Network measures of mixing

View Online   Export Citation   CrossMark

Ralf Banisch,[1,a] (iD) Péter Koltai,[1,b] (iD) and Kathrin Padberg-Gehle[2,c] (iD)

AFFILIATIONS

[1]Institute of Mathematics, Freie Universität Berlin, 14195 Berlin, Germany
[2]Institute of Mathematics and its Didactics, Leuphana Universität Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany

[a]ralf.banisch@fu-berlin.de
[b]peter.koltai@fu-berlin.de
[c]padberg@leuphana.de

ABSTRACT

Transport and mixing processes in fluid flows can be studied directly from Lagrangian trajectory data, such as those obtained from particle tracking experiments. Recent work in this context highlights the application of graph-based approaches, where trajectories serve as nodes and some similarity or distance measure between them is employed to build a (possibly weighted) network, which is then analyzed using spectral methods. Here, we consider the simplest case of an unweighted, undirected network and analytically relate local network measures such as node degree or clustering coefficient to flow structures. In particular, we use these local measures to divide the family of trajectories into groups of similar dynamical behavior via manifold learning methods.

Coherent structures play a ubiquitous role in mass transport and mixing in time-dependent dynamical systems. While most of the established identification methods require the knowledge of the underlying dynamical system or at least high-resolution trajectory data, this information may not be available in real-world scenarios. Network-based approaches applied to Lagrangian trajectories have been shown to be still successful in the analysis of transport and mixing processes even when data are sparse and incomplete. We focus on the local network measures of such trajectory networks and relate them to flow structures of the underlying system. Using manifold learning algorithms allows us to classify the phase space into regions of different dynamical behavior, complementing the frequently used spectral studies of flow networks.

I. INTRODUCTION

To set the scene for our approach, suppose we are given a nonautonomous ordinary differential equation (ODE)

$$\dot{x}(t) = f(t, x(t)), \tag{1}$$

with state $x \in \mathbb{M} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, time $t \in \mathbb{R}$, and sufficiently smooth right-hand side f to ensure existence and uniqueness of solutions $x(t) := x(t; t_0, x_0)$ to initial values (t_0, x_0). We may interpret f as a velocity field of a fluid flow and $x(t)$ as the position of an ideal

particle in that flow at time t. In this case, $\mathbb{M} \subset \mathbb{R}^2$ or $\mathbb{R}^3$. The time-parameterized family $(x(t))_t$ is referred to as a Lagrangian particle trajectory. We are interested in detecting coherent flow structures from the given trajectory data, that is, time-dependent regions in $\mathbb{M}$ that either inhibit or enhance transport and mixing processes of the underlying flow. In our setting, "transport" refers to the bulk movement of particles between different regions of the domain. Coherent structures trap particles for long times and in that way they determine the transport pathways of the underlying flow. "Mixing" refers to advective mixing (or stirring) by stretching and folding of fluid parcels. In contrast to transport, mixing describes a local flow property. Coherent structures resist mixing with neighboring regions of the flow. Thus, they are characterized by having mixing properties that are distinct from those of surrounding regions.

The mathematical definition and numerical study of coherent structures has been an area of intense research over the last two decades. In particular, different probabilistic and geometric methods have been proposed. These approaches have been discussed and systematically compared in a number of studies.[1,14,17] Geometric concepts aim at defining the boundaries between coherent sets, i.e., codimension-1 material surfaces in the flow that can be characterized by variational criteria.[20] Probabilistic methods are tailored to identify sets that are minimally dispersive while moving with the flow. Here, the main theoretical tools are transfer operators, i.e., linear Markov operators that describe the motion of probability densities under the

action of the nonlinear, time-dependent flow.[15] Both the geometric and the probabilistic approaches require high-resolution trajectory data from (1), that is, from a dense grid of initial conditions. This can be prohibitively expensive in complex systems, such as turbulent flows. Moreover, when the particle trajectories are obtained directly from measurements (e.g., from particle tracking experiments), then the Lagrangian data under consideration may even be sparse and incomplete.

To overcome these problems, different computational methods have been proposed to identify coherent behavior in flows directly from Lagrangian trajectory data. One of the earliest attempts is the braiding approach,[2] where trajectories are classified according to their intertwining pattern in space-time. This method is mathematically sound, but computationally demanding and currently restricted to two-dimensional flows. Other trajectory-based approaches use time-integrated quantities along trajectories.[4,22,25,26] Finally, there are attempts to reconstruct the transfer operator from limited amount of trajectory data[44] as well as the dynamic Laplacian,[13] which was recently introduced to study coherent sets as sets that keep an optimal boundary to volume ratio when evolved by the dynamics.[12]

Recent works focus on the use of spatiotemporal clustering algorithms.[3,16,19,37,39] There, the aim is to identify coherent sets as groups of trajectories that remain close and/or behave similarly in the time span under investigation. All these methods can deal with sparse and incomplete trajectory data, and their applicability has been demonstrated in several example systems.

Here, we revisit the framework introduced in Ref. 37, which is based on an unweighted, undirected network with the trajectories serving as nodes. A link is established between two nodes if the respective trajectories come close to each other at least once in the course of time. This construction is similar in spirit to the concept of recurrence networks,[9,10] but here in a spatiotemporal setting.

We note that the discretized transfer operator has also been viewed and treated as a network.[6,7,23,36,38,40] A recent review[8] addresses the different constructions of flow networks and their analyses.

In a previous work,[37] we have introduced the construction and have mainly considered spectral properties of our trajectory-based undirected and unweighted flow network, which allowed us to compare our approach with related spectral concepts.[3,19] We note that in these works the search for coherent sets is termed as a community detection problem for the resulting network and solved by a normalized cut method.[42]

In the present paper, our focus will be on the application and interpretation of "local network measures" such as node degrees or clustering coefficients. These and other quantities have been considered in previous work on recurrence networks,[10] where the authors could link the network measures to properties of the underlying dynamical system. In weighted, directed networks obtained from discretized transfer operators, the in- and out-degrees were found to highlight hyperbolic regions in the underlying flow,[23,40] whereas maxima of the local clustering coefficient could be related to regular or periodic behavior.[38] Similar properties appear to hold for the trajectory-based undirected and unweighted flow network, as demonstrated in example systems.[37]

Here, for the first time, we will draw an explicit analytical connection between these network measures and underlying flow structures. In particular, we will estimate the node degree in terms

of the finite-time Lyapunov exponent, an established quantity to measure stretching, and give some geometrical interpretation of the local clustering coefficient. Moreover, we will carry out an empirical flow classification based on further network measures using manifold learning methods. These complement the spectral approaches.[3,19,37,39]

The paper is organized as follows: In Sec. II, we review the construction of the simple trajectory-based network[37] as well as standard network local measures. In Sec. III, we will establish an analytical connection of some local network measures to the corresponding phase space structures. In particular, we will give analytical estimates of the node degree and of the local clustering coefficient. Further network measures are discussed in Sec. IV. In Sec. V, we will numerically demonstrate the estimates of Sec. III as well as an empirical network-based flow classification in a number of example systems, including the double-gyre flow and a real ocean surface flow from the AVISO data set.

II. A TRAJECTORY-BASED NETWORK

A. Construction of the network

Suppose we are given N trajectories from a flow simulation [i.e., as numerical solutions to (1)] or from a particle tracking experiment. In practice, the particle positions may be given at discrete times $\{0, 1, \ldots, T\}$. We denote the trajectories by x_i, $i = 1, \ldots, N$, and the positions at a certain time instance $t = 0, \ldots, T$ by $x_{i,t} \in \mathbb{R}^d$. We now fix some $\varepsilon > 0$ and set up a network with $x_1, \ldots, x_N$ as nodes. We link two nodes x_i and x_j if the respective trajectories come ε-close to each other at least once in the course of time $t \leq T$. Then, a symmetric adjacency matrix $A \in \{0, 1\}^{N \times N}$ describes the network, with

$$A_{ij} = \begin{cases} \max\limits_{0 \leq t \leq T} \mathbf{1}_{\mathbb{B}_\varepsilon(x_{i,t})}(x_{j,t}), & i \neq j \\ 0, & i = j \end{cases}, \qquad (2)$$

where $\mathbf{1}_{\mathbb{B}}$ denotes the indicator function of a set $\mathbb{B} \subset \mathbb{R}^d$. So $A_{ij} = 1$, that is, a link is established between trajectories x_i and x_j, if and only if at one or more time instances t, $x_{j,t}$ can be found in an ε-ball $\mathbb{B}_\varepsilon(x_{i,t})$ centered at $x_{i,t}$ and thus the trajectories x_i and x_j have come ε-close.

Naturally, other constructions are possible as well, e.g., setting A_{ij} to the number of ε-close encounters and similar. We will restrict our attention to the current one, as we consider it to be the simplest in terms of carrying the least quantitative information.

By an appropriate choice of ε, one ensures that the network defined by (2) is connected, and we will only consider connected networks from now on. Of course, the network topology depends also on other parameters, such as the particle density, the time-resolution, and the length of the trajectories. However, as these parameters are determined by the data, the only free parameter is ε, which can be reasonably chosen in relation to the particle density.[9,37]

B. Network analysis

The resulting network can be studied globally using the adjacency matrix. In particular, a normalized cut problem[42] is solved by considering leading eigenvectors of the generalized eigenvalue problem $Lv = \lambda Dv$, where D denotes the degree matrix (diagonal matrix with node degrees on the diagonal, i.e., $D_{ii} = \sum_j A_{ij}$, $i = 1, \ldots, N$). This spectral approach leads to the identification of

clusters in the network that correspond to coherent sets of the underlying flow,[3,19,37,39] that is, mobile regions in M that do not freely mix with the surrounding phase space regions.

As an alternative to these global, spectral approaches, we will attempt to tie in "standard local network measures"[27] with quantitative and qualitative dynamical behavior of the system. So far, such relations have been studied mainly heuristically,[10,23,36,40] but here we will draw some explicit analytical connections.

In the following, we will review some of the most frequently considered network measures.

The "node degree,"

$$d_i = \sum_j A_{ij}, \quad i = 1, \dots, N, \tag{3}$$

counts how many links are connected to node x_i. Similarly,

$$\langle d \rangle_{nn,i} = \frac{\sum_j A_{ij} d_j}{d_i}, \quad i = 1, \dots, N \tag{4}$$

denotes the "average node degree" of the neighbors of a node x_i.

The "local clustering coefficient" of a node x_i quantifies how tightly connected the subgraph induced by this node and its neighbors is (that is, how close this subgraph is to a complete graph). It is defined by

$$c_i = \frac{\text{\# triangles connected to } x_i}{\text{\# triples centered around } x_i} = \frac{(A^3)_{ii}}{d_i(d_i - 1)}, \tag{5}$$

where $i = 1, \dots, N$.

