Questioning the activity of active matter in natural flocks of birds

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*

The correlated motion of large bird flocks is an instance of self-organization where global order emerges from local interactions. Despite the analogy with ferromagnetic systems, a major difference is that flocks are active – animals move relative to each other, thereby dynamically rearranging their interaction network. Although the theoretical importance of this off-equilibrium ingredient has long been appreciated, its relevance to actual biological flocks remains unexplored. Here we introduce a novel dynamical inference technique based on the principle of maximum entropy, which takes into account network reshuffling and overcomes the limitations of slow experimental sampling rates. We apply this method to three-dimensional data of large natural flocks of starlings, inferring independently the strength of the social alignment forces, the range of these forces, and the noise. We show that the inferred timescale of bird alignment is much smaller than the timescale governing the rearrangement of the interaction network. We verify that, following from this observation, an equilibrium inference method assuming a fixed interaction network gives results that are fully consistent with the dynamical inference. We conclude that the birds' flight orientations are in a state of local equilibrium.

Animal groups moving in concert such as mammal herds, fish schools, insect swarms, and bird flocks provide striking biological examples of how local coordination can result in large-scale collective behavior [1–3]. Such examples differ radically from standard systems in statistical physics in that their constituents are themselves active: they constantly move by self-propulsion, pumping energy into the system and keeping it out of equilibrium [4–7]. However the motion of particles is not enough to make active matter genuinely active. Crystal particles translated together along parallel trajectories just describe a moving solid, and cannot be considered active. What makes a system truly active is the rearrangement of the interaction network as individuals diffuse within the group and change their neighbors. The effect of these rearrangements is very significant: among other things, theoretical studies predict that network motion in active systems enhances group-level alignment, lowering from 3 to 2 the minimal integer dimension at which an ordering transition occurs and changing the exponents governing the decay of long-range correlations [4, 8]. In short, network rearrangements could drastically change the group dynamics and the nature of their coherent motion. However, despite a deep theoretical understanding, their actual impact for biological groups has remained largely unexplored. Here we propose a dynamical inference method to investigate their role in natural flocks of starlings.

The impact of activity in biological groups depends on the interplay between two time scales: the first one is the scale of local relaxation, $\tau_{\rm relax}$, defined as the characteristic time needed to relax locally the order parameter if the interaction network remained fixed (in the case of flocks the order parameter is the orientation, *i.e.* the direction of motion); the second time scale is the network reshuffling time, $\tau_{\rm network}$, defined as the average time it takes for an individual to renew its interacting neighborhood. If $\tau_{\rm relax} \sim \tau_{\rm network}$, the system is out of equilibrium and the exotic physics of active matter applies. If, on the other hand, $\tau_{\rm relax} \ll \tau_{\rm network}$, the dynamics are adiabatic, closely following the equilibrium state of the network as it slowly evolves. In this work we investigate what regime applies to natural flocks of birds.

To address this issue we need to study the dynamical rules of alignment in a time-resolved manner, as they are carried out in the wild or in the laboratory. Recent work has allowed for a detailed description of the local dynamical rules underlying collective behavior using time-dependent data, in surf scoters [9], prawns [10] and fish schools [11–15], typically restricted to small groups in one or two dimensions. In these studies, the behavior (acceleration or turning) of individuals was mapped as a function of the parameters of their immediate environment, such as the distance or orientation of their neighbors. While the local rules of interaction were learned using small groups, they could often be used to predict some large-scale properties of group behavior [13, 15]. In a separate approach based on static data, snapshots of three-dimensional images of flocks of starlings in the wild were analysed and mathematically mapped onto statistical mechanics models of ferromagnets [16], lending insight into the ordering mechanism and the local topology of the interaction network [17, 18]. All these studies provide a quite detailed view of the rules underlying col-

lective behavior, but do not address the question of network activity. In fact, the impact of network reshuffling on emergent behavior in animal groups has been given surprisingly little attention, despite the importance conferred to it by the theory of active matter.

Our approach is based on the direct inference of dynamical rules from data, using maximum entropy models constrained to reproduce the *dynamical* correlation functions of the data, thus taking into account the reshuffling of the network [19]. A crucial and widespread problem faced by any inference approach, ours included, is that in all experiments motion is sampled at a finite rate. Most inference methods are based on approximated schemes of integration of continuous differential equations and they therefore fail when the sampling rate is slow compared to the relaxation time. The reason of this fact is somewhat intuitive: if two subsequent snapshots are too far in time from each other, any discrete inference method mistakes dynamical correlation for static interaction, thus yielding an inferred interaction range that is *larger* than the actual one.

Here we develop a new inference procedure based on the exact integration of the maximum-entropy dynamical equations that overcomes the issue of finite experimental sampling rates. We test the method against simulations for a wide range of sampling rates and apply it to data of starling flocks of sizes ranging from 50 to 596, obtained by three-dimensional reconstruction of stereoscopic images [20–22], where bird positions were tracked over a few seconds [23] (see Materials and Methods and Table S1 for a summary of the data). We infer the relevant parameters of alignment—the strength, range, and noise of the alignment forces between neighbors. We find that the alignment relaxation time, $\tau_{\rm relax}$, is an order of magnitude smaller than the network rearrangement time, $\tau_{\rm network}$. Accordingly, we show that the parameters learned from the dynamics agree remarkably well with those obtained by an equilibrium-like inference assuming a fixed network [17], indicating that network rearrangement plays a minor role in the flock's alignment. These results demonstrate that natural flocks are in a state of quasi-equilibrium, meaning that the relatively slow rearrangement of the interaction network does not affect their ordering dynamics.

Results

Dynamical inference.

