Statistical mechanics for natural flocks of birds

William Bialek,[a] Andrea Cavagna,[b] Irene Giardina,[b] Thierry Mora,[c]
Edmondo Silvestri,[b] Massimiliano Viale,[b] and Aleksandra M Walczak[d]

[a] *Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics,*
Princeton University, Princeton, New Jersey 08544 USA
[b] *Istituto Sistemi Complessi (ISC–CNR), Via dei Taurini 19, 00185 Roma, Italy*
[b] *Dipartimento di Fisica, "Sapienza" Universitá di Roma, P.le Aldo Moro 2, 00185 Roma, Italy*
[c] *Laboratoire de Physique Statistique de l'École Normale Supérieure,*
CNRS and Universities Paris VI and Paris VII,
24 rue Lhomond, 75231 Paris Cedex 05, France, and
[d] *Laboratoire de Physique Théorique de l'École Normale Supérieure,*
CNRS and University Paris VI, 24 rue Lhomond, 75231 Paris Cedex 05, France
(Dated: July 5, 2011)

Interactions among neighboring birds in a flock cause an alignment of their flight directions. We show that the minimally structured (maximum entropy) model consistent with these local correlations correctly predicts the propagation of order throughout entire flocks of starlings, with no free parameters. These models are mathematically equivalent to the Heisenberg model of magnetism, and define an "energy" for each configuration of flight directions in the flock. Comparing flocks of different densities, the range of interactions that contribute to the energy involves a fixed number of (topological) neighbors, rather than a fixed (metric) spatial range. Comparing flocks of different sizes, the model correctly accounts for the observed scale invariance of long ranged correlations among the fluctuations in flight direction.

The collective behaviour of large groups of animals is an imposing natural phenomenon, very hard to cast into a systematic theory [1]. Physicists have long hoped that such collective behaviours in biological systems could be understood in the same way as we understand collective behaviour in physics, where statistical mechanics provides a bridge between microscopic rules and macroscopic phenomena [2, 3]. A natural test case for this approach is the emergence of order in a flock of birds: out of a network of distributed interactions among the individuals, the entire flock spontaneously chooses a unique direction in which to fly [4], much as local interactions among individual spins in a ferromagnet lead to a spontaneous magnetization of the system as a whole [5]. Despite detailed development of these ideas [6–9], there still is a gap between theory and experiment. Here we show how to bridge this gap, by constructing a maximum entropy model [10] based on field data of large flocks of starlings [11–13]. We use this framework to show that the effective interactions among birds are local, and that the number of interacting neighbors is independent of flock density, confirming that interactions are ruled by topological rather than metric distance [14]. The statistical mechanics models that we derive in this way provide an essentially complete, parameter free theory for the propagation of directional order throughout the flock.

We consider flocks of European starlings *Sturnus vulgaris*, as in Fig 1A. At any given instant of time, following Refs [11–13], we can attach to each bird i a vector velocity $\vec{v}_i$, and define the normalized velocity $\vec{s}_i = \vec{v}_i/|\vec{v}_i|$ (Fig 1B). On the hypothesis that flocks have statistically stationary states, we can think of all these normalized velocities as being drawn (jointly) from a probability distribution $P(\{\vec{s}_i\})$. It is not possible to infer this full dis-
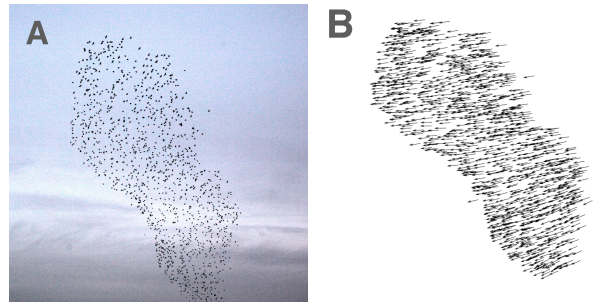

FIG. 1: The raw data. (A) One snapshot from flocking event 28–10, $N = 1246$ birds (see Table I in Methods). (B) Instantaneous vector velocities of all the individuals in this snapshot, normalized as $\vec{s}_i = \vec{v}_i/|\vec{v}_i|$.

tribution directly from experiments, since the space of states specified by $\{\vec{s}_i\}$ is too large. However, what we can measure from field data is the matrix of correlations between the normalized velocities, $C_{ij} = \langle \vec{s}_i \cdot \vec{s}_j \rangle$. There are infinitely many distributions $P(\{\vec{s}_i\})$ that are consistent with the measured correlations, but out of all these distributions, there is *one* that has minimal structure: it describes a system that is as random as it can be while still matching the experimental data. This distribution is the one with maximum entropy [10].

It should be emphasized that the maximum entropy principle is not a "modeling assumption;" rather it is the absence of assumptions. Any other model that accounts for the observed correlations will have more structure, and hence (explicitly or implicitly) assumes something about the nature of the interactions in the flock beyond what is required to match the data. Of course the fact

that the maximum entropy model is minimally structured does not make it correct. It could be, for example, that individual birds set their flight direction by computing a complicated nonlinear combination of the velocities from multiple neighbors, in which case correlations among pairs of birds would be insufficient to capture the essence of the ordering mechanism. We view the maximum entropy model as a powerful starting point, from which, as we will see, we can generate detailed and testable predictions.

The maximum entropy distribution consistent with the directional correlations C_{ij} is

$$P(\{\vec{s}_i\}) = \frac{1}{Z(\{J_{ij}\})} \exp\left[\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij}\ \vec{s}_i \cdot \vec{s}_j\right], \qquad (1)$$

where $Z(\{J_{ij}\})$ is the appropriate normalization factor, or partition function; the derivation follows Ref [10], as explained in Appendix A. Notice that there is one parameter J_{ij} corresponding to each measured element C_{ij} of the correlation matrix. To finish the construction of the model, we have to adjust the values of the J_{ij} to match the experimentally observed C_{ij},

$$\langle \vec{s}_i \cdot \vec{s}_j \rangle_P = \langle \vec{s}_i \cdot \vec{s}_j \rangle_{\exp}\ , \qquad (2)$$

where the symbol $\langle \cdot \rangle_P$ indicates an average using distribution P from Eq (1), whereas $\langle \cdot \rangle_{\exp}$ indicates an average over many experiments. This matching condition is equivalent to maximizing the likelihood that the model in Eq (1) will generate the data from which the correlations were computed.

The probability distribution in Eq (1) is mathematically identical to a model that is familiar from the physics of magnets—the Heisenberg model [5]—in which a collection of spins $\vec{s}_i$ interact so that their energy (or Hamiltonian) is $H(\{\vec{s}_i\}) = -(1/2) \sum_{i,j} J_{ij}\ \vec{s}_i \cdot \vec{s}_j$; Eq (1) then describes the thermal equilibrium or Boltzmann distribution at a temperature $k_B T = 1$. In this context, the constants J_{ij} are the strength of interaction between spins i and j, where $J > 0$ means that these elements tend to align. For many physical systems, once we know the Hamiltonian there is a plausible dynamics that allows the system to relax toward equilbrium, and this is the Langevin dynamics

$$\frac{d\vec{s}_i}{dt} = -\frac{\partial H}{\partial \vec{s}_i} + \vec{\eta}_i(t) = \sum_{j=1}^{N} J_{ij}\vec{s}_j + \vec{\eta}_i(t)\ , \qquad (3)$$

where $\vec{\eta}_i(t)$ is an independent white noise "force" driving each separate degree of freedom. Finding trajectories $\vec{s}_i(t)$ that solve Eq (3) produces samples that are drawn out of the probability distribution in Eq (1). The interesting point is that this kind of dynamical model also is well known in biology: the direction of motion of an individual evolves in time according to "social forces" reflecting a weighted sum of inputs from neighboring individuals, plus noise [4]. In this framework, J_{ij} measures

the strength of the force that tries to align the velocity of bird i along the direction defined by bird j. We emphasize that this is not an analogy, but a mathematical equivalence.

In contrast to most networks, the connectivity in a flock of birds is intrinsically dynamic—birds move and change their neighbors. Thus, it may not make sense to talk about matrix of correlations C_{ij} or interactions J_{ij} between labelled individuals. On the other hand, the continuous and rapid change of neighbors induced by motion implies that the interaction J_{ij} between bird i and bird j cannot depend directly on their specific identities, but only on some function of their relative positions.