Many further network measures exist in the literature, we will mention only a few here.

The (normalized) "closeness centrality" (or "closeness") of a node in a network is given by its reciprocal mean distance to all other nodes in the network, i.e.,

$$\text{cl}_i = \frac{N}{\sum_j \text{dist}(x_i, x_j)}, $$

where $\text{dist}(x_i, x_j)$ denotes the distance (or shortest path length) between nodes x_i and x_j in the graph.

"Betweenness centrality" is related to that quantity and measures how many shortest paths in the network traverse a certain node. It is given by

$$b_i = \sum_{j,k \neq i}^N \frac{p_{jk}(x_i)}{p_{jk}}, $$

where $p_{jk}(x_i)$ denotes the number of shortest paths between nodes x_j and x_k that include x_i and p_{jk} is the total number of shortest paths between x_j and x_k.

III. ANALYTICAL ESTIMATES OF LOCAL NETWORK MEASURES

In a number of previous studies in recurrence networks,[10] transfer operator-based[23,36,40] as well as trajectory-based networks,[37] it has been observed that high values of the node degree can be related to regions of strong stretching, indicating hyperbolic behavior of the underlying flow. Moreover, high values of the local clustering coefficient have been related to regular/elliptic dynamics and periodic

behavior.[10,38] Here, we will establish an analytic connection between these local network measures and the finite-time Lyapunov exponent, a frequently used indicator of hyperbolic dynamics.

Although we consider time-continuous dynamics for our analytic investigations, the results here "qualitatively" carry over to discrete-time systems as well, by replacing the corresponding discrete-time analogs of the objects in consideration. The quantitative result, Proposition 3, can be retained from the discrete-time system in the limit of vanishing sampling step size, if the discrete-time system is the finite-time flow map of a continuous-time system.

A. Preliminaries

Given a time-dependent flow $\phi(s, t)$ generated by the ODE (1), i.e., $x(t) = \phi(s, t)x(s)$ that solves $\dot{x}(t) = f(t, x(t))$, let us assume that this flow generates our trajectory data. Herein, $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is a sufficiently smooth vector field. For simplicity, we assume that the flow is volume-preserving for all times, i.e., $\text{div}_x f(t, \cdot) \equiv 0$ for all t, where div_x is the divergence operator for functions mapping $\mathbb{R}^d$ to itself.

Let $W(s, t) \in \mathbb{R}^{d \times d}$ be the "fundamental matrix" (Wronskian matrix) defined by

$$W(s, t) = D_x(\phi(s, t)x), \tag{6}$$

where D_x denotes the derivative with respect to x. Hence, the fundamental matrix is the derivative of the flow with respect to its initial condition. It is a first-order (linear) approximation of perturbation propagation along trajectories, since it holds that if $\tilde{x}$ satisfies (1) with $\tilde{x}(s) = x(s) + \delta x$, then

$$\tilde{x}(t) = x(t) + W(s, t)\delta x + \mathcal{O}(\|\delta x\|^2), \tag{7}$$

as $\|\delta x\| \to 0$. The fundamental matrix satisfies the initial value problem[28]

$$\dot{W}(s, t) = D_x f(t, x(t)) W(s, t), \qquad W(s, s) = I, \tag{8}$$

where the time-derivative on the left-hand side is with respect to t, and $I \in \mathbb{R}^{d \times d}$ denotes the identity matrix. We note the following properties:

Proposition 1. As the vector field f is divergence-free, we have $\det(W(s, t)) = 1$, for every x and $s \leq t$.[29]

The finite-time Lyapunov exponent (FTLE) for $x = x(s)$ is defined as

$$\Lambda(x, s; t) = \frac{1}{|t - s|} \log(\sigma_1(W(s, t))), \tag{9}$$

where $\sigma_1(W(s, t))$ denotes the largest singular value of the matrix $W(s, t)$.

Let us now turn our attention to the properties of the dynamic neighborhood network defined by (2). Fix some $\varepsilon > 0$. Let us recall our initial setting again, where we have some sampling points and their trajectories in a volume-preserving flow. Assuming the initial distribution of the sampling points being uniform (in some spatial domain of interest), and there are sufficiently many of them, the number of sampling points in the ε-neighborhood of any point is

proportional to the volume of the ε-ball around that point.[30] Denote

$$
\mathbb{G}_\varepsilon(x) = \bigcup_{t=t_0,\ldots,t_{T-1}} \left\{ \tilde{x} \,\middle|\, \phi(t_0,t)\tilde{x} \in \mathbb{B}_\varepsilon(x(t)) \right\}
$$

$$
= \bigcup_{t=t_0,\ldots,t_{T-1}} \phi(t_0,t)^{-1}\mathbb{B}_\varepsilon(x(t)), \tag{10}
$$

where $x(t)$ solves (1) with $x(t_0) = x^i_{t_0}$ and $\phi(t_0,t)^{-1} := \phi(t,t_0)$ denotes the time-reversed flow from time t back to t_0. Thus, $\mathbb{G}_\varepsilon(x)$ denotes the set of all initial states that are ε-close to the trajectory starting at x at some t_i for $i = 0, 1, \ldots, T$.

B. The degree

As the initial distribution is uniform, in the case of many data points (more precisely, in the $N \to \infty$ limit), the Lebesgue volume of a set is proportional to the number of data points lying in it. Thus, the degree d_i of node x_i in the network given by the adjacency matrix (2) (the row sum of A) is proportional to the set of initial conditions that get ε-close to $x_i(t)$ at some time in the observation interval, i.e.,

$$
d_i = \sum_{j=1}^{N} A_{ij} \propto \mathrm{vol}(\mathbb{G}_\varepsilon(x_i)). \tag{11}
$$

Here, the proportionality holds for large N and with the usual Monte Carlo error of $\mathcal{O}(N^{-1/2})$. We will now estimate the volume in (11) to first order in ε using the linear perturbation propagation relation (7). For a fixed t, let $\tilde{x}$ be such that $y = \phi(t_0,t)\tilde{x} \in B_\varepsilon(x_i(t))$ and $\delta y = y - x_i(t)$. Then, by (7), we obtain

$$
\|\tilde{x} - x_i(t_0)\| = \|W(t_0,t)^{-1}\delta y\| + \mathcal{O}(\varepsilon^2). \tag{12}
$$

Note that $W(s,t)$ is always invertible due to Proposition 1.
We have from (12) that

$$
d_i \overset{\cdot}{\propto} \mathrm{vol}\left(x_i(t_0) + \bigcup_{t=t_0,\ldots,t_{T-1}} W(t_0,t)^{-1}\mathbb{B}_\varepsilon(0) \right), \tag{13}
$$

where $\overset{\cdot}{\propto}$ denotes proportionality up to errors of size $\mathcal{O}(\varepsilon^2)$. Note that the translation by $x_i(t_0)$ does not change the volume and could be omitted. The sets $W(t_0,t)^{-1}\mathbb{B}_\varepsilon(0)$ are ellipsoids with semiaxes of length $\sigma_1^{-1}, \ldots, \sigma_d^{-1}$, the reciprocal singular values of the matrix $W(t_0,t)$, and these semiaxes are aligned with the corresponding right singular vectors.

We now restrict our considerations to the case of a two-dimensional area-preserving flow with states $(x,y) \in \mathbb{R}^2$.

1. Time-invariant singular vectors

For $d = 2$, we have by Proposition 1 that $\sigma_1\sigma_2 = 1$. We will make the following simplifying assumption.

Assumption 2. Let $\sigma(t)$ denote the larger of the both singular values of $W(t_0,t)$, and $v(t) \in \mathbb{R}^2$ the corresponding right singular vector. We assume that $v(t) = v$ is independent of t.

This means that the other right singular vector is also independent of t. In other words, we assume that the direction at initial time which undergoes the largest stretching (and also the one undergoing the largest squeezing) is independent of t.
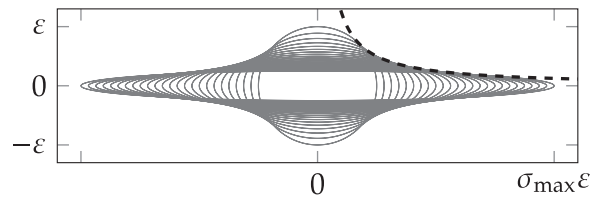
FIG. 1. Linearized pullbacks of ε-circles. Dashed line: the graph of $y = \frac{1}{2x}$, touching the union of the pullback circles for $x \in [\frac{1}{\sqrt{2}}, \frac{\sigma_{\max}}{\sqrt{2}}]$. In this visualization, $\sigma_{\max} = 4$.

In order to simplify notation, we fix initial time $t_0 = 0$ and final time $T > 0$. Let us consider the set

$$
\mathbb{S}(0,T) = \bigcup_{t\in[0,T]} W(0,t)^{-1}\mathbb{B}_\varepsilon(0),
$$

i.e., a union over continuous time. We would like to compute the two-dimensional volume of $\mathbb{S}(0,T)$. It is a union of ellipses with major semiaxes of length $\sigma \in [0, \sigma_{\max}]$, where we also assume $\sigma_{\max} = \sigma(T)$, i.e., the largest stretching appears at final time. Without loss, the larger semiaxis is assumed to be aligned with the x-axis. The resulting set $\mathbb{S}(0,T)$ is depicted in Fig. 1. Next, we will derive an analytic formula for the volume of $\mathbb{S}(0,T)$. Those readers not interested in the details of the derivation can skip to the result in Proposition 3.