Inferring the dynamical rules of behavior usually relies on the choice of a particular model of collective behavior adapted to the context at hand. Unfortunately, there are many candidates for such models. Instead of imposing a form for the rules of alignment in natural flocks, we follow Ref. [19] and apply the principle of maximum entropy to the trajectories of all birds in the group. We look for model distributions of the stochastic process that are as random as possible, while agreeing with the data on a key set of carefully chosen observables. In a flock of size N, call $\vec{s}_i(t)$ the three-dimensional flight orientation of bird i at time t. The maximum entropy distribution over possible flock trajectories that is consistent with the correlation functions $\langle \vec{s}_i(t)\cdot\vec{s}_j(t)\rangle$, as well as their derivatives $\langle d\vec{s}_i(t)/dt\cdot\vec{s}_j(t)\rangle$, can be exactly mapped, in the limit of strong polarization $P\equiv(1/N)\|\sum_i\vec{s}_i\|\approx 1$, onto the following stochastic differential equation (see SI and Ref. [19]):

$$\frac{d\vec{s}_i}{dt}=\left(\sum_j J_{ij}\vec{s}_j+\vec{\xi}_i\right)_\perp , \qquad (1)$$

where $\vec{\xi}_i$ is a random white noise, and where the projection $\vec{x}_\perp\equiv(\vec{s}_i\times\vec{x})\times\vec{s}_i=\vec{x}-\vec{s}_i(\vec{x}\cdot\vec{s}_i)$ onto the plane perpendicular to $\vec{s}_i$ ensures that $\vec{s}_i$ remains of norm 1.

The model is easily interpretable, and can be viewed as a generalization of the Vicsek model [24]. Each bird modifies its flight direction according to a weighted average of the directions of its neighbors, $\sum_j J_{ij}\vec{s}_i$, to which some noise $\vec{\xi}_i$ is added. The interaction matrix J_{ij} encodes how much bird i is influenced by (*i.e.* interacts with) bird j. Given the experimentally measured correlation functions, entropy maximization yields equations that fix the values of the noise amplitude, and the interaction matrix J_{ij}. This matrix has too many parameters to be reliably determined from the data, but we can reduce its complexity by parametrizing it [17]. Our choice of parametrization is guided by the results of Ref. [25], where it was shown that the interaction decays exponentially with the topological distance k_{ij} between birds,

$$J_{ij}=J\exp(-k_{ij}/n_c) , \qquad (2)$$

where k_{ij} denotes the (time-dependent) rank of bird j among the neighbors of bird i ranked by distance. This interaction form has just two parameters: n_c is the topological interaction range, while J is the overall strength of the interaction. Later we will see that our results are robust to other choices of the parametrization. For simplicity, the noise is taken to be uncorrelated among birds and of uniform magnitude T (in units of inverse time), by analogy with physical temperature: $\langle\vec{\xi}_i(t)\cdot\vec{\xi}_j(t')\rangle=2\,d\,T\,\delta_{ij}\delta(t-t')$, where d is the space dimension ($d=3$ in the following).

The parameter J, which has units of inverse time, quantifies the strength of the social force and sets the timescale for the relaxation of the system. Since there are approximately n_c neighbors acting on each individual, the total alignment force is of order Jn_c, and thus the characteristic time scale of relaxation of the orientational degrees of freedom is $\tau_{\rm relax}\equiv(Jn_c)^{-1}$. This is not entirely obvious, as the theory has a zero mode, implying long range correlations both in space and time; hence, one would expect the relaxation time to be infinite. However, the fluctuations associated to the zero mode have

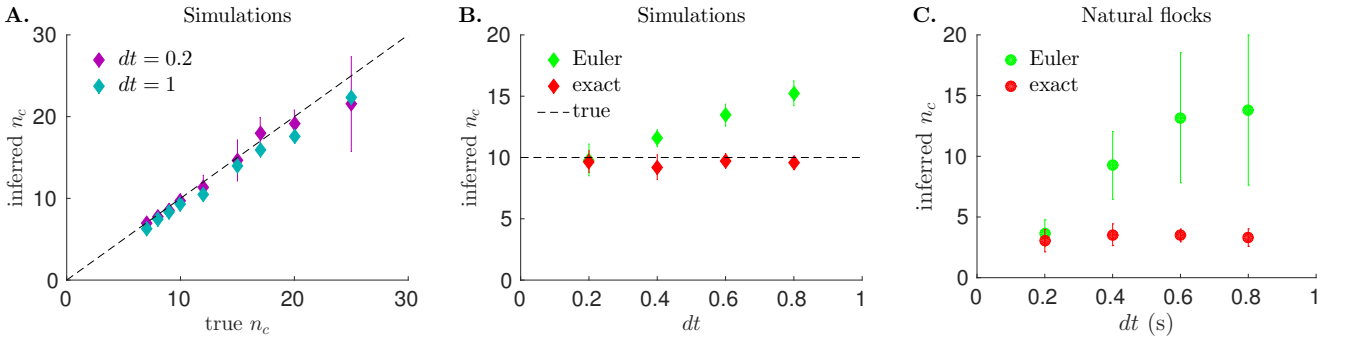

FIG. 1: Performance of the inference methods on the predicted interaction range n_c. **A.** Inferred versus real n_c obtained by applying our new inference method to simulated data generated with Eq. 1 at various interaction ranges. The method performs well for different values of the sampling rate dt. **B.** Dependence of the inferred n_c on the sampling time dt. On simulated data with $n_c = 10$ (dashed line), the inference method based on exact integration (red points) performs well regardless of the sampling time dt. By contrast, the inference method based on Euler's integration method (green points) overestimates the true interaction range at large dt. **C.** A similar trend is observed when we apply the two inference procedures to real flocking data, as illustrated here on one flocking event. Note that in this case the true value is not known.

very long wavelength, so that they change the direction of a bird in unison with that of its local neighbors and therefore they do not contribute to the competition between relaxation of the orientation and network reshuffling. The modes that disorder the network on the scale of the interaction have wavelengths much shorter than the size of the system. For this reason one can show that the relevant correlation function decays exponentially, with a finite scale $(Jn_c)^{-1}$ (see SI for a detailed discussion).

Coping with a finite sampling time.

