm L}\vec{n} + \vec{h}_{\rm i}^{\rm P} \quad \text{(S36)}$$

$$A_{\rm ij} \ = \ \delta_{\rm ij}\left(\sum_{k\in\mathcal{I}}J_{\rm ik} + h_{\rm i}^{L}\right) - J_{\rm ij} \qquad {\rm i,j}\in\mathcal{I} \quad \text{(S37)}$$

Here $\vec{h}_{\rm i}$ is a 'field' describing the influence of birds on the border on internal bird i. The effect of this field is to align bird i with the border birds that are within its direct interaction neighborhood $n_c^{\rm i}$. Thus, when n_c is small, this field only acts on individuals close to the border, while it is zero well inside the flock. We also note that, as compared to Eq (S19), the matrix A is now defined for internal birds only and gets an additional diagonal contribution coming from individuals on the border. As a result, A no longer has a zero mode. From a conceptual point of view, when we fix the direction of motion of birds on the border, not all directions in the bulk are a priori equivalent; rather, the boundary conditions explicitly break the symmetry. From a computational point of view, this implies that we cannot express in a simple way the constraint on the $\{\vec{\pi}_{\rm i}\}$'s as we did in the case of a free boundary.

To deal with the constraint, it is convenient to use an integral representation of the delta function

$$\delta\left(\sum_{\rm i}\vec{\pi}_{\rm i}\right) = \int\frac{d\vec{z}}{(2\pi)^2}\exp\left[i\vec{z}\cdot\sum_{\rm i}\vec{\pi}_{\rm i}\right]. \quad \text{(S38)}$$

Substituting into Eq (S35), we obtain

$$Z(\{J_{\rm ij}\};\mathcal{B}) = \int\frac{d\vec{z}}{(2\pi)^2}\int d^{\mathcal{I}}\vec{\boldsymbol{\pi}}\ \ \exp\left[-\frac{1}{2}\sum_{\rm i,j\in\mathcal{I}}A_{\rm ij}\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j} + \sum_{\rm i\in\mathcal{I}}\vec{\pi}_{\rm i}\cdot\left(\vec{h}_{\rm i}^{\rm P} + i\vec{z}\right) + i\vec{z}\cdot\sum_{\rm l\in\mathcal{B}}\vec{\pi}_{\rm l} + G(\mathcal{B})\right], \quad \text{(S39)}$$

where $G(\mathcal{B})$ is a function of boundary variables only. We notice that all the integrals are Gaussian, and we obtain, finally,

$$\ln Z\left(\{J_{\rm ij}\};\mathcal{B}\right) = \frac{1}{2}\sum_{\rm ij\in\mathcal{I}}(A^{-1})_{\rm ij}\ \vec{h}_{\rm i}^{\rm P}\cdot\vec{h}_{\rm j}^{\rm P} - \ln\det(A) - \ln\left[\sum_{\rm ij\in\mathcal{I}}(A^{-1})_{\rm ij}\right] - \frac{1}{2}\frac{\left[\sum_{\rm l\in\mathcal{B}}\vec{\pi}_{\rm l} + \sum_{\rm ij\in\mathcal{I}}(A^{-1})_{\rm ij}\vec{h}_{\rm j}^{\rm P}\right]^2}{\sum_{\rm ij\in\mathcal{I}}(A^{-1})_{\rm ij}}$$

$$+\frac{1}{2}\sum_{\mathrm{ij}\in\mathcal{I}} J_{\mathrm{ij}} + \sum_{\mathrm{i}\in\mathcal{I}} h_{\mathrm{i}}^{\mathrm{L}} + \sum_{\mathrm{lm}\in\mathcal{B}} J_{\mathrm{lm}}\vec{s}_{\mathrm{l}}\cdot\vec{s}_{\mathrm{m}}\ , \quad \text{(S40)}$$

where $G(\mathcal{B})$ is written explicitly. Recall that the matrix A is only defined on internal individuals and hence the number of eigenvalues that contribute to the computation of $\det(A)$ is given by the number of internal birds. In the same way, we can easily compute correlation functions. We find

$$\langle\vec{\pi}_{\mathrm{i}}\rangle = \sum_{\mathrm{j}\in\mathcal{I}}(A^{-1})_{\mathrm{ij}}\vec{h}_{\mathrm{i}}^{\mathrm{P}} - \frac{\sum_{\mathrm{j}\in\mathcal{I}}(A^{-1})_{\mathrm{ij}}}{\sum_{\mathrm{kj}\in\mathcal{I}}(A^{-1})_{\mathrm{kj}}}\left[\sum_{\mathrm{l}\in\mathcal{B}}\vec{\pi}_{\mathrm{l}} + \sum_{\mathrm{kj}\in\mathcal{I}}(A^{-1})_{\mathrm{kj}}\vec{h}_{\mathrm{j}}^{\mathrm{P}}\right], \quad \text{(S41)}$$

and

$$\langle\vec{\pi}_{\mathrm{i}}\cdot\vec{\pi}_{\mathrm{j}}\rangle = \langle\vec{\pi}_{\mathrm{i}}\rangle\cdot\langle\vec{\pi}_{\mathrm{j}}\rangle + 2\left[(A^{-1})_{\mathrm{ij}} - \frac{\sum_{\mathrm{kn}\in\mathcal{I}}(A^{-1})_{\mathrm{ik}}(A^{-1})_{\mathrm{nj}}}{\sum_{\mathrm{kn}\in\mathcal{I}}(A^{-1})_{\mathrm{kn}}}\right]. \quad \text{(S42)}$$