The simplest form of interaction that is independent of the birds' identity is one in which each bird interacts with the same strength, J, with the same number of neighbors, n_c (or with all birds within the same radius r_c; see below). If the interactions are of this form, then Eq (1) simplifies to

$$P(\{\vec{s}_i\}) = \frac{1}{Z(J, n_c)} \exp\left[\frac{J}{2} \sum_{i=1}^{N} \sum_{j \in n_c^i} \vec{s}_i \cdot \vec{s}_j\right]\ , \qquad (4)$$

where $j \in n_c^i$ means that bird j belongs to the first n_c nearest neighbors of i. This is, in fact, the maximum entropy model consistent with the average correlation among birds within the neighborhood defined by n_c,

$$C_{\text{int}} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j \in n_c^i} \langle \vec{s}_i \cdot \vec{s}_j \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_c} \sum_{j \in n_c^i} \vec{s}_i \cdot \vec{s}_j\ . \quad (5)$$

Biologically, Eq's (4) and (5) encapsulate the concept that the fundamental correlations are between birds and their directly interacting neighbors; all more distant correlations should be derivable from these. As in the more general problem, finding the values of J and n_c that reproduce the observed correlation C_{int} is the same as maximizing the probability, or likelihood, that model Eq (4) generates the observed configuration of flight directions $\{\vec{s}_i\}$ in a single snapshot. If this is correct, a model that appropriately reproduces the fundamental correlations up to the scale n_c must be able to describe correlations on all length scales.

Importantly, with large flocks we can estimate the correlations among interacting neighbors from a single snapshot of the birds' flight directions $\{\vec{s}_i\}$, as indicated in the second step of Eq (5). In contrast, if we were trying to estimate the entire correlation matrix in Eq (2), we would need as many samples as we have birds in the flock (see Appendix B), and we would have to treat explicitly the dynamic rearrangements of the interaction network during flight. This is an extreme version of the general observation that the sampling problems involved in the construction of maximum entropy models can be greatly reduced if we have prior expectations that constrain the structure of the interaction matrix [15, 16].
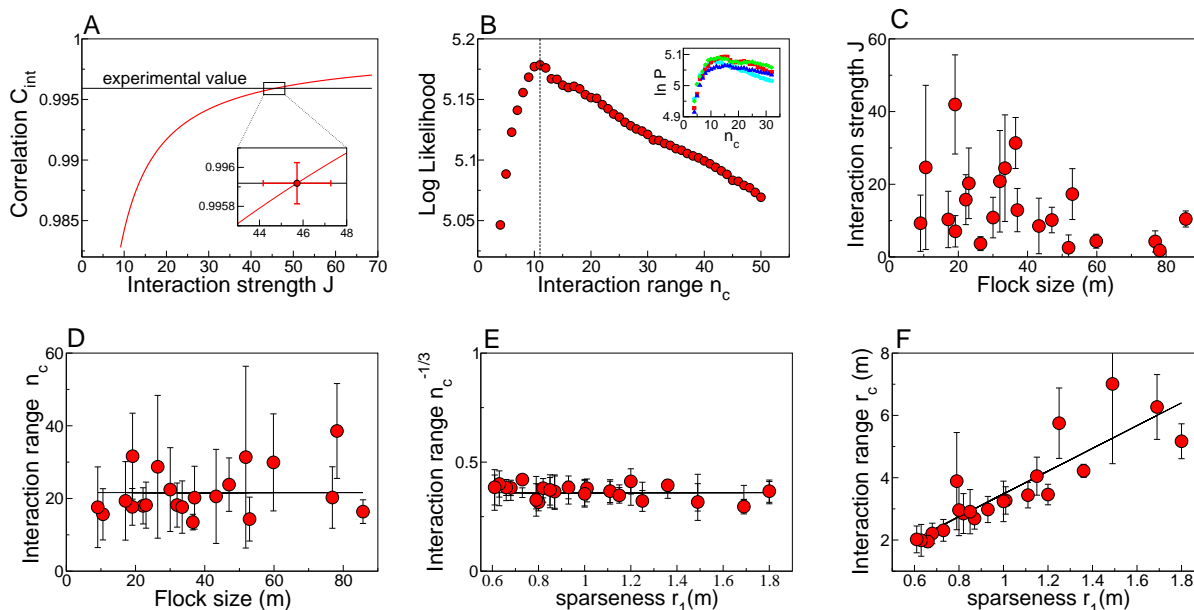
FIG. 2: Setting the strength and range of interactions. (A) The predicted strength of correlation, C_{int}, as a function of the interaction strength J, with $n_c = 11$, for the snapshot in Fig 1. Matching the experimental value of $C_{\text{int}} = 0.99592$ determines $J = 45.73$. Inset: Zoom of the crossing point; error bars are obtained from the model's predictions of fluctuations of $C_{\text{int}}(J, n_c)$. (B) The log–likelihood of the data per bird ($\langle \ln P(\{\vec{s}_i\})\rangle_{\text{exp}}/N$) as a function of n_c with J optimized for each value of n_c; same snapshot as in (A). There is a clear maximum at $n_c = 11$. Inset: the log–likelihood per bird for other snapshots of the same flocking event. (C) The inferred value of J for all observed flocks, shown as a function of the flock's size. Each point corresponds to an average over all the snapshots of the same flock. Error bars are standard deviations across multiple snapshots. (D) As in (C), but for the inferred values of n_c. Averaging over all flocks we find $n_c = 21.2 \pm 1.7$ (black line). (E) The inferred value of the topological range $n_c^{-1/3}$ as a function of the mean inter–bird distance in the flock, for all flocks. Error bars are standard deviations across multiple snapshots of the same flock. (F) As in (E), but for the metric range r_c. Interactions extend over some fixed metric distance r_0, then we expect $n_c^{-1/3} \propto r_1/r_0$ and $r_c = $ constant; we find the opposite pattern, which is a signature of interactions with a fixed number of topological neighbors [14].

We now apply this analysis to data on real flocks of starlings. Given a snapshot of the flock, we have the configuration $\{\vec{s}_i\}$, and we need to evaluate the probability $P(\{\vec{s}_i\})$ in Eq (4) for any value of J and n_c, then maximize this probability with respect to these parameters (see Methods and the Appendices for details of the calculation). Special care must be devoted to birds on the outer edge, or border, of the flock, since these individuals have very asymmetric neighborhoods and may receive inputs from signals outside the flock. If we take the flight directions of these border birds as given, we can study how information propagates through the flock, without having to make assumptions about how the boundary is different from the interior. Technically, then, we describe the flock with Eq (4), but with the flight directions of the border birds fixed (again, see Methods and especially Appendix C for details).

We proceed as follows: For a single flock, at a given instant of time, we compute the correlation C_{int} predicted by the model in Eq (4) as a function of the coupling strength J, and compare it with the experimental value of the correlation (Fig 2A). The equation $C_{\text{int}}(J, n_c) = C_{\text{int}}^{\text{exp}}$, fixes $J(n_c)$ for each value of n_c. Then

we fix the interaction range by looking at the overall probability of the data as a function of n_c. In general there is a clear optimum (Fig 2B), from which we finally infer the maximum entropy value of both parameters, n_c and J. We repeat this procedure for every snapshot of each flock, and compute the mean and standard deviation of the interaction parameters for each flock over time. Alternatively, we can average the log–likelihood over many snapshots, and then optimize, and this gives equivalent results for J and n_c (see Appendix E).

In Figures 2C and D we report the value of the interaction strength J and of the interaction range n_c for all flocks, as a function of the flock's spatial size, L. The inferred values of J and n_c are reproducible, although error bars are larger for smaller flocks. In particular, J and n_c do not show any significant trend with the flocks' linear dimensions, with the number of birds, or with the density. This did not have to be the case, nor is it in any way built in to our framework; for example, if the real interactions extended over long distances, then our fitting procedure would produce an increase of n_c and J with the size of the flock.

