Local equilibrium in bird flocks: Supplementary Information

Thierry Mora[1], Aleksandra M. Walczak[2], Lorenzo Del Castello[3,4], Francesco Ginelli[5], Stefania Melillo[3,4], Leonardo Parisi[6,4], Massimiliano Viale[3,4], Andrea Cavagna[4], Irene Giardina[3,4]

[1] *Laboratoire de physique statistique, CNRS, UPMC and École normale supérieure, 24, rue Lhomond, Paris, France*
[2] *Laboratoire de physique théorique, CNRS, UPMC and École normale supérieure, 24, rue Lhomond, Paris, France*
[3] *Dipartimento di Fisica, Università Sapienza, Rome, Italy*
[4] *Istituto Sistemi Complessi, Consiglio Nazionale delle Ricerche, UOS Sapienza, Rome, Italy*
[5] *SUPA, Institute for Complex Systems and Mathematical Biology, Kings College, University of Aberdeen, Aberdeen, UK and*
[6] *Dipartimento di Informatica, Università Sapienza, Rome, Italy*

Appendix A: Dynamical maximum entropy model

Call $\vec{s}_i(t)$ the d-dimensional flight orientation of bird i as a function of time, of unit norm $\|\vec{s}\| = 1$. We look for a probability disribution over whole flock trajectories, $(\vec{s}_1(t), \dots, \vec{s}_N(t))$, that has maximum entropy, but with the constraints that the correlation functions:

$$\langle \vec{s}_i(t) \cdot \vec{s}_j(t) \rangle \tag{A1}$$

and

$$\left\langle \frac{d\vec{s}_i(t)}{dt} \cdot \vec{s}_j(t) \right\rangle \tag{A2}$$

agree with the data. After time discretization, these constraints are equivalent to imposing the values of $\langle \vec{s}_i(t) \cdot \vec{s}_j(t) \rangle$ and $\langle \vec{s}_i(t+dt) \cdot \vec{s}_j(t) \rangle$, with dt an infinitesimal increment. Using the technique of Lagrange multipliers, one can show that the distribution over trajectories then takes the form [1, 2]:

$$P(\{\vec{s}_i(t)\}) = \frac{1}{\mathcal{Z}} \exp\left(\sum_{ij,t} J^{(1)}_{ij;t} \vec{s}_i(t) \cdot \vec{s}_j(t) + \sum_{ij,t} J^{(2)}_{ij;t} \vec{s}_i(t+dt) \cdot \vec{s}_j(t) \right) \prod_{i,t} \delta(\|\vec{s}_i(t)\| - 1) \tag{A3}$$

where sums and products over t run over a discrete set of times separated by dt, and where $\delta(\cdot)$ denotes the Dirac-delta function.

In [3], it was shown that, in the spin-wave approximation, the stochastic process described by this probability distribution is equivalent to a random walk:

$$\vec{s}_i(t) = \frac{\sum_j M_{ij;t} \vec{s}_j(t) + \vec{\eta}_i(t)}{\| \sum_j M_{ij;t} \vec{s}_j(t) + \vec{\eta}_i(t) \|}, \tag{A4}$$

with $\eta_i(t)$ is a Gaussian variable of zero mean and covariance $\langle \eta_i(t) \cdot \eta_j(t') \rangle = d(A_t^{-1})_{ij} \delta_{t,t'}$. The matrices $M_{ij;t}$ and $A_{ij;t}$ can be expressed in terms of the matrices $J^{(1)}_{ij;t}$ and $J^{(2)}_{ij;t}$. In order to take the limit $dt \to 0$, the matrices need reparametrizing as:

$$M_{ij;t} = \delta_{ij} + dt\, J_{ij;t} \tag{A5}$$

$$(A_t^{-1})_{ij} = dt\, X_{ij;t}. \tag{A6}$$

Then the random walk reduces to the Langevin equation:

$$\frac{d\vec{s}_i}{dt} = -\vec{s}_i \times \left(\vec{s}_i \times \left(\sum_j J_{ij}(t) \vec{s}_j + \vec{\xi}_i \right) \right) \tag{A7}$$

where $J_{ij}(t)$ denotes the influence of bird j on bird i's orientation, and $\vec{\xi}(t)$ is a Gaussian random d-dimensional noise with $\langle \vec{\xi}_i(t) \vec{\xi}_j(t') \rangle = dX_{ij}(t) \delta(t - t')$. To simplify, we assume that $X_{ij}(t) = 2T\delta_{ij}$; T quantifies the noise in alignment, and can be mapped onto a temperature, as we'll see later. In the following, for ease of notation we drop the dependency of J_{ij} on t.

The triple cross-product is easier to understand if we note that, for any vector $\vec{a}$, this cross-product reduces to

$$-\vec{s} \times (\vec{s} \times \vec{a}) = \vec{a} - (\vec{s} \cdot \vec{a}) \vec{s} \equiv \vec{a}_\perp, \tag{A8}$$

which is just the projection of $\vec{a}$ onto the hyperplane orthogonal to $\vec{s}$. Since $\vec{s}_i$ lives on the unit sphere, its variations must be perpendicular to itself. The triple cross-product just implements this projection by subtracting the parallel part. This projection ensures the conservation of the norm:

$$\frac{d\|\vec{s}_i\|^2}{dt} = 2\vec{s}_i \cdot \frac{d\vec{s}_i}{dt} = 0. \tag{A9}$$

The norm of $\vec{s}_i$ stays constant and equal to one.