At this point, to solve the maximum entropy model for the reduced case, we simply substitute the parametrization $J_{\mathrm{ij}} = Jn_{\mathrm{ij}}$. The log–likelihood takes the form

$$\left\langle \log P\left(\{\vec{s}_{\mathrm{i}}\}\right)\right\rangle_{\mathrm{exp}} = -\log Z(J, n_c; \mathcal{B}) + \frac{1}{2}Jn_cNC_{\mathrm{int}}^{\mathrm{exp}}\ . \quad \text{(S43)}$$

with $Z(J, n_c; \mathcal{B})$ as in Eq (S35). To find the optimal value for the parameters J and n_c we need to maximize the likelihood. Maximization with respect to J again is equivalent to matching the predicted correlations to the experimental ones, $C_{\mathrm{int}}(J, n_c; \mathcal{B}) = C_{\mathrm{int}}^{\mathrm{exp}}$. This equation is represented graphically in Fig 2A in the main text. It is worth noting that, as in the case with free boundary conditions, it is possible to solve this equation analytically. We can define

$$\tilde{A} = A/J, \quad \text{(S44)}$$

$$\tilde{\vec{h}}_{\mathrm{i}} = \vec{h}_{\mathrm{i}}/J, \quad \text{(S45)}$$

both of which are independent of J, and then, after some algebra, we obtain

$$\frac{(N_{\mathrm{in}}-1)}{J} = \frac{1}{2}\sum_{\mathrm{ij}\in\mathcal{I}}(\tilde{A}^{-1})_{\mathrm{ij}}\tilde{\vec{h}}_{\mathrm{i}}^{\mathrm{P}}\cdot\tilde{\vec{h}}_{\mathrm{j}}^{\mathrm{P}} + \sum_{\mathrm{i}\in\mathcal{I}}\tilde{h}_{\mathrm{i}}^{\mathrm{L}} - \frac{1}{2}\frac{\left[\sum_{\mathrm{l}\in\mathcal{B}}\vec{\pi}_{\mathrm{l}} + \sum_{\mathrm{ij}\in\mathcal{I}}(\tilde{A}^{-1})_{\mathrm{ij}}\tilde{\vec{h}}_{\mathrm{j}}^{\mathrm{P}}\right]^2}{\sum_{\mathrm{ij}\in\mathcal{I}}(\tilde{A}^{-1})_{\mathrm{ij}}} + \sum_{\mathrm{lm}\in\mathcal{B}} n_{\mathrm{lm}}\vec{s}_{\mathrm{l}}\cdot\vec{s}_{\mathrm{m}} + \frac{Nn_c}{2}(1 - C_{\mathrm{int}}^{\mathrm{exp}}) \quad \text{(S46)}$$

where N_{in} is the number of internal birds. Note that the right hand side is a function of n_c only, so we have an expression for $J(n_c; \mathcal{B})$. Substituting back into Eq (S43) we get the likelihood as a function of n_c only. Maximization can be performed numerically, as shown in Fig 2B in the main text.

Values of J and n_c for all flocks are collected in Fig 2 in the main text and in Fig S2. In this figure, we see the improvement in the prediction of the correlation function $C(r)$ that comes with fixed boundary conditions.

V. A GLOBAL MODEL

Given a flock of birds, so far we have solved the maximum entropy model for each individual snapshot independently, and then we have averaged the inferred values of the parameters J and n_c over all the snapshots. This is the most general procedure we can use, consistent with the dynamical nature of the interaction network. The inferred values of J and n_c fluctuate from snapshot to snapshot, due to several factors. It is possible that birds slightly adjust interaction strength and range during time, but there are other noisy contributions that might increase the fluctuations. The flocks we analyzed are finite groups, ranging from a few hundreds to a few thousands individuals, and we therefore expect finite size effects. The algorithmic procedure to reconstruct positions and velocities of individual birds in the flock is very efficient but not perfect, and there are fluctuations across snapshots in the number of reconstructed individuals; see Refs [19, 20, 22] for details on the 3D reconstructions. Finally, the log–likelihood can be very flat in the region of the maximum: in this case even small fluctuations can cause the value of the maximum to jump from a value of n_c to another one quite different. Averaging n_c and J over the snapshots, we get rid of these fluctuations. Alternatively, we can assume from the start that, given a flock, there is a unique value of n_c and J through time. In this case, the log–likelihood of each snapshot is a function of the *same* J and n_c and we need to optimize the global likelihood corresponding to all the snapshots, and not each one independently. In other terms, we first compute the average of the log–likelihood over the snapshots at J and n_c fixed, and then we maximize with respect to the two parameters. Note that we are inverting the procedure described in the previous sections, where, on the contrary, we first maximize each individual snapshot with respect to J and n_c and then we take the average over all the snapshots of the optimal parameters. The computation of the average log–likelihood can be easily done starting from the equations for the single snapshot. Let us denote, for future convenience, by

$$\phi_\alpha(J, n_c) = -\log Z(J, n_c; \mathcal{B}_\alpha) + \frac{1}{2}Jn_cNC_{\mathrm{int},\alpha}^{\mathrm{exp}} \quad \text{(S47)}$$

the log–likelihood of the snapshot α with parameters J and n_c (see Eq (S43)). Then, the average log–likelihood

