Electronic Supplementary Material to "Diffusion in starling flocks"

A. Cavagna[#,b], S. M. Duarte Queirós[#], I. Giardina[#,b], F. Stefanini[†], and M. Viale[#,b]

[#] *Istituto dei Sistemi Complessi, UOS Sapienza, CNR, via dei Taurini 19, 00185 Roma, Italy*
[b] *Dipartimento di Fisica, Universita Sapienza, P.le Aldo Moro 2, 00185 Roma, Italy and*
[†] *Institute of Neuroinformatics, University of Zurich and ETH Zurich,*
Winterthurerstrasse, 190 CH-8057, Zurich, Switzerland

I. MUTUAL DIFFUSION

As stated in the main text, we computed the mutual diffusion as,

$$\delta r_m^2(t) \equiv \frac{1}{T-t} \frac{1}{N} \sum_{t_0=0}^{T-t-1} \sum_{i=1}^{N} [\, \bar{s}_{ij}(t_0+t) \, - \, \bar{s}_{ij}(t_0) \,]^2 . \tag{S1}$$

where $\bar{s}_{ij}(t) \equiv \bar{r}_i(t) - \bar{r}_j(t)$ is the position of bird j (the nearest neighbour of i at time t_0) in the reference frame of i.

Applying this formula to our trajectories, we obtained a typical behaviour depicted in Fig. S1, with averages values given in (6) of the main text. The values of α_m and D_m for specific flocks are given in Table S1.
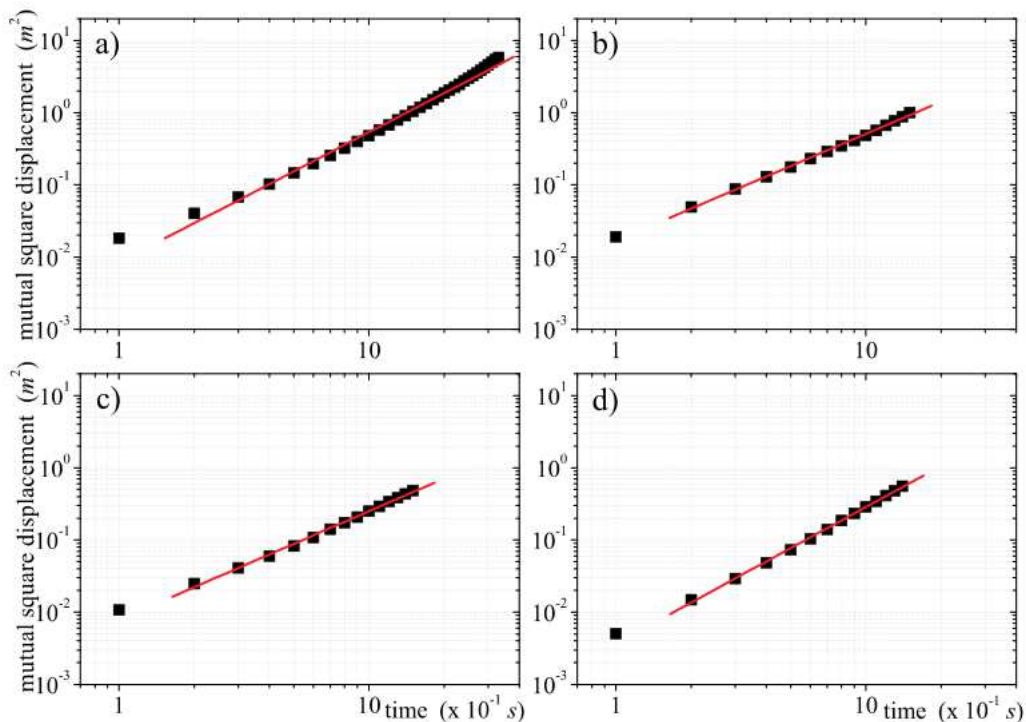

FIG. S1: Mutual diffusion of the nearest neighbour as defined in Eq. (4) of the main text for 4 different flocking events: a) 69-10; b) 48-17; c) 49-05; d) 28-10.