In principle, to learn the parameters of Eq. 1 from data requires to know the instantaneous derivative of each bird's flight orientation $d\vec{s}_i/dt$. However, in practice the configurations of the flock are sampled at some finite experimental rate, and we must infer the model using consecutive configurations of the birds' orientations separated by the sampling time dt: $\vec{s}_i(t)$ and $\vec{s}_i(t+dt)$. If dt is small enough (we shall see later what this means), inference can be performed by integrating Eq. 1 using Euler's approximate rule:

$$\vec{s}_i(t+dt) \approx \vec{s}_i(t) + dt \left(\sum_j J_{ij} \vec{s}_j \right)_\perp + \sqrt{2T\,dt}\,\vec{\eta}_i, \quad (3)$$

where $\vec{\eta}_i$ is a normally distributed vector orthogonal to $\vec{s}_i$ and of variance 1 in each direction. The conditional likelihood of the data given the model, $P[\{\vec{s}_i(t+dt)\}|\{\vec{s}_i(t)\}]$, can be written in Gaussian form after expanding Eq. (3) in the spin-wave approximation of strong polarizations (see Materials and Methods). Maximizing this likelihood yields best-fit values for the alignment parameters n_c, J and T. The procedure is easy to implement, as J and T have analytical expressions as a function of the observables (see Ref. [19] and SI).

Independently of the dynamical model actually used, Euler's rule is used by virtually all methods that try to fit a dynamical equation to a discrete time series of biological data [11–13]. Identifying the discrete experimental dynamics with the discretization of the continuous dynamical equations may seem a natural thing to do. However, the sampling time dt is not a natural, intrinsic time-scale of the biological system under investigation, but instead only depends on our experimental technique. Imagine that dt is much larger than the intrinsic relaxation time-scale, $\tau_{\rm relax}$: in this case, between two subsequent time frames information has spread far beyond the directly interacting neighbours, so we expect Euler's rule to overestimate the range of the interaction. In practice, by using Euler's rule we are coarse-graining the true dynamics over time, and hence get renormalized parameters that may differ considerably from the true ones. We stress that this issue is common to any inference procedure based on dynamical data sampled at relatively slow rates.

To go beyond Euler's approximation, we exploit the fact that Eq. 1 is linear in the spin-wave approximation and therefore exactly solvable for J_{ij} constant in time. To see how to solve it, it is convenient to rewrite Eq. 1 so that the alignment force consists of a sum with balanced weights, by subtracting $\sum_l J_{il}\vec{s}_i$ from it, with no consequence because $(\vec{s}_i)_\perp = 0$. The equation now takes the form:

$$\frac{d\vec{\mathbf{s}}}{dt} = \left(-J\mathbf{\Lambda}\vec{\mathbf{s}} + \vec{\boldsymbol{\xi}} \right)_\perp . \quad (4)$$

Bold symbols denote vectors and matrices over bird indices; the matrix $\Lambda_{ij} \equiv \delta_{ij} \sum_l n_{il} - n_{ij}$, where $n_{ij} = e^{-k_{ij}/n_c}$ is the (dimensionless) connectivity matrix according to parametrization (2). $\mathbf{\Lambda}$ is analogous to a Laplacian defined on a lattice, and has balanced weights: $\sum_j \Lambda_{ij} = 0$. In the spin-wave approximation, where all orientations $\vec{s}_i$ point in almost the same direction,

this balancing ensures that $\mathbf{\Lambda}\vec{\mathbf{s}}$ has almost no contribution along the common direction of flight, implying $(\mathbf{\Lambda}\vec{\mathbf{s}})_\perp \approx \mathbf{\Lambda}\vec{\mathbf{s}}$ (see Materials and Methods and SI). Equation 4 then becomes linear and can be integrated:

$$\vec{\mathbf{s}}(t+dt) = e^{-J\mathbf{\Lambda}dt}\vec{\mathbf{s}}(t) + \int_0^{dt} du\, e^{-J\mathbf{\Lambda}(dt-u)}\vec{\boldsymbol{\xi}}_\perp(t+u)\ . \quad (5)$$

This result assumes a constant J_{ij}, which is a good approximation if its variations are slow compared to the integration time, $dt \ll \tau_{\rm network}$. Conveniently, this same limit is required to be able to track individuals from frame to frame and thus to collect dynamical data in the first place (see discussion). From the exponentials in this equation we explicitly see that the experimental sampling time dt competes with a product of the interaction strength and of the connectivity matrix, confirming that the natural scale of relaxation is $\tau_{\rm relax} = (Jn_c)^{-1}$. As expected, expanding (5) for $dt \ll \tau_{\rm relax}$, gives back Euler's approximation, Eq. (3). The integrated noise in the right-hand side of (5) is Gaussian, of mean zero and covariance $4T\int_0^{dt} du\, e^{-J\mathbf{\Lambda}u}e^{-J\mathbf{\Lambda}^\dagger u}$. Using this exact solution allows us to write an explicit expression for the (still Gaussian) conditional likelihood $P[\{\vec{s}_i(t+dt)\}|\{\vec{s}_i(t)\}]$, which can then be maximized over the parameters of the model, as before (see Materials and Methods).

To test our new dynamical inference we simulated synthetic flocking trajectories using the model of Eq. 1, for various values of the interaction range n_c, while keeping $Jn_c = 1.5$, hence $\tau_{\rm relax} \approx 0.7$; the temperature T was chosen so as to have polarization $\Phi \approx 0.99$ similar to real flocks (see Materials and Methods). We then infer the parameters of the model using either Euler's rule, or the result of exact integration as explained above, for different values of the sampling time ranging from $dt = 0.2$ to $dt = 0.8$. We find that the our new method based on exact integration predicts the interaction range n_c well, regardless of dt (Fig. 1A and B). By contrast, the performance of the inference method based on the approximated Euler's formula depends strongly on dt (Fig. 1B). Its prediction is good at small dt, but it largely overestimates n_c at large dt.

We also test the effect of changing the sampling time dt on real data of natural flocks. The minimum sampling time allowed by the experimental conditions is $dt = 0.2\,$s because below this time individual bird flapping dominated the trajectories (see Materials and Methods). To assess the effect of changing the sampling time we artificially varied dt from 0.2 to 0.8 s. Remarkably, although we cannot compare the inferred value of n_c to the truth as in simulations, we observe a similar trend as a function of dt (Fig. 1C). The inference based on exact integration gives a consistent prediction for n_c regardless of dt, while that based on Euler's approximation only agrees with it for small dt. This suggests that the sampling time of 0.2 s is of the same order as the orientation relaxation time $\tau_{\rm relax}$, as we will confirm below. It also indicates that the inference method based on exact integration is extracting the parameters of alignment reliably.

