SUPPLEMENTARY NOTE 1: GENERALIZED KURAMOTO MODEL ON ELLIPSOIDS

In the main text, we remark that a seemingly more general Kuramoto model on ellipsoids [1]

\[
\dot{y}_i = \Omega_i A y_i + (I_{n+1} - y_i y_i^T) A \sum_{j=1}^{N} k_{ij} y_j \tag{1}
\]

can be reduced to the Kuramoto model on the \(n\)-sphere

\[
\dot{x}_i = \Omega_i x_i + (I_{n+1} - x_i x_i^T) \sum_{j=1}^{N} k_{ij} x_j. \tag{2}
\]

To see this, let an ellipsoid be given by

\[
\mathcal{E}^n = \{ y \in \mathbb{R}^{n+1} | y^T A y = 1 \}
\]

where \(A \in \mathbb{R}^{n+1 \times n+1}\) is a positive definite matrix. Then \(A\) has a Cholesky decomposition \(A = L L^T\). Introduce a change of variables \(x_i = L^T y_i\) in Eq. (1) whereby

\[
\dot{x}_i = L^T (\Omega_i L L^T y_i + (I_{n+1} - y_i y_i^T) L L^T) \sum_{j=1}^{N} k_{ij} y_j,
\]

\[
= L^T \Omega_i L x_i + (I_{n+1} - x_i x_i^T) \sum_{j=1}^{N} k_{ij} x_j,
\]

where the frequency term \(L^T \Omega_i L\) replaces \(\Omega_i\) in (2). Moreover, \(|x_i|^2 = y_i^T L L^T y_i = 1\).

SUPPLEMENTARY NOTE 2: ALMOST GLOBAL PRACTICAL SYNCHRONIZATION

In this section, we formally state and prove our result concerning the small-spread frequency condition that ensures the instability of dispersed equilibria, i.e., equilibria that are not contained in an open hemisphere. As mentioned in the main text, the gradient descent formulation of the homogeneous Kuramoto model on \(S^n\) allows us to consider the natural frequency terms \(\Omega_i\) in Eq. (2) as perturbations of a homogeneous model obtained when all frequencies are equal (which is equivalent to setting all frequencies to zero).

We use a result from matrix perturbation theory to obtain our instability condition. Given a Hermitian matrix \(X\), i.e., \(X = X^*\) (where \(^*\) denotes conjugate transpose), and an arbitrary matrix \(Y\), the following result relates the spectrum of \(X + Y\) to those of \(X\) and \(Y\):

**Theorem 1** (Kahan [2]). Let \(X \in \mathbb{C}^{n \times n}\) be a Hermitian matrix with eigenvalues \(\chi_1 \leq \ldots \leq \chi_n\) and \(Y \in \mathbb{C}^{n \times n}\) be an arbitrary matrix. Let \(\sigma_j\) with \(Re \sigma_1 \leq \ldots \leq Re \sigma_n\) denote the eigenvalues of \(X + Y\). Then

\[
\left( \sum_{j=1}^{n} (\chi_j - Re \sigma_j)^2 \right)^{\frac{1}{2}} \leq \frac{1}{2} \|Y + Y^*\|_2 + \left( \frac{1}{2} \|Y - Y^*\|_2 - \sum_{j=1}^{n} (Im \sigma_j)^2 \right)^{\frac{1}{2}}.
\]

Note that for a skew-symmetric matrix \(Y \in so(n)\), the above inequality can be simplified as

\[|\chi_n - Re \sigma_n| \leq \|Y\|_2,\]

which is the result that we will use later.

**Small frequency terms**—Let us first focus on the special case of frequency terms that are small in an absolute sense. Later we will show that frequencies which are small in a relative sense, i.e., that have small spread around their mean value, can be reduced to this case.