As commented in the main text, mutual diffusion is clearly suppressed with respect to diffusion in the centre of mass reference frame, due to the presence of correlations between birds. On the time scales accessible to our analysis, this is particularly evident in the values of the diffusion coefficients. Concerning the exponents, one has to be careful to draw conclusions. A linear regression between α and α_m shows that the two exponents are correlated (correlation coefficient 0.95). Then, we used a paired Wilcoxon test to assess whether the two exponents are significantly different. More precisely, given the set of α values obtained for each individual flock and the set of α_m values obtained for each flock, the test quantifies whether the hypothesis that the medians of the two sets are equal can or cannot be rejected. In our case, we find that the hypothesis cannot be rejected for a significance level of 0.05, but is rejected for

a significance level of 0.1. In other terms, if we use a stronger significance criterium then - given our data - we cannot conclude that the exponents are different. If we opt for a slightly less stringent criterium then we would consider them as different. However, we stress that it was not our intention to make strong claims about the diversity of the exponents. Longer time lags and a better statistics would be necessary to this task. Our main point is that mutual diffusion is anyway smaller than diffusion in the centre of mass reference frame for the time lags at our disposal.

Concerning this point, since the exponent of the mean square displacement is related to asymptotic behaviour, one might wonder whether the similarity between the exponents α and α_m increases when considering only the last part of the diffusion curves. For example, for one of the longest event that we have (69-10, panel a) in Fig S1) we can notice an upper bending in the last part of the curve and it is reasonable to ask whether taking this bending into account would change some of the conclusions. Therefore, we tried fitting the diffusion curves on the last interval $[1 - 3.4]$ (to be compared with the interval $[0.4 - 1.5]$ used in Fig. S1). We get in this case $\alpha = 1.66$ for diffusion and $\alpha_m = 2.07$ for mutual diffusion. While these values are both larger than the ones obtained using the interval $[0.4 - 1.5]$ (see Table S1), they are nonetheless less similar (rather than more similar) to each other. We note in this respect that considering only the last part of the diffusion curves for the fit is rather risky, since for those points the statistics is smaller: an example of how this might fictitiously affect the retrieved exponent is given in Fig. S4. For most of other flocks the interval used in the paper $[0.4 - 1.5]$ is already rather short to be further reduced.

Event	N	T (s)	N_{LL}	α	α_m	D $(\times 10^{-2})$	D_m $(\times 10^{-2})$
28-10	1246	1.5	785	1.83 ± 0.01	1.88 ± 0.02	3.8 ± 0.1	0.37 ± 0.04
48-17	871	1.6	350	1.73 ± 0.03	1.48 ± 0.02	3.5 ± 0.3	1.7 ± 0.03
49-05	797	1.6	146	1.71 ± 0.02	1.50 ± 0.02	3.9 ± 0.3	0.77 ± 0.06
58-06	442	3.1	140	1.69 ± 0.01	1.55 ± 0.02	3.6 ± 0.2	1.1 ± 0.04
69-09	239	4.6	62	1.64 ± 0.02	1.32 ± 0.01	4.1 ± 0.3	1.7 ± 0.04
69-10	1129	3.4	500	1.77 ± 0.02	1.72 ± 0.02	3.8 ± 0.2	0.89 ± 0.05

TABLE S1: Table of the analyzed flocks. The number of birds N is the number of individuals for which we obtained a 3D reconstruction of positions in space (average over all frames). The duration T of the event is measured in seconds = number of frames $\times 10^{-1}$s. N_{LL} indicates the number of retrieved trajectories that are as long as the entire time interval T. The last 4 columns give the values of diffusion and mutual diffusion parameters.