We rewrite $J_{ij} = J n_{ij}$, where J quantifies the aligning strength, and n_{ij} how j is taken into account by i (n_{ij} does not have to be an integer). J has the dimension of an inverse time, n_{ij} is dimensionless. Since anything inside the parentheses of Eq. A7 that is parallel to $\vec{s}_i$ is discarded, we can rewrite it as:

$$\frac{d\vec{s}_i}{dt} = J\vec{s}_i \times \left(\vec{s}_i \times \left(\sum_j \Lambda_{ij} \vec{s}_j \right) \right) + \vec{\xi}_{i\perp} \tag{A10}$$

where we have denoted $\Lambda_{ij} = \sum_k n_{ik}\delta_{ij} - n_{ij}$, and where now $\langle \vec{\xi}_{i\perp}(t) \vec{\xi}_{j\perp}(t') \rangle = 2(d-1)T\delta_{ij}\delta(t-t')$. The $(d-1)$ factor replaces d because of the projection of the noise term onto the hyperplane orthogonal to $\vec{s}_i$. The diagonal term in Λ_{ij} was chosen so as to balance each row of the matrix ($\sum_j \Lambda_{ij} = 0$).

Event ID	N	T (s)	P	v_0 (m/s)	r_0 (m)
20110208_ACQ3	179	5.5	0.984	8.7	0.85
20110211_ACQ1	595	4.5	0.971	8.5	0.95
20110217_ACQ2	407	2.1	0.986	11.0	0.70
20111124_ACQ1	125	1.8	0.993	11.1	0.66
20111125_ACQ1	50	5.6	0.987	12.4	1.21
20111125_ACQ2	530	4.4	0.957	9.2	0.85
20111201_ACQ3_1	137	2.9	0.987	10.1	0.74
20111201_ACQ3_4	489	2.3	0.9763	10.5	0.74
20111214_ACQ4_1	157	2.9	0.993	11.4	0.74
20111214_ACQ4_2	162	4.1	0.973	11.6	1.08
20111215_ACQ1	401	5.7	0.987	11.0	0.82
20111220_ACQ2	200	1.7	0.984	16.2	0.62
20111222_ACQ1	59	3.5	0.984	11.7	1.24
20120209_ACQ1	412	3.5	0.997	29.2	0.80

TABLE S1. Summary of the data used in the analysis. N is the number of birds, T the duration of the film, $P = (1/N)\|\sum_i \vec{s}_i\|$ the polarization of the flock, v_0 the average bird velocity, and r_0 the average interbird distance. The event ID contains its date and its acquisition index.

There is a link with the statistical description of flock configurations inferred in [4]. If Λ_{ij} is symmetric and constant in time, the steady-state probability distribution of the set of $(\vec{s}_1, \dots, \vec{s}_N)$ is given by the Boltzmann distribution

$$P(\vec{s}_1, \dots, \vec{s}_N) \propto \exp\left[-\frac{H(s)}{T}\right] \tag{A11}$$

with Hamiltonian:

$$H(s) = -\frac{J}{2}\sum_{ij} n_{ij}\vec{s}_i\vec{s}_j. \tag{A12}$$

We can expand Eq. A10 within the spin-wave approximation. In this limit, all vectors $\vec{s}_i$ almost point in a common direction, denoted by $\vec{n}$, so that we can write $\vec{s}_i = \vec{\pi}_i + \sqrt{1-\vec{\pi}_i^2}\vec{n} \approx \vec{\pi}_i + (1-\vec{\pi}_i^2/2)\vec{n}$, where $\vec{\pi}_i$ is the projection of $\vec{s}_i$ onto the hyperplane orthogonal to $\vec{n}$: $\vec{n}\cdot\vec{\pi}_i = 0$. Expanding at first order yields:

$$\frac{d\vec{\pi}_i}{dt} = -J\sum_j \Lambda_{ij}\vec{\pi}_j + \vec{\xi}_{i\perp}. \tag{A13}$$

In practice, this is the equation we will use for the inference.