**Theorem 2.** Every dispersed equilibrium of the \(n\)-dimensional network Kuramoto model on \(S^n\), i.e., Eq. (2), is exponentially unstable if the spread of frequencies \(\Omega_i\) around zero is small in the following sense

\[
\left( \sum_{i=1}^{N} ||\Omega_i||_2^2 \right)^{\frac{1}{2}} < \frac{2K}{N(n+1)} (n-1 - x_i \cdot x_j) (1 - x_i \cdot x_j) \tag{3}
\]

where \(K = \min_{(i,j) \in E} k_{ij}\) and \(n \geq 2\).

**Proof.** The linearized system—The \(n\)-dimensional network Kuramoto model, Eq. (2) is the sum of a drift term and a gradient descent term. The map from a system to its linearization is itself linear, so we can linearize the two terms separately. The drift term \(\Omega_i x_i\) is linear in \(x_i\). The gradient descent flow term was derived in
It is given by the \( N(n+1) \times N(n+1) \) block matrix 
\[
\mathbf{B}(x) = [B_{ij}],
\]
where each \((n+1) \times (n+1)\) block is
\[
B_{ij} = \begin{cases} 
-\sum_{k=1}^{N} k_{ij}(x_i, x_i) (I_{n+1} - x_i x_i^T) & \text{if } j = i, \\
 k_{ij}(x_i + x_j) (I_{n+1} - x_i x_j^T) & \text{otherwise.}
\end{cases}
\] 

(4)

The linearization of Eq. (2) is hence characterized by a block matrix \( \mathbf{A}(x) = [A_{ij}] \) where the blocks are given by
\[
A_{ij} = \begin{cases} 
\Omega_i + B_{ii} & \text{if } j = i, \\
B_{ij} & \text{otherwise.}
\end{cases}
\] 

(5)

We consider the linearized system Eq. (5) as a perturbation of the matrix \( \mathbf{B}(x) \), namely,
\[
\mathbf{A}(x) = \text{diag}(\Omega_1, \ldots, \Omega_N) + \mathbf{B}(x),
\]

(6)

where we assume that the perturbation term is small in terms of its 2-norm \( \| \text{diag}(\Omega_1, \ldots, \Omega_N) \|^2 \). To analyze the stability of Eq. (6), we first find a specific lower bound on the largest eigenvalue (the \textit{spectral abscissa}) of the symmetric part \( \mathbf{B}(x) \). We then use matrix perturbation theory in the case of small \( \| \Omega_i \|_2 \) to bound the spectral abscissa of \( \mathbf{A}(x) \) below by a small positive number.

**Eigenvalues of the symmetric matrix**—The spectral abscissa of \( \mathbf{B}(x) \) given by Eq. (4) is
\[
\beta(x) := \max_{\|v\|_2 = 1} \text{Re} \langle v, \mathbf{B}(x)v \rangle.
\]

Taking
\[
v = \frac{1}{\sqrt{N}} \left[ w^T \cdots w^T \right]^T,
\]

where \( \|w\| = 1 \) gives us a lower bound,
\[
\beta(x) \geq \max_{\|w\|_2 = 1} \frac{1}{N} \left( \sum_{i \in V} \sum_{j \in V} B_{ij} w \right)
\]
\[
\geq \frac{1}{N(n+1)} \sum_{i \in V} \sum_{j \in V} \text{tr} B_{ij} = \frac{1}{N(n+1)} \sum_{i \in V} \sum_{j \in N_i} \text{tr} B_{ij}
\]
\[
= \frac{2}{N(n+1)} \sum_{i} k_{ij} \left( -n(x_i, x_i) + n - 1 + (x_i, x_i)^2 \right)
\]
\[
= \frac{1}{N(n+1)} \sum_{i} \sum_{j} k_{ij} (n - 1 - (x_i, x_i))(1 - (x_j, x_j)),
\]

where \( N_i := \{ j \in V \mid \{ i, j \} \in E \} \) is the set of neighbors of agent \( i \). The second inequality above uses that the largest eigenvalue of \( \sum_{i \in V} \sum_{j \in V} B_{ij} \) is greater than the average eigenvalue. The lower bound is strictly positive for all \( n \geq 2 \), \( (x_i)_{i=1}^{N} \notin S \), where \( S \) is the synchronization manifold defined in the main text for the Stiefel manifold, which in the special case of the \( n \)-sphere reads:
\[
S := \{ (x_i)_{i=1}^{N} \mid x_i = x_i, \forall \{ i, j \} \in E \}.
\]