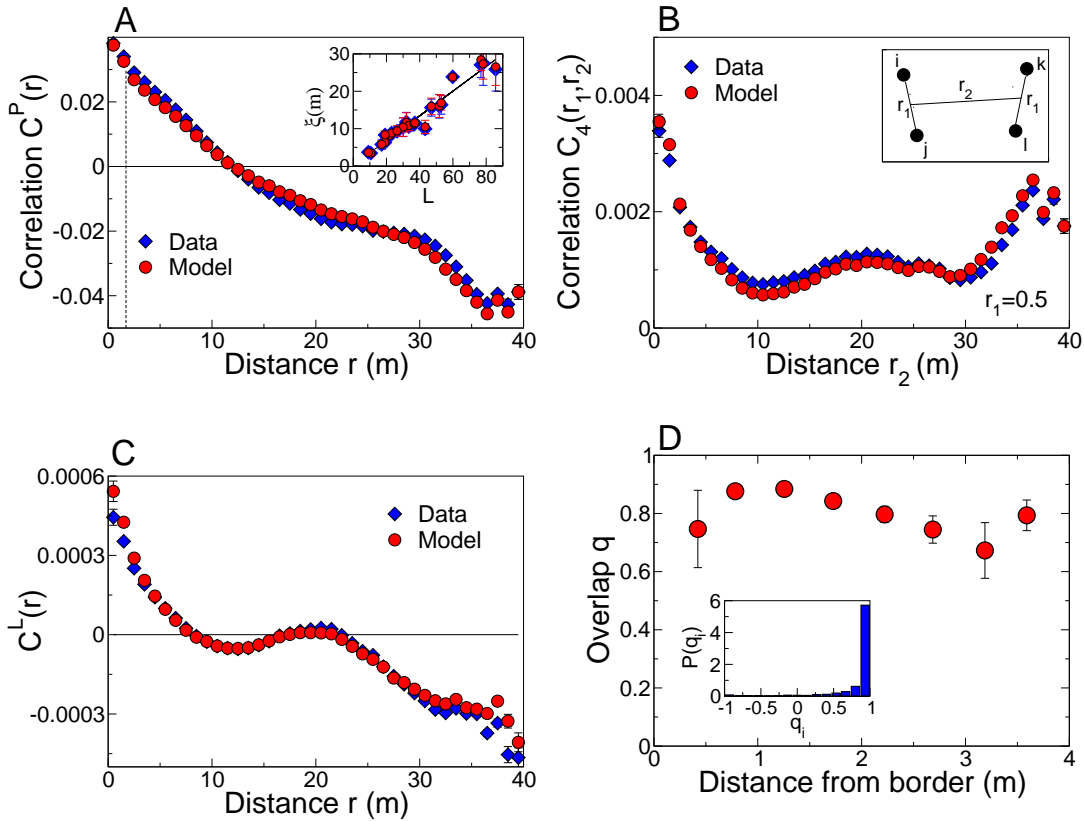In Figure 2E we also show that the interaction range

FIG. 3: Correlation functions predicted by the maximum entropy model vs experiment. The full pair correlation function can be written in terms of a longitudinal and a perpendicular component, i.e. $\langle \vec{s}_i \cdot \vec{s}_j \rangle = \langle s_i^{L} s_j^{L} \rangle + \langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle$. Since the two components have different amplitudes, it is convenient to look at them separately. (A) Perpendicular component of the correlation, $C^{P}(r) = \langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle$, as a function of the distance; the average is performed over all pairs ij separated by distance r. Blue diamonds refer to experimental data (for the sample in Fig 1), red circles to the prediction of the model in Eq (4). The dashed line marks the maximum r that contributes to C_{int}, which is the only input to the model. The correlation function is well fitted over all length scales. In particular, the correlation length ξ, defined as the distance where the correlation crosses zero, is well reproduced by the model. Inset: ξ vs. size of the flock, for all the flocking events; error bars are standard deviations across multiple snapshots of the same flocking event. (B) Four–point correlation function $C_4(r_1; r_2) = \langle (\vec{\pi}_i \cdot \vec{\pi}_j)(\vec{\pi}_k \cdot \vec{\pi}_l) \rangle$, where the pairs ij and kl are as shown in the inset (see also Appendix G). The figure shows the behaviour of $C_4(r_1; r_2)$ as a function of r_2, with $r_1 = 0.5$. (C) Longitudinal component of the correlation $C^{L}(r) = \langle s_i^{L} s_j^{L} \rangle - S^2$, as a function of distance. Note that in the spin wave approximation, $C^{L}(r) = 1 - C_4(0; r) - S^2$. (D) Similarity between the predicted mean value of flight direction, $\langle \vec{\pi}_i \rangle$, and real data, for all individual birds in the interior of the flock. The similarity can be quantified through the local overlap $q_i = \langle \vec{\pi}_i \rangle \cdot \vec{\pi}_i^{exp} / (|\langle \vec{\pi}_i \rangle||\vec{\pi}_i^{exp}|)$, which is plotted as a function of the distance of the individual from the border. Maximal similarity corresponds to $q_i = 1$. Inset: full distribution $P(q)$ for all the interior birds.

n_c does not depend on the typical distance between neighboring birds, r_1, which is closely related to the flock's density. Of course, we can run exactly the same method using a *metric* interaction range, r_c, rather than a *topological* range, n_c. We simply set $J_{ij} = J$ if and only if birds i and j lie within r_c meters. When we do this we find that the metric range r_c does depend on the nearest neighbor distance r_1, in contrast with the topological range n_c (Fig 2F). This result provides strong support for the claim put forward in Ref [14] that birds interact with a fixed number of neighbors, rather than with all the birds within a fixed metric distance.

Having fixed J and n_c by matching the scalar correlation in the flock, we have no free parameters—

everything that we calculate now is a parameter free prediction. We begin by computing the correlations between pairs of birds as a function of their distance, $C(r) \sim \sum_{ij} \vec{s}_i \cdot \vec{s}_j\, \delta(r - r_{ij})$, as shown in Fig 3A. There is extremely good agreement across the full range of distances. As we have seen, our maximum entropy calculation finds local interactions, i.e. a relatively small value of n_c ($n_c \sim 20$ for flocks of up to thousands birds). This implies that the scalar correlation C_{int}, used as an experimental input to the calculation, is the integral of $C(r)$ only over a very small interval close to $r = 0$: only the average value of pair correlations at very short distances is used as an input to the calculation, whereas all the long range part of $C(r)$ is not. Nevertheless, we have very
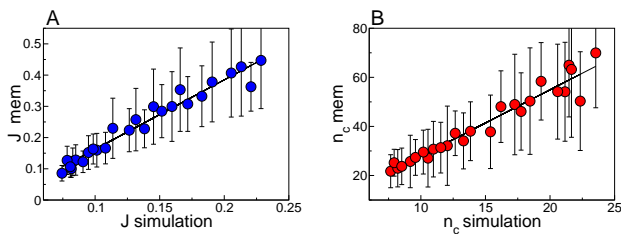
FIG. 4: Maximum entropy analysis for a model of self–propelled particles. (A) Inferred value of the parameter J vs. microscopic strength of alignment forces used in the simulation. (B) Inferred value of n_c vs. the true number of interacting neighbors in the simualtion. Slopes of the lines are 2.2 and 2.7, respectively. Error bars are standard deviations across 45 snapshots of the same simulation.

good agreement out to the overall extent of the flock itself. This confirms our expectation that a model for the local correlations is able to describe correlations on all length scales. We draw attention to the fact that the apparent correlation length, defined by $C(r = \xi) = 0$, is predicted to scale with the linear size of the flock ($\xi \propto L$, inset to Fig 3A), as observed experimentally.

Correlations exist not just between pairs of birds, but among larger n–tuplets. In Fig 3B we consider the correlations among quadruplets of birds. Although these correlations are small, their shape is nontrivial and quite noiseless. The model, which takes only local pairwise correlations as input, reproduces very accurately these 4-body correlations, including a non–monotonic dependence on distance, out to distances comparable to the full extent of the flock. Again, this is not a fit, but a parameter free prediction.

Finally, instead of measuring correlations, we can ask the model to predict the actual flight directions of individual birds in the interior of the flock, given the directions chosen by birds on the borders. This can't work perfectly, since the model predicts that individual birds have an element of randomness in their choice of direction relative to their neighbors, but as shown in Fig 3D the overlap between predicted and observed directions is very good, not just for birds close to the border but throughout the entire "thickness" of the flock.

The maximum entropy model has a mechanistic interpretation, from Eq (3), in terms of social forces driving the alignment of the flight directions. Given the success of the model in predicting the propagation of order throughout the flock, it is interesting to ask whether we can take this mechanistic interpretation seriously. As a test, we have simulated a population of self–propelled particles in three dimensions moving according to social forces that tend to align each particle with the average direction of its neighbors, as described by Eqs. (9) and (10) in the Methods. We then compared the simulation parameters $(J^{\mathrm{sim}}, n_c^{\mathrm{sim}})$ to the values $(J^{\mathrm{mem}}, n_c^{\mathrm{mem}})$ obtained by applying the maximum entropy method to

snapshots drawn from the simulation, just as we have analyzed the real data. Both the strength and the range of the interaction given by the maximum entropy analysis are proportional to the "microscopic" parameters used in the simulation (Figs 4A and B), although the maximum entropy interaction range n_c^{mem} is roughly $3\times$ larger than the true number of interacting neighbors, n_c^{sim}. We believe that this overestimation is due to the fact that birds (unlike spins) move through the flock, encountering new neighbors before losing memory of the earlier flight directions, and in so doing propagate information and correlation more effectively than if they were sitting on a fixed network. In other words, the maximum entropy model, where interactions are static by construction, compensates the dynamical nature of the true interaction network by giving a larger effective value of n_c. Hydrodynamic theories of flocking [6, 8] provide an analytic treatment of this effect, which is essential for collective motion of large two–dimensional groups. Indeed, in the limit of very large flocks, this ratio between the microscopic range of interactions and the effective range recovered by maximum entropy methods could become arbitrarily large, but the flocks we study here seem not to be big enough for this effect to take over. If we use the "calibration" of model from Fig 4B, then the observation of $n_c = 21.6$ in the real flocks (Fig 2) suggests that the true interactions extend over $n_c = 7.8$, in reasonable agreement with the result from [14, 20], $n_c = 7.0 \pm 0.6$ using very different methods.