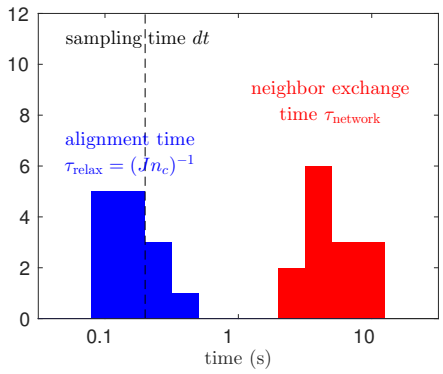

FIG. 2: Comparison between the two relevant time scales of active matter, as inferred in 14 natural flocks using our inference method based on exact integration. Histograms of the neighbor exchange time $\tau_{\rm network}$ versus the local alignment time $\tau_{\rm relax} = 1/Jn_c$, show that the relaxation of orientations is much faster than the turnover of neighbors. Note that the experimental sampling time $dt = 0.2\,$s (dashed line) is of the same order as the alignment time, justifying the use of exact integration.

Separation of time scales in natural flocks.

Confident with our dynamical inference method, we now move on to the analysis of its predictions on data of natural starling flocks. To assess the importance of network activity for the alignment dynamics, we need to compare the two time scales $\tau_{\rm network}$ and $\tau_{\rm relax}$ governing the evolution of the network and the relaxation dynamics of orientations.

We estimate, for each flocking event, the network reshuffling time $\tau_{\rm network}$ as the characteristic decay time of its autocorrelation function $C_{\rm network}(t) = \sum_{ij} n_{ij}(t_0)n_{ij}(t_0+t)$, by fitting $C_{\rm network}(t) \approx C_0 \exp(-t/\tau_{\rm network})$ (Fig. S2). We then compare its value to the relaxation time of the orientational dynamics, $\tau_{\rm relax} = (Jn_c)^{-1}$ learned from the dynamical inference method. The results are summarized in Fig. 2. The two time scales clearly separate, indicating that network variations are slow compared to the other time scales of the problem, in particular the relaxation of orientations. In addition, the estimate of $\tau_{\rm relax}$ is relatively close to the experimental sampling time $dt = 0.2\,$s, confirming in hindsight the need to use an exact integration method.

Local equilibrium in flocks.

The separation of time scales suggests that flocks are in a state of local equilibrium. The network of interactions does change over time, but it does so slowly enough for the dynamics of flight orientations to catch up before neighbours reshuffle. In other words, the orientation

dynamics tracks network changes adiabatically. It is important to stress that this statement is true in a local sense, as it holds within the interacting neighborhood. The network changes slowly *over the interaction length-scale* compared to the relaxation of the orientations over the same local scale.

This claim leads to a prediction: since flocks behave as if they were in local equilibrium, an equilibrium inference procedure, which takes as an input the local correlation in space calculated over a snapshot of the birds' flight directions [17], should be consistent with the results of dynamical inference. If, on the contrary, the system's activity strongly affected the alignment dynamics, or if our approach failed to capture important aspects of the matter, one should find different results. Is this prediction satisfied in natural flocks of birds?

To study this question, we first briefly recall the equilibrium-like inference method used in [17], and its relation to our dynamical framework. For symmetric J_{ij}, Eq. 1 is the Langevin equation derived from the Hamilonian of the Heisenberg model

$$\mathcal{H} = -\frac{1}{2}\sum_{i,j} J_{ij}\vec{s}_i \cdot \vec{s}_j. \tag{6}$$

When J_{ij} varies slowly in time, the fluctuations of $\vec{s}_i$ are in quasi-equilibrium and distributed according to Boltzmann's law:

$$P(\vec{s}_1, \ldots, \vec{s}_N) \sim \exp\left(-\mathcal{H}/T\right). \tag{7}$$

We recognize the maximum entropy distribution consistent with the local correlation index $\sum_{ij} n_{ij}\langle \vec{s}_i \vec{s}_j \rangle$ fitted in Ref. [17]. In practice, the equilibrium inference consists in maximizing the likelihood of Eq. 7 over its parameters n_c and J/T (see Materials and Methods and SI). It follows from the relation between the Lanvegin equation and Boltzmann's law that, if the variations of n_{ij} are slow compared to the dynamics of $\vec{s}_i$, $\tau_{\rm network} \ll \tau_{\rm relax}$, this inference procedure should give an accurate estimate of the alignment parameters.

Specifically, both equilibrium and dynamical inference methods give as an output the interaction range, n_c, whose value can therefore be directly compared in the two cases. On the other hand, while the dynamic inference provides the strength of the interaction, J, and the strength of the noise, T, separately, the equilibrium inference gives a single value for the ratio J/T, which is therefore the quantity to compare. Note that because the equilibrium inference only outputs this dimensionless ratio, it does not give any indication of timescales, making the validity of the equilibrium description impossible to assess self-consistently without recurring to dynamical information.

We proceed to the comparison between the two methods in natural flocks, by learning the model parameters from data using both the dynamical inference method and the equilibrium-like procedure we have just described, for each of the 14 flocking events. The results, plotted in Fig. 3, show that the dynamic inference gives parameters that are consistent, within error bars, with those of the equilibrium inference. This is true both for the interaction range, n_c, and for the coupling constant to noise ratio, J/T. This result confirms that the alignment dynamics of flocks are in an effective state of equilibrium.

It is important to stress that these two inference procedures are based on completely different information: the equilibrium inference uses the instantaneous statistics of orientations, while the dynamical inference exploits the difference between two snapshots to learn how they vary in time. Thus, their agreement not only means that neighborhood variations have little effect on the alignment properties of the flock, but it also makes a strong case for the validity of our approach. Finally, we note that using Euler's rule instead of the exact integration in the dynamical inference gives a significantly worse agreement, even for $dt = 0.2\,$s when considering all 14 flocks (Fig. S1).

Up to now we have used an exponential form for the interaction matrix, $J_{ij} = Je^{-k_{ij}/n_c}$, where k_{ij} is the topological distance between i and j. To test the robustness of our results against the choice of parametrization, we repeated the analysis with a step function: $J_{ij} = J$ if $k_{ij} \leq n_c$, and 0 otherwise. It was argued [25] that the interaction range $n_c^{\rm step}$ parametrizing the step form should be related to $n_c^{\rm exp}$ of the exponential form through the relation $n_c^{\rm step} \approx 2n_c^{\rm exp}$, in order to get the same mean rank between interaction partners (see SI). Our results agree very well with this prediction, both for the equilibrium and dynamical inference procedures (Fig. S3). Thus, the two parametrizations are interchangeable and give equivalent results.