**Matrix perturbation theory**—We have a lower bound for the spectral abscissa of the linearized homogeneous part of Eq. (2):
\[
\beta(x) \geq 2 \frac{1}{N(n+1)} \sum_{i,j} k_{ij} (n - 1 - x_i \cdot x_j)(1 - x_i \cdot x_j).
\]

(8)

It remains to relate this result to the heterogeneous model. This is done via the characterization in Eq. (6) of the linearization of the heterogeneous model as a perturbation of the homogeneous model, to facilitate the application of Theorem 1. The linearization matrix \( \mathbf{A}(x) \) of the heterogeneous model is separated from \( \mathbf{B}(x) \) by
\[
\mathbf{A}(x) - \mathbf{B}(x) = \text{diag}(\Omega_1, \ldots, \Omega_N).
\]

Recall that a skew-symmetric perturbation term \( \text{diag}(\Omega_1, \ldots, \Omega_N) \in \mathfrak{s}(n) \) allows for a simplification of the inequality in Theorem 1,
\[
|\lambda_n - \text{Re} \sigma_n| \leq \| \mathbf{Y} \|_2.
\]

Let \( \text{Re} \alpha(x) \) denote the spectral abscissa of \( \mathbf{A}(x) \). By the matrix perturbation result, Theorem 1, we have
\[
|\beta(x) - \text{Re} \alpha(x)| \leq \| \text{diag}(\Omega_1, \ldots, \Omega_N) \|_2.
\]

(9)

Note that \( \| \text{diag}(\Omega_1, \ldots, \Omega_N) \|_2 = \left( \sum_{j \in V} \| \Omega_j \|_2 \right)^2 \). By the assumption of Theorem 2,
\[
\left( \sum_{j \in V} \| \Omega_j \|_2 \right)^2 < \frac{2}{N(n+1)} \sum_{i,j} k_{ij} (n - 1 - x_i \cdot x_j)(1 - x_i \cdot x_j).
\]

(10)

Combine Eq. (8)–(10) to find
\[
|\beta(x) - \text{Re} \alpha(x)| < \beta(x),
\]

which implies \( \text{Re} \alpha(x) > 0 \). Since \( \alpha(x) \) is an eigenvalue of the linearization matrix \( \mathbf{A}(x) \) of Eq. (2) at any dispersed equilibrium \( x \), it follows that Eq. (2) is exponentially unstable at any dispersed equilibrium \( x \).

**SUPPLEMENTARY NOTE 3: INSTABILITY IN THE SENSE OF LYAPUNOV**

Theorem 2 establishes exponential instability. For Lyapunov instability of dispersed equilibria, recall that the indirect method of Lyapunov states a nonlinear system to be unstable if the matrix that characterizes its linearization has an eigenvalue with strictly positive real part [4]. However, this theorem is for systems on linear spaces; we must account for the fact that Eq. (2) evolves on \( S^n \).

**Corollary 3.** If the condition in Theorem 2 holds, the dispersed equilibria of Eq. (2) are not Lyapunov stable.

**Proof.** Extend the heterogeneous model from \( S^n \) to an open neighborhood of \( S^n \) in \( \mathbb{R}^{n+1} \setminus \{0\} \) as
\[
\tilde{n}_i = \Omega_i n_i + \left( I_{n+1} - \frac{n_i}{\|n_i\|_2} \left( \frac{n_i}{\|n_i\|_2} \right)^T \right) \sum_{j \in N_i} k_{ij} \frac{n_j}{\|n_j\|_2},
\]