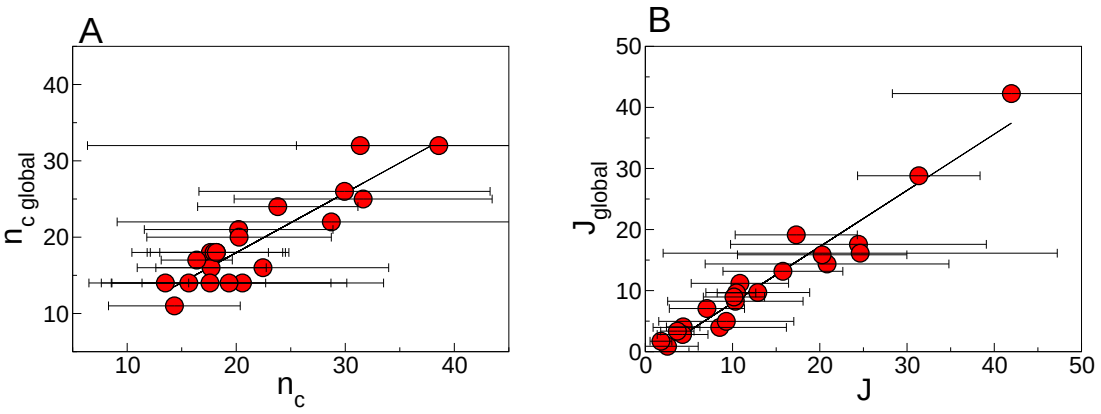

FIG. S3: Global models of flocking events. (A) Values of the neighborhood size n_c inferred from maximizing the log–likelihood averaged over snapshots, plotted vs. the mean values obtained from maximizing log–likelihood in individual snapshots. Error bars represent the standard deviation over snapshots for each flock. Black line has a slope 0.78. (B) As in (A), but for the interaction strength J; the black line has slope 0.92.

for all the snapshots is

$$\Phi_{\rm global}(J, n_c) = \frac{1}{N_{\rm snap}} \sum_\alpha \phi_\alpha(J, n_c) \ , \qquad \text{(S48)}$$

where $N_{\rm snap}$ is the number of snapshots available for that flock. At this point, we need to maximize $\Phi_{\rm global}$ over J and n_c. The maximization with respect to J leads, once again, to an explicit expression for the optimal J, that we shall call $J_{\rm global}$, as a function of n_c:

$$\frac{1}{J_{\rm global}} = \frac{1}{N_{\rm snap}} \sum_\alpha \frac{(N_{\rm in}^\alpha - 1)}{(N_{\rm in}^{\rm global} - 1)} \frac{1}{J_\alpha(n_c)} \qquad \text{(S49)}$$

where $J_\alpha(n_c)$ is the optimal value of J for the snapshot α, as above, $N_{\rm in}^\alpha$ is the number of internal individuals in the snapshot α, and $N_{\rm in}^{\rm global}$ is the corresponding average over snapshots. Substituting $J_{\rm global}(n_c)$ back in Eq (S48) we get an expression, which is a function of n_c only. The likelihood can then be maximized numerically with respect to n_c. The values $n_{c,{\rm global}}$ and $J_{\rm global}$ obtained in this way are plotted in Fig S3, where they are compared to the values inferred with the more general procedure (optimizing each snapshot independently and then averaging). There is a very strong correlation with slope close to one. This represents a strong consistency check on the inference procedure.

The same contributions that increase fluctuations from snapshot to snapshot might also affect some variability in the quality of the maximum entropy model prediction from flock to flock. In this respect, we also note that some flocks have lower polarization than others (see table S1), and that the spin wave approximation is a large polarization expansion working better the larger the polarization. An example of variability across flocks in the quality of the model predictions is shown in Fig S2 C and D, where the predicted correlation is computed for two different flocking events, both with and without fixed border. In Fig S4 we also plot the overlap distribution for three different flocking events. The first two events (blue and turquoise) are the same events as Fig S2. The other one (in blue) corresponds to flocking event 31-01. This flock is larger, the 3D reconstructions of velocities were less good (in terms of percentage of the reconstructed individuals), and the polarization is the lowest among all the flocking events we analyzed. Thus, this is one of those cases where we expect the model to give less good predictions. As it can be seen from the figure, however, the overlap distribution is nevertheless very peaked on 1 (even if less than the other two flocks), indicating that the large majority of velocities have been correctly reproduced.

VI. A MODEL FOR ORDER PROPAGATION

The maximum entropy model with fixed flight directions on the border gives excellent predictions for two–point and higher order correlation functions; see Fig 3 in main text, Fig S2 and Fig S5. In addition, it allows to infer—up to a calibration factor—the microscopic interactions in a numerical model of self–propelled particles. We can conclude that this model indeed offers a very good statistical description of the flight directions of individuals in a flock. Let us then look back at the model, and try to understand the kind of system that the model describes.

We recall that, in this version of the model, we take as fixed the flight directions of the individuals on the border. Therefore, the model does not aim at predicting properties of border individuals, which, as we noted, may depend on factors other than mutual interactions.