Discussion

We have presented a new dynamical method for inferring the range and strength of alignment interaction between birds in natural flocks. Applying it to trajectories of starling flocks has allowed us to learn the microscopic dynamical rules leading to collective behavior in the largest animal groups yet. Compared to previous work on dynamical inference [19], the present method can deal with any experimental sampling rate, even lower than the relaxation rate of the orientations. Thanks to this, the method can be applied to real data in which the relation between these two time scales is not known *a priori*. The new method only works as long as the experimental sampling time is faster than network rearrangement, $dt \ll \tau_{\rm network}$. This, however, is hardly a constraint, as this same condition is necessary for the very possibility to experimentally collect dynamical data in the first place. Any tracking procedure needs to follow each individual between consecutive time frames, which is only possible if individuals do not significantly change their neighborhood [23]. If the distance traveled by individuals in the centre of mass reference frame is of the

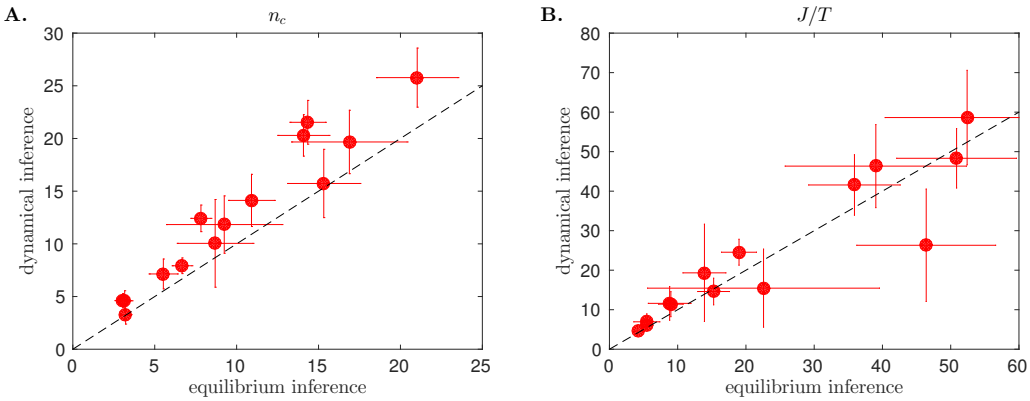

FIG. 3: Inference on natural flocks. For each of the 14 flocking events, the parameters of the model were inferred using either the dynamical inference method presented here, with $dt = 0.2\,\mathrm{s}$, or an equilibrium inference method as in [17]. **A**. Both methods agree well on the predicted value of the alignment range n_c. **B.** While the dynamical method infers the alignment strength J and the noise amplitude T separately, the equilibrium method only infers their ratio J/T, the value of which is consistent between the two methods.

same order (or larger) than the inter-particle distance, dynamical tracking becomes impossible. We therefore expect our method to work in the most diverse types of biological data.

Theoretical studies of active matter suggest that out-of-equilibrium effects induced by the rearrangement of the interaction network should play a major role in the alignment dynamics of flocks. In this light, any attempt to understand the properties of active biological matter based on an equilibrium approach seems questionable. The success of such an approach in Refs. [17], [18] and [25] is all the more surprising. Since these descriptions postulate an effective equilibrium distribution of the configurations of the flock, they are not informative of the out-of-equilibrium effects caused by network activity, and one could rightfully ask why they work and whether their parameters should be trusted. Our work provides an answer to this question by showing that flocks are in a state of local equilibrium, due to the rapid relaxation of orientations compared to the slow rearrangement of the network. A consequence of this separation of timescales is that the results of the full dynamical inference method are, within error bars, the same as a purely equilibrium method, which assumes that the interaction network is fixed. Thus, the equilibrium inference seems to be justified in this system, not only as a formal mathematical equivalence, but as tool to extract *bona fide* biological traits, including its ability to predict pairwise as well as higher-order correlations within the flock. Note that while the equilibrium approach is mathematically simpler and computationally less expansive than the dynamical one, thus making it possible to obtain results even in large groups that would be beyond the reach of a dynamical analysis, a dynamical approach is always needed to infer the timescale of the ordering mechanism as well as the noise amplitude.

Do our results mean that flocks are just like ferromagnets and that the dynamics of the interaction network has no role? The answer to this question is clearly no. What we have shown here is that the directions of motion over the local scale of interaction relax on a faster time scale than the network. But the network *does* move, neighbours are eventually reshuffled and this will have consequences on collective motion. To fully appreciate this point we must stress once again the difference between local, short-wavelength modes, which rules interaction and tune the balance between relaxation and reshuffling, and long-wavelength modes, which rule the long time and long distance correlations in the system. If we are interested in such large-scale properties, a hydrodynamic approach is always necessary [4] – even though orientations may relax fast, the long term motion of the active fluid needs to be described by taking into account the dynamics of the density field. Another important point to mention is that even when the network is rather stable (as in flocks), individuals do not sit on a regular lattice, so that heterogeneities in the network connectivity may arise. Their dynamics, even though slow, may be crucial to understanding long-term off-equilibrium properties. Also note that the slow rearrangement of the network can be seen as a consequence of the high degree of polarization in the studied flocks [26], and may be faster in less ordered groups. In conclusion, our work suggests to approach active systems without prior assumptions about their exotic properties. Depending on what our aim is in the study of collective motion, we should assess carefully the relevant time scales and choose accordingly the analytic tools that best suit our needs.

Materials and methods

Flocking data.

The three-dimensional trajectories of all birds were reconstructed using imaging techniques. Stereoscopic experiments on natural flocks of European starlings were performed in the field in Rome using three high speed machine vision cameras shooting at 170 fps . The stereoscopic video acquisitions were then processed using a novel purpose-built three-dimensional tracking algorithm based on a recursive global optimization method [23]. This algorithm is extremely powerful, allowing for the reconstruction of full 3D trajectories of all individuals in groups of several hundreds individuals. We collected 3D data from 12 flocking events with sizes ranging from 50 to 600 individuals, and lasting from 2s to 6s (for details on the experiments and the dataset see Table S1 and [21, 27]).