(11)
where \( \mathbf{n}_i(0) \in \mathbb{R}^{n+1} \setminus \{0\} \). Note that \( \| \mathbf{n}_i \|_2^2 \) is constant along trajectories of the system since \( \frac{d}{dt} \| \mathbf{n}_i \|_2^2 = 2 \langle \dot{\mathbf{n}}_i, \mathbf{n}_i \rangle = 0 \). Inspect Eq. (11) to see that its restriction to \((S^n)^N\) equals the heterogeneous model, Eq. (2). It follows that the linearization of Eq. (11) restricted to \( T_x(S^n)^N \), where \( x \in (S^n)^N \), equals the linearization of the heterogeneous model, Eq. (2).

The linearization of Eq. (11) at a dispersed equilibrium \( x \in (S^n)^N \) restricted to \( T_x(S^n)^N \) is unstable. How can we know that the unstable manifold of Eq. (11) entirely belongs to \((S^n)^N\)? To see this, follow a trajectory on the unstable manifold (i.e., one that satisfies \( \lim_{t \to -\infty} \mathbf{n}(t) = \mathbf{x} \)). Then \( \| \mathbf{n}_i(t) \|_2^2 = \| \mathbf{x}_i \|_2^2 = 1 \) or else we contradict \( \langle \dot{\mathbf{n}}_i, \mathbf{n}_i \rangle = 0 \). As such, the dispersed equilibrium \( x \) of the heterogeneous model, Eq. (2), which is also a dispersed equilibrium of Eq. (11) on \((S^n)^N\), is not Lyapunov stable.

**Frequency terms with small spreads**—We have detailed the central ideas leading to Theorem 2 for small frequencies centered around zero. The next step is to generalize this to the case presented in the main text, namely frequencies with a small spread around the mean frequency. Consider the case of large frequencies, \( \| \Omega_i \| \gg 1 \). Simulations reveal that the system behaviour is incoherent in general. A case of special interest is perturbations that are large but have small variance, i.e.,

\[
\| \Omega_i - \frac{1}{N} \sum_{j=1}^{N} \Omega_j \|_2 \ll 1.
\]

For ease of notation, we write \( \Omega := \frac{1}{N} \sum_{j=1}^{N} \Omega_j \).

Introduce the variables \( \mathbf{y}_i := e^{-\Omega t} \mathbf{x}_i \) that are rotated by a time-dependent amount \( e^{-\Omega t} \in SO(3) \) to find

\[
\dot{\mathbf{y}}_i = -\Omega \mathbf{y}_i + e^{-\Omega t} \dot{\mathbf{x}}_i
\]

\[
= -\Omega \mathbf{y}_i + e^{-\Omega t}(\Omega \mathbf{x}_i + (\mathbf{I}_{n+1} - \mathbf{x}_i \mathbf{x}_i^\top) \sum_{j \in N_i} k_{ij} \mathbf{x}_j)
\]

\[
= (-\Omega + e^{-\Omega t} \Omega \mathbf{e}^{\Omega t}) \mathbf{y}_i + (\mathbf{I}_{n+1} - \mathbf{y}_i \mathbf{y}_i^\top) \sum_{j \in N_i} k_{ij} \mathbf{y}_j.
\]

Let \( \Xi_i = e^{-\Omega t} \Omega \mathbf{e}^{\Omega t} \) which satisfies

\[
\Xi_i = -\Omega \Xi_i + \Xi_i \Omega.
\]

Together, \( \mathbf{y}_i \) and \( \Xi_i \) form a system given by

\[
\dot{\mathbf{y}}_i = (\Xi_i - \Omega) \mathbf{y}_i + (\mathbf{I}_{n+1} - \mathbf{y}_i \mathbf{y}_i^\top) \sum_{j \in N_i} k_{ij} \mathbf{y}_j,
\]

\[
\dot{\Xi}_i = [\Xi_i, \Omega],
\]

where \([\cdot, \cdot]\) denotes a commutator. We are only interested in the behavior of the \( \mathbf{y}_i \) variables since they contain all information about the original \( \mathbf{x}_i \) variables. As such, we consider the system to be in an equilibrium if \( \dot{\mathbf{y}}_i = 0 \).