II. NEIGHBOURS OVERLAP FROM DIFFUSION

The number M of neighbors within a radius R around some focal bird at some initial time t_0 is,

$$M \sim \rho R^d , \tag{S2}$$

where ρ is the density of the system and $d = 3$ is the dimension of space. As time evolves from t_0 to $t_0 + t$, the M nearest neighbors move due to diffusion, occupying an expanded sphere of radius $R(t) > R$ around the focal bird. This means that the effective density $\rho(t)$ of these M initial birds decreases,

$$\rho(t) = \frac{M}{R(t)^d} < \rho . \tag{S3}$$

At time $t_0 + t$, the number $M(t)$, out of the M initial birds, that *still* remain within a radius R (which is the distance that defines the M nearest neighbors), is given by,

$$M(t) = \rho(t)R^d \;\; = \frac{M}{R(t)^d}R^d . \tag{S4}$$

We can therefore work out the neighbours overlap,

$$Q_M(t) \equiv \frac{M(t)}{M} = \frac{R^d}{R(t)^d} . \tag{S5}$$

We now make a very crude deterministic approximation and assume that the value of the radius $R(t)$ depends on the relative diffusion of the birds, in the following way,

$$R(t) \sim R + \sqrt{D_{ij}}\, t^{\alpha_m/2}, \tag{S6}$$

where α_m is the mutual diffusion exponent and D_m is the mutual diffusion constant. Substituting Eq. (S6) into Eq. (S5), we get,

$$Q_M(t) = \left(1 + \sqrt{D_m}\,\frac{t^{\alpha_m/2}}{R}\right)^{-d}. \tag{S7}$$

In real finite flocks, the dimension $d = 3$ must be reduced to an effective value $\hat{d} < 3$ because of border effects, so that the formula, $M = 4/3\pi R^3 \rho$ gets modified to a more general expression, $M = a\,R^{\hat{d}}$, with $\hat{d} = 2.3$ and $a = 0.5$ (see Fig.S2). From this we finally obtain,

$$Q_M(t) = \left(1 + c\,\frac{t^{\alpha_m/2}}{M^{1/\hat{d}}}\right)^{-\hat{d}}, \tag{S8}$$

where $c = \sqrt{D_m}\,a^{1/\hat{d}}$. For flock 69-10, this expression gives $c = 0.07$, while a fit of the data in Fig.4 of the main text gives $c = 0.05$. Considering the crude approximation that we are using, eq.(S6), these two numbers are reasonably close to each other.
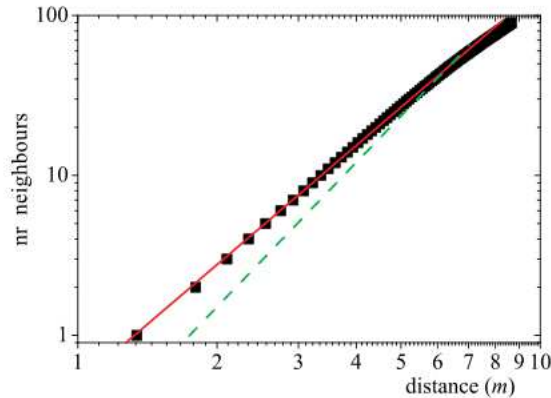

FIG. S2: Number of neighbors M as a function of the radius R of the sphere containing them. We fit this data to the formula, $M = aR^{\hat{d}}$. The full red line represents the best fit with fixed parameter $\hat{d} = 3$. The dashed green line represents the best fit where the exponent is let free. The result is $\hat{d} = 2.3 \pm 0.08$ $[R = 0.9998, \chi^2 = 0.0057, p < 0.0001]$. Both fits are performed up to $M = 50$.

III. DETERMINATION OF THE PRINCIPAL AXIS OF DIFFUSION

To determine the main axis of diffusion we consider a matrix generalization of Eq. (3),

$$\Delta_{\mu\nu}(t) \equiv \frac{1}{T-t}\frac{1}{N}\sum_{t_0=0}^{T-t-1}\sum_{i=1}^{N}\left[r_{i,\mu}(t_0+t) - r_{i,\mu}(t_0)\right]\left[r_{i,\nu}(t_0+t) - r_{i,\nu}(t_0)\right], \tag{S9}$$

where $r_{i,\mu}(t)$ represents the $\mu \in \{x,y,z\}$ Cartesian component of the position of bird i with respect to the center of mass at time t (and, in the same way, $\nu \in \{x,y,z\}$ indicates another component). The standard mean square displacement δr^2 defined in the main text is simply the trace of this matrix, i.e. $\delta r^2 = \sum_{\mu}\Delta_{\mu\mu}$.

