Dynamical Maximum Entropy Approach to Flocking

Andrea Cavagna[1,2,3], Irene Giardina[1,2,3], Francesco Ginelli[4], Thierry Mora[5], Duccio Piovani[2], Raffaele Tavarone[2] and Aleksandra M. Walczak[6]

[1] *Istituto Sistemi Complessi, Consiglio Nazionale delle Ricerche, UOS Sapienza, Rome, Italy*
[2] *Dipartimento di Fisica, Università Sapienza, Rome, Italy*
[3] *Initiative for the Theoretical Sciences, The Graduate Center, The City University of New York, New York*
[4] *SUPA, Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen, UK*
[5] *Laboratoire de physique statistique, CNRS, UPMC and École normale supérieure, Paris, France and*
[6] *Laboratoire de physique théorique, CNRS, UPMC and École normale supérieure, Paris, France*

(Dated: October 15, 2013)

We derive a new method to infer from data the out-of-equilibrium alignment dynamics of collectively moving animal groups, by considering the maximum entropy distribution consistent with temporal and spatial correlations of flight direction. When bird neighborhoods evolve rapidly, this dynamical inference correctly learns the parameters of the model, while a static one relying only on the spatial correlations fails. When neighbors change slowly and detailed balance is satisfied, we recover the static procedure. We demonstrate the validity of the method on simulated data. The approach is applicable to other systems of active matter.

Flocking, the highly coordinated motion displayed by large groups of birds, has attracted much attention over the last twenty years as a prototypical example of out-of-equilibrium collective behavior. It has been suggested that flocking is an emergent phenomenon resulting from mutual alignment of velocities between neighboring birds, much like the spontaneous symmetry breaking towards a magnetized state exhibited by ferromagnetic spins at low temperatures. Although this idea has been extensively studied from a theoretical view point [2–4], only recently have advances in the 3D imaging of large flocks of starlings [5] given empirical grounds supporting this picture. Interactions between individuals in the flock were shown to be topological and local [6], leading to the global ordering of flight orientations and scale-free correlation functions [7]. The analogy with ferromagnetic systems was made explicit by the quantitative inference of spin models from empirical data using the principle of maximum entropy [8, 9]. These analyses have focused on the steady state behaviour of flocks, by examining the flock configurations as drawn from a given statistical ensemble. This approach allows for an effective equilibrium-like description, without having to make detailed assumptions about the microscopic rules governing flock behaviour. Yet it is an incomplete picture as it does not take into account the dynamical, out-of-equilibrum nature of the process.

The major difference between flocks and equilibrium spin systems is that birds are like active particles, constantly moving within the flock along the direction given by their "spin", exchanging local interaction partners, thus extending their effective interaction range, and also breaking detailed balance. This qualitative difference between equilibrium spins and out-of-equilibrium active particles can dramatically affect the thermodynamic properties of the system, including the existence of an ordered phase in two dimensions, and the value of the critical exponents [3]. One can thus naively interpret the parameters of static descriptions of flocks as a renormalized version of some underlying and unknown out-of-equilibrium dynamical model.

In this paper we propose a general framework for learning the features of the out-of-equilibrium dynamics directly from data, while making minimal assumptions about the specific microscopic interaction rules. We generalize the principle of maximum entropy to account for multi-time correlations between birds, and show that maximizing the entropy under this constraint is equivalent to inferring a dynamical model of social forces. We test our dynamical inference method on synthetic data generated by a topological Vicsek model (VM), showing that its inferred interaction parameters are consistently better than the ones obtained in an equilibrium framework, especially when the relative mobility between individuals is high. When the interaction network is static, and the dynamics satisfies detailed balance, our method recovers the results of the static approach [8], additionally allowing us to separate the contributions of interaction strength and noise to the alignment dynamics.

Maximum entropy distributions are the least constrained distributions that are consistent with certain selected key observables of the data. They usually map onto equilibrium statistical mechanics problems and do not involve any assumptions about the system under study, besides the choice of the relevant observables, which should be selected accordingly to the fundamental symmetries of the underlying system. They have been particularly successful in describing collective and emergent phenomena in biological systems comprising many correlated degrees of freedom [10]. When considering flocks, where polar order is present, a natural choice of observables to be constrained by the data are

the equal time pairwise correlation functions between birds orientations: $\langle s_i s_j \rangle$, where s_i is a d-dimensional unit vector denoting the flight direction of bird i, with $i = 1, \ldots, N$. (Throughout the paper inner products over the physical space are implicit.) These correlations were found to exhibit scale-free behavior in natural flocks [7], and characterize the collective nature of flocking. The maximum entropy distribution $P(\mathbf{s})$ for the orientations can then be computed by maximizing the entropy $S[P] = -\sum_{\mathbf{s}} P(\mathbf{s}) \ln P(\mathbf{s})$, while constraining the equal-time correlations to their experimental values. The result is the stationary probability distribution for the equilibrium heterogeneous Heisenberg model [8]:

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(\frac{1}{2} \sum_{i \neq j} J_{ij}^{\text{stat}} s_i s_j \right), \tag{1}$$

where $\mathbf{s}$ is a shorthand for $(s_1, s_2, \ldots, s_N)$ and Z a normalization constant. The interaction parameters J_{ij}^{stat} are Lagrange multipliers that need to be tuned so that the probability distribution (1) matches the empirical correlation functions $\langle s_i s_j \rangle$. Using 3D, single individual resolution data of large bird flocks, this class of models was shown to recapitulate quantitatively the ordering properties of real flocks [8].

But infinitely many dynamical models may give rise to this steady-state distribution, most of which break detailed balance. In fact, the change of neighborhoods causes the interaction network to vary in time, keeping the system constantly out of equilibrium. Here we extend the maximum entropy framework to account for the non-equilibrium nature of flocking. We consider the set of entire trajectories $(\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^T)$, where the superscript index denotes time points separated by δt. We then look for the distribution $P(\mathbf{s}^1, \ldots, \mathbf{s}^T)$ that maximizes the entropy while reproducing some given experimental observables. Since we want to capture the dynamics, in addition to equal-time correlation functions, we also constrain the correlation functions between two consecutive time points $\langle s_i^{t+1} s_j^t \rangle$. Doing so yields the following form of the probability distribution over *trajectories* (see Appendix for details):

$$P(\mathbf{s}^1, \ldots, \mathbf{s}^T) = \frac{1}{\hat{Z}} \exp\left(-\mathcal{A}\right), \tag{2}$$

where $\hat{Z}$ is a normalization factor, and the "effective action" (or minus log-likelihood) reads:

$$\mathcal{A} = -\frac{1}{2} \sum_t \sum_{i \neq j} \left(J_{ij;t}^{(1)} s_i^t s_j^t + J_{ij;t}^{(2)} s_i^{t+1} s_j^t \right). \tag{3}$$

There now are two sets of time-dependent coupling parameters, for synchronous and consecutive times. We note that the probability (Eq. 2) corresponds to Markovian dynamics; non-Markov forms are possible if constraining more complex multi-time observables.