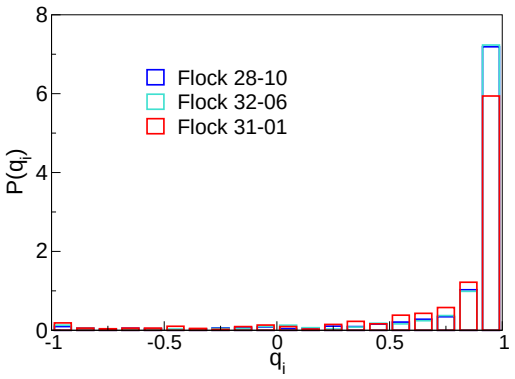

FIG. S4: Variability across flocks. Overlap distribution for three flocking events (one snapshot per event). The overlap $q_i = \langle\vec{\pi}_i\rangle \cdot \vec{\pi}_i^{\rm exp}/(|\langle\vec{\pi}_i\rangle||\vec{\pi}_i^{\rm exp}|)$ measures the similarity between real and predicted velocities (see main text and Fig 3D). The distributions are computed with a larger number of bins than in Fig 3D, for a better comparison between different flocks.

Rather, the model focuses on internal individuals and how ordering flows through the flock. The state of the birds on the border generate a 'field' ($\vec{h}_i$) on internal individuals, but this field is nonzero only for individuals interacting directly interacting with birds on the boundary (i.e. when $J_{ij} = Jn_{ij} \neq 0$). For the values of n_c retrieved by the model ($n_c \sim 20$), this is only a small shell close to the border: all individuals well inside the flock, on the contrary, do not experience any direct influence from the border.

Still, if the model does describe what happens in a real flock, it must predict collective coherence: *all* flight directions must strongly align and internal individuals must behave very much in unison with their exterior companions. Does the model reproduce this behaviour? If so, what is the mechanism leading to this kind of ordering? How do individuals on the border transmit information about their flight directions to distant individuals with whom they do not interact directly?

The formal answer to these questions can be read in Eq's (S41) and (S42). The first equation indicates that the model predicts a well defined perpendicular component of the flight direction $\langle\vec{\pi}_i\rangle$ for each internal individual i. Surprisingly, these perpendicular components agree remarkably well with the ones measured experimentally (see Fig 3D in main text), not only for birds close to boundary, but also well inside the group. The second equation provides a prediction for the correlation function. Visualization of these correlations as a function of distance shows that these predictions also are very good. We note that, since the longitudinal component of the flight direction is given by $\langle s_i^{\rm L}\rangle = 1 - 0.5\langle|\vec{\pi}_i|^2\rangle$, if we are getting the perpendicular components of the velocity right we must also be getting the longitudinal components right. Equations (S41) and (S42) therefore provide correct predictions of the full flight directions for all individuals in the flock.

The mechanism through which such ordering occurs, is the presence of long ranged correlations in the system. This can be seen more easily rewriting the equations in the following way:

$$\langle\vec{\pi}_i\rangle = \sum_{j\in\mathcal{I}} C_{ij}^{\rm con}\vec{h}_j^{\rm P} - \frac{\sum_{j\in\mathcal{I}}(A^{-1})_{ij}}{\sum_{kj\in\mathcal{I}}(A^{-1})_{kj}}\sum_{l\in\mathcal{B}}\vec{\pi}_l \quad \text{(S50)}$$

$$\langle\vec{\pi}_i\cdot\vec{\pi}_j\rangle = C_{ij}^{\rm con} + \langle\vec{\pi}_i\rangle\cdot\langle\vec{\pi}_j\rangle \quad \text{(S51)}$$

$$C_{ij}^{\rm con} = 2\left[A_{ij}^{-1} - \frac{\sum_{kn\in\mathcal{I}} A_{ik}^{-1}A_{nj}^{-1}}{\sum_{kn\in\mathcal{I}} A_{kn}^{-1}}\right] \quad \text{(S52)}$$

where we have separated the part of the correlation, $C_{ij}^{\rm con}$, which is locally connected (i.e. the covariance).

In Eq (S50) the first term describes a contribution coming from individuals on the border, while the second term is just a renormalization factor to ensure that $\sum_{i\in\mathcal{I}}\langle\vec{\pi}_i\rangle + \sum_{l\in\mathcal{B}}\vec{\pi}_l = 0$. We can see from Eq (S50) that an individual i far from the border can also feel the effect of birds on the border, provided there is a nonzero connected correlation $C_{ij}^{\rm con}$ between i and some individual j close to the border. In other terms, while *direct* mutual alignment occurs only between border individuals and immediate neighbors (for which $\vec{h}_i$ are non zero), *effective* alignment occurs also with internal birds that are indirectly correlated with them (for which $C_{ij}^{\rm con}\vec{h}_j$ are nonzero). If the connected correlations extend over sufficiently long distances, this mechanism ensures propagation of directional information trough the whole flock.

In Fig S5-B we show the behaviour of the connected correlation as a function of distance, for one flocking event. The scale over which this function decays, the correlation length, is of the order of the thickness of the flock (maximum distance between an internal point and the border), showing that $C^{\rm con}$ indeed is long ranged enough to propagate ordering well inside the group. In the inset, we show that the correlation length grows linearly with flock size, for all the flocking events we analyzed. Thus the correlation function is scale free: no matter how large the flock is, the correlation always extends over the whole flock.

VII. CORRELATION FUNCTIONS

In this section we summarize the definitions of all the correlation functions introduced in the paper and we comment on their behaviour.