We now want to compute the eigenvalues and eigenvectors of this diffusion matrix, for any given time lag t. The eigenvalues $\{\lambda_1(t), \lambda_2(t), \lambda_3(t)\}$ of   are given - at any time t - by the solutions of the equation

$$\det\left[\quad(t) - \lambda(t)\,\mathrm{I}\right] = 0, \tag{S10}$$

where I is the identity matrix and $\det|\mathrm{C}|$ represents the determinant of the matrix C. Each of the λ_a values (with $a = 1\cdots 3$) can be associated with a vector w_a such that,

$$\left[\quad\mathrm{w}_a - \lambda_a\,\mathrm{I}\,\mathrm{w}_a\right] = 0, \tag{S11}$$

with every $\mathbf{w}_a$ orthogonal to the others. In consequence, if $\lambda_1 > \lambda_2 > \lambda_3$, we are able to establish three independent (orthogonal) directions defined by the unitary eigenvectors, $\mathbf{u}_a \equiv \frac{\mathbf{w}_a}{|\mathbf{w}_a|}$, where $\mathbf{u}_1$ describes the direction of maximum diffusion, $\mathbf{u}_2$ the direction of second maximum diffusion and $\mathbf{u}_3$ the direction of minimum diffusion.

Since the trace (sum of the diagonal entries) of any matrix is invariant we therefore have that the standard mean square displacement (the trace of the diffusion matrix) is given by the sum of eigenvalues, i.e. $\delta r^2(t) = \sum_\mu \Delta_{\mu\mu}(t) = \lambda_1(t) + \lambda_2(t) + \lambda_3(t)$. The standard mean-square displacement can therefore be decomposed as the sum of the displacements arising along the three principal axis, each one given by the corresponding eigenvalue as a function of time. In Fig. S3 we report the behaviour of the mean-square displacement along the three principal axis as a function of time, for 4 flocking events. One can clearly see that the corresponding exponents (slopes of the curves) are different along the three axis, meaning that diffusion is anisotropic. The values of the diffusion exponents and diffusion coefficients along the three axis, averaged over all flocking events, are reported in the main text and confirm this conclusion.
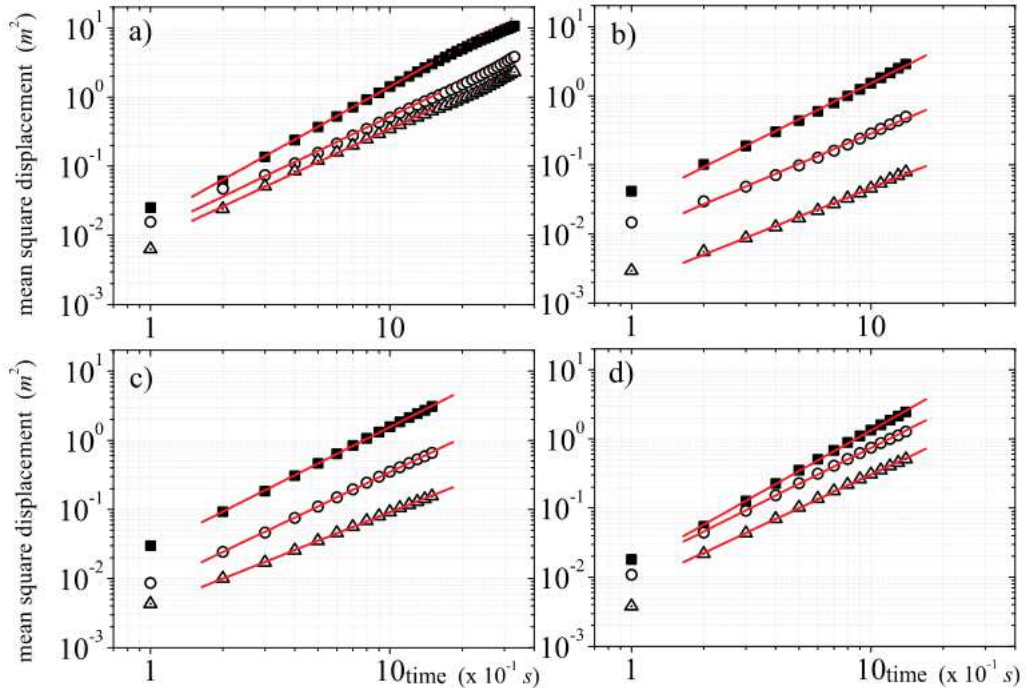