To summarize, we have constructed the minimal model that is consistent with a single number characterizing the interactions among birds in a flock, the average correlation between the flight directions of immediate neighbors. Perhaps surprisingly, this provides an essentially complete theory for the propagation of directional order throughout the flock, with no free parameters. The theory predicts major qualitative effects, such as the presence of long ranged, scale free correlations among pairs of birds, as well as smaller, detailed effects such as the non–monotonic distance dependence of (four–point) correlations among two pairs of birds. The structure of the model corresponds to interactions with a fixed number of (topological) neighbors, rather than with all neighbors that fall within a certain (metric) distance; the relevant number of neighbors and the strength of the interaction are remarkably constant across multiple flocking events. Our approach can be seen as part of a larger effort using maximum entropy methods to link the collective behaviour of real biological systems to theories grounded in statistical mechanics [21–33]. As with these other examples, we view the success of our theory as an encouraging first step. We have focused on the flight directions, taking the positions of the birds as given. A full theory must connect the velocities of the birds to their evolving positions, which requires more accurate measurements of trajectories over time, and we must consider the fluctuations in the speed as well as direction of flight. There are maximum entropy approaches to both

of these problems, and the mapping from maximum entropy models to statistical mechanics suggests that the observation of scale free correlations in speed fluctuations [17] will locate models of flocking at an especially interesting point in their parameter space [34].

Methods

Data. Analyzed data were obtained from experiments on large flocks of starlings (*Sturnus vulgaris*), in the field. Using stereometric photography and innovative computer vision techniques [12, 13] the individual 3D coordinates and velocities were measured in cohesive groups of up to 4268 individuals [11, 14, 17]. The data-set comprises 21 distinct flocking events (see Table I). Each event consists of up to 40 consecutive 3D configurations (individual positions and velocities), at time intervals of 1/10 s.

Analytic approach to the maximum entropy model. To apply the maximum entropy analysis, we need to compute the expected values of correlation functions according to the measure defined in Eq (1). This requires the computation of the partition function $Z(\{J_{ij}\})$, which is, in general, a hard task. Flocks, however, are very ordered groups, in that birds' velocities are neatly aligned to each other [17]. In this case we can use the "spin wave" approximation [18], which exploits the strong ordering condition. Let us call $\vec{S} = (1/N)\sum_i \vec{s}_i = S\,\vec{n}$ 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: $\vec{s}_i = s_i^L \vec{n} + \vec{\pi}_i$. If the system is highly polarized, $S \sim 1$, $|\vec{\pi}_i| \ll 1$, and $s_i^L \sim 1 - |\vec{\pi}_i|^2/2$; we verify this last condition for our data in Fig 5 of Appendix B. The partition function can be written as an integral over the $\{\vec{\pi}\}$, and if $S \sim 1$ the leading terms are (see Appendix B for details):

$$Z(\{J_{ij}\}) \approx \int d^N \vec{\pi} \exp\left[-\frac{1}{2}\sum_{i,j=1}^N A_{ij}\vec{\pi}_i \cdot \vec{\pi}_j + \frac{1}{2}\sum_{i,j=1}^N J_{ij} \right], \tag{6}$$

where $A_{ij} = \sum_k J_{ik}\delta_{ij} - J_{ij}$, $d^N\vec{\pi} = \prod_i d\vec{\pi}_i$ and the $\{\vec{\pi}_i\}$ satisfy the constraint $\sum_i \vec{\pi}_i = 0$. If we consider the flight directions of birds on the border as given, integration must be performed with respect to internal variables only (see Appendix D). After some algebra one gets

$$Z(\{J_{ij}\}; \mathcal{B}) = \int d^{\mathcal{I}} \vec{\pi} \exp\left[-\frac{1}{2}\sum_{i,j\in\mathcal{I}} A_{ij}\vec{\pi}_i \cdot \vec{\pi}_j + \sum_{i\in\mathcal{I}} \vec{\pi}_i \cdot \vec{h}_i + \frac{1}{2}\sum_{i,j\in\mathcal{I}} J_{ij} + \frac{1}{2}\sum_{i\in\mathcal{I}} h_i^L + \frac{1}{2}\sum_{i,j\in\mathcal{B}} J_{i,j}\vec{s}_i \cdot \vec{s}_j \right] \;. \tag{7}$$

Here $\mathcal{I}$ and $\mathcal{B}$ represent the subsets of, respectively, internal and border individuals; $\vec{h}_i = \sum_{l\in\mathcal{B}} J_{il}\vec{s}_l$ is a 'field' describing the influence of birds on the border on internal bird i; and, now, $A_{ij} = \delta_{ij}(\sum_{k\in\mathcal{I}} J_{ik} + \sum_{l\in\mathcal{B}} J_{il}s_l^L)$. The integral in Eq (7) can be carried out explicitly; see Eq (D6). The reduced model in Eq (4) corresponds to $J_{ij} = J\, n_{ij}$, with $n_{ij} = 1$, $1/2$, or 0 according to whether both individuals, just one, or none, belong to the local n_c-neighborhood of the other. Given the individual coordinates of birds in space, the matrix A_{ij} can be computed for any given snapshot, and $Z(J, n_c; \mathcal{B})$ (and correlation functions) can be calculated as a function of J and n_c. These two parameters must then be adjusted to maximize the log–likelihood of the data,

$$\left\langle \log P\left(\{\vec{s}_i\}\right) \right\rangle_{\exp} = -\log Z(J, n_c; \mathcal{B}) + \frac{1}{2}J n_c N C_{\text{int}}^{\exp} \;. \tag{8}$$

Maximizing with respect to J corresponds to equating expected and experimental correlations. In our case, this equation can be solved analytically, leading to an explicit expression of the optimal J vs. n_c; see Eq (D12). Minimization with respect to n_c can then be performed numerically. A graphical visualization of the solution can be found in Fig. 2.

Self-propelled particle model. We consider a model of self–propelled particles extensively studied in the literature [9]. Each particle moves with vector velocity $\vec{v}_i(t)$ according to the following equations:

$$\vec{v}_i(t+1) = v_0 \Theta \left[\alpha \sum_{j\in n_c^i} \vec{v}_j(t) + \beta \sum_{j\in n_c^i} \vec{f}_{ij} + n_c\, \vec{\eta}_i \right] \tag{9}$$

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t) \;, \tag{10}$$

where Θ is a normalization operator $\Theta(\vec{y}) = \vec{y}/|\vec{y}|$ that serves to keep the speed fixed at $|\vec{v}| = v_0$, and $j \in n_c^i$

means that j belongs to the n_c interacting neighbors of i. The distance–dependent force $\vec{f}_{ij}$ acts along the direction connecting i and j; following Ref [9], if $\vec{e}_{ij}$ is the unit vector between i and j, we take

$$\vec{f}_{ij}(r_{ij} < r_b) \; = \; -\infty \, \vec{e}_{ij} \tag{11}$$

$$\vec{f}_{ij}(r_b < r_{ij} < r_a) \; = \; \frac{1}{4} \cdot \frac{r_{ij} - r_e}{r_a - r_e} \vec{e}_{ij} \tag{12}$$

$$\vec{f}_{ij}(r_a < r_{ij} < r_0) \; = \; \vec{e}_{ij}, \tag{13}$$

Finally, $\vec{\eta}_i$ is a random unit vector, independent for each bird and at each moment of time. The parameters α and β tune the strength of the alignment and of the cohesion force, respectively; in particular, the strength of alignment is given by $J = v_0 \alpha / n_c$. To test the maximum entropy analysis, we modified the model in such a way that we could vary n_c. Specifically, we introduced an angular resolution μ such that only neighbors with mutual angles larger than μ were included in the neighborhood. When μ is of the order of the Voronoi angle the model is statistically equivalent to the original version (where Voronoi neighbors were considered), but increasing (decreasing) μ one can decrease (increase) the value of n_c. In this way both the number n_c of interacting neighbors and the strength of the interaction J can be arbitrarily tuned. Parameters were chosen as $r_0 = 1$ (to set the scale of distance), $r_b = 0.2$, $r_e = 0.5$, $r_a = 0.8$, $\alpha = 35$, $\beta = 5$, $v_0 = 0.05$, and we simulated a flock of $N = 512$ birds.