**Corollary 4.** Assume that spread of the frequencies \( \Omega_i \) is small,

\[
\left( \sum_{i=1}^{N} \| \Omega_i - \frac{1}{N} \sum_{j=1}^{N} \Omega_j \|_2 \right)^2 < \frac{2K}{N+1}(n - 1 - \langle x_j, x_i \rangle)(1 - \langle x_j, x_i \rangle),
\]

where \( K = \min_{(i, j) \in \mathcal{E}} k_{ij} \) and \( n \geq 2 \). Then, any dispersed equilibrium of the transformed heterogeneous model (12)–(13) on \( S^n \) is exponentially unstable.

**Proof.** Note that

\[
\| \Xi_i - \Omega \|_2 = \| e^{-\Omega t} \mathbf{e}^{\Omega t} - \Omega \|_2 = \| \Omega_i - \sum_{j=1}^{N} \Omega_j \|_2.
\]

The result is derived as that of Theorem 2.

**SUPPLEMENTARY NOTE 4: REMOVAL OF STATE-DEPENDENCE FROM THE SYNCHRONIZATION CONDITION**

The main text points out that the RHS in the synchronization condition that limits the spread of frequency matrices can be rendered state-independent. The bound becomes

\[
\left( \sum_{i=1}^{N} \| \Omega_i \|_2 \right)^2 < \sum_{(i, j) \in \mathcal{E}} \frac{2K}{N(n+1)}(n - 1 - \cos \frac{n}{N})(1 - \cos \frac{n}{N})
\]

where \( K = \min_{(i, j) \in \mathcal{E}} k_{ij} \) and \( n \geq 2 \). Note that Eq. (3) is in \( \mathcal{O}(K/N^4) \) for \( n = 2 \) and \( \mathcal{O}(K/N^2) \) for \( n \geq 3 \).

**Optimization of bound**—For future reference, we define the lower bound of \( \beta(\mathbf{x}) \) in Eq. (8) as a function

\[
f : \mathbf{x} \mapsto \frac{2}{N(n+1)} \sum_{(i, j) \in \mathcal{E}} k_{ij}(n - 1 - \langle x_j, x_i \rangle)(1 - \langle x_j, x_i \rangle),
\]

i.e., \( \beta(\mathbf{x}) \geq f(\mathbf{x}) \) holds. Let \( \theta_{ij} \) denote the angle between agent \( i \) and \( j \), i.e., \( \cos \theta_{ij} = \langle x_j, x_i \rangle \). Then we can write

\[
f : \mathbf{x} \mapsto \frac{2}{N(n+1)} \sum_{(i, j) \in \mathcal{E}} k_{ij}(n - 1 - \cos \theta_{ij})(1 - \cos \theta_{ij}).
\]

To remove the dependence of \( f \) on \( \mathbf{x} \), we minimize \( f \) over all dispersed equilibria. Recall that for an equilibrium to be regarded as dispersed, we require that the set \( \{x_1, \ldots, x_N\} \) is not contained in an open hemisphere of \( S^n \). The minimization must be done with respect to both the graph \( \mathcal{G} \) and all equilibria \( x = (x_i)_{i=1}^{N} \) that do not belong to a hemisphere. Since it is difficult to characterize equilibria, we minimize over all dispersed configurations.

Each term in \( f \) is positive, wherefore removing a link from \( \mathcal{G} \) leads to a decrease in \( f \). Since we require \( \mathcal{G} \) to be connected, it follows that the graph which minimizes \( f \) is a tree. By definition of \( \mathbf{x} \) being dispersed,
\( \{x_1, \ldots, x_N\} \) is not contained in a hemisphere. There must hence be a tuple \( P = (i_1, \ldots, i_k) \) for some \( k \leq N \) such that \( \sum_{j=1}^{k} \theta_{i_j i_{j+1}} \geq \pi \). Note that \( f \) increases with \( \theta_{ij} \). For any index \( i \notin P \) and its neighbor \( k \), it is optimal to set \( \theta_{ik} = 0 \). There is therefore no loss of optimality in assuming that \( P = (1, \ldots, N) \), i.e., \( G \) is a path.