FIG. S3: Mean square displacement along the 3 principal axes of diffusion. Same flocks labels as in the previous figures.

IV. STATISTICAL SIGNIFICANCE

As briefly discussed in the main text, given a flocking event, we considered the subset of retrieved long-lasting trajectories and for each frame we calculated the center of mass coordinates R_{CM}. Then, we computed the mean squared displacement and mutual square displacement, following Eqs (3) and (5) of the main text. We note that - for what concerns diffusion - we could have used a larger sample of trajectories: when computing the mean-square displacement on a time lag equal to t we can indeed consider all trajectories that are long at least as t (and not only the long-lasting ones). Results do not change much and one would get very similar exponents.

To estimate the diffusion exponents, we fitted the resulting time dependence in log-log scale. The results in Table 1 correspond to averages of the numerical adjustments between time lags of 0.4 and 1.5 seconds, which takes into account the length of all the data at our disposal. The exponents that we find are much larger than the value $\alpha = 1$, indicating that flocks do exhibit super rather than standard diffusion. To check that this finding is not an artifact due to the finiteness of the time series (which causes a small number of samples as the lag approaches the series length), and assess the statistical significance of our results, we produced synthetic data obeying standard diffusion on the same time lags as our data. In this way, we verified that the exponents that we find for real flocks are consistently greater than the exponents corresponding to the percentile 95 obtained for normal diffusion series. Indeed, for the series we have studied, the critical values associated with this percentile have its maximal value equal to 1.64.

Let us now explain in detail the procedure that we followed. In order to analyze the statistical hypothesis of Brownian motion as well as the determination of the critical diffusion exponent, α^*, the percentile 95 of which corresponds to the value obtained by numerical adjustment of a time series of length T, we have carried out the generation of long series from which a patch of length T taken. The diffusion, $\delta r^2(t)$, is analytically obtained from,

$$\delta r^2(t) = \int \int \mathcal{C}_v(\tau', \tau)\, d\tau\, d\tau',$$

where the covariance of the velocities $\mathcal{C}_v$ is defined as in Eq. (12) in the main text, and with $\tau' = \tau + t$. If correlations decay in time as power-law $\mathcal{C}_v(\tau + t, \tau) \sim t^{-\xi}$, then the diffusion is a power function with respect to t, $\delta r^2(t) \sim t^{2-\xi}$, and thus $\alpha = 2 - \xi$. To generate power-law correlated velocities in a savvy way we have resorted to the Wiener-Khinchin theorem relating the correlation function and the spectral density and proceeded as follows;

- Generate a series of Gaussian time series, $\{u\}$, of length N (with N being a odd number, $e.g.$, $10^6 + 1$);[1]

- Compute its Fourier transform, where the element u_i corresponds to a value $\bar{u}\left(f = \frac{i-1}{N} - \frac{1}{2}\right)$;