Appendix A: Maximum entropy approach

The maximum entropy method 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 [21–26], ensembles of amino acid sequences [27–30], biochemical and genetic networks [31, 32], and the statistics of letters in words [33]. 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 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

Event	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

TABLE I: Summary of experimental data. 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 and in Appendix B. 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.

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 $\{\langle f_\nu(\mathbf{x}) \rangle_{\exp}\}$, where $\langle \cdots \rangle_{\exp}$ 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 [35, 36], and this is the

entropy

$$S[P] = -\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x}) . \qquad (A1)$$

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

$$\langle f_\mu(\mathbf{x}) \rangle_{\exp} = \langle f_\mu(\mathbf{x}) \rangle_P \equiv \sum_{\mathbf{x}} P(\mathbf{x}) f_\mu(\mathbf{x}) \qquad (A2)$$

for all μ [10]. 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 [19]: we introduce a generalized entropy function,

$$S[P; \{\lambda_\nu\}] = S[P] - \sum_{\mu=0}^{K} \lambda_\mu \left[\langle f_\mu(\mathbf{x}) \rangle_P - \langle f_\mu(\mathbf{x}) \rangle_{\exp} \right] ,$$
$$(A3)$$

where a multiplier λ_μ appears for each constraint to be satisfied, and then we maximize 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

$$0 = \frac{\partial S[P; \{\lambda_\nu\}]}{\partial P(\mathbf{x})}$$

$$= \frac{\partial S[P]}{\partial P(\mathbf{x})} - \sum_{\mu=0}^{K} \lambda_\mu \frac{\partial \langle f_\mu(\mathbf{x}) \rangle_P}{\partial P(\mathbf{x})}$$

$$= -\ln P(\mathbf{x}) - 1 - \sum_{\mu=0}^{K} \lambda_\mu f_\mu(\mathbf{x}), \qquad (A4)$$

$$\Rightarrow P(\mathbf{x}) = \frac{1}{Z(\{\lambda_\nu\})} \exp \left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x}) \right], \quad (A5)$$

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,

$$Z(\{\lambda_\nu\}) = \sum_{\mathbf{x}} \exp \left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x}) \right]. \qquad (A6)$$

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

$$0 = \frac{\partial S[P; \{\lambda_\nu\}]}{\partial \lambda_\mu}$$

$$= \langle f_\mu(\mathbf{x}) \rangle_{\exp} - \langle f_\mu(\mathbf{x}) \rangle_P$$

$$\Rightarrow \langle f_\mu(\mathbf{x}) \rangle_{\exp} = \frac{1}{Z(\{\lambda_\nu\})} \sum_{\mathbf{x}} f_\mu(\mathbf{x}) \exp \left[-\sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x}) \right] .$$
$$(A7)$$

Thus, when we optimize 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 (A5) back into Eq (A3), we obtain

$$S[P; \{\lambda_\nu\}] = \ln Z(\{\lambda_\nu\}) + \sum_{\mu=0}^{K} \lambda_\mu \langle f_\mu(\mathbf{x}) \rangle_{\exp} , \quad (A8)$$

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 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 (A6) 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

$$P(\mathbf{x}) = \frac{1}{Z(\beta)} e^{-\beta E(\mathbf{x})}, \qquad (A9)$$

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

$$Z(\beta) = \sum_{\mathbf{x}} e^{-\beta E(\mathbf{x})}. \qquad (A10)$$

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 (A5) and (A9) gives us

$$E(\mathbf{x}) = \sum_{\mu=1}^{K} \lambda_\mu f_\mu(\mathbf{x}), \qquad (A11)$$

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 $\langle \vec{s}_i \cdot \vec{s}_j \rangle$ are measured experimentally. Thus we can use the general maximum entropy formulation, identifying $\mathbf{x} = \{\vec{s}_i\}$ and $f_\mu(\mathbf{x}) = \vec{s}_i \cdot \vec{s}_j$. Since the quantities that will be measured refer to pairs, it is useful to set $\lambda_\mu = -J_{ij}$, and we obtain Eq (1),

$$P(\{\vec{s}_i\}) = \frac{1}{Z(\{J_{ij}\})} \exp \left[\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} \vec{s}_i \cdot \vec{s}_j \right] . \quad (A12)$$

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

The model defined by Eq (A12) 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_{ij}. In this context, the effective energy is

$$E(\{\vec{s}_i\}) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} \vec{s}_i \cdot \vec{s}_j. \qquad (A13)$$

For $J_{ij} > 0$, the energy is lowered when the vectors $\vec{s}_i$ and $\vec{s}_j$ are parallel.

Appendix B: The spin wave approximation

The most demanding step in evaluating the probability distribution in Eq (A12) is the computation of the partition function

$$Z(\{J_{ij}\}) = \int d^N \mathbf{\vec{s}} \, \exp\left[\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} \vec{s}_i \cdot \vec{s}_j \right], \qquad (B1)$$

where we recall that the $\{\vec{s}_i\}$ are real, three dimensional vectors of unit length and $d^N \mathbf{\vec{s}} = \prod_i d\vec{s}_i$.

In presence of strong ordering, we can use the "spin wave" approximation to compute analytically the partition function of the Heisenberg model, Eq (B1). Let us call $\vec{S} = (1/N) \sum_i \vec{s}_i = S\vec{n}$ 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 $\vec{n}$,

$$\vec{s}_i = s_i^L \vec{n} + \vec{\pi}_i \,, \qquad (B2)$$

where, by construction, $\sum_i s_i^L = SN$, $\vec{\pi}_i \cdot \vec{n} = 0$, and $\sum_i \vec{\pi}_i = 0$. The partition function then reads

$$Z(\{J_{ij}\}) = \int d^N \mathbf{s^L} \, d^N \mathbf{\vec{\pi}} \left[\prod_i \delta\left((s_i^L)^2 + |\vec{\pi}_i|^2 - 1 \right) \right] \delta\left(\sum_i \vec{\pi}_i \right) \exp\left[\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} \left(s_i^L s_j^L + \vec{\pi}_i \cdot \vec{\pi}_j \right) \right] \,, \qquad (B3)$$

where $d^N \mathbf{s^L} = \prod_i ds_i^L$ and $d^N \mathbf{\vec{\pi}} = \prod_i d\vec{\pi}_i$. The delta functions implement the constraint on the length of each vector $\vec{s}_i$ and the global constraint on the $\vec{\pi}_i$. Note that since the $\{\vec{\pi}_i\}$'s belong to the subspace perpendicular to $\vec{n}$, in Eq (B4) there are only two independent degrees of freedom for each integration variable.

If the system is highly polarized, $S \sim 1$ and $|\vec{\pi}_i| \ll 1$.

The constraint on the norm of the vectors can then be written as $s_i^L \sim 1 - |\vec{\pi}_i|^2/2$. Note that indeed flocks are very polarized groups and this expression is very well satisfied by the data (see Fig 5 and Table I). Using this expansion the longitudinal components can be integrated out easily. The partition function then becomes, to leading order in the $\vec{\pi}$'s,

$$Z(\{J_{ij}\}) = \int d^N \mathbf{\vec{\pi}} \left[\prod_i \frac{1}{\sqrt{1 - |\vec{\pi}_i|^2}} \right] \delta\left(\sum_i \vec{\pi}_i \right) \exp\left[-\frac{1}{2} \sum_{i,j=1}^{N} A_{ij} \vec{\pi}_i \cdot \vec{\pi}_j + \frac{1}{2} \sum_{i,j=1}^{N} J_{ij} \right] \qquad (B4)$$

with

$$A_{ij} = \sum_k J_{ik} \delta_{ij} - J_{ij} \,. \qquad (B5)$$

The product over $1/\sqrt{1 - |\vec{\pi}_i|^2}$ in Eq (B4) is the Jacobian coming from the integration over the s_i^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_k and eigen-

vectors $\mathbf{w}^k$ as usual through

$$\sum_j A_{ij} w_j^k = a_k w_j^k. \tag{B6}$$

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

$$\sum_j A_{ij} w_i^1 = \frac{1}{\sqrt{N}} \sum_j A_{ij} = 0. \tag{B7}$$

The argument of the delta function in Eq (B4) is re-

lated only to the projection of the $\{\vec{\pi}_i\}$ 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 (B4) in the orthonormal basis defined by $\{\vec{w}^k\}$:

$$Z(\{J_{ij}\}) = \int d^N \vec{\boldsymbol{\pi}}' \, \delta\left(\vec{\pi}'_1\right) \exp\left[-\frac{1}{2}\sum_{k=1}^{N} a_k |\vec{\pi}'_k|^2 + \frac{1}{2}\sum_{i,j=1}^{N} J_{ij}\right], \tag{B8}$$

where $\vec{\pi}'_k = \sum_i w_i^k \vec{\pi}_i$. This leads to

$$\log Z(\{J_{ij}\}) = -\sum_{k>1} \log(a_k) + \frac{1}{2}\sum_{i,j=1}^{N} J_{ij}, \tag{B9}$$