The pairwise correlation. Let us start by recalling the definition of the pairwise correlation,

$$\langle\vec{s}_i\cdot\vec{s}_j\rangle = \langle s_i^{\rm L}s_j^{\rm L}\rangle + \langle\vec{\pi}_i\cdot\vec{\pi}_j\rangle \quad \text{(S53)}$$

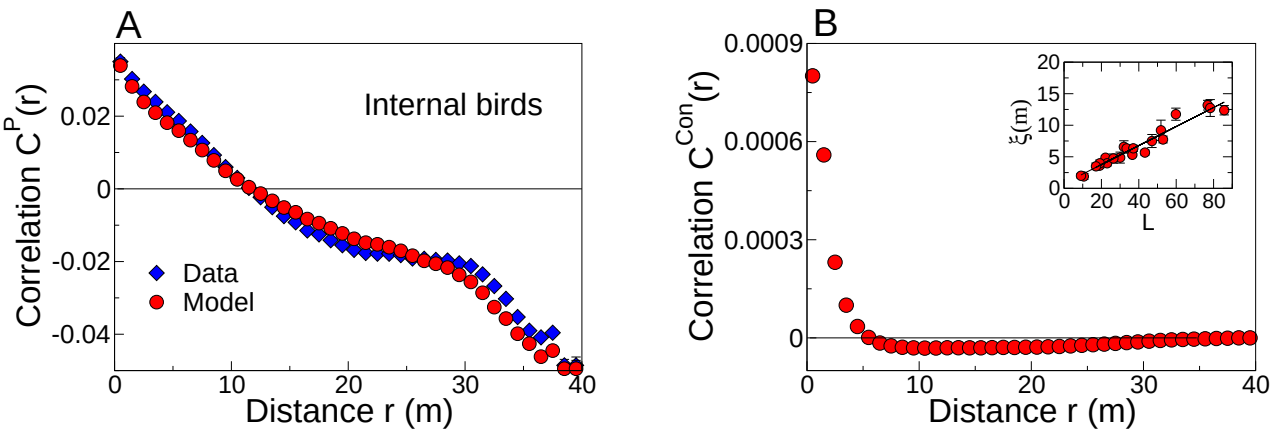

FIG. S5: Correlations in the interior of the flock. (A) Perpendicular component of the two–point correlation function (as in Fig 2A) for internal birds only, as a function of distance. (B) Connected correlation function predicted in the model, as a function of distance. Inset: correlation length vs. flock size, for all the flocks that we analyzed.

where, for future convenience, we have separated the longitudinal part of the correlation from the perpendicular one. We note that while the sample average of the perpendicular flight direction is zero, $(1/N)\sum_{\rm i}\vec{\pi}_{\rm i}=0$, the same is not true for the longitudinal direction. Rather, we have $(1/N)\sum_{\rm i} s^{\rm L}_{\rm i} = S$, and the longitudinal correlation is dominated by a contribution from the global polarization S. To better investigate the degree of correlation in the system, it is then convenient to focus on fluctuations of the individual flight directions with respect to the sample average. To this end, in all the figures in this paper we consider the following correlations, where we have subtracted the sample average contribution:

$$C^{\rm P}_{\rm ij} = \langle \vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j}\rangle \tag{S54}$$

$$\begin{aligned} C^{\rm L}_{\rm ij} &= \langle (s^{\rm L}_{\rm i}-S)(s^{\rm L}_{\rm j}-S)\rangle \\ &= \langle (1-S-\frac{1}{2}\langle\pi^2_{\rm i}\rangle)(1-S-\frac{1}{2}\langle\pi^2_{\rm j}\rangle)\rangle \end{aligned} \tag{S55}$$

$$C_{\rm ij} = C^{\rm P}_{\rm ij} + C^{\rm L}_{\rm ij}\ . \tag{S56}$$

The last identity in Eq (S55) is a consequence of the spin wave approximation.

Connected correlations. In Section IV of the SI we have described a theory where we get nonzero expectation values for the flight directions of individual birds, $\langle\vec{\pi}_{\rm i}\rangle \neq 0$. In this case, it may be useful to look at correlation functions which are *locally* connected, i.e. that describe how the individual bird flight direction fluctuates with respect to its own average value and not—as in the previously defined correlations—with respect to the sample average. To this end, we have introduced in Section VI the following connected correlation function

$$C^{\rm con}_{\rm ij} = \langle\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j}\rangle - \langle\vec{\pi}_{\rm i}\rangle\cdot\langle\vec{\pi}_{\rm j}\rangle\ . \tag{S57}$$

We note that in our case $C^{\rm con}_{\rm ij}$ is purely a theoretical construct. Indeed, we have applied the maximum entropy approach to each single snapshot independently. For a single snapshot, the experimental measurement of the correlation only consists in one configuration (the velocity field at that instant of time) and we cannot distinguish between connected and non–connected correlations. The only quantity that can be compared between theory and experiments is therefore $\langle\vec{\pi}_{\rm i}\cdot\vec{\pi}_{\rm j}\rangle$.

The degree of direct correlation. One important quantity entering our computation is the degree of direct correlation,

$$C_{\rm int} = \frac{1}{N}\sum_{\rm i}\frac{1}{n_c}\sum_{{\rm j}\in{\rm n^i_c}}\langle\vec{s}_{\rm i}\cdot\vec{s}_{\rm j}\rangle, \tag{S58}$$

which measures the average correlation between an individual and its first n_c neighbors. This degree of direct correlation is a single scalar quantity, and represents the input observable used by our maximum entropy approach to build a statistical model for the flight directions.