When flight orientations are highly polarized (as in the case of starling flocks [7]), one can use the spin-wave (SW) approximation [11] to explicitly rewrite the action as a sum of Markov terms which are quadratic in the spin-wave variables. Specifically, we denote $s_i = \pi_i + n\sqrt{1-(\pi_i)^2}$, where n is an abitrary unit vector close to the average flight direction of the flock, and π_i is the perpendicular component of the orientation, $\pi_i n = 0$. (When there is no ambiguity we drop the time superscript.) When the flock is highly polarized, we have $\pi_i^2 \ll 1$, and we may expand at small π_i. The action may then be written as a sum of terms corresponding to the transition probabilities $P(\boldsymbol{\pi}'|\boldsymbol{\pi})$ between successive time points (see Appendix for technical details): $\mathcal{A} + \ln \hat{Z} = -\ln P(\mathbf{s}^1) + \sum_t \mathcal{L}_t$, with:

$$\mathcal{L}_t(\boldsymbol{\pi}^{t+1}, \boldsymbol{\pi}^t) \equiv -\log P(\boldsymbol{\pi}^{t+1}|\boldsymbol{\pi}^t) = -\frac{d-1}{2} \ln\left(\frac{\det \mathbf{A}_t}{(2\pi)^N} \right) + \frac{1}{2} \left(\boldsymbol{\pi}^{t+1} - \mathbf{M}_t \boldsymbol{\pi}^t \right)^\dagger \mathbf{A}_t \left(\boldsymbol{\pi}^{t+1} - \mathbf{M}_t \boldsymbol{\pi}^t \right), \tag{4}$$

where $\mathcal{L}_t$ is formally equivalent to a Lagrangian density. In (4) we have defined: $\mathbf{M}_t = \mathbf{A}_t^{-1} \mathbf{J}_t^{(2)}/2$ with $A_{ij;t} = -K_{ij;t} + \delta_{ij} \sum_k K_{ik;t} + \delta_{ij} \sum_k J_{ik;t}^{(2)}/2$, where $\mathbf{K}_t$ is a calculation intermediate obtained by a descending recursion enforcing normalization at each time step: $\mathbf{K}_{t-1} = \mathbf{J}_t^{(1)} + \mathbf{J}_t^{(2)\dagger} \mathbf{A}_t^{-1} \mathbf{J}_t^{(2)}/4$.

The Gaussian form of the transition probabilities Eq. (4), corresponds to a spin-wave dynamics described by the following stochastic equation:

$$\pi_i^{t+1} = \sum_j M_{ij;t} \pi_j^t + \epsilon_i^t, \tag{5}$$

with ϵ^t being a random, isotropic Gaussian noise perpendicular to n, of zero mean and covariance: $\langle \epsilon^t (\epsilon^{t'})^\dagger \rangle = 2(d-1) \mathbf{A}_t^{-1} \delta_{t,t'}$, where $\delta_{t,t'}$ is the Kronecker delta.

Eq. (5) can be interpreted as follows. At each time, individual i computes its new orientation from a weighted average over the orientation of other individuals, including itself, at the previous time point with weights encoded in the matrix $\mathbf{M}_t$ (one can check that, by construction, $\sum_j M_{ij} = 1$). Noise $\boldsymbol{\epsilon}^t$ added to this average determines the level of error in the alignment. Without it, all individuals would be perfectly aligned. This model may be viewed as the spin-wave expansion of a generalized Vicsek model [1] with arbitrary weights and noise.

Tuning the parameters to match the correlation functions is equivalent to maximizing the likelihood, Eq. (2) (see Appendix), or equivalently maximizing the log-likelihood $-\sum_t \mathcal{L}_t$, with which we will work from now on. To maximize the likelihood with respect to the two equivalent sets of parameters $\{\mathbf{J}_t^{(1)}, \mathbf{J}_t^{(2)}\}$ or $\{\mathbf{M}_t, \mathbf{A}_t\}$, we would need to observe a large number of random realizations of the same flock dynamics. This is impossible

in practice due to limited data compared to prohibitively large number of potential configurations of the bird positions that one would need to sample.

To overcome this problem, we need to introduce some additional assumptions about the interaction network and the form of the noise in order to simplify the parameter space and the number of observables. These simpifications come naturally in the Markovian description parametrized by $\mathbf{M}_t$ and $\mathbf{A}_t$. From a biological standpoint, it is reasonable to assume that birds treat information from each interacting neighbor (the precise definition of "neighborhood" being left unspecified for the moment) equally, while keeping memory of their own direction. Mathematically this translates into:

$$M_{ij} = (1 - J\delta t n_i)\delta_{ij} + J\delta t n_{ij}\,, \tag{6}$$

where $n_{ij}=1$ if j is one of i's neighbours, and 0 otherwise, and $n_i=\sum_j n_{ij}$ is the global number of neighbors interacting with bird i. (For ease of notation we omit the t index, even though n_{ij} depends on t.) The scalar parameter J now measures the alignemnt interaction strength. Errors made by different birds when trying to align with their neighbours can be assumed to be of the same amplitude and independent of each other, so that noise is uncorrelated and $\mathbf{A}$ is proportional to the identity, $A_{ij}=[1/(2\delta t T)]\delta_{ij}$. Here T is a squared noise amplitude (the out-of-equilibrium equivalent of a temperature) that sets the level of disorder in the system. The scaling in δt ensures a well-defined continuous limit when $\delta t\to 0$, described by a Langevin equation.

We can reconcile this dynamical description with the static inference [8] in the special case of equilibrium dynamics, which is realized when n_{ij} is symmetric and constant in time. In this case, the spins can be described for $\delta t \to 0$ by a stationary distribution with the same form as in Eq. (1) and the steady-state couplings take the simple equilibrium value [8], $J_{ij}^{\rm stat}=(J/T)n_{ij}$ (see Appendix).

Taking the specific form of $\mathbf{M}_t$ and $\mathbf{A}_t$ above for a given network of neighbours, we obtain a formula for $\mathcal{L}_t$ that only depends on two parameters, the interaction strength J and the "effective temperature" T

$$\mathcal{L}_t = \frac{d-1}{2}\ln(2T\delta t) - 2Jn_c\delta t\left[\tilde{C}_s - C_{\rm int} - \tilde{G}_s + G_{\rm int}\right] + (Jn_c\delta t)^2\left[\hat{C}_s - 2\tilde{C}_{\rm int} + C'_{\rm int}\right] + C_s^1 + C_s - 2G_s\,, \tag{7}$$

with $n_c = (1/N)\sum_i n_i$. Also, the number of independent observables appearing in $\mathcal{L}_t$ is drastically reduced, to a handful of empirical integrated pair correlation functions defined in Table I. These correlations can be evaluated over pairs of consecutive configurations, or averaged over the entire sequence if we work with time-independent parameters and steady state dynamics.