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

$$\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle_{\exp} = 2 \sum_{k>1} \frac{w_i^k w_j^k}{a_k}, \tag{B10}$$

where the right hand side, the expectation value $\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle_P$, can be obtained from Eq (B8) 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. This motivates, as discussed in the text, the analysis of a more restricted problem, in which we know only the average local correlations C_{int} from Eq (5).
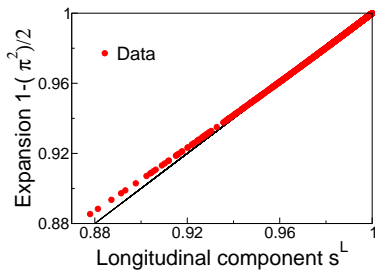

FIG. 5: Longitudinal components of the flight directions vs. prediction of the spin wave expansion, for all individuals in the snapshot of Fig 1. The black line has slope 1. Note that 95% of the birds have $s_i^L > 0.94$, and lie well on the line.

Appendix C: Computation with free boundaries

Let us now address more in details the reduced model, Eq (5) in main text, where each individual interacts with constant strength with its first n_c neighbors. In this case the J_{ij}'s have a particularly simple form:

$$J_{ij} = J \, n_{ij} \tag{C1}$$

with

$$
n_{\mathrm{ij}} = \begin{cases} 1 & \text{if } j \in n_c^i \text{ and } i \in n_c^j \,, \\[2mm] \frac{1}{2} & \text{if } j \in n_c^i \text{ and } i \notin n_c^j \,, \text{ or vice versa, and} \\[2mm] 0 & \text{otherwise.} \end{cases}
$$
(C2)

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 this case, therefore, $A_{\mathrm{ij}} = J\tilde{A}_{\mathrm{ij}}$, where $\tilde{A}$ depends only on the neighborhood relations.

Before proceeding with the full computation, let us briefly look at the simplest case, where we allow all the $\vec{\pi}_i$'s to freely fluctuate according to Eq (B8). The result can be read directly from Eq (B9), giving

$$
\log Z(J, n_c) = -\frac{1}{2} \sum_{k>1} \log(J\lambda_k) + \frac{N J n_c}{2} \,,
$$
(C3)

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

$$
\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle = \frac{2}{J} \sum_{k>1} \frac{w_i^k w_j^k}{\lambda_k}
$$

$$
\langle s_i^{\mathrm{L}} s_j^{\mathrm{L}} \rangle = 1 - \frac{1}{J} \sum_{k>1} \frac{(w_i^k)^2 + (w_j^k)^2}{\lambda_k}
$$
(C4)

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. As discussed in Appendix A, this amounts to maximizing the log–likelihood of the experimental data given the model. This can be written simply as

$$
\left\langle \log P\left(\{\vec{s}_i\}\right) \right\rangle_{\exp} = -\log Z(J, n_c) + \frac{1}{2} J N n_c C_{\mathrm{int}}^{\exp} \,,
$$
(C5)

where, following the notation in the main text,

$$
C_{\mathrm{int}} = \frac{1}{N} \sum_i \frac{1}{n_c} \sum_{j \in n_c^i} \langle \vec{s}_i \cdot \vec{s}_j \rangle
$$
(C6)

indicates the degree of correlation up to the interaction range n_c.

Maximizing with respect to J, which corresponds to setting $C_{\mathrm{int}}(J, n_c) = C_{\mathrm{int}}^{\exp}$, gives immediately

$$
\frac{1}{J} = \frac{n_c}{2}\left(1 - C_{\mathrm{int}}^{\exp}\right) \,.
$$
(C7)

This equation provides an explicit relationship between J and n_c. Substituting into Eq (C5), the likelihood becomes a function of n_c only, and its maximum can be found numerically.

The values of J and n_c, obtained from this computation with free boundaries, are displayed in Fig 6, 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 $J n_c$ is approximately the same. On the contrary, the prediction for the perpendicular correlation as a function of distance (Fig 6C 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

$$
C_{\mathrm{int}} = C_{\mathrm{int}}^{\mathrm{P}} + S + \left(1 - \frac{1}{N} \sum_i \frac{1}{n_c} \sum_{j \in n_c^i} \langle |\vec{\pi}_j|^2 \rangle \right) \,,
$$
(C8)

where, we recall, S is the polarization. The first term in this decomposition of C_{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_{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 Figs 6C 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 6C is evidence that if we ignore these differences we really do fail to predict correctly the correlation structure of the flock as a whole.

Appendix D: 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.
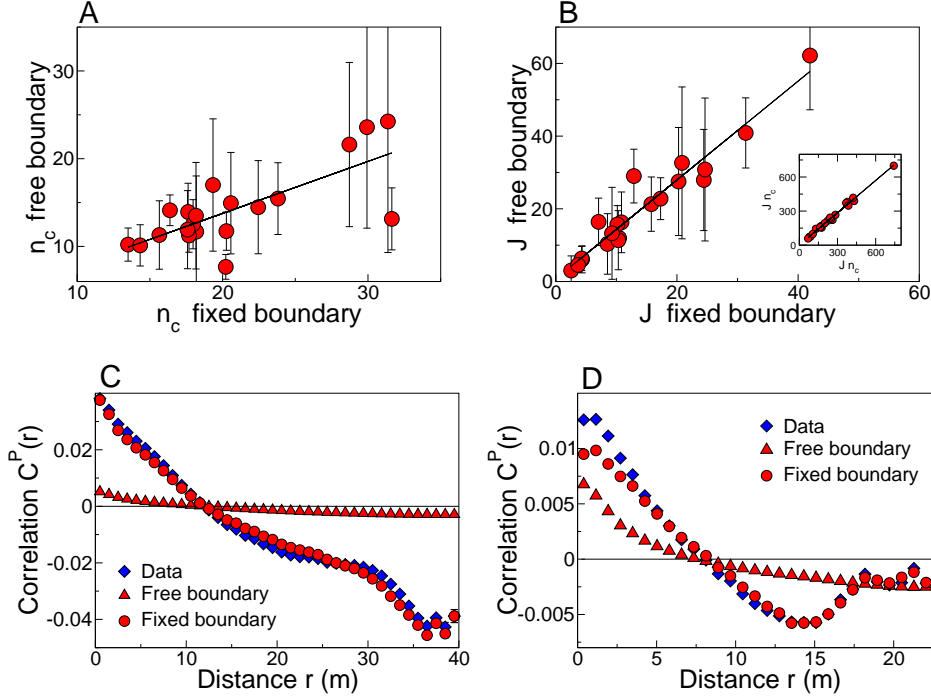
FIG. 6: Computation with free boundary conditions vs computation with fixed flight directions on the border. (A) Values of the parameter n_c for all the flocking events; the black line is the linear regression. Error bars are standard deviations across multiple snapshots of the same flock. (B) Values 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.