If $\sigma(t), \sigma(t)^{-1}$ are the singular values of $W(0,t)$, with $\sigma(t) \geq 1$, then the ellipse $W(0,t)^{-1}\mathbb{B}_\varepsilon(0)$ is given in the xy-plane by the equation

$$
\frac{x^2}{\sigma(t)^2} + \sigma(t)^2 y^2 \leq \varepsilon^2.
$$

To obtain the boundary point (x,y) of $\mathbb{S}(0,T)$ for fixed x, we need to maximize y over the possible values $\sigma(t) \in [1, \sigma(T)]$, because the outermost ellipse defines the boundary point of the union. A straightforward computation shows that $y = \frac{\varepsilon^2}{2x}$, which is realized by the ellipse with semimajor axis length $\sigma = \sqrt{2}x$. This holds for $|x| \in \left[\frac{\varepsilon}{\sqrt{2}}, \frac{\sigma(T)}{\sqrt{2}} \right]$. For smaller x (in magnitude), the boundary coincides with that of the circle of radius ε; for larger x, it is the ellipse with semimajor axis length $\sigma(T)$ constituting the boundary. This suggests to compute the volume of $\mathbb{S}(0,T)$ as sum over the three distinct intervals of x-values,

$$
\mathrm{vol}(\mathbb{S}(0,T)) = 4(V_1 + V_2 + V_3), \qquad V_i = \mathrm{vol}(\mathbb{A}_i),
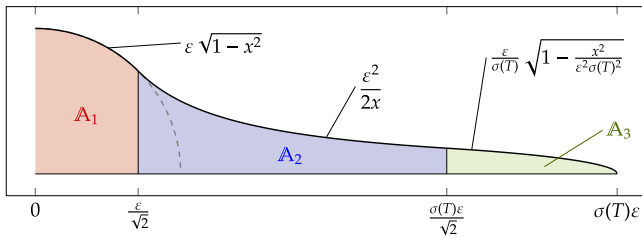$$

where, as depicted in Fig. 2,

- $\mathbb{A}_1 = \left\{ (x,y) \,\middle|\, 0 \leq x \leq \frac{\varepsilon}{\sqrt{2}},\, 0 \leq y \leq \varepsilon\sqrt{1-x^2} \right\}$;
- $\mathbb{A}_2 = \left\{ (x,y) \,\middle|\, \frac{\varepsilon}{\sqrt{2}} \leq x \leq \frac{\sigma(T)\varepsilon}{\sqrt{2}},\, 0 \leq y \leq \frac{\varepsilon^2}{2x} \right\}$; and
- $\mathbb{A}_3 = \left\{ (x,y) \,\middle|\, \frac{\sigma(T)\varepsilon}{\sqrt{2}} \leq x \leq \sigma(T)\varepsilon,\, 0 \leq y \leq \frac{\varepsilon}{\sigma(T)}\sqrt{1-\frac{x^2}{\varepsilon^2\sigma(T)^2}} \right\}$.

Direct computation gives[31]

$$
V_1 = \frac{\pi+2}{8}\varepsilon^2, \quad V_2 = \log(\sigma(T))\frac{\varepsilon^2}{2}, \quad V_3 = \frac{\pi-2}{8}\varepsilon^2.
$$

Thus, we obtain

FIG. 2. Partitioning the union of linearized pullbacks of the ε-circle in the idealized situation where the right singular vectors of the fundamental matrix are not changing in time.

Proposition 3. Under Assumption 2, one has

$$\mathrm{vol}(\mathbb{S}(0,T)) = \big(\pi + 2\log(\sigma(T))\big)\varepsilon^2.$$

That means, in a linear regime, where the node degree approximates $\mathrm{vol}(\mathbb{S}(0,T))$, one can expect an affine-linear relationship between node degree and the FTLE $\Lambda(\cdot, 0, T) = \frac{1}{T}\log(\sigma(T))$, at least when the singular vectors do not change much in time. $\mathrm{vol}(\mathbb{S}(0,T)) \gg \big(\pi + 2\log(\sigma(T))\big)\varepsilon^2$ is to be expected, when $\sigma(T) \gg 1$ and the singular vectors are strongly time-dependent.

2. Singular vectors with changing direction

Assumption 2 is not realistic in practice. However, in the following, we will show reasons that it is also unrealistic that it is strongly violated in a "quantitative sense," as this would require unlikely strong vector fields.

Let $W(0,t) = U(t)\Sigma(t)V(t)^T$ be the singular value decomposition of the fundamental matrix. As above, none of its singular values are zero, and thus $W(0,t)^{-1} = V(t)\Sigma(t)^{-1}U(t)^T$, yielding that the semiaxes of $W(0,t)^{-1}\mathbb{B}_\varepsilon(0)$ are aligned with the columns of $V(t)$, i.e., the right singular vectors of $W(0,t)$.

Let us consider now the conditions on the dynamics that are necessary for the right singular vectors to change. As we are interested in the action of $W(0,t)^{-1}$ on $\mathbb{B}_\varepsilon(0)$, and $U(t)^T\mathbb{B}_\varepsilon(0) = \mathbb{B}_\varepsilon(0)$ (orthogonal transformations keep the unit ball unchanged), we set $U(t) = I$ for simplicity. Also, by continuity of $t \mapsto V(t)$, we have $\det(V(t)) = 1$, and in two dimensions, it means that $V(t)$ is a rotation.

The main insight can be seen in the following prototypical example, where we rotate the columns of $V(t)$ with angular frequency $\omega(t)$ such that

$$\dot{V}(t) = \omega(t)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}V(t).$$

Further, let $\sigma_1 = \sigma_1(t)$ and $\sigma_2 = \sigma_2(t) = \frac{1}{\sigma_1(t)}$ be the diagonal elements of $\Sigma(t)$. Thus, $W(0,t) = \Sigma(t)V(t)^T$. Under our

assumptions, we obtain

$$\dot{W}(0,t) = \begin{pmatrix} \dot{\sigma}_1 & 0 \\ 0 & \dot{\sigma}_2 \end{pmatrix}V(t)^T + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}\begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}V(t)^T$$

$$= \underbrace{\begin{pmatrix} \dot{\sigma}_1/\sigma_1 & \omega\sigma_1/\sigma_2 \\ -\omega\sigma_2/\sigma_1 & \dot{\sigma}_2/\sigma_2 \end{pmatrix}}_{=D_x v,\ \mathrm{cf.}\ (8)} W(0,t),$$

so the spatial derivative of the vector field v has a component $\omega\sigma_1/\sigma_2 = \omega\sigma_1^2$. That is, for the right singular vector of the fundamental matrix to change its direction with unit speed, the vector field v needs to have a large spatial derivative, provided $\sigma_1(t)$, the accumulated stretching from time t_0 to time t is large. For hyperbolic trajectories $x(t)$, the singular value $\sigma_1(t)$ of the fundamental matrix grows exponentially in time. Thus, either the spatial derivative of the velocity field grows exponentially as well or the change in the direction of the singular vector is exponentially slow.

We conclude that the larger the already present local stretching in the system is, the more unlikely it is that smooth vector fields change the direction of the corresponding right singular vector significantly. With this, even if Assumption 2 is violated, it is likely that the direction of the dominant singular vector (for trajectories showing considerable hyperbolic behavior) shows a step-function like behavior in time, as this direction is likely to change only in time intervals where $\sigma_1(t) \gg \sigma_2(t)$. By superposing the corresponding linearized pullback of the unit circle, one obtains a superposition of single "galaxies," each as in Fig. 1 with different sizes and major axis directions. This is what we often observe in the examples below. In summary, the degree correlates largely with the FTLE but, in addition, it takes rotation into account as well as nonlinear effects. In particular, the FTLE measures only stretching in the dominant direction, whereas the degree also captures expansion in the other directions. Moreover, if the singular vectors change their directions significantly, this may lead to a moderate degree even if the FTLE is small.

These theoretical considerations are underlined by the numerical example in Sec. V B.

Remark 4. We note that the only results that require the dynamics to be area-preserving are Proposition 3, the considerations in Sec. III 2 and those in Fig. 3. These results rely on the fact that the fundamental matrix of the system has two singular values whose product is one. The qualitative classification of dynamical regimes in Table I is independent of area-preservingness.
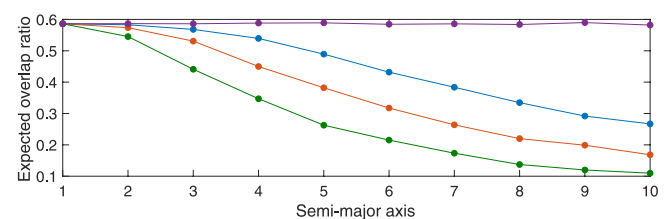
FIG. 3. Expected relative overlaps of two ellipses depending on the semimajor axis and the rotation angle [0 (purple), up to $\pm\frac{\pi}{32}$ (blue), $\pm\frac{\pi}{16}$ (red), $\pm\frac{\pi}{8}$ (green)].

TABLE I. Network measures and finite-time dynamical regimes.

Degree	Clustering coefficient	Finite-time dynamical regime
Small–moderate	Large	Elliptic or parabolic motion
Moderate	Small	Filamentation (finite-time hyperbolicity), no mixing
Moderate	Moderate	Stickiness (mixing close to regular regions)
Large	Large	Mixing

C. Clustering coefficient

Recall that the clustering coefficient (5) of a node x_i is defined as

$$c_i = \frac{\sum_{j,k}^{N} A_{ij} A_{jk} A_{ki}}{d_i(d_i - 1)}, \tag{14}$$

where A_{ij} is as in (2). It counts the number of triangles with vertex x_i divided by the total number of possible triangles. In the limit of very many data points (and thus large degree), we can approximate the clustering coefficient by simplifying the denominator above, yielding

$$\tilde{c}_i = \frac{\sum_{j,k}^{N} A_{ij} A_{jk} A_{ki}}{d_i^2}. \tag{15}$$

From now on, this equation will be used and the tilde will be dropped. It is useful to write this as

$$c_i = \frac{1}{d_i} \sum_j A_{ij} f_j^{(i)}, \qquad f_j^{(i)} = \frac{1}{d_i} \sum_k A_{jk} A_{ki}.$$

1. Geometrical interpretation of the clustering coefficient

We already know that $N^{-1} d_i$ converges to $\mathrm{vol}\,(\mathbb{G}_\varepsilon(x_i))$, i.e., the volume of the galaxy neighborhood, in the infinite data limit $N \to \infty$. Further, we may write $A_{ij} = \mathbf{1}_{\mathbb{G}_\varepsilon(x_i)}(x_j)$ and then $f_j^{(i)}$ converges in the Monte Carlo sense to the function $f^{(i)}$ evaluated at the point x_j,

$$f^{(i)}(x_j) = \frac{1}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))} \int \mathbf{1}_{\mathbb{G}_\varepsilon(x_j)}(y) \mathbf{1}_{\mathbb{G}_\varepsilon(x_i)}(y) dy$$

$$= \frac{1}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))} \int \mathbf{1}_{\mathbb{G}_\varepsilon(x_j) \cap \mathbb{G}_\varepsilon(x_i)}(y) dy$$

$$= \frac{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i) \cap \mathbb{G}_\varepsilon(x_j))}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))}.$$

This is the proportion of $\mathbb{G}_\varepsilon(x_i)$ that overlaps $\mathbb{G}_\varepsilon(x_j)$. Finally, the clustering coefficient converges to

$$c(x_i) = \frac{1}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))} \int \mathbf{1}_{\mathbb{G}_\varepsilon(x_i)}(y) f^{(i)}(y) dy$$

$$= \frac{1}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))} \int_{\mathbb{G}_\varepsilon(x_i)} \frac{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i) \cap \mathbb{G}_\varepsilon(y))}{\mathrm{vol}(\mathbb{G}_\varepsilon(x_i))} dy. \tag{16}$$

In other words, $c(x)$ is the expected relative overlap of the neighborhood $\mathbb{G}_\varepsilon(x)$ and a second neighborhood $\mathbb{G}_\varepsilon(y)$, where y is drawn from $\mathbb{G}_\varepsilon(x)$.