The choice of agent placements that minimize \( f \) solves

\[
\min f = \frac{2}{N(n+1)} \sum_{i \in V} k_{ij}(n - 1 - \cos \theta_{i,i+1})(1 - \cos \theta_{i,i+1}) \sum_{i \in V} \theta_{i,i+1} = \pi,
\]

where addition of indices is modulo \( N \). The inequality in the constraint has been replaced by an equality since \( f \) increases with increasing \( \theta_{ij} \), i.e., it is sub-optimal to choose \( (x_i)_{i=1}^N \) such that \( \sum_{i \in V} \theta_{i,i+1} > \pi \). The coupling gains \( k_{ij} \) complicate the next step of the analysis, wherefore we let \( K := \min_{(i,j) \in E} k_{ij} \) and replace the minimization of \( f \) by another problem,

\[
\min g(x) := \frac{2K}{N(n+1)} \sum_{i \in V} (n - 1 - \cos \theta_{i,i+1})(1 - \cos \theta_{i,i+1}) c(x) := \sum_{i \in V} \theta_{i,i+1} - \pi = 0, \tag{14}
\]

where \( f(x) \geq g(x) \) with equality if \( k_{ij} = K \).

Solution to optimization problem for large \( N \)—To solve the nonlinear programming problem (14), we use the Lagrange conditions for optimality [5]. There is only one constraint \( c(x) = 0 \), the gradient of which is non-zero. The optimal solution hence satisfies

\[
\nabla g(x) + \nabla \lambda c(x) = \nabla g(x) + \lambda 1 = 0, \tag{15}
\]

where \( \lambda \) is a Lagrange multiplier and \( 1 = [1, \ldots, 1]^T \). It follows that

\[
\sin \theta_{i,i+1}(n - 2 \cos \theta_{i,i+1}) = -\frac{N(n+1)}{2K} \lambda, \tag{16}
\]

where \( \lambda \leq 0 \) is required for \( \theta_{i,i+1} \in [0, \pi] \) to exist.

For ease of notation, introduce \( h(\theta) := \sin \theta - 2 \cos \theta \), \( c := -N(n+1)\lambda/(2K) \). Note that \( h \) is the product of two factors. The first factor, \( \sin \theta \), increases on \([0, \pi/2]\). The second factor, \( n - 2 \cos \theta \), increases on \([0, \pi]\). As such, \( h \) increases on \([0, \pi/2]\), see Supplementary Fig. 1 for an illustration. There is then at most one intersection of \( h \) with the constant function \( c \) on \([0, \pi/2]\). Moreover, at most one \( \theta_i > \pi/2 \) due to the constraint \( \sum_{i=1}^N \theta_i = \pi \). There are hence only two solutions to Eq. (15):

(i) either \( \theta_{i,i+1} = \pi/N \) for all \( i \in V \) or

(ii) \( \theta_{j,j+1} = \varphi > \pi/2, \theta_{i,i+1} = (\pi - \varphi)/(N - 1) \), for all \( i \in V \setminus \{j\} \) (this is one solution up to permutations of the large angle \( \theta_{j,j+1} \) among \( \{1, \ldots, N\} \)).