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 (B1), but the integration must be performed with respect to internal variables only. It is then convenient to separate, in the exponent of Eq (B1), 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 (B4):

$$
Z(\{J_{ij}\}; \mathcal{B}) = \int d^{\mathcal{I}} \vec{\pi} \; \delta \left(\sum_i \vec{\pi}_i \right) \exp \left[-\frac{1}{2} \sum_{i,j \in \mathcal{I}} A_{ij} \vec{\pi}_i \cdot \vec{\pi}_j + \sum_{i \in \mathcal{I}} \vec{\pi}_i \cdot \vec{h}_i + \frac{1}{2} \sum_{i,j \in \mathcal{I}} J_{ij} + \frac{1}{2} \sum_{i \in \mathcal{I}} h_i^L + \frac{1}{2} \sum_{i,j \in \mathcal{B}} J_{ij} \vec{s}_i \cdot \vec{s}_j \right],
$$
(D1)

where

$$
\vec{h}_i = \sum_{l \in \mathcal{B}} J_{il} \vec{s}_l = \sum_{l \in \mathcal{B}} J_{il} \left(s_l^L \vec{n} + \vec{\pi}_l \right) = h_i^L \vec{n} + \vec{h}_i^P
$$
(D2)

$$
A_{ij} = \delta_{ij} \left(\sum_{k \in \mathcal{I}} J_{ik} + h_i^L \right) - J_{ij} \qquad i, j \in \mathcal{I}
$$
(D3)

Here $\vec{h}_i$ 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^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 (B4), the matrix A is now defined only for internal birds and gets an additional diagonal contribution coming from individuals on the border. As a result, A has no longer 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 computa-

tional point of view, this implies that we cannot express in a simple way the constraint on the $\{\vec{\pi}_i\}$'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

$$\delta\left(\sum_i \vec{\pi}_i\right) = \int \frac{d\vec{z}}{(2\pi)^2} \exp\left[i\vec{z}\cdot\sum_i \vec{\pi}_i\right]. \tag{D4}$$

Substituting into Eq (D1), we obtain

$$Z(\{J_{ij}\};\mathcal{B}) = \int \frac{d\vec{z}}{(2\pi)^2} \int d^{\mathcal{I}}\vec{\pi} \ \exp\left[-\frac{1}{2}\sum_{i,j\in\mathcal{I}} A_{ij}\vec{\pi}_i\cdot\vec{\pi}_j + \sum_{i\in\mathcal{I}} \vec{\pi}_i\cdot\left(\vec{h}_i^P + i\vec{z}\right) + i\vec{z}\cdot\sum_{l\in\mathcal{B}} \vec{\pi}_l + G(\mathcal{B})\right], \tag{D5}$$

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

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

$$+\frac{1}{2}\sum_{ij\in\mathcal{I}}J_{ij} + \sum_{i\in\mathcal{I}}h_i^L + \sum_{lm\in\mathcal{B}}J_{lm}\vec{s}_l\cdot\vec{s}_m, \tag{D6}$$

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

$$\langle\vec{\pi}_i\rangle = \sum_{j\in\mathcal{I}}(A^{-1})_{ij}\vec{h}_i^P - \frac{\sum_{j\in\mathcal{I}}(A^{-1})_{ij}}{\sum_{kj\in\mathcal{I}}(A^{-1})_{kj}}\left[\sum_{l\in\mathcal{B}}\vec{\pi}_l + \sum_{kj\in\mathcal{I}}(A^{-1})_{kj}\vec{h}_j^P\right], \tag{D7}$$

and

$$\langle\vec{\pi}_i\cdot\vec{\pi}_j\rangle = \langle\vec{\pi}_i\rangle\cdot\langle\vec{\pi}_j\rangle + 2\left[(A^{-1})_{ij} - \frac{\sum_{kn\in\mathcal{I}}(A^{-1})_{ik}(A^{-1})_{nj}}{\sum_{kn\in\mathcal{I}}(A^{-1})_{kn}}\right]. \tag{D8}$$

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

$$\left\langle \log P\left(\{\vec{s}_i\}\right) \right\rangle_{exp} = -\log Z(J,n_c;\mathcal{B}) + \frac{1}{2}Jn_c NC_{int}^{exp}. \tag{D9}$$

with $Z(J,n_c;\mathcal{B})$ as in Eq (D1). 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, $C_{int}(J,nc;\mathcal{B}) = C_{int}^{exp}$. 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

$$\tilde{A} = A/J, \tag{D10}$$
$$\tilde{\vec{h}}_i = \vec{h}_i/J, \tag{D11}$$

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

$$\frac{(N_{in}-1)}{J} = \frac{1}{2}\sum_{ij\in\mathcal{I}}(\tilde{A}^{-1})_{ij}\tilde{\vec{h}}_i^P\cdot\tilde{\vec{h}}_j^P - \frac{1}{2}\frac{\left[\sum_{l\in\mathcal{B}}\vec{\pi}_l + \sum_{ij\in\mathcal{I}}(\tilde{A}^{-1})_{ij}\tilde{\vec{h}}_j^P\right]^2}{\sum_{ij\in\mathcal{I}}(\tilde{A}^{-1})_{ij}} + \sum_{i\in\mathcal{I}}\tilde{h}_i^L + \sum_{lm\in\mathcal{B}}n_{lm}\vec{s}_l\cdot\vec{s}_m + \frac{Nn_c}{2}(1-C_{int}^{exp}) \tag{D12}$$

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 (D9) 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 Figs 2 in the main text and in Fig 6. In this figure, we see the improvement in the prediction of the correlation function $C(r)$ that comes with fixed boundary conditions.

Appendix E: 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 [14, 17] 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

$$\phi_\alpha(J, n_c) = -\log Z(J, n_c; \mathcal{B}_\alpha) + \frac{1}{2} J n_c N C_{\text{int},\alpha}^{\exp} \quad \text{(E1)}$$

the log–likelihood of the snapshot α with parameters J and n_c (see Eq (D9)). Then, the average log–likelihood

for all the snapshots is

$$\Phi_{\text{global}}(J, n_c) = \frac{1}{N_{\text{snap}}} \sum_\alpha \phi_\alpha(J, n_c) , \quad \text{(E2)}$$

where N_{snap} is the number of snapshots available for that flock. At this point, we need to maximize Φ_{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_{global}, as a function of n_c:

$$\frac{1}{J_{\text{global}}} = \frac{1}{N_{\text{snap}}} \sum_\alpha \frac{(N_{\text{in}}^\alpha - 1)}{(N_{\text{in}}^{\text{global}} - 1)} \frac{1}{J_\alpha(n_c)} \quad \text{(E3)}$$

where $J_\alpha(n_c)$ is the optimal value of J for the snapshot α, as above, N_{in}^α is the number of internal individuals in the snapshot α, and $N_{\text{in}}^{\text{global}}$ is the average over snapshots. Substituting $J_{\text{global}}(n_c)$ back in Eq (E2) 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 $n_{c,\text{global}}$ and J_{global} obtained in this way are plotted in Fig 7, 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.
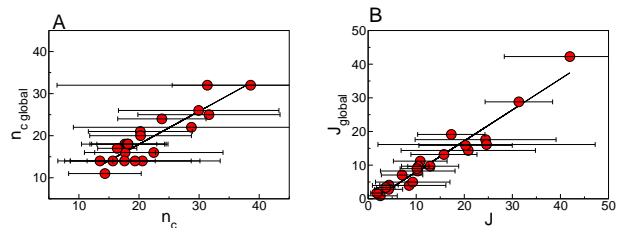

FIG. 7: Global models of flocking events. (A) Values of the neighborhood size n_c inferred from maximizing the log–likelihood averaged over snapshots, plotted vs. the mean values obtained from maximizing log–likelihood in individual snapshots. Error bars represent the standard deviation over snapshots 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.

Appendix F: 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 Figs 3, 6 and 8. 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. 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.
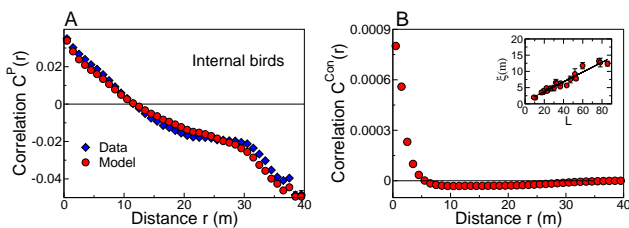

FIG. 8: Correlations in the interior of the flock. (A) Perpendicular component of the two–point correlation function (as in Fig 3A) for internal birds only, as a function of distance. (B) Connected correlation function predicted in the model, as a function of distance. Inset: correlation length vs. flock size, for all the flocks that we analyzed.

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 (D7) and (D8). 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 $\langle s_i^L\rangle = 1 - 0.5\langle|\vec{\pi}_i|^2\rangle$, if we are getting the perpendicular components of the velocity right we must also be getting the longitudinal

components right. Equations (D7) and (D8) 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:

$$\langle\vec{\pi}_i\rangle = \sum_{j\in\mathcal{I}} C_{ij}^{con}\vec{h}_j^P - \frac{\sum_{j\in\mathcal{I}}(A^{-1})_{ij}}{\sum_{kj\in\mathcal{I}}(A^{-1})_{kj}}\sum_{l\in\mathcal{B}}\vec{\pi}_l \quad \text{(F1)}$$

$$\langle\vec{\pi}_i\cdot\vec{\pi}_j\rangle = C_{ij}^{con} + \langle\vec{\pi}_i\rangle\cdot\langle\vec{\pi}_j\rangle \quad \text{(F2)}$$

$$C_{ij}^{con} = 2\left[A_{ij}^{-1} - \frac{\sum_{kn\in\mathcal{I}}A_{ik}^{-1}A_{nj}^{-1}}{\sum_{kn\in\mathcal{I}}A_{kn}^{-1}}\right] \quad \text{(F3)}$$

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

In Eq (F1) 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 (F1) 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}^{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 $C_{ij}^{con}\vec{h}_j$ are nonzero). If the connected correlations extend over sufficiently long distances, this mechanism ensures propagation of directional information trough the whole flock.