Appendix B: Inference from data

1. Static inference

We start by recalling how to do the steady-state inference based on the steady-state distribution of Eqs. A11 and A12. We assume that the flock is very polarized, so that the spin-wave approximation is valid. In this approximation, the steady-state distribution reads:

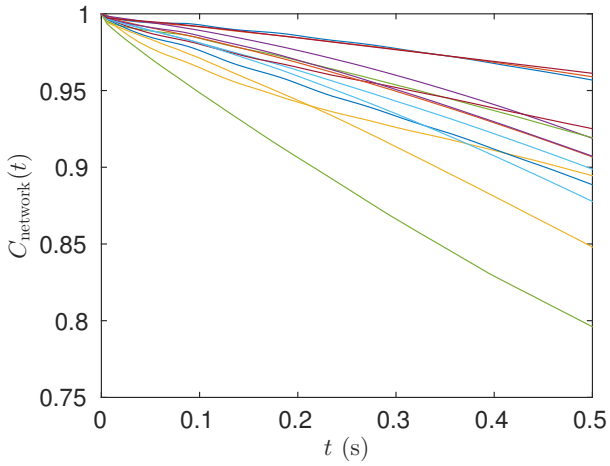

FIG. S1. Normalized autocorrelation function of the network for all 14 flocking events. The decay is approximately exponential, allowing for the definition of a characteristic decay time τ_{relax} for each event.

$$P(\vec{\boldsymbol{\pi}}|\vec{n}) = \frac{1}{Z}\exp\left(-\frac{J}{2T}\sum_{ij}\Lambda_{ij}\vec{\pi}_i\vec{\pi}_j\right)\delta\left(\sum_i \vec{\pi}_i\right) \tag{B1}$$

where the common direction $\vec{n}$ is chosen so that $\sum_i \vec{\pi}_i = \vec{0}$, and where for simplicity n_{ij} is assumed to be symmetric. Integrating over $\vec{\boldsymbol{\pi}}$ satisfying that condition gives the normalization constant:

$$Z = \left(\frac{2\pi T}{J}\right)^{(N-1)(d-1)/2}\prod_{\lambda_k>0}\lambda_k^{-(d-1)/2} \tag{B2}$$

where λ_k are the eigenvalues of the matrix Λ_{ij}. Since $\sum_j \Lambda_{ij} = 0$ for all i, we know that one of these eigenvalues is 0. It is the one corresponding to variations along the direction $(1, \dots, 1)$. These variations are entirely suppressed by the condition $\sum_i \vec{\pi}_i = 0$, and this direction does not contribute to the Gaussian integral, hence the condition $\lambda_k > 0$.

In summary, the minus-log-likelihood of the data reads:

$$-\ln P(\vec{\boldsymbol{\pi}}|\vec{n}) = \frac{J}{2T}\mathrm{Tr}(\mathbf{C}\boldsymbol{\Lambda}^\dagger) - \frac{(d-1)(N-1)}{2}\ln\left(\frac{J}{T}\frac{1}{2\pi}\right) - \frac{d-1}{2}\sum_{\lambda_k>0}\ln\lambda_k, \tag{B3}$$

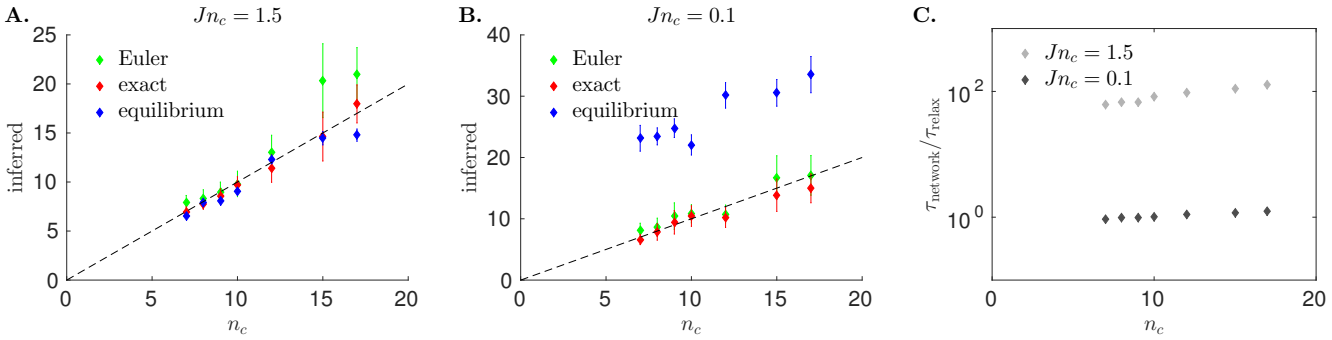

FIG. S2. Simulations of fast versus slow relaxation. **A.** Inferred interaction range n_c using dynamical Euler (green), dynamical exact integration (red), or equilibrium-like inference (blue), versus the true n_c for fast relaxation dynamics relative to network rearrangement. The parameters are: $Jn_c = 1.5$, $\sqrt{2T} = 0.15$, bird speed $v_0 = 1$, unit bird density (512 birds in an 8 x 8 x 8 box with periodic boundary conditions), inference $dt = 0.2$. Polarization is ≈ 0.99. The equilibrium inference gives the same result as the dynamical one, since the orientation dynamics is fast compared to network reshuffling. **B.** Same as A., but with slow relaxation of orientations. The parameters are chosen to keep a similar polarization of 0.99: $Jn_c = 0.1$, $\sqrt{2T} = 0.05$, bird speed $v_0 = 1$, unit bird density, inference $dt = 1$. The equilibrium inference systematically overestimates the true n_c, while the dynamical inferences predict it accurately. **C.** Comparison of $\tau_{\rm network}$ and $\tau_{\rm relax}$ in the two simulations of A. and B. The relaxation time $\tau_{\rm relax}$ is taken to be $1/(Jn_c)$, while $\tau_{\rm network}$ is estimated as explained in the main text, by fitting an exponential decay to the overlap autocorrelation function, as in Fig. S1.