To avoid interference from birds flapping, which occurs at frequency $\approx$ 10 Hz, we subsampled all the 3D sequences so that two snapshots are separated by $dt' = 0.1$ s. The instantaneous flight orientations were estimated by $\vec{s}_i(t) = [\vec{r}_i(t+dt') - \vec{r}_i(t)]/\|r_i(t+dt') - \vec{r}_i(t)\|$. To avoid overlap between two subsequent evaluations of $\vec{s}_i(t)$, we used $dt = 2dt' = 0.2$ s. The lower sampling rates of Fig. 1C, were obtained by taking $dt' = 0.2$, 0.3, and 0.4 s.

Simulated data.

Data were simulated in three dimensions with the continuous Vicsek model of Eq. 1 with the interaction matrix of Eq. 2. The positions $\vec{r}_i$ of individuals are updated according to $d\vec{r}_i/dt = v_0\vec{s}_i$, with $v_0 = 1$. The simulations were set in a $8 \times 8 \times 8$ box with periodic boundary conditions, and $N = 512$ birds, so that density is exactly 1. We set $\sqrt{2T} = 0.15$ to obtain a polarization $P \approx 0.99$ similar to natural flocks. Eq. 1 was integrated using Euler's method with a simulation step $dt_{\rm sim} = 0.01$ that is much smaller than any other time scale in the system. The interaction range n_c varied from 7 to 25, and the interaction strength was picked so that $Jn_c = 1.5$. The flocks were first brought to a steady state before taking snapshots for analysis.

Spin-wave approximation.

The polarization P quantifies the level of order in the system. When $P \approx 1$, we can expand each $\vec{s}_i$ around the common direction of flight $\vec{n} \equiv (1/NP)\sum_i \vec{s}_i$. This expansion gives $\vec{s}_i = \vec{\pi}_i + \sqrt{1-\vec{\pi}_i^2}\vec{n} \approx \vec{\pi}_i + (1-\vec{\pi}_i^2/2)\vec{n}$, with $\vec{n}\cdot\vec{\pi}_i = 0$. At leading order in $\vec{\pi}_i \ll 1$, Eq. 4 becomes

$$\frac{d\vec{\pi}_i}{dt} = -J\sum_j \Lambda_{ij}\vec{\pi}_j + \vec{\xi}_{i\perp}, \tag{8}$$

with $\langle\vec{\xi}_{i\perp}(t)\vec{\xi}_{j\perp}(t')\rangle = 4T\delta_{ij}\delta(t-t')$. Similarly, the equilibrium distribution (Eq. 7) can be expanded into

$$P(\vec{\boldsymbol{\pi}}) = \frac{1}{Z}e^{-(J/T)\sum_{ij}\Lambda_{ij}\vec{\pi}_i\cdot\vec{\pi}_j}. \tag{9}$$

Since this distribution is Gaussian, Z can be calculated analytically and reads: $Z = (2\pi T/J)^{(N-1)}\prod_{\lambda_k>0}\lambda_k^{-1}$, where λ_k are the eigenvalues of the matrix Λ_{ij}.

Maximum likelihood Inference.

The *equilibrium inference* is performed by maximizing the likelihood of the data given by Eq. 9 over the parameters n_c and (J/T) (see SI for detailed formulas).

The *dynamical inference based on Euler's rule* is implemented by maximizing the likelihood $P(\{\vec{\pi}_i(t+dt)\}|\{\vec{\pi}_i(t)\})$ calculated from Euler's formula (Eq. 3). This likelihood reads

$$(4\pi T dt)^{-N}e^{-\frac{1}{4Tdt}\sum_i[\vec{\pi}_i(t+dt)-\vec{\pi}_i+Jdt\sum_j\Lambda_{ij}\vec{\pi}_j]^2}. \tag{10}$$

The *dynamical inference based on exact integration* uses Eq. 5, rewritten as $\vec{\boldsymbol{\pi}}(t+dt) = e^{-J\boldsymbol{\Lambda}dt}\vec{\boldsymbol{\pi}}(t) + \vec{\boldsymbol{\epsilon}}$, where $\vec{\boldsymbol{\epsilon}}$ is a zero-mean Gaussian vector of covariance $\langle\vec{\boldsymbol{\epsilon}}\vec{\boldsymbol{\epsilon}}^\dagger\rangle = 4T\int_0^{dt}du\,e^{-J\boldsymbol{\Lambda}u}e^{-J\boldsymbol{\Lambda}^\dagger u} = \mathbf{X}^{-1}$. The conditional likelihood $P(\{\vec{\pi}_i(t+dt)\}|\{\vec{\pi}_i(t)\})$ now reads

$$\frac{\det(\mathbf{X})}{(2\pi)^N}e^{-\frac{1}{2}[\vec{\boldsymbol{\pi}}(t+dt)-e^{-J\boldsymbol{\Lambda}dt}\vec{\boldsymbol{\pi}}(t)]^\dagger\mathbf{X}[\vec{\boldsymbol{\pi}}(t+dt)-e^{-J\boldsymbol{\Lambda}dt}\vec{\boldsymbol{\pi}}(t)]}. \tag{11}$$

Depending on whether one uses Euler's or exact integration rules, Eq. 10 or 11 is maximized over J, T and n_c (see SI for detailed formulas).

In all three inference procedures, the parameters are learned for each time t. Then the median and the associated standard error are calculated for each flocking event.

Acknowledgements.

Work in Paris was supported European Research Council Starting Grant 306312. Work in Rome was supported by IIT-Seed Artswarm, European Research Council Starting Grant 257126, and US Air Force Office of Scientific Research Grant FA95501010250 (through the University of Maryland). F.G. acknowledges support from EU Marie Curie ITN grant n. 64256 (COSMOS) and Marie Curie CIG PCIG13-GA-2013-618399.

[1] Camazine S, et al. (2001) *Self-Organization in Biological Systems* (Princeton University Press, Princeton, NJ, USA).

[2] Krause J, Ruxton GD (2002) *Living in groups* (Oxford University Press).