Maximizing the log-likelihood with respect to J and T, $\partial\mathcal{L}_t/\partial T=0$ and $\partial\mathcal{L}_t/\partial J=0$, yields simple analytical expressions for the parameters as a function of the empirical correlation functions:

C_s^1	$(1/N)\sum_i(\pi_i^{t+1})^2$	$C_{\rm int}$	$(1/Nn_c)\sum_{ij} n_{ij}\pi_i^t\pi_j^t$
C_s	$(1/N)\sum_i(\pi_i^t)^2$	$C'_{\rm int}$	$(1/Nn_c^2)\sum_{ijk} n_{ij}n_{ik}\pi_j^t\pi_k^t$
G_s	$(1/N)\sum_i \pi_i^{t+1}\pi_i^t$	$G_{\rm int}$	$(1/Nn_c)\sum_{ij} n_{ij}\pi_i^{t+1}\pi_j^t$
$\tilde{C}_s$	$(1/Nn_c)\sum_i n_i(\pi_i^t)^2$	$\hat{C}_s$	$(1/Nn_c^2)\sum_{ij}(n_i\pi_i^t)^2$
$\tilde{G}_s$	$(1/Nn_c)\sum_i n_i\pi_i^{t+1}\pi_i^t$	$\tilde{C}_{\rm int}$	$(1/Nn_c^2)\sum_{ij} n_i n_{ij}\pi_i^t\pi_j^t$

TABLE I. Empirical correlation functions used in the text.

$$J = \frac{1}{n_c}\frac{\Omega + (d-1)T_0}{C'_{\rm int} + \hat{C}_s - 2\tilde{C}_{\rm int}}, \tag{8}$$

$$T = T_0 + \frac{C_s^1 - C_s}{2(d-1)\delta t} - \frac{J\,n_c\delta t}{2(d-1)}\left(\frac{\tilde{C}_s - \tilde{G}_s}{\delta t} + \Omega\right), \tag{9}$$

where

$$T_0 = \frac{C_s - G_s}{\delta t(d-1)}, \quad \Omega = \frac{G_{\rm int} - C_{\rm int}}{\delta t}. \tag{10}$$

The leading-order temperature T_0 is the derivative of a self-correlation function, and obeys the standard fluctuation-dissipation relationship found in equilibrium dynamics. The term Ω is related to the dynamics of the network. In particular, at steady state $dC_{\rm int}/dt = 0$ implies $\Omega \propto \sum_{ij}\pi_i\pi_j dn_{ij}/dt$.

In order to apply Eqs. (8)-(9) to data, one still needs to specify the neighboring matrix n_{ij}. In absence of prior information, the simplest possibility is to assume that each bird interacts with the first n_c neighbors [8]. An alternative choice would be to define neighbors according to a metric rule, each bird interacting with neighbors within a given distance r_c. In both cases an extra parameter is introduced, either the 'topological' interaction range n_c or the metric range r_c, that can also be inferred by likelihood maximization. Another scheme is to define neighbors through a Voronoi tassellation [12], as in the Topological VM [4]. Likelihoods between different neighborhood definitions may also be compared to find the one closest to optimality.

We tested our dynamical inference method on synthetic data generated from a slight generalization of the Topological VM on a two dimensional torus of linear size $L = 32$ with $N = 1024$ particles:

$$\theta_i^{t+\delta t} = {\rm Arg}[s_i^t + J_V\delta t\sum_j n_{ij}s_j^t] + \sqrt{\delta t}\,\xi_i^t \tag{11}$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\,s_i^{t+\delta t} \tag{12}$$

where $s_i = (\cos\theta_i, \sin\theta_i)$, and ${\rm Arg}(s)$ is the angle of vector s. The delta-correlated angular noise ξ_i^t is uniformly distributed in $[-\eta\pi, +\eta\pi]$, corresponding to an effective temperature $T_V = (\eta\pi^2)/6$ for $\delta t \to 0$. The Voronoi adjacency matrix n_{ij} has a non-uniform degree

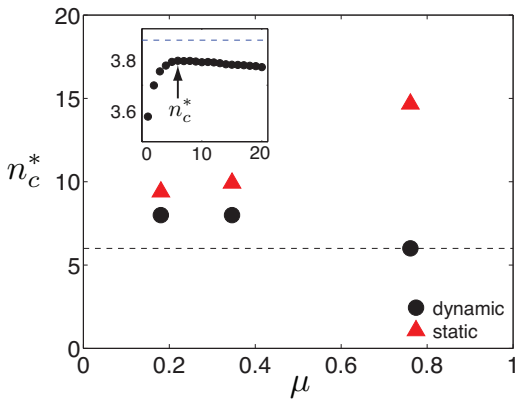

FIG. 1. (Color online.) Comparison between dynamical and static inference. Data was generated using Voronoi neighborhood. The inference was performed by using either a Voronoi rule or a nearest-neighbor (NN) topological rule, parametrized by the number n_c of interacting neighbors. Main panel: The inferred number of interacting neighbors n_c^* is shown as a function of the mixing rate μ. Circles: dynamical inference; triangles: static inference. The dashed lines marks the real average value, $n_V = 6$. Static inference badly overestimates the number of interacting neighbors at large mixing, while dynamical inference does a much better job. Inset: Dynamical normalized log-likelihood $-\mathcal{L}_t/N$ as a function of n_c for the NN topological rule (circles). The maximum of this function gives the NN value of n_c^* reported in the main panel. The Voronoi likelihood (dashed line) is larger than the NN one, revealing that Voronoi was the actual generating rule. Data are for high mixing.

n_i, of mean n_V=6. A spin-wave expansion of Eq. (11) leads to an expression of the form of (5)-(6), with $J \approx J_V/(1 + J_V n_V \delta t)$ (see Appendix). The degree of neighbor mixing is characterized by a single mixing parameter $\mu = \langle 1/(N n_c) \sum_{ij} |dn_{ij}/dt| \rangle$, which quantifies how fast birds exchange neighbors. We performed simulations with time step $\delta t = 0.01$ in three regimes with slow, medium and fast neighbor mixing ($\mu = 0.18, 0.35, 0.76$, $v_0 = 0.5, 1.0, 2.0$, $J_V = 1.0, 1.0, 0.1$ and $\eta = 0.3, 0.2, 0.12$ respectively), all of which display the same level of polarization, $N^{-1} \| \sum_i s_i \| \approx 0.97$.

We then applied the inference procedure described in Eqs. (8)-(9) to the synthetic dataset generated by the simulations. In the inference we tried the choices for n_{ij} discussed above: the n_c nearest-neighbor (NN) topological rule, the metric rule where $n_{ij} = 1$ within a metric range r_c (and 0 outside), and the Voronoi rule (actually used to generate the data). Correlation functions were averaged over 10^3 different configurations in the stationary state, sampled from a single run at 100 time unit intervals, ensuring independent sampling.