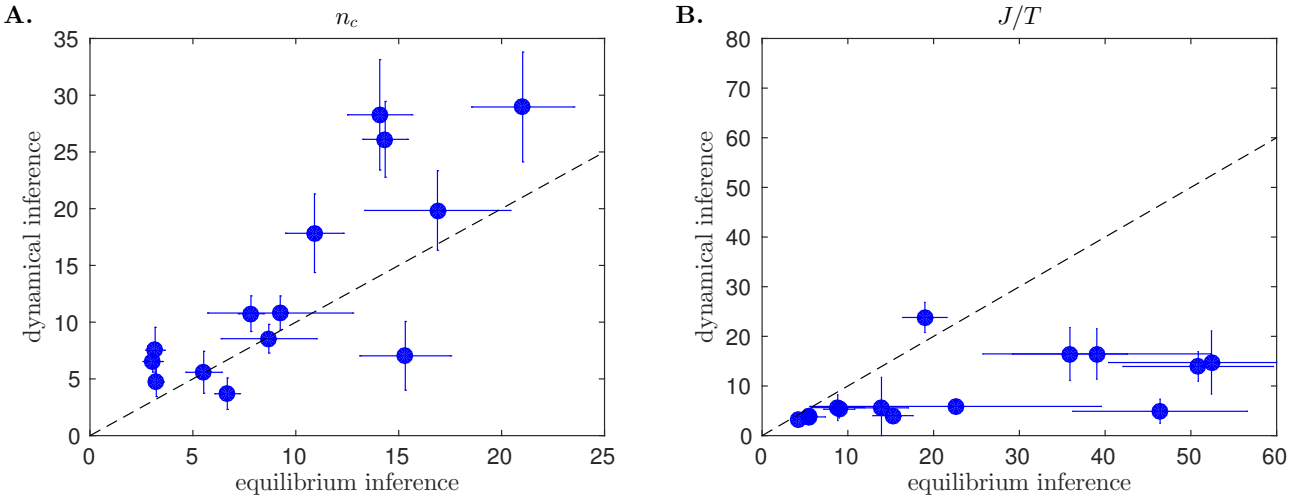

FIG. S3. Comparison between the equilibrium inference method (abcissa) and the dynamical inference method using Euler's rule (ordinate), for (**A**) the interaction range n_c and (**B**) the interaction parameter J/T. The agreement is relatively poor, especially for the prediction of J/T.

where $\mathbf{C} = \vec{\boldsymbol{\pi}}\vec{\boldsymbol{\pi}}^\dagger$.

We want to minimize this quantity according to the principle of maximum likelihood. Taking the derivative with respect to J/T gives:

$$(J/T)^* = \frac{(d-1)(N-1)}{\mathrm{Tr}(\mathbf{C}\boldsymbol{\Lambda}^\dagger)} \approx \frac{d-1}{C_{\rm int}} \tag{B4}$$

with the definition $C_{\rm int} = (1/N)\mathrm{Tr}(\mathbf{C}\boldsymbol{\Lambda}^\dagger)$.

Replacing into Eq. B3 gives:

$$-\ln P(\vec{\boldsymbol{\pi}}|\vec{n}, (J/T)^*) = \frac{(d-1)(N-1)}{2}\left[1 + \ln C_{\rm int} + \ln(2\pi/(d-1))\right] - \frac{d-1}{2}\sum_{\lambda_k>0}\ln\lambda_k. \tag{B5}$$

Finally, this quantity must be minimized over the parameters defining Λ_{ij}, or equivalently, ignoring the constants