[3] Sumpter DJ (2010) *Collective animal behavior* (Princeton University Press).

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

[5] Ramaswamy S (2010) The mechanics and statistics of active matter. *Annu. Rev. Condens. Matter Phys.* 1:323.

[6] Vicsek T, Zafeiris A (2012) Collective motion. *Physics Reports* 517:71–140.

[7] Marchetti M, et al. (2013) Hydrodynamics of soft active matter. *Reviews of Modern Physics* 85:1143.

[8] Toner J, Tu Y (1995) Long-range order in a Two-Dimensional Dynamical XY Model: How Birds Fly Together. *Phys Rev Lett* 75:4326.

[9] Lukeman R, Li YX, Edelstein-Keshet L (2010) Inferring individual rules from collective behavior. *Proc Natl Acad Sci USA*.

[10] Mann R, et al. (2012) Multi-scale inference of interaction rules in animal groups using bayesian.

[11] Katz Y, Tunstrom K, Ioannou CC, Huepe C, Couzin ID (2011) Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl. Acad. Sci.* 108:18720–18725.

[12] Herbert-Read JE, et al. (2011) Inferring the rules of interaction of shoaling fish. *Proc. Natl. Acad. Sci.* 108:18726–18731.

[13] Gautrais J, et al. (2012) Deciphering Interactions in Moving Animal Groups. *PLoS Comput. Biol.* 8:e1002678.

[14] Strandburg-Peshkin A, et al. (2013) Visual sensory networks and effective information transfer in animal groups. *Current Biology* 23:R709–R711.

[15] Rosenthal SB, Twomey CR, Hartnett AT, Wu HS, Couzin ID (2015) Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proc. Natl. Acad. Sci.* 112:201420068.

[16] Parisi G (1988) *Statistical field theory*, Frontiers in Physics (Addison-Wesley, Redwood City, CA).

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

[18] Bialek W, et al. (2014) Social interactions dominate speed control in poising natural flocks near criticality. *Proc. Natl. Acad. Sci. U. S. A.* 111:7212–7.

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

[20] Ballerini M, et al. (2008) Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim. Behav.* 76:201–215.

[21] Cavagna A, et al. (2008) The STARFLAG handbook on collective animal behaviour: 1. Empirical methods. *Anim. Behav.* 76:217–236.

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

[23] Attanasi A, et al. (2015) GReTA a novel Global and Recursive Tracking Algorithm in three dimensions. *IEEE Trans. Pattern Anal. Mach. Intell.* X:1–14.

[24] Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O (1995) Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys Rev Lett* 75:1226.

[25] Cavagna A, et al. (2015) Short-range interactions versus long-range correlations in bird flocks. *Phys Rev E* 012705:1–15.

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

[27] Attanasi A, et al. (2014) Information transfer and behavioural inertia in starling flocks. *Nat. Phys.* 1:1–6.

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

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

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, $(\vec{s}_1(t), \ldots, \vec{s}_N(t))$, 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 [28, 29]:

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

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 [19], 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 $J_{ij} = Jn_{ij}$, 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 ($\sum_j \Lambda_{ij} = 0$).

There is a link with the statistical description of flock configurations inferred in [17]. If Λ_{ij} is symmetric and constant in time, the steady-state probability distribution of the set of $(\vec{s}_1, \ldots, \vec{s}_N)$ is given by the Boltzmann distribution

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

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:

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

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 $(1, \ldots, 1)$. 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:

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

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:

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

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

Replacing into Eq. B3 gives:

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

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

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

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:

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

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:

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

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

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

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 S3 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?

To fix ideas we work on a regular lattice in the continuum limit; the following arguments, though, are valid in general. We consider a fixed lattice, as we want to assess the contribution to $\tau_{\rm relax}$ from the relaxation of the orientations only. 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:

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

The local time correlation function $C_{\rm relax}(t) = \langle \vec{\pi}(t_0) \cdot \vec{\pi}(t_0 + t) \rangle$ is thus given by

$$C_{\rm relax}(t) = 2(d-1)T \int_{1/L}^{1/a} d^d k \int d\omega \ \frac{e^{-i\omega t}}{(i\omega + J a^2 n_c k^2)(i\omega - J a^2 n_c k^2)} = 2(d-1)T \int_{1/L}^{1/a} d^d k \ \frac{e^{-J a^2 n_c k^2 t}}{J a^2 n_c k^2} \ . \tag{C5}$$

where L is the size of the system. The absence of a mass term (zero mode) in a theory with spontaneously broken continuous symmetry is due to Goldstone's theorem and it affects the integration for small k (long wavelengths). In two dimensions the effect of the zero mode is so strong that the integral in (C5) diverges in the $L \to \infty$ limit, meaning that there cannot be long range order in $d = 2$ (Mermin-Wagner theorem). In $d = 3$, on the other hand, the integral is finite for any time t in the $L \to \infty$ limit; however, the resulting correlation function is a power law, so that the relaxation time is infinite.

The largest contribution to the correlation in the integral (C5) comes from the small k modes, those near the lower extreme of integration, $1/L$: for large systems the local spin relaxes very slowly because even for very long times it is crossed by long wavelength fluctuations, $\lambda \sim L$. The crucial point is that these long wavelength fluctuations *do not* contribute to the disordering of the local interaction network: since the wavelength is much larger than the interaction range, $r_c = a n_c^{1/d}$, all spins in the local neighborhood fluctuate in sync, with no change in the mutual positions. On the other hand, short wavelength fluctuations, $\lambda < r_c$, do contribute to the disordering of the local interaction network, so that their relaxation time is the one that must be compared to the network reshuffling time. Therefore, the relevant part of the correlation receives contribution only by the modes $k > 1/r_c$; we therefore define the effective correlation function,

$$C^*_{\rm relax}(t; r_c) \equiv 2(d-1)T \int_{1/r_c}^{1/a} d^d k \frac{e^{-J a^2 n_c k^2 t}}{J a^2 n_c k^2} \ . \tag{C6}$$

Due to the elimination of the $k \sim 1/L$ modes 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 $\tau_{\rm relax} = 1/J n_c$, which is what we study in the main text.

The argument we just provided 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 [26]). In a similar way, when we integrate

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.

in (C5) 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.