In case (i) where \( \theta_{i,i+1} = \pi/N \), the objective function value is

\[
g_1 := \frac{2K}{n+1}(n - 1 - \cos \frac{\pi}{N})(1 - \cos \frac{\pi}{N}).
\]

In case (ii) set \( \theta_{j,j+1} = \varphi \). Then

\[
g_2 := \frac{2K(N-1)}{N(n+1)}(n - 1 - \cos \frac{\pi - \varphi}{N-1})(1 - \cos \frac{\pi - \varphi}{N-1}) + \frac{2K}{N(n+1)}(n - 1 - \cos \varphi)(1 - \cos \varphi)
\]

The value \( g_1 \) belongs to \( O(K/N^2) \) if \( n \geq 3 \) (or \( O(K/N^4) \) if \( n = 2 \)) since \( 1 - \cos \pi/N \in O(1/N^2) \), whereas \( g_2 \) belongs to \( O(K/N) \) due to the second term being bounded below by \( 2K(n-1)/(N(n+1)) \). As such, \( g_1 \) is optimal for sufficiently large \( N \). Note that \( \theta_{i,i+1} = \pi/N < 1 \) and \( \pi^2 \leq 2N \) for \( N \geq 5 \) whereby

\[
g_1 \leq \frac{2K}{n+1}(n - 1)\frac{1}{2} < \frac{2K}{n^2}(n - 1)\frac{1}{2} < g_2.
\]

Hence \( \theta_{i,i+1} = \pi/N \) is optimal for \( N \geq 5 \).

Three special cases—Consider the remaining cases of \( N \in \{2, 3, 4\} \). If \( N = 2 \), then

\[
g_1 = \frac{2K}{n+1}(n - 1),
\]

\[
g_2 = \frac{K}{n+1}(n - 1 + \cos \varphi)(1 + \cos \varphi) + \frac{2K}{n+1}(n - 1 - \cos \varphi)(1 - \cos \varphi)
\]

\[
= g_1 + \frac{2K}{n+1}\cos^2 \varphi.
\]

If \( N = 3 \), then

\[
g_1 = \frac{K}{n+1}(n - \frac{3}{2}),
\]

\[
g_2 = \frac{4K}{3(n+1)}(n - 1 - \sin \frac{\varphi}{2})(1 - \sin \frac{\varphi}{2}) + \frac{2K}{3(n+1)}(n - 1 - \cos \varphi)(1 - \cos \varphi)
\]

\[
= \frac{K}{n+1}[n(2 - \frac{3}{2}(2\sin \frac{\varphi}{2} + \cos \varphi))
\]

\[
- 2(1 - \frac{1}{2}(1 - \cos \varphi + \cos^2 \varphi))]
\]

\[
\geq \frac{K}{n+1}(n - \frac{3}{2}) = g_1,
\]

where we omitted a few steps. The minimum of \( g_2 \) occurs near \( \varphi = 1 \), which contradicts the assumption \( \varphi \geq \pi/2 \). Hence the solution \( \theta_{i,i+1} = \pi/N \) is optimal. Finally, if \( N = 4 \), then

\[
g_1 = \frac{2K}{n+1}(n - 1 - \frac{1}{\sqrt{2}})(1 - \frac{1}{\sqrt{2}})
\]

\[
= \frac{K}{n+1}(2 - \sqrt{2})n - 1,
\]

\[
g_2 = \frac{3K}{2(n+1)}(n - 1 - \cos \frac{\pi - \varphi}{3})(1 - \cos \frac{\pi - \varphi}{3}) + \frac{2K}{2(n+1)}(n - 1 - \cos \frac{\pi - \varphi}{3})(1 - \cos \frac{\pi - \varphi}{3})
\]

\[
- \frac{3K}{2(n+1)}(n - 1 - \cos \varphi)(1 - \cos \varphi).
\]

Supplementary Fig. 1. The function \( h := \sin \theta(n - 2 \cos \theta) \) for \( n \in \{2, \ldots, 9\} \). Higher curves correspond to larger \( n \).
\[
\frac{K}{n+1}(n-1-\cos \varphi)(1-\cos \varphi) \\
= \frac{K}{n+1} \left[ (2 - \frac{3}{2} \cos \frac{\varphi}{3} - \frac{3\sqrt{3}}{4} \sin \frac{\varphi}{3} - \frac{1}{2} \cos \varphi) n - \\
\left( \frac{3}{2} \sin^2 \frac{\varphi}{3} + \frac{1}{2} \sin^2 \varphi \right) \right] \\
> \frac{K}{n+1} \left[ (2 - \frac{3\sqrt{3}}{4} n - \frac{3}{8} (3 \sin^2 \frac{\varphi}{3} + \sin^2 \varphi) \right] \\
> \frac{K}{n+1} \left[ (2 - \frac{3\sqrt{3}}{4} n - \frac{7}{8} ) > g_1, 
\right]
\]

where we have omitted a few steps. It follows that \(\theta_{i,i+1} = \pi/N\) is optimal for all \(N \geq 2\).