The likelihood as a function of n_c can be computed with the NN rule using Eqs. (8)(9) and (7). The result is shown in the inset of Fig. 1 for the high mixing regime. Its maximum n_c^* corresponds to the most likely interaction range, from which the optimal J^* and T^* are computed via Eqs. (8)-(9). Fig. 1 shows that the new dynamical procedure systematically outperforms the static approach described in [8] in predicting the mean interaction range n_c. The error made by the static inference is larger when neighbor mixing is higher and the dynamics is strongly out-of-equilibrium. That is because in the high-mixing case, the *effective* number of interacting neighbors, as inferred by the static approach, includes neighbors visited in the recent past in addition to the current ones, and thus is larger than the true n_c. Overall, the dynamical inference based on NN interactions performs reasonably well, considering that the model used for the inference incorrectly assumes a constant n_c. Not surprisingly, the log-likelihood computed with the (correct) Voronoi topology is larger than with the (incorrect) NN one. The temperature T is well inferred in both cases (8% error), while the alignment strength J is well recovered when assuming Voronoi neighbors (3% error), and approximately with a NN topology (20% error).

Performing the dynamical inference using a metric rule gives significantly worse results, giving $n_c \sim 3$ (see figure 2), a factor 2 smaller than the correct value. Hence the dynamical method not only gives us the correct interaction parameters, but also distinguishes the rule used to build the interaction network. The method exploits the different ways in which spatial density fluctuations translate into fluctuations in the number of neighbors. In the Voronoi network (the generating one), the number of neighbors n_i of each point fluctuates weakly around its mean value of 6. In the NN case, n_i does not fluctuate at all, whereas with the metric rule n_i exhibits *very* large fluctuations, directly linked to the VM giant density fluctuations [4]. The large fluctuations of n_i make the correlation functions of Table I very different from their correct (Voronoi) values.

In summary, we have derived a dynamical maximum entropy method to infer the alignment dynamics of highly-ordered animal groups from just two consecutive snapshots. Tests on synthetic data confirm the validity of our method. Our approach is very general and makes minimal, symmetry-based assumptions on the structure of the dynamics under investigation, alternative to other inference methods [13]. Related approaches have been proposed in the context of Ising spins or spiking neurons [14]; however, in that case it is hard to relate a simple interaction form of the Markovian transition probabilities to a principle of maximum entropy. Our work emphasizes the need for a dynamical inference approach to out-of-equilibrium active matter systems, especially when there is no *a priori* knowledge of the timescales in the system, which is usually the case when dealing with experimental data.

Our approach is applicable to many systems where collective motion is observed, including moving animal groups [15], bacterial colonies [16], motility assays [17],

collective motion of epithelial cells [18], or vibrated polar disks [19]. Throughout this work we have assumed that δt is equal to (or smaller than) the real update time lag, namely the biological timescale. This may not be true for some datasets, as the sampling time of the experimental equipment is likely to be larger than the neural update time actually used by animals. This is certainly the case for the starling data of [8]. When this happens, the experimental time series is a coarse-grained version of the real dynamics, so that the present method would probably provide a time-renormalized value of the interaction parameters. It would therefore be important to generalize our equations to deal with this issue. Other generalizations include the analysis of other symmetries than the polar one (as in systems with nematic order [20]), or the extension to second-order dynamics describing systems characterized by linear, not diffusive, dispersion relations [21].

Acknowledgements. We thank Martin Weigt for helpful discussions. I.G. was supported by grants IIT–Seed Artswarm, ERC–StG n.257126. A.C was supported by grant US-AFOSR FA95501010250 (through the University of Maryland). FG acknowledges support from grants EPSRC First Grant EP/K018450/1 and MC Career Integration Grant PCIG13-GA-2013-618399. Work in Paris was supported by grant ERCStG n. 306312.

[1] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, PRL **75**, 1226 (Aug 1995)
[2] Y. Tu, J. Toner, and M. Ulm, PRL **80**, 4819 (1998); G. Grégoire and H. Chaté, PRL **92**, 025702 (2004); E. Bertin, M. Droz, and G. Grégoire, PRE **74**, 022101 (2006); J. Phys. A **42**, 445001 (2009); T. Ihle, PRE **83**, 030901 (2011); H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, PRE **77**, 046113 (2008); P. Szabó, M. Nagy, and T. Vicsek, PRE **79**, 021908 (2009); A. Peshkov, S. Ngo, E. Bertin, H. Chaté, and F. Ginelli, PRL **109**, 098101 (2012); J. Toner, PRE **86**, 031918 (2012); S. Ramaswamy, Annu. Rev. Condens. Matter Phys. **1**, 323 (2010)
[3] J. Toner and Y. Tu, PRL **75**, 4326 (1995); PRE **58**, 4828 (1998)
[4] F. Ginelli and H. Chaté, PRL **105**, 168103 (2010)
[5] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Anim Behav **76**, 217 (2008); A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, and A. Procaccini, Anim Behav **76**, 237 (2008); M. Ballerini et al., Anim Behav **76**, 201 (2008)
[6] M. Ballerini et al., PNAS **105**, 1232 (2008)
[7] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, PNAS **107**, 11865 (2010)
[8] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, PNAS **109**, 4786 (2012)
[9] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. Walczak, arXiv:1307.5563v1 (2013)
[10] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006); J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, J Neurosci **26**, 8254 (2006); M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, PNAS **106**, 67 (2009); T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, **107**, 5405 (2010); M. Santolini, T. Mora, and V. Hakim, arXiv:1302.4424v1 (2013); T. Mora and W. Bialek, J Stat Phys **144**, 268 (2011)
[11] F. Dyson, Physical review **102**, 1217 (1956)
[12] G. Voronoi, J Reine Angew Math **133**, 97 (1907)
[13] J. E. Herbert-Read, A. Perna, R. P. Mann, T. M. Schaerf, D. J. T. Sumpter, and A. J. W. Ward, PNAS **108**, 18726 (2011); Y. Katz, K. Tunstrøm, C. C. Ioannou, C. Huepe, and I. D. Couzin, **108**, 18720 (2011); J. Gautrais et al. *et al.*, PLoS Comp. Bio. **8**, e1002678 (2012)
[14] O. Marre, S. E. Boustani, Y. Frégnac, and A. Destexhe, PRL **102**, 138101 (2009); Y. Roudi and J. Hertz, PRL **106**, 048702 (2011); J. C. Vasquez, O. Marre, A. G. Palacios, M. J. Berry II, and B. Cessac, J. Physiol. Paris **106**, 120 (2012)
[15] J. K. Parrish and W. M. Hamner, *Animal Groups in Three Dimensions* (Cambridge University Press, Cambridge, 1997)
[16] H. P. Zhang, A. Be'er, E.-L. Florin, and H. L. Swinney, PNAS **107**, 13626 (2010); X. Chen, X. Dong, A. Be'er, H. L. Swinney, and H. P. Zhang, PRL **108**, 148101 (2012)
[17] Y. Sumino. *et al.*, Nature (London) **483**, 448 (2012)
[18] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim, PLoS Comput Biol **9**, e1002944 (2013)
[19] J. Deseigne, O. Dauchot, and H. Chaté, PRL **105**, 098001 (2010); C. Weber, T. Hanke, J. Deseigne, S. Léonard, O. Dauchot, E. Frey, and H. Chaté, PRL **110**, 208001 (2013)
[20] F. Ginelli, F. Peruani, M. Bär, and H. Chaté, PRL **104**, 18452 (2010)
[21] A. Attanasi et al., arXiv:1303.7097v1 (2013)