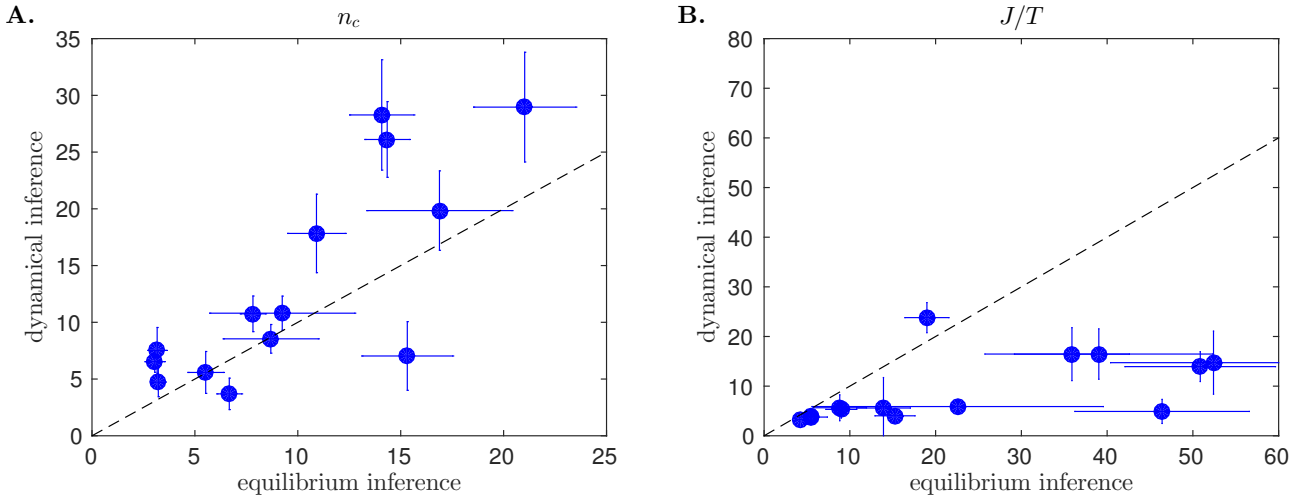

FIG. S1: Comparison between the equilibrium inference method (abcissa) and the dynamical inference method using Euler's rule (ordinate), for (**A**) the interaction range n_c and (**B**) the interaction parameter J/T. The agreement is relatively poor, especially for the prediction of J/T.

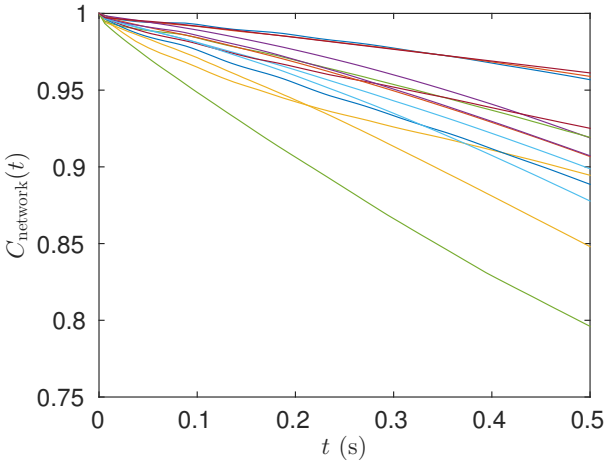

FIG. S2: Normalized autocorrelation function of the network for all 14 flocking events. The decay is approximately exponential, allowing for the definition of a characteristic decay time τ_{relax} for each event.

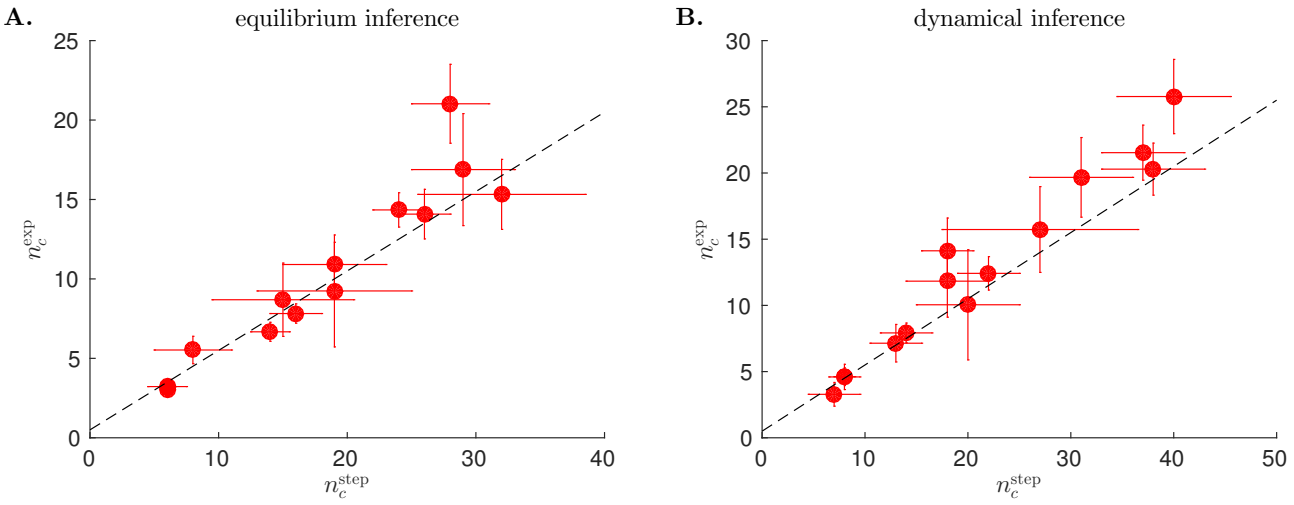

FIG. S3: Comparison of the interaction range n_c inferred assuming a step-function interaction function (n_c^{step}, abscissa) or an exponentially decaying interaction function (n_c^{exp}, ordinate), using (**A**) the equilibrium inference method and (**B**) the dynamical inference method. We expect a correspondance between n_c^{step} and n_c^{exp}: $n_c^{\mathrm{exp}} = n_c^{\mathrm{step}}/2$. Here this correspondance is verified for both inference methods.