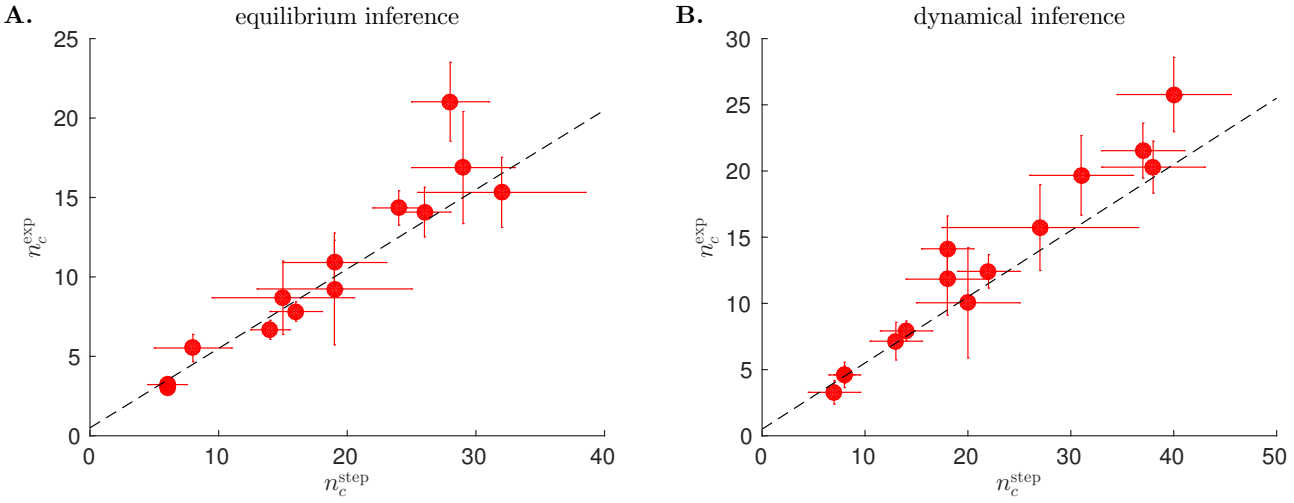

FIG. S4. Comparison of the interaction range n_c inferred assuming a step-function interaction function ($n_c^{\rm step}$, abscissa) or an exponentially decaying interaction function ($n_c^{\rm exp}$, ordinate), using (**A**) the equilibrium inference method and (**B**) the dynamical inference method. We expect a correspondance between $n_c^{\rm step}$ and $n_c^{\rm exp}$: $n_c^{\rm exp} = n_c^{\rm step}/2$. Here this correspondance is verified for both inference methods.

and prefactors:

$$\ln C_{\rm int} - \frac{1}{N-1}\sum_{\lambda_k>0}\ln\lambda_k. \tag{B6}$$

2. Dynamical inference using Euler's method

We now move to the dynamical inference from data using Eq. A13. Let us start by assuming that we have a series of data points separated by a small dt. We can write Euler's approximation to the stochastic differential equation:

$$\vec{\pi}_i(t+dt) = \vec{\pi}_i(t) - Jdt\sum_j \Lambda_{ij}\vec{\pi}_j + \vec{\epsilon}_i \tag{B7}$$

where $\vec{\epsilon}_i$ is Gaussian noise of variance $2(d-1)Tdt$.

Or, in matrix form:

$$\vec{\boldsymbol{\pi}}(t+dt) = \vec{\boldsymbol{\pi}}(t) - Jdt\boldsymbol{\Lambda}\vec{\boldsymbol{\pi}} + \vec{\boldsymbol{\epsilon}}. \tag{B8}$$

Let us denote $\vec{\boldsymbol{\pi}}' = \vec{\boldsymbol{\pi}}(t+dt)$. Then the probability of $\vec{\boldsymbol{\pi}}'$ given $\vec{\boldsymbol{\pi}}$ is:

$$P(\vec{\boldsymbol{\pi}}'|\vec{\boldsymbol{\pi}}) = (4\pi Tdt)^{-N(d-1)/2}\exp\left[-\frac{1}{4Tdt}(\vec{\boldsymbol{\pi}}' - \vec{\boldsymbol{\pi}} + Jdt\boldsymbol{\Lambda}\vec{\boldsymbol{\pi}})^2\right]. \tag{B9}$$

The associated minus-log-likelihood, $\mathcal{L} = -\ln P(\vec{\boldsymbol{\pi}}'|\vec{\boldsymbol{\pi}})$, is thus given by:

$$\mathcal{L} = N\frac{d-1}{2}\ln(4\pi Tdt) + \frac{1}{4Tdt}{\rm Tr}\left[\mathbf{C}' + \mathbf{C} - 2\mathbf{G} + 2Jdt(\mathbf{G}-\mathbf{C})\boldsymbol{\Lambda}^\dagger + (Jdt)^2\boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}^\dagger\right], \tag{B10}$$

where $\mathbf{C} = \vec{\boldsymbol{\pi}}\vec{\boldsymbol{\pi}}^\dagger$, $\mathbf{C}' = \vec{\boldsymbol{\pi}}'\vec{\boldsymbol{\pi}}'^\dagger$ and $\mathbf{G} = \vec{\boldsymbol{\pi}}'\vec{\boldsymbol{\pi}}^\dagger$. Or, in short-hand:

$$\frac{\mathcal{L}}{N} = \frac{d-1}{2}\ln(4\pi Tdt) + \frac{1}{4Tdt}\left[C'_{\rm s} + C_{\rm s} - 2G_s + 2Jdt(G_{\rm int} - C_{\rm int}) + (Jdt)^2 C_{\rm int^2}\right] \tag{B11}$$

$$\equiv \frac{d-1}{2}\ln(4\pi Tdt) + \frac{\hat{\mathcal{L}}}{4Tdt}, \tag{B12}$$

with $C'_{\rm s} = {\rm Tr}(\mathbf{C}')/N$, $C_{\rm s} = {\rm Tr}(\mathbf{C})/N$, $G_{\rm s} = {\rm Tr}(\mathbf{G})/N$, $G_{\rm int} = {\rm Tr}(\mathbf{G}\boldsymbol{\Lambda}^\dagger)/N$, $C_{\rm int} = {\rm Tr}(\mathbf{C}\boldsymbol{\Lambda}^\dagger)/N$, and $C_{\rm int^2} = {\rm Tr}(\boldsymbol{\Lambda}\mathbf{G}\boldsymbol{\Lambda}^\dagger)/N$