APPENDIX

Maximum entropy approach

In the maximum entropy approach, one looks for the maximally disordered probability distribution consistent with carefully chosen observables of the data. In practice, given a stochastic variable $\mathbf{s}$, and a set of observables $\{\mathcal{O}_\mu(\mathbf{s})\}$, with $\mu = 1, \dots, K$, one looks for the model distribution P of maximum entropy

$$S[P] = -\sum_{\mathbf{s}} P(\mathbf{s}) \ln P(\mathbf{s}), \tag{13}$$

that coincides with the data for the average values of each of the observables:

$$\langle \mathcal{O}_\mu \rangle_{\rm data} = \langle \mathcal{O}_\mu \rangle_P. \tag{14}$$

Using the technique of Lagrange multipliers, one shows that the distribution takes the exponential form:

$$P(\mathbf{s}) = \frac{1}{\mathcal{Z}(\{\lambda_\mu\})} \exp\left(-\sum_{\mu=1}^{K} \lambda_\mu \mathcal{O}_\mu(\mathbf{s})\right), \tag{15}$$

where $\{\lambda_\mu\}$ are Lagrange multipliers that need to be set to satisfy (14), and $\mathcal{Z}(\{\lambda_\mu\})$ is a normalization factor enforcing $\sum_{\mathbf{s}} P(\mathbf{s}) = 1$. By analogy with the Boltzman distribution from equilibrium statistical mechanics, the sum inside the exponential may be interpreted as an energy.

Conveniently, the Lagrange multipliers that match the mean value of the observables are also those that maximize the likelihood of the data given the exponential form (15). Given M data points $\mathbf{s}^1, \ldots, \mathbf{s}^M$, the log-likelihood of the data reads:

$$\begin{aligned} \ln \mathcal{P}(\{\lambda_\mu\}) &\equiv \ln \prod_{a=1}^{M} P(\mathbf{s}^a) \\ &= -\sum_{a=1}^{M}\sum_{\mu=1}^{K} \lambda_\mu \mathcal{O}_\mu(\mathbf{s}^a) - M \ln \mathcal{Z}(\{\lambda_\mu\}). \end{aligned} \tag{16}$$

Maximizing the log-likelihood with respect to the parameters $\{\lambda_\mu\}$ implies:

$$\begin{aligned} \frac{\partial \ln \mathcal{P}(\{\lambda_\mu\})}{\partial \lambda_\mu} = M\left[-\frac{\partial \ln \mathcal{Z}}{\partial \lambda_\mu} - \langle \mathcal{O}(\mathbf{s})\rangle_{\text{data}}\right] = 0 \\ M\left[\langle \mathcal{O}(\mathbf{s})\rangle_P - \langle \mathcal{O}(\mathbf{s})\rangle_{\text{data}}\right] = 0. \end{aligned} \tag{17}$$

By virtue of this equivalence, we will maximize the expression of the log-likelihood with respect to the parameters to find the correct maximum entropy distribution.

Let us now consider the specific case of bird flocks. Denote $\mathbf{s} = (s_1, \ldots, s_N)$ the flight directions of birds in a flock of size N. The maximum entropy distribution consistent with the *synchronous* pairwise correlation functions $\langle s_i s_j \rangle$, for all (i, j), reads:

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(\frac{1}{2}\sum_{ij} J_{ij}^{\text{stat}} s_i s_j\right), \tag{18}$$

where $\{J_{ij}\}$ are (minus) the Lagrange multipliers associated to the constraints on the correlation functions.

Generalizing the set of constrained obsersables to both synchronous and consecutive-time correlation functions, $\{s_i^t s_j^t\}$ and $\{s_i^{t+1} s_j^t\}$, for all pair (i, j), and for all times t in the trajectory, yields a time-dependent maximum entropy distribution:

$$P(\mathbf{s}^1, \ldots, \mathbf{s}^T) = \frac{1}{\hat{Z}} \exp\left(-\mathcal{A}\right). \tag{19}$$

with $\hat{Z}$ again a normalization factor, and

$$\mathcal{A} = -\frac{1}{2}\sum_t \sum_{i\neq j} \left(J_{ij;t}^{(1)} s_i^t s_j^t + J_{ij;t}^{(2)} s_i^{t+1} s_j^t\right), \tag{20}$$

where $\{J_{ij;t}^{(1)}\}, \{J_{ij;t}^{(2)}\}$ are the Lagrange multipliers associated to the constraints on the synchronous and consecutive-time correlation functions. Here, $\mathcal{A}$ is more appropriately interpreted as an action, in a path-integral representation of the stochastic trajectories of the whole flock.

Markovian description

Because the action only involves cross-terms between consecutive times, it underlies a Markov process

$$P(\mathbf{s}^1, \ldots, \mathbf{s}^T) = P(\mathbf{s}^1) \prod_{t=1}^{T-1} P(\mathbf{s}^t | \mathbf{s}^{t-1}) \tag{21}$$

and can be rewritten as:

$$\mathcal{A} + \ln \hat{Z} = -\ln P(\mathbf{s}^1) + \sum_t \mathcal{L}_t(\mathbf{s}^{t+1}, \mathbf{s}^t), \tag{22}$$

where

$$\mathcal{L}_t(\mathbf{s}^{t+1}, \mathbf{s}^t) \equiv -\ln P(\mathbf{s}^{t+1} | \mathbf{s}^t) \tag{23}$$

may be interpreted as a Lagrangian density in the path integral formalism.