In Fig 8B 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^{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.

Appendix G: 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,

$$\langle\vec{s}_i\cdot\vec{s}_j\rangle = \langle s_i^L s_j^L\rangle + \langle\vec{\pi}_i\cdot\vec{\pi}_j\rangle \quad \text{(G1)}$$

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, $(1/N)\sum_i \vec{\pi}_i = 0$, the same is not true for the longitudinal direction. Rather, we have $(1/N)\sum_i s_i^{\mathrm{L}} = S$, 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:

$$C_{\mathrm{ij}}^{\mathrm{P}} = \langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle \qquad (G2)$$

$$
\begin{aligned}
C_{\mathrm{ij}}^{\mathrm{L}} &= \langle (s_i^{\mathrm{L}} - S)(s_j^{\mathrm{L}} - S) \rangle \\
&= \langle (1 - S - \tfrac{1}{2}\langle \pi_i^2 \rangle)(1 - S - \tfrac{1}{2}\langle \pi_j^2 \rangle) \rangle
\end{aligned}
\qquad (G3)
$$

$$C_{\mathrm{ij}} = C_{\mathrm{ij}}^{\mathrm{P}} + C_{\mathrm{ij}}^{\mathrm{L}} . \qquad (G4)$$

The last identity in Eq (G3) is a consequence of the spin wave approximation.

Connected correlations. In Appendix D we have described a theory where we get nonzero expectation values for the flight directions of individual birds, $\langle \vec{\pi}_i \rangle \neq 0$. 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 Appendix F the following connected correlation function

$$C_{\mathrm{ij}}^{\mathrm{con}} = \langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle - \langle \vec{\pi}_i \rangle \cdot \langle \vec{\pi}_j \rangle . \qquad (G5)$$

We note that in our case $C_{\mathrm{ij}}^{\mathrm{con}}$ 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 $\langle \vec{\pi}_i \cdot \vec{\pi}_j \rangle$.

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

$$C_{\mathrm{int}} = \frac{1}{N}\sum_i \frac{1}{n_c} \sum_{j \in n_c^i} \langle \vec{s}_i \cdot \vec{s}_j \rangle, \qquad (G6)$$

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
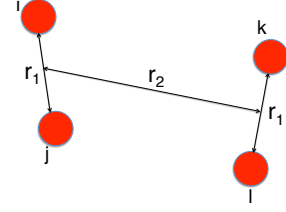

FIG. 9: Sketch of the structure of the four point correlation function. Red circles represent birds. Birds i and j have mutual distance $r_{\mathrm{ij}} = r_1$; birds k and l also have mutual distance $r_{\mathrm{kl}} = r_1$. The distance between the mid-point of the ij pair and the midpoint of the kl pair is r_2.

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

$$C(r) = \frac{\sum_{\mathrm{ij}} C_{\mathrm{ij}} \delta(r_{\mathrm{ij}} - r)}{\sum_{\mathrm{ij}} \delta(r_{\mathrm{ij}} - r)} , \qquad (G7)$$

where $r_{\mathrm{ij}} = |\vec{\mathbf{r}}_i - \vec{\mathbf{r}}_j|$ 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,

$$C(r) = C^{\mathrm{L}}(r) + C^{\mathrm{P}}(r), \qquad (G8)$$

describing the contributions relative to, respectively, longitudinal and perpendicular fluctuations in the flight directions. Figure 3 in the main text and Fig 8 in the Appendix 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_{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 (G7), 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 (G7), but where we only count contributions from individuals inside the flock ($i, j \in \mathcal{I}$); the result is in Fig 8A. 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

$$C_4(r_1; r_2) = \frac{\sum_{ijkl}\langle(\vec{\pi}_i \cdot \vec{\pi}_j)(\vec{\pi}_i \cdot \vec{\pi}_j)\rangle\Delta_{ijkl}}{\sum_{ijkl}\Delta_{ijkl}}, \qquad (G9)$$

$$\Delta_{ijkl} = \delta(r_{ij} - r_1)\delta(r_{kl} - r_1)\delta(r_{ij-kl} - r_2) \qquad (G10)$$

where r_{ij-kl} indicates the distance between the midpoints of the pairs ij and pair kl; see Fig 9. 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^{L} is nothing else than a particular case of the four–point correlation,

$$C^{\mathrm{L}}(r) = 1 - C_4(0; r) - S^2. \qquad (G11)$$

[1] S Camazine, J–L Deneubourg, NR Franks, J Sneyd, G Theraulaz & E Bonabeau. *Self-Organization in Biological Systems.* (Princeton University Press, Princeton, 2003).

[2] PW Anderson, More is different. *Science* **177,** 393–396 (1972).

[3] KG Wilson, Problems in physics with many scales of length. *Sci Am* **241**, 158–179 (1979)

[4] ID Couzin & J Krause, Self-organization and collective behaviour in vertebrates.*Adv Study Behav* **32**, 1–75 (2003)

[5] J Binney, NJ Dowrick, AJ Fisher & MEJ Newman. *The Theory of Critical Phenomena* (Oxford University Press, New York, 1992).

[6] J Toner & Y Tu, Long–range order in a two–dimensional dynamical XY model: How birds fly together. *Phys Rev Lett* **75,** 4326–4329 (1995).

[7] T Vicsek, A Czirók, E Ben–Jacob, I Cohen & O Shochet, Novel type of phase transition in a system of self–driven particles. *Phys Rev Lett* **75,** 1226–1229 (1995).

[8] J Toner & Y Tu, Flocks, herds, and schools: A quantitative theory of flocking. *Phys Rev E* **58,** 4828–4858 (1998).

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

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

[11] M Ballerini, N Cabibbo, R Candelier, A Cavagna, E Cisbani, I Giardina, A Orlandi, G Parisi, A Procaccini, M Viale & V Zdravkovic, Empirical investigation of starling flocks: A benchmark study in collective animal behaviour. *Animal Behaviour* **76,** 201–215 (2008); arXiv:0802.1667.

[12] A Cavagna, I Giardina, A Orlandi, G Parisi, A Procaccini, M Viale & V Zdravkovic, The STARFLAG handbook on collective animal behaviour: 1. Empirical methods, *Animal Behaviour* **76,** 217–236 (2008); arXiv:0802.1668.

[13] 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); arXiv:0802.1674.

[14] 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); arXiv:07091916 (2007).

[15] S Cocco & R Monasson, Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys Rev Lett* **106,** 090601 (2011); arXiv.org:1102.3260 [physics.data–an].

[16] S Cocco, R Monasson & V Sessak, High–dimensional inference with the generalized Hopfield model: Principal component analysis and corrections. *Phys Rev E* in press; arXiv.org:1104.3665 [cond–mat.stat–mech] (2011).

[17] A Cavagna, A Cimarelli, I Giardina, G Parisi, R Santigati, F Stefanini & M Viale, Scale–free correlations in starling flocks. *Proc Nat'l Acad Sci (USA)* **107,** 11865–11870 (2010); arXiv:0911.4393 (2009).

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

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

[20] 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).

[21] 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), arXiv:q-bio/0512013v1.

[22] G Tkačik, E Schneidman, MJ Berry II & W Bialek,

Ising models for networks of real neurons. arXiv:q-bio/0611072 (2006).

[23] 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).

[24] 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).

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

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

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

[28] 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).

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

[30] T Mora, AM Walczak, W Bialek & CG Callan, Maximum entropy models for antibody diversity. *Proc Nat'l Acad Sci (USA)* **107,** 5405–5410 (2010); arXiv:0912.5175 [q–bio.GN] (2009).

[31] 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).

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

[33] GJ Stephens & W Bialek, Statistical mechanics of letters in words. *Phys Rev E* **81,** 066119 (2010); arXiv:0801.0253 [q–bio.NC] (2008).

[34] T Mora & W Bialek, Are biological systems poised at criticality? *J Stat Phys* in press (2011); arXiv:1012.2242 [q–bio.QM] (2010).

[35] 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).

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