Following the principle of maximum likelihood, which is equivalent to solving the inverse maximum entropy model in the spin-wave approximation, we minimize this quantity over the parameters J, T, and the parameters of Λ_{ij}. Let us start with the temperature T. $\partial\mathcal{L}/\partial T = 0$

gives:

$$T^* = \frac{\hat{\mathcal{L}}}{2(d-1)dt}. \tag{B13}$$

We can now minimize $\mathcal{L}$ taken at that value of $T = T^*$,

$$\frac{\mathcal{L}(T^*)}{N} = \frac{d-1}{2}\left[1 + \ln\hat{\mathcal{L}} + \ln(2\pi/(d-1))\right]. \tag{B14}$$

In other words, we want to minimize $\hat{\mathcal{L}}$ over the remaining parameters J and n_c. Writing the condition for J, $\partial\hat{\mathcal{L}}/\partial J = 0$ gives:

$$J^* = \frac{C_{\rm int} - G_{\rm int}}{dt C_{\rm int^2}}. \tag{B15}$$

And replacing into $\hat{\mathcal{L}}$ gives:

$$\hat{\mathcal{L}}(J^*) = C'_{\rm s} + C_{\rm s} - 2G_s - \frac{(G_{\rm int} - C_{\rm int})^2}{C_{\rm int^2}}. \tag{B16}$$

The first three terms do not depend on the choice of Λ. The last step is to maximize $(G_{\rm int} - C_{\rm int})^2/C_{\rm int^2}$ over the paramters defining Λ_{ij}.

3. Dynamical inference using exact integration

In general n_{ij} and Λ_{ij} may depend on time, because they will evolve with the local neighbours of each birds. But on short time scales such that neighbours do not change significantly, we can view them as constant. If on this time scale the main direction of the flock has not changed much, we can consider Eq. A13 as valid with constant Λ_{ij}. This linear stochastic equation can actually be solved analytically:

$$\vec{\boldsymbol{\pi}}(t+dt) = e^{-J\boldsymbol{\Lambda}dt}\vec{\boldsymbol{\pi}}(t) + \int_0^{dt} du\, e^{-J\boldsymbol{\Lambda}(dt-u)}\vec{\boldsymbol{\xi}}_\perp(t+u). \tag{B17}$$

We define the integrated noise term as:

$$\vec{\boldsymbol{\epsilon}} = \int_0^{dt} du\, e^{-J\boldsymbol{\Lambda}(dt-u)}\vec{\boldsymbol{\xi}}_\perp(t+u). \tag{B18}$$

Since it is a sum of Gaussian variables, $\vec{\boldsymbol{\epsilon}}$ is also Gaussian, of mean zero and covariance:

$$\langle\vec{\boldsymbol{\epsilon}}\vec{\boldsymbol{\epsilon}}^\dagger\rangle = 2(d-1)T\int_0^{dt} du\, e^{-J\boldsymbol{\Lambda}u}e^{-J\boldsymbol{\Lambda}^\dagger u} \tag{B19}$$

In the limit $dt \to 0$, we recover Euler's approximation, Eq. B7.

With this new, exact integration formula, we can write the minus-log-likelihood:

$$\mathcal{L} = N\frac{d-1}{2}\ln(4\pi T dt) + \frac{d-1}{2}\ln\det\mathbf{B} + N\frac{\hat{\mathcal{L}}}{4Tdt}, \tag{B20}$$

with:

$$\hat{\mathcal{L}} = \frac{1}{N}\mathrm{Tr}\left[\mathbf{C}'\mathbf{A} - 2\mathbf{G}e^{-J\boldsymbol{\Lambda}^\dagger dt}\mathbf{A} + e^{-J\boldsymbol{\Lambda}dt}\mathbf{C}e^{-J\boldsymbol{\Lambda}^\dagger dt}\right], \tag{B21}$$

$$\mathbf{A} = \mathbf{B}^{-1} \quad \text{and} \quad \mathbf{B} = \frac{1}{dt}\int_0^{dt} du\, e^{-J\boldsymbol{\Lambda}u}e^{-J\boldsymbol{\Lambda}^\dagger u}. \tag{B22}$$

As before, we can solve for T easily:

$$T^* = \frac{\hat{\mathcal{L}}}{2(d-1)dt}, \tag{B23}$$

yielding:

$$\frac{\mathcal{L}(T^*)}{N} = \frac{d-1}{2}\left[1 + \ln\hat{\mathcal{L}} + \frac{1}{N}\ln\det\mathbf{B} + \ln(2\pi/(d-1))\right]. \tag{B24}$$

Note that now $\mathbf{A}$ and therefore $\mathbf{B}$ depend on J as well as Λ_{ij}. The sum $[\ln\hat{\mathcal{L}} + (1/N)\ln\det\mathbf{B}]$ must be minimized numerically with respect to both J and the parameters defining $\boldsymbol{\Lambda}$.