- Set apart the absolute value and multiply $\exp\left[i \arg(\tilde{u}(f))\right]$ by the square root of the Fourier transform of $\mathcal{C}_v(\tau + t, \tau)$, which in this case is a power-law function as well, $S(f) \sim |f|^{-\xi-1}$;
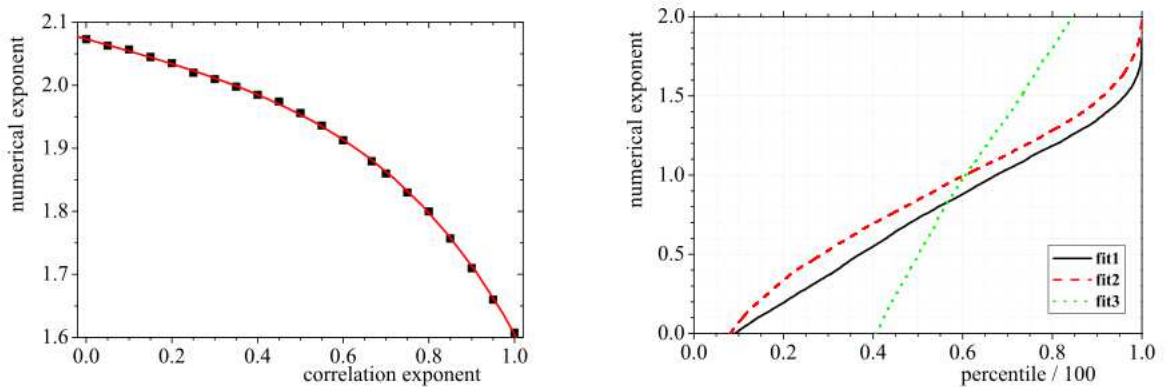

FIG. S4: Left panel: exponent obtained by the numerical adjustment of the diffusion, which has been computed in patches of Brownian motion with the same length of the flock 69–10 and averaged over the 500 samples (one sample cycle), as a function of the percentile (divided by 100). The line **fit1** takes into account lags from 1 until 33, **fit2** corresponds to the interval used in the manuscript and **fit3** is obtained considering the interval of lags between 16 and 33. Right panel: The percentile 95 as a function of the correlation exponent for the dataset 69–10 and considering the lag interval **fit2**. The points have been obtained from 10^4 sampling cycles and the line has been obtained by interpolation using the points.

- Invert the Fourier transform and multiply the outcome by $(-1)^{i+1}$ to finally obtain the correlated series $v(i)$.

Afterwards, the series is summed to create position sequences. We have considered values of ξ between 0 and 1, which after integration bear ballistic and Brownian trajectories, respectively.

As an illustration, in Fig. S4, we present the exponent obtained by the numerical adjustment of the diffusion, which has been computed in patches of Brownian motion with the same length of the flock 69–10 and averaged over the same number of birds, as a function of the percentile and the exponent corresponding to the percentile 95 as a function of the correlation exponent, ξ.

We have also considered the hypothesis that velocities follow a Langevin stochastic equation with a typical scale equal to k^{-1}, the correlation function of which decays in the form of an exponential. This case also leads to a functional dependence diffusion given by $kt + \exp[-kt] - 1$, which that does fit for our field values.

---

[1] The odd number is a simple trick to avoid the frequency 0 which is associated with a singularity of the power spectrum.

V. BORDER DEFINITION AND BORDER DIFFUSION

Given a flock, in order to compute several global and statistical quantities it is necessary to appropriately define its exterior border (see [1] for a thorough discussion of this problem). Flocks are typically non-convex systems. Therefore, standard methods to define the border, like the convex-hull, are inadequate because they are unable to detect concavities. To overcome this problem, we used the so-called 'α-shape algorithm [2]. The main idea of this method is the following: given a set of 3D points, one 'excavates' the set of points with spheres of radius α, so that all concavities of size larger than α are detected. Formally, one selects the sub-complex of the Delaunay triangulation on scale α (the α-complex) and the external surface of this triangulation defines the border.

The scale α must be appropriately chosen. If α is too large, some concavities are neglected and void regions are included as being part of the flock. Too small values of α, on the other hand, might cause the sphere to penetrate the flock and break it into sub-connected components. A robust criterion is to look at the density of the internal points as a function of α [1][3]. This quantity typically has a maximum, which defines a natural scale for α.