The two–point correlation function. To describe the behaviour of the correlation at different scales, it is convenient to define the two–point correlation function

$$C(r) = \frac{\sum_{\rm ij} C_{\rm ij}\delta(r_{\rm ij}-r)}{\sum_{\rm ij}\delta(r_{\rm ij}-r)}\ , \tag{S59}$$

where $r_{\rm ij} = |\vec{\mathbf{r}}_{\rm i}-\vec{\mathbf{r}}_{\rm j}|$ is the distance between bird i and bird j and the delta function selects pairs of individuals that have mutual distances in a small interval around r (the denominator representing the number of pairs in such an interval). This function measures the average degree of correlation between individuals separated by a distance r. Again it is possible to distinguish a longitudinal and a perpendicular component of these correlations,

$$C(r) = C^{\rm L}(r) + C^{\rm P}(r), \tag{S60}$$

describing the contributions relative to, respectively, longitudinal and perpendicular fluctuations in the flight directions. Figure 3 in the main text and Fig S5 in the SI show the two–point correlation function computed from the maximum entropy model with fixed boundary conditions. The prediction agrees nicely with the experimental one, on all scales. We stress that the maximum entropy model uses as an input only $C_{\rm int}$, which measures the average degree of correlation up to scale n_c. With the values of n_c retrieved for our events ($n_c = 21.2 \pm 1.7$), this corresponds to a scale of the order of a few meters in r. In contrast, the two–point correlation function measures the correlation on all possible scales, from close neighbors (a few meters) to the whole extension of the flock (hundreds of meters, for some flocks). Therefore, the good agreement with experiments represents a highly nontrivial prediction of the model. From Eq (S59), the correlation function takes into account the contribution coming from all pairs of individuals, independent of whether they reside on the border or in the bulk of the flock. Yet, when adopting fixed flight directions on the border of the flock, the contribution coming from birds on the border is by construction identical in the predicted and observed correlation functions. To test more explicitly whether the model provides good predictions for the correlations of internal individuals, we can consider an *internal* correlation function, defined as in Eq (S59), but where we only count contributions from individuals inside the flock ($\mathrm{i,j} \in \mathcal{I}$); the result is in Fig S5-A. Again, the prediction of the model is nicely consistent with the experimental correlation.

The four–point correlation function. We can define correlation functions not only between pairs of individuals, but for more complicated arrangements of birds. For example, let us consider a pair of birds i, j separated by a distance r_1, and measure their mutual alignment. We might want to compare this degree of alignment to the one that another pair of birds k, l, also separated from one another by a distance r_1, that are located in another position in the flock.

We can then define the following four–point correlation

$$C_4(r_1; r_2) = \frac{\sum_{\rm ijkl} \langle (\vec{\pi}_{\rm i} \cdot \vec{\pi}_{\rm j})(\vec{\pi}_{\rm i} \cdot \vec{\pi}_{\rm j}) \rangle \Delta_{\rm ijkl}}{\sum_{\rm ijkl} \Delta_{\rm ijkl}}, \qquad \text{(S61)}$$

$$\Delta_{\rm ijkl} = \delta(r_{\rm ij} - r_1)\delta(r_{\rm kl} - r_1)\delta(r_{\rm ij-kl} - r_2) \qquad \text{(S62)}$$

where $r_{\rm ij-kl}$ indicates the distance between the mid-points of the pairs ij and pair kl; see Fig S6. We can plot $C_4(r_1; r_2)$ as a function of the two distances r_1 and r_2. For example, in Fig 3C in the main text, it is shown for a fixed value of r_1 as a function of r_2. We also note that, in the spin wave approximation, the longitudinal correlation $C^{\rm L}$ is nothing else than a particular case of the four–point correlation,

$$C^{\rm L}(r) = 1 - C_4(0; r) - S^2. \qquad \text{(S63)}$$

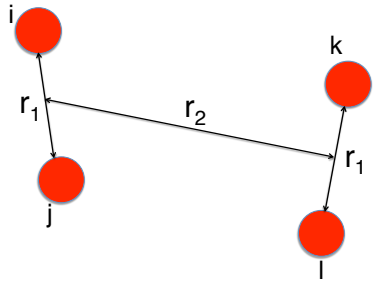

FIG. S6: Sketch of the structure of the four point correlation function. Red circles represent birds. Birds i and j have mutual distance $r_{\rm ij} = r_1$; birds k and l also have mutual distance $r_{\rm kl} = r_1$. The distance between the mid-point of the ij pair and the midpoint of the kl pair is r_2.

VIII. DATA ANALYSIS AND BORDER DEFINITION

Three-dimensional data have been obtained from stereoscopic experiments on large flocks of starlings, during pre-roosting aerial display [19][20][22]. Digital high-resolution stereoscopic images were processed using innovative computer vision techniques [20] and 3D coordinates and velocities reconstructed for individual birds in the flock. A summary of the global properties of the analyzed flocking events can be found in Table S1.

Given a flock, in order to compute global and statistical quantities it is necessary to appropriately define its exterior border. Flocks are typically non-convex systems. Standard methods to define the border, like the convex-hull, are therefore inadequate because they are unable to detect concavities. To overcome this problem, we used the so-called 'α-shape algorithm [23][21]. The main idea of this method is the following: given a set of 3D points, one 'excavates' the set of points with spheres of radius α, so that all concavities of size larger than α are detected. Formally, one selects the sub-complex of the Delaunay triangulation on scale α (the α-complex) and the external surface of this triangulation defines the border. The scale α must be appropriately chosen. If α is too large, some concavities are neglected and void regions are included as being part of the flock. Too small values of α, on the other hand, might cause the sphere to penetrate the flock and break it into sub-connected components. A robust criterion is to look at the density of the internal points as a function of α [21][24]. This quantity typically has a maximum, which defines a natural scale for α.