4. Two parametrizations for n_{ij}

We now need to specify the matrix Λ_{ij}. Here we only consider topological distance for the interaction matrix. Let us denote k_{ij} the rank of j among the neighbors of i, from the closest in distance to the farthest.

In the first parametrization, already used in previous work, we say that a bird interacts with its $n_c^{\rm step}$ closest neighbours. This corresponds to:

$$\text{step:} \qquad n_{ij} = \Theta(n_c^{\rm step} - k_{ij}), \tag{B25}$$

where $\Theta(x) = 1$ if $x \geq 0$ and 0 otherwise. Numerically, J^* is calculated for each integer value of $n_c^{\rm step}$ using a simple iterative 1D optimization algorithm.

In the second parametrization, we assume an exponentially decaying interaction as a function of rank:

$$\text{exp:} \qquad n_{ij} = \exp(-k_{ij}/n_c^{\rm exp}). \tag{B26}$$

Numerically, we implement a 1D iterative optimization algorithm for $n_c^{\rm step}$, where $J^*(n_c^{\rm exp})$ is calculated for each $n_c^{\rm step}$ as before, in a nested loop.

Can we compare the two parametrizations? In the first case, the average rank of an interacting neighbour is $(n_c^{\rm step}+1)/2 \approx n_c^{\rm step}/2$. In the second case, this average rank is $\approx n_c^{\rm exp}$. It makes sense to hypothesize this average rank should be invariant, regardless of the choice of parametrization. Then, if we infer models with data using the two parametrizations, we expect:

$$n_c^{\rm exp} \approx \frac{n_c^{\rm step}}{2}. \tag{B27}$$

The second important effective parameter is the total interaction strength $J\sum_j n_{ij}$, equal to $J_{\rm step}n_c^{\rm step}$ is the

first case, and to $\approx J_{\rm exp} n_c^{\rm exp}$ in the second one. Requiring that these quantities are equal in the two parametrizations yields:

$$J_{\rm exp} \approx 2 J_{\rm step}. \tag{B28}$$

Figure S4 shows that the effective $n_c^{\rm step}$ and $n_c^{\rm exp}$ learned from data follow these relations accurately.

Appendix C: Orientation relaxation time

In our work we compare the relaxation time of the orientational degrees of freedom, $\tau_{\rm relax}$, to the reshuffling time of the network, $\tau_{\rm network}$, finding the first one to be much smaller than the second one. This may seem an odd result, as in a fixed-lattice theory with spontaneously broken continuous symmetry both the correlation length and the relaxation time *diverge* with the system size L. Hence, in what sense can $\tau_{\rm relax}$ be small?

In the following we consider a fixed lattice for the following reason: we need to compare the relaxation time to the network reshuffling time; to do this consistently, we need to work out the relaxation time of the order parameter it in absence of the effect of network reshuffling. To fix ideas we also work on a regular lattice in the continuum limit; the following arguments, though, are valid in general. In this limit Eq. A13 now reads:

$$\frac{d\vec{\pi}}{dt} = J n_c a^2 \Delta \vec{\pi} + \vec{\xi}_\perp. \tag{C1}$$

where Δ is the Laplacian operator and a the lattice spacing. In Fourier space, this equation becomes:

$$i\omega \vec{\pi}(k,\omega) = -J n_c (ka)^2 \vec{\pi}(k,\omega) + \vec{\xi}_\perp(k,\omega) \tag{C2}$$

and its solution is:

$$\vec{\pi}(k,\omega) = G(k,\omega)\vec{\xi}_\perp(k,\omega), \tag{C3}$$

were the dynamical propagator (or dynamic response) of the Gaussian spin-wave theory in Fourier space is:

$$G(k,\omega) = \frac{1}{i\omega + J a^2 n_c k^2} \ , \tag{C4}$$

We need now to compute the dynamical self-correlation function, that is the correlation of the fluctuations at the *same* position x (or site i), namely,

$$C_{\rm relax}(t) = \langle \vec{\pi}(x,t_0) \cdot \vec{\pi}(x,t_0+t) \rangle \ . \tag{C5}$$

From (C3) and (C4) we have,

$$C_{\rm relax}(t) = 2(d-1)T \int_{1/L}^{1/a} d^d k \ \int d\omega \ \frac{e^{-i\omega t}}{(i\omega + J a^2 n_c k^2)(i\omega - J a^2 n_c k^2)} = 2(d-1)T \int_{1/L}^{1/a} d^d k \ \frac{e^{-J a^2 n_c k^2 t}}{J a^2 n_c k^2} \ , \tag{C6}$$

which (up to constant prefactors) is the self-correlation function reported in the main text. The absence of a mass term (zero mode) implies that in $d = 3$ the function $C_{\rm relax}(t)$ is a power law, so that the self-relaxation time diverges with L. However, as we explain the main, the modes that contribute to the rearrangement of the network are only those with short wavelength, comparable with the interaction range r_c; hence, only k larger to $1/r_c$ contributes to the network reshuffling in the integral above, and we therefore define the effective correlation function,

$$C^*_{\rm relax}(t) \equiv 2(d-1)T \int_{1/r_c}^{1/a} d^d k \ \frac{e^{-J a^2 n_c k^2 t}}{J a^2 n_c k^2} \ . \tag{C7}$$

This correlation function has now an exponential behavior for large t, with finite relaxation time equal to $(1/J n_c) \cdot (r_c/a)^2$. The ratio between interaction range and lattice spacing, (r_c/a), is in general of order 1 for short range interaction (as it is the case in flocks) and therefore the time scale of relaxation of the orientational degrees of freedom is $\tau_{\rm relax} = (J n_c)^{-1}$, which is what we study in the main text.