In the linear regime, that is, for small ε and appropriate time spans, these neighborhoods and their volumes can be approximated using the variational equation and the estimates in Subsection III B. When both $\mathbb{G}_\varepsilon(x)$ and $\mathbb{G}_\varepsilon(y)$ are balls, the expected overlap can be explicitly computed, but this is already no longer possible when ellipsoids have to be taken into account.

In Fig. 3, we have numerically (via a Monte Carlo approach) estimated the expected relative overlaps of equally sized ellipses in the 2D case depending on the length of the semimajor axis $\sigma \geq 1$, with the semiminor axis being σ^{-1}. The purple curve corresponds to the (unrealistic) case, when the overlapping ellipses are exactly axis-parallel. Here, the numerical estimation is independent of σ and nicely matches the theoretical value for two overlapping circles of $1 - \frac{3\sqrt{3}}{4\pi} \approx 0.59$.[32] For the other curves, the overlapping ellipses are allowed to be slightly rotated [rotation angles up to $\pm \frac{\pi}{32}$ (blue), $\pm \frac{\pi}{16}$ (red), $\pm \frac{\pi}{8}$ (green)], and in these cases, the expected relative overlap decreases as σ is increased. In particular, an increase in the maximum rotation angle also leads to a decrease in the expected overlap ratio when the length of the semimajor axis are kept fixed.

This confirms the frequent observation that the local clustering coefficient takes large values where the dynamics is elliptic.[10,38] In this case, the corresponding galaxy neighborhood is ball-like and thus the FTLE and also the degree are small. Moreover, the local clustering coefficient is small when there is strong stretching and the FTLE is large. So, we expect that the FTLE and the local clustering coefficient are strongly negatively correlated. However, due to the finite ε in the network construction, the local clustering coefficient also measures nonlinear effects as we will discuss in the following.

2. Time-dependent behavior

Note that by construction, the set of neighbors in the network increases in a nested manner as the observation time interval increases: If $A_{ij} = 1$ for trajectories observed for the time interval $[t_0, t]$, naturally $A_{ij} = 1$ holds also for the time interval $[t_0, t']$ with $t' > t$. Unlike the degree, which thus increases monotonically in time, we expect the "qualitative" behavior of the clustering coefficient to change in time considerably.

In a dynamically mixing region, the clustering coefficient starts for small times with a moderate to large value, as the network for small times is based on vicinity of initial points. Then, as the observation time grows, it decreases, as new neighbors are introduced which are not necessarily neighbors of neighbors (due to hyperbolic stretching). This holds in an intermediate time interval for which the image of small balls under the dynamics is a filamented set, but does not yet fold back to itself and "cover" full-dimensional subsets of the state space. Then, as time increases, we expect the clustering coefficient to grow again, as eventually any two points get close-by in a mixing region again,[33] and the filaments tend to become "space-filling." To be more precise, once the filamentation is so strong that in a measure-theoretic sense $\mathbb{G}_\varepsilon(x_i) \cap \mathbb{G}_\varepsilon(y) \approx \mathbb{G}_\varepsilon(x_i)$ for every $y \in \mathbb{G}_\varepsilon(x_i)$, then by (16), the clustering coefficient gets large again.

In regular regions—where the dynamics is not distorting strongly and the mutual distances between points barely change—we expect the clustering coefficient to stay approximately constant all the time.

We summarize the expected characteristics of the network measures which we expect for different finite-time dynamical behavior in Table I.

IV. DISCUSSION OF FURTHER NETWORK MEASURES

Let us now briefly discuss the expected behavior of the other network measures from above. To this end, it is helpful to differentiate two kinds of dynamical behaviors that are very characteristic of complicated flows we are interested in. The first we connect to "mixing regions," where (weak) mixing is understood in the measure-theoretic sense.[43] The second kind is connected to "regular regions" and refer to those regions in state space that are not mixing, and we think of them as regions performing a rigid-body motion—up to slight distortions. Naturally, this is not a complete or well-defined partition of the flow domain, as it uses notions (like mixing) that are defined for infinite time, and we are looking at finite-time intervals; so there could be a whole homotopy of characteristics connecting these two. Still, as a descriptor of the two "extreme cases," it will prove very useful in the following.

A. Closeness

As network measures get more complicated, it gets significantly more involved to connect them directly to the dynamical behavior of a system. It is safe to claim that in a mixing region most trajectories eventually "meet" one another, their mutual distances will be small, giving a large closeness value. In a regular region, the mutual arrangement of trajectories stays similar in time, thus their distance stays moderate to large, and hence we expect them to have a moderate closeness.

Because the set of neighbor nodes increases in a nested manner in time, the length of shortest paths decrease monotonically as the observation time increases, and thus closeness increases. In contrast to the clustering coefficient, we do not expect closeness to change its qualitative behavior in time: Based on the above considerations, closeness increases more rapidly in mixing regions than in regular regions, but the qualitative picture with respect to this network measure does not change with the length of the time interval in consideration.

Note that closeness here works with distances of trajectories with respect to shortest paths in the graph given by A. A concept in its nature similar to this was put forward in Ref. 21, where a "semidistance of mixing"[34] for a finite set of Lagrangian trajectory data is derived. Further, it is shown that it can be computed by shortest paths in time-dependent graphs that comprise similar information to the ε-neighbor adjacency graphs at some fixed time instant t [without accumulating the neighbors in time, as done in (2)]. More crucially, it is demonstrated that coherent sets—sets that move with the flow and do not mix with their exterior while doing so—are regions "maximally far" from one another with respect to this distance. This connection suggests that coherent regions have a large mean distance to the rest of the network and that their closeness is thus smaller.

Our observations for the one-dimensional example in Sec. V A confirm this. As closeness is much more expensive to compute than the other network measures considered here [as shortest paths between "all" pairs of node need to be computed, giving a best-case complexity of $\mathcal{O}(n^3)$ for the Floyd–Warshall algorithm and $\mathcal{O}(n^2 \log(n))$ for the Dijkstra algorithm with Fibonacci heap], we refrain from further numerical analysis of closeness.

Finally, we note that closeness is influenced also by the relative position of a trajectory with respect to the other trajectories; toward the "boundary" of the trajectory ensemble, closeness is expected to be smaller. This is nicely reflected in the experimental example of Sec. V C.

B. Betweenness centrality

The betweenness centrality b_i of a node x_i measures the proportion of shortest paths of the network that traverse x_i. Thus—like closeness—it takes global information of the network into account, and computationally it has the same complexity. In general networks, b is large for nodes that connect different almost-decoupled subgraphs. Such nodes, often called "hubs," play a crucial role in the transfer of information or mass transport. So, nodes with high betweenness centrality in our trajectory-based network correspond to structures that connect the different coherent regions of the underlying system.

C. Average node degree

While the local degree of a node x_i only takes the direct neighboring nodes of x_i into account, by taking their average degree $\langle d \rangle_{nn,i}$ an extended neighborhood is considered. The resulting average degree field is a smoothed version of the degree field, with similar properties and with d and $\langle d \rangle_{nn}$ strongly correlated. Significant quantitative differences in the fields may only occur at nodes x_i where the node degree is locally maximal or minimal. Such a difference is measured by the degree anomaly $\Delta d_i = d_i - \langle d \rangle_{nn,i}$, which may serve as an indicator of the local heterogeneity of the phase space structures.[10]
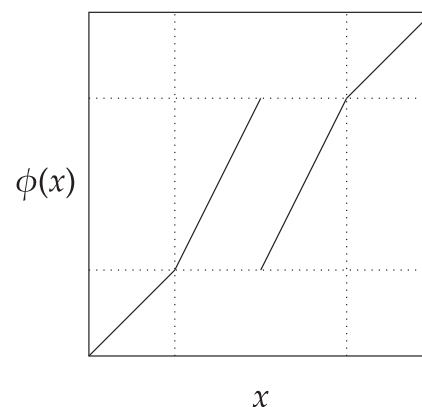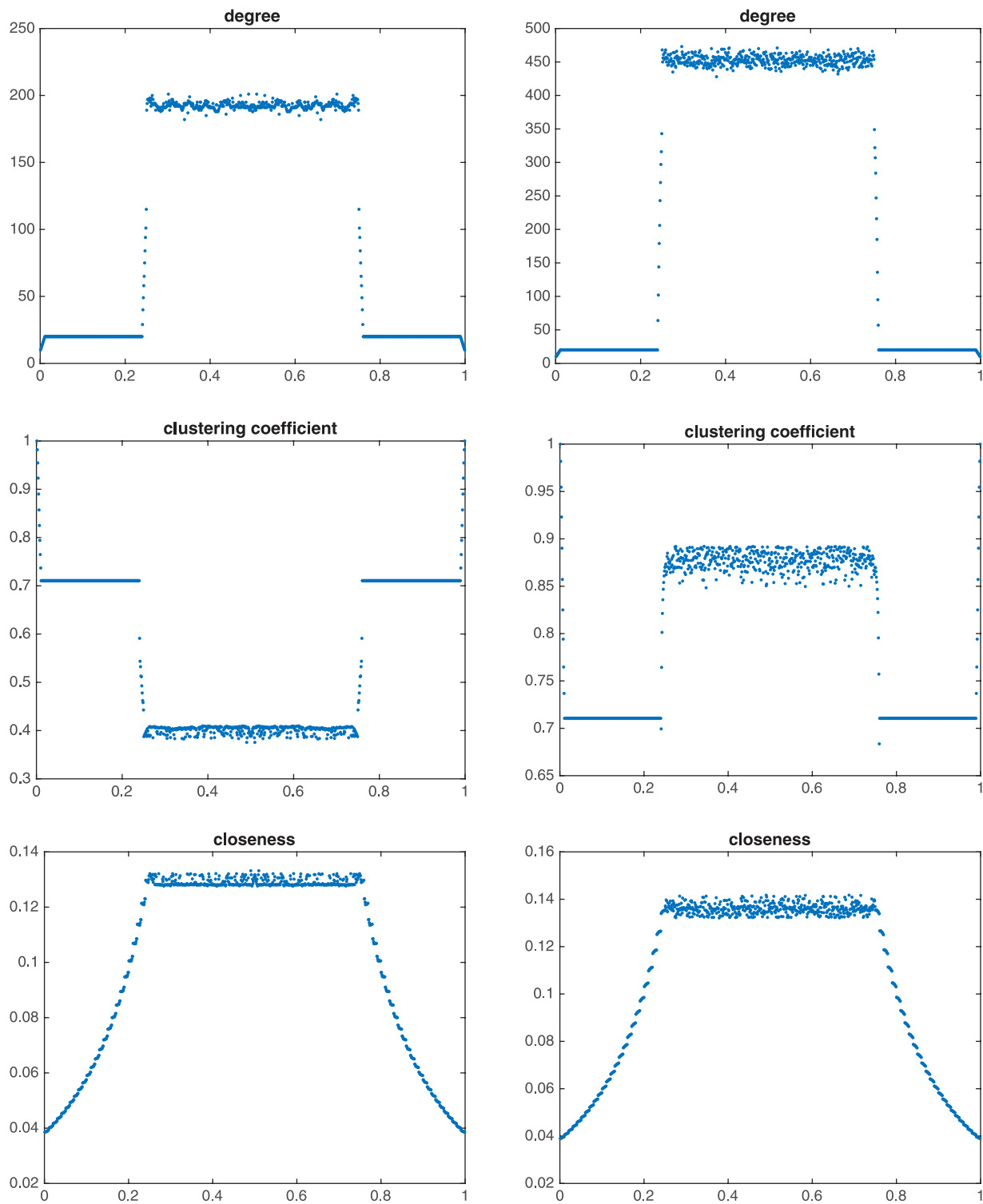