For all the analyzed flocking events, the border has been computed following the above procedure. We note that, since flocks change shape in time, the border must be computed and re-defined at each instant of time. Be-

TABLE S1: Summary of experimental data.

Event[a]	N	S	v_0 (m/s)	L (m)
17–06	552	0.935	9.4	51.8
21–06	717	0.973	11.8	32.1
25–08	1571	0.962	12.1	59.8
25–10	1047	0.991	12.5	33.5
25–11	1176	0.959	10.2	43.3
28–10	1246	0.982	11.1	36.5
29–03	440	0.963	10.4	37.1
31–01	2126	0.844	6.8	76.8
32–06	809	0.981	9.8	22.2
42–03	431	0.979	10.4	29.9
49–05	797	0.995	13.9	19.2
54–08	4268	0.966	19.1	78.7
57–03	3242	0.978	14.1	85.7
58–06	442	0.984	10.1	23.1
58–07	554	0.977	10.5	19.1
63–05	890	0.978	9.9	52.9
69–09	239	0.985	11.8	17.1
69–10	1129	0.987	11.9	47.3
69–19	803	0.975	13.8	26.4
72–02	122	0.992	13.2	10.6
77–07	186	0.978	9.3	9.1

[a]Flocking events are labelled according to experimental session number and to the position within the session they belong to. The number of birds N is the number of individuals for which we obtained a 3D reconstruction of positions in space. The polarization S is defined in the Methods. The linear size L of the flock is defined as the maximum distance between two birds belonging to the flock. The speed v_0 is that of the centre of mass, i.e. the mean velocity of the group. All values are averaged over several snapshots during the flocking event.

sides, due to the continuous movement of individuals through the group, the individuals belonging to the border change from time to time.

IX. ADDITIONAL NUMERICAL SIMULATIONS

The self-propelled particle model defined in Eqs. (9)(10) has been studied extensively in the literature, in the case where the interacting neighbors are chosen as the Voronoi neighbors [25][26]. From these works we know that the alignment term in eq. (9) is the most relevant one in determining the properties of the velocity fields. The distance dependent attraction-repulsion force, on the other hand, acts predominantly on the structure of the group, fixing the density and preventing collisions on the short scale. As long as the system remains in a flock-like state, i.e. it does not crystalize and diffusion of individuals occurs throughout the group, one would not expect a significant role of this term on the inference procedure described in this paper. Still, one might ask how much the metric dependency of the attraction-repulsion force affects the relationship between real and inferred parameters (Fig 4 in main text).

To investigate this point, we run a few additional simulations with the model of eqs. (9),(10). We considered a number of interacting neighbors of order 13 (corresponding to $\mu = 0.49$), and varied the parameters entering the distance dependent term (all other parameters being fixed as specified in the main text). We recall that this term has the following form:

$$\vec{f}_{\rm ij}(r_{\rm ij} < r_b) = -\infty\, \vec{e}_{\rm ij} \tag{S64}$$

$$\vec{f}_{\rm ij}(r_b < r_{\rm ij} < r_a) = \frac{1}{4} \cdot \frac{r_{\rm ij} - r_e}{r_a - r_e} \vec{e}_{\rm ij} \tag{S65}$$

$$\vec{f}_{\rm ij}(r_a < r_{\rm ij} < r_0) = \vec{e}_{\rm ij}, \tag{S66}$$

where r_b sets the hard core below which repulsion occurs, r_e is an "equilibrium" preferred distance (where the attraction-repulsion force is zero), and $[r_a, r_0]$ defines a region where the force is constant [25][26].

We tried 5 different values of β (the parameter modulating the strength of the attraction-repulsion force), and 4 different values of the set of parameters r_e, r_a, r_b. In all these cases we checked that the simulated flocks had appropriate structure and polarization (i.e. the group had to exhibit internal diffusion, as natural flocks do, and polarization had to be large). The results for these new simulations are shown in Fig S7: here green points are the new ones, red and blue points are as in Fig 4 of the main text. As can be seen from this figure, new points lie on the same lines as old ones, demonstrating that the metric dependency does not affect the value of the proportionality constant between estimated and real parameters.