Let us check that this Markovian decomposition is possible. Identifying the two expressions of $\mathcal{A}$ (20) and (22), we may write $\mathcal{L}_t$ in the form:

$$\mathcal{L}_t(\mathbf{s}', \mathbf{s}) = -\frac{1}{2}\sum_{ij}\left(J_{ij;t}^{(2)} s_i' s_j + J_{ij;t}^{(1)} s_i s_j\right) - K_t(\mathbf{s}') + K_{t-1}(\mathbf{s}), \tag{24}$$

with the constraint that, for all $\mathbf{s}$, the transition probability be normalized,

$$1 = \sum_{\mathbf{s}'} \exp\left[-\mathcal{L}_t(\mathbf{s}', \mathbf{s})\right] \tag{25}$$

which entails:

$$K_{t-1}(\mathbf{s}) = \ln \sum_{\mathbf{s}'} \exp\left[\frac{1}{2}\sum_{ij}\left(J_{ij;t}^{(2)} s_i' s_j + J_{ij;t}^{(1)} s_i s_j\right) + K_t(\mathbf{s}')\right]. \tag{26}$$

Eq. (26) defines a descending recursion, by which K_t is calculated from the next time point. Thus the Markovian form of the action is fully specified using (24).

Equivalence with a generalized Vicsek model in the spin-wave approximation

In general, the integral in (26) cannot be calculated analytically, and K_t does not have a simple quadratic form as a function of $\mathbf{s}$. However things simplify in the spin-wave approximation, where the flock is very polarized, as

we will show now. Denote $s_i = \pi_i + n\sqrt{1-(\pi_i)^2}$, where n is an abitrary unit vector, and π_i is the perpendicular component of the orientation, $\pi_i n = 0$. n is chosen to be close the flock's main direction of flight, so that $\pi_i \ll 1$. Let us assume a quadratic form for K_t:

$$K_t(\mathbf{s}) = \frac{1}{2}\sum_{ij} K_{ij;t} s_i s_j + U_t. \tag{27}$$

The integral in (26) can be expanded at small π:

$$\begin{aligned} K_{t-1}(\boldsymbol{\pi}) &= \frac{1}{2}\sum_{ij} J^{(1)}_{ij;t}(1+\pi_i\pi_j - \pi_i^2) \\ &+ U_t + \frac{1}{2}\sum_{ij}\left(J^{(2)}_{ij;t} + K_{ij;t}\right) - \frac{1}{4}\sum_{ij} J^{(2)}_{ij;t}\pi_j^2, \\ &+ \ln\int d\boldsymbol{\pi}' \exp\left[-\frac{1}{2}\sum_{ij} A_{ij;t}\pi'_i\pi'_j + \frac{1}{2}\sum_{ij} J^{(2)}_{ij;t}\pi'_i\pi_j\right] \end{aligned} \tag{28}$$

with

$$A_{ij;t} = -K_{ij;t} + \delta_{ij}\sum_k K_{ik;t} + \frac{1}{2}\delta_{ij}\sum_k J^{(2)}_{ik;t}. \tag{29}$$

This Gaussian integral can be calculated exactly. Doing so, and expanding the left-hand side of (26) at small π, yields

$$\begin{aligned} K_{ij;t-1} - \delta_{ij}\sum_k K_{ik;t-1} &= J^{(1)}_{ij;t} - \delta_{ij}\sum_k J^{(1)}_{ik;t} \\ &+ \frac{1}{4}\left[\mathbf{J}^{(2)\dagger}_t \mathbf{A}_t^{-1}\mathbf{J}^{(2)}_t\right]_{ij} - \frac{1}{2}\delta_{ij}\sum_k J^{(2)}_{ik;t}, \end{aligned} \tag{30}$$

$$\begin{aligned} U_{t-1} + \frac{1}{2}\sum_{ij} K_{ij;t-1} &= \frac{1}{2}\sum_{ij} J^{(1)}_{ij;t} + U_t \\ &+ \frac{1}{2}\sum_{ij}\left(J^{(2)}_{ij;t} + H_{ij;t}\right) - \frac{d-1}{2}\ln\left(\frac{\det \mathbf{A}_t}{(2\pi)^N}\right). \end{aligned} \tag{31}$$

Focusing on the non-diagonal terms of the matrix $\mathbf{K}_t$, we obtain a simple expression for the recursion:

$$\mathbf{K}_{t-1} = \mathbf{J}^{(1)}_t + \frac{1}{4}\mathbf{J}^{(2)\dagger}_t \mathbf{A}_t^{-1}\mathbf{J}^{(2)}_t. \tag{32}$$

We can now replace the expression of K_t (24), and thus rewrite the transition probability in terms of π in a Gaussian form:

$$\begin{aligned} \mathcal{L}_t(\boldsymbol{\pi}', \boldsymbol{\pi}) = &-\frac{d-1}{2}\ln\left(\frac{\det \mathbf{A}_t}{(2\pi)^N}\right) \\ &+\frac{1}{2}\left(\boldsymbol{\pi}' - \mathbf{M}_t\boldsymbol{\pi}\right)^\dagger \mathbf{A}_t \left(\boldsymbol{\pi}' - \mathbf{M}_t\boldsymbol{\pi}\right), \end{aligned} \tag{33}$$

with

$$\mathbf{M}_t = \frac{1}{2}\mathbf{A}_t^{-1}\mathbf{J}^{(2)}_t. \tag{34}$$

This transition probability rule describes a random walk in the joint space of bird directions, described by:

$$\pi_i^{t+1} = \sum_j M_{ij;t}\pi_j^t + \epsilon_i^t, \tag{35}$$

with $\boldsymbol{\epsilon}^t$ a random, isotropic Gaussian noise perpendicular to n, of zero mean and covariance:

$$\langle \boldsymbol{\epsilon}^t (\boldsymbol{\epsilon}^{t'})^\dagger \rangle = (d-1)\mathbf{A}_t^{-1}\delta_{t,t'}. \tag{36}$$

Note that the $(d-1)$ factor, here and in previous equations, corresponds to the dimensionality of the perpendicular component π.

$\mathbf{M}_t$ defines a well-balanced weighted average, as it satisfies:

$$\sum_j M_{ij;t} = 1. \tag{37}$$

To show this, let us rewrite this identity in a matrix form:

$$\frac{1}{2}\mathbf{A}_t^{-1}\mathbf{J}^{(2)}_t\mathbf{u} = \mathbf{u} \tag{38}$$

where $\mathbf{u}$ is a vector of ones, $u_i = 1$. Proving (37) is therefore equivalent to showing: $\mathbf{J}^{(2)}_t\mathbf{u} = 2\mathbf{A}_t\mathbf{u}$, which follows from the definition of $\mathbf{A}_t$ (29).

This identity also allows us to check that the diagonal components in the equality (30) are consistent with the off-diagonal components. This is done by checking that on both sides of the (30), contraction with $\mathbf{u}$ gives zero.

The equation describing the collective random walk in terms of the perpendicular component π holds almost the same for the flight direction s itself. Starting from the update equation:

$$s_i^{t+1} = \theta\left[\sum_j M_{ij;t}s_j^t + \eta_i^t\right], \tag{39}$$

where $\theta(x) = x/\|x\|$ is the normalization operator, and expanding in the spin-wave approximation ($\pi_i \ll 1$), one recovers (35) with $\epsilon_i^t = \eta_i^t - (n\cdot\eta_i^t)n$ the perpendicular component of the vectorial noise η.