FIG. 4. The system (17).

FIG. 5. Network measures and resulting classification for 20 (left) and 100 (right) steps of the prototypical one-dimensional system (17). Note that only the behavior of the clustering coefficient changes "qualitatively," as we also expected.

V. NUMERICAL EXAMPLES

A. One-dimensional prototypical example

To gain some intuition about regular and mixing regions in a "controlled environment," let us now consider the discrete-time system on $\mathbb{M} = [0, 1]$ given by (see Fig. 4)

$$\phi(x) = \begin{cases} x, & x \in [0, \frac{1}{4}] \cup (\frac{3}{4}, 1] \\ \left(2(x - \frac{1}{4}) \mod \frac{1}{2}\right) + \frac{1}{4}, & x \in [\frac{1}{4}, \frac{3}{4}], \end{cases} \quad (17)$$

This map has three invariant sets. The left and right ones are static such that the mapping restricted to them is the identity, and are meant to model regions of the state space in complicated flows, that are "regular" in the sense that the mutual distance of points is not changed (or just barely) by the dynamics. We will consider these as one kind of prototype for coherent vortices. The third region physically separates the other two, and the dynamics on it is mixing (as it is the circle doubling map).

We carry out our computations for a network with 1000 initially equispaced trajectories and $\varepsilon = 0.01$. By this choice, all points (neglecting the boundaries) have initially 18 neighbors, resulting in $\frac{18 \cdot 17}{2}$ triples. The initial number of triangles of a node x_i is 108, and thus the proportion of triples that are triangles is $\frac{2 \cdot 108}{18 \cdot 17} \approx 0.706$. These quantities, 18 and 0.706, coincide with the degree and the local clustering coefficient, respectively, for all times in the static regions, as shown in Fig. 5. Moreover, this study confirms what we expected from the analysis in Sec. III: the degree and closeness are monotonic in the observation time span, but do not change qualitatively, while the clustering coefficient shows a very clear qualitative change between the regular and mixing region as the time span grows. This

is highlighted in Fig. 6, where the time-evolution of the degree and clustering coefficient is shown, averaged over the mixing region.

B. Periodically driven double-gyre flow

As a benchmark problem for analyzing flow structures, we consider the double-gyre flow,[41] a time-dependent system of differential equations,

$$\dot{y} = -\pi A \sin(\pi f(y, t)) \cos(\pi z), \quad (18)$$

$$\dot{z} = \pi A \cos(\pi f(y, t)) \sin(\pi z) \frac{df}{dy}(y, t),$$

where $x = (y, z) \in \mathbb{R}^2$ is the state, $f(y, t) = \delta \sin(\omega t) y^2 + (1 - 2\delta \sin(\omega t)) y$. We choose parameter values $A = 0.25$, $\delta = 0.25$, $\omega = 2\pi$, and fix $t_0 = 0$. We obtain a flow of period $\tau = 1$ on the domain $\mathbb{M} = [0, 2] \times [0, 1]$.

Figure 7 shows $\mathbb{S}(0, T)$ for $T = 5$ and different initial conditions,

$$x_{1,0} = \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix}, \quad x_{2,0} = \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}, \quad x_{3,0} = \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix},$$

$$x_{4,0} = \begin{pmatrix} 0.86 \\ 0.25 \end{pmatrix}, \quad x_{5,0} = \begin{pmatrix} 0.99 \\ 0.01 \end{pmatrix}, \quad x_{6,0} = \begin{pmatrix} 0.98 \\ 0.25 \end{pmatrix}.$$

Recall that the set $\mathbb{S}(0, T)$ is the union over the observation times of pullbacks of the unit circle with respect to the linearized dynamics, and thus we color the single pullbacks with respect to the time they belong to. The inner white regions are not filled because we only plot the boundaries of the respective ellipses.

Assumption 2 extended by the simplified considerations in Sec. III B seems to be valid: For the trajectories showing strong finite-time hyperbolic behavior (x_1, x_5, and x_6 moderately), the direction
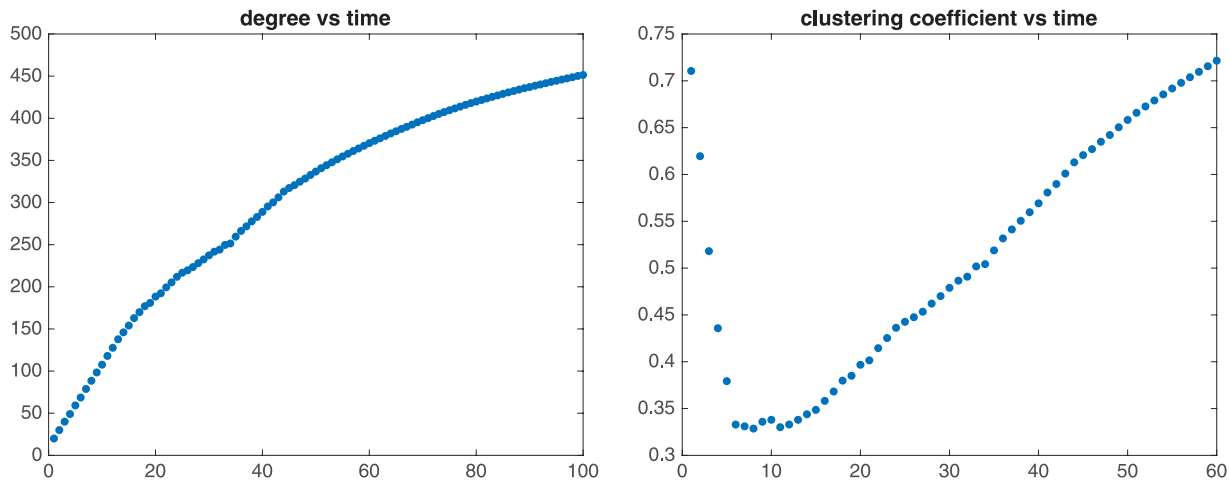

FIG. 6. The evolution of degree and clustering coefficient (averaged for the mixing region $\mathbb{M}_0 = [1/4, 3/4]$) as the time window $[0, t]$ grows. As the mapping on $\mathbb{M}_0$ is the circle doubling map, in every iteration, half of the neighborhood of every point is exchanged, thus the degree grows initially by $\sim$9 per iteration. As the time window grows, this value increases, as by mixing every point eventually gets arbitrarily close to any other in $\mathbb{M}_0$, and eventually there are no new neighbors to find. The clustering coefficient also shows the behavior we expect; first dropping, as the neighborhood grows rapidly by mixing, then growing, as arbitrary two neighbors are likely to meet at some point, also due to mixing.
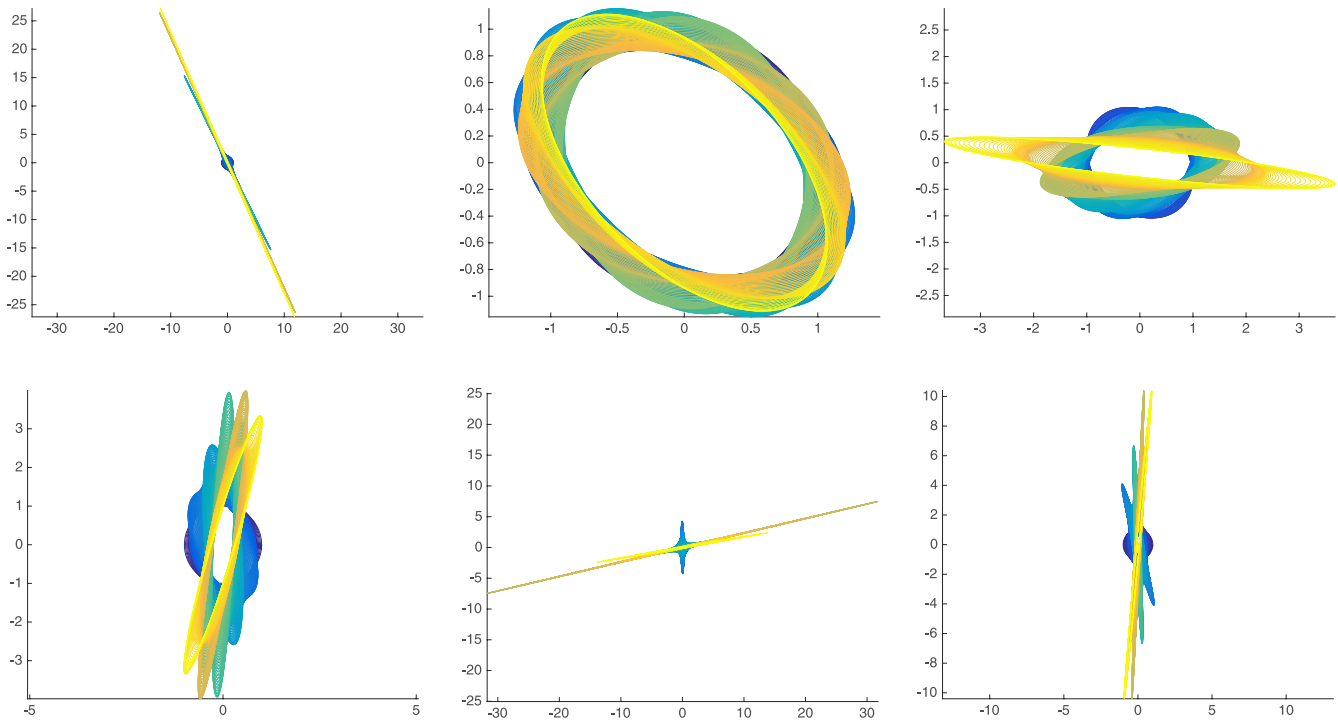
FIG. 7. The linearized pullback set $\mathbb{S}(0, T)$ of the periodically driven double-gyre system for $T = 5$ and different initial conditions. The color of the contour line corresponds to the time t (blue: initial, yellow: final).

of the pullback ellipses stay constant or change rapidly in short time intervals when the associated ellipse has small eccentricity (i.e., when $\sigma_1/\sigma_2 \ggg 1$). This is confirmed by Fig. 8. Therein, the top row shows the time evolution of the orientation of the first singular vector $v_1(t)$ as an angle in $[0, 2\pi)$ for the trajectories x_1, x_5, x_6 (left to right), while the bottom row shows the time evolution of $\sigma_1(t)$ on a logarithmic scale.