For all the analyzed flocking events, the border has been computed following the above procedure. We note that, since flocks change shape in time, the border must be computed and re-defined at each instant of time. Besides, due to the continuous movement of individuals through the group, the individuals belonging to the border change from time to time.

Once we have defined the border, we can ask whether the diffusion properties that we have described and quantified in the main text are similar for birds belonging to the border and birds well inside the bulk of the group. Let us consider for example the behaviour of the mean-square displacement. Clearly, since individuals on the border do not remain there forever, but at some point leave the boundary, on large time scales we cannot even distinguish border and internal birds. However, on shorter time lags we can try to investigate such a difference.

To do this, for any time-lag t we divided the statistical sample (segments of trajectories that are long t steps) into trajectories where the bird belonged to the border for the whole period t, and internal birds (the complementary set). We then computed the mean-square displacement as a function of t. For statistical reasons we did this on the largest flock that we have (event 28-10). The results are shown in Fig. S5 for diffusion in the centre of mass, and in Fig. S6 for mutual diffusion.
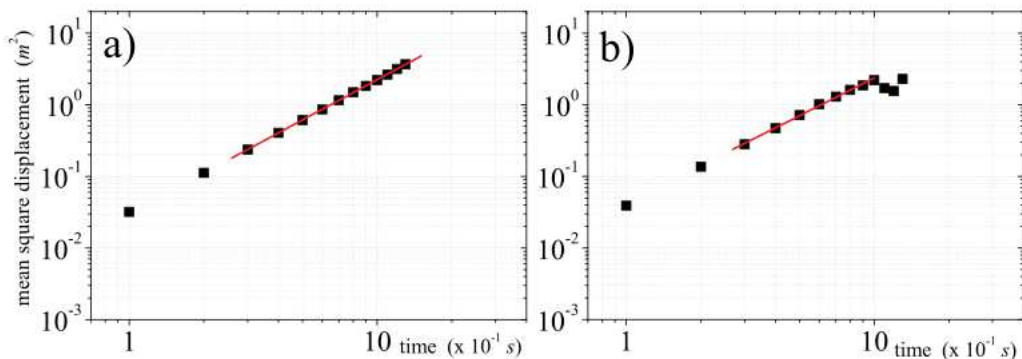

FIG. S5: Mean square displacement in the centre of mass reference frame for a) internal birds and b) birds on the border of the flock. In panel b) only individuals that remain on the border for the entire time lag t are considered when computing $\delta r^2(t)$ (see text). The exponents are $\alpha = 1.86 \pm 0.02$ for internal birds, and $\alpha = 1.78 \pm 0.03$ for border birds, to be compared with the value $\alpha = 1.83 \pm 0.01$ obtained with all the birds. For the diffusion coefficients we get (values divided for 10^{-2} as in Table S1) $D = 3.1 \pm 0.03$ (internal) and $D = 4.2 \pm 0.05$ (border), to be compared with $D = 3.8 \pm 0.1$ (all birds).

As one can see from these figure, the exponents of internal diffusion are very similar to the ones computed with all the trajectories. Both the exponents and the diffusion coefficients of the boundary birds are slightly smaller (even if not too much). We note that - due to the procedure we used - the results for mutual diffusion on the border refer to individuals who *both* are on the border for the time lag t. Thus, they describe how a bird on the border moves with respect to its nearest-neighbour *on the border*. However, they do not provide any information on how birds on the border move with respect to their first *internal* nearest-neighbour. We tried also to investigate internal-border diffusion, but the statistics is really poor to draw any conclusions. For this reason we preferred to look at the survival probability of the permanence on the border (see main text), which is statistically more robust.