Parametrization

The matrices $\mathbf{M}_t$ and $\mathbf{A}_t$ are parametrized as follows:

$$M_{ij} = (1 - J\delta t n_i)\delta_{ij} + J\delta t n_{ij}, \tag{40}$$

where $n_{ij} = 1$ if j is one of i's neighbours, and 0 otherwise, and $n_i = \sum_{ij} n_{ij}$. (We drop the t index, even though n_{ij} depends on t in general); and

$$A_{ij} = [1/(2\delta t T)]\delta_{ij}. \tag{41}$$

J is interpreted as an alignment strength, and T as a temperature.

Continuous time limit and equivalence with static maximum entropy

The parametrization has a well defined continuous-time limit. When $\delta t \to 0$, (35):

$$\frac{d\boldsymbol{\pi}}{dt} = -J\boldsymbol{\Lambda}\boldsymbol{\pi} + \boldsymbol{\xi}(t), \tag{42}$$

where $\Lambda_{ij} = n_i\delta_{ij} - n_{ij}$, and $\xi_i(t)$ are i.i.d Gaussian white noises with $\langle \xi_i(t)\xi_i(t')\rangle = 2T(d-1)\delta(t-t')$, where $\delta(x)$ is Dirac's delta function.

When $\boldsymbol{\Lambda}$ varies slowly with time, (42) can be formally integrated:

$$\boldsymbol{\pi}(t) = \int_{-\infty}^{t} dt'\, e^{-J\boldsymbol{\Lambda}(t-t')}\boldsymbol{\xi}(t') \tag{43}$$

If, in addition, $\boldsymbol{\Lambda}$ is symmetric, the system reaches some equilibrium steady state. More precisely, the collective mode that is parallel to $\mathbf{u}$, which corresponds to the average direction of the flock $(1/N)\sum_i \pi_i$, follows an unconstrained random walk, as it corresponds to a the zero mode of $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}\mathbf{u} = 0$. All the other modes that are orthogonal to $\mathbf{u}$ are bounded by a restoring force. The steady-state distribution of $\boldsymbol{\pi}$ is therefore Gaussian, with $C_{ij} = \mathrm{Cov}(\pi_i, \pi_j)$ satisfying:

$$J\boldsymbol{\Lambda}\mathbf{C} = (d-1)T\left(\mathbf{1} - \frac{\mathbf{u}\mathbf{u}^\dagger}{N}\right), \tag{44}$$

where $\mathbf{1}$ is the identity matrix.

Remarkably, in the spin-wave approximation, this distribution is the same as the one obtained by the principle maximum entropy constrained by the static correlation functions:

$$P(\mathbf{s}) = \frac{1}{Z}\exp\left(\frac{1}{2}\sum_{i\neq j} J_{ij}^{\mathrm{stat}} s_i s_j\right), \tag{45}$$

with

$$J_{ij}^{\mathrm{stat}} = \frac{J}{T} n_{ij}. \tag{46}$$

One can check this by expanding (45) at small π, after setting n to be the average direction of the flock, so that $\sum \pi_i = 0$, and

$$P(\boldsymbol{\pi}) \propto \delta\left(\sum_i \pi_i\right)\exp\left(-\frac{J}{2T}\sum_{ij}\Lambda_{ij}\pi_i\pi_j\right). \tag{47}$$

By virtue of Gaussian integration rules, this distribution has the same covariance as (44), and therefore is identical.

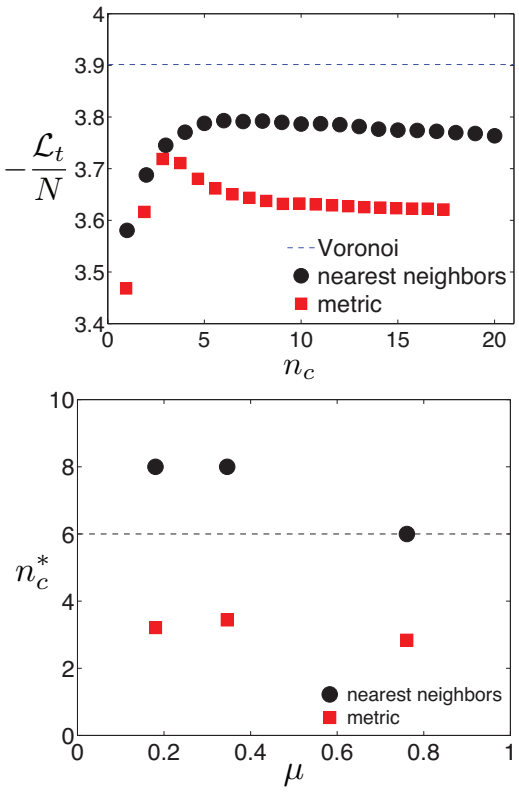

FIG. 2. Upper: comparison of the normalized log-likelihood for the nearest-neighbor and metric rules, as a function of n_c. For the metric case, for increasing values of r_c, the empirical $n_c = (1/N)\sum_i n_i$ is shown. The dashed line corresponds to the log-likelihood calculated with the Voronoi rule. Lower: Inferred interaction range n_c^* for the nearest-neighbor and metric cases, as a function of the mixing parameter μ.

Parameter inference

We can rewrite the Lagrangian (33) in a slightly different manner:

$$\begin{aligned}\mathcal{L}_t(\boldsymbol{\pi}^{t+1}|\boldsymbol{\pi}^t) = &-\frac{d-1}{2}\ln\left(\frac{\det\mathbf{A}_t}{(2\pi)^N}\right) + \frac{1}{2}\mathrm{Tr}(\mathbf{C}_{t+1}\mathbf{A}_t^\dagger)\\ &- \frac{1}{2}\mathrm{Tr}(\mathbf{J}_t^{(2)}\mathbf{G}_t^\dagger) + \frac{1}{8}\mathrm{Tr}(\mathbf{J}_t^{(2)\dagger}\mathbf{A}_t^{-1}\mathbf{J}_t^{(2)}\mathbf{C}_t^\dagger),\end{aligned} \tag{48}$$

where $\mathbf{C}_t = \boldsymbol{\pi}^t(\boldsymbol{\pi}^t)^\dagger$ and $\mathbf{G}_t = \boldsymbol{\pi}^{t+1}(\boldsymbol{\pi}^t)^\dagger$.

Under the parametrization (40),(41), the minus-log-likelihood (48) becomes (the time index is implicit from now on):

$$\frac{\mathcal{L}}{N} = \frac{d-1}{2}\ln 2T\delta t + \frac{\hat{\mathcal{L}}}{4T\delta t}, \tag{49}$$

where

$$\begin{aligned}\hat{\mathcal{L}} =& C_s^1 + C_s - 2\alpha\tilde{C}_s + \alpha^2\hat{C}_s + 2\alpha(C_{\mathrm{int}} - \alpha\tilde{C}_{\mathrm{int}})\\ &+ \alpha^2 C'_{\mathrm{int}} - 2\alpha G_{\mathrm{int}} - 2(G_s - \alpha\tilde{G}_s),\end{aligned} \tag{50}$$

and $\alpha = Jn_c\delta t$.