We also construct a network with 500×251 trajectories spaced equally on a grid in the domain $[0, 2] \times [0, 1]$. We choose $\varepsilon = 0.03$, flow time $T = 20$, and $\Delta t = 0.1$. Degree, clustering coefficient, and the sets $\bigcup_{t \in [0,5]} \phi(t_0, t)^{-1} \mathbb{B}_\varepsilon(x(t))$ for $x_1, \ldots, x_6$, of which $\mathbb{S}(0, 5)$ are linearized approximations, shown in Fig. 9.

For comparison, we also compute the FTLE field (9) with the same resolution [see Fig. 9 (bottom left)]. We see that the affine-linear
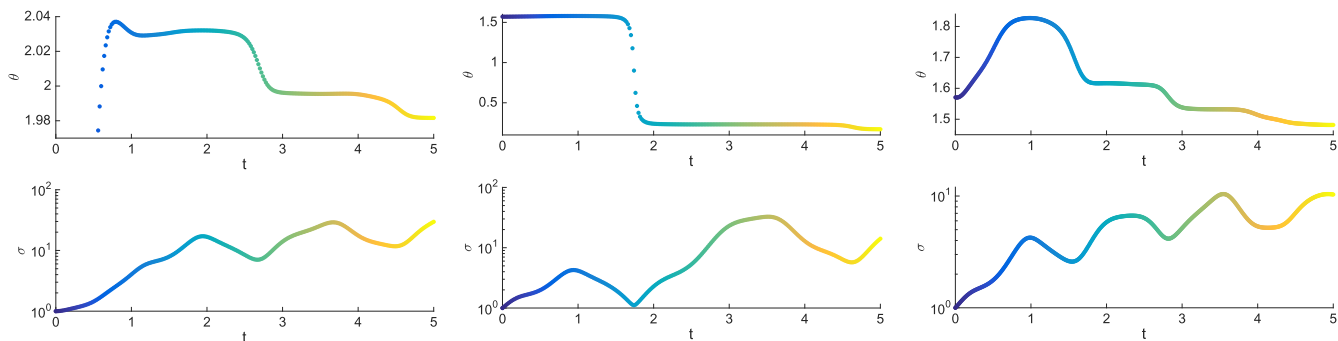

FIG. 8. Top row: orientation $\theta(t)$ of the first singular vector of the fundamental matrix $W(0, t)$ of the double-gyre system for $t \in [0, 5]$ for different initial conditions: $x = x_{1,0}, x_{5,0}, x_{6,0}$, from left to right. Bottom row: time evolution of the associated singular value $\sigma(t)$ on a logarithmic scale. The color corresponds to the time t (blue: initial, yellow: final).
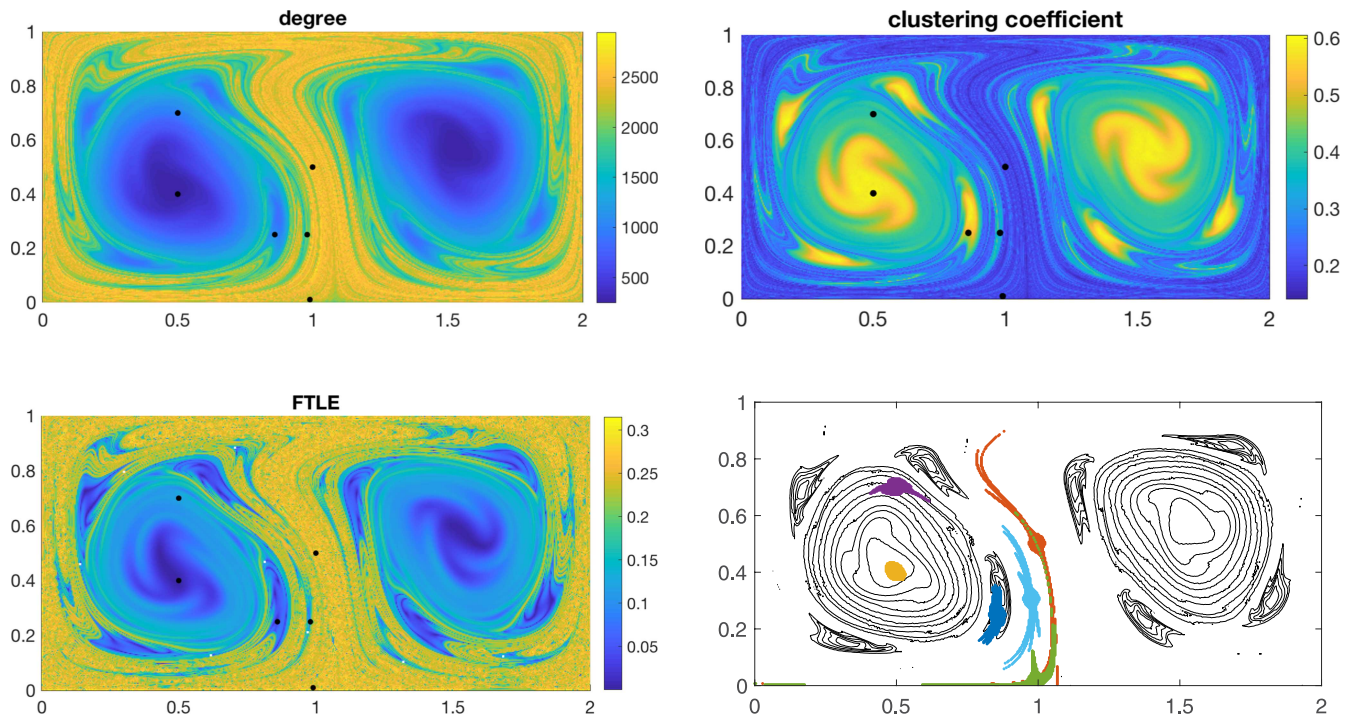
FIG. 9. Degree (top left), clustering coefficient (top right), and FTLE field (bottom left) of the double gyre, with markers for the different initial conditions used in Fig. 7. For the degree and clustering coefficient, $\varepsilon = 0.03$ and flow time $T = 20$ with $\Delta t = 0.1$ was used. Bottom right: The initial conditions of all trajectories that come ε-close to one of the x_i, $i = 1, \ldots, 6$ for $T = 5$, i.e., $\mathbb{G}_\varepsilon(x_i)$ (the shorter time span was taken for better visual comparability).

relation between FTLE and degree, predicted by Proposition 3 in an idealized (linearized) setting, holds only up to a substantial spread in the values (see Fig. 10). Some quantitative agreement is clearly visible though, and the correlation coefficient between them is 0.95

for a smoothed FTLE field.[35] The clustering coefficient and FTLE are negatively correlated with a correlation coefficient of -0.93, indicating that the region with chaotic dynamics is already in a filamenting regime, but did not yet reach well-mixedness; cf. Table I. In this
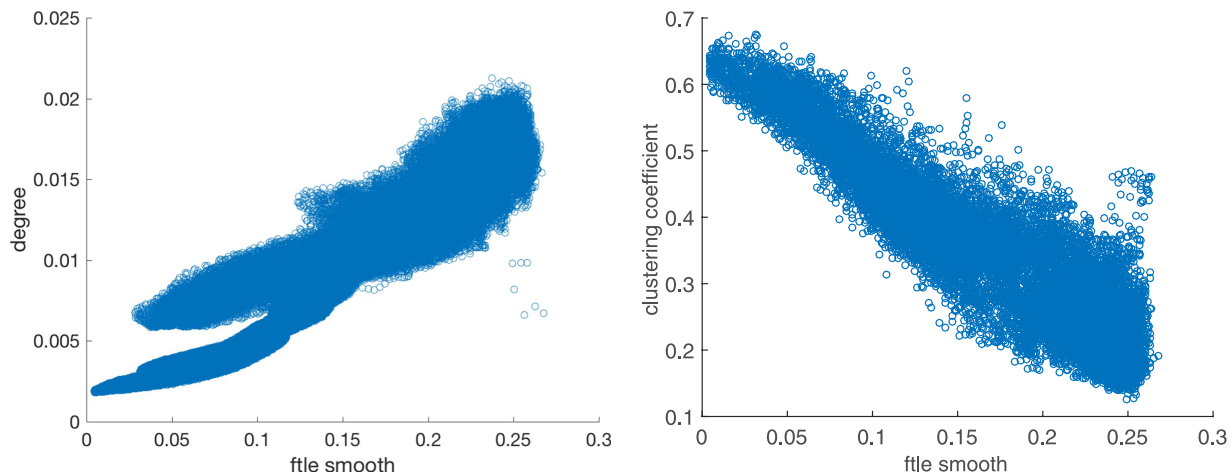
FIG. 10. Top: smoothed FTLE field vs the relative degree (degree normalized by n). The correlation coefficient is 0.95. Bottom: smoothed FTLE field vs the local clustering coefficient. The correlation coefficient is -0.93.
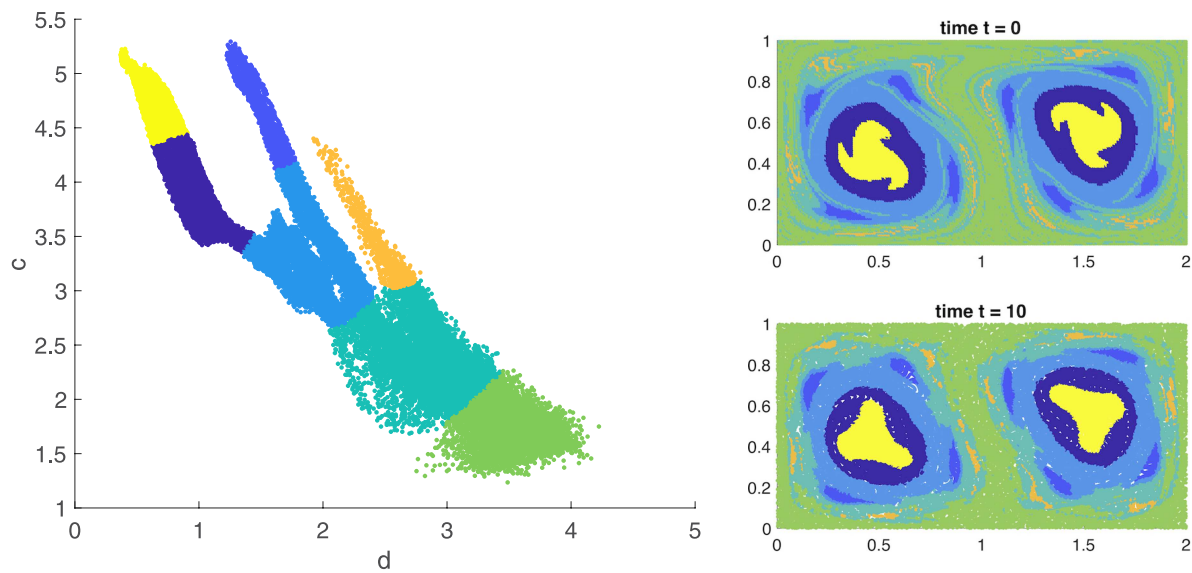
FIG. 11. Left: When the degree d of trajectories is plotted against the clustering coefficient c, one can see that they are not perfectly correlated: The points with large c fall into three branches. The colors are a classification produced with diffusion maps and subsequent clustering by k-means.[24] Right: Same classes as on the left, trajectories plotted in state space at initial time $t = 0$ and halftime $t = 10$.