As already underlined, the above measurements of border diffusion suffer from a very much reduced statistical sample, especially at the larger times (only individuals that remained on the border up to t are included). For smaller flocks this make the above analysis not feasible.
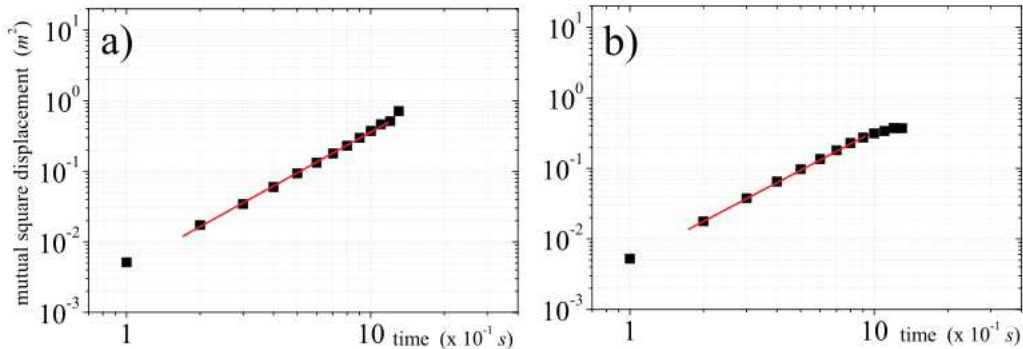
FIG. S6: Mean square displacement relative to mutual diffusion for a) internal birds and b) birds on the border of the flock. In panel b) only pairs of individuals (a bird and its nearest-neighbour) that remain on the border for the entire time lag t are considered when computing $\delta r_m^2(t)$ (see text). The exponents are $\alpha_m = 1.92 \pm 0.02$ for internal birds, and $\alpha_m = 1.84 \pm 0.04$ for border birds, to be compared with the value $\alpha_m = 1.88 \pm 0.02$ obtained with all the birds. Diffusion coefficients are (values divided for 10^{-2} as in Table S1) $D = 0.42 \pm 0.06$ (internal) and $D = 0.35 \pm 0.03$ (border), to be compared with $D = 0.37 \pm 0.04$ (all birds).

VI. COMPUTING THE SURVIVAL PROBABILITY FOR INTERNAL INDIVIDUALS

As discussed in the main text, we can define the survival probability of internal individuals as the probability that in the centre of mass reference frame a bird has moved less than a distance l_B in a time t, where l_B is the typical distance of the nearest-neighbour. This probability gives an estimate of how much time is needed for a bird to exchange location with respect to the first shell of neighbours, and is a relevant benchmark to be compared with the border survival probability.

We know that birds at the interior of the flock obey super-diffusive behaviour. Thus, we would like to compute the survival probability starting from the diffusion properties of the birds. In the case of a generic diffusion exponent α there are not analytic exact computations. However, there are some rigorous bounds for the short and large time limits of the survival probability [4, 5]. More precisely, the theory predicts that

$$ P(t) \sim \exp\left[-\frac{k(t)}{(l_B^2/D)^{1/\alpha}} \right] , \tag{S12} $$

where $k(t)$ is an unknown function, for which however we know that $k(t) = k_1 t$ in the short time regime, and $k(t) = k_2 t$ in the long time regime. For Brownian diffusion ($\alpha = 1$), one has $k_1 = k_2$ and this function becomes a simple exponential. For super-diffusion ($\alpha > 1$) one has $k_2 > k_1$, i.e. there is an initial exponential decay at short times and a faster decay in the large time regime. The survival probability for internal individuals that we computed on the data is qualitatively consistent with this $\alpha > 1$ prediction. Besides, if we fit the initial part of the curve (first 8 points) with an exponential function of the form (S12) with $k(t) = k_1 t$, and we use the experimental values for α and D in each flock, we find that the constant k_1 is fairly consistent in all flocks, and equal to 1.42 ± 0.15. In other terms, the short time regime of the survival probability for internal birds is well described by Eq. (S12) with a flock independent constant k_1.

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

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

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

[4] Monrad D. and Rootzén H. (1995) Small values of Gaussian process and functional laws of the iterated logarithm. *Probab Theory Relat Fields* 101:173–192.

[5] Metzler R. and Klafter J. (2000) The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys Rep* 339:1–77.