If we do not assume a regular lattice, instead of a differential Laplacian operator, we have to deal with the generic Laplacian matrix $\mathbf{\Lambda}$ in equation (4) in the main text, and with its eigenvalues, let us call them Λ. In this case the self-correlation function is given by,

$$C_{\rm relax}(t) \equiv 2(d-1)T \int_{\Lambda_{min}}^{\Lambda_{max}} d\Lambda \ \rho(\Lambda) \ \frac{e^{-J n_c \Lambda t}}{J n_c \Lambda} \ , \tag{C8}$$

where $\rho(\Lambda)$ is the eigenvalue spectrum of $\mathbf{\Lambda}$. In a spatially homogeneous network Λ scales as an inverse length squared, playing the same role as k^2 in a regular lattice. Thus, $\Lambda_{min} \sim 1/L^2$ and $\Lambda_{max} \sim 1/a^2$, a being the average nearest neighbors distance. The absence of a Λ-independent term at the denominator is equivalent to the absence of a k-independent term the case of a regular lattice (zero mode). Similarly, the largest contribution to the integral comes from the modes near the lower extreme of integration, $\Lambda_{min} \sim 1/L^2$. The previous argument then requires to restrict the integral for $\Lambda > 1/r_c^2$, hence giving

$$C^*_{\rm relax}(t) \equiv 2(d-1)T \int_{1/r_c^2}^{1/a^2} d\Lambda \ \rho(\Lambda) \ \frac{e^{-J n_c \Lambda t}}{J n_c \Lambda} \ . \tag{C9}$$

which (as in the regular lattice case) gives exponential relaxation with $\tau_{\rm relax} = (Jn_c)^{-1}$.

Our argument to restrict the k integral in the self-correlation function to short wavelength modes, $k > 1/r_c$, finds a strong consistency check in the following fact: even the network correlation function, $C_{\rm network}(t)$, does depend on a local scale, exactly as $C^*_{\rm relax}$ depends on r_c. When we ask what is the degree of reshuffling of the interaction network within a time t, we are effectively asking how much the network changes over a spatial scale n_c. We could, for example, ask what is the time needed to disrupt the entire network, i.e. the reshuffling over a scale N, and this would give a much larger time, scaling with N (for a computation of this time and its connetcion to mutual diffusion in space see [5]). In a similar way, when we integrate in (C6) down to $1/L$ we get a time scale which scales with L. Hence, when comparing orientation relaxation and network reshuffling we need to fix a scale for both phenomena. Since we are interested here in inferring the interaction rules, the right scale is the scale of interaction, namely r_c or n_c. On the other hand, as we discuss in the conclusions of the main text, were we interested in studying (or predicting) the large size behaviour in the long time limit, we should assess the divergence of both time scales with the size, which is the realm of the hydrodynamic theory. In general, both timescales τ_{relax} and $\tau_{\rm network}$ can be defined on a given spatial scale r (or the equivalent topological scale n). What we expect is that, as this scale r increases, τ_{relax} and $\tau_{\rm network}$ become closer and, at given $r^\star$, one has $\tau_{relax}(r^\star) \sim \tau_{\rm network}(r^\star)$. This lengthscale $r^\star$ represents a crossover scale above which the motility of individuals becomes relevant and the system behaves in a non-equilibrium way. When $r^\star \gg r_c$ we are in the condition of local equilibrium that we discussed in this paper. Note that an estimate of the crossover length can also be computed using scaling arguments within the hydrodynamic approach, see e.g. [6]

[1] Jaynes ET (1957) Information theory and statistical mechanics. *Physical Review* 106:620.

[2] Jaynes ET (1957) Information theory and statistical mechanics. ii. *Physical Review* 108:171.

[3] Cavagna A, et al. (2014) Dynamical maximum entropy approach to flocking. *Phys. Rev. E* 89:1–10.

[4] Bialek W, et al. (2012) Statistical mechanics for natural flocks of birds. *Proc. Natl. Acad. Sci. U. S. A.* 109:4786–91.

[5] Cavagna a, Duarte Queirós SM, Giardina I, Stefanini F, Viale M (2013) Diffusion of individual birds in starling flocks. *Proc. Biol. Sci.* 280:20122484.

[6] Toner J, Tu Y, Ramaswamy S (2005) Hydrodynamics and phases of flocks. *Annals of Physics* 318:170–244.