example, we do not consider the measure "closeness," as it does not contribute additional insights.

Next, we investigate which dynamical structures can be identified and distinguished from another by looking at the two network measures, degree and clustering coefficient, "simultaneously." To this end, we consider the point cloud

$$\mathbb{E} = \left\{ (\bar{d}_i, \bar{c}_i) \,\middle|\, i = 1, \ldots, n \right\} \subset \mathbb{R}^2, \tag{19}$$

where $\bar{d}$ and $\bar{c}$ are the degree and clustering coefficient normalized by their respective standard deviations such that they cover a comparable numerical range and analyze this set by the established manifold learning tool, the so-called "diffusion maps"[5] together with clustering. Diffusion maps, in a nutshell, find intrinsic coordinates on a point-cloud approximation of a manifold such that these coordinates are monotonic in the geodesic distance along the manifold. As an effect, if the point cloud has a complicated topology in its original space, the diffusion-map coordinates tend to "disentangle" it, and clustering in this new space reveals regions of the point cloud that are close-by with respect to the "intrinsic geodesic" distance of this set. This is shown in Fig. 11 for proximity parameter $\epsilon = 0.01$ in the diffusion maps algorithm and clustering its seven dominant eigenvectors into seven clusters.

It is interesting that the classification of trajectories by their proximity in the set $\mathbb{E}$ (with respect to its geodesic distance) can be connected to different qualitative dynamical behavior. This is shown in Fig. 12, where the classification is compared with a "Poincaré plot" of the double-gyre system. The classification separates regular regions in the gyre core, KAM tori, and the chaotic region around them. Also, the invariant "inner" and "outer" gyre cores are

distinguished. Although they all consist of trajectories evolving on invariant cycles, the outer cycles are longer. On the one hand, since trajectories on close-by but different cycles do not keep in phase, the longer cycles have a larger ε-neighborhood, thus a larger degree. On the other hand, on intermediate time intervals (like ours, with $T = 20$), this means a smaller clustering coefficient.

To gain some additional intuition of the structure of the network and its temporal change, we visualize a part of the adjacency matrix A for the end times $T = 5, 10, 20$. We do not show the entire matrix due to its size (see Fig. 13).
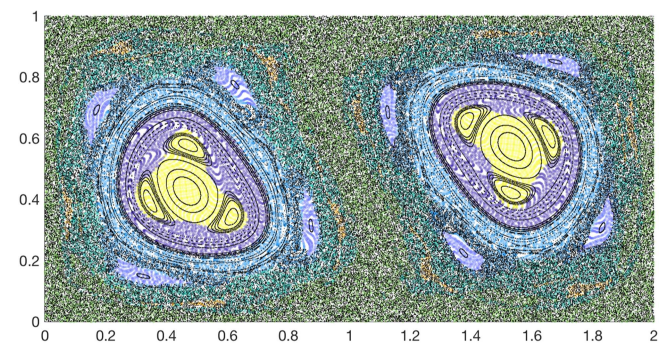
FIG. 12. The classification from Fig. 11 with trajectory positions at halftime $t = 10$ superimposed with a Poincaré plot. The classification identifies gyre cores, KAM tori, and the chaotic region.
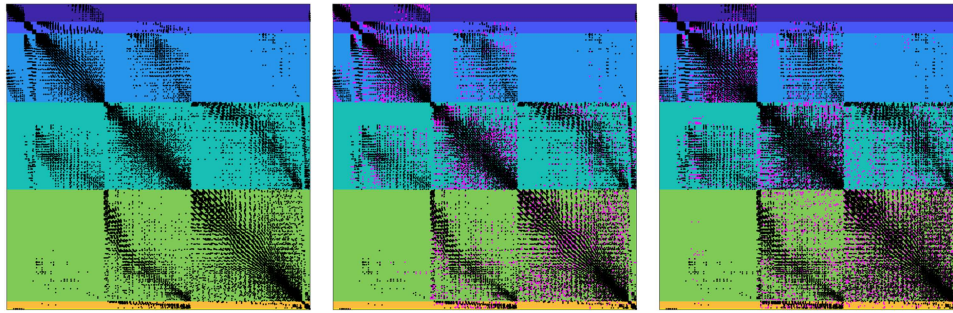
FIG. 13. Occupation structure (nonzero entries are show by magenta and black dots) of a 1000 × 1000 submatrix of the adjacency matrix A for final times $T = 5, 10, 20$, from left to right. The background color of a row is according to the clustering of $\mathbb{E}$, and the same as in the previous figures. The entries that are new compared with the previous final-time matrix are shown in magenta. We note that the largest growth of neighbors for increasing final time T is experienced by the nodes in the regions mixing most strongly, i.e., green and teal.
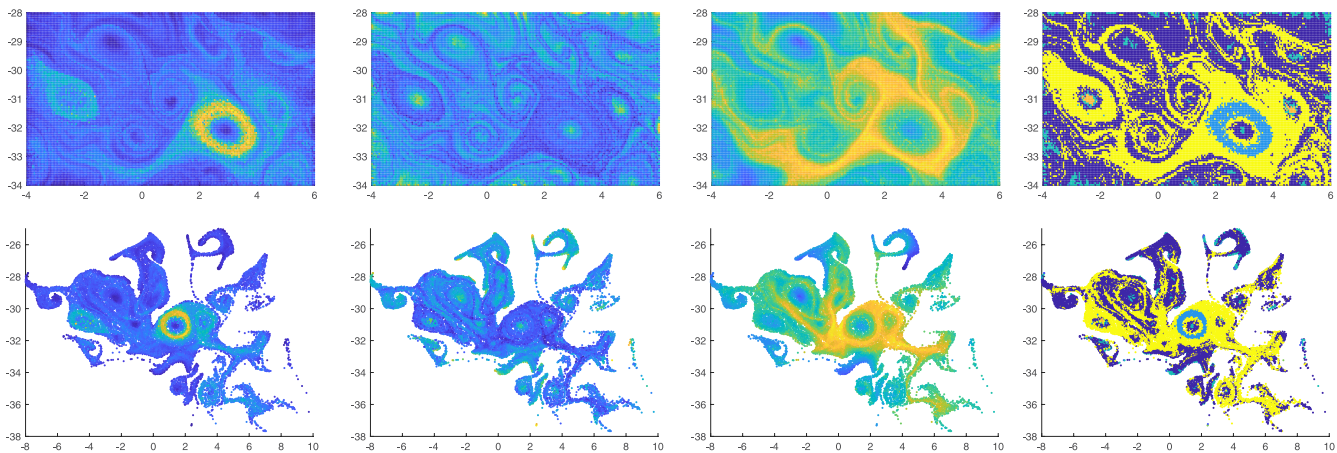
C. Ocean flow

As a last experimental case, we will now analyze an actual ocean flow. We consider a velocity field of the surface water derived from AVISO satellite altimetry measurements. The flow is area-preserving on a spherical surface. We focus on the region of the Agulhas leakage in the South Atlantic Ocean, using the same data set as Refs. 13 and 18. We initialize a 200 × 100 array of drifters in $[-4, 6] \times [-34, -28]$ advected by the flow and observed at times $t = 0, 1, \ldots, 90$, where t_0 corresponds to November 11, 2006.[13,18] With $\varepsilon = 0.1$, we calculate the adjacency matrix and the so far discussed network measures, then compute a classification of the trajectories based on degree and clustering coefficient, as done in the previous example (i.e., we perform a clustering of the diffusion-map embedding of the degree-clustering coefficient point cloud). The results are shown and described in Fig. 14.

First, we observe that the highest degree attained—in contrast to the double-gyre flow—is on the outer perimeter of an eddy. This is underlined by closeness, as the neighborhood of this eddy seems to be some sort of "hub" for transport; many trajectories from different regions pass by this eddy. Second, we also observe trajectories of high clustering coefficient value near the (time-evolving) boundary of the region of consideration. This is due to the dynamics-induced filamentation; subsets of trajectories are separated from the "main region" and build islands or peninsula that do not return to an ε-proximity of other trajectories. Thus, this subset maintains a low degree and high internal connectivity, giving a large clustering coefficient value.

Both of these behaviors arise, because this flow, in contrast to the previous ones, is considered on a "free domain", meaning that the state space (the region we have trajectory data from) evolves with the flow. Thus, we are not taking dynamical information into account from the neighborhood of our set of trajectories. However, this neighborhood interacts with our observations, as the region where we have trajectory information starts to mix with the white region, where

FIG. 14. Ocean flow data set. From left to right: degree, clustering coefficient, closeness, classification. Top: initial time, bottom: final time. For the first three columns lighter yellow colors indicate higher values, while darker blue indicate lower ones. In the rightmost column, the colors indicate seven clusters.

we do not have any. In summary, this "free domain" situation pollutes our analysis with spurious structures. The derivation of sensible network measures that account for this dynamical situation is the next challenge on the way to being able to apply these methods in real-world situations.