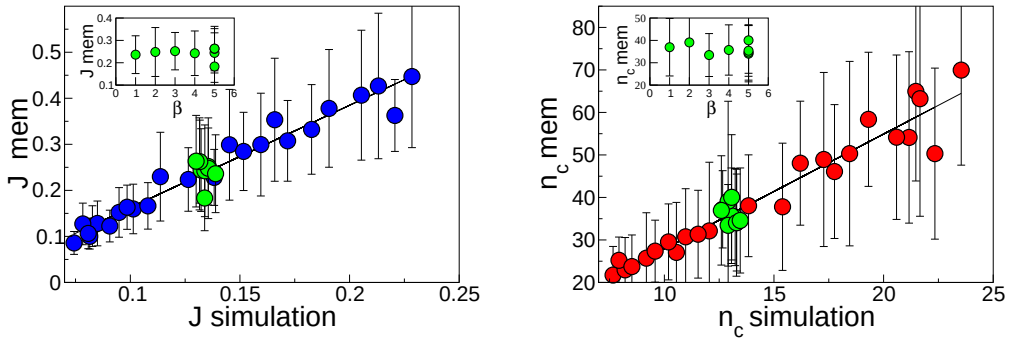

FIG. S7: Maximum entropy analysis for a model of self–propelled particles - Additional simulations. (A) Inferred value of the parameter J vs. microscopic strength of alignment forces used in the simulation. Blue points correspond to $\beta = 5$, $r_0 = 1$, $r_b = 0.2$, $r_e = 0.5$, $r_a = 0.8$, $\alpha = 35$ (and are identical to Fig 4 in main text). Green points correspond to new numerical simulations where we fixed $\mu = 0.46$ (corresponding to $n_c \sim 13$) and varied β, r_a, r_b, r_e. Inset: inferred value of J for the new simulations as a function of β. The points at $\beta = 5$ have $r_e \in [0.4, 0.7]$, $r_a \in [0.64, 1.12]$, $r_b \in [0.8, 1.4]$. No significative dependence of the value of J is observed on any of these parameters. (B) Inferred value of n_c vs. the true number of interacting neighbors in the simulation. Red points correspond to the simulations of Fig 4. Green points correspond to new numerical simulations. Inset: inferred value of n_c as a function of β. All parameters as in (A). Slopes of the lines are 2.2 and 2.7, respectively. Error bars are standard deviations across 45 snapshots of the same simulation.

[1] ET Jaynes, Information theory and statistical mechanics. *Phys Rev* **106,** 620–630 (1957).

[2] E Schneidman, MJ Berry II, R Segev & W Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440,** 1007–1012 (2006).

[3] G Tkačik, E Schneidman, MJ Berry II & W Bialek, Ising models for networks of real neurons. arXiv:q-bio/0611072 (2006).

[4] J Shlens, GD Field, JL Gauthier, MI Grivich, D Petrusca, A Sher, AM Litke & EJ Chichilnisky, The structure of multi–neuron firing patterns in primate retina. *J Neurosci* **26,** 8254–8266 (2006).

[5] A Tang, D Jackson, J Hobbs, W Chen, A Prieto, JL Smith, H Patel, A Sher, A Litke & JM Beggs, A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro. J Neursoci* **28,** 505–518 (2008).

[6] S Yu, D Huang, W Singer & D Nikolić, A small world of neural synchrony. *Cereb Cortex* **18,** 2891–2901 (2008).

[7] G Tkačik, E Schneidman, MJ Berry II & W Bialek, Spin glass models for networks of real neurons. arXiv:0912.5409 [q–bio.NC] (2009).

[8] W Bialek & R Ranganathan, Rediscovering the power of pairwise interactions. arXiv:0712.4397 [q–bio.QM] (2007).

[9] M Weigt, RA White, H Szurmant, JA Hoch & T Hwa, Identification of direct residue contacts in protein–protein interaction by message passing. *Proc Nat'l Acad Sci (USA)* **106,** 67–72 (2009).

[10] N Halabi, O Rivoire, S Leibler & R Ranganathan, Protein sectors: Evolutionary units of three–dimensional structure. *Cell* **138,** 774–786 (2009).

[11] T Mora, AM Walczak, W Bialek & CG Callan, Maximum entropy models for antibody diversity. *Proc Nat'l Acad Sci (USA)* **107,** 5405–5410 (2010) (2009).

[12] TR Lezon, JR Banavar, M Cieplak, A Maritan & NV Federoff, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Nat'l Acad Sci (USA)* **103,** 19033–19038 (2006).

[13] G Tkačik, *Information Flow in Biological Networks* (Dissertation, Princeton University, 2007).

[14] GJ Stephens & W Bialek, Statistical mechanics of letters in words. *Phys Rev E* **81,** 066119 (2010).

[15] CE Shannon, A mathematical theory of communication. *Bell Sys. Tech. J.* **27,** 379–423 & 623–656 (1948). Reprinted in CE Shannon & W Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[16] TM Cover & JA Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[17] CM Bender & SA Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, New York, 1978).

[18] FJ Dyson, General theory of spin–wave interactions. *Phys Rev* **102**, 1217–1230 (1956).

[19] M Ballerini, N Cabibbo, R Candelier, A Cavagna, E Cisbani, I Giardina, V Lecomte, A Orlandi, G Parisi, A Procaccini, M Viale & V Zdravkovic, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proc Nat'l Acad Sci (USA)* **105,** 1232–1237 (2008).

[20] A Cavagna, I Giardina, A Orlandi, G Parisi, A Procaccini, M Viale & V Zdravkovic, The STARFLAG handbook on collective animal behaviour: 1. Empirical meth-

ods, *Animal Behaviour* **76,** 217–236 (2008).

[21] A Cavagna, I Giardina, A Orlandi, G Parisi & A Procaccini, The STARFLAG handbook on collective animal behaviour: 2. Three–dimensional analysis, *Animal Behaviour* **76,** 237–248 (2008).

[22] A Cavagna, A Cimarelli, I Giardina, G Parisi, R Santagati, F Stefanini & M Viale, Scale–free correlations in starling flocks. *Proc Nat'l Acad Sci (USA)* **107,** 11865–11870 (2010).

[23] H Edelsbrunner & Mucke E. P. ACM Trans. Graphics 13, 43-72 (1994).

[24] A Cavagna, A Cimarelli, I Giardina, G Parisi, R Santagati, F Stefanini & R Tavarone. From empirical data to inter-individual interactions: unveiling the rules of collective animal behaviour, *Math Models Methods Appl Sci* **20**, 1491–1510 (2010).

[25] G Grégoire & H Chaté, Onset of collective and cohesive motion. *Phys Rev Lett* **92**, 025702 (2004).

[26] G Grégoire, H Chaté & Y Tu, Moving and staying together without a leader, Physica D **181**, 157-170 (2003)