The various correlated functions used in this expression are defined in Table I in the main text.

In the case of non-constant n_i, n_c is defined as $(1/N)\sum_i n_i$. Note that in the case of constant $n_i = n_c$, as in the case of the nearest-neighbor model, $\tilde{C}_s = \hat{C}_s = C_s$, $\tilde{G}_s = G_s$ and $\tilde{C}_{\rm int} = C_{\rm int}$.

There are three parameters to optimize over: the interaction strengh J, the interaction range n_c, and the "temperature" T which sets the strength of noise. This last one is simply given by the condition $\partial\mathcal{L}/\partial T = 0$, which yields:

$$T = \frac{\hat{\mathcal{L}}}{2(d-1)\delta t}. \tag{51}$$

At this optimum value of T, we have

$$\frac{\mathcal{L}}{N} = \frac{d-1}{2}\{\ln[\hat{\mathcal{L}}/(d-1)] + 1\}. \tag{52}$$

Minimizing $\hat{\mathcal{L}}$, $\partial\hat{\mathcal{L}}/\partial\alpha$, then yields the optimum value of α:

$$\alpha = \frac{C_{\rm int} - \tilde{C}_s + \tilde{G}_s - G_{\rm int}}{2\tilde{C}_{\rm int} - C'_{\rm int} - \hat{C}_s}. \tag{53}$$

At this optimum, one has $\hat{\mathcal{L}} = C_s^1 + C_s - 2G_s + \tilde{\mathcal{L}}$, where

$$\tilde{\mathcal{L}} = \frac{\left(C_{\rm int} - \tilde{C}_s + \tilde{G}_s - G_{\rm int}\right)^2}{2\tilde{C}_{\rm int} - C'_{\rm int} - \hat{C}_s} \tag{54}$$

is the only term that depends on the interaction matrix n_{ij}. Therefore, to find the optimum interaction range n_c in the case of the nearest-neighbor model, one just needs to minimize $\tilde{\mathcal{L}}(n_c)$.

Consistency with the static approach

To recover the static inference equations, we start by rewriting the dynamical inference equations, Eqs. (51) and (53), explicitly:

$$J = \frac{1}{n_c}\frac{\Omega + (d-1)T_0}{C'_{\rm int} + \hat{C}_s - 2\tilde{C}_{\rm int}}, \tag{55}$$

$$T = T_0 + \frac{C_s^1 - C_s}{2(d-1)\delta t} - \frac{J\,n_c\delta t}{2(d-1)}\left(\frac{\tilde{C}_s - \tilde{G}_s}{\delta t} + \Omega\right) \tag{56}$$

with $n_c = (1/N)\sum_i n_i$ and

$$T_0 = \frac{C_s - G_s}{\delta t(d-1)}, \quad \Omega = \frac{G_{\rm int} - C_{\rm int}}{\delta t}. \tag{57}$$

When the system is at steady state, we have $C_s^1 \approx C_s$ and $\Omega \approx (2Nn_c)^{-1}\sum_{ij}\pi_i\pi_j\frac{dn_{ij}}{dt}$ (directly from definitions in Table I of the main text and Eq. 57); the second term in Eq. (56) cancels and $T \approx T_0$ for small δt. If we further assume that data was actually generated by *exactly* the class of models we are trying to infer (which may not be the case in general, as we are looking at effective descriptions), we have exactly $T = T_0$. If in addition neighbor changes are slow, then $\Omega \approx 0$ and Eq. (44) implies $\tilde{C}_{\rm int} \approx C'_{\rm int}$. Eq. (55) thus gives

$$\frac{Jn_c}{T} \approx \frac{d-1}{\hat{C}_s - \tilde{C}_{\rm int}}, \tag{58}$$

which is the result of the static inference [8]. Note however that in addition to recovering the alignment strength, the dynamical inference procedure allows us to separate the interaction coupling J from the temperature T.

Spin wave expansion of the Topological Vicsek model

As described in the main text, to test our dynamical inference method we generated synthetic data with the Topological VM defined by

$$\theta_i^{t+\delta t} = {\rm Arg}[s_i^t + J_V\delta t\sum_j n_{ij}s_j^t] + \sqrt{\delta t}\,\xi_i^t, \tag{59}$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\,s_i^{t+\delta t}. \tag{60}$$

In this section, we show that Eq. (59) is in fact equivalent in the spin-wave limit to an update equation of the same kind as Eqs. (35),(40) and (41). To this aim, it is convenient to rewrite Eq. (59) in the following equivalent form

$$s_i^{t+\delta t} = \frac{s_i^t + J_V\delta t\sum_j n_{ij}s_j^t}{\|s_i^t + J_V\delta t\sum_j n_{ij}s_j^t\|} + \sqrt{\delta t}\,\epsilon_i^t, \tag{61}$$

$$r_i^{t+\delta t} = r_i^t + v_0\,\delta t\,s_i^{t+\delta t}, \tag{62}$$

where ϵ_i is a delta-correlated noise perpendicular to s_i with variance $2(d-1)T_V$ (i.e. whose effect is the same as the angular noise appearing in Eq. (59)).

In the large polarization regime we can perform a spin wave expansion $s_i = \pi_i + n\sqrt{1-\pi_i^2}$, where n is a vector representing the global direction of motion and π_i is the component of the direction s_i perpendicular to n. We can now expand the normalization at the r.h.s. in Eq. (61) with respect to π_i^2 to get

$$\|s_i^t + J_V\delta t\sum_j n_{ij}s_j^t\| = 1 + \delta t J_V n_i + {\rm O}(\pi^2) \tag{63}$$

where $n_i = \sum_j n_{ij}$. Eq. (61) then leads to the following

update equation for the $\{\pi_i\}$

$$\begin{aligned}\pi_i^{t+\delta t} &= \frac{\pi_i^t + \delta t J_V \sum_j n_{ij}\pi_j^t}{1+\delta t J_V n_i} + \sqrt{\delta t}\epsilon_i \\ &= \left(1 - \delta t \frac{J_V}{1+\delta t J_V n_i}\right)\pi_i^t \\ &\quad + \delta t \frac{J_V}{1+\delta t J_V n_i}\sum_j n_{ij}\pi_j^t + \sqrt{\delta t}\epsilon_i. \end{aligned} \tag{64}$$

When δt is small, we can disregard fluctuations in n_i and Eq. (64) is of the same form of Eqs. (35) with the parametrization defined in (40)-(41) and

$$J = \frac{J_V}{1+\delta t J_V n_V}. \tag{65}$$