**SUPPLEMENTARY NOTE 5: PHASE SYNCHRONIZATION**

We provide a formal statement and proof of our second synchronization condition which concerns phase synchronization in the case when all frequency matrices share a nullspace.

**Theorem 5 (Phase synchronization).** Consider the heterogeneous model (2) on the 2-sphere for which:

(i) There exists a nonzero vector \(v \in S^n\) such that

\[
\text{span} \{v\} = \bigcap_{i=1}^{N} \ker \Omega_i,
\]

(ii) and it also holds that

\[
\{x_1(0), \ldots, x_N(0)\} \subset H^n(v),
\]

where \(H^n(v) = \{y \in \mathbb{R}^{n+1} | \langle v, y \rangle > 0\}\) is an open hemisphere,

Then \(x_i = v\) for all \(i \in V\) is an asymptotically stable equilibrium. The region of attraction contains \(H^n(v)\).

**Proof.** Introduce a potential function

\[
V(x) := \sum_{i=1}^{N} \langle v, x_i \rangle,
\]

where \(\langle v, x_i \rangle > 0\) since \(x_i \in H^n(v)\). Calculate

\[
\dot{V} = \sum_{i=1}^{N} \langle v, \dot{x}_i \rangle \\
= \sum_{i \in I} \langle v, \Omega_i x_i \rangle + \sum_{i,j} k_{ij} \langle x_j - \langle x_i, x_j \rangle x_i \rangle \\
= \sum_{i=1}^{N} \sum_{j \in N_i} k_{ij} \langle v, x_j \rangle - \sum_{i=1}^{N} \sum_{j \in N_i} k_{ij} \langle x_i, x_j \rangle \langle v, x_i \rangle \\
= \sum_{i=1}^{N} \sum_{j \in N_i} k_{ij} \langle x_i, x_j \rangle \langle v, x_i \rangle \\
= \sum_{i=1}^{N} \sum_{j \in N_i} k_{ij} (1 - \langle x_i, x_j \rangle) \langle v, x_i \rangle \geq 0. \tag{17}
\]

Change the potential function \(V\) to

\[
U(x) := \max_{y \in (S^n)^N} V(y) - V(x).
\]

Note that \(U \geq 0\) and \(\dot{U} = -V \leq 0\). The inequality in Eq. (17) only holds with equality if \(\{x_i, x_j\} = 1\) for all \(\{i,j\} \in E\). By LaSalle’s principle, the system in Eq. (2) converges to an invariant set such that \(x_i = x_j\) for all \(\{l,k\} \in E\). Setting the time-derivative of agent \(i\) in (2) to zero yields \(x_i \in \ker \Omega_i\). This holds for all \(i\) whereby we obtain \(x_i = v\). \(\square\)

**SUPPLEMENTARY REFERENCES**

[1] J. Zhu, High-dimensional Kuramoto model limited on smooth curved surfaces, Physics Letters A 378, 1269 (2014).
[2] W. Kahan, Spectra of nearly hermitian matrices, Proceedings of the American Mathematical Society 48, 11 (1975).
[3] J. Markdahl, J. Thunberg, and J. Gonçalves, Almost global consensus on the n-sphere, IEEE Transactions on Automatic Control 63, 1664 (2018).
[4] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*, Vol. 3 (Prentice hall Upper Saddle River, NJ, 2002).
[5] S. Wright and J. Nocedal, Numerical optimization, Springer Science 35, 7 (1999).