VI. CONCLUSION

We have studied an unweighted and undirected trajectory-based network.[37] Simply computable network measures allow us to infer valuable information about the dynamics of the underlying system,[11] even if a full global analysis of the system is out of reach—due to the dimensionality of the system or because only a finite amount of trajectory data is available.

The palette of network measures is broad, both in complexity and computational efficiency. We have focused on simpler ones here and were able to show analytic connections between the local degree of a network and quantitative dynamical descriptors, like FTLE, in the large-data limit.

More complex network measures, such as clustering coefficient and closeness can be linked to qualitative dynamical behavior. We have experimentally verified these connections and have shown how classification with respect to multiple network measures separates regions exhibiting different (topological) dynamical behavior.

The general aim is to identify structurally different dynamical behavior from large sets of possibly high-dimensional trajectory data. Further developments need to be done to understand how consistent estimators of dynamical descriptors can be derived from trajectory networks and how to deal with the "free domain problem" above, and with missing data, in general.

REFERENCES

[1]M. R. Allshouse and T. Peacock, "Lagrangian based methods for coherent structure detection," Chaos **25**(9), 097617 (2015).
[2]M. R. Allshouse and J.-L. Thiffeault, "Detecting coherent structures using braids," Physica D **241**(2), 95–105 (2012).
[3]R. Banisch and P. Koltai, "Understanding the geometry of transport: Diffusion maps for Lagrangian trajectory data unravel coherent sets," Chaos **27**, 035804 (2017).
[4]M. Budišić and I. Mezić, "Geometry of the ergodic quotient reveals coherent structures in flows," Phys. D Nonlinear Phenom. **241**(15), 1255–1269 (2012).
[5]R. R. Coifman and S. Lafon, "Diffusion maps," Appl. Comput. Harmon. Anal. **21**(1), 5–30 (2006).
[6]M. Dellnitz, O. Junge, W. Koon, F. Lekien, M. Lo, J. Marsden, K. Padberg, R. Preis, S. Ross, and B. Thiere, "Transport in dynamical astronomy and multibody problems," Int. J. Bifurcat. Chaos **15**(3), 699–727 (2005).
[7]M. Dellnitz and R. Preis, "Congestion and almost invariant sets in dynamical systems," in *Symbolic and Numerical Scientific Computation (Proceedings of SNSC'01),*

[8]R. V. Donner, M. Lindner, L. Tupikina, and N. Molkenthin, *Characterizing Flows by Complex Network Methods* (Springer International Publishing, Cham, 2019), pp. 197–226.
[9]R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Ambiguities in recurrence-based complex network representations of time series," Phys. Rev. E **81**(1), 015101 (2010).
[10]R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networksa novel paradigm for nonlinear time series analysis," New J. Phys. **12**(3), 033025 (2010).
[11]D. Faranda and S. Vaienti, "Correlation dimension and phase space contraction via extreme value theory," Chaos **28**(4), 041103 (2018).
[12]G. Froyland, "Dynamic isoperimetry and the geometry of Lagrangian coherent structures," Nonlinearity **28**, 3587–3622 (2015).
[13]G. Froyland and O. Junge, "Robust FEM-based extraction of finite-time coherent sets using scattered, sparse, and incomplete trajectories," SIAM J. Appl. Dyn. Syst. **17**(2), 1891–1924 (2018).
[14]G. Froyland and K. Padberg, "Almost-invariant sets and invariant manifolds—Connecting probabilistic and geometric descriptions of coherent structures in flows," Phys. D Nonlinear Phenom. **238**(16), 1507–1523 (2009).
[15]G. Froyland and K. Padberg-Gehle, "Almost-invariant and finite-time coherent sets: Directionality, duration, and diffusion," in *Ergodic Theory, Open Dynamics, and Coherent Structures*, Proceedings in Mathematics and Statistics Vol. 70, edited by W. Bahsoun, C. Bose, and G. Froyland (Springer, 2014), Chap. 9, pp. 171–216.
[16]G. Froyland and K. Padberg-Gehle, "A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data," Chaos **25**(8), 087406 (2015).
[17]A. Hadjighasem, M. Farazmand, D. Blazevski, G. Froyland, and G. Haller, "A critical comparison of Lagrangian methods for coherent structure detection," Chaos **27**(5), 053104 (2017).
[18]A. Hadjighasem and G. Haller, "Level set formulation of two-dimensional Lagrangian vortex detection methods," Chaos **26**(10), 103102 (2016).
[19]A. Hadjighasem, D. Karrasch, H. Teramoto, and G. Haller, "Spectral-clustering approach to Lagrangian vortex detection," Phys. Rev. E **93**, 063107 (2016).
[20]G. Haller, "Lagrangian coherent structures," Annu. Rev. Fluid Mech. **47**(1), 137–162 (2015).
[21]P. Koltai and D. M. Renger, "From large deviations to semidistances of transport and mixing: Coherence analysis for finite Lagrangian data," J. Nonlinear Sci. **28**, 1–43 (2018).
[22]Z. Levnajić and I. Mezić, "Ergodic theory and visualization. I. Mesochronic plots for visualization of ergodic partition and invariant sets," Chaos **20**(3), 033114 (2010).
[23]M. Lindner and R. Donner, "Spatio-temporal organization of dynamics in a two-dimensional periodically driven vortex flow: A Lagrangian flow network perspective," Chaos **27**(3), 035806 (2017).
[24]S. Lloyd, "Least squares quantization in PCM," IEEE Trans. Inf. Theory **28**(2), 129–137 (1982).
[25]A. M. Mancho, S. Wiggins, J. Curbelo, and C. Mendoza, "Lagrangian descriptors: A method for revealing phase space structures of general time dependent dynamical systems," Commun. Nonlinear Sci. Numer. Simul. **18**(12), 3530–3557 (2013).
[26]I. Mezić, S. Loire, V. A. Fonoberov, and P. Hogan, "A new mixing diagnostic and gulf oil spill movement," Science **330**(6003), 486–489 (2010).
[27]M. E. J. Newman, "The structure and function of complex networks," SIAM Rev. **45**(2), 167–256 (2003).
[28]Sometimes "any" solution $\tilde{W}(s,t)$ of (8) for an arbitrary initial condition is called fundamental matrix. In this case, our definition from (6) is recovered by $W(s,t) = \tilde{W}(s,t)\tilde{W}(s,s)^{-1}$.
[29]This follows from the chain rule, from the fact that $\frac{d}{dt}\det(A(t)) = \mathrm{tr}\left(\frac{1}{\det(A(t))}\right.$ $\left. A(t)^{-1}\frac{d}{dt}A(t)\right)$ and that the trace is invariant under similarity transformations.
[30]This statement is made rigorous by considering independently uniformly sampled points, and then their relative ratio in the ε-neighborhood converges to the relative volume of this neighborhood almost surely as the number of samples goes

Lecture Notes in Computer Science Vol. 2630, edited by F. Winkler (Springer, 2003), pp. 183–209.

to infinity. This is the consequence of the law of large numbers, exploited in Monte Carlo methods.

[31] See http://www.wolframalpha.com/input/?i=integrate+sqrt(a%5E 2-x%5E2)+from+x%3D0+to+a%2Fsqrt(2) for V_1. Note that $V_3 = \sigma(T)(\frac{\pi}{4} - V_1)$, since it is congruent with the patch given by the intersection of the quarter disc and A_2 (i.e., in Fig. 2 the part of the blue shaded region under the dashed curve) stretched in the x and contracted in the y direction by $\sigma(T)$, respectively. Since the stretching and contracting are by the same factor, the area of the patch under this transformation is unchanged.

[32] The integral, $\frac{1}{\pi} \int_0^1 4x \arctan \frac{x}{2} - x^2\sqrt{4 - x^2}dx = 1 - \frac{3\sqrt{3}}{4\pi}$, exactly describes the expected relative overlap of two circles $\mathbb{B}_\varepsilon(x)$ and $\mathbb{B}_\varepsilon(y)$, where y is uniformly drawn from $\mathbb{B}_\varepsilon(x)$.

[33] For dynamics that is mixing in the measure-theoretic sense, this is shown, e.g., in footnote 5 of Koltai and Renger.[21]

[34] This semidistance is defined as the shortest path in a time-dependent graph, where the weight of edge (i,j) at the k-th time instance is the squared distance of the trajectories i and j at time t_k, $\|x_i(t_k) - x_j(t_k)\|^2$, while in every step, it is allowed to stay in the same node (self-transitions have zero weight). Thus, this semidistance is short between two trajectories that eventually come close during the time of consideration.

[35] Structures that are on scales of higher order in ε are not captured by the degree anyway, and the linearized relations (12) and (13) disregard errors on higher than first order too; without smoothing the FTLE field, it is 0.88.

[36] K. Padberg, B. Thiere, R. Preis, and M. Dellnitz, "Local expansion concepts for detecting transport barriers in dynamical systems," Commun. Nonlinear Sci. Numer. Simul. **14**(12), 4176–4190 (2009).

[37] K. Padberg-Gehle and C. Schneide, "Network-based study of Lagrangian transport and mixing," Nonlinear Process. Geophys. **24**, 661–671 (2017).

[38] V. Rodriguez-Mendez, E. Ser-Giacomi, and E. Hernandez-Garcia, "Clustering coefficient and periodic orbits in flow networks," Chaos **27**(3), 035803 (2017).

[39] K. L. Schlueter-Kuck and J. O. Dabiri, "Coherent structure coloring: Identification of coherent structures from sparse data using graph theory," J. Fluid Mech. **811**, 468–486 (2017).

[40] E. Ser-Giacomi, V. Rossi, C. Lopez, and E. Hernandez-Garcia, "Flow networks: A characterization of geophysical fluid transport," Chaos **25**(3), 036404 (2015).

[41] S. C. Shadden, F. Lekien, and J. E. Marsden, "Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows," Physica D **212**, 271–304 (2005).

[42] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern. Anal. Mach. Intell. **22**(8), 888–905 (2000).

[43] P. Walters, *An Introduction to Ergodic Theory* (Springer Science & Business Media, 2000), Vol. 79.

[44] M. O. Williams, I. I. Rypina, and C. W. Rowley, "Identifying finite-time coherent sets from limited quantities of Lagrangian data," Chaos **25**(8), 087408 (2015).